EFRON, B. (1971). Does an observed sequence of numbers follow a simple rule? (Another look at Bode's law). *J. Amer. Statist. Assoc.* **66** 552–559.

GILULA, Z. and HABERMAN, S. (1984). Canonical analysis of contingency tables by maximum likelihood. Unpublished manuscript.

GOODMAN, L. A. (1969). On partitioning χ^2 and detecting partial association in three-way contingency tables. *J. Roy. Statist. Soc. Ser. B* **31** 486–498.

GOODMAN, L. A. (1970). The multivariate analysis of qualitative data: Interactions among multiple classifications. *J. Amer. Statist. Assoc.* **65** 226–256.

GOODMAN, L. A. (1971). The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications. *Technometrics* **13** 33–61.

GOODMAN, L. A. (1978). *Analyzing Qualitative/Categorical Data: Log-Linear Models and Latent-Structure Analysis*. Abt Books, Cambridge, Mass.

GOODMAN, L. A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* **74** 537–552.

GOODMAN, L. A. (1981). Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* **76** 320–334.

GOODMAN, L. A. (1984). *The Analysis of Cross-Classified Data Having Ordered Categories*. Harvard University Press, Cambridge, Mass.

GOODMAN, L. A. (1985). The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *Ann. Statist.* **13** 10–69.

GOODMAN, L. A. and HABERMAN, S. J. (1985). On the analysis of non-additivity in two-way analysis of variance. Unpublished manuscript.

GREENACRE, M. J. (1984). *Correspondence Analysis*. Academic, New York.

HOTELLING, H. (1939). Tubes and spheres in n-spaces, and a class of statistical problems. *Amer. J. Math.* **61** 440–460.

JOHNSON, D. E. and GRAYBILL, F. A. (1972). An analysis of a two-way model with interaction and no replication. *J. Amer. Statist. Assoc.* **67** 862–868.

LEBART, L., MORINEAU, A. and WARWICK, K. M. (1984). *Multivariate Descriptive Statistical Analysis*. Wiley, New York.

MANDEL, J. (1971). A new analysis of variance model for non-additive data. *Technometrics* **13** 1–18.

NISHISATO, S. (1980). *Analysis of Categorical Data: Dual Scaling and Its Applications*. University of Toronto Press, Toronto.

SNEE, R. (1974). Graphical display of two-way contingency tables. *Amer. Statist.* **28** 9–12.

TUKEY, J. W. (1949). One degree of freedom for nonadditivity. *Biometrics* **5,** 232–224.

TUKEY, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, Mass.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637

TOM LEONARD[1]

University of Wisconsin-Madison

This exciting and imaginative paper promises substantial impact upon the practical application of contingency table methodology. It highlights the difficulty for the user in making an overall judgement concerning a reduced model, by reference to the tricky (and dare I say virtually impossible?) interpretation,

[1] This work was supported under Contract N00014-83-C-0514 between the Office of Naval Research and the University of Iowa, Melvin R. Novick, Principal Investigator, while the author was visiting the Lindquist Center for Measurement, University of Iowa.

regarding the smallness of a single p-value, e.g., after considering numbers of dimensions, sample size, sampling scheme, scientific background, scientific objective, and so on. The broad ranging synthesis of ideas described by the authors promises advances beyond the existing classical log-linear/chi-square methodology and brings fresh philosophy into an area otherwise getting a bit inbred since the simple theoretical developments of the late 1960s.

When I introduced the authors' exponential alternative in a Bayesian context in Leonard (1977) via derivations relying on similar classical procedures to those described in Sections 4 and 5, I realized that the main result (that under appropriate mixtures the asymptotic distribution of chi-square is a scale multiple of chi-square) was well known in the frequency literature. See, for example, Paul and Plackett (1978) and Bateman (1950). The new p-value interpretations are an easy consequence of this and uniformity is just a limiting case. Therefore, Diaconis and Efron's volume test seems to be a standard test in the literature. However, their fascinating mathematical development, particularly in later sections, is to be highly commended.

My own statistical philosophy is to permit as general an alternative model as possible a priori, then use the data, together with the practical background, to suggest possible modifications to the hypothesised model. The best sort of mechanism for this type of investigation seems to be a residual/interaction analysis which considers the adequacy of the reduced model for each individual cell. Unfortunately, the classical log-linear approach does not permit precise marginal inferences concerning individual residuals or interactions since the various, e.g., normal, approximations in the literature are fairly inaccurate for practical sample sizes. I believe that an accurate residual analysis would solve many of the practical problems. The emphasis should be on posterior inferences about alternative hypotheses, rather than possibly artificial prior hypotheses. For example, uniformity seems a bit out of place for the author's hair/eye color example.

One possibility is an approach synthesised by Leonard and Novick (1985) but also relating to the many Bayesian developments in this area during the 1970s. Take the (conditional) prior mean ξ_{ij} of the (i, j)th cell probability to satisfy

$$\xi_{ij} = \xi_{ij}(\beta) \quad (i = 1, \cdots, I; j = 1, \cdots, J) \tag{1}$$

where the function $\xi_{ij}(\cdot)$ is specified according to the reduced form of the model, and the $\ell \times 1$ vector β represents the parameters within this model. Suppose that, given α and β, the θ_{ij} follow the Dirichlet distribution which possesses these prior means and prior variances $\xi_{ij}(1 - \xi_{ij})/(\alpha + 1)$. Then the posterior mean of θ_{ij}, given α and β is

$$E(\theta_{ij} \mid y_{ij}, \alpha, \beta) = (1 - \tau)(y_{ij}/n) + \tau\xi_{ij}(\beta) \tag{2}$$

where $\tau = \alpha/(n + \alpha)$ represents a proportionate mixing between the reduced model $\xi_{ij}(\cdot)$ and a general alternative as represented by the y_{ij}. The procedure can be rendered largely data analytic by assigning appropriate ignorance priors to α and β at the lower stage of a hierarchical model. The class of alternative models is then highly diffuse, and involves a broad mixture of multinomial-

Dirichlet distributions. Diaconis and Efron's apparently weak notion of uniformity is in fact a much stronger assumption and corresponds to the choice $\alpha\xi_{ij} \equiv 1$.

The marginal posterior density of τ now provides a mechanism for an overall inferential check for the adequacy of the reduced model, compensating for sample size and dimensionality and providing much more information to the user than given by a single p-value; the user should combine this with his or her practical experience. The marginal posterior densities of the *parametric residuals*.

$$(3) \qquad \rho_{ij} = \log \theta_{ij} - \log \xi_{ij}(\beta) \quad (i = 1, \cdots, I; j = 1, \cdots, J)$$

permit a detailed residual analysis investigating specific deviations from the reduced model and suggesting meaningful alternative models.

Leonard and Novick use the conditional maximization procedure recommended by Leonard (1982) to closely approximate all these marginal posterior densities under the assumptions above described. In the special case where $\xi_{ij}(\cdot)$ represents the independence model, ρ_{ij} in (3) is just the (i, j)th interaction effect under log-linear assumptions.

The data in Table 1 are more fully described and analysed by Leonard and Novick. The observed frequency for the (i, j)th cell gives the number of Marine Corps students obtaining the jth grade on a common aptitude test prior to entry at one of 12 clerical schools. The table is a cross-section of a $12 \times 8 \times 5$ three-way table also classifying according to final grade obtained at each of the 12 schools.

The above hierarchical Bayesian analysis was performed with the null hypothesis in (1) representing independence of rows and columns and the ρ_{ij} in (3) denoting the log-linear interaction effects. For an overall check, an approximate posterior density of the mixing proportion τ in (2) is given in Figure 1, with a mean of 0.172. This gives substantial evidence that the independence model is inappropriate. In general, a similar conclusion would seem reasonable if the posterior probability that $\tau \leq 0.5$ were greater than 0.5. The chi-square value is

TABLE 1
The Marine Corps data

	Grade							
School	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
A	20	179	276	316	123	27	10	4
B	10	80	112	112	6	5	0	1
C	25	293	390	337	126	66	32	18
D	3	32	55	51	17	7	3	2
E	2	41	46	43	12	4	0	0
F	10	81	138	242	145	32	6	2
G	9	131	270	263	100	39	39	12
H	3	35	57	64	21	7	3	2
I	2	38	69	45	15	8	1	2
J	1	28	49	81	32	28	10	4
K	1	29	51	56	37	18	6	2
L	0	48	87	162	62	71	21	7

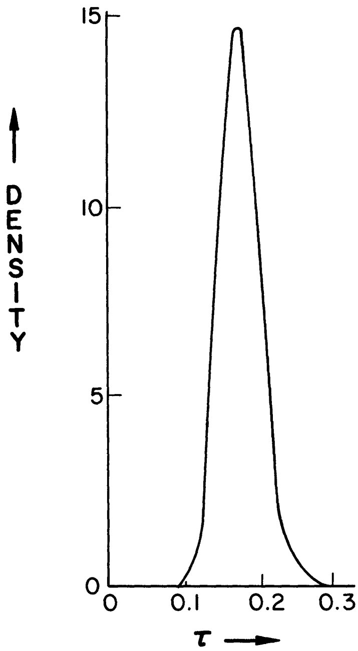

FIG. 1. *Posterior Density of Mixing Proportion J*

$X^2 = 456.93$ with 77 degrees of freedom again refuting independence. It is, however, possible to avoid consideration of the usual p-value by instead assessing the whole posterior curve for τ which summarises the information in the data regarding overall deviations from the independence model, given the reasonability of the distributional assumptions. If required, a variety of special decision rules could be based on this curve, but formal decision theory seems unnecessary in this practical context.

For individual deviations from the null model consider the eight approximate posterior densities in Figure 2, which corresponds to the interaction effects for the eight grades obtainable by students at school L. For the highest three grades there are clear negative interactions, and for the lowest three grades there are strong positive interactions, suggesting a clear individual deviation from the null hypothesis of independence of the entry performances for the 12 schools. For the fourth and fifth grades, the interactions seem positive but the evidence is not overwhelming. Judgement of the tail probabilities seems less important than an overall appraisal of the curves and the detection of patterns in the interactions when compared with other schools.

In Table 2 a summary of the complete interaction analysis is described. A boxed +, 0, or − means that the judgement on this particular interaction is not completely clear. The patterns in the residuals suggest to me that

(a) The first three grades are good for judging superior performances and the last five grades for judging inferior performances.

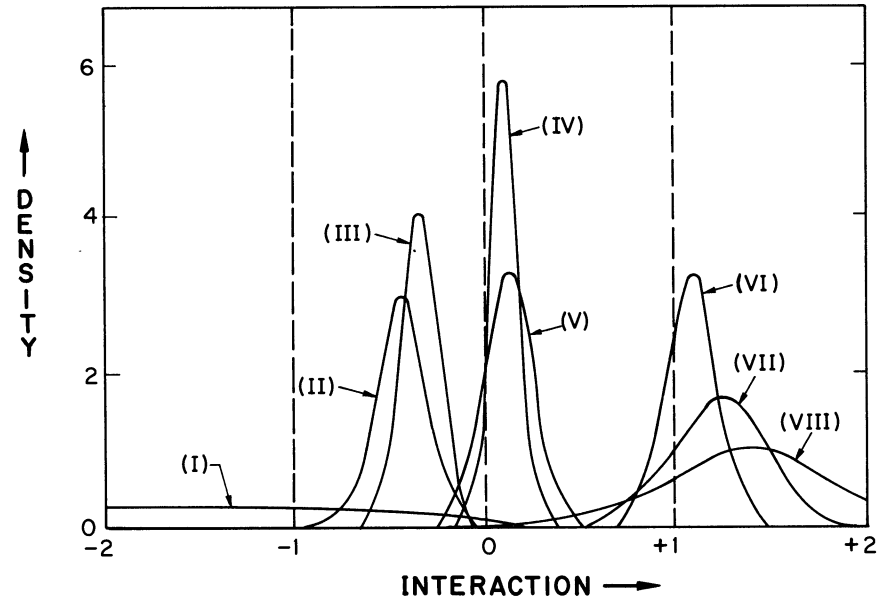

FIG. 2. *Eight Posterior Densities of Interaction Effects for School L*

TABLE 2
Interaction analysis

	1	2	3	4	5	6	7	8
B	+	+	+	[0]	−	−	−	0
C	[+]	+	0	−	[−]	0	+	+
E	0	+	0	0	0	0	−	−
I	0	[+]	+	[−]	0	0	0	0
A	+	0	0	0	0	−	[−]	0
D	0	0	0	0	0	0	0	0
G	0	0	[+]	0	0	0	0	+
H	0	0	0	0	0	0	0	0
F	0	−	−	+	+	0	0	0
J	0	[−]	[−]	[+]	[+]	+	+	+
K	0	0	0	0	+	+	[+]	0
L	−	−	−	[+]	[+]	+	+	+

(b) Four of the schools (B, C, E, I) should be grouped as possessing superior entry levels. Four schools (A, D, G, H) are average and four (F, J, K, L) are inferior.

This interaction analysis finally led us to a choice of alternative model where (i) the first three columns and final five columns of the table are collapsed and (ii) there is independence of entry level within the three groups of schools in (b) but not between groups. A further technical analysis confirmed the adequacy of our simplified model; the posterior mean of τ was now 0.60. This enabled us to obtain a simple analysis for the full $12 \times 8 \times 5$ table and to investigate the associations between entry levels and final grades.

My overall conclusion is that most observed contingency tables possess intrinsically individualistic structures which should not be concealed by unduly constraining alternative hypotheses in advance. Diaconis and Efron are, however, taking us in a good perceptive direction which should yield fresh advances in the future.

Acknowledgements. The collaboration of Melvin R. Novick, Stephen B. Dunbar, and Ming-Mei Wang on the analysis of the Marine Corps data is gratefully acknowledged.

REFERENCES

BATEMAN, G. I. (1950). The power of the χ^2 index of dispersion test when Neyman's contagious distribution is the alternative hypothesis. *Biometrika* **37** 59–63.

LEONARD, T. (1977). A Bayesian approach to some multinomial estimation and pre-testing problems. *J. Amer. Statist. Assoc.* **72** 869–874.

LEONARD, T. (1982). Comment on the paper by Lejeune and Faulkenberry. *J. Amer. Statist. Assoc.* **77** 657–658.

LEONARD, T. and NOVICK, M. R. (1985). Bayesian full rank marginalization for two-way contingency tables. To appear in *J. of Educ. Statist.*

PAUL, S. R. and PLACKETT, R. L. (1978). Inference insensitivity for Poisson Mixtures. *Biometrika* **65** 591–601.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN-MADISON
MADISON, WISCONSIN 53706

PETER MCCULLAGH AND DARYL PREGIBON

AT&T Bell Laboratories

Some years ago, a well-known Irish newspaper carried a series of advertisements by an eccentric entrepreneur known as "The Brother," offering correspondence courses in the art of "periastral peregrinations on the aes ductile," more commonly, but less accurately, known as tight-rope walking. Despite the incentive of generous course credit, the University of the Air, as it was known, had few registered students and no known graduates. In the present paper, Diaconis and Efron give a superb exhibition of The Brother's singular art in its metaphorical form, by attempting to dance on two ropes at once—and almost succeeding!

Diaconis and Efron have chosen, quite sensibly, not to argue against modelling departures from independence, noting that such models can often give deeper insights into the data. Instead, they emphasize the common χ^2 statistic, here denoted by X^2, as "an effective device for preliminary data analysis, particularly when the statistician has many two-way tables under review." This point of view seems difficult to comprehend because the most common and compelling objection to the use of X^2 in applications is that it gives no information regarding the