Editor's Note: Toolboxes are intended to describe and evaluate methods that are becoming widely relevant to the neuroscience community or to provide a critical analysis of established techniques. For more information, see http://www.jneurosci.org/misc/ifa_minireviews.dtl.

Toward an Efficient and Integrative Analysis of Limited-Choice Behavioral Experiments

Karl Schilling,¹ John Oberdick,² and René L. Schilling³

¹Anatomisches Institut, Anatomie und Zellbiologie, Rheinische Friedrich-Wilhelms-Universität, D-53115 Bonn, Germany, ²Department of Neuroscience, The Ohio State University Wexner Medical Center, Columbus, Ohio 43210, and ³Department of Mathematics, Institut für Mathematische Stochastik, Technische Universität Dresden, D-01062 Dresden, Germany

Introduction

Quantitative approaches are increasingly used to compare behavioral traits and relate them to genetic and neural mechanisms. Key to this are simple, yet powerful and reproducible behavioral tests applicable to small model organisms. In such tests, experimental subjects often have to choose between a limited number of alternatives. Examples include local preference tests such as the light-dark preference test; the two-temperature choice test; the bedding preference test; the elevated p-maze; some versions of the open field test; implementation of the free choice exploration by Crestani et al. (1999); mate choice copying tests (Galef et al., 2008; Mery et al., 2009); and tests of social novelty preferences and sociability (for an authoritative description of such tests and further reference, see Crawley 2007; Nadler et al., 2004). Statistical analysis is critical in the interpretation of such experiments. Traditionally, the outcome of such limited-choice tests is presented as the arithmetic mean and its standard error (SE) and probed using classical parametric tests, such as Student's *t* test or ANOVA for means (Nadler et al., 2004; Moy et al.,

2007; Saverino and Gerlai, 2008; Hines et al., 2008; Mery et al., 2009; Blundell et al., 2010; Peça et al., 2011).

Yet, this analytical approach severely limits the efficient and integrative interpretation of the data. This is a consequence of the specific structure of data resulting from the type of behavioral tests described above, i.e., tests built around a limited-choice procedure. Such data are typically relative in nature, bounded, and highly correlated. Technically, data of this structure are referred to as compositional (Aitchison and Egozcue, 2005 and references therein). That is, the primary interest of the experimenter is not on absolute measures of how often one choice was selected, or how much time was spent on it, but on how these values compare to those measured for the alternative choices available. In fact, total observational time is typically limited, say to 5 or 10 min. Then, the data are not only compositional, but also referred to as closed (Pawlowsky-Glahn and Egozcue, 2006): times measured in such an experiment sum up to a constant value. Karl Pearson noted as early as 1897 that this data structure—the fact that we do not deal with absolute, independent measures—precludes a direct analysis with standard parametric methods and makes the latter at best inefficient, if not plainly misleading (Pearson, 1897; Kronmal, 1993; Pawlowsky-Glahn and Egozcue, 2006). This caveat seems well known in fields that typically deal with data of obviously compositional structure, such as geology (where the interest

may be in the how percentages of pure substances that make up a rock differ between samples) (Montero-Serrano et al., 2010) or social sciences (where the interest may be in the relative parts of a total budget spent on various consumer choices) (Aitchison, 1982), but less so among behavioral biologists (but for examples to the contrary, see Solberg et al., 2004; Pierotti et al., 2009).

Here, we draw attention to how recent developments in statistical techniques allow experimenters to overcome this impasse. Proper consideration of data structure not only results in a formally correct, hence more reliable, analysis, but also allows a much more efficient use of behavioral experimental data, as it allows experimenters to handle, and probe, their intrinsically multivariate structure, which is inaccessible with the currently preferred styles of analysis. This is particularly useful when it comes to comparing behavioral preferences of animals that differ, say due to defined genetic or pharmacological interventions (Saverino and Gerlai, 2008; Blundell et al., 2010; Satoh et al., 2011; Walton et al., 2012), and substantially adds to the sensitivity of such tests. In the following, (1) we briefly exemplify the specific structure of compositional data as may arise from behavioral experiments; (2) we discuss how such data may be transformed so that they can be effectively interpreted in biological terms and, at the same time, may be handled statistically, including a short discussion of potential limitations of the approach proposed; (3)

Received March 25, 2012; revised June 17, 2012; accepted July 16, 2012.

We thank S. S. Moy and her coauthors for making their data publicly available through the Jackson Laboratory Mouse Phenome Database, and those responsible for this repository for this unique resource.

Correspondence should be addressed to Dr. Karl Schilling, Anatomisches Institut, Anatomie und Zellbiologie, Rheinische Friedrich-Wilhelms-Universität, Nussallee 10, D-53115 Bonn, Germany. E-mail: karl.schilling@uni-bonn.de.

DOI:10.1523/JNEUROSCI.1452-12.2012

Copyright © 2012 the authors 0270-6474/12/3212651-06\$15.00/0

we sketch efficient ways to visualize such data and propose a powerful procedure for testing hypotheses based on them; and (4) we conclude by pointing out how our approach allows experimenters to put the type of behavioral tests discussed in a broader framework of behavioral analysis.

The structure of experimental behavioral data—some background

The relative character and bounded structure of data resulting from limited-choice tests become apparent when preferences between choices are presented as percentages or ratios, which add up to 100, or unity. They may be harder to spot when preferences are expressed, e.g., as absolute times spent with any one choice, or as counts. Yet as the total time of observation in such tests, or total counts, are limited, times spent with any one choice or counts representing them are also bounded, and hence are indicators of relative preferences rather than free variables (Pennington et al., 2009). A hypothetical numerical example construed to model the standard test for social novelty preference for mice (Nadler et al., 2004) should clarify the consequences of this situation for statistical analyses, and why ANOVA is not appropriate for data of this structure. For this test, animals are placed in the neutral zone of a tripartite cage, from where they can choose to enter two additional partitions (zones), each containing a stimulus mouse: one of these is familiar and the other novel (a stranger) to the test animal. Suppose we want to compare the behavior of two groups of animals, A and B, which, over the observational period of 600 s, partition their times as follows: animals of group A spent, on average 360 s with the stranger and 180 s with the familiar mouse; the remaining time (60 s) they passed in the neutral zone. Animals of group B spent, on average, 120 s with the stranger and 20 s with the familiar mouse, and 460 s in the neutral zone. We do not have to bother about variances in this simplifying example.

To assess social novelty preference, time spent in the neutral zone is not really relevant, and so novelty preference is typically quantified by comparing the times spent in each of the two compartments holding the stimulus mice. Standard parametric testing (including ANOVA) does so by assessing the differences of the times. So we are led to conclude that group A animals have a greater preference for the novel (stranger) mouse ($360 - 180 \text{ s} = 180 \text{ s}$) than group B animals, for which this difference is only 100 s.

While this approach does not explicitly mention times spent in the neutral zone, these directly, if somewhat furtively, enter the analysis: given the overall limited observational period, the time an animal spends in the neutral compartment automatically limits the times it may spend with either the stranger or familiar mouse, and hence the difference of these values.

To avoid this hideous interference, we may recast our question and ask: how does a test animal partition the total time it spends on social contacts—i.e., the time it spends outside the neutral zone—between the stranger and familiar mouse? In our example, group A animals spent a total of 540 s on social contacts, of which they allocated 66.7% (360 s) to the stranger, and 33.3% (180 s) to the familiar mouse. Group B animals spent a total of 140 s on social contacts, 85.7% (120 s) with the stranger, and only 14.3% (20 s) with the familiar mouse. So this ratio-based perspective suggests that group B animals show a greater preference for the stranger than those of group A. Before considering how this may be tested formally, we would like to point out that times spent in the neutral zone again entered the analysis, but this time not as a nuisance parameter, as in the ANOVA-based approach, but rather as a sensible parameter to “normalize” the social contact times of various test animals.

We may take this a step further and ask whether time spent in the neutral zone is not an interesting variable in its own right. In our example, times spent in the neutral zone may just as well be used as differential preferences for the stranger mouse to distinguish group A animals (60 s) from those of group B (460 s), and indeed suggest that group B animals might be loners. Note that this perspective allows the experimenter to extract, from a test set up to assess social novelty preference, a measure of sociability, which is typically assessed in a separate test in which the test animal is presented with a choice between a stimulus mouse and an inanimate object.

This exemplifies the point that data resulting from limited-choice tests are typically multivariate—i.e., constituted by multiple outcome variables, as, in our example, descriptors of social novelty preference and sociability. Consequently, such data call for appropriate multivariate methods. ANOVA (and *t* tests), though, are strictly univariate techniques (Tabachnick and Fidell, 2006, chapter 2) and cannot account for the data in full (Weltje, 2002; see also Filzmoser et al., 2009).

Making behavioral data from limited-choice tests accessible to analysis

Fortunately, there do exist methods appropriate for the fundamentally compositional structure of data resulting from limited-choice tests that also allow the experimenter to take full advantage of their genuinely multivariate nature. These have been developed largely in fields like geology and chemistry, but also in the social sciences, where compositional data are common (Aitchison and Egozcue, 2005; Pawlowsky-Glahn and Egozcue, 2006; Pawlowsky-Glahn and Buccianti, 2011 and further references contained therein). We may readily tap and adapt these methods for behavioral analysis, since they are conveniently implemented in the freely available R computational environment (van den Boogaart and Tolosana-Delgado, 2008; Hron et al., 2010; Rizzo and Székely, 2010; R Development Core Team, 2011). Indeed, there is precedence for the application of compositional statistics to biologically motivated questions including animal and human behavioral biology [Solberg et al., 2004; Pierotti et al., 2009; Pennington et al., 2009 (note that the latter two studies use implementations of the method in STATA and Matlab, respectively)].

At the heart of this methodology is a familiar and time-honored approach. That is, the data are transformed to conform to the (distributional) requirements of the statistical procedures one wants to use, such that the restrictions inherent in the original data are removed. The transformation that simultaneously removes the bounds of compositional data and allows the experimenter to tackle their correlation structure, thus making them amenable to standard parametric statistics, is referred to as the log-ratio transform. This has been pioneered and popularized by Aitchison and Egozcue (2005; and the extensive bibliography therein). The rationale and the formalism of this approach, its technical details, and its benefits have been extensively described (for review, see Billheimer et al., 2001; Weltje, 2002; Egozcue et al., 2003; Aitchison and Egozcue, 2005; Pawlowsky-Glahn and Buccianti, 2011).

Of the three variants of the log-ratio transformation commonly used, the isometric log-ratio (ILR) transform yields the most suitably transformed data for standard multivariate techniques (Egozcue et al., 2003; Pawlowsky-Glahn and Egozcue, 2006). The key step of this transform is that original data are systematically related to each other. Several technical descriptions of how to calculate the IRL are available (for

review, see Egozcue et al., 2003; Egozcue and Pawłowsky-Glahn, 2005), and it may be readily implemented in standard spreadsheet software or any statistical program that allows basic mathematical operations on the data. Convenient implementations are available, for example, in the R packages “compositions” or “robCompositions” (van den Boogaart and Tolosana-Delgado, 2008; Hron et al., 2010). So, we may focus here on the conceptual advantage that the log-ratio transform and the ILR, in particular, offer. If we subject the three times recorded in our above example for each experimental animal—i.e., those spent with a familiar mouse, a stranger mouse, and in the neutral zone—to ILR transformation, two derived variables are obtained, which, for obvious reasons, are also referred to as contrasts or coordinates. The first relates the times spent with the familiar and the stranger mouse to each other, and this serves as a measure of social novelty preference. The second relates the combined times spent with the stranger and familiar mouse to the time spent in the neutral compartment and may be sensibly interpreted as a measure of the overall tendency to seek social contacts (i.e., sociability). Note that derived variables are built on ratios rather than differences as in standard ANOVA.

In addition to transforming the original data so that they become suitable for standard multivariate analysis, the ILR transform also allows the experimenter to contrast the original data such that it becomes easier to give a meaningful biological interpretation. Actually, how exactly the original variates are contrasted with each other depends on the order they are entered into the ILR transform. There exist several mathematically equivalent and consistent ILR transforms for a given dataset. These differ by the perspective they give us on the data. Significantly, the experimenter may choose which perspective to take and which contrasts to form simply by defining the order in which the original variates are entered into the ILR transform. This choice could be informed by prior knowledge and thus focus attention on those variates of the original data that have the greatest diagnostic relevance. Conversely, we may calculate several ILR transforms for a given dataset and then ask which of these result in the contrasts with the best diagnostic characteristics (however we may define them) and which original variates enter those contrasts. For a formalized procedure to determine how all possible contrasts may be systematically formed, see Egozcue and Pawłowsky-Glahn (2005).

While log-ratio transformation provides a unique tool to make compositional data accessible to standard multivariate analysis and at the same time provides an approach for their sensible biological interpretation, there is one shortfall, of direct practical relevance, that needs to be briefly discussed. That is, log-ratio transformations, including ILR, cannot handle data that contain zero values, since neither division by zero nor logarithms of zero are mathematically defined. At first sight, this looks like a mathematical nuisance, but it actually takes us to a central methodological and biological issue. Simply to exclude samples that contain zero measurements (in our example, these could arise if a test animal would not visit one cage compartment at all) would not only be a waste of data, but might even substantially distort subsequent analyses. So, it seems well advised to ask how zero values originated in the first place. If test subjects do not select one of the choices available during the observational period, is this a defining behavior, or does it reflect some shortcoming of our experimental setup? Could we have observed the subject take this choice if we had extended our observational period appropriately? Both situations result in zero values, and in technical parlance the first is often called a structural zero, whereas the second is a case of censored data, sometimes also referred to as a rounded, or trace, zero.

In a realistic experimental setting, structural zeros and those due to measurement sensitivity may be hard to discern. Although definitive advice cannot be given here, a workable approach may be to handle, in a first step, a limited number of randomly interspersed zero values as due to experimental sensitivity and thus replace them by carefully imputed small values. However, any clustering of zeros should alert the experimenter to the possible existence of structural zeros, defining a behavioral pattern. These should not be replaced, but suggest the use of an alternative model (see third paragraph in the next section). A more extensive discussion of this issue, along with advice and sensible methods for replacing zeros, may be found, for example, in the studies by Fry et al. (2000), Palarea-Albaladejo and Martín-Fernández (2008), and Hron et al. (2010).

Efficient display and analysis of data from limited-choice tests

To briefly document the conceptual advantages granted by consideration of the specific structure of data resulting from limited-choice tests, and to introduce some concepts about how to proceed after data transformation, we now shall discuss,

and reanalyze, some selected data on sociability and social novelty preference originally published by Moy et al. (2007, 2008). These data are publicly available from the Jackson Laboratory Mouse Phenome Database (<http://phenome.jax.org/>; Moy1 dataset).

As so often, a sensible first step in the analysis is to plot the data. In Figure 1, bar diagrams traditionally used to represent the outcomes of tests for sociability and social novelty preference are shown alongside two diagrams particularly suitable for compositional data formed by three variables. Ternary diagrams (Billheimer et al., 2001) allow visualization of the original data. They are based on equilateral triangles, where each corner (vertex) represents one variable of the composition. That is, in our example, one of the behavioral choices offered. Each experimental subject is represented by a single point within this triangle, whose distance from any one corner reflects the relative preference for the behavior represented by this very corner. A subject avoiding one choice altogether is represented by a point on the line opposite to the vertex representing this choice. ILR-transformed data are efficiently displayed in biplots (Aitchison and Greenacre, 2002), which, in the case of a ternary composition, are simple x - y scatterplots in which the two contrasts [say, “familiar/stranger” and “(familiar, stranger)/neutral” in our example above] are plotted against each other. Both ternary diagrams and biplots give an ad hoc impression of data variability (or spread), and also of how individual variates and contrasts formed by them may relate to each other (Fig. 1). Of course, summary statistics (as the mean or median, and measures of data variability) may also be included in such diagrams.

Figure 1A–C documents the sociability of A/J mice reported by Moy et al. (2007). When displayed as means $\pm$ SEM (Fig. 1A, which reproduces Fig. 4A in Moy et al., 2007), the data suggest that these animals spent about equal times with the stranger mouse, in the object cage, and in the neutral cage partition. Consistently, testing by ANOVA suggests that A/J mice, on average, do not show a preference for social contacts (i.e., show no sociability). Inspection of the data displayed in a ternary plot (Fig. 1B) reveals how misleading this conclusion can be. Indeed, of 19 A/J mice tested for sociability, 7 did not spend any time in the cage partition holding the stranger mouse, and another 7 did not enter the partition holding the object. One mouse did not enter either. Only four animals actually visited

Figure 1. *A–C*, Data display to detect potentially extreme behaviors. Traditional summary statistics (*A*, mean $\pm$ SEM) suggest that, in the sociability test, A/J mice partition their time about equally between the neutral compartment and the cage compartments housing the object and the stranger mouse. Plotting the data in a ternary diagram (*B*) reveals that most animals either avoid the object, or the stranger-side altogether, suggesting that a dual-choice model (neutral vs non-neutral zones) may be more appropriate than the classical representation and the ternary plot. This is also suggested by the observation that in the biplot (*C*), most data fall on an essentially straight line. Note that for numerical reasons, zero values had to be replaced by imputation (see main text) to allow generation of the biplot. Data are from Moy et al. (2007). *D–F*, Multivariate display and analysis allow the experimenter to use intragroup variability to discern DBA/2J and NOD/LtJ mice. The traditional univariate representation (*D*; mean $\pm$ SEM) does not give any clues that DBA/2J and NOD/LtJ mice might differ in their social novelty-seeking behavior. Although variability appears somewhat more pronounced for DBA/2J mice, this eludes formal statistical detection for univariate statistics calculated for untransformed data. In contrast, plotting the data in a ternary diagram (*E*) or, following ILR transformation, in a biplot (*F*), readily draws attention to the much more pronounced behavioral variability of DBA/2J mice (gray triangles) compared with NOD/LtJ mice (black circles), though both have identical mean performances (gray “x” and black cross, respectively). This difference in variability is also detectable using appropriate formal testing (see main text). Data are from Moy et al. (2008). *G–I*, Multivariate compositional analysis allows the experimenter to use an internal correlation structure of behavioral data to probe for intergroup differences. Synthetic data describing two groups of mice for which times spent in the individual compartments of a three-compartment cage and the extent of data variability are virtually identical, as seen either in the traditional (*G*, mean $\pm$ SEM), the ternary (*H*), or the biplot display (*I*). However, the biplot (*I*) clearly documents the essentially inverse relationship between the familiar/stranger variate and the variate that contrasts times spent with either the familiar or a stranger mouse with that in the neutral (center) zone. This is also clearly visible in the ternary plot (*H*), which documents that animals of group 2, represented by gray triangles, show a positive correlation between social novelty seeking and leaving the center (which may be interpreted as a measure of overall sociability). For group 1, represented by black circles, increased sociability goes along with an increased tendency to contact the familiar mouse; i.e., there is a negative correlation between sociability and social novelty seeking.

all three cage partitions, and these rare animals also strongly favored one partition. This perspective on the data may suggest that A/J mice comprise two subgroups with distinct sociability. A more parsimo-

nous explanation is that they simply stick to their first random choice and that this is a characteristic for this strain. The somewhat larger variability of the sociability data of A/J mice seen when data are fitted

to a three-choice model (be that by conventional ANOVA or following ILR transformation; see Fig 1*A–C*; and compare variability with that of the other examples shown in Fig 1*D–I*) would thus hint that

this model be inappropriate. Indeed, the occurrence of multiple zero values in the data describing A/J sociability behavior should also prompt the experimenter to consider a binary-choice model, rather than a ternary-choice model (as done here). In fact, this dataset seems to be a clear-cut example of a behavior characterized by structural zeros (see the last two paragraphs of the preceding section). The ternary display clearly draws attention to this situation.

How to proceed after data transformation and inspection is, of course, primarily dictated by the experimental questions one wants to address. Here, we focus on how to take advantage of the multivariate data structure for the efficient and sensible comparison of two or more sets of data, as is often the focus of interest when assessing, say, behavioral effects of genetic or pharmaceutical manipulations (Blundell et al., 2010; Peça et al., 2011; Walton et al., 2012). A highly efficient test to do so was recently presented by Székely and Rizzo (2009) and Rizzo and Székely (2010) (implemented in the R-package “energy”), and we use social novelty preference data recorded with DBA/2J and NOD/LtJ mice (Moy et al., 2008) to exemplify its usefulness. Group averages alone, whether calculated from non-transformed or ILR-transformed data, do not suggest that these two strains might differ in their choices of the three cage partitions (Fig. 1*D,E*). Although the error bars in the classical representation of nontransformed data (Fig. 1*D*) suggest that intra-group variability might vary, this is much better documented in the ternary diagram and also in the biplot of ILR-transformed data (Fig. 1*E,F*). Yet testing by ANOVA (Fig. 1*D*, data displayed) or MANOVA (and also Anderson’s robust variant; Anderson, 2001) would not even hint at this obvious difference, simply because these tests cannot discern data based on differences in variability. MANOVA actually presupposes that variability be similar among the groups compared. Conversely, the Rizzo–Székely test clearly allows to distinguish NOD/LtJ and DBA/2J mice based on the variability of their social novelty preference ($p = 0.002$). Clearly, knowledge about the behavioral diversity within a group may be as informative as any measure of group average to guide the search for potential behavioral determinants.

The Rizzo–Székely test not only allows for the simultaneous testing of multivariate datasets based on their means and variability, it also allows the experimenter to distinguish sets that differ by how individual variates relate (or correlate) with each other within the groups compared. A

hypothetical example devised to convey this concept within a biological context is shown in Figure 1*G–I*. In contrast to the DBA/2J and NOD/LtJ data shown in Figure 1*D–F*, those of the two groups of animals shown in Figure 1*G–I* not only have comparable group means, but also have the same overall variability, as may be directly guessed from the error bars in Figure 1*G*, and is clearly visible in Figure 1, *H* and *I*. Actually, the data were constructed so as to have identical variability. However, as is clearly visible in Figure 1*I*—but also Figure 1*H*—the two groups differ in how their preference for either the familiar or stranger mouse (Fig. 1*I*, *y*-axis) and the overall tendency to leave the neutral zone (i.e., sociability; Fig. 1*I*, *x*-axis) relate to each other. This striking behavioral difference goes completely un-noticed in the classical display (Fig. 1*G*), and it cannot be detected using ANOVA or MANOVA. However, it is readily revealed by the Rizzo–Székely test ($p = 0.036$).

This example underscores the importance, and substantial profit, of selecting a test appropriate to the data structure, and also the critical value of visual inspection of original data in this process. While the Rizzo–Székely test does not tell us whether data differ due to differences in means, their variability (spread), or the internal correlation between individual variates, it allows the diagnostic use of the latter two parameters. Clearly, the potential to probe group differences beyond mean performance, as provided by the Rizzo–Székely test (Székely and Rizzo, 2009; Rizzo and Székely, 2010), enhances overall test sensitivity and thus contributes to an ethical, and more economic, use of experimental subjects. The graphical display afforded by the ternary diagram and biplot should help to decide how to proceed once a significant result is obtained with the Rizzo–Székely procedure. This may include applying a formal test for differences in scatter, or fitting (linear) models to describe the correlations of individual variates. Here, sensitive model choice may provide substantial insight into the behavioral difference diagnosed.

Conclusions and perspectives

Recognition and proper handling of the particular, compositional structure of data resulting from tests built around a limited-choice and/or forced-choice procedure not only allows a formally correct analysis, but also a more efficient and coherent use of data than has hitherto been appreciated. When combined with novel and efficient testing procedures that can account for the intrinsically multivariate

structure of the data (Székely and Rizzo, 2009; Rizzo and Székely, 2010), this adds substantially to overall sensitivity. Significantly, this gain in sensitivity results from a deep probing of information encoded in the extent and form of variance (or covariance) of choice behaviors, thus opening up an avenue for better description, and understanding, of behavioral structure.

While the examples chosen here to introduce the methodology focus on a three-choice procedure, the techniques are readily extended to situations with more than three choices (and also simplified for analysis of a dual-choice situation; see Heslop, 2009), although efficient visualization seems to be limited to a maximum of four choices (that can be displayed in a regular tetrahedron, i.e., the three-dimensional equivalent of the ternary diagram).

The realization that many behavioral tests for small animal models, including the tests for sociability and social novelty preference used as an example in the present communication, are incarnations of a much more general class of limited-choice procedures has consequences beyond statistical analysis. In fact, it allows us to view these tests in a much broader framework, and to see their direct conceptual and structural relationships with testing procedures applied in quite diverse settings of behavioral analysis. These include, for example, tests aimed at understanding consumer preferences, say the partitioning of disposable income among an exhaustive list of mutually exclusive consumer choices in commodities (say: “food,” “travel,” “electronics” and “other”) (Aitchison, 1982; Fry et al., 2000). We would also like to point out the conceptual proximity of experimental behavioral testing to document classification using dimension-reducing conditions (e.g., relying on the frequency of a few selected “key” words) to probe author characteristics (Lebanon, 2005).

We close with a cautionary note. While at times one may be tempted to forego the advantages of compositional and multivariate analysis for the convenience of more familiar univariate methods (like ANOVA, *t* tests), this comes at a considerable price. In doing so, one disregards, and in fact discards, valuable information. Worse, this price does not buy what it may promise. By turning a blind eye to some parts of the data, those parts chosen for analysis do not become independent of the neglected ones: they are still bounded, and they show a restricted correlation structure (i.e., they are still compositional, or closed). As one author concluded (and nicely documented for geological data),

“much of the valuable information that could be gained from quantitative analysis [...] is lost forever if one attempts to ‘summarize’ the results of one’s painstaking labor in the form of univariate statistics” (Weltje, 2002; see also Filzmoser et al., 2009). We hope that this primer attracts the attention of behavioral scientists to the power and potential of compositional data analysis, and vice versa motivates its further development.

References

- Aitchison J (1982) The statistical analysis of compositional data. *J R Stat Soc Series B Stat Methodol* 44:139–177.
- Aitchison J, Egozcue JJ (2005) Compositional data analysis: where are we and where should we be heading? *Math Geol* 37:829–850.
- Aitchison J, Greenacre M (2002) Biplots of compositional data. *J R Stat Soc Ser C Appl Stat* 51:375–392.
- Anderson MJ (2001) A new method for non parametric multivariate analysis of variance. *Austral Ecol* 26:32–46.
- Billheimer D, Guttorp P, Fagan WF (2001) Statistical interpretation of species composition. *J Am Stat Assoc* 96:1205–1214.
- Blundell J, Blais CA, Etherton MR, Espinosa F, Tabuchi K, Walz C, Bolliger MF, Südhof TC, Powell CM (2010) Neuroligin-1 deletion results in impaired spatial memory and increased repetitive behavior. *J Neurosci* 30:2115–2129.
- Crawley JN (2007) What’s wrong with my mouse? Behavioral phenotyping of transgenic and knockout mice. Hoboken, NJ: Wiley.
- Crestani F, Lorez M, Baer K, Essrich C, Benke D, Laurent JP, Belzung C, Fritschy JM, Lüscher B, Mohler H (1999) Decreased GABA_A-receptor clustering results in enhanced anxiety and a bias for threat cues. *Nat Neurosci* 2:833–839.
- Egozcue JJ, Pawłowsky-Glahn V (2005) Groups of parts and their balances in compositional data analysis. *Math Geol* 37:795–828.
- Egozcue JJ, Pawłowsky-Glahn V, Mateu-Figueras G, Barcelo-Vidal C (2003) Isometric logratio transformations for compositional data analysis. *Math Geol* 35:279–300.
- Filzmoser P, Hron K, Reimann C (2009) Univariate statistical analysis of environmental (compositional) data: problems and possibilities. *Sci Total Environ* 407:6100–6108.
- Fry JM, Fry TRL, McLaren KR (2000) Compositional data analysis and zeros in micro data. *Appl Econ* 32:953–959.
- Galef BG, Lim TCW, Gilbert GS (2008) Evidence of mate choice copying in Norway rats, *Rattus norvegicus*. *Anim Behav* 75:1117–1123.
- Heslop D (2009) On the statistical analysis of the rock magnetic S-ratio. *Geophys J Int* 178:159–161.
- Hines RM, Wu L, Hines DJ, Steenland H, Mansour S, Dahlhaus R, Singaraja RR, Cao X, Sammler E, Hormuzdi SG, Zhuo M, El-Husseini A (2008) Synaptic imbalance, stereotypies, and impaired social interactions in mice with altered neuroligin 2 expression. *J Neurosci* 28:6055–6067.
- Hron K, Templ M, Filzmoser P (2010) Imputation of missing values for compositional data using classical and robust methods. *Comput Stat Data Anal* 54:3095–3107.
- Kronmal RA (1993) Spurious correlation and the fallacy of the ratio standard revisited. *J R Stat Soc Ser A Stat Soc* 156:379–392.
- Lebanon G (2005) Information geometry, the embedding principle, and document classification. Paper presented at 2nd International Symposium on Information Geometry and Its Applications, Tokyo, Japan, December 12–16.
- Mery F, Varela SA, Danchin E, Blanchet S, Parejo D, Coolen I, Wagner RH (2009) Copying versus personal information for mate copying in an invertebrate. *Curr Biol* 19:730–734.
- Montero-Serrano JC, Palarea-Albaladejo J, Martín-Fernández JA, Martínez-Santana M, Gutiérrez-Martín JV (2010) Sedimentary chemofacies characterization by means of multivariate analysis. *Sediment Geol* 228:218–228.
- Moy SS, Nadler JJ, Young NB, Perez A, Holloway LP, Barbaro RP, Barbaro JR, Wilson LM, Threadgill DW, Lauder JM, Magnuson TR, Crawley JN (2007) Mouse behavioral tasks relevant to autism: phenotypes of 10 inbred strains. *Behav Brain Res* 176:4–20.
- Moy SS, Nadler JJ, Young NB, Nonneman RJ, Segall SK, Andrade GM, Crawley JN, Magnuson TR (2008) Social approach and repetitive behavior in eleven inbred mouse strains. *Behav Brain Res* 191:118–129.
- Nadler JJ, Moy SS, Dold G, Trang D, Simmons N, Perez A, Young NB, Barbaro RP, Piven J, Magnuson TR, Crawley JN (2004) Automated apparatus for quantitation of social approach behaviors in mice. *Genes Brain Behav* 3:303–314.
- Palarea-Albaladejo J, Martín-Fernández JA (2008) A modified EM algorithm for replacing rounded zeros in compositional data sets. *Comput Geosci* 34:902–917.
- Pawłowsky-Glahn V, Buccianti A (2011) Compositional data analysis: theory and applications. Chichester, UK: Wiley.
- Pawłowsky-Glahn V, Egozcue JJ (2006) Compositional data and their analysis: an introduction. Geological Society, London, Special Publications 264:1–10.
- Pearson K (1897) Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc R Soc Lond* 60:489–498.
- Peça J, Feliciano C, Ting JT, Wang W, Wells MF, Venkatraman TN, Lascola CD, Fu Z, Feng G (2011) Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* 472:437–442.
- Pennington L, James P, McNally R, Pay H, McCoshachie H (2009) Analysis of compositional data in communication disorders research. *J Commun Disord* 42:18–28.
- Pierotti ME, Martín-Fernández JA, Seehausen O (2009) Mapping individual variation in male mating preference space: multiple choice in a color polymorphic cichlid fish. *Evolution* 63:2372–2388.
- R Development Core Team (2011) R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; <http://www.R-project.org/>.
- Rizzo ML, Székely GJ (2010) DISCO analysis: a nonparametric extension of analysis of variance. *Ann Appl Stat* 4:1034–1055.
- Satoh Y, Endo S, Nakata T, Kobayashi Y, Yamada K, Ikeda T, Takeuchi A, Hiramoto T, Watanabe Y, Kazama T (2011) ERK2 contributes to the control of social behaviors in mice. *J Neurosci* 31:11953–11967.
- Saverino C, Gerlai R (2008) The social zebrafish: behavioral responses to conspecific, heterospecific, and computer animated fish. *Behav Brain Res* 191:77–87.
- Solberg LC, Baum AE, Ahmadiyeh N, Shimomura K, Li R, Turek FW, Churchill GA, Takahashi JS, Redei EE (2004) Sex- and lineage-specific inheritance of depression-like behavior in the rat. *Mamm Genome* 15:648–662.
- Szekely GJ, Rizzo ML (2009) Brownian distance covariance. *Ann Appl Stat* 3:1236–1265.
- Tabachnick BG, Fidell LS (2006) Using multivariate statistics. Upper Saddle River, NJ: Allyn and Bacon (Pearson).
- van den Boogaart KG, Tolosana-Delgado R (2008) “Compositions”: a unified R package to analyze compositional data. *Comput Geosci* 34:320–338.
- Walton JC, Schilling K, Nelson RJ, Oberdick J (2012) Sex-dependent behavioral functions of the Purkinje-cell specific Gai/o binding protein, Pcp2(L7). *Cerebellum*. Advance online publication. Retrieved August 7, 2012. doi: 10.1007/s12311-012-0368-4.
- Weltje GJ (2002) Quantitative analysis of detrital modes: statistically rigorous confidence regions in ternary diagrams and their use in sedimentary petrology. *Earth Sci Rev* 57:211–253.