These two types of models are difficult to distinguish, because they are mathematically very similar. In fact, models of type (1) can—under balancedness assumptions—be interpreted as models of type (2) with some inequality restrictions on the eigenvalues of the covariance matrix. Conversely, models of type (2) can be forced into the framework (1) if one allows for negative variance components, i.e., formally negative values of the variances on some of the error terms. Thus, (1) and (2) are just two different ways of stating almost the same model, and this has been the origin of much confusion. However, all difficulties seem to disappear if one thinks of the type (1) models as models for *finite subarrays* of infinite arrays, in the spirit of Section 5. The variances on random effect terms in a type (1) model then become interpretable as proper population variances, which certainly cannot be negative. For example, consider measurements y_{ij} of the same quantity with similar instruments $i = 1, \dots, m$, each repeated n times, $j = 1, \dots, n$. In this case, the variance on the error term ε_i has a straightforward interpretation as the variance on the baseline error of an instrument from the population of instruments of this kind, whereas the variance on ε_{ij} is the variance on measurements corrected for baseline error. Conversely, consider an example of a finite array which cannot be extended arbitrarily in the j-direction, say a field trial with blocks $i = 1, \dots, m$ and plots $j = 1, \dots, n$ within each block. In this case, the random effects interpretation is usually not meaningful, because the plots within a given block cannot be regarded as a sample from some infinite population. The intrablock variance component may very well turn out to be negative (negative correlation between plots in the same block), and even if it is positive, it cannot be taken as a measure of variation to be transferred to some other design with a different block size.

Model (2) is, mathematically, the nice one. Model (1) is a model for incomplete data from an infinite array model of type (2), and as such implies some mathematical problems (estimates on the boundary, etc.). These problems are similar to (though less serious than) those coming up when a stationary time series (also a nice model) is restricted to a finite time interval.

Institute of Mathematical Statistics
University of Copenhagen
Universitetsparken 5
2100 Copenhagen Ø
Denmark

Randall D. Tobias

SAS Institute, Inc.

It is a pleasure to have the opportunity to discuss this paper. Dr. Speed lucidly presents a concise and consistent notion of ANOVA which is yet broad enough to touch on time series and harmonic analysis of groups. His most important contribution, however, may be to have reminded us that the title question is still an open one. It is indeed astonishing how many apparently distinct general mathematical formulations the analysis of variance has received since it became a standard component in the statistical repertoire. I attribute the confusion that still persists about the nature of this essential technique largely to the fact that there are three different ways to construct and/or to justify a particular sum of squares decompositon. A decomposition may be intuitively reasonable: This is the view of ANOVA taken by many practitioners. It may be useful under an assumed stochastic model for testing certain hypotheses of interest about the data: This is the view of ANOVA which we as mathematical statisticians like to believe is the only correct one. And, finally, a decomposition may arise naturally from an abstract mathematical structure which has more or less explicit connections to both intuition and a stochastic model, the view of ANOVA examined by Dr. Speed in this paper. Moreover, theoretically and in particular troublesome cases, none of the three approaches commands absolute precedence. It is an empirical fact, for example, that the model which is assumed for a collection of data often depends on the ease and intuitive simplicity of the consequent analysis. In many cases, though, the analyses obtained by the three approaches coincide, and Speed's research can be regarded as a useful summary and unification of such situations.

However, an important conception of ANOVA is conspicuous by its absence from Speed's discussion, and the reader may be left wondering how all of this fits into what comes out on the *computer* as

ANALYSIS OF VARIANCE.

Rightly or wrongly, the idea that ANOVA is *identical* to regression analysis is probably the most widespread conception of what ANOVA is in general, a fact which is both revealed by and partly due to the acceptance of the results of popular general regression software, such as PROC GLM on SAS or ANOVA on GENSTAT, as general ANOVAs. The input for such programs is, formally, a linear model for the expected value of the data, of the form

$$E(\mathbf{y}) \in \mathscr{U}_0 + \mathscr{U}_1 + \cdots + \mathscr{U}_r, \tag{1}$$

where the $\mathscr{U}_i$'s are vector subspaces spanned by different subsets of the independent variables. The program then fits the data to the concentric spaces $\mathscr{V}_i = \sum_{j=0}^{i}\mathscr{U}_j$. Let $V_i\mathbf{y}$ denote the predicted data at the ith stage of the fit, where V_i is the projector onto $\mathscr{V}_i$. The ANOVA is taken to consist of the differences between the prediction sums of squares at successive stages of the fit, which is the same as

$$\|\mathbf{y}\|^2 = \|T_0\mathbf{y}\|^2 + \cdots + \|T_{r+1}\mathbf{y}\|^2, \tag{2}$$

where T_0 is the projector onto $\mathscr{T}_0 = \mathscr{V}_0$, $T_i = V_i - V_{i-1}$ is the projector onto $\mathscr{T}_i = \mathscr{V}_i \cap \mathscr{V}_{i-1}^{\perp}$ for $i = 1, \ldots, r$, and T_{r+1} the projector onto the residual space $\mathscr{T}_{r+1} = \mathscr{V}_r^{\perp}$. In what sense does Speed claim to have answered his title question, then, if he has not covered this common case, beyond passing such analyses off as "arbitrary?" I will try to indicate how a connection can be drawn.

Speed's definition of ANOVA consists essentially of restricting to association schemes a matrix formulation equivalent to the definition of Graybill and Hultquist (1961). Now, the association scheme approach is indisputably the correct one for summarizing, unifying and categorizing a large class of ANOVAs, as Speed convincingly demonstrates. I believe he errs, however, in separating ANOVA so completely from the traditional conception that he can call the mathematical constructions he derives for infinite random arrays true ANOVAs even though there are no sums of squares decompositions in sight. He justifies this separation by implying that the variance we are analyzing is a *parameter of the model* [(2.3) and (3.5)], whereas the word has traditionally been applied in this context to the observed variance, in the sense of variation, of the data—that is, the total sum of squares. That this latter interpretation prevails is evidenced by the notorious lack, for better or for worse, of explicit model assumptions in much practical ANOVA.

Well, as Speed says, one can define ANOVA to mean whatever one wishes. But the point is that direct consideration of sums of squares and their properties can lead to a much wider class of ANOVA than those which Speed covers, including in particular those with nontrivial expectation terms such as (1). Such ANOVAs are indeed arbitrary with respect to the general matrix formulation: They are associated with a particular but nonunique decomposition of the relationship algebra into one-sided ideals [see James (1957) and Tobias (1986)]. However, as decompositions of the total sum of squares, these ANOVAs can be uniquely specified by the characteristics of the component sums of squares. For example, under a Gaussian model with the trivial dispersion structure $V = \sigma^2 I$, the decomposition (2) satisfies the following:

(3a) all the component sums of squares are multiples of (possibly noncentral) χ^2 random variates;

(3b) $\|T_{r+1}\mathbf{y}\|^2/\sigma^2$ is a central χ^2 random variate, with maximal degrees of freedom subject to this centrality; and

(3c) $E\|T_i\mathbf{y}\|^2/\text{trace}(T_i) \geq \sigma^2$ where the inequality is strict if and only if $E(\mathbf{y})$ has a component in $\mathscr{T}_i$, for $i = 0, \dots, r$.

Thus, we can compare $\|T_i\mathbf{y}\|^2/\text{trace}(T_i)$, the mean square due to the ith expectation component, to an unbiased estimate of σ^2, $\|T_{r+1}\mathbf{y}\|^2/\text{trace}(T_{r+1})$, the residual mean square, to see whether $\mathscr{U}_i$ is needed in the model after fitting $\mathscr{U}_0, \dots, \mathscr{U}_{i-1}$. Furthermore, (2) is the *only* decomposition satisyfing (3). A corresponding result holds for the more complicated dispersion models which Speed considers, and it forms the basis of nonorthogonal mixed-effect ANOVA. It should be recognized that Speed himself has participated in some of the most important recent work on nonorthogonal ANOVAs, those associated with *generally balanced* designs [Speed (1983) and Houtman and Speed (1983)]. In the spirit of Speed's equation (3.2), a necessary and sufficient matrix condition for the existence of ANOVA's such as we are thinking of can be given, and it turns out to be a natural extension of the condition of general balance. See Tobias (1986) for details.

So what is the relation between expectation-model ANOVA (E-ANOVA) and dispersion-model ANOVA (D-ANOVA)? Superficially, they share characterization as decompositions of the identity matrix into orthogonal idempotents, but the decompositions arise in different ways: For an E-ANOVA the idempotents are projectors onto spaces which depend on the order in which model terms are fit; while for D-ANOVA they are projectors onto common eigenspaces and are independent of any computational method. The connection is just this: For many common structures there is a natural order in which to fit terms which gives rise to the same projectors as those derived from the full spectral analysis. Such structures include, in particular, all of the classical factorial structures and their finite extensions discussed by Speed in Section 6, in which the terms of the expectation model (1) correspond to filters. The coincidence of the two approaches is due to several special characteristics of these structures. For one thing, all of the parameters associated with any particular term in the model are equally replicated, so that the projector for that term is a scalar multiple of the relationship matrix associated with the corresponding filter [Speed's R_b's in (6.2)]. The commutativity of these matrices is again also important, and finally, there is the fact that for any two terms $\mathscr{U}_i$ and $\mathscr{U}_j$ in the model, $\mathscr{U}_i \cap \mathscr{U}_j$ is also in the model: This is equivalent to the closure of $L(F)$ with respect to intersection. See Tjur (1984) for details. In fact, in this case a fitting order which will make the E-ANOVA and the D-ANOVA coincide is just any well-ordering of the terms which contains the partial-ordering on the filters. The emphasis on a particular "natural" order is important: There are other conceivable fitting orders which will not give rise to the eigenspace projectors, and in some situations such alternatives will be appropriate [see Nelder (1977), where the problem is discussed under the heading of *marginality*]. Only in this sense are E-ANOVAs arbitrary, that they require, in addition to the terms themselves, specification of the fitting order.

I have two further observations on the paper. First, I regard the restriction of the general matrix formulation of D-ANOVA to association schemes as unnecessarily limiting: It adds no substantial mathematical or intuitive apparatus to the problem, and if it comes to approaches which include "almost all examples" a great deal less generality is really needed, probably no more than Nelder's *simple block structures* [Nelder (1965)]. Personally, I find the core mathematical requirement for the existence of such ANOVA's in (3.3) and assumption (iv) for association schemes, which can be stated succintly as:

The symmetric matrices A_a span their algebraic closure.

Finally, the discussion in Section 7 of the connection between ANOVA and decompositions of groups is especially fortunate. Since James' forward-looking 1957 paper on the similar relation between ANOVA and decompositions of matrix algebras, the idea has been whittled at by various people, but the explicit relation between Gel'fand pairs and ANOVA in general has not, to my knowledge, been drawn before. It is a bit discouraging to find that this mathematically interesting correspondence does not appear to be statistically fruitful—that the group theory is much harder than the standard combinatorial approach and does not buy anything in the end. Still, it should not be ignored that the *symmetries* of a collection of data—that is, the permutations with respect to which it is invariant, in the sense of the statistical information it contains—are often the most basic way of getting at the structure of the collection. Certainly the dispersion model basis matrices A_a do not appear intuitively in the mind of the experimenter; the individual factors as equivalence relations might, but not their nesting structure, at least not for structures of any complexity. But, I assert, what the experimenter should always be able to answer, upon a little reflection, is the question, "In which ways can we arbitrarily swap around the data without affecting the conclusions we should make?" Thus, the group theory approach may have useful ramifications for practical statistical consulting, in discovering in the first place from the experimenter what the structure of the data is.

REFERENCES

GRAYBILL, F. A. and HULTQUIST, R. A. (1961). Theorems concerning Eisenhart's model II. *Ann. Math. Statist.* **32** 261–269.

HOUTMAN, A. M. and SPEED, T. P. (1983). Balance in designed experiments with orthogonal block structure. *Ann. Statist.* **11** 1069–1085.

JAMES A. T. (1957). The relationship algebra of an experimental design. *Ann. Math. Statist.* **28** 993–1002.

NELDER, J. A. (1965). The analysis of randomised experiments with orthogonal block structure. I and II. *Proc. Roy. Soc. London Ser. A* **283** 147–178.

NELDER, J. A. (1977). A reformulation of linear models (with discussion). *J. Roy. Statist. Soc. Ser. A* **140** 48–76.

SPEED, T. P. (1983). General balance. In *Encyclopedia of Statistical Sciences* (S. Kotz and N. L. Johnson, eds.) **3** 320–326. Wiley, New York.

TJUR, T. (1984). Analysis of variance models in orthogonal desings. *Internat. Statist. Rev.* **52** 33–81.

TOBIAS, R. D. (1986). The algebra of a multi-stratum design and the application of its structure to analysis. Ph.D. thesis, Univ. of North Carolina.

SAS INSTITUTE INC.
BOX 8000
CARY, NORTH CAROLINA 27511-8000

JOHN W. TUKEY

Princeton University

We must respect the many long steps that Speed has taken to understand, focus and describe a mathematical structure for what R. A. Fisher may have sensed in introducing the analysis of variance before 1925. But we dare not regard it as telling us why the analysis of variance deserves the great practical importance that it has held throughout recent decades.

I am not equipped to comment adequately on the mathematical niceties and careful craftmanship of Speed's paper. I do have an obligation, however, to point out why what he describes as the analysis of variance is *not* the core of what is practiced in so many areas of application.