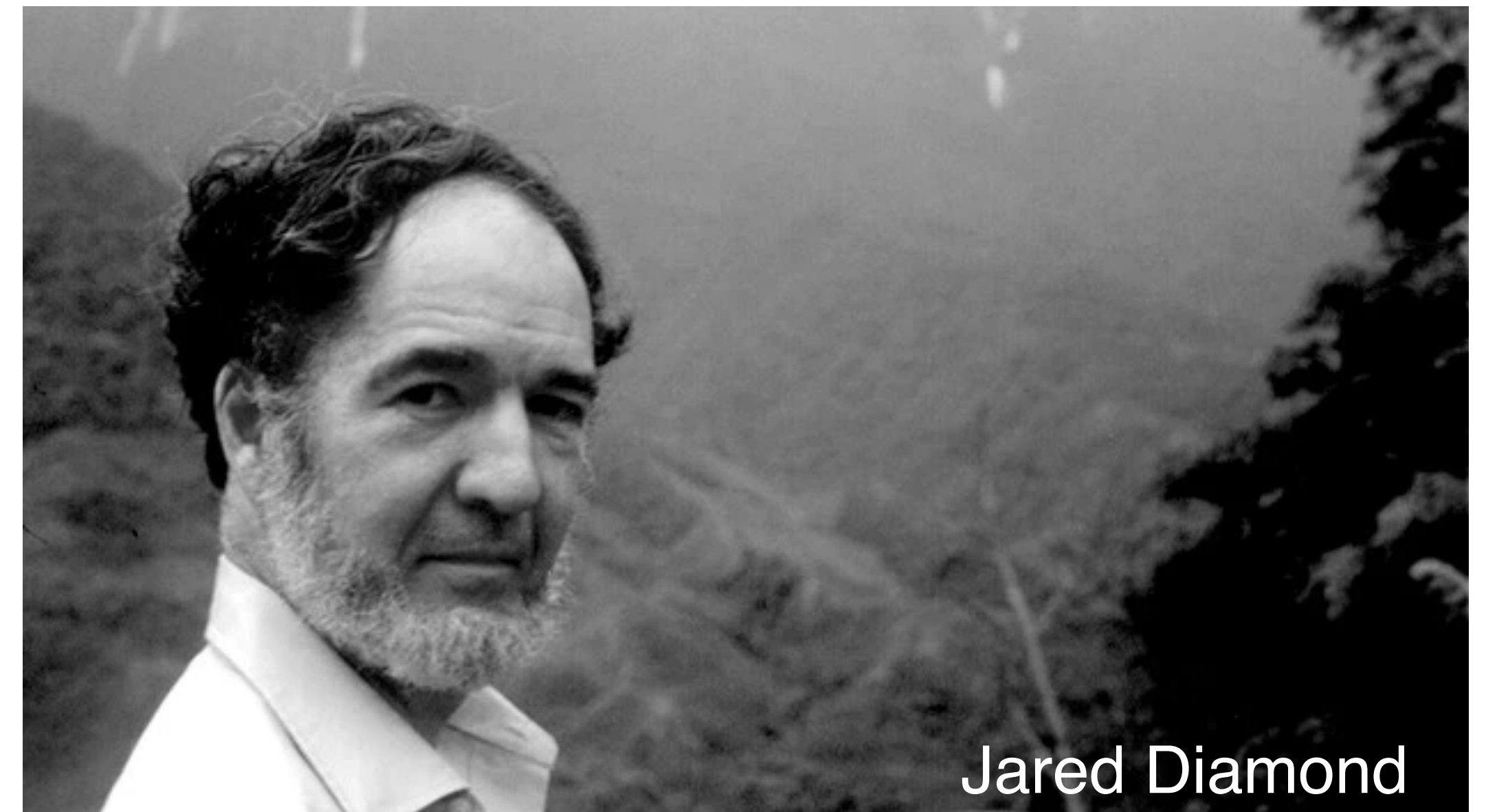


null models

foreword: a tale of islands



Jared Diamond

assembly rules: forbidden species combinations



Bismarck black myzomela
(found in 23 islands)

exclusion →



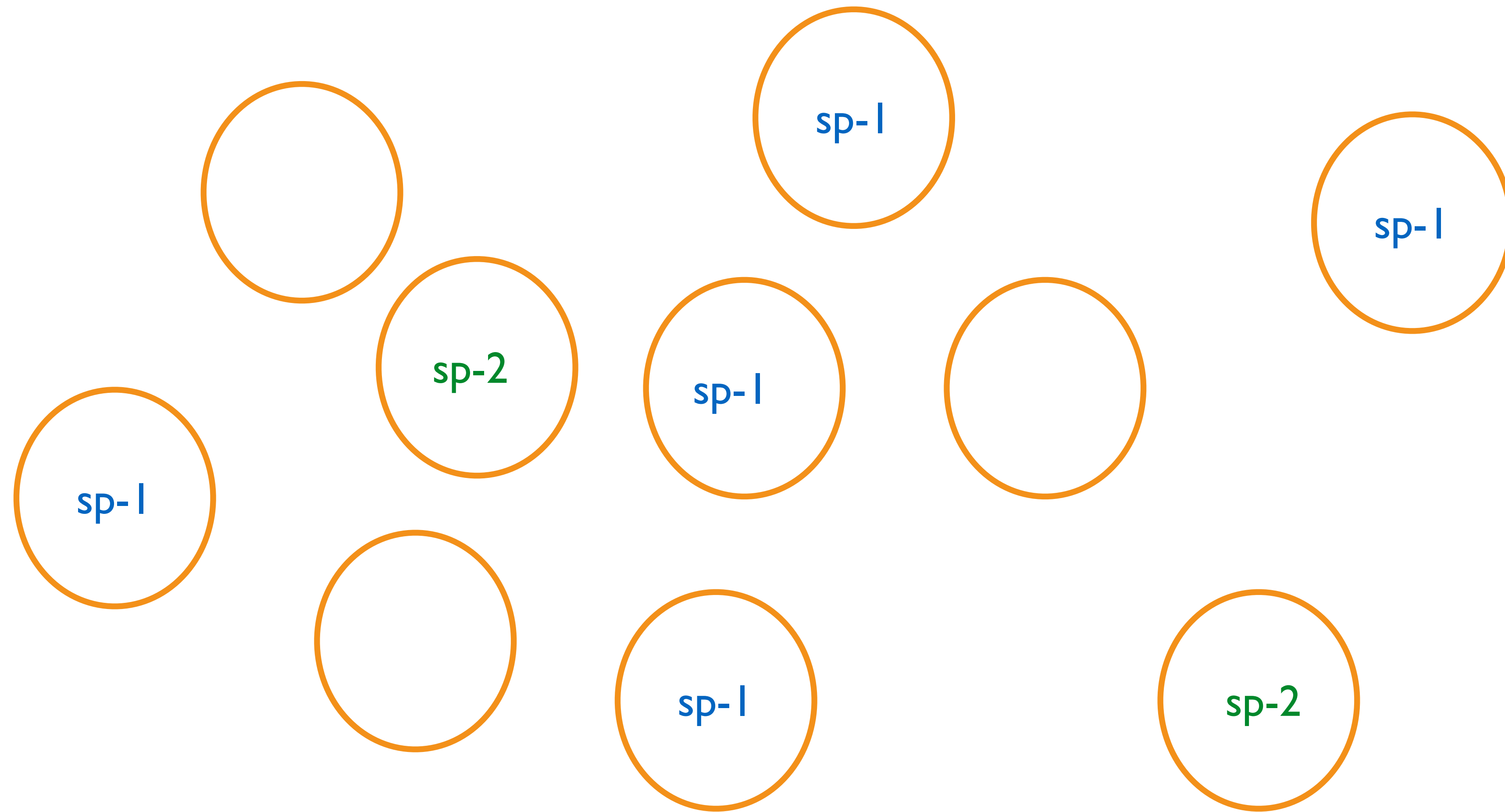
black sunbird
(found in 14 islands)



Daniel Simberloff

“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”

(Connor and Simberloff, 1979, *Ecology* **60**: 1132-1140)

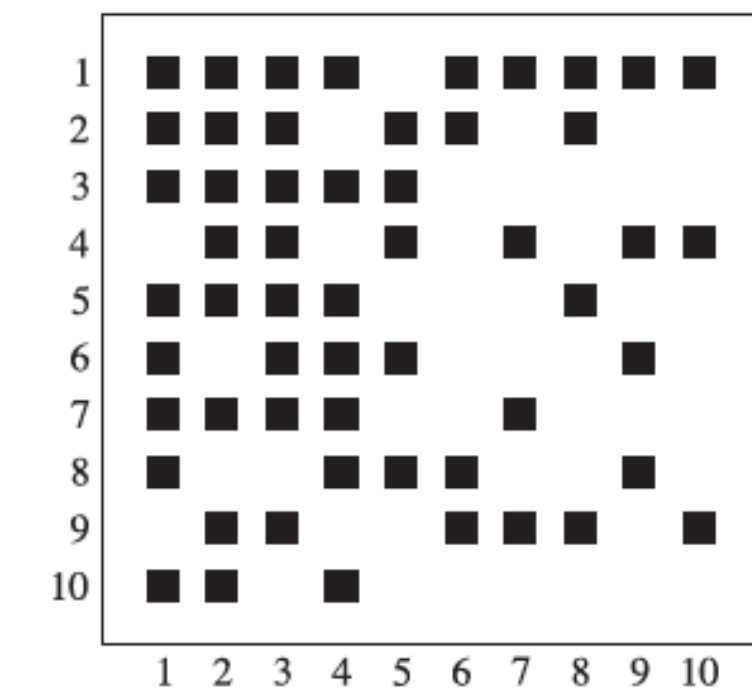
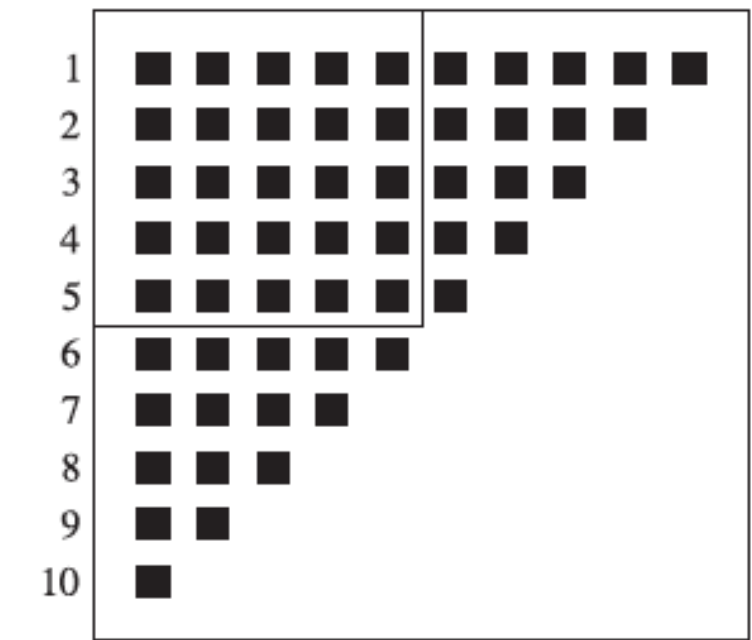
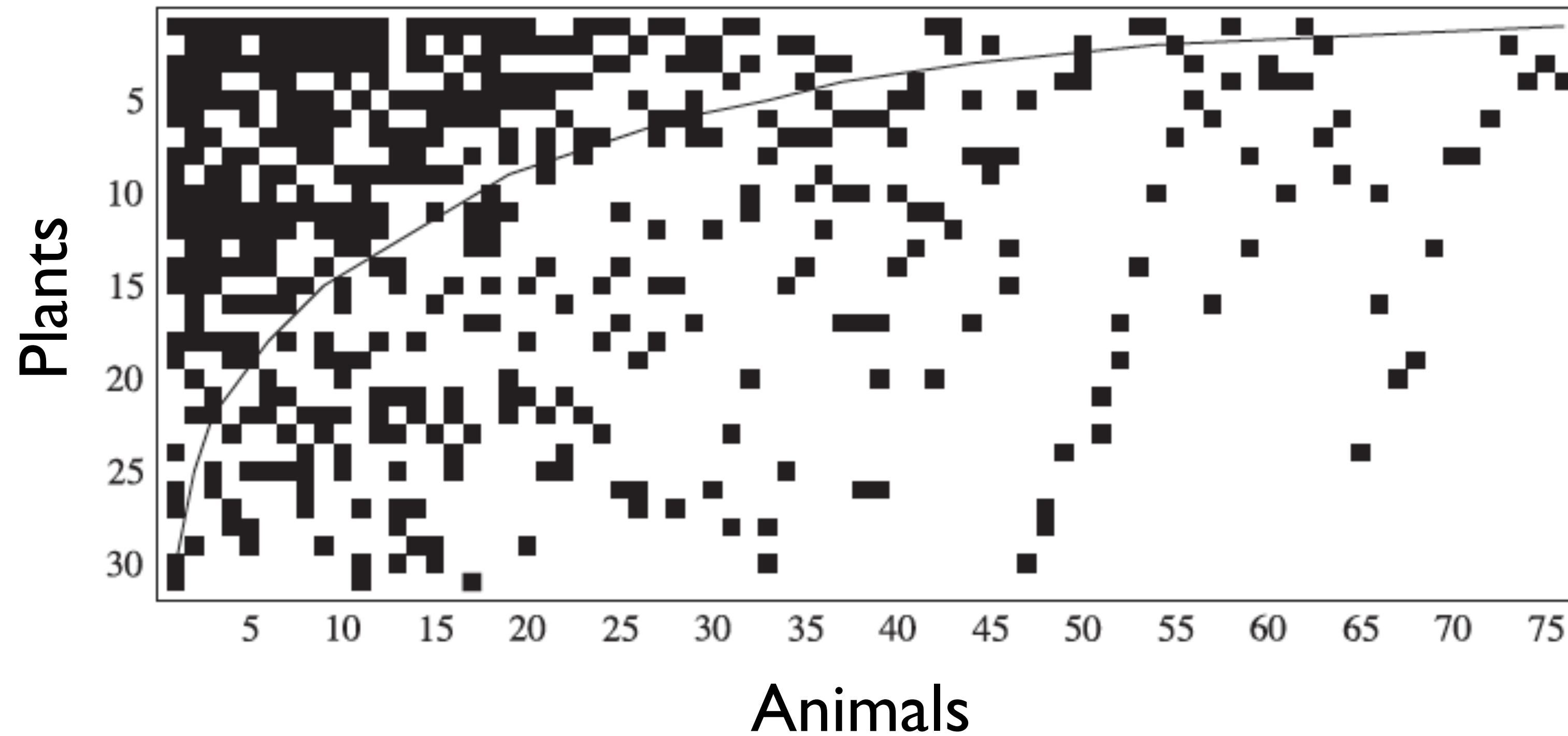


what is the expectation?

$N = 10$ islands; fraction sp-1 = 0.5; fraction sp-2 = 0.2

Probability of overlap if they are independent = $P_1 * P_2 = 0.1$

falsifying a null hypothesis in network research



$N = 0.74, \dots$ so what?

null models

a null model is an appropriate randomisation of the observed matrix intended to serve as a benchmark to assess the significance of a given pattern. It is a pattern-generating model that deliberately excludes a mechanism of interest, and so it serves to test whether the observed level of structure can be explained out of chance.

I. equiprequent null model

each cell ij has the same probability of having a “I” given by the total fraction of “I”s in the observed matrix

I	I	I	I	I
I	I	I		
I	I			
I				
I				

observed

	I		I	
I		I		I
	I		I	I
I	I			
		I	I	

randomization

$$\rho_{ij} = \frac{\text{total number of "I"s}}{\text{potential number of "I"s}} = \frac{12}{25} = 0.48$$

2. equiproquent column null model

the probability ρ_{ij} of drawing an interaction in cell ij is the fraction of “I”s in column j

$$\rho_{ij} = \begin{array}{l} \text{fraction of “I”s} \\ \text{in column } j \end{array} = \frac{1}{n} \sum_{i=1}^n M_{ij}$$

I	I	I		I
I	I	I		
I	I			
			I	
I				

observed

		I?		

randomization

$$\rho_{2,3} = \frac{2}{5} = 0.4$$

3. probabilistic cell null model

the probability ρ_{ij} of drawing an interaction in cell ij is proportional to the degree of both row i and column j

fraction of "I"s in row i fraction of "I"s in column j

$$\rho_{ij} = \frac{p_i + q_j}{2}$$

I	I	I		I
I	I	I		
I	I			
			I	
I				

observed

$$\rho_{2,3} = \frac{3/5 + 2/5}{2} = 0.5$$

		I?		

randomization

$$p_i = \frac{1}{n} \sum_{j=1}^n M_{ij} \quad q_j = \frac{1}{m} \sum_{i=1}^m M_{ij}$$

4. swap null model

it maintains exactly the number of “1”s in both rows and columns by sequentially reshuffling 2x2 sub matrices with the same row and column total as follows:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

1	1	1	0	1
1	1	1		
1	1			
0			1	
1				

observed

0	1	1	1	1
1	1	1		
1	1			
1			0	
1				

randomization

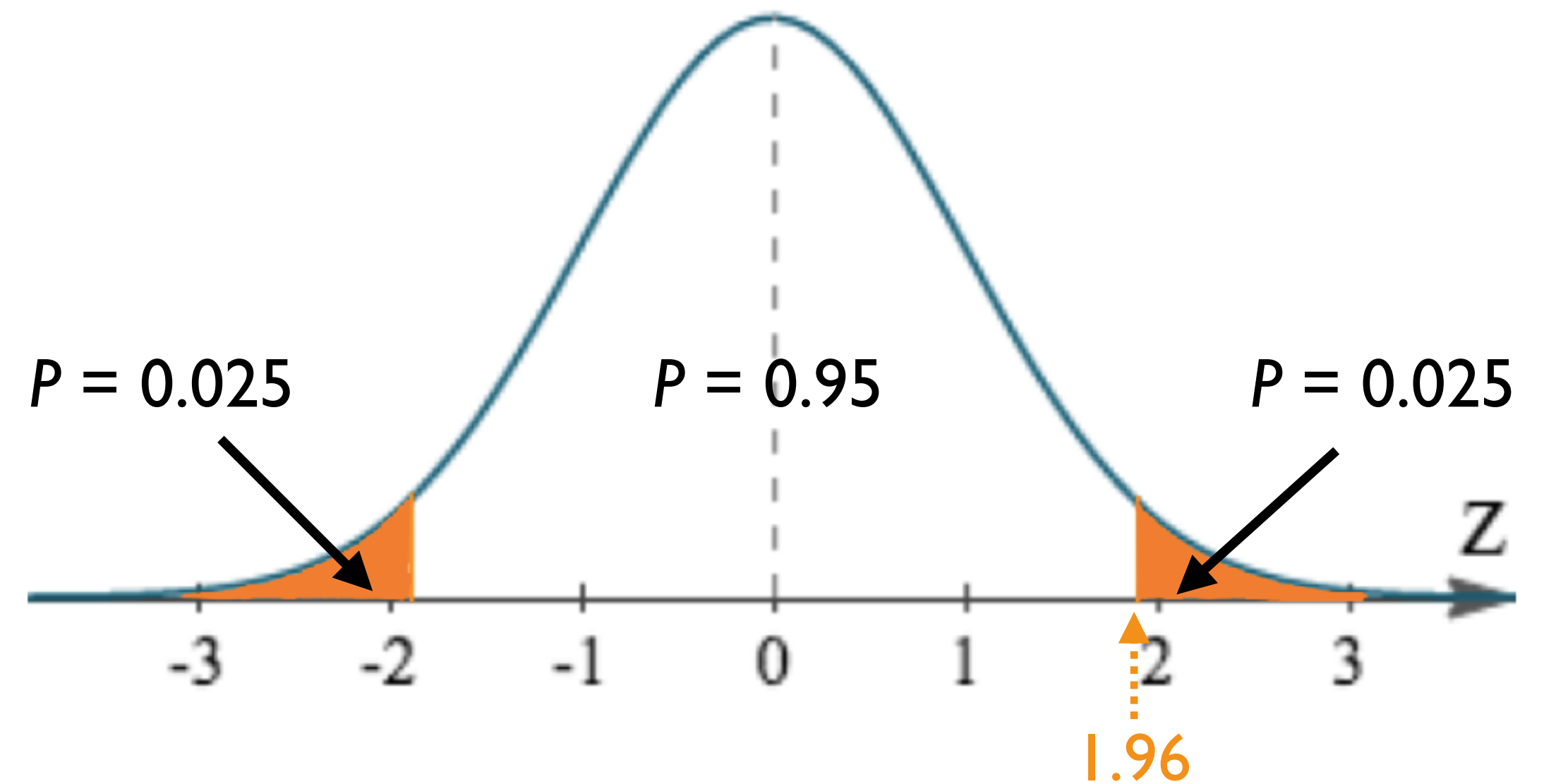
significance of a network pattern

observed value

average across replicates

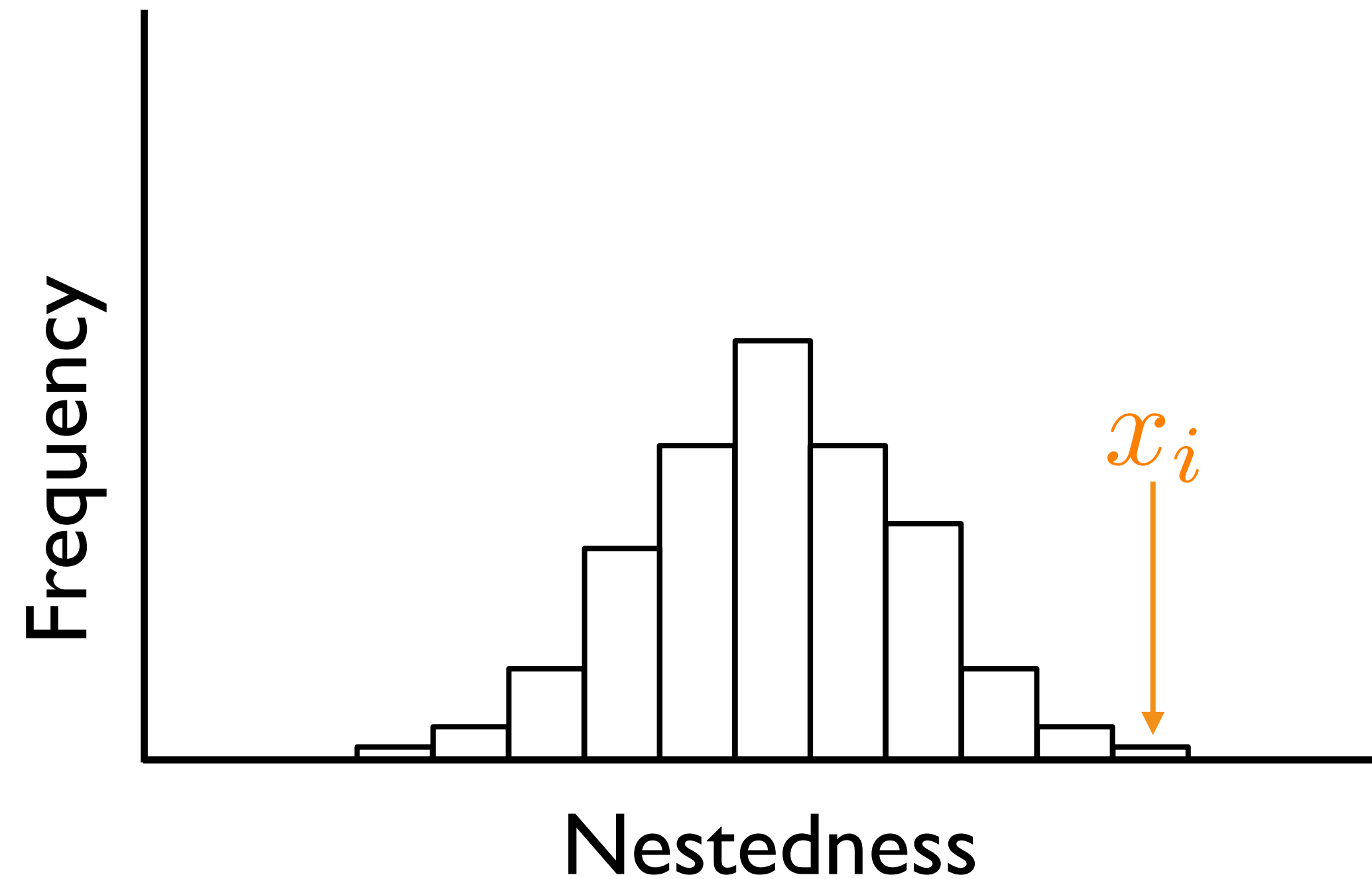
$$Z = \frac{x_i - \bar{x}}{\sigma}$$

standard deviation across replicates



- a z-score is a very useful statistic for calculating the probability of a value occurring within a normal distribution with a mean of 0 and a standard deviation of 1.
- a pattern is statistically significant ($P < 0.05$) if $Z > 1.96$
- a p -value tells us how likely it is that the pattern we had seen could have arisen by chance.

significance of a network pattern (II)



significance of nestedness (p -value): relative fraction of randomisations with a value of nestedness higher or equal than the observed one (e.g., 2 out of 100, $P = 0.02$)

comparing across networks

$$z = \frac{x_i - \bar{x}}{\sigma}$$

- a z-score is a very useful statistic for normalising a value and therefore making it comparable across systems. The raw value of nestedness, for example, depends on matrix size, shape, and filling. It can not be compared across communities. Relative nestedness (z-score) can be used to compare across networks (not perfect, but good enough).

which null model should one use?

- keep in mind what do we want to fix (e.g., only total number of “I”s, number of “I”s per rows, number of “I”s per rows and columns, ...)
- understand their goodness/limitations in the context of the trade-off between type I error (mistakenly rejecting a true null hypothesis) and type II error (mistakenly accepting a wrong null hypothesis)

hypothesis testing and type I vs type II error

decision about null hypothesis (H_0)

		null hypothesis (H_0) is	
		true	false
accept	correct	type II error probability = β	
reject	type I error probability = α	correct	

Example:

H_0 : water is contaminated

H_1 : there is no contaminant

The type I error rate or significance level is the probability of rejecting the null hypothesis given that it is true. Often, the significance level is set to 0.05 (5%),

A type II error occurs when the null hypothesis is false, but erroneously fails to be rejected. The rate of the type II error is related to the power of a test.

p -values and the prosecutor's fallacy

	match	no match
guilty	1 person	0 people
innocent	5 people	50,000,000 people

50,000,000 fingerprints in FBI database

if, according to FBI, the $p(\text{match} | \text{innocent})$ is $1/10,000,000$, we should expect 5 matches of innocent people in the database

prosecutor's comparison:
 $p(\text{match} | \text{innocent}) = 1/10,000,000$
(p -value)

A p -value tells us how likely it is that the pattern we had seen could have arisen by chance alone.

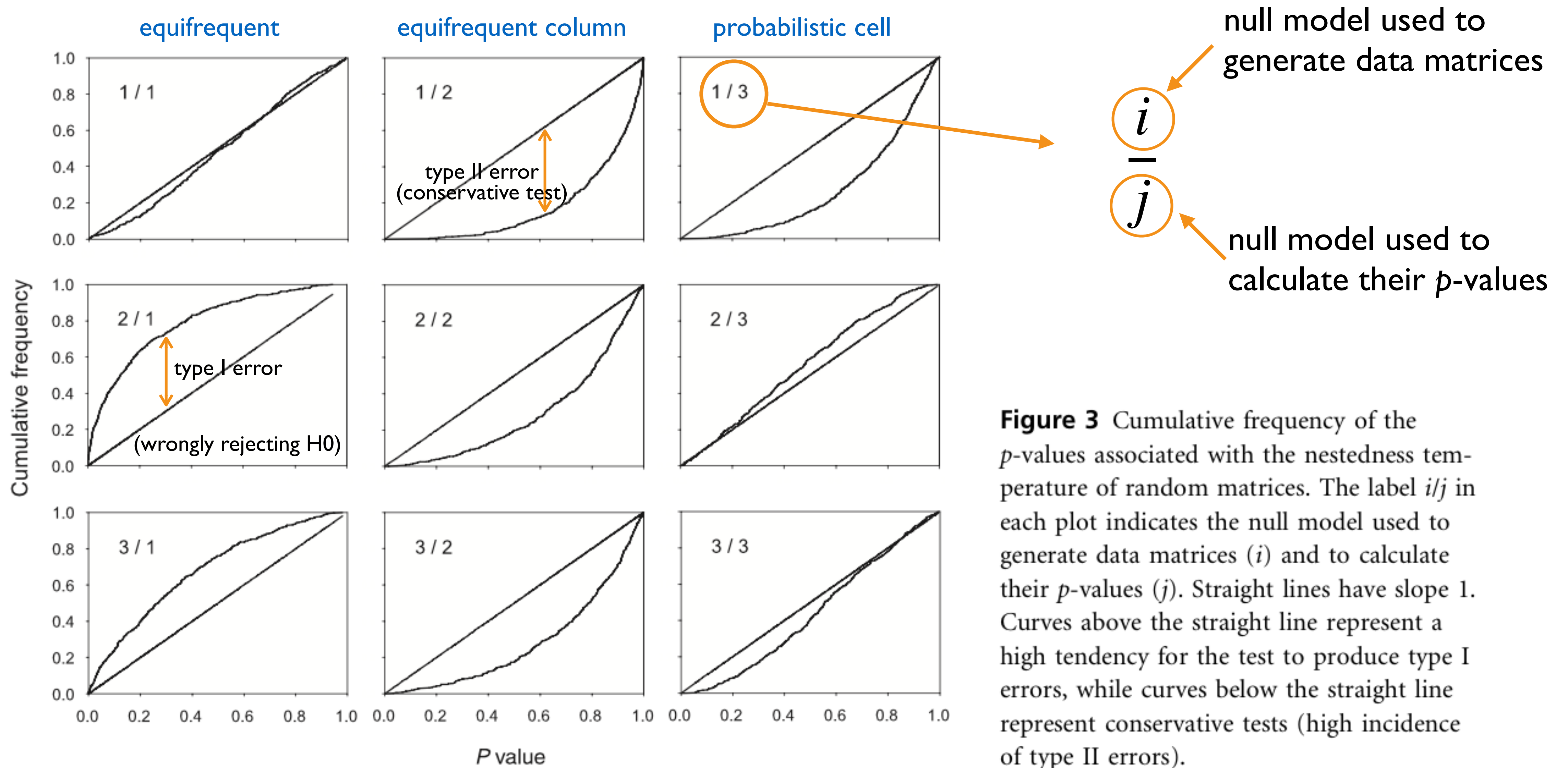
p -values and the prosecutor's fallacy

	match	no match
guilty	1 person	0 people
innocent	5 people	50,000,000 people

defence's comparison:
 $p(\text{innocent}|\text{match}) = 5/6$

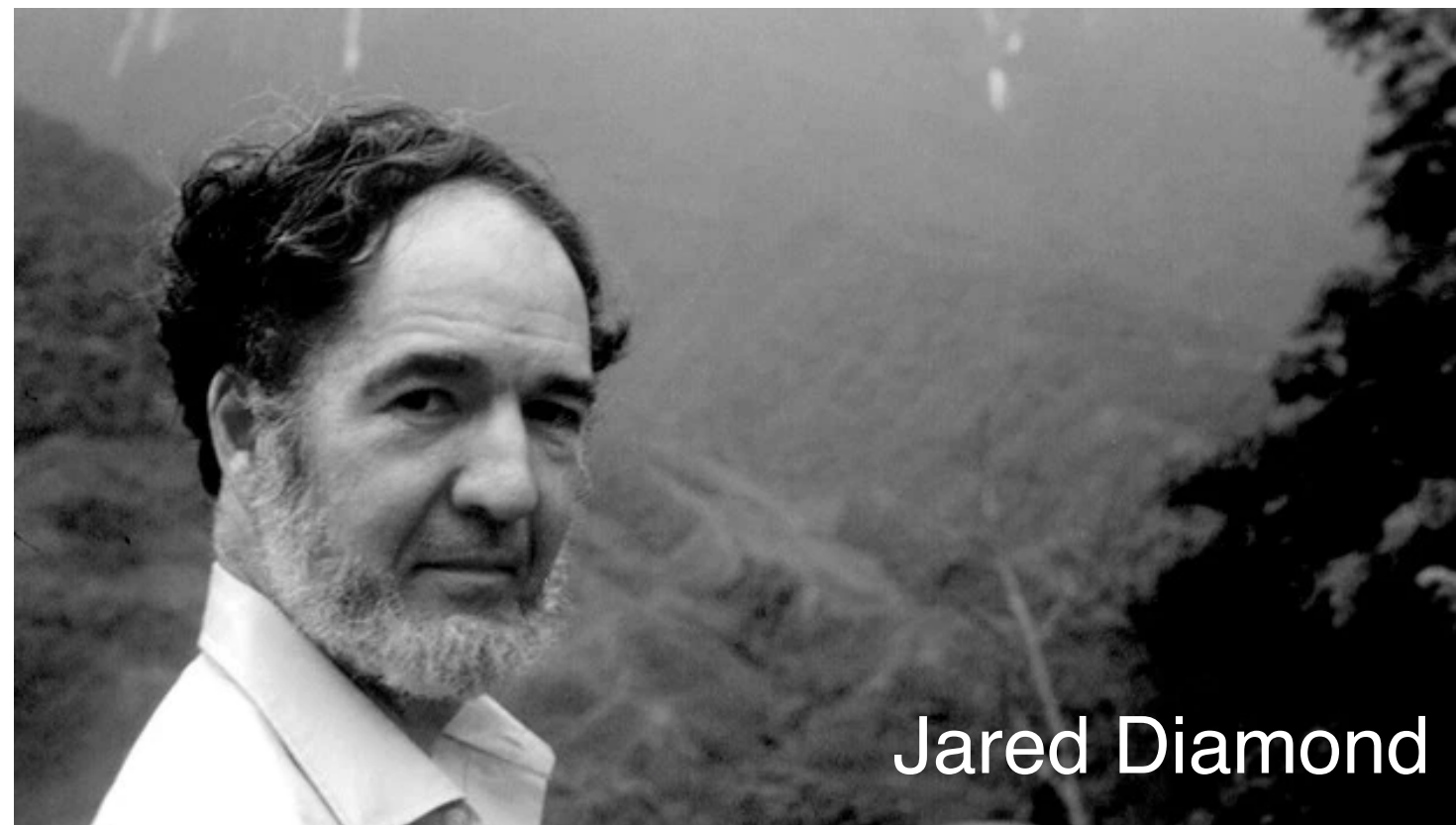
50,000,000 fingerprints in FBI database

if, according to FBI, the $p(\text{match}|\text{innocent})$ is $1/10,000,000$, we should expect 5 matches of innocent people in the database



(Rodríguez-Gironés and Santamaría 2006)

back to Diamond's assembly rules



Jared Diamond

Ecology, 83(8), 2002, pp. 2091–2096
© 2002 by the Ecological Society of America

SPECIES CO-OCCURRENCE: A META-ANALYSIS OF J. M. DIAMOND'S ASSEMBLY RULES MODEL

NICHOLAS J. GOTELLI¹ AND DECLAN J. MCCABE²

Department of Biology, University of Vermont, Burlington, Vermont 05405 USA

Abstract. J. M. Diamond's assembly rules model predicts that competitive interactions between species lead to nonrandom co-occurrence patterns. We conducted a meta-analysis of 96 published presence–absence matrices and used a realistic “null model” to generate patterns expected in the absence of species interactions. Published matrices were highly nonrandom and matched the predictions of Diamond's model: there were fewer species combinations, more checkerboard species pairs, and less co-occurrence in real matrices than expected by chance. Moreover, nonrandom structure was greater in homeotherm vs. poikilotherm matrices. Although these analyses do not confirm the mechanisms of Diamond's controversial assembly rules model, they do establish that observed co-occurrence in most natural communities is usually less than expected by chance. These results contrast with previous analyses of species co-occurrence patterns and bridge the apparent gap between experimental and correlative studies in community ecology.

Key words: community assembly rules; meta-analysis; null models; presence–absence matrix; species co-occurrence.