THE LACK OF CONSISTENCY FOR STATISTICAL DECISION PROCEDURES

D.B. Haunsperger and D.G. Saari
Department of Mathematics, Northwestern University, Evanston, IL, USA

Research Reports, which record research conducted at IIASA, are independently reviewed before publication. However, the views and opinions they express are not necessarily those of the Institute or the National Member Organizations that support it.

Foreword

My colleagues at IIASA, Sergei Scherbov, Wolfgang Lutz and Evgenii Andreev, were engaged in projecting future Soviet population about three years ago, and they came up with a remarkable result: that period life expectancy when calculated for the USSR as a whole was lower than life expectancy for any one of the republics taken individually. (The result is shortly to appear in print, "Sensitivity of Aggregate Period Life Expectancy to Different Averaging Procedures" by W. Lutz and S. Scherbov. Forthcoming in *Population Bulletin of the United Nations*.) Their first reaction was that an arithmetical error had been made, but after thoroughly verifying the arithmetic the paradox remained.

It seemed, and still seems, extraordinary that if the populations of two republics are joined together and their statistics consolidated, mortality should go up just because of the consolidation. No one's chance of dying has increased, so the phenomenon must be a statistical artifact, perhaps an intrinsic defect of the method of demographic calculation. It is possible, because life expectancy is a nonlinear function of the deaths and population from which it is calculated.

Such paradoxes are disturbing to those who work with statistical data. According to intuition an average ought to be higher than the lowest of the items averaged, and lower than the highest. Donald Saari of Northwestern University has worked on similar matters at Northwestern University and at IIASA, and has turned up a variety of such dramatic paradoxes. Deanna Haunsperger, as a student in the Young Scientists' Summer Program at IIASA in 1990, took up the problem in its more general form, and made enough progress to win a Peccei award that enabled her to continue the work in the summer of 1991.

The outcome is the present paper giving the conditions under which the phenomenon can arise, and the direction it will take. It relieves the discomfort scholars feel when using arithmetic methods which are not fully

understood. By showing the very general conditions under which the phenomenon can arise, Haunsperger has made a contribution to statistical and demographic technique.

NATHAN KEYFITZ
Leader
Population Program

The Lack of Consistency for Statistical Decision Procedures

DEANNA B. HAUNSPERGER and DONALD G. SAARI*

Simpson's paradox exhibits seemingly deviant behavior where the data generated in independent experiments support a common decision, but the aggregated data support a different outcome. It is shown that this kind of inconsistent behavior occurs with many, if not most, statistical decision processes. Examples are given for the Kruskal–Wallis test and a Bayesian decision problem. A simple theory is given that permits one to determine whether a given decision process admits such inconsistencies, to construct examples, and to find data restrictions that avoid such outcomes.

KEY WORDS: Bayesian decision theory; Kruskal–Wallis test; Simpson's paradox.

1. THE BASIC PROBLEM AND SOME EXAMPLES

For excellent reasons, Simpson's (or Yule's) paradox continues to attract attention. This is because it exhibits the perverse behavior where the data for each of the "parts" support a common decision, yet the data for the "whole" imply a different conclusion. [See, for example, Bickel, Hammel, and O'Connell (1975), Blyth (1972), Cohen (1986), Good and Mittal (1987), Saari (1987, 1988, 1990), Vaupel and Yashin (1985), Wagner (1982), and their cited references.] For instance, suppose that when treatment X is compared with a standard approach in Evanston, 90 of the 240 X subjects (over 1/3) regained health as compared to only 20 of the 60 standard approach subjects (1/3). In Chicago, 30 of the 60 X subjects (1/2) regained health as compared to 110 of the 240 standard approach subjects (less than 1/2). Although X does better in both locales, the aggregated data support the standard treatment. (For the standard treatment, 130/300 regained health compared to 120/300 of the X subjects.) There are several explanations for this paradox; we offer a new, geometric explanation that is particularly simple.

One might wonder whether the perverse behavior exhibited by Simpson's paradox is restricted to procedures of the above kind or whether it identifies an inherent problem with statistical decision methods. It is the latter; we contend that many, if not most, statistical decision processes can be plagued by this behavior where the conclusion of the aggregated data differs from a common

conclusion of the subpopulations. We prove this with examples and with a simple theory that explains the source of the difficulties. As we show, one cause is the nonlinear form of the equations that necessarily arises when data are aggregated. Also, we indicate how to use this theory to generate examples and to find conditions that avoid such outcomes.

To illustrate our assertion we show that the well-known Kruskal–Wallis (KW) test from nonparametric statistics also exhibits an inconsistency between the parts and the whole. In the KW test, observed data are replaced with "ranks." The final ranking is found by summing the assigned ranks; a larger total represents a more preferred choice. Thus the data representing the performance of three machines are

Machine 1	Machine 2	Machine 3
5.89	5.81	5.80
5.98	5.90	5.99

and they define the KW table of ranks

Machine 1	Machine 2	Machine 3
3	2	1
5	4	6

in which the KW ranking is Machine 1 > Machine 3 > Machine 2 with the tally of $8:7:6$—Machine 1 is the top choice. An identical KW table of ranks, with Machine 1 as top ranked, arises with

Machine 1	Machine 2	Machine 3
5.69	5.63	5.62
5.74	5.71	6.00

With the combined data, however, Machine 3 is top ranked as the KW ranking is Machine 3 > Machine 1 > Machine 2 with the tally $30:26:22$. Thus the flavor of Simpson's paradox extends to the KW test.

2. WEAK CONSISTENCY

To model this behavior for the $n \geq 2$ alternatives $A^n = \{a_1, \ldots, a_n\}$, let $P(A^n)$ be the set of all $2^n - 1$ nonempty subsets of A^n. Here, $P(A^n)$ is the *decision space* for a procedure. For instance, if a_j denotes Machine j, then a singleton in $P(A^3)$ identifies the single selected machine while a pair identifies the two selected machines. The domain (sample space, space of probability distributions, space of ranks, and so on) is represented by a set S. On S, a closed, commutative, associative binary operation "$\circ$" describes how elements of S are combined.

Example. For the procedure of Simpson's paradox, an element of S—a *profile*—is a four-tuple $(u_1, x_1; u_2, x_2)$, where u_j $(j = 1, 2)$ is the number of subjects assigned to the jth treatment and x_j is the fraction of these

subjects judged successful. The binary operation, defined by Bayes's rule, is

$$(u_1^1, x_1^1; u_2^1, x_2^1) \circ (u_1^2, x_1^2; u_2^2, x_2^2)$$
$$= \left(u_1^1 + u_1^2, \frac{x_1^1 u_1^1 + x_1^2 u_1^2}{u_1^1 + u_1^2}; \quad u_2^1 + u_2^2, \frac{x_2^1 u_2^1 + x_2^2 u_2^2}{u_2^1 + u_2^2} \right). \quad (1)$$

For a Bayesian decision process, S is a space of probability distributions for a relevant parameter. The binary operation is based on how a data set determines a posterior distribution from a given prior and a likelihood function. If the distribution is based on θ with values $\{\theta_i\}_{i=1}^m$, if $\xi(\theta) \in S$ is a prior distribution, if $\{\mathbf{y}^j\}_{j=1,2}$ represent the data from independent experiments, and if the posterior distribution based on the data set $\mathbf{y}^j$ is $\xi(\theta \mid \mathbf{y}^j) \in S$, then the binary operation is

$$\xi(\theta \mid \mathbf{y}^1) \circ \xi(\theta \mid \mathbf{y}^2) = \xi(\theta \mid \mathbf{y}^1 + \mathbf{y}^2). \quad (2)$$

Here $\mathbf{y}^1 + \mathbf{y}^2$ represents the aggregated data of both data sets.

Definition. A mapping

$$f : S \to P(A^n) \quad (3)$$

is called a *statistical choice function*. A choice function satisfies the *weak consistency conditions* if (a) f is nonconstant (i.e., at least two sets from $P(A^n)$ can be selected) and if (b) $\mathbf{x}^1, \mathbf{x}^2 \in S$ are such that $f(\mathbf{x}^1) = f(\mathbf{x}^2)$, then $f(\mathbf{x}^1) = f(\mathbf{x}^1 \circ \mathbf{x}^2)$.

In other words, a statistical choice function is weakly consistent if when the parts agree $(f(\mathbf{x}^1) = f(\mathbf{x}^2))$, then this is the conclusion for the aggregated data $(f(\mathbf{x}^1) = f(\mathbf{x}^2) = f(\mathbf{x}^1 \circ \mathbf{x}^2))$. The weakly consistent statistical choice functions are precisely the procedures that avoid the kind of difficulties exhibited by Simpson's paradox.

Definition. A set $\Omega \subset S$ is *algebraically closed* with respect to "$\circ$" if when $\mathbf{x}^1, \mathbf{x}^2 \in \Omega$, then $\mathbf{x}^1 \circ \mathbf{x}^2 \in \Omega$.

The cone of a set Ω, denoted by Co(Ω), is the set $\{n\mathbf{x}^1 \circ m\mathbf{x}^2 : \mathbf{x}^1, \mathbf{x}^2 \in \Omega, n, m$ are positive integers$\}$. Here, $n\mathbf{x}$ is the *n*-fold replication $\mathbf{x} \circ \cdots \circ \mathbf{x}$.

The cone Co(Ω) identifies all the ways profiles from Ω can be combined to obtain new profiles. Thus for $\alpha \in P(A^n)$, the portion of Co$(f^{-1}(\alpha))$ outside of $f^{-1}(\alpha)$ identifies examples where weak consistency is violated. It follows, therefore, that Ω is algebraically closed if and only if $\Omega = $ Co(Ω). This leads to the following easily proved theorem used in Saari (1990) to study general classes of decision and election procedures.

Theorem. A statistical choice function $f : S \to P(A^n)$ satisfies the weak consistency condition iff for each $\alpha \in P(A^n)$, the corresponding "decision region" $f^{-1}(\alpha) \subset S$ is algebraically closed with respect to the binary operation. This condition holds iff

$$f^{-1}(\alpha) = \text{Co}(f^{-1}(\alpha)). \quad (4)$$

3. ILLUSTRATING THE THEOREM WITH SIMPSON'S PARADOX

To see how this simple theorem explains Simpson's paradox, assume that two treatments a_1, a_2 are compared, where x_j is the fraction of success with treatment a_j. The variables (x_1, x_2) define a point $\mathbf{x}$ in the unit square $U = [0, 1] \times [0, 1]$, where $a_1 > a_2$ iff $x_1 > x_2$, $a_2 > a_1$ iff $x_2 > x_1$, and $a_1 \sim a_2$ iff $x_1 = x_2$. Each decision region is determined by the diagonal line $x_1 = x_2$ in U; see Figure 1. For instance, $f^{-1}(a_1)$ is the convex set $\Omega = \{(x_1, x_2) \in U : x_1 > x_2\}$.

We know from Simpson's paradox that this procedure is not weakly consistent, so $\Omega \neq \text{Co}(\Omega)$. To see why, let $\mathbf{x}^j = (x_1^j, x_2^j) \in U$ represent the set of data $(u_1^j, x_1^j; u_2^j, x_2^j)$ for the jth site $(j = 1, 2)$. In this set, the x_i^j values remain fixed, but the integers u_i^j vary over all positive values. Thus $\mathbf{x}^j$ represents all choices of the u_i^j's, while the fraction of success for each treatment is given by the specified components of $\mathbf{x}^j$. According to the definition of "$\circ$,"

$$\text{Co}(\mathbf{x}^1, \mathbf{x}^2) = \left\{ (sx_1^1 + (1-s)x_1^2, tx_2^1 + (1-t)x_2^2) \; \middle| \right.$$
$$\left. s = \frac{u_1^1}{u_1^1 + u_1^2}, t = \frac{u_2^1}{u_2^1 + u_2^2} \right\}, \quad (5)$$

where integers u_i^j $(i, j = 1, 2)$ can assume any positive value. This means that s is the fraction of all subjects assigned to treatment one that are at the first locale, while t is a similar fraction for the treatment two. Without restrictions on the assignment process, there is no relationship between the values of s and t. Consequently the s variable can assume any (rational) value in $[0, 1]$, while t has a fixed value; this defines a horizontal line segment, depicted in Figure 1, where the height of the line is determined by t. By varying the value of t, it follows that Co$(\mathbf{x}^1, \mathbf{x}^2)$ is the unique rectangle in U (with edges parallel to the coordinate axis) with $\{\mathbf{x}^1, \mathbf{x}^2\}$ as vertices; see Figure 1. So, if $B \subset U$, then Co(B) is the union of
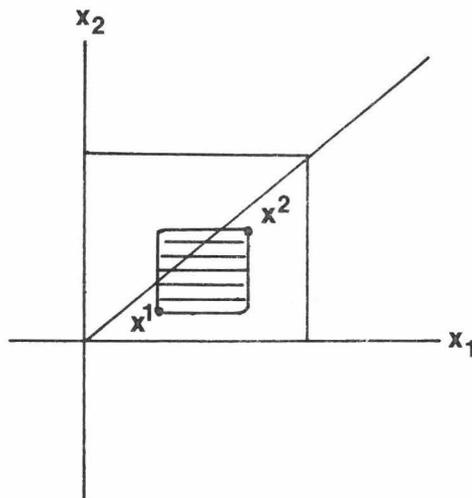
Figure 1. Decision Region. The cone defined by x^1, x^2 extends into the $a_2 > a_1$ region.

all rectangles defined by pairs of points in B. In particular the cone condition does not hold because

$$\text{Co}(\Omega) = U \neq \Omega. \qquad (6)$$

In addition to proving that this decision process is not weakly consistent, the geometry of the cone can be used to construct examples that illustrate the paradox. To do so, select $\mathbf{x}^1, \mathbf{x}^2 \in \Omega = f^{-1}(a_1)$, where $\text{Co}(\mathbf{x}^1, \mathbf{x}^2)$ extends into $f^{-1}(a_2)$, $x_2 > x_1$. Choose $\mathbf{x} \in \text{Co}(\mathbf{x}^1, \mathbf{x}^2) \cap f^{-1}(a_2)$ and find its unique (s, t) coordinates. (Recall that s is the fractional horizontal position of $\mathbf{x}$ in $\text{Co}(\mathbf{x}^1, \mathbf{x}^2)$, while t is the fractional vertical position.) These (s, t) values determine [from Eq. (5)] a choice of the values $(u_1^1, u_2^1; u_1^2, u_2^2)$ that creates the example $\mathbf{x}$. Extreme examples occur if $\{\mathbf{x}^j\}_{j=1,2} \in \Omega$ are close to the diagonal of U but far from one another—here the set $\text{Co}(\mathbf{x}^1, \mathbf{x}^2)$ almost equals U.

The cone property can also be used to derive conditions for the design of an experiment to avoid the paradox. The idea is simple; just impose the appropriate restrictions so that the cone property is satisfied. As the $\{x_i^j\}$ values represent experimental data from the treatment, they are not subject to control. Thus the only variables at our "design" disposal are $\{u_i^j\}$—the numbers of subjects assigned to each treatment at each site. According to Equation (5), restrictions on the $\{u_i^j\}$ variables impose restrictions on the s, t variables. This permits the analysis to be reduced to a geometric one; find those restrictions on the s, t variables that permit the cone condition to be satisfied always.

Let $\text{Co}_{\{R\}}(\Omega)$ represent the algebraically closed set defined by Ω when the restrictions R are imposed upon the values of s, t. For instance, if $R_L = \{(s, t) \mid s = t, t \in [0, 1]\}$, then $\text{Co}_{\{R_L\}}(\mathbf{x}^1, \mathbf{x}^2)$ is the line segment $\{t\mathbf{x}^1 + (1 - t)\mathbf{x}^2 \mid t \in [0, 1]\}$ connecting $\mathbf{x}^1, \mathbf{x}^2$. To avoid Simpson's paradox, the objective, then, is to determine those restrictions, R, such that whenever $\mathbf{x}^1, \mathbf{x}^2$ are in the same decision region, $\text{Co}_{\{R\}}(\mathbf{x}^1, \mathbf{x}^2)$ is contained in the same decision region. For instance, because $\text{Co}_{\{R_L\}}(\mathbf{x}^1, \mathbf{x}^2)$ is the straight line segment connecting $\mathbf{x}^1, \mathbf{x}^2$, this condition is always satisfied by R_L. Thus $R_L = \{(s, t) \mid s = t\}$ is a sufficient condition to avoid Simpson's paradox.

The R_L restriction also is a necessary condition to avoid the paradox. This is because the constraint on the choice of the restrictions, R, must hold for all values of $\mathbf{x}^1, \mathbf{x}^2$ when both are in the same decision region. In particular, this must be true when both $\mathbf{x}^1, \mathbf{x}^2$ are on the diagonal (the decision region for $a_1 \sim a_2$). But if $\mathbf{x}^1, \mathbf{x}^2$ are on the diagonal, then R must be such that $\text{Co}_{\{R\}}(\mathbf{x}^1, \mathbf{x}^2)$ is also on the diagonal; thus it is necessary for $\text{Co}_{\{R\}}(\mathbf{x}^1, \mathbf{x}^2)$ to be on the straight line segment joining $\mathbf{x}^1, \mathbf{x}^2$. It now is easy to see that, if no other restrictions are imposed on the values of $\{u_i^j\}$, then R_L is the most general set of restrictions that avoids Simpson's paradox. With the use of simple algebra, this necessary and sufficient restriction $(s = t)$ to avoid Simpson's paradox is equivalent to requiring that in each locale, the same ratio of subjects is assigned to treatment one.

Other kinds of results follow from the geometry. For instance, one might wish to replace the definition $a_i \sim$ a_j iff $|x_i - x_j| = 0$ with the more relaxed conditions $a_i \sim a_j$ iff $|x_i - x_j| \leq \beta$ and $a_i > a_j$ iff $x_i > x_j + \beta$. The idea is to choose a value of $\beta \geq 0$ to lessen the impact of Simpson's paradox. Namely, choose β so that if the conclusion of both subpopulations is $a_i > a_j$, then the aggregated outcome is either $a_i > a_j$ or $a_i \sim a_j$ but not $a_j > a_i$. The geometric description of the decision region for $a_1 \sim a_2$ is the diagonal band in U defined by $|x_i - x_j| \leq \beta$. Using the geometry of the cones and simple algebra it follows that this goal is attained iff $\beta \geq 1/3$.

This approach of using the geometry of the cone to explain Simpson's rule extends to other statistical decision processes. As examples, this approach can be used to rederive the interesting conclusions in Good and Mittal (1987), to explain the above KW example, and to characterize the component example of Kaigh (1989). Notice that the decision regions for many statistical procedures are convex sets, while the binary operation of aggregating data necessarily is a highly nonlinear process. It is this nonlinearity that creates the difficulties.

4. ROBUSTNESS AND BAYES'S DECISION PROCEDURES

To illustrate how robust these difficulties of inconsistency can be, consider the problem of choosing alternatives from $\{a_1, \ldots, a_m\}$ based on a parameter θ from $\{\theta_i\}_{i=1}^k$. The loss incurred should a_j be chosen when $\theta = \theta_i$ is $w_{i,j}$, so $\mathbf{w}^j = (w_{1,j}, \ldots, w_{k,j})$ is the *loss vector* associated with a_j. If $\mathbf{p} = (p_1, \ldots, p_k)$ is a probability distribution where $p_j = \Pr(\theta = \theta_j)$ $(j = 1, \ldots, k)$, then the expected risk, r_j, associated with choosing a_j is

$$r_j = \langle \mathbf{p}, \mathbf{w}^j \rangle = \sum_{i=1}^k p_i w_{i,j}, \qquad j = 1, \ldots, m. \qquad (7)$$

The selected alternative is the one that minimizes the risk.

For this model, $S = \Psi^k = \{\mathbf{p} = (p_1, \ldots, p_k) \mid p_i \geq 0, \Sigma_{j=1}^k p_i = 1\}$ is a probability space. The decision region in Ψ^k for a single alternative is determined by its boundaries, and this boundary is where the outcome is two or more alternatives. For example, if $\{a_1, a_2\}$ is selected, then $r_1 = r_2 \leq r_j$ for $j \geq 2$. The equality $r_1 = r_2$ defines the set

$$\{\mathbf{p} \in \Psi^k \mid \langle \mathbf{p}, \mathbf{w}^1 \rangle = \langle \mathbf{p}, \mathbf{w}^2 \rangle\}. \qquad (8)$$

This decision region, defined by $\langle \mathbf{p}, \mathbf{w}^1 - \mathbf{w}^2 \rangle = \Sigma_{i=1}^k p_i(w_{i,1} - w_{i,2}) = 0$, is in a linear subspace of Ψ^k. Thus, a single alternative decision region is a convex region in Ψ^k.

If the process is weakly consistent, then for $\mathbf{y}^1, \mathbf{y}^2$, where $\xi(\theta \mid \mathbf{y}^j) \in f^{-1}(\{a_1, a_2\})$, it must be that $\xi(\theta \mid \mathbf{y}^1) \circ \xi(\theta \mid \mathbf{y}^2) = \xi(\theta \mid \mathbf{y}^1 + \mathbf{y}^2)$ is in the same $r_1 = r_2$ linear subspace. For $k = 3$, this requires $\xi(\theta \mid \mathbf{y}^1 + \mathbf{y}^2) = a\xi(\theta \mid \mathbf{y}^1) + b\xi(\theta \mid \mathbf{y}^2)$ for scalars a, b; thus the likelihood function $f(\mathbf{y}^j \mid \theta_i)$ must have a linearity that is not enjoyed by most models. So, as true for many procedures, although the decision regions are convex sets, the binary operation of aggregating data introduces nonlinearities

that force the process to violate weak consistency. This can be shown with a simple example for $k = 2$.

Example. To create a robust example that violates weak consistency, let $\{\theta_1 = .1, \theta_2 = .3, \theta_3 = .5\}$, where θ is the probability of success in a Bernoulli trial; let $Y^j = y_j$ $(j = 1, 2)$ be the number of successes in two independent random samples of n_j items, and let $(p_1, p_2, p_3) = (.5, .3, .2)$ be a prior distribution for θ. When $\theta = \theta_i$, the likelihood function is $f(y_j \mid \theta_i) \propto \theta_i^{y_j}(1 - \theta_i)^{n_j - y_j}$ for $y_j = 0, \ldots, n_j$. Thus the posterior distribution is $\xi(\theta \mid y_j) = (\xi_1(\theta \mid y_j), \xi_2(\theta \mid y_j), \xi_3(\theta \mid y_j))$, where $\xi_i(\theta \mid y_j) \propto p_i f(y_j \mid \theta = \theta_i)$. By choosing $n_1 = n_2 = 20$ and $y_1 = 4$, $y_2 = 9$, we have that $\xi(\theta \mid y_1) = (.5285, .4606, .0109)$ and $\xi(\theta \mid y_2) = (.0005, .3795, .6200)$, while $\xi(\theta \mid y_1 + y_2) = (8.737 \times 10^{-4}, .9445, 5.466 \times 10^{-2})$. After plotting these distributions on Ψ^3, it becomes obvious that there is an enormous degree of flexibility in selecting the $\mathbf{w}_j$'s to create examples—the violation of weak consistency can be quite robust. A simple choice in which the parts have a common decision different from that of the aggregated data is $\mathbf{w}^1 = (1, 1, 1)$, $\mathbf{w}^2 = (2, 0, 2)$, and $\mathbf{w}^3 = (3/2, 1, 3/2)$.

5. CONCLUSION

Weak consistency is a natural requirement to impose on a decision procedure. After all, if the independent "parts" support a common conclusion, then why should not this same conclusion hold for the whole? One might view such inconsistencies as constituting a serious flaw in a statistical decision procedure. It may be, but, as we show, weak consistency is violated by many, if not most, procedures. A critical source of these difficulties is the nonlinearity necessarily introduced by the binary process of combining data.

In an earlier version of this article, we noted that even though weak consistency can be violated by wide classes of decision procedures, there do not seem to be real-data examples other than those illustrating Simpson's paradox. Thanks to subsequent correspondence with N. Keyfitz, A. Rogers, and others, we learned of recent examples. We point to Rogers (1989) and, in particular, to the papers associated with the Population Program at IIASA such as those of Andreev, Lutz, and Scherbov (1989) and Vaupel and Yashin (1985). We also thank W. Kruskal for calling some references to our attention.

REFERENCES

Andreev, E., Lutz, W., and Scherbov, S. (1989), "Averaging Life Expectancy," IIASA Working Paper 89-35, Laxenburg, Austria.

Bickel, P. J., Hammel, E. A., and O'Connell, J. W. (1975), "Sex Bias in Graduate Admissions: Data from Berkeley" *Science*, 187, 398–404. Reprinted in *Statistics and Public Policy*, eds. W. B. Fairley and F. Mosteller, Reading, MA: Addison-Wesley, 1977.

Blyth, C. (1972), "On Simpson's Paradox and the Sure-Thing Principle," *Journal of the American Statistical Association*, 67, 364–366.

Cohen, J. E. (1986), "An Uncertainty Principle in Demography and the Unisex Issue," *The American Statistician*, 40, 32–39.

Good, I. J., and Mittal, Y. (1987), "The Amalgamation and Geometry of Two-by-Two Contingency Tables," *The Annals of Statistics*, 15, 694–711.

Kaigh, W. D. (1989), "A Category Representation Paradox," *The American Statistician*, 43, 92–94.

Rogers, A. (1989), "Requiem for the Net Migrant," Working Paper 89-5, University of Colorado, Population Program.

Saari, D. G. (1987), "On the Source of Some Paradoxes From Social Choice and Probability," *Journal of Economic Theory*, 41, 1–22.

—— (1988), "Symmetry, Voting, and Social Choice," *The Mathematical Intelligencer*, 10, 32–42.

—— (1990), "Consistency of Decision Processes," *Annals of Operation Research*, 23, 103–137.

Vaupel, J. W., and Yashin, A. I. (1985), "Heterogeneity's Ruses: Some Surprising Effects on Selection in Population Dynamics," *The American Statistician*, 39, 176–185.

Wagner, C. H. (1982), "Simpson's Paradox in Real Life," *The American Statistician*, 36, 46–48.