

# Distinguishing Biological Trends from Adaptation

Lucas John Mix\*

Models of evolution commonly decompose change into deterministic and stochastic components. Different models, however, produce different concepts of signal and noise. Excursion tests correct for two distinct types of noise, sampling error and phenotypic drift, resulting in two distinct types of signal or trend. A *minimal trend* is a signal of historical change after correcting for sampling error; a *directed trend* is a model dependent measure of confidence that selection has occurred (correcting for both sampling error and drift). The history of random walks and excursion tests in paleobiology highlights the conceptual middle ground between historical pattern and underlying processes. In paleobiology, both minimal and directed trends reflect a causally agnostic, trait-level product of evolutionary and environmental processes. Directed trends provide evidence for directional causes, but the identity and relative strength of those causes requires a deeper understanding. Thus, the magnitude of minimal and directed trends should not be interpreted as the strength of selection for a specified trait. They can be a useful stepping stone in that direction when paired with a map of the middle ground and agreement about how genetic, macroevolutionary, and environmental processes contribute to evolutionary change.


## Keywords

excursion test • drift • MBL model • natural selection • random • trend

## 1 Introduction

There has been great confusion about the meaning of “trend” in biology. Empirical questions come deeply entwined with methodological and philosophical issues that can affect how we move between observed patterns, modeled patterns, and inferences about natural selection (McShea 2000; Sheets and Mitchell 2001; Gregory 2008; Turner 2009, 2015). Biologists regularly contrast “trends” as natural selection with “random” change based on a null model. Null models vary, however, with regard to which processes they include (Bookstein 1987; 2013; Alroy 2000; Millstein 2000; Sheets and Mitchell 2001; Hunt 2006). Thus, both foreground and background vary from study to study. Different random backgrounds result in different foreground

\*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA,  
lucasjmix@gmail.com

 <https://orcid.org/0000-0003-1217-5476>

Received 16 May 2020; Revised 7 December 2021; Accepted 21 December 2021  
doi:10.3998/ptpbio.2104



concepts of “selection” (Kaplan 2013). More broadly, different models serve different research agendas. Two models can correct for bias in different ways, each sufficient for one particular question, but not for another. “The fidelity of a data model must be judged *relative to a particular purpose*” (Bokulich 2018, emphasis in original). A close look at the models used in excursion tests reveals two kinds of trends relevant to paleobiology: minimal trends and directed trends. Both are causally agnostic and cannot be interpreted simply as strength of selection.

Paleobiologists model historical changes in biodiversity as the sum of stochastic and deterministic factors. Such changes result from a variety of causes, often unknown or poorly characterized. The stochastic factor, sometimes called phenotypic drift, includes variation due to genetic drift as well as macroevolutionary processes (e.g., population dynamics) and environmental processes (e.g., mass extinctions). Meanwhile, the deterministic factor, often called a trend, describes directional change in a phenotype over time as selection acts to increase or decrease the frequency of a specified trait. These models are “causally agnostic” because they work in the absence of details about specific processes and their relative contributions. Multiple processes can contribute to a single factor, and individual processes can contribute to both.

In Matthen’s (2009) language, paleobiology is explanandum-oriented. It is focused on the phenomenon which is to be explained (the explanandum) rather than the detailed mechanism of explanation (the explanans). With regard to *evolutionary* questions, the explanandum is a historical pattern of biodiversity while the explanans includes specific causes, such as selection. Paleobiologists (often) model the former before—sometimes in place of—modeling the latter. Excursion tests can reveal trends produced by deterministic selection, but they can also reveal accidental trends, resulting from stochastic variation acting under macroevolutionary and environmental constraints. The stochasticity of phenotypic drift can reflect ontologically probabilistic causes, but it can also reflect uncertainty about the contribution of known deterministic causes which vary across time and space. Here, “trend” refers to the deterministic factor in the model. It can describe a historical pattern without making causal claims.

This paper looks at two types of trends. The first is a signal of historical change that rises above the noise of sampling error. When the deterministic factor is non-zero, there is a *minimal trend*. The second is a signal of selection that rises above the noise of both sampling error and phenotypic drift. When the deterministic factor is large enough to infer a directional cause, there is a *directed trend*. And yet, the directed trend remains causally agnostic. The magnitude of the directed trend measures the strength of the evidence for selection and not the strength of selection for a specified trait. Thus, all directed trends are minimal trends. One describes a real historical change in phenotype frequency. The other describes such a change unlikely to arise without a directional cause for a given model. The semantic issue of which properly deserves the name “trend” distracts from important epistemological issues about how we define and detect each. Both can be useful, and both warrant attention, however we label them.

A closer look at the MBL model and excursion tests helps to reveal a middle ground between the historical pattern and causal factors. Minimal and directed trends can and should be used as part of a larger research project of inferring evolutionary causes. Biologists often seek to map historical patterns of biodiversity (i.e., the sum of stochastic and deterministic factors when describing changes in phenotypic variation through time) onto specific causes via microevolutionary, macroevolutionary, and environmental processes. Many debates about “trends” arise from the difficulty of moving between these two types of explanations. They will not be resolved without consensus about how one maps onto the other. This could be ontological consensus about the nature of biological causation, but it is much more likely to be a pragmatic consensus about how concepts relate across disciplines. Thus, my primary goal is to reveal the ground to be covered.

I start with a brief history of paleobiology, largely drawn from Sepkoski's (2012) *Rereading the Fossil Record* (Section 2). The MBL model and excursion tests were created to utilize historical and ecological data that could not readily be accommodated within population genetic models. This need not entail incompatibility, though some authors would emphasize that possibility. Rather, it identifies the explanatory middle ground that could not be crossed between the fossil record and genetic explanations.<sup>1</sup> Subsections provide greater detail on stochastic modelling of extinction (a middle ground process) and the earliest excursion tests.

Section 3 uses analyses by Lande (1976) and Sober (2008) to identify features of the middle ground. I contrast the phenotypic model of stochasticity with a common genetic drift/selection test to highlight the process/product distinction and discuss implications for understanding the MBL model and excursion tests.

Section 4 looks more closely at how I use “selection” and “drift” in this paper. Sober's (1984) definition of adaptation as selection *for* a trait, in contrast to selection *of* an organism (with the trait) or *of* a gene (for the trait) reveals an important distinction between minimal and directed trends. A minimal trend reflects selection of organisms with the trait—a change in frequency through time—while a directed trend reflects something more—a change due to selection for the trait in question. Millstein's (2000, 2002) discussion of “drift” as process and outcome similarly clarifies this distinction. “Drift” as outcome is always involved in evolution; in a directed trend, it proves insufficient to explain the observed pattern of change.

Section 5 returns to the question of how we define trends and contrasts my minimal/directed distinction with McShea's (1994) passive/driven distinction. I argue that McShea's terms jump straight to the causal question, which obscures the middle ground so important to understanding approaches in paleobiology.

Section 6 states my recommendations for future work on trends across fields of biology.

## 2 The Origin and Intent of Stochastic Models in Paleobiology

David Sepkoski (2012) traces the origins of paleobiology in *Rereading the Fossil Record: The Growth of Paleobiology as an Evolutionary Discipline*. He presents a history in three stages: literal, idealized, and generalized approaches to the fossil record. As a starting point, he looks at “Darwin's Dilemma,” the incompleteness of the fossil record. If evolution really is gradual, why do we not see gradual transitions between species? He then suggests an identity crisis in the late nineteenth and early twentieth century as biology transitioned to quantitative methods, formalized in the Modern Synthesis and population genetic models. This created a challenge for paleontologists, who worked with phenotypic data. Looked down on as qualitative “stamp collectors,” they sought ways to become more quantitative, and thus respectable, scientists. In the 1960s paleontologists begin exploring statistical models appropriate to their phenotypic data sets.

Sepkoski presents Stephen Jay Gould and Niles Eldredge as the key figures of the literal phase. They looked at periods of stasis and rapid change and proposed differing rates of evolution. Eldredge (1971) proposed an alternation of morphological stasis with periods of rapid evolution spurred by allopatric speciation. His proposal was both defended and popularized in Eldredge and Gould (1972) as “punctuated equilibrium.” Gould spoke of rooting paleontology in present-day biology, linking long-term patterns with short-term, ecologically informed,

<sup>1</sup>This ground could not be covered during the early days of paleobiology. I believe it has still not been covered; however, the thesis of the paper remains in either case. The MBL model and excursion tests have causal ignorance built in.

and population-minded adaptation. Both authors were concerned with the impact of ecological and geographic factors, particularly when they resulted in different effects in isolated sub-populations. They each also had a critical approach to theory, arguing that “models are never neutral: *any* model rests implicitly on assumptions that cannot, ultimately, be acquired solely from the data” (Sepkoski 2012, 162). In Sepkoski’s view (following Michael Ruse), Eldredge was cautious both ontologically and epistemologically, while Gould varied from an uncontroversial reflection on evolutionary mode to a strong saltationism (and back again).

In the literal phase, paleobiologists took the fossil record “literally” and sought to fit processes to the observed pattern. In the idealized phase, they attempted to correct for “noise” in the fossil record, resulting from biases in deposition, preservation, and collection of specimens. Models came to include a stochastic background, separating an idealized model of change from the observed record. Sepkoski links this phase to the MBL model, named for a series of meetings at the Marine Biological Laboratory in Woods Hole, Massachusetts. David Raup, Stephen Jay Gould, Thomas Schopf, and Daniel Simberloff (1973) turned to stochastic modeling for a generalized and idealized model of evolution. The collaboration was short lived, but it formalized two important features of paleobiology. First, it involved a null model of random change and, second, it utilized discrete (in Sepkoski’s language, “particulate”) sub-populations or lineages that could be shaped by diverse external forces. This allowed for a new distinction to be made—within-lineage versus between-lineage processes—and contributed to a differentiation of microevolutionary and macroevolutionary descriptions.

Building on this foundation, Raup and Gould began to argue that some predictions could be made based solely on the null model, stochastic change which occurred even in the absence of named directional processes such as selection (Sepkoski 2012, 242; Raup and Gould 1974). Invoking equilibrium models in genetics, Raup et al. (1973) demonstrated that models of random speciation and extinction could produce phylogenies similar to those observed in nature. They emphasized that “randomness does not imply disorder, contrary to vernacular usage. A high degree of apparent order can arise in purely stochastic systems” (526–27). Random processes are neither completely predictable nor utterly unpredictable. They are partially predictable. Some predictions can be made, usually about changes in summary statistics, such as minimum, maximum, and mean values for a given trait. Species level effects and environmental factors contribute to such models as well as genetic drift. Raup et al. included natural selection in each generation but allowed for the direction of selection to change stochastically on larger timescales. All of these processes (speciation, extinction, environmental variation, mutation, drift, ...) may contribute to statistically regular behavior, even if they cannot be, in practice, characterized independently. Raup et al. hoped that such “random” models would enable researchers to identify non-random factors when looking at data.

Raup (1977) went on to use a random walk as a null model against which to test long-term trends. If an observed pattern of consistent change is inconsistent with the random walk, he argued, it must be the effect of selection. The random walk reflects causal ignorance and not a specific model of genetic drift, speciation, or other processes. Raup lays out his null model as a “pseudorandom” process: “We either cannot know or do not choose to know the actual deterministic basis of the process we are dealing with” (61). He highlights extinction as a product of deterministic factors acting at the level of individual organisms. It may, nonetheless, be modelled, above the species level, as a stochastic process. Causal ignorance need not reflect ignorance about which causes are involved. In many cases, the relevant processes can be named,

but the researcher claims no knowledge of their relative contributions.<sup>2</sup> For Raup, stochasticity describes the volatility of effects.

The MBL model showed that stochastic models could be effective at describing a collection of processes whose individual contributions remained unknown. It also showed that stochastic variation can result in apparently deterministic changes. It was, however, too simple and idealized. Starting in the 1980s, paleobiologists began to look more closely at which biasing processes could be identified and isolated. This included geological and anthropogenic factors that can bias the data set (sampling error) as well as macroevolutionary and environmental processes that can obscure the signal of selection.

Sepkoski describes this third phase of paleobiology as generalized, making use of the stochastic models, but not taking for granted either the completeness of the fossil record nor the ideal and particulate quality of lineages.<sup>3</sup>

Bookstein (1987) explores stochastic modeling in great detail, again emphasizing causal ignorance. He uses great care in delineating a null model and a null hypothesis. The null model is a stochastic model of change through time. The null hypothesis includes the null model alongside a second stochastic process representing sampling error. If stochastic change and stochastic error are sufficient to explain the data, then “evolutionary rate” becomes meaningless and no selection can be inferred. For Bookstein, evolutionary rate is a directional vector necessary to complete the model. It is not observed change over time and cannot be found in the data. It must be inferred from the data in conjunction with the null model and a model of how error comes in. The random walk should always be included; if it is a biased random walk, then and only then can the bias be understood as an evolutionary rate.

Bookstein’s (1987, 461) null model includes two kinds of stochastic noise: sampling error and drift. Sampling error disguises the historical signal. If error alone can explain an observed change, no trend exists. The pattern of data may not reflect a historical pattern. If error alone cannot explain the change, a trend exists, which may or may not be directed; that determination requires further analysis. Evolutionary volatility, or “drift” broadly construed, disguises the signal of selection. If error plus phenotypic drift can explain a trend, there is no evidence for selection. The population changed through time, but a directional cause may not have been involved—a *minimal trend*. The magnitude of the trend must be considered. When improbably high, relative to a random walk, it indicates directional selection and a *directed trend*. When the magnitude is improbably low, it indicates stabilizing selection and stasis. Controlling for error, the trend describes an observed pattern. The trend magnitude is always the rate *at which a trait has changed*. When it is improbably high or low (for a random walk), it can also be viewed as the force or strength of selection acting on the trait.<sup>4</sup>

Bookstein (1987, 446) stresses the dangers of interpreting a trend or evolutionary rate naively: “Rate is a property not of the empirical data but of a particular mathematical model for those

---

<sup>2</sup>Eble (1999) discusses multiple uses of “chance” in evolutionary biology. “Causal ignorance” is the first of two statistical notions of chance and may describe situational ignorance (we do not know), fundamental ignorance due to some barrier (we cannot know), or tychism (it cannot be known; the universe is essentially probabilistic). Millstein (2000) usefully explores indiscriminate sampling in concepts of drift (discussed in section 4.2). The causal mechanisms may be fully known, but the outcome remains unpredictable. Raup’s method is open to all these interpretations.

<sup>3</sup>Sepkoski presents the development of paleobiology as dialectical with the generalized approach (synthesis) taking the best of the literal (thesis) and idealized (antithesis) approaches. I find it difficult to see the stages as distinct in this fashion. Instead, I would interpret this as a novel idea whose implications and limitations were slowly worked out over time, wearing away the more extreme and less-functional aspects.

<sup>4</sup>Note that selection of organisms with a trait is not necessarily selection for that trait, but only selection for organisms with the trait, as discussed in section 4.1.

data.” We begin with a ratio—change over time—and infer the rate as a derivative with respect to time. That derivative does not exist in the case of a random walk. The problem is amplified by the recognition that stochastic processes in biology occur stepwise by generation and, as such, must be interval rates and not instantaneous rates (Gingerich 1993). Generation time must be a factor in any null model. The temporal scale of time-intervals and overall change have an impact. For Bookstein, the rate is a directional vector necessary to complete the model. It must be inferred from the data in conjunction with the null model and error model.

To me, it makes sense to speak of an observed rate of change as a genuine historical signal, but I remain agnostic about causes without further information. The semantic issue can be sidestepped by noting that all trends (directed and undirected) are inherently artefacts of the models used to infer them. When those models are stochastic, then the details of the models matter. In the case of directed trends, the rate or magnitude of the trend can vary from the strength of selection in a manner dependent on the variables used.

Bokulich (2018) explores sampling error, or the question of “correcting the data,” in much greater detail. She notes several filters that distort the historical signal. They include taphonomic filters or biases in preservation (e.g., soft bodies leave fewer fossils); ecological biases (e.g., species living in a lake are preserved in the sediment); geological biases (e.g., erosion, heat, and tectonic shifts can erase some groups); and anthropogenic biases (e.g., large bones attract attention). Paleobiologists attempted to correct for these biases in three ways: subsampling approaches, residuals approaches, and phylogenetic approaches. Bokulich discusses all three.

Bokulich argues that all data is, in some sense, model dependent because models have been built into collection strategies and experimental design. “So it is not the ‘purity,’ but rather the *fidelity* of the data that matters. However, it is also important to remember that in assessing fidelity, what counts as signal and what counts as noise depends on the particular uses to which the data set will be put (i.e., what hypotheses the data will be used to provide evidence for or against)” (Bokulich 2018, 2). Her examples focus on sampling error, but the conclusions apply equally to other types of noise found in phylogenetic drift. Causal agnosticism may be unsatisfying, but it is well suited for the minimal and directed trends for which excursion tests were designed. She discusses *vicarious controls*, by which the effects of bias can be removed after data collection through estimating and accounting for them directly or by data collection. Phylogenetic approaches, including variation among lineages, have been particularly important for trends research.

While paleobiologists often move from data to historical pattern to causal inference, this need not be the only approach. Causal assumptions are frequently (if not universally) built into models. These include theories about the number and relative contribution of various processes, about generation time, and about population size and structure but also, critically, about the operation of selection and drift. Signal and noise are inescapably linked and model dependent. If drift is to be a priori excluded as an explanation—as, for example, for large populations—it cannot, then, be a hypothesis rejected by the experiment. If population structure is excluded a priori, the experiment cannot be used to show that population structure is irrelevant. Such theories make up the middle ground between cause and historical pattern.

The earliest paleobiologists sought to detect directed trends amidst a background of stochastic noise, thus a particular understanding of null models is baked into their models. The MBL model and its conceptual descendants describe a historical pattern as the product of multiple processes and remain causally agnostic. They allow for multiple probabilistic processes in addition to random genetic drift, notably species level processes including (but not limited to) speciation, extinction, patterns of migration, and fluctuations in environmental variables. These

background processes all contribute to the stochastic factor of phenotypic change—phenotypic drift.

### 2.1 *Extinction as an Exemplary Environmental Process*

Mass extinctions provide the prime example of environmental processes with deterministic, abiological causes that can, nonetheless, be modeled stochastically. Jablonski (1986) discusses mass extinctions, potentially driven by global climate cycles or periodic meteorite impacts, as macroevolutionary regimes: “The alternation of these macroevolutionary regimes disrupts any smooth extrapolation of microevolutionary or macroevolutionary processes across the sweep of geological time; a complete theory of evolution must incorporate the different sets of selective and random processes that characterize the background and mass extinction regimes” (133). Mass extinctions result in a fluctuating background condition that shapes variation in ways inexplicable without appeal to large-scale, abiological factors.

Over time, biologists have explored less extreme processes and more sophisticated stochastic models for extinction (Ovaskainen and Meerson 2010; Ellner, Childs, and Rees 2016). Leigh (1981) looks at ways a fluctuating environment could drive fluctuations in population size and uses this to predict mean time to extinction for a population. Lande (1993) spells out differences between demographic stochasticity, environmental stochasticity, and catastrophes and how they affect extinction. He notes that stochasticity, here, reflects a summation of lower-level processes, which may be deterministic or probabilistic, rare or common. Foley (1994, 125) continues on this theme, comparing demographic stochasticity to genetic drift and noting that a probabilistic Poisson process models births and deaths, likely due to idiosyncratic deterministic causes. Conversely, Henson et al. (2003) explore the use of low-dimensional deterministic models to predict the outcome of stochastic processes. Benton et al. (2006) review related work on the complex relationship between environmental variation and population dynamics. They note that multiple causal mechanisms are capable of producing the same outcome and argue for the importance of considering multiple models. Looking at empirical examples, they note complex interactions due to environmental fluctuation and diversity: “Therefore, extracting mean demographic parameters using simple statistical modelling of time-series may provide some insight, but is unlikely to lead to full understanding as many of the determinants of individual performance are linked in complex ways to past environments” (1177). Natural selection may not be separable from environmental fluctuation.

### 2.2 *Excursion Tests*

Sepkoski (2012) argues that paleobiologists share a set of approaches rather than a unified theory of evolution. Not all paleobiologists agree on punctuated equilibrium or the correct set of causal processes. Sheets and Mitchell (2001) describe a variety of statistical tests used to infer evolutionary trends. Excursion tests provide one popular approach that models evolution through stochastic and deterministic factors or phenotypic drift and trend.

“Scaling of excursion” tests begin with a random walk null model and calculate the expected excursion over a fixed period. If a variable changes according to an unbiased random walk, then its expected value will remain the same after any interval. Its final value has a Gaussian distribution with mean equal to the initial value and variance proportional to time. McKinney (1990), following Raup and Bookstein, suggests that apparent trends should not be compared to the expected value, but to the expected peak excursion—the point of greatest difference from the initial value. Very large excursions are highly improbable, but so are very small excursions. As a

difference over time has already been observed, researchers should ask whether this is consistent with the maximum excursion expected for a random walk.

Bookstein's (1987) Theorem of the Scaled Maximum locates a directed trend threshold based on a simple random walk. He starts by calculating the range of maximum excursions consistent with a random walk of  $n$  steps with finite variance  $\sigma^2$ . Above this range, the excursion suggests a directed trend, below it, the excursion suggests stasis. Bookstein comments on the low power of the test and advocates for a wider range. For example, to achieve a statistical significance level of 0.1% (or 0.2%, for a two-tailed test), the excursion should fall outside the range of roughly  $0.08\sigma n^{1/2}$  to  $6\sigma n^{1/2}$  (Appendix 2). Gingerich's (1993) Log Rate versus Log Interval (LRI) method measures "rates of change" in haldanes for each pair of time points in the data set.<sup>5</sup> These rates are plotted against intervals on a log-log plot. The resulting distribution has a structure that can be used to estimate the intrinsic rate of a random walk (the distribution of steps in a single generation). It also has a slope that can be compared to the expected slope for a randomly generated time series ( $-0.5$ ). Together they can be used to create a confidence interval for random change. Slopes significantly smaller (closer to 0) indicate stasis while slopes significantly larger (closer to  $-1$ ) indicate directed change. In either case, the null hypothesis of random walk can be rejected. Sheets and Mitchell (2001) also describe a test based on the Hurst exponent, comparing short-term and long-term change in the series. Values outside a threshold range are inconsistent with a random walk null hypothesis. In each case, the parameters of the random walk are generated from available data and agnostic to cause.

### 3 Deterministic and Stochastic Components in Evolution

The MBL model and excursion tests demonstrate a popular approach in evolutionary biology, modeling change as a function of stochastic and deterministic elements. A common drift/selection test displays the ubiquity of the comparison. Random *genetic* drift has often been treated as a background and null model for evolutionary hypotheses. Alternatively, Lande (1976) and Sober (2008) discuss random *phenotypic* "drift" as a background for phenotypic selection. Differences among the three approaches reveal more about the middle ground between historical pattern and causal factors, specifically divides between genetic and phenotypic accounting and between process and product. All three mathematical formalisms (the drift/selection test, Lande's model, and Sober's model) include stochastic backgrounds, but they define them differently and, thus, foreground different types of selection. Sober's model provides a generalized understanding of assumptions built into excursion tests.

A common test in genetics contrasts natural selection and random genetic drift, comparing the strength of two processes. Kimura (1968) and King and Jukes (1969) reported unexpectedly high levels of neutral mutation. Both provided formal stochastic descriptions of non-selective "random" change in genes. This genetic volatility came to be viewed as a background or null hypothesis against which to test claims of selection.<sup>6</sup> Drift can dominate when the product of effective population size ( $N_e$ ) and strength of selection ( $s$ ) is less than one half (Graur and Li 2000, 62).

$$N_e|s| < \frac{1}{2}$$

Phenotypic volatility can be treated in a similar fashion. Lande (1976) contrasted selection and random genetic drift, treating both as forces. He described phenotype evolution as an

<sup>5</sup>Here, "rate" simply refers to change in a period of time. Pairs with a rate of zero are discarded as hiding very small rates of change.

<sup>6</sup>Hartl and Clark (1997, 180) and Graur and Li (2000, 63), for example—both standard population genetics textbooks—refer to the neutral theory as a common null hypothesis.



Ornstein-Uhlenbeck process, with a deterministic component presented as selection and a stochastic component presented as random genetic drift. Sober (2008, 192–99) provides a more recent discussion, with useful commentary. For a trait ( $x$ ) evolving in time ( $t$ ),

$$dx_t = a(\theta - x_t)dt + \sigma dB_t.$$

Change in  $x$  as a function of time can be described with two factors. The first, the deterministic component, depends on how far a trait is from the optimal phenotype ( $\theta$ ), and the change in selection per unit difference ( $a$ ). The second, the stochastic component reflects a Wiener process, a continuous-time stochastic process frequently used by physicists to describe Brownian motion. Here,  $dB_t$  is “a vector of independent and identically distributed normal random variables.” The variance ( $\sigma$ ) is the volatility, or magnitude of random fluctuations. Thus, the stochastic component describes diffusion, the “noise” of random genetic drift. In the absence of selection ( $a = 0$ ), the deterministic term goes to zero, leaving pure drift. Without volatility ( $\sigma = 0$ ), the stochastic term goes to zero, leaving pure selection.

Sober (2008) adds three valuable comments. First, he presents the deterministic component as the *effect* of selection, not the *force* of selection. It is a product and not a process. Second, he notes that evolution in the real world always occurs in finite populations; pure selection does not occur. Stochastic change is the background on which the hypothesis (selection) and the null hypothesis (no selection) can be judged. In Sober’s terms, the appropriate comparison is between pure drift and drift plus selection. Third, he speaks of a “purely phenotypic notion of drift” (193). The underlying cause is random genetic drift, but the product is a distinct stochastic process.

In an earlier discussion of evolution, Sober (1984, 37–38) describes a conceptual gap when causes are described in terms of phenotypes while effects are described in terms of genotype (allele frequency). This gap necessitates a transformation from phenotype space to genotype space. Transformation is, likewise, necessary when moving in the opposite direction. When causes are conceived at the gene level (random genetic drift) and effects at the trait level (patterns of phenotypic change), an inverse mapping will be necessary.

The common test described by Graur and Li (2000) distinguishes two causal processes occurring at the gene level. Lande (1976) attempts to apply the same distinction to processes at the trait level. Sober (2008) does similar work but pays close attention to the gap between gene level and trait level explanation and the gap between describing process and describing product. Sober focuses on the trait level product. Both deterministic and stochastic components in his equation describe a single outcome. They need not align with deterministic and stochastic causes.

Excursion tests rely on a model of evolution with two components. The stochastic component resembles the phenotypic drift of Lande and Sober, whose models are not so much different as more general.<sup>7</sup> Paleobiologists observe trait distributions across a time series and construct a model that describes phenotypic volatility as a cause agnostic random walk. This phenotypic volatility must be distinguished from sampling error. Both are stochastic and normally distributed, but one arises historically from evolutionary causes while the other arises artificially during data collection and interpretation. Correcting for sampling error *but momentarily ignoring phenotypic volatility*, a minimal trend reflects observed consistent change in a trait over time. The rate of change or trend magnitude can then be compared to the range of maximum excursions consistent with the random walk. When the rate is within that range, the

<sup>7</sup>The continuous-time Wiener process can be viewed as the limit of the discrete-time random walk as time interval approaches zero.

stochastic component is sufficient to describe both evolution and trend. If the rate is too high or too low, a deterministic component is added to the model, reflecting the effects of selection driving or inhibiting change. Thus, the implicit model of evolution is the same as in Sober. When the rate is low, the deterministic term  $[a(\theta - x_t)dt]$  describes selection. The farther the trait is from the optimum ( $\theta$ ) the stronger the deterministic effect. When the rate is consistent with a random walk,  $a = 0$  and the deterministic term goes away.

Excursion tests reveal why the trend or deterministic component must be a product and not a process (or cause) in paleobiology. The stochastic component contributes to observed change in all three modes of evolution.<sup>8</sup> In the random walk mode, the stochastic component fully describes the system, including any trend. Therefore, the stochastic component will affect the trend for stasis and directed trends as well. Even when sampling error can be eliminated, evolutionary volatility ensures causal ignorance. The relative contributions of selection and “drift” cannot be disentangled using these methods. Pure drift can only be distinguished from selection plus drift. Deviation of the rate from a random walk can suggest the extent to which selection drives. And yet, when selection plays a role, the exact contribution of drift still cannot be determined. The deterministic component of such models will always come with a margin of error based on the expected excursion of a random walk. In this sense, directed trends (inferred from excursion tests) always come with the limitations of stochastic modeling. The relative contribution of multiple causes remains unknown and the concept is inherently stochastic.<sup>9</sup>

## 4 Selection and Drift as Cause and as Effect

The limits of causal inference when dealing with trends have been noted in numerous contexts. In molecular phylogenetics, Uyeda, Zenil-Ferguson, and Pennell (2018, abstract) note that the “field has, at times, been sloppy when weighing evidence in support of causal hypotheses.” In paleontology, Hunt (2006, 581) addresses random walks: “Because we usually do not know the detailed mechanisms mediating each phenotypic change in the past, we limit ourselves to the more modest but attainable goal of inferring something about the aggregate qualities of a set of evolutionary changes, i.e., their directionality and volatility.” In philosophy of biology, McConwell and Currie (2017) emphasize the significance of identifying—or not identifying—causes when discussing contingency. Causal conclusions may be desired, but we must think carefully about terms like selection and drift. Some models make causal claims when applying these terms. Other models speak of them as effects or products.

### 4.1 Adaptation

Elliott Sober (1984, 208) links adaptation with the causal question of selection *for* a specified trait:

*A* is an adaptation for task *T* in population *P* if and only if *A* became prevalent in *P* because there was selection for *A*, where the selective advantage of *A* was due to the fact that *A* helped perform task *T*.

The deterministic component of directed trends only requires that organisms with trait *A* increased in frequency within the population. There has been selection of those organisms (and

<sup>8</sup>A rate of change exactly equal to zero is extremely unlikely.

<sup>9</sup>Returning to the discussion of Eble (1999) and Millstein (2000) in note 1, this does not entail fundamental ignorance or tychism. Hidden variables may exist, and other methods may reveal the relative contributions of selection and drift. Finite populations and indiscriminate sampling demonstrate, however, that it may be impossible to fully disentangle causes even when the relevant processes are known and physically deterministic.

of the relevant genes). Independent evidence would be required to conclude that this occurred because *A* helped to perform a specific task. The trait may have become prevalent due to random fluctuations at the gene level (e.g., neutral mutation, indiscriminate gene segregation in gamete formation) or at the individual level (e.g., indiscriminate mate sampling) or at the species level (e.g., indiscriminate sampling of populations due to extinction).<sup>10</sup>

A trait may also increase due to indirect forms of selection. Selection *of* genes and organisms with a trait need not entail selection *for* that trait. Selection *of* a gene may occur because it is linked to a gene that codes for an adaptive trait. Similarly, selection *of* a subpopulation may occur because it lives within (or closely associated with) a species bearing an adaptive trait. Thus, selection as effect need not have the same magnitude as adaptation or selection as cause. Both minimal and directed trends describe effects or outcomes; therefore, they need not have the same magnitude as adaptation for the trait in question. Minimal trends can occur without any selection. Directed trends reflect adaptation amplified or dampened by stochastic processes.

#### 4.2 Drift

Millstein (2002, 38) highlights the distinction between process and outcome in discussing drift.

In distinguishing a process from its outcome, I mean to distinguish the kinds of changes that occur over time (the process) from the ‘ending’ state that occurs at one point in time (the outcome). Of course, in a population undergoing evolution, designation of an ending state (or a beginning state, for that matter) at any particular point in time is arbitrary.

She argues that drift is best understood as the process, not the outcome. Random drift describes numerous kinds of indiscriminate sampling. Millstein (2002) considers seven stochastic processes, each of which involves an unbiased sampling of genes: parent sampling, gene segregation (into gametes), gamete sampling, population bottlenecks (including founder effects), rate fluctuations in evolutionary processes (mutation, migration, and selection), and autonomous indeterminacy (as in Brandon and Carson 1996).<sup>11</sup> She notes that different authors include different members of this list in “drift,” leading to confusion about the term and, by extension, about selection as the non-random alternative.

The phenotypic drift described in this paper reflects the drift-as-outcome, contrary to Millstein’s preference for drift-as-process. Nonetheless, Millstein’s analysis reveals the important distinction between the two approaches. Once again, the question of privilege—which one deserves the name drift—is less important than understanding the work each is intended to do.

Matthen (2009, 486) takes a very different approach: “Obviously, their view is radically at odds with mine, for they view drift as physical, whereas I take it to be an artifact of the theoretician’s acts of statistical abstraction.” Matthen, however, glosses over several critical points. He usefully notes the significance of statistical abstraction in explanations and that “chance” can be used to refer to intentional ignorance of certain factors. He does not address the possibility

<sup>10</sup>Volatility occurs at every level. The strength of effects due to certain processes, such as neutral selection, deserves independent discussion.

<sup>11</sup>Brandon and Carson (1996) note that drift can be a *necessary* result of sampling in some populations, including many of large size, due to the numerical impossibility of maintaining gene frequency when population size shifts. Alleles and loci are both counted with natural numbers (1, 2, 3). Any allele frequency across 1,000 loci must change when shifted to a frequency across 999 loci. Finite populations, regardless of size, can experience random shifts. “Put another way, when we are concerned with finite populations, and all real biological populations are obviously finite, it is legitimate to consider drift without selection but not vice versa” (325).

that different evolutionary theories may use statistical abstraction in different ways. Throughout the paper, he refers specifically to genetic drift, consistently equates drift-as-process with drift-as-cause, and explicitly defines drift-as-cause and selection-as-cause as mutually exclusive causal claims. Millstein's (2002) examples make it clear she is thinking more broadly in a way that can (though need not) include phenotypic drift. Her approach is broader descriptively and humbler in its ontological claims. Even if we accept Matthen's conclusions with respect to population genetic evolutionary models (which I would be reluctant to do), they simply highlight the profound differences between population genetic models and paleobiological models. The latter describe outcomes probabilistically, rather than describing probabilistic causes.

Kaplan (2013) reflects further on statistical abstraction in evolutionary theory, suggesting general criteria for judging between abstractions, but makes no concrete proposal. Currently, paleobiological models cannot be reduced to population genetic causal theories, largely due to competing abstractions and competing bridge theories for how to cover the explanatory middle ground.

Millstein (2000) looks more closely at the role of chance in macroevolutionary processes and the relationship between random genetic drift and a more generalized phenotypic "drift." She argues in both articles that uncritical dualism, wherein selection and drift are treated as mutually exclusive causes of evolutionary change, leads to confusion. Indiscriminate sampling always occurs, and the outcome of selection may be indistinguishable from the outcome of drift. Trends, both minimal and directed, describe outcomes. Indiscriminate sampling or "drift" contributes to both. Discriminate sampling, that is selection *for* traits, may contribute as well, even when not detected. Observers of outcomes frequently lack sufficient information to overcome causal ignorance.

Sober and Millstein identify a key middle ground between the clearly defined processes of gene-level accounting in population genetics and the clearly defined products of trait-level accounting in paleobiology. Although the stochastic/deterministic distinction is made in both and the drift/selection distinction is made in both, the different fields do not make the distinctions in the same way.

## 5 What is a Trend?

There is consensus around biological trends at a basic level. There exist biological populations in which a summary statistic (representing some observable trait) changes consistently over a long period of time. McKinney (1990) proposed a minimal definition of trend along these lines, and several recent authors have followed his lead (e.g., McShea 2005; Gregory 2008). It captures certain desirable characteristics that were contentious historically. Evolutionary trends should describe changes at the population level, not within individual lineages. They include changes in minima, maxima, and averages across multiple lineages. They do not include short-term fluctuations and statistical outliers. The minimal trend described in this paper is the same as the trend defined by McKinney.

Even at this minimal level, it is important to ask two clarifying questions. First, what kind of population counts? And second, what is the appropriate time frame? Answers to both questions will be built into models of phenotypic drift, either explicitly or implicitly. Only when explicit can they be used to move between gene-level and trait-level approaches.

In population genetics, a local or breeding population refers to a collection of individuals capable of mating—or in an asexual population, directly competing in some meaningful way. It may be a sub-population or deme, within a larger meta-population. A breeding population appears to be too strict a requirement for paleontological work, however. A collection of

fossils—even fossils of homologous traits in closely related organisms—may represent a breeding population, a non-representative sub-sample of a breeding population, or a collection of breeding populations with various levels of mixing. In a minimal trend, “population” should be understood in the most generic mathematical sense: it is a set of biological objects from which we collect data to generate statistics. Mapping that population onto a set of breeding populations requires significant work, as displayed for example by extinction models discussed in section 2.1.

How long is long enough for a trend to be worthy of attention? Both phenotypic and genotypic patterns of change can have structure across a broad range of temporal scales, from one generation to the history of life on Earth. Evolutionary trends should be, at some minimal level, consistent across scales and should persist for a considerable period of time. Within a breeding population of fixed size, we might exclude random “trends” below a certain threshold, proportional to effective population size and strength of selection. It is unclear how such a threshold should be determined for alternative populations. Palstra and Fraser (2012), Gilbert and Whitlock (2015), and Ryman et al. (2019) discuss one aspect of this problem, the challenges involved in estimating effective population accurately for structured populations. Identification of phylogenetic and temporal range will play a crucial role.

Some authors have proposed stricter definitions for “trend”—they add additional requirements to the simple pattern of historical change involved in a minimal trend. These additional requirements have been extremely controversial. Many biologists implicitly assume that selection must be involved; however, as we have seen above, the discovery of minimal trends can be an important step in the identification of directed trends. And, more significantly for the history of paleobiology, a minimal trend may be detected, even when the underlying processes and their relative contributions remain obscure. A minimal trend represents genuine knowledge, whether or not we call it a “trend.”

### 5.1 *Driven and Passive Trends*

McShea’s (1994) distinction between driven and passive trends has been popular and raises an important issue—differential causation—well summarized in the abstract.

In a driven trend, the distribution mean increases on account of a force (which may manifest itself as a bias in the direction of change) that acts on lineages throughout the space in which diversification occurs. In a passive system, no pervasive force or bias exists, but the mean increases because change in one direction is blocked by a boundary, or other inhomogeneity, in some limited region of the space. (1747)

He credits paleobiologists Steven Stanley (1973) and Daniel Fisher (1986) with introducing and refining the concept of passive trends. McShea’s language is clearly causal, even agential. Selection, or some other force, drives evolution in a particular direction. This cause is active in driven trends, but not in passive trends.

Alroy (2000) and Gregory (2008) critique simple interpretations of this as a causal claim. The driven/passive distinction does not correspond to random and non-random processes: “The real question, then, is not whether evolution is random, but instead exactly what mechanisms do govern demonstrable trends, and at what hierarchical level these mechanisms operate” (Alroy 2000, 319). Both authors emphasize the distinction between processes that occur within species and those that occur between species. These might be called micro- and macroevolutionary forces as in McConwell and Currie (2017) or anagenetic and cladogenetic trends as in McKinney (1990). Gregory (2008) refers more generally to a variety of taxonomic and historical

scales. Alroy and Gregory also highlight the ubiquitous influence of both stochastic processes and deterministic drivers. It is difficult to find changes that do not involve both, albeit with different levels of influence.

McShea starts with an ontology of forces and asks which are involved, but this immediately pushes him into the explanatory gap discussed in previous sections. We are often ignorant of which causes are involved and how much they contribute to the outcome. More problematically, known stochastic processes, including drift, are not “passive.” They require energy and drive change. The important distinction, which McShea identifies (as does Millstein; see sec. 4.2) is between biased and unbiased processes, not between active and passive processes or between driven and random processes. By placing his emphasis on passivity, McShea has obscured the importance of the stochastic background and our frequent causal ignorance. In evolutionary biology, environments and populations never intend or steer with an eye toward increasing the frequency of a trait. They do not “drive.” Natural selection is always random or accidental in an Aristotelian sense; it admits of neither cosmic necessity in the environment nor substantial form in the species; it never involves intent or prospect and is always contingent upon circumstances.<sup>12</sup> Multiple internal and external forces act on every population, including drift and low levels of selection. The key distinction is not whether a given sampling process acts or occurs, but whether it is consistently biased over some phylogenetic and temporal span. The emphasis should be on consistency rather than agency.

Causal claims move the discussion from statistical description to causal inference, from a mathematical characterization of data to an attempt to describe history: events and their efficient causes. They require epistemic and ontological commitments regarding sufficiency, necessity, and contingency that have proven problematic in biology. One must ask how strong a claim is being made on this front and what backing (both philosophical and empirical) would be required to justify it. Linguistic quibbles aside, I think McShea’s distinction is important and valuable when achievable. I suspect, however, that the history of paleobiology has been far more involved with situations where the explanatory gap has not been bridged. Nor is this simply a matter of moving from model-free data or a clear causal scheme. As Bokulich (2018) notes, the symbiosis between data and model pervades the process from beginning to end.

## 6 Conclusion

Trends have been invoked throughout the history of modern biology with a wide range of meanings, from the unnatural agency of “Spirit” in Alfred Russell Wallace and the German Idealists to simple reporting on statistical patterns. Most biologists seek something more than post hoc reporting; they want to know about historical and consistent causes in nature. Philosophical reflection can be useful at precisely this juncture, where causal inference moves us from patterns of data to general rules about evolution. Excursion tests provide a concrete example, requiring biological insight in addition to statistical skill. Both minimal trends and directed trends contribute to investigations of natural selection, but neither should be equated with natural selection or, more specifically with the process of adaptation for a particular trait. Instead, a minimal trend should be viewed as a signal of real historical change and a directed trend should be viewed as a measure of confidence that selection occurred. Both are model dependent and cause agnostic and both depend critically on a null model of evolutionary volatility.

In the context of excursion tests in paleobiology, the trend should never be equated with the magnitude of selection for a given trait. In some cases, drift suffices to explain statistical

---

<sup>12</sup>Curiously, this aligns well with Aristotle’s distinction between accidental causation and growth. Growth follows an evolved program or “teleonomy,” but no such program exists for the succession of species.

shifts through time. In other cases, drift amplifies or dampens the signal of selection. These descriptions of the outcome of evolution can be decomposed into deterministic and stochastic components, but they do not exactly mirror the deterministic and stochastic causes of evolutionary change.

Biologists should never take the deterministic/stochastic divide for granted. It occurs in different places for different models and different tests. The common drift/selection test of section 2 works to distinguish causes or processes within a very narrow context where evolutionary causes have been explicitly and exhaustively enumerated (as mutation, migration, selection, and drift) and mutation and migration controlled for.

Moving to paleobiology, phenotypic drift represents a composition of multiple stochastic factors, including random genetic drift (transformed into phenotype space) as well as temporal variation in environment and species level stochastic processes. A Gaussian process in one space need not map to a Gaussian process in another. More problematically, a collection of deterministic processes can be modeled as a stochastic process. Temporal, geographic, and phylogenetic scale all impact estimates of variation and thus, the expected excursion of traits through time. Careful research and reporting will note these factors and how they contribute to null models of evolutionary volatility. A directed trend always entails comparison with such a theoretical background.

Like sampling error, causal ignorance reflects an unavoidable bias which can, nonetheless be minimized when addressed transparently. The causes of evolutionary change can often be listed, and some will be eliminable. For example, independent knowledge of population size and structure may rule out random genetic drift. Or independent knowledge about gene location and expression may rule out pleiotropy, linkage, and associated questions about selection *of* versus selection *for* a trait. These distinctions cannot be taken for granted, however, especially when considering long-term trends that extend over multiple species and environments. Minimal and directed trends remain useful as stepping stones to estimating the strength of selection and the causes of phenotypic change. They open the door to a fuller theory of evolutionary causation and more explicit models for mapping the middle ground between causal stories and historical patterns.

## Acknowledgments

This paper arose out of the 2019 Long-Term Trends in Evolution Workshop at the University of Arizona supported by the John Templeton Foundation. The discussion revealed considerable variety in how different thinkers approach trends. The conclusions above are mine alone, but I am grateful for insights and provocations provided by the participants, especially Joanna Masel, Josef Uyeda, Roberta Millstein, and Daniel McShea. Fred Bookstein, Joseph Felsenstein, and John Wakely read drafts and provided useful comments.

## Literature cited

- Alroy, John. 2000. "Understanding the Dynamics of Trends Within Evolving Lineages." *Paleobiology* 26 (3): 319–29.
- Benton, Tim G., Stewart J. Plainstow, and Tim N. Coulson. 2006. "Complex Population Dynamics and Complex Causation: Devils, Details and Demography." *Proceedings of the Royal Society B: Biological Sciences* 273 (1591): 1173–81.
- Bokulich, Alisa. 2018. "Using Models to Correct Data: Paleodiversity and the Fossil Record." *Synthese* 198 (24): 5919–40.

- Bookstein, Fred L. 1987. "Random Walk and the Existence of Evolutionary Rates." *Paleobiology* 13 (4): 446–64.
- Bookstein, Fred L. 2013. "Random Walk as a Null Model for High-Dimensional Morphometrics of Fossil Series: Geometrical Considerations." *Paleobiology* 39 (1): 52–74.
- Brandon, Robert N., and Scott Carson. 1996. "The Indeterministic Character of Evolutionary Theory: No 'No Hidden Variables Proof' but No Room for Determinism Either." *Philosophy of Science* 63 (3): 315–37.
- Eble, Gunther J. 1999. "On the Dual Nature of Chance in Evolutionary Biology and Paleobiology." *Paleobiology* 25 (1): 75–87.
- Eldredge, Niles. 1971. "The Allopatric Model and Phylogeny in Paleozoic Invertebrates." *Evolution*, 25 (1): 156–167.
- Eldredge, Niles, and Stephen J. Gould. 1972. "Punctuated Equilibria: An Alternative to Phyletic Gradualism." In *Models in Paleobiology*, edited by Thomas J. M. Schopf, 82–115. San Francisco: Freeman, Cooper & Co.
- Ellner, Stephen P., Dylan Z. Childs, and Mark Rees. 2016. *Data-Driven Modelling of Structured Populations: A Practical Guide to the Integral Projection Model*. Cham: Springer.
- Fisher, Daniel C. 1986. "Progress in Organismal Design." In *Patterns and Processes in the History of Life*, edited by David M. Raup and David Jablonski, 99–117. Berlin: Springer.
- Foley, Patrick. 1994. "Predicting Extinction Times from Environmental Stochasticity and Carrying-Capacity." *Conservation Biology* 8 (1): 124–37.
- Futuyma, Douglas J. 1986. *Evolutionary Biology*. Sunderland, MA: Sinauer.
- Gilbert, Kimberly J., and Michael C. Whitlock. 2015. "Evaluating Methods for Estimating Local Effective Population Size with and without Migration." *Evolution* 69 (8): 2154–66.
- Gingerich, Philip D. 1993. "Quantification and Comparison of Evolutionary Rates." *American Journal of Science* 293 (A): 453–78.
- Gregory, R. Ryan. 2008. "Evolutionary Trends." *Evolution: Education and Outreach* 1:259–73.
- Graur, Dan, and Wen Hsiung Li. 2000. *Fundamentals of Molecular Evolution*, 2nd Ed. Sunderland, MA: Sinauer.
- Hartl, Daniel L., and Andrew G. Clark. 1997. *Principles of Population Genetics*, 3rd Ed. Sunderland, MA: Sinauer.
- Henson, Shandelle M., Aaron A. King, R. F. Costantino, J. M. Cushing, Brian Dennis, and Robert A. Desharnais. 2003. "Explaining and Predicting Patterns in Stochastic Population Systems." *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270 (1524): 1549–53.
- Hunt, Gene. 2006. "Fitting and Comparing Models of Phyletic Evolution: Random Walks and Beyond." *Paleobiology* 32 (4): 578–601.
- Jablonski, David. 1986. "Background and Mass Extinctions: The Alternation of Macroevolutionary Regimes." *Science* 231 (4734): 129–33.
- Kaplan, Jonathan M. 2013. "'Relevant Similarity' and the Causes of Biological Evolution: Selection, Fitness, and Statistically Abstractive Explanations." *Biology & Philosophy* 28 (3): 405–21.
- Kimura, Motoo. 1968. "Evolutionary Rate at the Molecular Level." *Nature* 217 (5129): 624–26.
- King, Jack Lester, and Thomas H. Jukes. 1969. "Non-Darwinian Evolution." *Science* 164 (3881): 788–98.



- Lande, Russell. 1976. "Natural Selection and Random Genetic Drift in Phenotypic Evolution." *Evolution* 30 (2): 314–34.
- Lande, Russell. 1993. "Risks of Population Extinction from Demographic and Environmental Stochasticity and Random Catastrophes." *American Naturalist* 142 (6): 911–27.
- Leigh, Egbert G., Jr. 1981. "The Average Lifetime of a Population in a Varying Environment." *Journal of Theoretical Biology* 90 (2): 213–39.
- Matthen, Mohan. 2009. "Drift and 'Statistically Abstractive Explanation.'" *Philosophy of Science* 76 (4): 464–87.
- McConwell, Alison K., and Adrian Currie. 2017. "Gouldian Arguments and the Sources of Contingency." *Biology and Philosophy* 32 (2): 243–61.
- McKinney, Michael L. 1990. "Classifying and Analysing Evolutionary Trends." In *Evolutionary Trends*, edited by Kenneth J. McNamara, 28–58. Tucson: University of Arizona Press.
- McShea, Daniel W. 1994. "Mechanisms of Large-Scale Evolutionary Trends." *Evolution* 48 (6): 1747–63.
- McShea, Daniel W. 2000. "Trends, Tools, and Terminology." *Paleobiology* 26 (3): 330–33.
- Millstein, Roberta L. 2000. "Chance and Macroevolution." *Philosophy of Science* 67 (4): 603–24.
- Millstein, Roberta L. 2002. "Are Random Drift and Natural Selection Conceptually Distinct?" *Biology and Philosophy* 17 (1): 33–53.
- Palstra, Friso P., and Dylan J. Fraser. 2012. "Effective/Census Population Size Ratio Estimation: A Compendium and Appraisal." *Ecology and Evolution* 2 (9): 2357–65.
- Raup, David M., and Stephen J. Gould. 1974. "Stochastic Simulation and Evolution of Morphology—Towards a Nomothetic Paleontology." *Systematic Biology* 23 (3): 305–22.
- Ovaskainen, Otso, and Baruch Meerson. (2010) "Stochastic Models of Population Extinction." *Trends in Ecology & Evolution* 25 (11): 643–52.
- Raup, David M., Stephen Jay Gould, Thomas J. M. Schopf, and Daniel S. Simberloff. 1973. "Stochastic Models of Phylogeny and the Evolution of Diversity." *The Journal of Geology* 81 (5): 525–42.
- Raup, David M. 1977. "Stochastic Models in Evolutionary Paleobiology." In *Patterns of Evolution as Illustrated by the Fossil Record*, edited by Anthony Hallam, 59–78. Amsterdam: Elsevier.
- Ryman, Nils, Linda Laikre, and Ola Hössjer. 2019. "Do Estimates of Contemporary Effective Population Size Tell Us What We Want to Know?" *Molecular Ecology* 28 (8): 1904–18.
- Sepkoski, David. 2012. *Rereading the Fossil Record: The Growth of Paleobiology as an Evolutionary Discipline*. Chicago: University of Chicago Press.
- Sheets, David H., and Charles E. Mitchell. 2001. "Why the Null Matters: Statistical Tests, Random Walks, and Evolution." *Genetica* 112:105–25.
- Sober, Elliott. 2008. *Evidence and Evolution*. New York: Cambridge University Press.
- Sober, Elliott. 2014. *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. Chicago: University of Chicago Press.
- Stanley, Steven M. 1973. "An Explanation for Cope's Rule." *Evolution* 27 (1): 1–26.
- Turner, Derek D. 2009. "How Much Can We Know About the Causes of Evolutionary Trends?" *Biology and Philosophy* 24 (3): 341–57.
- Turner, Derek D. 2015. "Historical Contingency and the Explanation of Evolutionary Trends." In *Explanation in Biology*, edited by Christophe Malaterre and Pierre-Alain Braillard, 73–90. Dordrecht: Springer.

---

Uyeda, Josef C., Rosana Zenil-Ferguson, and Matthew W. Pennell. 2018. "Rethinking Phylogenetic Comparative Methods." *Systematic Biology* 67 (6): 1091–1109.

---

© 2022 Author(s)

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits anyone to download, copy, distribute, display, or adapt the text without asking for permission, provided that the creator(s) are given full credit.

ISSN 2475-3025