

# Modeling: Neutral, Null, and Baseline

William C. Bausman\*†

---

Two strategies for using a model as “null” are distinguished. *Null modeling* evaluates whether a process is causally responsible for a pattern by testing it against a null model. *Baseline modeling* measures the relative significance of various processes responsible for a pattern by detecting deviations from a baseline model. When these strategies are conflated, models are illegitimately privileged as accepted until rejected. I illustrate this using the neutral theory of ecology and draw general lessons from this case. First, scientists cannot draw certain conclusions using null modeling. Second, these conclusions follow using baseline modeling, but doing so requires more evidence.

---

**1. Introduction.** Nitecki and Hoffman begin the volume *Neutral Models in Biology* by saying, “‘Neutral model’ belongs to a whole family of terms, which are sometimes . . . used interchangeably; these are ‘null hypothesis,’ ‘null model,’ ‘random model,’ ‘baseline model,’ ‘stochastic approach,’ ‘neutral theory,’ etc.” (1987, 3). This ambiguity remains today. This article concerns one problematic way that scientists reason when they are using reason strategies associated with this family of terms.

Van Valen identified the same problem in a short scientific correspondence in *Nature*:<sup>1</sup> “In the past decade or so a subtle misuse of null hypotheses has become almost standard in ecology, biogeography, functional mor-

Received May 2017; revised January 2018.

\*To contact the author, please write to: Department of Philosophy, University of Geneva, CH-1211 Geneva, Switzerland; e-mail: wcbausman@gmail.com.

†Versions of this argument were presented at the Eastern APA 2016 in Washington, DC, and PSA 2016 in Atlanta. I thank Adrian Currie, Stephen Hubbell, Andrew Inkpen, Oliver Lean, Shay Logan, Alan Love, James Rosindell, Lauren Ross, Raphael Scholl, and Ken Waters for comments on this manuscript. The research and writing of this article was supported in part by grant 50191 from the John Templeton Foundation, From Biological Practice to Scientific Metaphysics.

1. Thanks to Chris Haufe for showing this letter to me.

Philosophy of Science, 85 (October 2018) pp. 594–616. 0031-8248/2018/8504-0003\$10.00

Copyright 2018 by the Philosophy of Science Association. All rights reserved.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), which permits non-commercial reuse of the work with attribution. For commercial use, contact [journalpermissions@press.uchicago.edu](mailto:journalpermissions@press.uchicago.edu).

phology and theoretical paleontology. . . . The difficulty here comes when a null hypothesis is placed in a privileged position, to be accepted at least provisionally until disproved” (1985, 230). Yet he explained neither why scientists make this reasoning nor why it is problematic. This misuse has surrounded every case of a neutral theory in and around the biological sciences. Hubbell, lead developer of the neutral theory of ecology, stepped right into this morass when called his theory “neutral” (2001) and said it played the role of the “null” (2005, 2006).

In this article, I focus on the case of the neutral theory of ecology and show how Hubbell and other ecologists privilege the neutral model qua “null” as accepted until rejected. To show why this reasoning does not establish its conclusions, and to enable ecologists to state their intended methods, I introduce a distinction between two important methodological uses of models: *null modeling* and *baseline modeling*. When scientists conflate these reasoning strategies, the weaker argument appears the stronger. But when these uses are respected, scientists can reason clearly and make strong conclusions.

*Null modeling* evaluates whether a process is causally responsible for type of pattern by testing it against a null model lacking that process. *Baseline modeling* measures the relative significance of various processes responsible for a token instance of a pattern by detecting deviations from a baseline model. While the difference between evaluating the evidence for a hypothesis and measuring the strength of a process is clear in the abstract, in practice these activities are easily run together. In both reasoning strategies, the two models used are treated asymmetrically. And the same model may be appropriate for use as a null model in one context and as a baseline model in another context. But failing to distinguish these strategies leads to running them together and invalid inferences being drawn. This is especially likely when using a *neutral* model, as continues to happen in community ecology.

Community ecologists investigate questions such as: How many species of tree are there on an island, and why are most of them rare? There is methodological disagreement about how to approach answering these questions, and methodological positions track the theory being used. Ecologists using competition theory look for the differences between species that make a difference to their geography and abundance. They first try to understand observed patterns as the result of competition between differentially adapted species for limited resources. Every species has its own fundamental niche, and each individual works to extend its realized niche at the cost of others. Why is one tree species so much more abundant than another? They first hypothesize that the species is better adapted to the conditions.

Ecologists using neutral theory, in contrast, start by assuming that there are no relevant differences between species and that every individual, regardless of species, is functionally equivalent. They first try to understand

observed patterns as the result of history and chance. They hypothesize that the reason why one tree species is more abundant than another is because it was either more abundant in the past or in the surrounding area.

The methodological issue I address in this article arises when the ecologists invoke the neutral model as a “null” with respect to competition models. As I explain in detail below, because a neutral model is neutral and so leaves out species differences, some ecologists treat it and a competition model including species differences asymmetrically. I claim the users of neutral theory sometimes fail to distinguish when they are null modeling from when they are baseline modeling.

The following argument represents the reasoning, made by various users of neutral theory, that I critique in this article:

1. The neutral theory of ecology supplies the appropriate null model for testing competition theory with respect to a pattern of interest.
2. The neutral model fits the pattern of interest well enough and so fails to be rejected for that pattern.
3. Therefore, history and chance, and not interspecific competition, are the dominant causes of that pattern of interest.

This argument privileges the neutral theory as accepted until rejected. I argue that this argument does not follow because the reasoning strategy that produces it conflates null modeling and baseline modeling. Distinguishing between null and baseline modeling explains why the reasoning fails and how it may arise.

The same faulty reasoning is liable to emerge whenever scientists employ any of the terms mentioned by Nitecki and Hoffman. We should examine cases in which a scientific theory supplies the “null hypothesis” but classical statistics seems to be absent (Bausman and Halina, forthcoming). But this illegitimate privileging is liable to accompany the use of any “neutral theory” in biology. Kimura’s (1983) neutral theory of molecular evolution and the Woods Hole Group’s Marine Biological Laboratory (MBL) model of paleobiology (Raup et al. 1973) are only the most prominent examples of neutral theories. Scientists continue to develop neutral theories for life histories (Steiner and Tuljapurkar 2012), wealth accumulation (Fargione et al. 2011), and language dynamics (Blythe 2012), and the reasoning strategies used in each of these cases should be analyzed.

This article proceeds as follows. Section 2 introduces the history of these methodological issues. Sections 3–5 characterize and distinguish null modeling and baseline modeling. Section 6 introduces the neutral theory of ecology. Section 7 shows how its use has shifted between null modeling and baseline modeling. Section 8 argues for how and why a hybrid reasoning strategy that shifts from null modeling to baseline modeling can privi-

lege a neutral model and why the corresponding argument fails. It fails because not only are null modeling and baseline modeling distinct, but baseline modeling requires more support than null modeling. This support cannot, however, come from the neutrality of the model.

**2. Background.** Null modeling is distinct from statistical null hypothesis testing. The relevant difference for this article is that null modeling tests the hypothesis that a process is causally responsible for a type of pattern, while null hypothesis testing tests the hypothesis that a pattern was due to sampling error, measurement error, or randomness and independent causes. When the null hypothesis cannot be rejected, no more explanation is sought. Null modeling and baseline modeling are properly pursued after the appropriate null hypothesis has been rejected. However, the distinction between null hypothesis testing and null modeling was slow to emerge. And so many critiques of null hypotheses carry over to null modeling.

When Van Valen identified null hypotheses being illegitimately privileged, Strong, Simberloff, and colleagues were attempting to put ecology on sounder scientific footing. Strong, Szyska, and Simberloff base their image of good science on Popper's conjecture-and-refutation conceptual scheme, urging ecologists to adopt falsificationist practices by emphasizing the role of testing null hypotheses and null models: "We propose another possibility with logical primacy over other hypotheses, that other hypotheses must first be tested against. . . . This is the null hypothesis that community characteristics are apparently random" (1979, 910). They meant that, until you exclude the appropriate null hypothesis, the fit between your alternative hypothesis and the data gives no evidence for your hypothesis.

Quinn and Dunham (1983) demonstrate the mistake made by Simberloff and Strong's argument for using null hypotheses. Rather than assigning logical primacy to randomness when testing hypotheses, applied statisticians use null hypotheses to estimate the relative importance of factors not included in the null hypothesis, factors already assumed to be relevant. Sloep (1986) argues that Strong and Simberloff's proposed methodology for testing null hypotheses is mistaken because their null and alternative hypotheses overlap while statistical hypothesis testing requires that they be exclusive. But these critiques have not stuck in biology.

Wimsatt (1987), in his contribution to *Neutral Models in Biology*, distinguishes 12 different uses of "false" models, of which he takes "neutral models" to be an instance. Some of these uses are similar to how I characterize null modeling and baseline modeling.<sup>2</sup> But Wimsatt is interested in neither

2. Null modeling is closest to Wimsatt's use 7, and baseline modeling is closest to uses 4 and 5.

explaining these reasoning strategies in detail nor cases in which they are mixed together to ill effect. He also continues the ambiguous terminology of the volume, saying “Neutral models in biology represent ‘baseline models’ or ‘null hypotheses’ for testing the . . . efficacy of selection processes by trying to estimate what would happen in their absence” (52).

Beatty (1987) critiques the use of Kimura’s neutral theory of molecular evolution as a “null hypothesis” in the context of the debate between the neutralists and adaptationists. Beatty (1997) shows that while the debate between the neutralists and adaptationists was framed as an exclusive choice, in fact both neutral drift and selection were being apportioned relative significance. Huss (2004) shows how the Woods Hole Group used the MBL model in paleobiology as both a null hypothesis and a null model. Gotelli and Graves (1996, 3) give clear examples of models used as null models across ecology, and theirs remains the only general book on null modeling in any science. But they define a null model, and I do not find their definition useful because to be a null model is just to be used in null modeling.

Regarding the neutral theory of ecology, Gotelli and McGill (2006) distinguish using it as *null* from its being *neutral*, correcting the common conflation of these two terms. Rosindell and colleagues (2012) distinguish between using the neutral theory of ecology as a null model and as an approximation and argue that you cannot do both at the same time. This is similar to my distinction, but baseline modeling is more than use as an approximation, and their analysis does not distinguish these uses and their associated reasoning strategies. Odenbaugh (forthcoming) distinguishes between being null and being causal and argues for using the neutral theory as a causal theory much as I characterize baseline modeling. But no one has appreciated how these strategies conflated into a single reasoning strategy that privileges a neutral theory. Therefore, I continue this critical dialectic and extend it to the case of the neutral theory of ecology. The distinction between null modeling and baseline modeling needs to be sharpened and honed. More cases can then be analyzed using this distinction.

**3. Null Modeling.** Null modeling tests a hypothesis that a set of processes is causally responsible for a set of types of patterns. Null modeling takes place in the following context. Scientists identify a type of target pattern (or set of patterns) that calls for causal explanation. Scientists form their original hypothesis that a set of original processes is causally responsible for the target pattern. One way to evaluate this hypothesis is to first formalize the processes into a mathematical model. If they formalize the original hypothesis into the original model, then they can determine whether the original model suffices to produce the target patterns. If it is sufficient, this is evidence for the original hypothesis. But even if the original model suffices, the original hypothesis might still be false. Stronger evidence is needed.

Enter null modeling. Null modeling begins by finding or constructing a model that excludes the proposed original process and that formalizes a set of processes that, if sufficient, would undermine the need for the original processes. The processes formalized in the null model may be either simpler in some way than the original processes or else a strict subset of the processes in the original model. The null model is null with respect to the original processes and is appropriate for playing the role of the null model for some set of original models. Null modeling tests the processes included in the original model and excluded from the null model.

Null modeling is based on two core ideas. First, if the null model suffices to produce the target patterns, this undermines the evidence for the original hypothesis inferred from the fit of the original model to the pattern, as the original processes have been shown to be unnecessary. Second, if the null model cannot produce the target patterns, then the evidence for the original hypothesis inferred from the fit of the original model to the pattern stands, as it shows that a set of processes that are weaker than the original processes is insufficient.

I characterize the Null Modeling Reasoning Strategy as follows. Assume scientists have already shown that the original model suffices to produce the pattern of interest and set standards for whether a model suffices to produce a pattern.

1. Select a model as an appropriate null model to test whether the original processes are causally relevant to a target pattern.
2. Compare the null model's outputs with the target pattern. The null model is either sufficient or insufficient to produce the target pattern.
3. If the null model is sufficient, then
  - 3.1. Reject the original model, and
  - 3.2. Undermine the evidence for the original hypothesis gained from model fit.
4. If the null model is insufficient, then
  - 4.1. Reject the null model, and
  - 4.2. Let stand the evidence for the original hypothesis gained from model fit.

Null modeling is not a method for choosing between two hypotheses, theories, or models. It is a method for evaluating one hypothesis about the causal responsibility of a set of processes for certain patterns. The two models are treated asymmetrically: the original hypothesis is tested by the null model but not vice versa. Only the original hypothesis can gain evidence. However,

there is nothing preventing a model from playing the role of the null model in one case and playing the role of the original model in another case. Models are not inherently null, although a model may be appropriate or inappropriate for playing the role of a null with respect to another model.

For example, null modeling was first used in ecology to evaluate whether competition between species of different genera are causally relevant to the ratio of the number of species to the number of genera in a community (species to genus ratio, or S/G).<sup>3</sup> The null model used for testing the original process of intrageneric competition was a model that included no generic differences across individuals and that distributed individuals to areas randomly (with respect to genus). The null model produced similar S/Gs to those observed, and this undermined the evidence for intrageneric competition on S/Gs.

**4. Baseline Modeling.** Scientists use baseline modeling to apportion relative causal responsibility for token patterns in the following context. Scientists accept that multiple processes are jointly responsible for producing types of patterns. They want to determine the relative contribution of each process in token cases.

Baseline modeling proceeds by first establishing what we would expect to observe if only a subset of accepted processes were contributing to a pattern. The baseline model formalizes this subset of processes. The *additional model* formalizes the remainder of the accepted processes. Deviations from the baseline model's output and token patterns are explained by invoking the additional processes as causally responsible on top of the baseline processes. This estimates the relative significance of baseline and additional processes to the token pattern of interest. If the additional processes understood and modeled, then the strength of these processes can be measured. There can be more than two sets of processes—baseline and additional. The additional processes can be subdivided and measured separately by iterating the same reasoning.

I characterize the Baseline Modeling Reasoning Strategy as follows.

1. Select a model as an appropriate baseline model to apportion relative significance to multiple processes, each of which is relevant to the type of target pattern.
2. Compare the baseline model's outputs with a token target pattern and prove they are either sufficient or insufficient to produce it.<sup>4</sup>
3. If the outputs are sufficient, then

3. For more detail on this example and its development, see Gotelli and Graves (1996, 13) and Huss (2004, 123).

4. It is sufficient or not depending on the standards set for what model fit is good enough. I characterize the reasoning here in binary terms because it is often used this

- 3.1. Conclude that the baseline processes dominate the additional processes in producing the token pattern.
4. If the outputs are insufficient, then
  - 4.1. Conclude that the additional processes dominate the baseline modeling in producing the token pattern.

The two models are treated asymmetrically: baseline modeling measures the baseline responsibility first and then uses the additional processes to fill the gap but not vice versa. In this way, baseline modeling proceeds by assuming that the baseline model accurately describes the system, both in the complete absence of additional processes and to a degree that is inversely proportional to the strength of additional processes. Models are not inherently baselines, although a model may be appropriate or inappropriate for playing the role of a baseline with respect to certain additional processes.

For example, physicists use baseline modeling to detect and measure the force of gravity on the motion of heavenly bodies. The principle of inertia supplies the baseline model with which accelerations qua deviations are detected. The law of gravitation supplies the additional model with which the distance and mass of the influencing body can be estimated from the strength of the force.

**5. Null Modeling versus Baseline Modeling.** The formal similarity of null and baseline modeling facilitates their conflation. Both reasoning strategies use two models and treat them asymmetrically. The null and baseline models are both compared with data before their counterparts, and model fit obtained informs a judgment about the original hypothesis and additional processes. But they differ from each other along four dimensions: empirical target, purpose, context, and commitment.

*Empirical Target:* The empirical target of null modeling is a type of pattern and its variation across cases. Examples of types of patterns include the motions of bodies, the behaviors of gases, and climate change.

The empirical target of baseline modeling is a token instance of a pattern. Examples of types of patterns include the current motion of the earth around the sun, how the pressure of hydrogen changes as the temperature rises and the volume is fixed, and the warming trend in earth's climate.

*Purpose:* The purpose of null modeling is to test for the causal relevance of a set of processes to a type of pattern. A process is causally relevant to a pattern just in case it is sometimes a cause of the pattern. Examples of such

---

way, and it makes the formal similarity to null modeling clear. But the reasoning can be made more complex to consider varying ratios of the baseline to additional processes.



tests include whether there is a fifth force (beyond the electromagnetic, strong, weak, and gravitational forces) relevant to the motion of bodies, whether intermolecular forces are relevant to the behavior of gases, and whether anthropogenic activities are relevant to climate change.

The purpose of baseline modeling is to apportion relative responsibility across multiple processes relevant to token patterns. A process is responsible to a token pattern to the degree to which it makes a difference in the case. Examples of such apportioning include gravitational and electromagnetic forces acting on the motion of the earth around the sun, collisions and intermolecular forces acting on the pressure of hydrogen, and anthropogenic and nonanthropogenic causes of the warming trend.

*Context:* The methodological context of null modeling is evidence. Null modeling evaluates the evidence for the causal responsibility of a process for a type of pattern. Judgments of such evaluations include that there is insufficient evidence for the existence of a fifth force, that intermolecular forces are relevant to the behavior of gases, and that anthropogenic activities are relevant to climate change.

The methodological context of baseline modeling is explanation. Baseline modeling measures and causally explains a token pattern in terms of the relative responsibility of multiple processes. Examples of such causal explanation include that gravitation dominates electromagnetic forces in the orbit of the earth around the sun, that collisions dominate intermolecular forces in the behavior of hydrogen gas, and that anthropogenic activities make a difference to the warming trend of the earth.

*Commitment:* Null modeling is agnostic concerning whether the null model accurately represents, or can feature in a causal explanation of, the target system. Therefore, possible outcomes of null modeling make positive claims only about the original hypothesis. The criteria for being an appropriate null model depend only on the relationship between the original model and the null model. This is why null models are null with respect to their original model.

Baseline modeling requires that the baseline model accurately represents, or can feature in a causal explanation of, the behavior of the target system in the absence of additional causal processes. Therefore, baseline processes are afforded some degree (possibly zero) of causal responsibility, no matter the model fit of the baseline model to the target system. The criteria for being an appropriate baseline model depend only on the relationships between the baseline model and the target system.

It is easy to see how one could fail to distinguish between null modeling and baseline modeling. But failing to make this distinction is not hypothetical. Ecologists using neutral theory continue to fail to make it.

**6. The Neutral Theory of Ecology.** Community ecologists study biodiversity and biogeography. They are interested in the abundances and diver-

sity of organisms and how these distributions vary across space and time. Some restrict themselves to studying single trophic-level communities. A single trophic-level community comprises all the organisms at the same energy level in the same place. One contains only organisms that eat, and are eaten by, the same kinds of things. For example, all the trees in a forest are studied as a single trophic-level community, as are all the birds on an island with raptors excluded. In what follows, I refer only to single trophic-level communities.

One pattern investigated in communities are relative species abundance distributions (SADs). There are three stages to constructing a SAD. First, ecologists sample a community, or in rare cases census it, and record the species of every individual. Second, they drop names of the species from the data to permit comparisons across communities that may not possess the same mix of species. Now the data include only the number of species and the number of individuals in each species.

Third, they represent these data as a curve on a graph. A dominance-diversity plot, one such type of representation, has axes of percentage relative species abundance (on  $\log_{10}$  scale) and species rank in abundance. Figure 1 compares five SADs on a dominance-diversity plot. A point on a line is read as: ‘the  $x$ th ranked species makes up  $y\%$  of the community’.

The patterns of abundance and diversity, including SADs, are types of general patterns. They exhibit variation in form across token instances. When you compare SADs of distinct communities, you typically see an ‘S’ shape—steep, then flatter, then steeper—showing that most species are rare and few are common. The S shape of SADs is robust across and within many types of communities and calls for explanation.

Community ecologists want to describe, predict, and explain tokens and types of SADs. Competition theory is the traditional way of approaching such problems.<sup>5</sup> Ecologists using competition theory hypothesize that the competitive differences between individuals of distinct species are the most important causal factors relevant to SADs: every species is best adapted to a set of resources and conditions, called its *niche*, by natural selection. When two species live in the same area, the species that better succeeds at using the available resources will likely outcompete the other. The *competitive exclusion principle* generalizes this idea, stating that, if there are insufficient differences between how two species use resources and affect their environment, then those two species cannot stably coexist. One species will eventually exclude the other. The general process responsible for this is *interspecific competition* and is increasingly referred to as selection (Vellend 2016).

5. An example of a competition theory is Tilman’s (1982) R\* Theory. For a defense of theorizing in terms of competition and niches against the neutral theory, see Chase and Leibold (2003).

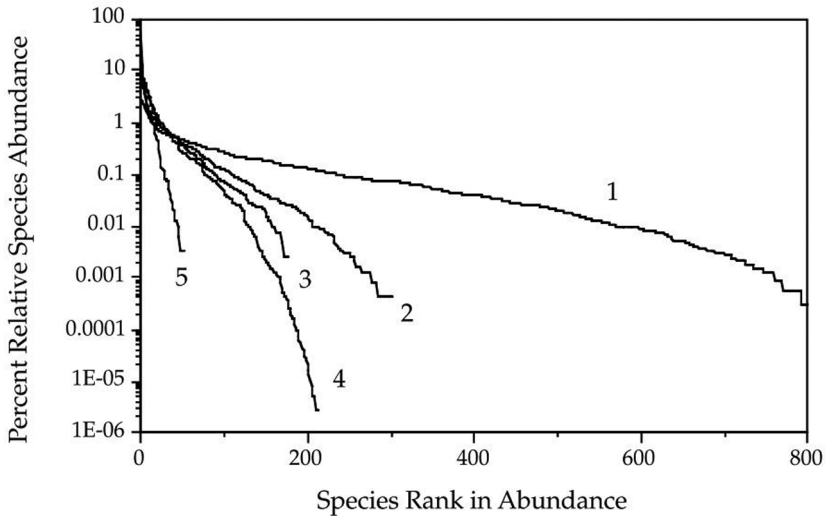


Figure 1. Reproduced with permission from Princeton University Press from Hubbell (2001).

In direct contrast, users of neutral theory hypothesize that patterns of biodiversity and biogeography result from chance and history. According to neutral theory, a tree lives where it does, not because its species is best adapted to specific local conditions but because that spot was recently vacated and seeds happened to land there. The ability to disperse seeds or young to areas with open resources is one of the most important causal factors. To a first approximation, organisms of every species are equally able to disperse to and colonize any open space.

Hubbell constructed neutral theory to counter proposals that interspecific competition causes the patterns of abundance and diversity. The *neutral theory of ecology* has three independent characteristics. First, it hypothesizes three ecological processes: drift, immigration, and speciation. Ecological drift is random death and birth in a community, also called demographic stochasticity. Immigration disperses individuals between different regions. And speciation results in the birth of an individual of a novel species. This set of demographic and dispersal processes controls how SADs can change over time.<sup>6</sup> Demographic and dispersal processes are chancy, historical processes because the SAD at any given time results from both its previous state and the prob-

6. There is a philosophical debate over whether genetic drift is really a causal process and distinct from natural selection. Ecologists certainly treat drift as a causal process in their reasoning. For defenses of genetic drift as a causal process that transfer to ecology, see Millstein (2002) and Clatterbuck (2015).

abilities that some individual dies, gives birth, immigrates, and gives birth to an individual of a new species.

Second, neutral theory assumes *neutrality*. Neutrality states that all individuals in a community, regardless of species, are functionally equivalent. In a neutral-demographic-dispersal theory, individuals have identical chances of dying, giving birth, immigrating, and giving birth to an individual of a new species. Assuming neutrality rules out the process of interspecific competition since this requires competitive differences between species.

Third, neutral theory assumes *saturation*. Saturation implies that a community's size is approximately fixed. This ensures competition for space between species. But, coupled with neutrality, this is competition between equals. In a saturated neutral community, interspecific competition does not differ from intraspecific competition.

A neutral model of the neutral theory of ecology is a mathematical model that formalizes the assumptions made and processes hypothesized by the theory. The mathematics of a neutral model can be interpreted as follows: imagine a community as a finite checkerboard with  $J$  spaces where an individual can live.<sup>7</sup> Fill the board with pieces of various colors drawn from a bag. Each piece represents an individual, and each color represents a species. The board represents the island or local community, and the bag represents the mainland or metacommunity. Ascertain the SAD of the board at a given time by counting the number of colors and individuals of each color on the board and representing these data on a graph.

The neutral model describes how the board's SAD can change over time and predicts the time-independent equilibrium SAD obtained in the long run. The neutral model is a discrete-time model. Each turn consists of removing a piece and refilling the space with a new piece to maintain constant community size. With the board filled, begin by removing a piece according to the *death rule*: each piece on the board has an equal chance of being removed. To satisfy saturation, fill the opening. Fill the open space either by a local birth or by immigration. With probability  $1 - m$ , replace the death with a piece according to the *birth rule*: each piece on the board has an equal chance of reproducing and filling the space. To satisfy neutrality, the probability that a piece of a given color will fill the space is just proportional to the current relative abundance of that color. With probability  $m$ , replace a death according to the *immigration rule*: each piece in the bag has an equal chance of immigrating and filling the space. Under only drift (birth and death), local species diversity will tend to decrease. Immigration increases local diversity by drawing individuals from the metacommunity, where it decreases diversity.

7. Depicting the neutral model of ecology as a checkerboard has been used by Rosindell et al. (2012).

Altogether then, the SAD of the board evolves through the following steps:

1. Remove a piece from the board following the death rule and go to step 2.
2. Fill the gap on the board with either step 2.1 or step 2.2 and then return to step 1.
  - 2.1. With probability  $1 - m$ , fill the gap with a piece on the board following the birth rule—each piece on the board has an equal chance of reproducing and filling the space.
  - 2.2. With probability  $m$ , fill the gap with a piece from the immigrating bag following the immigration rule—each piece in the bag has an equal chance of immigrating and filling the space.

Notice that the dynamics of the board involve the SAD of the bag. For immigration to work, the dynamics of a metacommunity must also be tracked, and this follows analogous death and birth rules. Here drift and emigration occur, balanced by speciation instead of immigration. The simplest *speciation rule* (used in the point mutation model) states: replace a death in the bag of size  $J_M$  by an individual from a novel species with probability  $v$ . Speciation is the ultimate diversity source. The diversity produced by speciation is dispersed to the local community, where it is counters drift as the ultimate diversity sink.

After many rounds of death and replacement in both the bag and on the board, a time-independent equilibrium SAD will probably be reached. In equilibrium, diversity is the balance of speciation and drift, like the level of water in a sink is at equilibrium is the balance of inflow and outflow rates. This equilibrium SAD does not depend on the initial configuration of the board, only on the values of the free parameters  $J$ ,  $m$ , and  $\theta$  (a function of  $J_M$  and  $v$ ). The neutral model outputs a statistical distribution that can be compared to observed SADs in two ways.<sup>8</sup> First, variations in the outputs of the neutral model are studied to understand the variation in outputs possible and the sensitivity of the outputs on the free parameters. Second, parameter values are set, and the output distribution is compared with token SADs so that model fit can be measured. Cases of very good model fit have are obtained, especially for tropical forest plots. Poor model fits to SADs are also obtained. The conclusions that can be inferred from good and bad model fit depend on the reasoning strategy being used.

8. For rigorous descriptions of the mathematics of the neutral theory, see Hubbell (2001), Volkov et al. (2003), and Etienne and Alonso (2007).

**7. Hubbell's Uses of the Neutral Theory.** I want to understand both why privileging a neutral theory as accepted until rejected is illegitimate and how such a privileging could come about. Section 8 will explain both as the result of conflating null and baseline modeling. But this conflation is not hypothetical. This section shows that null and baseline modeling are inseparable from the development of the neutral theory. Hubbell shifted from using the neutral model as a null model for testing competition theory to using the neutral model as a baseline model for detecting interspecific competition. And when another ecologist presented Hubbell with a choice between these two reasoning strategies, Hubbell failed to choose.

*7.1. Hubbell's Use of the Community Drift Model.* Hubbell (1979) developed the ancestor of the neutral theory, the *community drift model*, for tropical forest ecology. The community drift model includes only drift and immigration and assumes saturation and neutrality. As we can see from the conclusions he draws, Hubbell used the community drift model only as a null model for testing the role of interspecific competition in shaping tropical forest patterns.

Hubbell drew two conclusions from the comparison of the community drift model with SADs. First, he stated, "we may expect to observe substantial differentiation of the relative abundance of species in natural communities as a result of . . . a kind of 'community drift'" (Hubbell 1979, 1307). By varying the free parameters of the model, a wide variety of SADs can be approximated. This shows that the community drift model can fit the general diversity of observed SADs. Hubbell also fit the community drift model's output to the token SAD of the tropical dry forest studied and showed that a close fit could be obtained by inputting the known community size and fitting the immigration parameter value. These results were surprising given the expectation that interspecific competition dominates.

Second, he claimed, "we cannot necessarily conclude that, just because a species is of rank-1 importance in a community, its current success is due to competitive dominance . . . stemming from some superior adaptation to the local environment" (Hubbell 1979, 1307). The ability of the community drift model to fit observed SADs and their variation shows that interspecific competition is not required to explain either SADs or the dominance of a species. This undermined the evidence for interspecific competition as a cause of those patterns based on model fit. Hubbell drew no further causal conclusions about the responsibility of random drift or immigration from the model fit of the neutral model. His conclusions follow from null modeling.

*7.2. Hubbell's Use of the Neutral Theory of Ecology.* Hubbell (2001) later developed the community drift model into a general, positive theory of abundance and diversity in all communities, named the Unified Neutral

Theory of Biodiversity and Biogeography. This coincides with his first use of baseline modeling. But he did not distinguish this from null modeling.

Terborgh, Foster, and Nuñez (1996, 564) used the community drift model as a null model for testing interspecific competition in a series of forest plots in Peru. Their target pattern was the observation that the same species were most abundant across a set of five forest plots. They argued that such correlated abundances were very unlikely given the community drift model. Therefore, using it as a null model, they claimed to show interspecific competition to be a cause of abundance patterns and that the most abundant species are the best competitors in that region.

Hubbell responded to this challenge in 2001. He showed that, because immigration connected the five sites in a river valley, neutral theory almost always predicts that the most abundant species in one local community will be the commonest species in another local community. Therefore, he did not reject the neutral model (with or without speciation) qua null model and did not invoke interspecific competition. So far, his usage is consistent with neutral theory's supplying only a null model.

Yet, where Hubbell had stopped at the negative conclusion regarding the lack of support for interspecific competition, he now pressed the argument further. While some correlation between the commonest species did not show the importance of interspecific competition, the degree of correlation in which the commonest species dominate the other abundant species did show this. By comparing the observed correlation with the expectation given neutrality, Hubbell estimated that this *ecological dominance deviation* showed a 6% fitness advantage in the commonest species (2001, 337). He thereby explained a token pattern of correlated abundance using neutral theory as a baseline model.

Hubbell's baseline use is the only time in the book he describes his reasoning using "null": "What they did not comment on, however, was the excessive dominance itself—presumably because they had no prior statistical hypothesis of what null relative abundance distribution to expect" (2001, 336). Looking forward, Hubbell and other ecologists continue to use the neutral theory to draw two kinds of causal conclusions about ecological communities. These causal conclusions will feature in section 8, where I use them to characterize and critique the reasoning strategy being used to privilege the neutral theory. First, Condit, Chisholm, and Hubbell draw negative conclusions: "Species interactions, niche partitioning, or density-dependence, while they may be present, do not appear to enhance tree species richness at Barro Colorado" (2012, 1). This is a negative conclusion about the lack of responsibility of several processes linked to interspecific competition in one community.

Second, Condit and colleagues draw positive conclusions: "The neutral model predicts diversity and abundance at Barro Colorado because it prop-

erly describes what matters most—species input [neutral speciation]—while ignoring irrelevant details” (2012, 5). This is a positive causal conclusion about the relative significance of multiple processes in one community, measured by detecting deviations from the neutral model’s predictions. The strength of both kinds of conclusions depends on the reasoning strategy used to infer them.

At the time of writing his 2001 book, Hubbell had so far remained practical, never explicitly addressing the methodological issues raised by neutral theory. But this soon changed.

*7.3. Bell’s Weak versus Strong Distinction.* Just as Hubbell debuted the unified neutral theory in 2001 as a positive and general theory, ecologist Bell introduced an important methodological choice and explained the difference the choice would make.

Bell said that the success of the neutral model raised an important question with practical importance for community ecologists: What should ecologists infer from the success of the neutral theory? He gave two answers: ecologists can use the neutral theory in either a “weak” or a “strong” way (2001, 2418). The weak use is agnostic regarding whether the success of the neutral model shows that it captures the underlying processes of real communities and justifies causal explanations. “The role of the NCM [neutral community model] is then restricted to providing the appropriate null hypothesis [null model<sup>9</sup>] when evaluating patterns of abundance and diversity” (2418). Ecologists can use the neutral theory in the weak way to evaluate the evidence of other hypotheses and theories of patterns of abundance and diversity.

The strong use makes a stronger commitment. “The strong version is that the NCM is so successful precisely because it has correctly identified the principal mechanism underlying patterns of abundance and diversity” (Bell 2001, 2418). Ecologists can use the neutral theory in the strong way to explain observed patterns and to unify disparate ecological phenomena in terms of neutral drift, immigration, and speciation. The strong use does not require that the neutral theory explain everything in its domain perfectly. Bell acknowledged that species differences are sometimes causally relevant to patterns of abundance and argued that the strong use enables detecting when interspecific competition is causally important by seeing when the theory fails to give adequate predictions. The strong use is almost baseline modeling.

9. Bell distinguished null hypothesis testing from null modeling (in other terminology) in his other methodological paper about neutral theory (Bell 2000).



Hubbell (2006) responded to Bell's choice.<sup>10</sup> Hubbell began by answering the question, 'what can ecologists do with a neutral model?': "We obtain a quantitative null hypothesis against which to test when, to what extent, and for which species demographic differences among species are necessary to explain observed community patterns. The UNT [unified neutral theory] is a nontraditional approach because it does not postulate differences among species as the point of departure, but only adds them when necessary to explain the observed data" (1387). While he couched this in 'null hypothesis' terminology, he intended determining the extent to which species differences matter when detecting deviation with neutral models. Ecologists should first explain that observed patterns are using the neutral drift, immigration, and speciation and then add species differences. This is baseline modeling.

Next, Hubbell segued into the general problem that neutral patterns do not imply neutral processes, a version of the underdetermination problem. He said, "I agree, but obtaining acceptable fits from neutral models shifts the burden of proof to those who would assert that more complex theory is required to explain nature and with what level of detail and generality" (Hubbell 2006, 1387). Because the theory is neutral, it shifts the burden of proof onto competition models to reject the neutral model first. The neutrality assumption breaks the underdetermination problem. Here he also shifts back from explaining one SAD to undermining the evidence for interspecific competition as a general cause of SADs.

Hubbell then argued that the empirical success owing to the great model fits is strong evidence that the neutral theory accurately captures something about patterns of abundance and diversity, supporting the strong use and baseline modeling.<sup>11</sup> But he closed the opening section by saying that, irrespective of its eventual fate, it will remain useful as a null model.

What then are we to make of Hubbell's response to Bell? This depends on his position on the weak versus strong question. One position he defends is that the neutral theory will remain useful as a null model even if it is not useful as a baseline model. This *Weak-if-not-Strong* position is defensible. But he also defends using the neutral theory as both a null model and a baseline model. This *Weak-and-Strong* position is vulnerable, and I will undermine it in section 8. There is a basic tension between the weak use and the strong use. The results of using the neutral model qua null model can un-

10. Leigh (2007) observed that this was the first place Hubbell uses "null," which is almost correct. Hubbell discusses the issue in his article on neutral theory and Stephen Jay Gould (Hubbell 2005), in which he draws the connection to the MBL model, and once in Hubbell (1979), as mentioned in sec. 7.1.

11. The current consensus in ecology is that this is extremely weak evidence for any theory. Fitting patterns such as SADs is too easy and can be done with too many models.

dermine support for ever invoking interspecific competition. But using the neutral model qua baseline model always invokes interspecific competition as an actual cause. There would be no conflict if the neutral model perfectly fit every SAD, but the world does not abide.

Despite the rhetoric of undermining interspecific competition as a cause, interspecific competition remains accepted, including by Hubbell. But thanks in part to Hubbell, many more ecologists accept drift, immigration, and speciation as important causes of patterns of abundance.<sup>12</sup> At stake is the relative significance of these various processes in particular cases within the context of explanation. This is part of the resolution sought. Hubbell shifted from using the community drift model as a null model to using the neutral theory as a baseline model. But he continued to couch his baseline modeling within the framework of null modeling. Hubbell should have respected the distinction between null modeling and baseline modeling. We can at last diagnose and cure the combined reasoning strategy that begins by using the neutral theory as a null model and ends by measuring the relative significance of all the processes.

**8. Prognosis Negative.** I introduced the distinction between null modeling and baseline modeling to diagnose how theories are illegitimately privileged as accepted until rejected. Section 7 showed that this distinction is not foreign to the neutral theory of ecology but that they have grown up together. It also showed that Hubbell switched between and sometimes conflated this distinction.

I can now show my reconstruction of the reasoning strategy used to privilege the neutral theory of ecology. Hubbell and various other ecologists have used it when discussing the neutral theory as a “null.” It begins with using the neutral theory as a “null” and ends with drawing the kinds causal conclusions about relative significance shown at the end of section 7. And no one should use it anymore.

### *Neutral-Null-Baseline Reasoning Strategy*

1. The neutral theory supplies the appropriate null model for testing whether interspecific competition is causally relevant to patterns of abundance and diversity, including SADs.
2. When comparing the neutral model with the targeted SADs, it is either sufficient or insufficient to produce the targeted SADs.

12. See Vellend (2016) for the theory containing all four processes, called the Theory of Community Ecology after Hubbell’s theory.

3. If it is sufficient, then
  - 3.1. Conclude that some combination of random drift, immigration, and speciation dominates interspecific competition in producing the SADs targeted.
4. If it is insufficient, then
  - 4.1. Conclude that interspecific competition dominates random drift, immigration, and speciation.

This reasoning strategy is problematic, first, because it serves rhetorically to privilege the neutral theory over its rivals (nonneutral, selection-based theories) in a way that covers its weaknesses. It is the privilege of a neutral theory to shift the *burden of disproof* onto its rivals while retaining the *benefit of the doubt* for itself. Because neutral theory ignores species differences, it provides an appropriate null model for testing interspecific competition. This shifts the burden of disproof to competition theory, forcing the neutral model to be rejected before interspecific competition can ever be invoked. Also, because neutral theory ignores species differences, we expect the processes of random drift, immigration, and speciation to act in the absence of interspecific competition. This gives the neutral theory the benefit of the doubt, allowing it to identify neutral demographic and dispersal processes as the dominant causes of the patterns of interest without comparing a neutral model with a competition model. Taken together, these two features mean that neutral processes are accepted as dominating until the neutral model can be rejected.

Second, this reasoning strategy is problematic because it does not yield strong reasons for accepting its conclusions. The privilege is not a legitimate sign of the virtues of neutral theories because it produces bad arguments. Take an application of the neutral-null-baseline reasoning strategy to some targeted patterns, such as the example I gave in section 1. The argument generated begins with the neutral theory supplying the null model for testing competition theory. The fit will be either sufficient or insufficient, and so either one set of processes or the other dominates. This reasoning strategy sanctions inferences to causal responsibility irrespective of how well the neutral model fits the patterns of interest. Why are these conclusions unjustified? This depends on how the neutral-null-baseline reasoning strategy is analyzed. There are four cases to consider.

Case 1: The reasoning could follow from null modeling. But if we were null modeling, the only possible positive conclusion would be that interspecific competition is causally relevant to the patterns of interest. Instead,

no matter whether the neutral model is sufficient or insufficient to fit the patterns of interest, positive causal conclusions about the relative significance of multiple processes follow. Neutral-null-baseline reasoning strategy is invalid if understood as null modeling.

Case 2: The reasoning could follow from baseline modeling. In this case, we should reinterpret the “null” language of the first premise as baseline language. If we do this, the conclusion follows because it matches baseline modeling. So, the reasoning is valid if we understand it as baseline modeling. There are, however, two problems. First, the “null” language is both meant to do work and actually doing work. Remember its rhetorical role in privileging the neutral theory. Reinterpreting “null” out of the neutral-null connection misrepresents the actual reasoning.

Second, users of neutral theory should admit when they are baseline modeling and drop the “null” language when measuring relative significance. But the support required for baseline modeling goes well beyond that required for null modeling. If they are baseline modeling, then they are making two assumptions: random drift, immigration, and speciation are causally relevant to patterns of abundance and diversity and random drift, immigration, and speciation are acting in the absence of interspecific competition. Reasons need to be given for making these two assumptions—reasons obscured by the neutral-null connection. I do not deny that they exist, but they are not clear and must go beyond assuming neutrality. In this case, the reasoning has an unjustified premise and is unsound.

Case 3: The reasoning could follow from a combination of null modeling and baseline modeling. But null modeling is distinct from baseline modeling. And the methods are in tension with each other when used in sequence. If we begin by testing interspecific competition against the neutral model qua null model, then we doubt the causal relevance of interspecific competition. If the reasoning invokes it in the conclusion of the argument, it is incoherent.

Case 4: The reasoning could follow from some other reasoning strategy. I know of no promising suggestions and so I leave this option open for others to elaborate.

Therefore, the neutral-null-baseline reasoning strategy produces arguments that are invalid, unsound, incoherent, or mysterious. The most charitable interpretation of this reasoning is case 2. When using “null” and “testing” and drawing causal conclusions about relative significance, ecologists are baseline modeling. However, using a neutral model qua baseline model requires more justification than using a neutral model qua null model.

What evidence and reasoning supports using the neutral theory to supply the baseline model? This could come from null modeling. It is possible to gain support for the neutral theory qua explanatory via null modeling but only when the neutral model plays the role of the original model. Gotelli and McGill (2006) discuss this procedure. The neutral model would have to be tested by an appropriate null model—null with respect to random drift, immigration, and speciation. For example, you might construct a model analogous to the Hardy-Weinberg equilibrium in genetics in which the SAD stays constant over time.

But there is no reason to limit the sources of evidence for a process's being causally relevant to the model fit of a pattern.<sup>13</sup> Independent support for the neutral theory's assumptions is also a source of evidence.<sup>14</sup> In the argumentation critiqued here, being neutral is the support given. But neutrality as a mere assumption of the neutral theory does not warrant causal explanations.

**9. Conclusion.** By calling his theory “neutral” and claiming it plays the role of a “null,” Hubbell revived methodological issues that have surrounded the use of this family of terms since at least the 1980s. This made possible privileging the neutral theory as accepted until rejected and using the neutral model to make causal explanations no matter how well it fits the data. To clarify these methodological issues, I distinguish two reasoning strategies: null modeling and baseline modeling. When any theory, neutral or otherwise, is privileged and called a null model but used to measure relative significance, the conclusions drawn with the neutral model do not follow. Ecologists want to use the neutral theory to supply the baseline model so that they can apportion relative responsibility to demographic, dispersal, and competition processes. To do so, they need to identify their justification and distinguish it from the neutrality of the theory.

#### REFERENCES

- Bausman, William C., and Marta Halina. Forthcoming. “Not Null Enough: Pseudo-Null Hypotheses in Community Ecology and Comparative Psychology.” *Biology and Philosophy*.
- Beatty, John. 1987. “Natural Selection and the Null Hypothesis.” In *The Latest on the Best*, ed. John Dupre. Cambridge, MA: MIT Press.

13. See Lloyd (1994) for a useful framework for understanding the kinds of evidence that are relevant to the neutral theory.

14. For the same reason, null modeling can directly undermine only the evidence for the tested model gained through model fit. It does not, at least obviously, undermine independent sources of support.

- . 1997. “Why Do Biologists Argue Like They Do?” *Philosophy of Science* 64 (Proceedings): S432–S443.
- Bell, Graham. 2000. “The Distribution of Abundance in Neutral Communities.” *American Naturalist* 155 (5): 606–17.
- . 2001. “Neutral Macroecology.” *Science* 293 (5539): 2413–18.
- Blythe, Richard A. 2012. “Neutral Evolution: A Null Model for Language Dynamics.” *Advances in Complex Systems* 15:1150015.
- Chase, Jonathan M., and Mathew A. Leibold. 2003. *Ecological Niches: Linking Classical and Contemporary Approaches*. Chicago: University of Chicago Press.
- Clatterbuck, Hayley. 2015. “Drift beyond Wright-Fisher.” *Synthese* 192 (11): 3487–3507.
- Condit, Richard, Ryan A. Chisholm, and Stephen P. Hubbell. 2012. “Thirty Years of Forest Census at Barro Colorado and the Importance of Immigration in Maintaining Diversity.” *PLoS ONE* 7 (11): e49826.
- Etienne, Rampal S., and David Alonso. 2007. “Neutral Community Theory: How Stochasticity and Dispersal-Limitation Can Explain Species Coexistence.” *Journal of Statistical Physics* 128 (1–2): 485–510.
- Fargione, Joseph E., Clarence Lehman, Stephen Polasky, and Pieter H. M. van Baal. 2011. “Entrepreneurs, Chance, and the Deterministic Concentration of Wealth.” *PLoS ONE* 6 (7): e20728.
- Gotelli, Nicholas J., and Gary R. Graves. 1996. *Null Models in Ecology*. Washington, DC: Smithsonian Institution.
- Gotelli, Nicholas J., and Brian J. McGill. 2006. “Null versus Neutral Models: What’s the Difference?” *Ecography* 29 (5): 793–800.
- Hubbell, Stephen P. 1979. “Tree Dispersion, Abundance, and Diversity in a Tropical Dry Forest.” *Science* 203 (4387): 1299–1309.
- . 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, NJ: Princeton University Press.
- . 2005. “The Neutral Theory of Biodiversity and Biogeography and Stephen Jay Gould.” *Paleobiology* 31 (S2): 122–32.
- . 2006. “Neutral Theory and the Evolution of Ecological Equivalence.” *Ecology* 87 (6): 1387–98.
- Huss, John Edward. 2004. “Experimental Reasoning in Non-experimental Science: Case Studies from Paleobiology.” PhD diss., University of Chicago.
- Kimura, Motoo. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
- Leigh, Egbert Giles. 2007. “Neutral Theory: A Historical Perspective.” *Journal of Evolutionary Biology* 20 (6): 2075–91.
- Lloyd, Elisabeth A. 1994. *The Structure and Confirmation of Evolutionary Theory*. Princeton, NJ: Princeton University Press.
- Millstein, Roberta L. 2002. “Are Random Drift and Natural Selection Conceptually Distinct?” *Biology and Philosophy* 17 (1): 33–53.
- Nitecki, Matthew H., and Antoni Hoffman. 1987. *Neutral Models in Biology*. New York: Oxford University Press.
- Odenbaugh, Jay. Forthcoming. “Neutrality, Niche, and Nulls: Causal Relevance in Ecology.” In *Philosophical Perspectives on Causal Reasoning in Biology*, ed. C. Kenneth Waters and James Woodward. Minneapolis: University of Minnesota Press.
- Quinn, James F., and Arthur E. Dunham. 1983. “On Hypothesis Testing in Ecology and Evolution.” *American Naturalist* 122 (5): 602–17.
- Raup, David M., Stephen Jay Gould, Thomas J. M. Schopf, and Daniel S. Simberloff. 1973. “Stochastic Models of Phylogeny and the Evolution of Diversity.” *Journal of Geology* 81 (5): 525–42.
- Rosindell, James, Stephen P. Hubbell, Fangliang He, Luke J. Harmon, and Rampal S. Etienne. 2012. “The Case for Ecological Neutral Theory.” *Trends in Ecology and Evolution* 27 (4): 203–8.
- Sloep, Peter B. 1986. “Null Hypotheses in Ecology: Towards the Dissolution of a Controversy.” In *PSA 1986: Proceedings of the 1986 Biennial Meeting of the Philosophy of Science Association*, ed. Arthur Fine and Peter K. Machamer. East Lansing, MI: Philosophy of Science Association.

- Steiner, Ulrich Karl, and Shripad Tuljapurkar. 2012. "Neutral Theory for Life Histories and Individual Variability in Fitness Components." *Proceedings of the National Academy of Sciences* 109 (12): 4684–89.
- Strong, Donald R., Jr., Lee Ann Szyska, and Daniel S. Simberloff. 1979. "Test of Community-Wide Character Displacement against Null Hypotheses." *Evolution* 33 (3): 897–913.
- Terborgh, John, Robin B. Foster, and Percy V. Nuñez. 1996. "Tropical Tree Communities: A Test of the Nonequilibrium Hypothesis." *Ecology* 77 (2): 561–67.
- Tilman, David. 1982. *Resource Competition and Community Structure*. Princeton, NJ: Princeton University Press.
- Van Valen, Leigh M. 1985. "Null Hypotheses and Prediction." *Nature* 314 (6008): 230.
- Vellend, Mark. 2016. *The Theory of Ecological Communities*. Princeton, NJ: Princeton University Press.
- Volkov, Igor, Jayanth R. Banavar, Stephen P. Hubbell, and Amos Maritan. 2003. "Neutral Theory and Relative Species Abundance in Ecology." *Nature* 424 (6952): 1035–37.
- Wimsatt, William C. 1987. "False Models as Means to Truer Theories." In Nitecki and Hoffman 1987.