The use and limitations of null-model-based hypothesis testing

Mingjun Zhang

Abstract: In this article I give a critical evaluation of the use and limitations of null-model-based hypothesis testing as a research strategy in the biological sciences. According to this strategy, the null model based on a randomization procedure provides an appropriate null hypothesis stating that the existence of a pattern is the result of random processes or can be expected by chance alone, and proponents of other hypotheses should first try to reject this null hypothesis in order to demonstrate their own hypotheses. Using as an example the controversy over the use of null hypotheses and null models in species co-occurrence studies, I argue that null-model-based hypothesis testing fails to work as a proper analog to traditional statistical null-hypothesis testing as used in well-controlled experimental research, and that the random process hypothesis should not be privileged as a null hypothesis. Instead, the possible use of the null model resides in its role of providing a way to challenge scientists' commonsense judgments about how a seemingly unusual pattern could have come to be. Despite this possible use, null-model-based hypothesis testing still carries certain limitations, and it should not be regarded as an obligation for biologists who are interested in explaining patterns in nature to first conduct such a test before pursuing their own hypotheses.

Keywords: null model, null hypothesis, checkerboard distribution, interspecific competition, random colonization, control of variables

1. Introduction

In a study of bird communities on New Guinea and nearby islands, especially the Bismarck Archipelago, Diamond (1975) claimed to have found certain "assembly rules" with respect to the distribution of bird species. One of them says that "[s]ome pairs of species never coexist, either by themselves or as part of a larger combination" (Diamond 1975, p. 344). For example, the Mackinlay's cuckoo-dove (*Macropygia mackinlayi*) and the bar-tailed cuckoo-dove (*Macropygia nigrirostris*), while ecologically similar and overlapping in their geographical ranges, never co-occur on any island of the Bismarck Archipelago. [1] This pattern, called by Diamond a "checkerboard distribution" because the geographical ranges of the two species interlace like black and white squares on a checkerboard, seems to require an explanation.

Diamond (1975) proposed the hypothesis that the existence of such checkerboard distribution patterns can be explained by, among other things, interspecific competition for limited resources. Some other ecologists, however, thought that it is premature to infer the existence of competitive exclusion among relevant species just by looking at their distribution patterns. For example, Connor and Simberloff (1979) argued that in order to demonstrate that competition is responsible for the formation of checkerboard distribution patterns, one needs to first falsify a null hypothesis that those patterns are due to the random colonization of bird species. Their strategy is to build a null model in which bird species are randomly assigned to different islands of an archipelago under

---

[1] In species co-occurrence studies, when claiming that a species exists, occurs, or is present on an island, ecologists typically mean that the species has established a breeding population on that island instead of just having several vagile individuals.

certain constraints and then calculate the values of certain statistics such as the number of species pairs that never co-exist on any island. If the expected number of forbidden pairs (or the value of other chosen statistics) is not statistically significantly different from what is obtained based on the empirical data collected from the actual archipelago, then there is no strong reason to appeal to interspecific competition as the explanation of checkerboard distributions.

Connor and Simberloff's challenge triggered sharp rebuttals from Diamond and his colleagues (e.g., Diamond and Gilpin 1982; Gilpin and Diamond 1984), which in turn incurred further rejoinders from Simberloff and his colleagues (e.g., Connor and Simberloff 1983, 1984). This disagreement has led to a schism between two camps that persists even today (for a recent round of debate, see Connor et al. 2013, 2015; Diamond et al. 2015).

This debate is worth attention because the strategy used by Simberloff and his colleagues, which I will call *null-model-based hypothesis testing*, has been widely used in the biological sciences, especially in ecological research such as studies of species co-occurrences, species/genus (S/G) ratios, food webs, and character displacement. To be clear, the use of the term "null model" is somewhat inconsistent in the biological literature. For example, Hubbell (2001) and Rosindell et al. (2011) argue that the neutral model of biodiversity provides a useful null model in community ecology, in which the term "null model" just means a model providing a null hypothesis. While the practice of using the neutral model as a null model has been criticized by Bausman (2018), the neutral model that he considers is different from the null model that I will focus on here.[2] The kind of null model that I will consider in this article is also called "random model" or "stochastic model," which is constructed based on a *randomization* procedure and used to test the hypothesis that the existence of a pattern is the result of random processes or can be expected by chance alone. This

---

[2] For a detailed discussion of the differences between neutral models and null models, see Gotelli and McGill (2006).

type of model is regarded by its proponents as playing a role similar to the "control" as used in traditional statistical hypothesis testing in experimental settings, and it is said to provide a null hypothesis against which other hypotheses should be tested.

At the heart of the debate concerning null models is whether null hypotheses and null models are useful at all in the biological sciences. Since an appropriate null model is essential for the validity of null-model-based hypothesis testing, most discussions have been focused on whether it is possible (and if so, how) to construct a model that is genuinely null. Many constructive discussions have taken place, but many technical details remain controversial. This article contributes to the literature from a somewhat different perspective. Assuming that the technical controversies about the construction of null models can be resolved, what would be the possible use of null-model-based hypothesis testing, if it is useful at all, and how could that use be justified? What limitations does this strategy have? These are the questions that I aim to answer in this article.

My strategy is to use as an example the controversy over Connor and Simberloff's advocacy of null hypotheses and null models in species co-occurrence studies and draw some general lessons from it. I will argue that null-model-based hypothesis testing as a research strategy can be useful – but in a limited sense. It is useful not because it works as an analog or approximation to traditional statistical hypothesis testing in well-controlled experimental settings, but because it provides a possible way to challenge scientists' commonsense judgments about how a seemingly unusual pattern could have come to be. To better demonstrate this point, I will compare the null model used by Connor and Simberloff with Schelling's model of segregation. I will also analyze the limitations on the use of null models and discuss some lessons drawn from the debate concerning the use of null-model-based hypothesis testing in species co-occurrence studies.

2. Null-model-based hypothesis testing in species co-occurrence studies

In order to answer the general question of the use and limitations of null-model-based hypothesis testing, it is necessary to properly detail how it is actually used in scientific research. In this section, I will use Connor and Simberloff's (1979) null model of species co-occurrences as a representative example to illustrate how null-model-based hypothesis testing has been employed as a research strategy in the biological sciences.[3]

2.1 Constructing the null model

Table 1. A 3-by-4 random matrix

	Island A	Island B	Island C	Island D	Row sum
Species a	1	0	1	0	2
Species b	1	0	0	1	2
Species c	0	1	0	1	2
Column sum	2	1	1	2	

Connor and Simberloff (1979) want to use their null model to construct a simulated archipelago in which bird species are randomly assigned to the islands. The null model used in their species co-occurrence analyses is actually a set of *random matrices*. In the rest of this section I will proceed by explaining what a random matrix is, how to construct it, and in what sense it is "random."

---

[3] In species co-occurrence studies, the null models constructed by different ecologists may be more or less different from each other. Even Connor and Simberloff themselves keep modifying their null models in later publications. Nevertheless, the version I will introduce here, which appears in one of their earliest and also most-cited publications on this subject, helps demonstrate the key features of null-model-based hypothesis testing.

A random matrix used in species co-occurrence analyses is a rectangular array of 1's and 0's, in which each row corresponds to a species, and each column an island within an archipelago (see Table 1). The number "1" means that a species is present on an island, while "0" means that a species is absent. Hence, the process of randomly assigning species to islands can be simulated by the process of randomly placing 1's and 0's in a matrix.

Since each element in the matrix has two choices (1 or 0), for a matrix with m rows and n columns, there will be 2^{mn} unique matrices if there is no constraint about how to place 1's and 0's, which correspond to 2^{mn} different ways of distributing m species on n islands. But in the relevant literature, the placement of species on islands in a random matrix is almost always subject to some constraints. Connor and Simberloff (1979, p. 1133) use the following three constraints:

> (i) For each island, there is a fixed number of species, namely, that which is observed.
> (ii) For each species, there is a fixed number of occurrences, namely, that which is observed.
> (iii) Each species is placed only on islands with species numbers in the range for islands which that species is, in fact, observed to inhabit. That is, the "incidence" range convention is maintained.

The first constraint requires that each island hold a fixed number of species, which is equal to the number of species actually observed on each island. For example, if a real island holds two bird species, it should also hold two in the random matrix. Since a column sum in a random matrix corresponds to the number of species present on an island, constraint (i) is equivalent to keeping all the column sums fixed at certain numbers according to the number of species actually observed on each island. Similarly, constraint (ii) requires that each species occur on a fixed number of islands, which is equivalent to keeping all the row sums fixed at certain numbers according to the number of islands on which each species actually occurs. The third constraint is more complicated. Some species may occur on islands with a wide range of species numbers, while some others may

be present only on species-poor (or, conversely, species-rich) islands. Hence, constraint (iii) is a further restriction on the kind of islands to which each species can be assigned. Note that although the matrices are called by Connor and Simberloff as "random matrices," the constraints used in these matrices are not themselves random. Rather, they are decided according to the empirical data collected from the actual archipelago. From a statistical point of view, these constraints amount to very strong parametric restrictions. The matrices are called "random" because Connor and Simberloff randomize the placement of bird species on islands in the simulation given the preset constraints. Accordingly, a distribution pattern is said to be generated by the species randomly colonizing an archipelago if their colonization of islands is not subject to any other constraints besides the ones already mentioned above.

If all the non-equivalent random matrices satisfying the preset constraints are collected together, they form the *full null space*. Given the number of species and islands in an actual archipelago (e.g., there are at least 147 species of land birds and 50 islands in the Bismarck Archipelago), the number of random matrices in this space can be very large. When it is computationally intractable to get all the possible random matrices, one can collect a sample of them. Random matrices in this sample constitute the *sample null space* (see Figure 1).
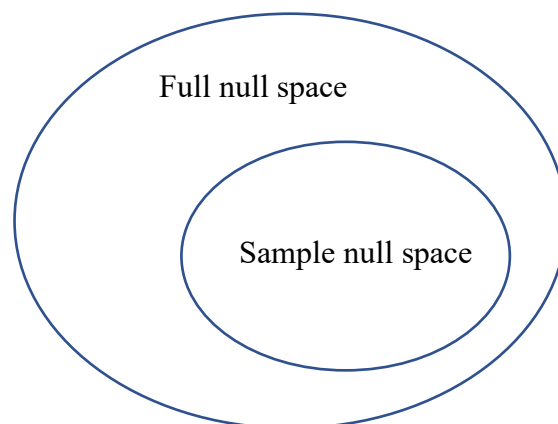

Figure 1. The relationship between the full null space and the sample null space

2.2 Comparing simulated data with empirical data

The test is conducted by comparing the empirical data collected from the actual archipelago with the simulated data generated by the null model. In order to conduct this test, one first needs to choose a statistic (i.e., a measure of some attribute of the sample) so that it can be used to make the comparison. There is more than one kind of statistic that can be used for the comparison, but what is most directly relevant to checkerboard distributions is the number of pairs of species that never co-occur on any island within an archipelago.

When understanding each random matrix in the sample null space as a simulated archipelago, one can count how many pairs of non-coexisting bird species there are on each of those archipelagos. Since these numbers may not be exactly the same, what one actually obtains is a *null distribution*, which describes the probabilities of all the expected values of the number of non-coexisting pairs based on the null model. Once this distribution is obtained, one can calculate its mean and standard deviation, compare them with the number of non-coexisting species pairs in the actual archipelago, and see whether there is a statistically significant difference between them. If no such difference is found, it means that the existence of checkerboard distributions is consistent with the result of random colonization. If so, Connor and Simberloff (1979) argue, then there is no strong reason to appeal to interspecific competition as the explanation of checkerboard distributions. It is worth noting that depending on the target phenomena and statistics of interest, researchers may choose measures other than the means and standard deviations of null distributions to conduct the test.

2.3 Technical controversies concerning the test

Many discussions have taken place concerning Connor and Simberloff's null model. Diamond and Gilpin (1982), for example, provide seven criticisms of Connor and Simberloff's (1979) approach. Here I only recapitulate two issues that have drawn the most attention. The first problem is called the *dilution effect*. According to Diamond and Gilpin, interspecific competition most likely happens within guilds, i.e., groups of ecologically similar species overlapping in resource utilization. Connor and Simberloff, however, conduct their test in whole faunas. As a result, the effects of interspecific competition within guilds are diluted by a mass of irrelevant data from ecologically distinct species. But this should be understood as a criticism of the specific model constructed by Connor and Simberloff (1979), rather than the use of null models as a whole. In fact, Simberloff and his colleagues have accepted this criticism and revised their models accordingly in later publications (Connor et al. 2013).

Another, perhaps more serious, problem is the incorporation of *hidden effects of competition* into the null model. In Connor and Simberloff's random matrices, the placement of species on islands is subject to three constraints. Diamond and Gilpin argue that since constraints (ii) and (iii) are themselves strongly influenced by competition, a null model subject to these constraints would already contain some effects of competition in its structure, making the model not as "null" as it needs to be. Since the construction of a null model always involves some constraints, the problem always exists that some effects of competition have already been included in the null model. If this is true, it would be a criticism of the use of null models as a whole, at least in studies of species co-occurrences. But whether this is a fatal issue is still controversial.

Technical issues like these, although important, are not the focus of this article.[4] In what follows, I will put aside these technical issues for a moment and consider instead some more general questions about the use and limitations of null-model-based hypothesis testing as a research strategy in the biological sciences. The benefit of doing so is that it can deepen our understanding of this research strategy in a more general way without being trapped in still unsolved technical controversies.

3. A critical evaluation of null-model-based hypothesis testing

3.1 Connor and Simberloff's interpretation

As mentioned before, advocates of null models think that null models provide a null hypothesis against which other hypotheses should first be tested. Null-hypothesis testing is not a native research method in ecology. Instead, it is borrowed from statistics.

The term "null hypothesis" was probably first introduced by R. A. Fisher. In his seminal book *The Design of Experiments*, Fisher (1935) considered a case where a lady claimed that by tasting a cup of tea mixed with milk she could tell whether the milk or the tea was first added to the cup. He designed an experiment to test the hypothesis that the lady did not have such discrimination ability, and labelled this hypothesis as the "null hypothesis." According to Fisher, a null hypothesis like this is characteristic of all experimentation, but one may choose any hypothesis as the null hypothesis as long as it is exact and free from vagueness and ambiguity. One important feature of the null hypothesis is that it is "never proved or established, but is possibly disproved, in the course of experimentation" (Fisher 1935, p. 19). Whether the null hypothesis in an experiment should be rejected can be decided by conducting a test of significance, which was developed by Fisher in the

---

[4] For reviews of the technical issues in the construction of null models, see Gotelli and Graves 1996; Sanderson and Pimm 2015.

1920s (Fisher 1925, 1926). The rationale behind this test is to calculate the *p*-value, i.e., the probability of obtaining a result at least as extreme as what is actually observed in the experiment given that the null hypothesis is true. If the *p*-value is smaller than a certain threshold chosen by the researcher, such as 0.05, the result of the test is said to be statistically significant and the null hypothesis is rejected.

Notice that in Fisher's tests of significance, only a single hypothesis – the null hypothesis – is considered. By contrast, in the Neyman–Pearson approach to statistical hypothesis testing, it is required that the null hypothesis be tested against an alternative hypothesis. While the debate between these two approaches is not the focus of this article, it is worth mentioning that the current approach to statistical hypothesis testing, which is widely used in statistical education and scientific research today, involves both the null and alternative hypotheses.

Fisher (1925, 1926, 1935) also played an important role in advocating the use of randomization in experimental design, which is used to avoid systematic biases and proves to be essential for the validity of statistical hypothesis testing in experimental research. Nowadays, randomized controlled trials[5] have become the gold standard in experimental studies, especially those that aim to test the cause-and-effect relationship between two variables. For example, in order to test whether an antihypertensive drug really has the effect of reducing blood pressure, subjects participating in the experiment are randomly assigned to two groups – the control group and the treatment group. Subjects in the treatment group take the drug every day, while subjects in the control group take placebo tablets, which look the same as the drug but do not contain the drug's

---

[5] Although the term "randomization test" is often used interchangeably with "permutation test," actually they are different. A randomization test is based on random assignment involved in experimental design; the procedure of random assignment is conducted before empirical data are collected. By contrast, a permutation test is a nonparametric method of statistical hypothesis testing based on data resampling.

active ingredients. Then the blood pressure data of the treatment group are compared with those of the control group statistically.

Notice that a "null model" is not needed for traditional statistical hypothesis testing used in well-controlled experimental settings (Figure 2a). By contrast, an appropriate null model is essential for the validity of the kind of testing conducted by Connor and Simberloff in their study of species co-occurrences – that is why I call it *null-model-based* hypothesis testing (Figure 2b). So the first question that we need to answer is: why is a null model needed in this kind of testing?
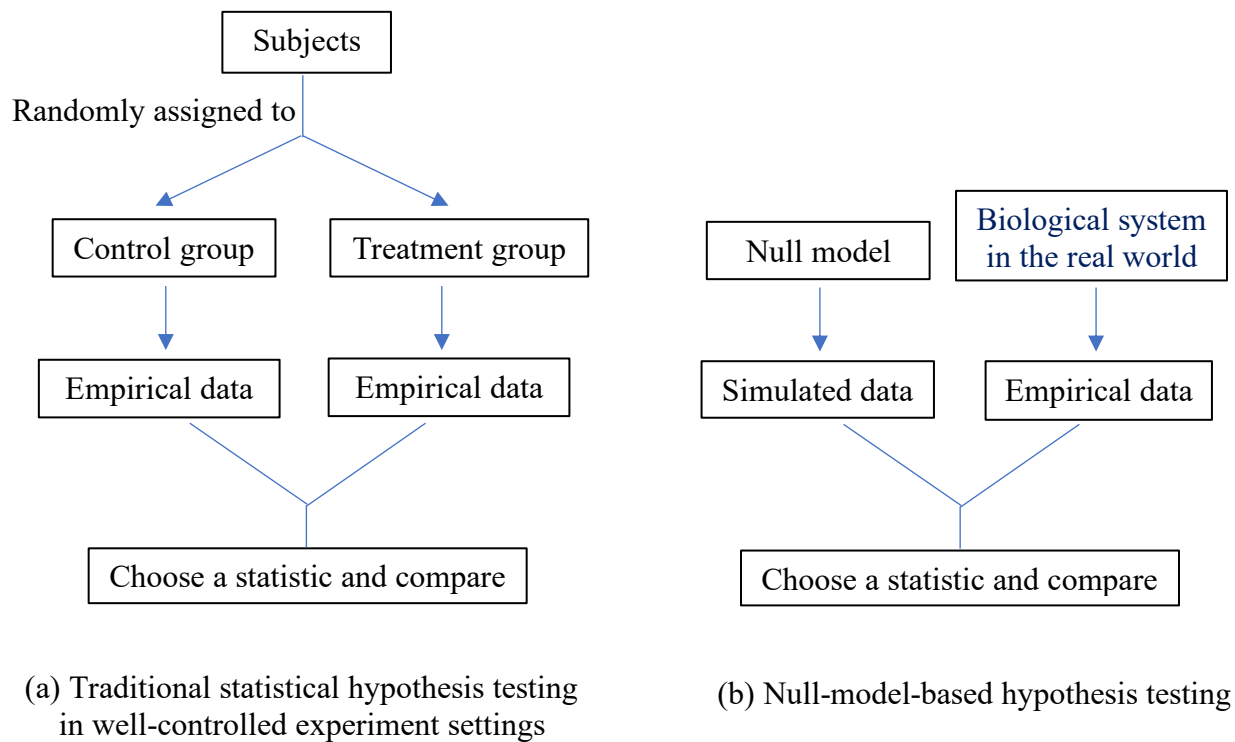
Figure 2. Procedures of two types of hypothesis testing

(a) Traditional statistical hypothesis testing in well-controlled experiment settings

(b) Null-model-based hypothesis testing

The term "null model" does not belong to the terminology of traditional statistical hypothesis testing. Instead, it was coined for the use of null hypothesis testing in non-experimental research such as many studies in ecology and biogeography. In well-controlled experimental studies,

researchers collect data from both the control and treatment groups, compare two data sets with respect to a statistic of interest, and see whether there is a statistically significant difference between them. But in an ecological case, the contrast between the control and treatment groups is usually lacking. For example, in the debate about checkerboard distributions, what is of interest is whether competition is responsible for the formation of those patterns. However, it is impossible to find an actual archipelago as the "control" in which bird species are manipulated in a way such that the level of interspecific competition is set as zero.

Given this, the role of the null model, as envisioned by Connor and Simberloff (1983), is to construct an imaginary archipelago in which bird species are randomly assigned to islands within the archipelago so that the effect of interspecific competition can be excluded. By using the null model as an analog of the control in experimental research, Connor and Simberloff (1983) aim to generate *simulated data* via the null model so that they can be compared with the *empirical data* collected from the actual archipelago. If this strategy works, then null-model-based hypothesis testing seems to be justified as a legitimate extension of traditional statistical hypothesis testing.

But is this interpretation of null-model-based hypothesis testing really justified? In the following sections, I will give a critical evaluation of this interpretation.

3.2 Evaluating the interpretation

3.2.1 The mismatch problem between the null and alternative hypotheses

Traditional statistical hypothesis testing used in well-controlled experimental research involves a pair of hypotheses – a null hypothesis and an alternative hypothesis. These two hypotheses must be *mutually exclusive* and *collectively exhaustive* of all the admissible hypotheses *given the question of interest* (Sloep 1986). For example, if researchers want to test the effect of an antihypertensive drug, then the question of interest is whether there is a cause-and-effect

13

relationship between the use of this antihypertensive drug and the reduction of blood pressure. Given this question of interest, the null hypothesis is that such a cause-and-effect relationship holds, and the alternative hypothesis is that it doesn't. Although the reduction of blood pressure can also be caused by many other factors, hypotheses involving those factors are not admissible hypotheses given the question of interest, and hence do not undermine the "collectively exhaustive" criterion.

Now consider the null and alternative hypotheses involved in null-model-based hypothesis testing. In their widely cited paper, Connor and Simberloff (1979, p. 1132) argue that:

> In order to demonstrate that competition is responsible for the joint distributions of species, one would have to falsify a null hypothesis stating that the distributions are generated by the species randomly and individually colonizing an archipelago.

If we focus on the controversy over checkerboard distributions, it is reasonable to formulate the null (H_0) and alternative (H_1) hypotheses used in Connor and Simberloff's test as follows:

> H_0: The checkerboard distributions of certain pairs of closely related bird species on the islands of the Bismarck Archipelago are the result of random colonization.

> H_1: The checkerboard distributions of certain pairs of closely related bird species on the islands of the Bismarck Archipelago are caused by interspecific competition.

Before analyzing the logical relationship between these two hypotheses, it is necessary to clarify their meanings. In the interspecific competition hypothesis (H_1), competition refers to interspecific competition for limited resources, which does not include interspecific differences in dispersal abilities. As to Connor and Simberloff's null hypothesis (H_0), what exactly it means depends on one's interpretation of the null model. Gotelli and Graves (1996, pp. 3–4) provide an operational definition of a null model:

> A null model is a pattern-generating model that is based on randomization of ecological data or random sampling from a known or imagined distribution. The null model is designed with respect to some ecological or evolutionary process of interest. Certain elements of the data are held constant, and others are allowed to vary stochastically to create new assemblage patterns. The randomization is

designed to produce a pattern that would be expected in the absence of a particular ecological mechanism.

This definition consists of both a *description* and an *interpretation* of a null model. The description part indicates how a null model is actually constructed: It is constructed based on the randomization of empirical data given certain constraints or random sampling from a species pool.

The interpretation part is about what a null model is intended to represent. According to the definition above, a null model in ecology is intended to simulate an ecological system where a particular ecological mechanism is excluded. Under this interpretation, Connor and Simberloff's null model is intended to simulate an archipelago where only interspecific competition is excluded, with their null hypothesis being simply the denial of the interspecific competition hypothesis. But this interpretation faces a serious problem. As Diamond and Gilpin (1982, p. 73) rightly point out, in order to use a null model to test the effect of a particular factor, the null model must both exclude the factor to be tested and include all the other important factors that could structure the actual data. However, it is not clear how Connor and Simberloff's null model can satisfy this requirement. On the one hand, as mentioned earlier, Diamond and Gilpin (1982) argue that the constraints in Connor and Simberloff's null model contain hidden effects of competition. On the other hand, even if this problem could be somehow managed, there is no way to guarantee that all the other relevant factors are included in the null model.

Another, perhaps better, interpretation of the null model is that it is intended to provide an alternative explanation to the competing hypothesis by taking into account chance processes. Under this interpretation, Connor and Simberloff's null model is a simulation of bird species' random colonization of islands, and the constraints used in the null model are structural assumptions about the ecological system under investigation. Accordingly, their null hypothesis is

an alternative explanation to, rather than a simple denial of, the interspecific competition hypothesis.

Sloep (1986, p. 309) claims that the interspecific competition hypothesis (H_1) and Connor and Simberloff's random colonization hypothesis (H_0) cannot be an appropriate pair of null and alternative hypotheses in the statistical sense because the first hypothesis does not *logically exclude* the second. This is not true. H_0 and H_1 are indeed mutually exclusive because they cannot both be true at the same time.

The real problem is that they are not *collectively exhaustive*. The question of interest in this case is how the checkboard distribution patterns have come into being. There are multiple possible answers to this question because whether two species will co-occur on the same island can be influenced by many factors other than interspecific competition, such as their dispersal abilities, habitat preferences, and responses to predators, parasites, or pathogens. Since H_0 and H_1 together do not exhaust all the admissible hypotheses given the question of interest, rejecting one of them does not necessarily provide support for the other. More specifically, rejecting that the checkerboard distribution patterns are the result of random colonization does not force us to conclude that they are caused by interspecific competition. By the same token, even if it can be shown that the patterns are not caused by interspecific competition, it does not mean that they are the result of random colonization.

My analysis above has demonstrated an important disanalogy between traditional statistical hypothesis testing and null-model-based hypothesis testing: the null and alternative hypotheses in the former satisfy the condition of being collectively exhaustive given the question of interest, while those in the latter do not. This difference is due to the fact that traditional statistical hypothesis testing and null-model-based hypothesis testing are faced with different tasks.

Hypotheses involved in traditional statistical hypothesis testing are usually statements about a relationship. Since a relationship either holds or not, the "collectively exhaustive" criterion is readily satisfied. By contrast, hypotheses involved in null-model-based hypothesis testing are usually explanations of a phenomenon. Since a phenomenon usually has multiple possible explanations, the "collectively exhaustive" criterion is typically unsatisfied.

It may be argued that the mismatch problem between the null and alternative hypotheses is not a fatal issue for null-model-based hypothesis testing. For example, in the case of checkerboard distributions, one can claim that when the random colonization hypothesis (H_0) is chosen as the null hypothesis, its appropriate alternative hypothesis is *not* the interspecific competition hypothesis (H_1), but the logical negation of H_0, i.e., the hypothesis that the checkerboard distributions are *not* the result of random colonization. Let's call this updated alternative hypothesis H_1^*. As shown in Figure 3, while H_0 and H_1 do not satisfy the collectively exhaustive criterion, H_0 and H_1^* do.
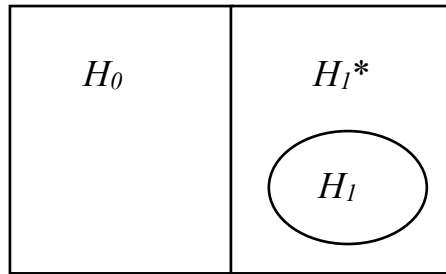

Figure 3. The relation between H_0, H_1, and H_1^*

Notice that the question of interest has been changed after the alternative hypothesis is reframed this way. What is of interest is no longer how the patterns have come into being, but whether those patterns are the result of random colonization. While this kind of reframing seems to circumvent the mismatch problem, it does not guarantee that null-model-based hypothesis testing is indeed an

appropriate null hypothesis test in a strict statistical sense. I will explain why in the following two sections.

3.2.2 The principle of equifinality and the lack of proper control of variables

In this section, I will examine Connor and Simberloff's (1983, p. 464) claim that in null-model-based hypothesis testing, a null hypothesis and a null model are contrived to "usefully approximate the role of the 'control' in order to test a hypothesis involving non-experimental evidence," and see whether they are actually able to play the role as envisioned.

In traditional statistical hypothesis testing, what counts as a proper control group is not solely determined by the setting of the supposed control group per se, but also influenced by how the treatment group is set up. Let's again use the test of the effect of antihypertensive drugs as an example. Suppose that the subjects participating in the experiment are randomly assigned to two groups. Subjects in the supposed treatment group take the antihypertensive drug and another kind of drug every day, while subjects in the supposed control group take placebo tablets. It is clear that the supposed control group does not count as a proper control because it does not guarantee that the only independent variable in this test is whether the subjects take the antihypertensive drug or not. Without proper control of variables, one is not justified to make any causal inference about the effect of the antihypertensive drug. More specifically, one is not justified to reject the alternative hypothesis that the antihypertensive drug has the effect of reducing blood pressure even when there is no statistically significant difference between the supposed control and treatment groups with respect to the subjects' blood pressure. This is because it is possible that the antihypertensive drug is indeed effective, but the other drug has the effect of increasing blood pressure, thus counteracting its effect.

Now consider null-model-based hypothesis testing. In the last section, I have argued that there exists a mismatch between the null and alternative hypotheses commonly used in null-model-based hypothesis testing, namely, they do not satisfy the "collectively exhaustive" criterion given that the question of interest is how to explain a phenomenon. One possible way to circumvent this problem is to reframe the question of interest as whether the pattern under investigation is the result of random processes. Since the answer to this question is either yes or no, the null and alternative hypotheses after this reframing will satisfy the "collectively exhaustive" criterion. Hence, if the empirical data is consistent with the simulated data generated by the null model, one seems to be able to retain the null hypothesis and reject the alternative hypothesis, just as what researchers would do in traditional statistical hypothesis testing.

I will use the rest of this section to show that this is not the case. I argue that in null-model-based hypothesis testing, because of the lack of proper control of variables, one is not justified to reject the alternative hypothesis even when there is no statistically significant difference between the simulated and empirical data with respect to the statistics of interest. The case of checkerboard distribution patterns will still be used as an example for demonstration.

In ecology, scientists make an important distinction between *pattern* and *process* (Gotelli and Graves 1996, p. 5). In the introduction to his book *Geographical Ecology*, MacArthur (1972, p. 1) wrote: "To do science is to search for repeated patterns, not simply to accumulate facts […]." This quote serves as a nice summary of a common approach adopted by ecologists – they start by identifying patterns in nature and then try to reveal the underlying processes that have produced these patterns. Since ecological systems are thermodynamically open systems, the same final state may be reached from different initial states and in different ways. This has been described as the principle of equifinality (von Bertalanffy 1968). Due to the equifinality of ecological systems, the

same pattern can be produced through different (or different combinations of) processes, just as there are many roads to Rome.

In Connor and Simberloff's null model of species occurrences, bird species are randomly assigned to islands of the simulated archipelago. This procedure is intended to exclude the effects of interspecific competition, but at the same time it also ignores many other factors that may influence species' co-occurrences such as the varied dispersal abilities of different species. Hence, there is no proper control of variables in the construction of the null model. Since closely related species tend to have similar dispersal abilities, they are more likely to colonize the same islands than what they would do if they were randomly distributed, which leads to an opposite effect of interspecific competition (Colwell and Winkler 1984; Harvey 1987). Given the multiple producibility of patterns, the joint effects of dispersal abilities and interspecific competition may produce a similar number of pairs of non-coexisting species as would be expected by random colonization. In other words, even if the simulated data generated by the null model are not significantly different from the empirical data with respect to the number of non-coexisting species pairs, one is not justified to reject the supposed alternative hypothesis (H_1*) that the pattern is not the result of random processes.

It may be argued that if a model based on a simpler hypothesis, such as the null model based on the random colonization hypothesis in the case of checkerboard distributions, is sufficient to produce a pattern statistically indistinguishable from the empirical pattern, then this hypothesis should be favored because of its simplicity. Although the principle of simplicity is often used to privilege simpler hypotheses over more complex ones, its use is not without controversy. For example, Sober (1988, 1994) has argued that there is no global justification for the principle of simplicity in general. Instead, its legitimacy depends on researchers' substantive background

assumptions about the way the world is in specific research contexts. Hence, what makes the principle of simplicity reasonable in one research context may be different from that in another. If proponents of different hypotheses have different background assumptions about the phenomenon under investigation, appeals to simplicity alone may not be able to justify privileging more parsimonious hypotheses. In the case of checkerboard distributions, proponents of the interspecific competition hypothesis tend to think that in a real biological context, it is very unlikely that the colonization of islands by birds is a purely random process. By contrast, proponents of null-model-based hypothesis testing think that it is problematic to exclude the random colonization hypothesis from the very beginning without testing it, which is why they urge researchers to conduct null-model-based hypothesis testing. However, if one chooses to privilege the random colonization hypothesis by appealing to simplicity, she needs to explain why the simpler hypothesis is more likely to be the true explanation of the phenomenon in this particular context.

As a brief summary, my analysis in this section reveals another important disanalogy between traditional statistical hypothesis testing and null-model-based hypothesis testing. For traditional statistical hypothesis testing used in well-controlled experimental research, since the proper control of variables is in place, if there is no statistically significant difference between the control and treatment groups with respect to the statistics of interest, one is justified to maintain the null hypothesis and reject the alternative hypothesis. In null-model-based hypothesis testing, however, because of the equifinality of thermodynamically open systems and the lack of proper control of variables, one is not justified to reject the alternative hypothesis even when the simulated data match the empirical data.

3.2.3 The issue with the size of datasets

In null-model-based hypothesis testing, phrases such as "comparing simulated data with empirical data" are rhetorically powerful and epistemically appealing because they leave people with the impression that the comparison involved here is analogous to the comparison between data from the control and treatment groups in an experimental setting. However, while a comparison procedure does exist in null-model-based hypothesis testing, the datasets that are compared may not be a good analog to those in traditional statistical hypothesis testing.

To test a cause-and-effect relationship between two variables, the size of both the control and treatment groups should be large enough in order to guarantee the statistical power of the test. The specific minimum requirement is based on the evaluation of the context of each specific test. In any case, if the sample size is too small, one is not justified to make any causal inference based on the test. For example, if there are thirty subjects in the control group and only one subject in the treatment group, one is not justified to reject the alternative hypothesis even if there is no statistically significant difference between the results of the two groups, because the statistical power of such a test is so weak.

In null-model-based hypothesis testing, sample size is not inherently a problem for the simulated dataset because one can conduct the randomization simulation multiple times, and each round of simulation will contribute a data point for the statistic of interest. In the case of species co-occurrence studies, for example, each random matrix works as a simulated archipelago and contributes a data point about the number of non-coexisting species pairs. The minimal size of the simulated dataset can be guaranteed by examining multiple random matrices in the sample null space. But in the empirical dataset, there is actually only one data point for comparison in each test, which is the number of pairs of species that never co-exist in the actual archipelago. In other

words, the comparison is conducted between a *null distribution*, which describes the probabilities of all the expected values of the number of non-coexisting pairs based on the simulated data generated by the null model, and a *single* empirical data point. While one can still test whether the value of the single empirical data point falls within the null distribution, it does not count as a proper analog to the comparison involved in traditional statistical hypothesis testing in experimental research.

3.3 Applying the analysis results

I have provided three reasons to show that null-model-based hypothesis testing is not an appropriate approximation of traditional statistical hypothesis testing in experimental research. Maybe this is an obvious point, but it does not always get enough attention in the discussion of null-hypothesis testing in the biological sciences, even among philosophers who aim to critically evaluate its use in specific cases. In this section, I will apply the results of my critical analysis to a particular philosophical discussion of null hypotheses and null models.

In a recent paper, Bausman and Halina (2018) evaluate the use of what they call the "pseudo-null strategy" in explaining relative species abundance (hereafter RSA) distributions, i.e., the relative commonness or rarity of different species on the same trophic level of an ecological community. The traditional way to explain RSA patterns is to appeal to niche differences. Every species has acquired a unique set of traits that allow it to be adapted to a particular environment and occupy a unique niche. Different species can coexist in the same environment because they have different niches and make use of resources in different ways (Chase and Leibold 2003). Hubbell (2001, 2006) has challenged this explanation by proposing a neutral theory of biodiversity. He assumes that all the individuals within a particular trophic level are ecologically similar

regardless of their species identity, i.e., they have the same chance of reproduction, death, immigration, and speciation. Hubbell uses this neutral assumption as a null hypothesis against which the niche-based hypothesis should be tested.

Bausman and Halina (2018) rightly point out that the neutral hypothesis is a pseudo-null hypothesis because compared with the hypothesis based on niche differences, it just appeals to a different set of processes, which include reproduction, death, immigration, and speciation, to explain RSA distributions. What is more relevant to this article is their suggestion about what counts as a genuine null hypothesis in this case:

> Community ecologists have sampled communities from all over the world and used the observed distributions to test whether *ecological selection* is operating. The appropriate statistical null hypothesis for such tests is that the observed species distributions will not differ significantly from what we would expect if individuals of different species were *distributed at random* in space and time […]. (Bausman and Halina 2018, p. 11, my italics)

Their suggestion is actually very similar to Connor and Simberloff's (1979) strategy, i.e., building a model in which individuals of different species are randomly distributed and then comparing the simulated data generated by this model with the empirical data collected from actual ecological communities. And the random distribution hypothesis is regarded as the null hypothesis. Not surprisingly, then, it suffers from the same problems as Connor and Simberloff's strategy. First, the authors' claim about the appropriate null hypothesis could be potentially misleading because they do not explicitly point out what the alternative hypothesis is given the statistical null hypothesis they suggest. If the alternative hypothesis is that the observed species distributions are the result of ecological selection, then it suffers from the same problem as the case of species co-occurrence studies, because the random distribution hypothesis that the authors suggest (the supposed null hypothesis) and the ecological selection hypothesis (the supposed alternative hypothesis) are not collectively exhaustive. This example reminds us again that whether a

hypothesis counts as an appropriate null hypothesis is not solely determined by its own content, but is also influenced by the choice of its alternative hypothesis. One cannot take for granted that a hypothesis is an appropriate statistical null so long as it claims that a phenomenon is the result of some random processes or can be expected by chance alone.

Bausman and Halina may reply to this critique by claiming that the corresponding alternative hypothesis is not the ecological selection hypothesis, but the hypothesis that RSA distributions are *not* the result of random distribution. But this change does not solve the problem of lacking proper control of variables. RSA distributions can be influenced by many processes, such as selection due to niche differences, ecological drift (random birth and death), immigration, and speciation. It is possible that the same distribution pattern of RSA can be produced through different combinations of those processes. So even if the simulated data match the empirical data, one is not justified to reject the alternative hypothesis and maintain the null.

The issue of dataset size also exists. In RSA studies, the unit of analysis is not the value of the point estimation of a statistic, but a *distribution pattern*. Through the random distribution model, one may be able to collect a set of distribution patterns by conducting the random simulation multiple times. But in the actual ecological community, one can only obtain a single distribution pattern of RSA for each trophic level. So what is compared is a set of distribution patterns from the simulated data and a single distribution pattern from the empirical data. This comparison fails to be a proper approximation to the comparison conducted in traditional statistical hypothesis testing in experimental research.

4. The possible use and limitations of null-model-based hypothesis testing

In the last section, I argued that null-model-based hypothesis testing cannot be justified as an appropriate analog to traditional statistical hypothesis testing in well-designed experimental research. But this does not mean that null-model-based hypothesis testing cannot be of any use at all. To better demonstrate the possible use of this strategy, I will make a comparison between Connor and Simberloff's (1979) null model of species co-occurrences and Schelling's (1978) model of segregation. I will argue that although these two kinds of models are constructed in very different ways, both of them can work as a way to challenge our commonsense judgments about what could have produced a seemingly unusual pattern.

4.1 Challenging "common sense" by providing how-possibly explanations

In his book *Micromotives and Macrobehavior*, Schelling (1978) presents a very simple model to study the segregation of people in neighborhoods. His model is based on a grid representing a community. Suppose that residents in this community belong to two different groups, and the distinction between the groups can be made based on any reasonable criterion, such as people's racial identity, language, religion, or occupation. Each square in the grid can be either empty or occupied by a resident, and each resident has a minimum requirement of their neighborhood, such as that at least one-third of immediate neighbors be of the same group as the resident. The simulation starts from a random distribution of two groups on the grid. Each individual takes turns to decide whether they are content with their neighborhood. For those who are not, as a rule they will move to the closest empty square that satisfies their requirement. It turns out that the initial integrated distribution of two groups becomes strongly segregated when all the residents are content with their neighborhood after a number of rounds. And the outcome is robust when one

alters the overall ratio of two groups, residents' minimum requirement of their neighborhood, the initial distribution, etc.

So what is the relevance of Schelling's model of segregation to our evaluation of the possible use of null models? After all, they are constructed in very different ways. The answer is that both Connor and Simberloff's null model of species co-occurrences and Schelling's model of segregation can be interpreted as a way to challenge the commonly held explanation of a seemingly unusual pattern by providing an alternative how-possibly explanation.

As I mentioned earlier, Schelling's model can be used to study the segregation of any two different groups in a community. Suppose that we specify the two groups in his model as two racial groups (such as black and white people), then this model can be used to study racial residential segregation in neighborhoods. More specifically, Schelling's model can be used to show that patterns of racial residential segregation can result from individual choices based on small preferences for similarity that are not racist in the sense of being entirely intolerant of racial diversity. This challenges the commonly held view that racial residential segregation is due to the presence of extreme racist attitudes. By using the word "challenge," I do not mean that Schelling's model refutes or falsifies the categorical racism explanation; likewise, I do not claim that his model describes how racial residential segregation *actually* forms. Nor does it follow that racist problems do not exist or are not severe. Rather, it just provides another possible explanation. If both explanations are compatible with the emergence of segregation patterns, then *without further evidence* one cannot take for granted that the observation of racial residential segregation implies categorical intolerance of racial diversity.

Similarly, in the case of species co-occurrence studies, if the simulated data generated by Connor and Simberloff's null model are consistent with the empirical data collected from the

actual archipelago with respect to the statistic of interest, then their model shows that the existence of checkerboard distributions might just be the result of random colonization. This alternative explanation challenges the prevailing view that checkerboard distribution patterns result from interspecific competition. Again, it does not follow that checkerboard distributions are *indeed* the result of random colonization, or that relevant species do not compete in reality. Nevertheless, it shows that *without further evidence* one cannot take for granted that exclusive competition among bird species actually happens and causes checkerboard distributions.

A possible objection to my analysis is that since in this article I do not aim to deal with the technical controversies over how to build an appropriate null model, if it turns out that those issues cannot be solved, then the possible use that I propose for this strategy would not be available. I agree. Whether the strategy of null-model-based hypothesis testing can *actually* be useful does rely on whether an appropriate null model can be constructed. This is why I employ the term "possible use" throughout this article. Nevertheless, my analysis can still contribute to the discussion by showing the *scope* of the use of this strategy, i.e., what it might be able to do and what it absolutely cannot do. As I have shown, even if it is possible to construct a truly null model, null-model-based hypothesis testing still fails to be an appropriate analog of traditional statistical hypothesis testing, given the reasons provided in Section 3. Instead, its possible use resides in its role of providing a way to challenge scientists' commonsense judgments about how a seemingly unusual pattern could have come to be.

4.2 The limitations of null-model-based hypothesis testing

Despite the possible use, null-model-based hypothesis testing still carries severe limitations. In Section 3, I showed that null-model-based hypothesis testing cannot test whether a particular

process or mechanism such as interspecific competition is responsible for the formation of certain patterns. Connor and Simberloff (1983, p. 464) explicitly acknowledge this limitation by noting that "[w]ithout further evidence, probably of an experimental nature, one can neither eliminate any particular causal mechanism, nor conclude that a particular mechanism has operated."

Another limitation of this strategy is that the null model lacks the potential for dealing with other problems via further refinements and modifications. The null model is constructed to test the hypothesis that the existence of a pattern is the result of random processes or can be expected by chance alone. No matter whether this hypothesis can be rejected, once the test is conducted, the null model's mission is done. It may be argued that the null model can be further revised or made more complex for other uses by changing its constraints. This method is possible in principle, but difficult to use in practice. The difficulty lies in how to construe the updated model when the constraints are changed, and it is usually unclear what the updated model is supposed to represent. It may also be argued that lacking the potential for dealing with other problems hardly amounts to a fatal issue. Since the null model is a bespoke model that is designed to test a particular hypothesis (i.e., the random process hypothesis), it does not matter so much if it cannot be further developed to deal with other problems. I agree that this is not a fatal problem for null-model-based hypothesis testing, but it is still a limitation; recognizing this limitation can help us understand why many biologists are not so interested in null-model-based hypothesis testing. Understood this way, my discussion of the null model's limitations is not so much a criticism as an observation.

5. Lessons from the debate

The controversy surrounding null-model-based hypothesis testing is commonly interpreted as a debate about the usefulness of null hypotheses and null models in the biological sciences,

especially in ecology and biogeography. But it is also important to take notice of the deeper origin of this debate. The motivation behind Connor and Simberloff's advocacy of null-model-based hypothesis testing is their general dissatisfaction with the then prevailing "competitionist paradigm" in community ecology. According to some ecologists, "competition has been invoked as an explanation for patterns to such a degree that it is in danger of becoming a panchreston, a concept that can explain everything" (Rathcke 1984, p. 383). In this extreme case, one might invent a competitionist explanation for almost any seemingly unusual pattern in ecological communities without sufficient evidence. Connor and Simberloff (1979, 1983) have repeatedly argued that the existence of checkerboard distributions per se is not evidence for one species being actively resisted by another, and that more compelling evidence, such as evidence based on detailed autecological study, field observation, or experimentation, should be provided to support the interspecific competition hypothesis.

Against this background, null-model-based hypothesis testing is used by its proponents as a way to challenge the "competitionists." If one can show that the empirical data is consistent with what would be expected by a null model, then merely identifying certain distribution patterns is not enough to support the commonly proposed competition hypothesis. Hence, ecologists like Connor and Simberloff are not really against the role of competition as a *possible* mechanism for producing certain patterns. What concerns them is the lack of compelling evidence for many competitionist explanations of ecological patterns. As Simberloff emphasized, "I've never said that there is no competition, that competition isn't important in generating patterns in nature […]. All I've been addressing is the canons of evidence" (Simberloff as quoted in Lewin 1983, p. 639).

But this is not the only thing that they want to contend. Connor and Simberloff (1979) were deeply inspired by Karl Popper's falsificationism and thought that science progresses by proposing

testable hypotheses and then trying to falsify them. Hence, besides requiring more compelling evidence, they also argued that proponents of the competition hypothesis (or any non-random hypothesis) should first try to falsify or reject a null hypothesis stating that the patterns under investigation are generated by random processes. It is this requirement that stirred up the controversy because it is intended to work as a *normative* claim about how research *should* be conducted, which puts extra epistemic responsibility on the side of competitionists.

We should be careful about the relationship between Connor and Simberloff's concern about evidence and their normative claim about using null-model-based hypothesis testing. First, while null-model-based hypothesis testing might be useful by working as a way to challenge competitionist hypotheses of ecological patterns, its usefulness is not a necessary condition in order to criticize the "competitionist paradigm" with respect to the lack of compelling evidence. In other words, even if it turns out that the technical problems of constructing an appropriate null model cannot be solved, and null-model-based hypothesis testing is not useful at all, Connor and Simberloff's call for more compelling evidence will remain valid. Second, if Connor and Simberloff's primary concern is about the evidence for competitionist explanations, then null-model-based hypothesis testing is neither a sufficient nor a necessary way for competitionists to deal with this critique. It is not necessary because competitionists can reply to their critics by collecting more compelling evidence through detailed autecological study, field observation, or experimentation without having to conduct a null-hypothesis test. It is not sufficient because even if competitionists successfully reject the null hypothesis, they might still lack the kind of evidence for their competitionist hypothesis demanded by their critics. Hence, competitionists do not have the responsibility to first conduct null-model-based hypothesis testing before pursuing their own hypotheses.

In sum, by focusing their attention excessively on null hypotheses and null models, both sides, including Connor and Simberloff themselves, deviate from the initial problem of the lack of direct evidence. Although Connor and Simberloff are absolutely right in claiming that more direct evidence is needed in order to support the interspecific competition hypothesis, null-model-based hypothesis testing is not the way to obtain the kind of evidence that they demand.

6. Conclusion

Null-model-based hypothesis testing has been used in many fields of biology, but its usefulness remains a concern. Although proponents of this method usually think that it is analogous to traditional statistical null-hypothesis testing in experimental research, I have shown that this analogy is not justified. Also, the random process hypothesis should not be privileged as a null hypothesis that biologists must try to reject before pursuing other hypotheses. But this does not mean that null-model-based hypothesis testing is necessarily useless. When trying to explain patterns in nature, biologists usually assume that they are produced by specific causal factors, while ignoring the possibility that those seemingly unusual patterns are the result of random processes. By explicitly testing the random process hypothesis, null-model-based hypothesis testing might work as a way to challenge scientists' commonsense judgments about what count as unusual patterns and how those patterns could have come to be. Despite this possible use, researchers should also pay attention to the limitation of null-model-based hypothesis testing: the null model cannot be used to test whether a particular process or mechanism is responsible for the formation of certain patterns; it lacks the potential for dealing with other problems via further development; it cannot provide direct evidence for any non-random hypothesis.

References

Bausman W, Halina M (2018) Not null enough: pseudo-null hypotheses in community ecology
and comparative psychology. Biol Philos 33:1–20

Bausman WC (2018) Modeling: Neutral, Null, and Baseline. Philos Sci 85:594–616

Chase JM, Leibold MA (2003) Ecological niches: linking classical and contemporary
approaches. University of Chicago Press, Chicago and London

Colwell RK, Winkler DW (1984) A Null Model for Null Models in Biogeography. In: Strong Jr
DR, Simberloff D, Abele LG, Thistle AB (eds) Ecological Communities: Conceptual Issues
and the Evidence. Princeton University Press, Princeton, NJ, pp 344–359

Connor EF, Collins MD, Simberloff D (2013) The checkered history of checkerboard
distributions. Ecology 94:2403–2414

Connor EF, Collins MD, Simberloff D (2015) The checkered history of checkerboard
distributions: reply. Ecology 96:3388–3389

Connor EF, Simberloff D (1979) The Assembly of Species Communities: Chance or
Competition ? Ecology 60:1132–1140

Connor EF, Simberloff D (1983) Interspecific Competition and Species Co-Occurrence Patterns
on Islands: Null Models and the Evaluation of Evidence. Oikos 41:455–465

Connor EF, Simberloff D (1984) Neutral models of species' co-occurrence patterns. In: Strong Jr
DR, Simberloff D, Abele LG, Thistle AB (eds) Ecological Communities: Conceptual Issues
and the Evidence. Princeton University Press, Princeton, NJ, pp 316–331

Diamond J, Pimm SL, Sanderson JG (2015) The checkered history of checkerboard distributions:
comment. Ecology 96:3386–3388

Diamond JM (1975) Assembly of Species Communities. In: Cody ML, Diamond JM (eds)

Ecology and Evolution of Communities. Harvard University Press, Cambridge, MA and London, England, pp 342–444

Diamond JM, Gilpin ME (1982) Examination of the "Null" Model of Connor and Simberloff for Species Co-occurrences on Islands. Oecologia 52:64–74

Fisher RA (1925) Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh and London

Fisher RA (1926) The arrangement of field experiments. J Minist Agric 33:503–513

Fisher RA (1935) The Design of Experiments. Oliver and Boyd, Edinburgh and London

Gilpin ME, Diamond JM (1984) Are species co-occurrences on islands non-random, and are null hypotheses useful in community ecology? In: Strong Jr DR, Simberloff D, Abele LG, Thistle AB (eds) Ecological Communities: Conceptual Issues and the Evidence. Princeton University Press, Princeton, NJ, pp 297–315

Gotelli NJ, Graves GR (1996) Null Models in Ecology. Smithsonian Institution Press, Washington and London

Gotelli NJ, McGill BJ (2006) Null versus neutral models: What's the difference? Ecography 29:793–800

Harvey PH (1987) On the Use of Null Hypotheses in Biogeography. In: Nitechi MH, Hoffman A (eds) Neutral Models in Biology. Oxford University Press, New York and Oxford, pp 109–118

Hubbell SP (2001) The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, Princeton and Oxford

Hubbell SP (2006) Neutral Theory and the Evolution of Ecological Equivalence. Ecology 87:1387–1398

Lewin R (1983) Santa Rosalia Was a Goat. Science 221:636–639

MacArthur R (1972) Geographical Ecology: Patterns in the Distribution of Species. Harper & Row, Publishers, Inc., New York

Rathcke BJ (1984) Patterns of Flowering Phenologies: Testability and Causal inference Using a Random Model. In: Strong Jr DR, Simberloff D, Abele LG, Thistle AB (eds) Ecological Communities: Conceptual Issues and the Evidence. Princeton University Press, Princeton, pp 383–396

Rosindell J, Hubbell SP, Etienne RS (2011) The Unified Neutral Theory of Biodiversity and Biogeography at Age Ten. Trends Ecol Evol 26:340–348

Sanderson JG, Pimm SL (2015) Patterns in Nature: The Analysis of Species Co-occurences. The University of Chicago Press, Chicago and London

Schelling TC (1978) Micromotives and Macrobehavior. W. W. Norton & Company, New York and London

Sloep PB (1986) Null Hypotheses in Ecology : Towards the Dissolution of a Controversy. Philos Sci 1:307–313

Sober E (1988) Reconstructing the Past: Parsimony, Evolution, and Inference. The MIT Press, Cambridge, Massachusetts, and London, England

Sober E (1994) Let's Razor Ockham's Razor. In: From a Biological Point of View. Cambridge University Press, Cambridge, pp 136–157

von Bertalanffy L (1968) General System Theory: Foundations, Development, Applications. George Braziller, New York