Design and analysis of experiments testing for biodiversity effects in ecology

R. A. Bailey[a], Julia Reiss[b]

[a]*School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK*
[b]*Department of Life Sciences, Whitelands College, University of Roehampton, London SW15 4JD UK*

Abstract

It is now widely believed that biological diversity is good for the natural environment. One way that ecologists test this is to place random collections of species in mini-environments and then measure some outcome. Statisticians have been working with fresh-water ecologists to improve this in two ways. The first is that the subsets of species are carefully chosen, not random. The second is that a nested family of plausible models is fitted. The results of three experiments suggest that biodiversity can have no effect at all, but that there are other plausible underlying mechanisms.

Implications for the design of such experiments, the understanding of the family of models, and the analysis of the data are discussed.

Keywords: Biodiversity, Design of experiments, Family of models, Hasse diagram
2008 MSC: 62K99, 62J10, 62P12

1. Introduction

There are many experiments in ecology whose results seem to suggest that biodiversity is generally a good thing. Often a large collection of different species is considered, and random subsets of these species are used

Email addresses: r.a.bailey@qmul.ac.uk (R. A. Bailey), Julia.Reiss@roehampton.ac.uk (Julia Reiss)
URL: http://www.maths.qmul.ac.uk/~rab (R. A. Bailey), http://www.roehampton.ac.uk/Staff/JuliaReiss (Julia Reiss)

as treatments and put into some artificial set-up that mimics nature. The measured response is some eco-desirable outcome. Very often, the conclusion is that the greater the number of different species the better the outcome.

For example, Bell et al. (2005) used random subsets from a collection of 72 bacterial species. These were grown on sterile leaf discs in a sterile fluid to resemble bacterial assemblages found on decomposing beech leaves. The authors found that bacteria were "more active" when species richness was high (they showed higher respiration); as a consequence species-rich assemblages will be able to decompose leaves faster than species-poor ones. Another example comes from Cardinale (2011), who showed that biodiversity of stream algae improves stream water quality because diverse algae assemblages can take greater advantage of differences in their environment (so called *niche partitioning*); that is, they can grow better when species richness is high and take up more pollutants as a consequence. He used a laboratory set-up consisting of cultured algae that were added to artificial "mini-streams".

The experiment by Cardinale was very much in the tradition of those plant ecologists who were the first to design "biodiversity and ecosystem functioning" (B-EF) experiments in the 1990s. These experiments specifically addressed the effects of biodiversity (species richness) on particular "ecosystem processes", such as nutrient uptake (see Loreau et al., 2002). Because of global change and the species loss it causes, B-EF experiments are now a major research topic in Ecology. These experiments all vary in terms of the species used (often they use either plant or animal species and do not mix them), response variables measured (for example, how productive a plant assemblage is or how good an animal assemblage is at using food resources) and in terms of their statistical analysis and experimental design (for example, short term studies where species do not reproduce or long term studies where species grow and reproduce). Many studies have "failed" to show that biodiversity is important for ecosystem processes (for example, McKie et al., 2008; Perkins et al., 2010). In general, the statistical analysis does not seem to address the mechanisms that could govern biodiversity effects or explain why no biodiversity effects can be observed.

The experiments we describe here were specifically designed to address how mixtures of animals "perform" with regard to how efficiently they consume food and generate fine particular matter (that can be used by other organisms as food) in a "short term" experiment. In addition, two of the experiments described addressed not only the role of species richness, but also that of size within species.

In Section 2 we describe three experiments in which the subsets of species were carefully and deliberately chosen. Section 3 gives the models that we fitted to the data. These included not only the 'biodiversity' model where the number of different species is considered to be a quantitative factor, but also models that would be more familiar to people running experiments on mixtures of different ingredients—see Cornell (2002). We found that Hasse diagrams helped the biologists to understand the relationships between the different models.

Section 4 briefly summarizes our conclusions from the data analysis, and suggests a new graphical method of summarizing the analysis-of-variance table. Finally, Section 5 considers some questions about how such experiments should be designed.

2. The experiments

The first two experiments are described by Reiss et al. (2011). Six types of freshwater organisms called invertebrate detritivore shredders were used. These types were three species, with two size classes within each species. For simplicity, they are referred to here as A, B, ..., F. The experimental unit was a jar. Twelve organisms were put into each jar. The treatments were thus the combinations of types put into each jar.

Table 1 shows the treatments used in the first experiment. There were six treatments called *monocultures* where all twelve organisms in the jar were of the same type. There were 15 further treatments called *dicultures*: in these, there were six organisms of one type and six of another. Finally, there were 20 treatments called *tricultures* in which three different types of organism were used, four of each. Thus there were 41 treatments altogether. (There was also a 'control' treatment with no organisms, but we ignore that here.)

The experiment was carried out in four blocks of 41 jars. Each block was in a slightly different place in the laboratory, and it was expected that there would be block-to-block differences because of differences in temperature, ambient lighting etc. Carefully measured amounts of stream water and of alder leaf litter were put into each jar. Then one treatment was added to each jar, in such a way that each treatment occurred in exactly one jar in each block. The jars were left for 28 days: then the amount of leaf litter eaten was measured. A secondary measure was the quantity of fine particulate organic matter (FPOM) in the jar after 28 days.

Number	Treatment	Name	Example	Richness Level
6	$A, \ldots, F$	monoculture	12 of type A	1
15	$AB, \ldots, EF$	diculture	6 of A, 6 of B	2
20	$ABC, \ldots, DEF$	triculture	4 of A, 4 of B, 4 of C	3

Table 1: The 41 treatments in the first experiment

Number	Treatment	Name	Example	Evenness Level
6	$A, \ldots, F$	monoculture	12 of type A	1
15	$AB, \ldots, EF$	even diculture	6 of A, 6 of B	2
30	$AAB, \ldots, FEE$	uneven diculture	8 of A, 4 of B	3

Table 2: The 51 treatments in the second experiment

The first experiment was designed to find out not only if the responses were affected by the number of types of organism present but also if it mattered which combinations of types were present. The second experiment took this a little further by replacing the tricultures by so-called *uneven dicultures*, in which there were eight organisms of one type and four of the other. Thus there were the 51 treatments shown in Table 2. This experiment was run in three blocks of 51 jars. The same two responses were measured as in the first experiment.

The third experiment is described by Reiss et al. (2010). It concerned a slightly different ecological setting, and different organisms, but was otherwise similar to the first experiment, in that the treatments were monocultures, dicultures and tricultures. However, there were now seven types of organism rather than six. There are 35 possible tricultures from seven organisms. In order to cut down the number of treatments without favouring any one type over another, only seven tricultures were used, those corresponding to the lines and circle in Figure 1: see Table 3. This collection of lines and a circle is known as the Fano plane: see van Lint and Wilson (2001). They have the property that each pair of types occurs in exactly one triculture: in other words, these seven tricultures form a *Steiner triple system*, or *balanced incomplete-block design* with blocks of size 3. (These abstract blocks are not to be confused with those in which the experiment was run.)

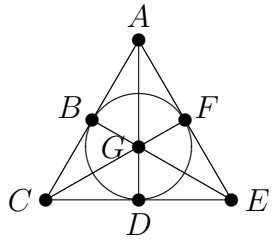

Figure 1: The lines and circle show the seven tricultures used in the third experiment

Number	Treatment	Name	Example	Richness Level
7	$A, \ldots, G$	monoculture	12 of type A	1
21	$AB, \ldots, FG$	diculture	6 of A, 6 of B	2
7	see Figure 1	triculture	4 of A, 4 of B, 4 of C	3

Table 3: The 35 treatments in the third experiment

The experiment was conducted in three blocks of 35 jars. Two responses were measured, leaf mass loss and FPOM, the same as in the first two experiments.

3. Families of models

In each experiment it was assumed that the response Y_ω in jar ω had the following form:

$$Y_\omega = \tau_{f(\omega)} + \beta_{g(\omega)} + \varepsilon_\omega. \tag{1}$$

Here $f(\omega)$ denotes the treatment in jar ω and $g(\omega)$ denotes the block containing jar ω. In addition, τ_θ is an unknown constant depending on treatment θ, while the β_Γ and the ε_ω are uncorrelated normal random variables with zero expectation; the random variables ε_ω all have the same unknown variance σ^2, while the random variables β_Γ all have variance σ_B^2, which may be different from σ^2. Thus the expectation part of Equation (1) is just the parameter $\tau_{f(\omega)}$.

In the first and third experiments there was a factor called Richness whose levels were the numbers of different types of organism present: see Tables 1 and 3. The second experiment had a similar factor called Evenness: see Table 2.

In the analysis of data from biodiversity experiments, it is common to fit Richness. Sometimes this is treated as a factor, with one parameter per

level: for example, see Jonsson and Malmqvist (2005), Bell at al. (2009) or Perkins et al. (2010). Other authors, such as Bell et al. (2005), treat it as a linear covariate with one parameter for the intercept and another for the slope (for example when species richness is estimated during rather than at the start of the experiment; see also Schmid et al., 2002).

We now describe the linear models that we considered for expectation in the first experiment. Because they are linear models, we can identify each one with the subspace of $\mathbb{R}^{164}$ containing all possible vectors of fitted values. The dimension of the subspace will be called the dimension of the model.

The model most familiar to ecologists has one parameter for each level of Richness. We called this model Richness: it has dimension 3.

A model more obvious to statisticians has parameters, α_A, ..., α_F, one for each type of organism. The expected response on monoculture A is α_A; that on diculture AB is $(\alpha_A + \alpha_B)/2$, while that on triculture ABC is $(\alpha_A + \alpha_B + \alpha_C)/3$. Thus there are six linearly independent parameters, and the model has dimension 6. We called this model Type.

There is another way of writing this model. Let x_1, ..., x_6 denote the number of organisms of type A, ..., type F, and put $a_1 = \alpha_A/12$, ..., $a_6 = \alpha_F/12$. Then the expectation in model Type is just

$$a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_6x_6. \tag{2}$$

Because every treatment contained the same number of organisms, there is no need for an intercept in expression (2). This model simply says that each organism of type i eats a_i amount of leaf litter in 28 days, irrespective of the identity of other organisms present.

Bell et al. (2005) include an effect called 'Composition' which is analogous to our model Type. However, they do not consider any model which includes Composition without Richness. Likewise, Model 2 of Kirwan et al. (2009) includes expression (2) only *in addition* to the effect of overall abundance.

Bailey (2008) recommends that if linear models M_1 and M_2 are both considered as possible explanations of the data then their intersection $M_1 \cap M_2$ and their sum $M_1 + M_2$ should also be considered. Note that $\dim(M_1 + M_2) = \dim(M_1) + \dim(M_2) - \dim(M_1 \cap M_2)$. The intersection of the models Richness and Type is just the 1-dimensional model Constant in which τ_θ has the same value for all treatments θ. Their sum is the model Richness + Type, whose dimension is 8: the expectation is similar to that shown in (2), except that changing from one level of Richness to another adds a single constant to all the coefficients a_1, ..., a_6.

This suggests a further model, in which all the coefficients $a_1, \ldots, a_6$ can vary with the level of Richness. It has dimension 18. In some sense, this allows for 'interaction' between Richness and Type, so we call this model Richness $*$ Type. However, even in this model the different types of organism are *additive* in the sense that the amount of leaf litter eaten by an organism of one type is not affected by which other types are present: all that matters is how many other types are present.

Our final model has one parameter per treatment, so it has dimension 41. It does not have any simple expression like (2) or any correspondingly simple biological explanation. It contains all the other models discussed so far, and is the only one which allows a type of organism to be affected by the identities of the other organisms present. The model is called Assemblage Identity by Reiss et al. (2011): here we call it Treatment for brevity.

Bailey (2008) recommends showing the family of considered models in a Hasse diagram. There is one dot for each model. It is useful to show the dimension of each model as well as its name. The diagram also contains edges linking some dots. The convention is that if model M_1 contains model M_2 then the dot for M_1 is higher than the dot for M_2 and there is a chain of generally downwards edges linking the dot for M_1 to the dot for M_2.

Figure 2 shows the Hasse diagram for the models considered in the first experiment.

Six of the models for the second experiment are analogous to those for the first experiment, with the factor Richness replaced by the factor Evenness. However, there is an extra possibility for the uneven dicultures: there may be one parameter c_i for type i when it is in the majority, and a different parameter d_i when it is in the minority. We called this extra model Dominance. It has dimension 23, because $8c_i + 4d_j = 8(c_i + \kappa) + 4(d_j - 2\kappa)$ for any constant κ. The Hasse diagram for the models considered in the second experiment is in Figure 3.

The third experiment used the same models as the first. Of course, the dimensions are slightly different from those shown in Figure 2, because there were seven types instead of six.

Although Hasse diagrams seem to be less familiar to statisticians than to pure mathematicians, they are a very helpful way of showing the relationships between the models. Using such a diagram helps the researcher to think about models rather than parameters, and so to respect marginality (Nelder, 1977, 1994). It may be that biologists find them easier to understand than equations. Kirwan et al. (2009, Figure 2) gave a version of a Hasse diagram

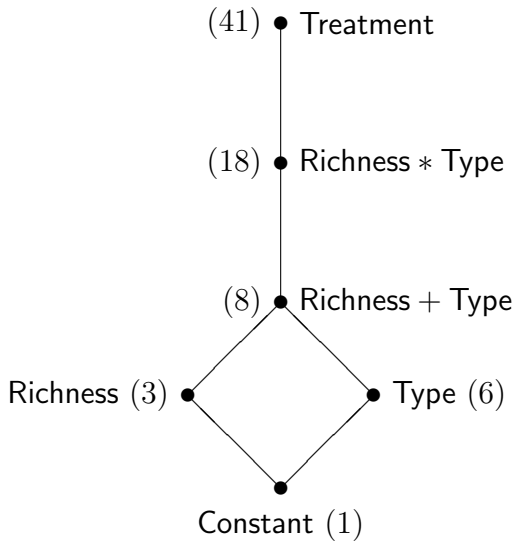

Figure 2: Hasse diagram of expectation models considered in the first experiment: dimensions are shown in parentheses

with the convention that the larger models are at the bottom; also, each edge had an arrow pointing to the larger model. Although biological journals often place much of the technical detail in web appendices, versions of Figures 2 and 3 both appeared in the paper by Reiss et al. (2011).

With one exception (see Section 4), no other models were considered by Reiss et al. (2011). However, we have more recently observed that there are some other plausible models. The data from these three experiments did not need them, but they might be useful in some future experiment.

In the first experiment, the six types actually consisted of two different sizes of each of three different species. This cannot be considered as factorial combinations of factors Size and Species, because the sizes varied with the species, but it does make sense to think of Species as a submodel of the model Type. It has one parameter per species and hence has dimension 3. This leads to the expanded collection of models in Figure 4.

In the third experiment, the seven types actually consisted of three species of shredder, as before, together with four species of fungi. Ingenuity was required to decide what quantity of fungi was equivalent to a single shredder! In this case the three-parameter model Species in Figure 4 could be replaced

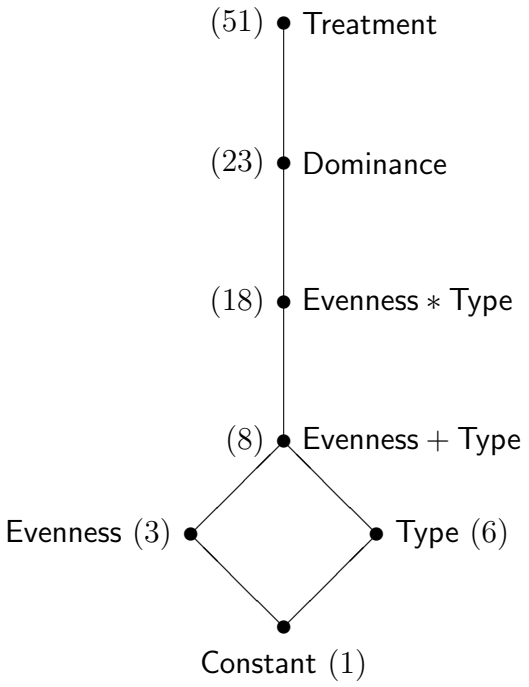

Figure 3: Hasse diagram of expectation models considered in the second experiment: dimensions are shown in parentheses

by a two-parameter model which distinguishes between the shredders and the fungi.

The other plausible model comes from comparing the treatments with the sorts of mixtures of ingredients used in industrial processes. In that context, a diagram like the one in Figure 5 is often used to show the possible proportions of three ingredients: see Atkinson et al. (2007, Chapter 16) or Cornell (2002). The small dots in Figure 5 show all possible ways of forming a collection of twelve organisms if only three types (A, B and C) are available. The seven larger dots indicate the combinations used in the first experiment. The similarity of Figure 5 to Figure 1 is a confusing coincidence.

Expression (2) is linear in the variables $x_1, \ldots, x_6$. It is natural to extend this to include product terms $x_i x_j$ for $i \neq j$, thus giving the model where the

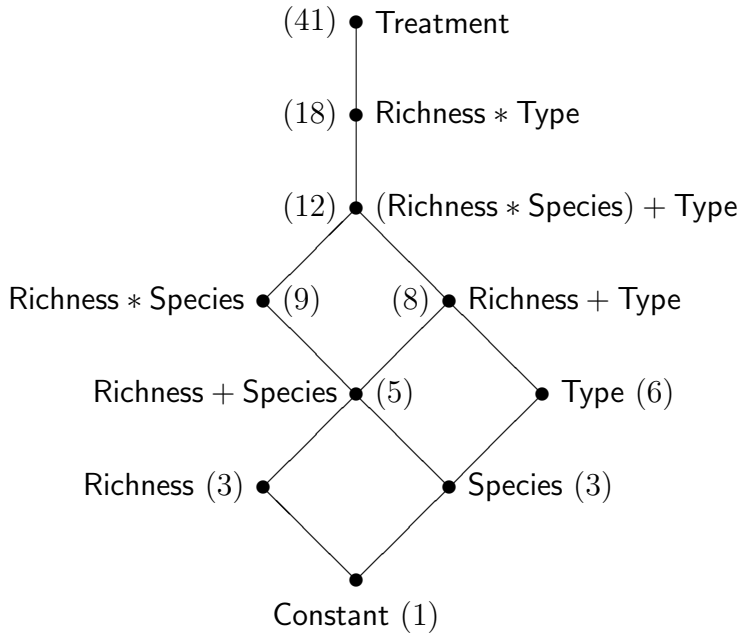

Figure 4: Modification of Figure 2 to allow for the fact that the six types were grouped into three species

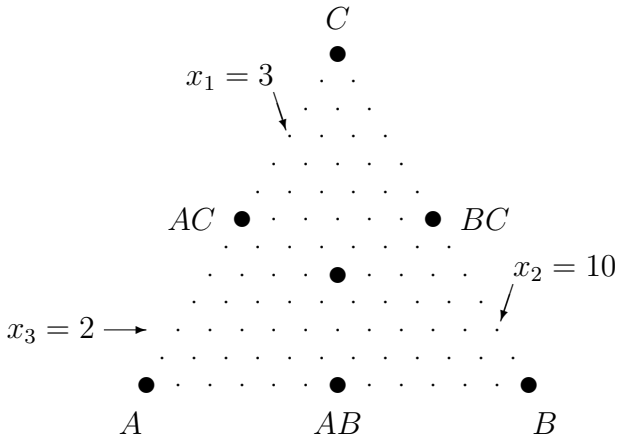

Figure 5: All possible ways of assembling twelve organisms from the three different types A, B and C: the central dot is the triculture ABC

expectation is

$$\sum_{i=1}^{6} a_i x_i + \sum_{i=1}^{5} \sum_{j=i+1}^{6} b_{ij} x_i x_j. \tag{3}$$

(There is no need for quadratic terms, because $x_i^2 = 12x_i - \sum_{j \neq i} x_i x_j$.) The coefficient b_{ij} is sometimes called the *interaction* between types i and j; model (3) contains all of these interactions *in addition* to the effects in model (2). Here we call model (3) Competition for simplicity, but in fact types i and j compete if b_{ij} is negative whereas they facilitate each other if b_{ij} is positive.

The model Richness $\cap$ Competition has two parameters, a and b: the expectation is $12a$, $12a + 36b$ and $12a + 48b$ for monocultures, dicultures and tricultures respectively. The models (Richness $\cap$ Competition) + Type and (Richness + Type) $\cap$ Competition are the same: the expectation is given by formula (3) with $b_{ij} = b$ for all i and j. Kirwan et al. (2009) call this model 'Evenness', but it has no relation to our model Evenness. The model (Richness $*$ Type) $\cap$ Competition has expectation given by (3) with b_{ij} replaced by $b_i + b_j$. Kirwan et al. (2009) call the terms $(b_i + b_j)x_i x_j$ 'additive type-specific contributions to interaction'.

The new expanded collection of models in shown in Figure 6. This looks complicated, but contains fewer models than are needed in a standard factorial design with three treatment factors (Bailey, 2008, Figure 5.11).

It would be possible to go further, and build the collection of models which includes both Species and Competition. Kirwan et al. (2009) go some way towards this by including a version of model (3) in which the coefficients b_{ij} depend only on which species types i and j belong to, but they do not include the simpler model Species. (In their case, it is actually functional group rather than species, as in our third experiment.)

4. Data analysis

Reiss et al. (2011) analysed four data sets: two responses from each of the first and second experiments. In three cases out of the four, the mean square for blocks was at least twice the error mean square; in the fourth case it was of a similar size. Thus the effect of carrying out the experiment in blocks, and allowing for this in the data analysis, was to reduce the error mean square, thus reducing the variance of the estimators of parameters and increasing the power of F tests.

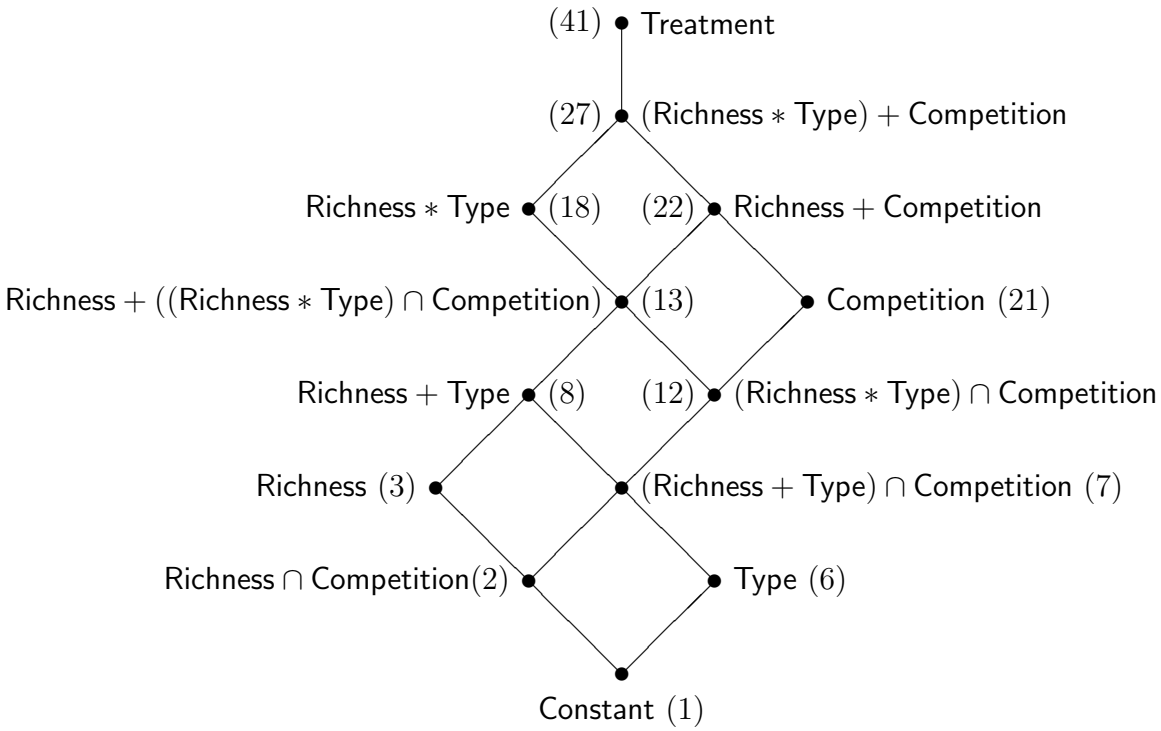

Figure 6: Modification of Figure 2 to include the model in (3)

Since each treatment occurred exactly once in each block, all the expectation models were orthogonal to blocks. It therefore made no difference whether the block effects β_Γ in Equation (1) were regarded as fixed or random.

As Reiss et al. (2011) stated, "each row in the ANOVA table represents not a model but the difference between a larger model and the next smaller one." Initially, one of the biological referees disputed this, but we were able to provide confirmation from other biologists (Grafen and Hall, 2002). In fact, Bell et al. (2005) give a similar statement about models in a single chain. In explaining this statement, we found it helpful to refer to the Hasse diagrams. Indeed, Kirwan et al. (2009) label the edges of their Hasse diagram by the corresponding hypothesis tests.

We now suggest a way of using the Hasse diagram directly to show the information in the ANOVA table. Suppose that there is an edge joining

models M_1 and M_2, where M_1 is larger than M_2. For $i = 1, 2$, let SS_i be the sum of the squares of the fitted values for M_i, and let $d_i = \dim(M_i)$. The ANOVA table will contain one row like the following:

d.f.	SS	MS
$d_1 - d_2$	$\mathrm{SS}_1 - \mathrm{SS}_2$	$(\mathrm{SS}_1 - \mathrm{SS}_2)/(d_1 - d_2)$

We propose drawing an 'anova' version of the Hasse diagram in which the length of the edge between M_1 and M_2 is proportional to the mean square $(\mathrm{SS}_1 - \mathrm{SS}_2)/(d_1 - d_2)$. For data with a single error mean square, the size of this (or of a simple multiple of it) should be shown beside the diagram, to indicate the scale, that is, the denominator for F tests.

Figure 7 shows the 'anova' version of the Hasse diagram in Figure 2, for the response amount of leaf litter eaten. Two results stand out from this diagram, with no need for formal hypothesis testing. The first is that the model Richness does not explain the data *at all*. The second is that the model Type explains the data so well that there is no need for further investigation of any of the larger models.

The results from the other three analyses reported by Reiss et al. (2011) were qualitatively similar.

Figure 8 gives the 'anova' version of the Hasse diagram for the first response in the third experiment. That for the second response is qualitatively similar. Again, the factor Richness appears to play no role, and the model Type explains a large part of the data. However, the line between the models Richness $*$ Type and Treatment is now long enough to warrant further investigation. It was discovered that one of the shredder types was interacting with some of the other types (positively with some, negatively with others).

In these experiments, the expectation models were *geometrically orthgonal* in the sense of Tjur (1984). One consequence is that certain edges in the 'anova' version of the Hasse diagram have the same length: for example, in Figure 7, the edge between Richness and Richness + Type has the same length as the edge between Constant and Type. Under non-orthogonality, the diagram would still give useful information: the length of the former shows the mean square for fitting Type *in addition* to Richness, while that of the latter shows the mean square for fitting Type on its own (assuming that Constant is always fitted).

One further model was fitted to all six responses. The biomass of each individual organism was measured at the start of each experiment. The response variable was regressed on the sum of the biomasses in each jar.

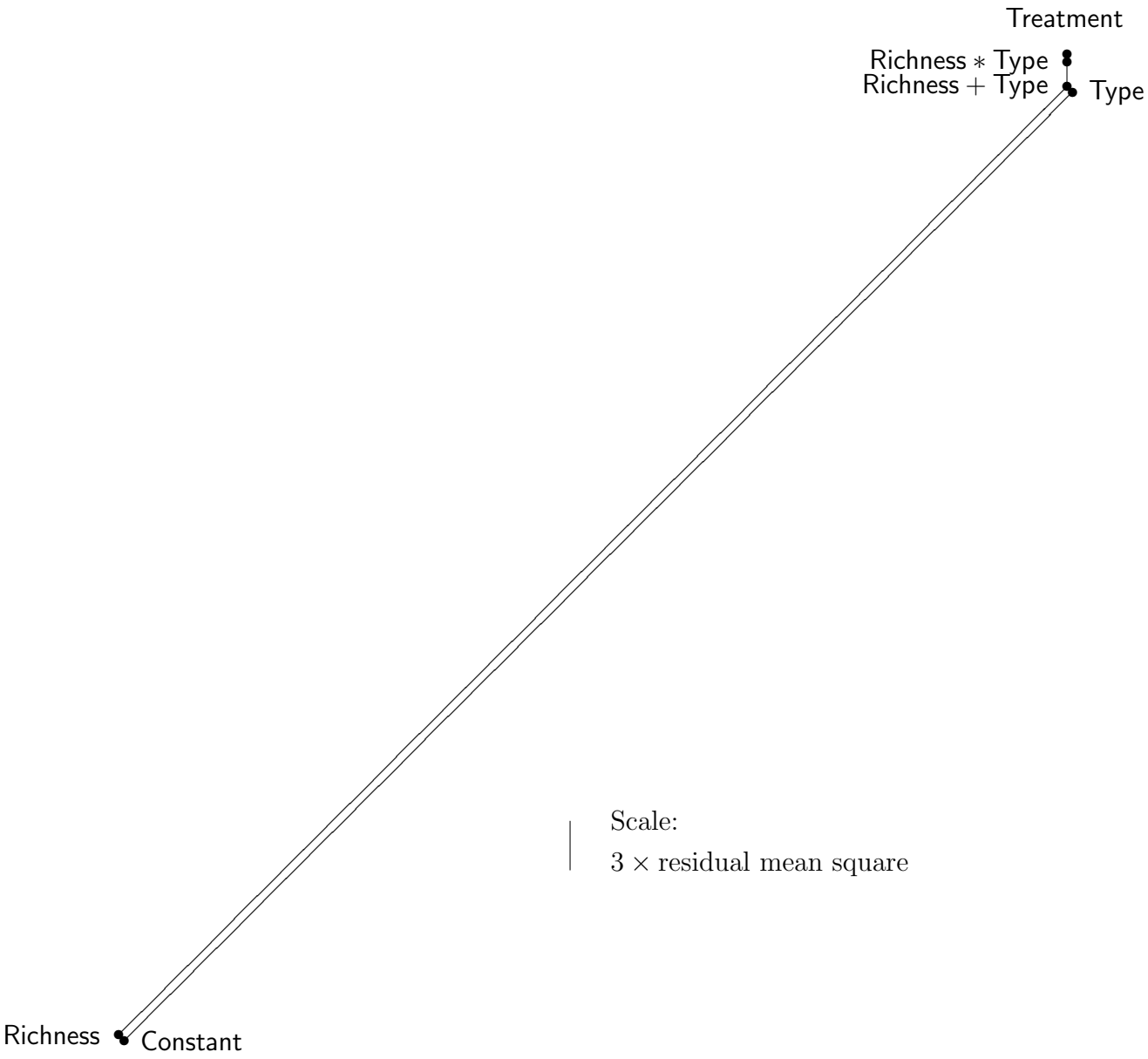

Figure 7: Anova version of the Hasse diagram in Figure 2 for the leaf decomposition response

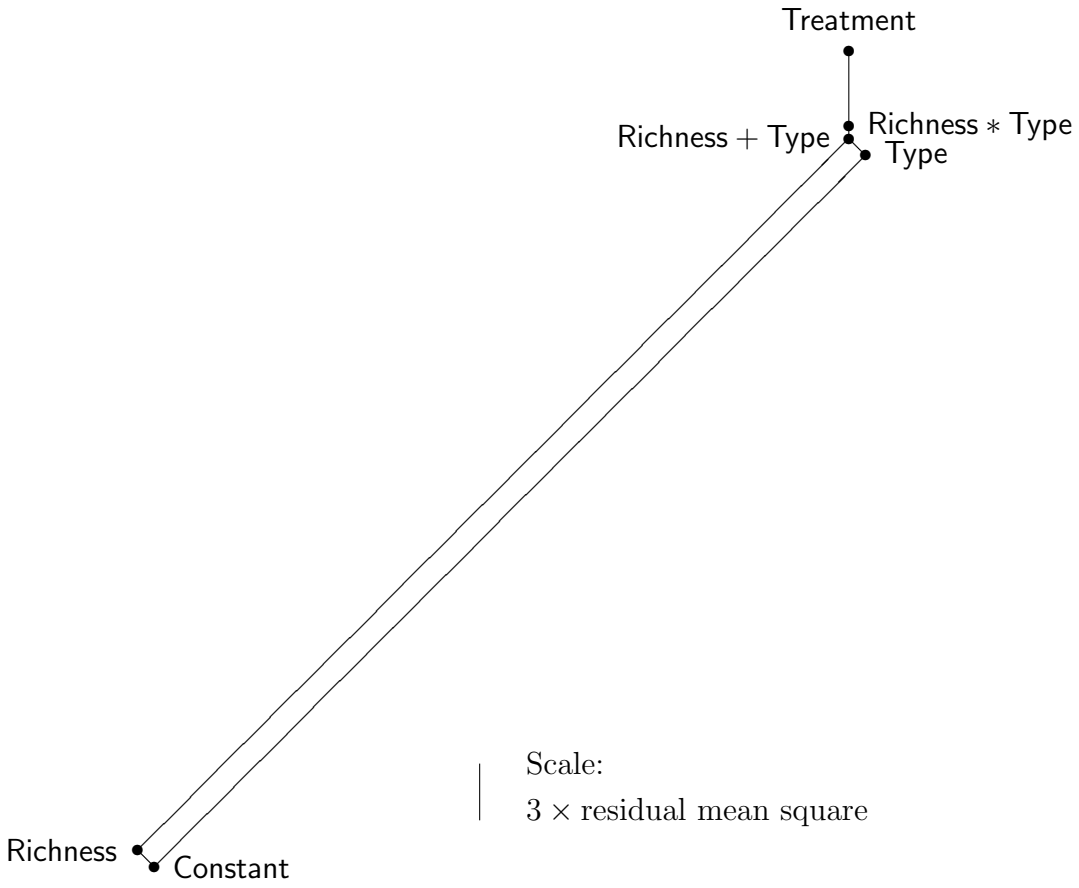

Figure 8: Anova version of the Hasse diagram for the leaf decomposition response in the third experiment

Because jars with the same treatment did not necessarily have the same total biomass, this two-dimensional model was not contained in, nor geometrically orthogonal to, any of the other models we have discussed. Nevertheless, it is not too far from a submodel of Type. It did not give a completely adequate explanation of the data, but did very well for such a small model.

5. How should such experiments be designed?

These experiments were all designed with some idea of fairness and balance, but no statistical theory was involved. So, how should such experiments

be designed?

To answer this question, we need to know what the purpose the experiment is. Here are some possible reasons for conducting an experiment like the first one.

(i) To find out whether the response changes with different levels of Richness.

(ii) To estimate the differences between the different levels of Richness.

(iii) To discriminate between the models Richness and Type.

(iv) To discriminate between the model Type and the more general model Richness $*$ Type.

(v) To estimate the parameters (response per individual for each type) for the model Type.

Aims (i) and (ii) are different. Even for aim (i), there is a difference between merely checking that the model is adequate (Atkinson et al., 2007, Sections 20.1–20.5) and giving ourselves a chance to find out if something else causes the response to change. The first experiment was explicitly designed to find out whether the presence of any one type affected the performance of any other types. For aim (ii), we want an experiment which can provide unbiased estimators of these differences, with as small variance as possible. Such designs are called *A-optimal* (Atkinson et al., 2007, Section 10.1). The sum of the variances of the estimators of differences is minimized when the replications of the levels are as equal as possible, so we should replicate them more equally than $6 : 15 : 20$ if this is our aim.

Aims (iii) and (iv) are quite similar: we have two explicit possible models and want to discriminate between them. The best designs for this situation are called *T-optimal*, and are discussed by Atkinson et al., (2007, Sections 20.6–20.11). However, there is a subtle difference between case (iv), when one design is a special case of the other, and the more general situation in case (iii). In fact, it is rare to be considering only two explanatory models in an experiment: none of the Hasse diagrams in this paper has so few.

Aim (v) assumes that we know that the model Type is all that is needed to explain the data and that we simply need an A-optimal design for the parameters in expression (2). An A-optimal design would include all monocultures equally often and exclude all polycultures. Such a design would give

no possibility of discovering whether the factor Richness has any effect or whether types interact with each other.

What seems to be needed for real experimentation is a theory about what is best for the process of first choosing an adequate model from a family of several models and secondly estimating the parameters of the chosen model. So far, there does not seem to be a general theory for this process.

Even without such a general theory, there is the pertinent question about which subsets of types we should choose as treatments. To pose this generally, suppose that there are t types in all. For any given level k of Richness, each treatment consists of equal numbers of each type in some subset Γ of k types. Suppose that we can use n such treatments. How should we choose the subsets to include?

It seems that ecologists traditionally choose subsets at random. In our third experiment, we chose the Fano plane, which would have been the A-optimal design for an experiment with seven treatments in seven blocks of size three. Is the design which is best for an incomplete-block experiment necessarily best as a design for polycultures?

Unfortunately, the answer is 'No'. Consider incomplete-block designs for t treatments in n blocks of size k. In the usual model for an incomplete-block design, the expected response on any experimental unit with treatment i in block Γ is

$$\alpha_i + \beta_\Gamma. \tag{4}$$

For polycultures, our model Type states that the expected response on any experimental unit whose treatment is an equal mixture of the types in subset Γ is

$$\frac{1}{k}\sum_{i\in\Gamma}\alpha_i. \tag{5}$$

It can be shown that if there is a balanced incomplete-block design for t treatments in n blocks of size k then it is A-optimal for both situations. Otherwise, a design which is best for one situation may be worst for the other.

For example, let $t = n = 6$ and $k = 2$. Figure 9 shows two incomplete-block designs for six treatments in six blocks of size two: each edge denotes a block consisting of the letters at its two ends. If either of these designs is used as a block design in the usual way then there are twelve experimental units and their expected responses are given by (4). The design in Figure 9(b) is A-optimal for the 15 differences $\alpha_A-\alpha_B, \ldots, \alpha_E-\alpha_F$, while that in Figure 9(a) is *disconnected* because the difference $\alpha_A - \alpha_D$ cannot be estimated. On the

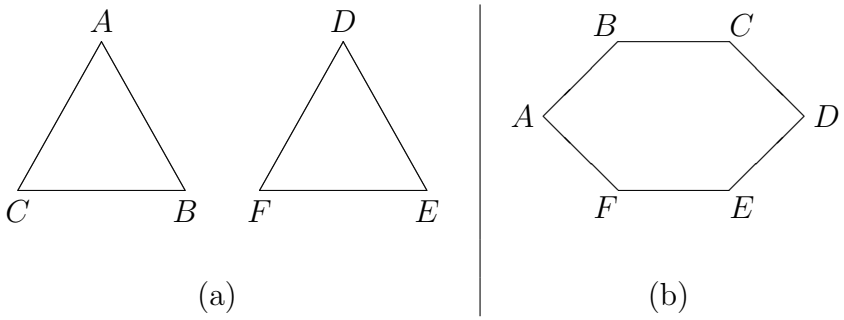

Figure 9: Two collections of six subsets of size two from $A, \ldots, F$: each edge denotes the pair consisting of the letters at its ends

other hand, suppose that one of these designs is used for an experiment with six dicultures, one per experimental unit, with no monocultures and no other polycultures. Then the expected response is given by (5), and the design in Figure 9(a) is A-optimal for the parameters $\alpha_A, \ldots, \alpha_F$. However, the design in Figure 9(b) is disconnected, because the difference $\alpha_A - \alpha_B$ cannot be estimated.

Acknowledgements. The experiments reported here were funded by a grant (reference NE/D013305/1) from the Natural Environment Research Council, UK. This paper was written while RAB was at the Isaac Newton Institute, Cambridge, UK during the 2011 programme Design and Analysis of Experiments.

References

Atkinson, A. C., Donev, A. N., Tobias, R. D., 2007. Optimum Experimental Designs, with SAS, Oxford University Press, Oxford.

Bailey, R. A., 2008. Design of Comparative Experiments, Cambridge University Press, Cambridge.

Bell, T., Lilley, A. K., Hector, A., Schmid, B., King, L., Newman, J. A., 2009. A linear model method for biodiversity-ecosystem functioning experiments. The American Naturalist 174, 826–849.

Bell, Thomas, Newman, Jonathan A., Silverman, Bernard W., Turner, Sarah L., Lilley, Andrew K., 2005. The contribution of species richness and composition to bacterial services. Nature 436, 1157–1160.

Cardinale, B. J., 2011. Biodiversity improves water quality through niche partitioning. Nature 472, 86–89.

Cornell, John A., 2002. Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data (3rd edition), John Wiley and Sons, New York.

Grafen, A., Hall, R., 2002. Modern Statistics for the Life Sciences, Oxford University Press, Oxford.

Jonsson, M., Malmqvist, B., 2005. Species richness and composition effects in a detrital processing chain. Journal of the North American Benthological Society 24, 798–806.

Kirwan, L., Connolly J., Finn, J. A., Brophy, C., Luscher, A., Nyfeler, D., Sebastia, M.-T., 2009. Diversity-interaction modeling: estimating contributions of species identities and interactions to ecosystem function. Ecology 90, 2032–2038.

van Lint, J. H., Wilson, R. M., 2001. A Course in Combinatorics (2nd edition), Cambridge University Press, Cambridge.

Loreau, M., Naeem, S., Inchausti, P., 2002. Biodiversity and Ecosystem Functioning: Synthesis and Perspectives, Oxford University Press, Oxford.

McKie, B. G., Woodward, G., Hladyz, S., Nistorescu, M., Preda, E., Popescu, C., Giller, P. S., Malmqvist, B., 2008. Ecosystem functioning in stream assemblages from different regions: contrasting responses to variation in detritivore richness, evenness and density. Journal of Animal Ecology 77, 495–504.

Nelder, J. A., 1977. A reformulation of linear models. Journal of the Royal Statistical Society, Series A 140, 48–77.

Nelder, J. A., 1994. The statistics of linear models: back to basics. Statistics and Computing 4, 221–234.

Perkins, D. M., McKie, B. G., Malmqvist, B., Gilmour, S. G., Reiss, J., Woodward G., 2010. Environmental warming and biodiversity-ecosystem functioning in freshwater microcosms: partitioning the effects of species identity, richness and metabolism. Advances in Ecological Research 43, 177–208.

Reiss, Julia, Bailey, R. A., Cássio, Fernanda, Woodward, Guy, Pascoal, Cláudia, 2010. Assessing the contribution of micro-organisms and macrofauna to biodiversity-ecosystem functioning relationships in freshwater microcosms. Advances in Ecological Research 43, 151–176.

Reiss, Julia, Bailey, R. A., Perkins, Daniel M., Pluchinotta, Angela, Woodward, Guy, 2011. Testing effects of consumer richness, evenness and body size on ecosystem functioning. Journal of Animal Ecology 80, 1145–1154.

Schmid, B., Hector, A., Huston, M. A., Inchausti, P., Nijs, I., Leadley, P. W., Tilman, D., 2002. The design and analysis of biodiversity experiments. In Loreau, M., Naeem, S., Inchausti, P. (editors), Biodiversity and Ecosystem Functioning: Synthesis and Perspectives, Oxford University Press, Oxford.

Tjur, T., 1984. Analysis of variance models in orthogonal designs. International Statistical Review 52, 33–81.