Table 1
American Psychologists Who Have Served at the Office of Naval Research London or Tokyo Branch

Name	Year	Main field of interest	Home institution
Henry Imus	1951–1952	Experimental	Office of Naval Research
Clarence Graham	1952–1953	Experimental	Columbia University
W. D. Neff	1953–1954	Physiological	University of Chicago
Dean Farnsworth	1958–1959	Experimental	Navy Medical Service Corps
Lee Cronbach	1955–1956	Educational	University of Illinois
Frank A. Geldard	1956–1957	Experimental	University of Virginia
Alphonse Chapanis	1960–1961	Engineering	Johns Hopkins University
John Lanzetta	1962–1963	Social	University of Delaware
John Rasmussen	1964–1967	Industrial/ Organizational	Navy Medical Service Corps
John Nagay	1965–1966	Social	Office of Naval Research
H. Wallace Sinaiko	1967–1968	Engineering	Institute for Defense Analyses
Newell Berry	1967–1969	Clinical	Navy Medical Service Corps
Joseph Zeidner	1968–1969	Industrial/ Organizational	Army Research Institute for the Social & Behavioral Sciences
Ivan Mensh	1969–1970	Medical	UCLA
James T. Lester	1972–1974	Social	Bridgewater State College
James W. Miller	1975–1976	Social	National Oceanographic and Atmospheric Administration
Morton Bertin[a]	1975–1980	Engineering	Office of Naval Research
Jack A. Adams	1977–1978	Experimental	University of Illinois
Nicholas Bond[b]	1981–	Industrial/ Organizational	Sacramento State University
Richard Snow	1983–	Educational	Stanford University

[a] Bertin served only in the Tokyo office. [b] Bond is also currently in the Tokyo office.

by his Beefeater-garbed aide; being invited to toast the health of a British society (which I did in verse supplied by the president of the counterpart American society); observing experiments on animal aggression in a Finnish laboratory; learning first-hand of an exciting program in Wales that dealt with the design of facilities for the aged and the disabled; hearing a great deal of debate, throughout the year, on the measurement of mental workload; and listening to a Bulgarian industrial psychologist's theories of color in the workplace and his citations of Marx and Goethe as primary sources.

Did the investment of ONR, and of the individuals who spent a year or so in London or Tokyo, do much for the problems of world psychology? Speaking only for myself, but in ways that I suspect would be endorsed by most of the other liaison scientists, it was a remarkable experience. As foreigners, we certainly learned a great deal about the state of European psychology. In some instances, we provided our hosts with new and useful information about what was going on in the United States. We arranged for exchange visits to American institutions. Given the number of psychologists in Europe and throughout the world, the ONR liaison science program could not be expected to have a major influence. But in its own way the effort worked, as was evidenced by the reception we received and by the organization's reputation that preceded us. Asked to summarize in a few words what the experience meant to me, I have many times replied, "After ONR London it's all downhill."

REFERENCE

Rosenzweig, M. R. (1984). U.S. psychology and world psychology. *American Psychologist, 39,* 877–884.

Relationships Among Models of Salary Bias

Michael H. Birnbaum
University of Illinois, Urbana-Champaign

Studies of sex bias in faculty salaries often yield paradoxical results: On the average, women are paid less than men with the same measured qualifications, while at the same time, women have lower average qualifications than men with the same salaries (Birnbaum, 1979a, 1979b, 1981). These paradoxical sex differences in salary and merit can be explained without postulating any sex bias, because they follow from lack of perfect correlations in a one-mediator model.

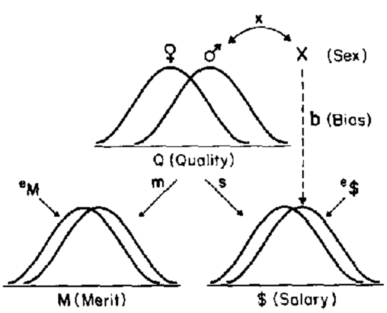
As Birnbaum (1979b, 1981) noted, the one-mediator model (which assumes no bias) can be rejected (perhaps in favor of a theory of bias), when women have higher average qualifications than men with the same salaries. The one-mediator model has received attention—pro and con—as a null hypothesis for studies of salary bias (Birnbaum, 1981, 1982; Humphreys, 1981; McFatter, 1982; McLaughlin, 1980, 1982; Veit, 1981).

Gollob (1984) proposed a difference index for salary equity. The purpose of this comment is to clarify the relationships among various models of salary equity, including Gollob's difference index, from the viewpoint of Birnbaum's (1979b; 1981) mediated models of salary. It will be argued that the difference index, as presented by Gollob (1984), has properties that make it unattractive as a measure of group bias. It will also be shown that there is a special case of Birnbaum's (1981) one-mediator model that is closely related to a special case of Gollob's (1984) difference index. Situations in which the models reach the same conclusion or may disagree will be discussed.

Birnbaum's Mediated Model

Birnbaum's (1979b) model of sex (X), salaries ($), and measured qualifications or merit (M), can be represented as in Figure 1. Suppose that employees differ in true deservingness for salary. Deservingness depends on such factors as seniority, productivity, and quality of work. True quality (deservingness) may not be perfectly correlated with either measured qualifications, salary, or sex. The correlations of true deservingness (Q, the mediator) with sex (X), salary ($), and measured qualifications (M), are denoted by x, s, and m, respectively. Note that the model allows a mean sex difference in salary deservingness, as well as overlap of the distributions. If the sex difference in quality (x) were zero, then women would be equal to men on the measured qualifications. However, most studies have found that women are lower than men on the average (less seniority, fewer publications, etc). With x coded so that men have the higher score, the three observed correlations are typically positive.

Figure 1
Path Analysis of Salary Bias



Note. In this model, there may be a group difference in true quality of performance (Q). If bias (b) equals 0, then people of equal quality have equal average pay, with no group difference. The correlations of the three observed variables with quality are denoted x, m, and s for sex, merit, and salary, respectively.

If the errors in measured qualifications and salary are mutually uncorrelated and uncorrelated with sex, then the correlations among the three variables are as follows:

$$\rho_{M\$} = ms + mxb \quad (1)$$

$$\rho_{MX} = mx \quad (2)$$

$$\rho_{XS} = sx + b. \quad (3)$$

If $b = 0$, then the model assumes no direct impact of sex on salary, apart from differences in salary due to sex differences in true quality of work. In other words, if $b = 0$, the sexes receive equal average pay for equal true quality of work. However, when $b = 0$, it follows that if true quality of work is not perfectly correlated with measured qualifications or salary ($m < 1$ and $s < 1$), then women will receive lower salaries on the average than men with the same measured qualifications, and in the same population women will have lower average qualifications than men with the same salaries (see Birnbaum, 1979b, 1981).

However, if there is a direct effect of sex on salary (suppose $b > 0$), then it is possible that women would have higher qualifications than men with the same salaries. As Birnbaum (1981) has shown, if the correlation between merit and salary is high, the range of possible values of b is small. The next sections show that so-called "direct" and "reverse" regression can be viewed as consistent with special cases of Birnbaum's (1979b, 1981) model. A difference of standard differences model that relates to Gollob's (1984) difference index will be developed, also as a special case of Figure 1.

Special Cases: Regression Models

1. *Forward regression.* Suppose the measure of qualifications is presumed to be perfect (i.e., $m = 1$). It follows that if $b = 0$, then there will be no sex differences in salary with qualifications held constant (Birnbaum, 1979b, 1981).

But no measure of qualifications can be perfectly correlated with true deservingness because true deservingness involves quality of work as well as quantity. In practice, investigators try to measure the scholarship of faculty members, for example, by counting publications in some interval, such as five years. But the number of publications in five years would not correlate perfectly with lifetime publications, let alone with the true scholarly contribution represented by those publications. In practice, no measure or combination of measures can reasonably be supposed to be a *perfect* measure of true scholarship. There are always unmeasured aspects of quality; therefore, it seems plausible that $m < 1$.

2. *Reverse regression.* Suppose salary is perfectly correlated with true quality ($s = 1$). If so, then men and women with the same salaries should be equal in average qualifications (Birnbaum, 1979b, 1981). This approach is termed "reverse" regression by Roberts (1980).

However, salary is unlikely to be perfectly correlated with true deservingness. Salaries are determined by many random factors: For example, faculty salaries are often determined by subjective judgments of a department head or advisory committee. Because human judgments are not perfectly reliable, they cannot be perfectly valid. Salaries are also affected by market forces, by outside offers, by negotiating skill, and by lucky or unlucky correspondence from year to year between legislative appropriations to universities and promotion or merit reviews. Therefore, it is implausible to assume that either salary or measured merit is perfectly correlated with true quality of work.

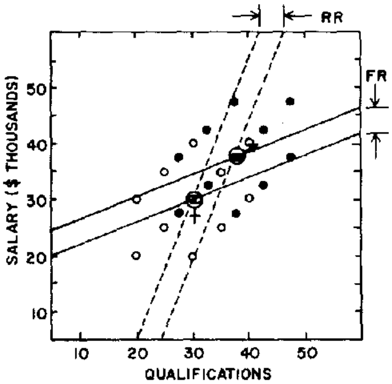
Regression Paradoxes

Figure 2 depicts the usual situation in which the within-group correlation between measured qualifications and salary is less than perfect. In Figure 2, forward regression and "reverse" regression lead to different conclusions. Note that forward regression (FR) indicates bias against women, because women are paid less on the average than equally qualified men. Reverse regression (RR) draws the opposite conclusion, because men have higher qualifications than women with the same salaries. The models of Birnbaum (1979b), Gollob (1984), and McFatter (1982) all could accept this situation as unbiased.

Difference Index

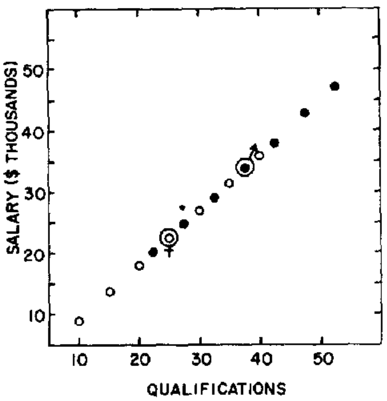
Gollob's (1984) index of bias is the mean difference between the sexes on salary minus the mean sex difference in worth of true qualifications. This

Figure 2
Paradoxical Group Differences



Note. The models of Birnbaum, McFatter, and Gollob can accommodate this pattern without bias. Forward regression (FR) indicates bias against women; reverse regression (RR) indicates bias against men. Salary ($) is plotted against measured qualifications (M), with a solid point for each male and an open circle for each female.

Figure 3
A Problem With the Difference Index



Note. Despite the fact that salary is perfectly predictable from qualifications, with no difference between the sexes, Gollob's difference index implies discrimination against men. The other models would not indicate bias in this case. Furthermore, if everyone were given a 20% raise, Gollob's difference index would reverse the conclusion to bias against women.

index requires calibration of the worth of true quality, which is not required by the other models. Gollob argued that before one can contend that a situation is unfair, one should be able to specify what people should be paid (in dollars). This calibration of worth gives the difference index unfavorable properties as an index of sex bias.

To illustrate the problem, Figure 3 depicts a case similar to one presented by Gollob (1984, p. 449, Case 1). In this case, salary is perfectly predictable from qualifications. Everyone is paid 90% of what they are worth, regardless of sex. The other models of bias, including both regression models, would all describe Figure 3 as unbiased with respect to sex, and this conclusion would hold under any linear transformation of the salaries and qualifications. However, Gollob's difference index indicates sex bias *against men* because they are more underpaid on the average than are women. Now, suppose the employer gave everyone a 20% raise. Unless the worth of qualifications was refigured, the difference index would now switch and indicate bias *against women.* In sum, the difference index is not invariant under multiplication of all of the salaries by a positive constant.

It seems useful to know that people are paid 90% of their true worth and, therefore, that people of higher qualifications are more underpaid. However, it does not seem useful to mix this scaling issue with the issue of sex bias. In this case, the employer could be described as being unfair to people with high qualifications, but should not be described as paying the sexes differently. It is because men have higher qualifications on the average that Gollob's index identifies men as the underpaid group. However, this employer underpays highly qualified women to the same extent as comparable men, and so the bias against the highly qualified should not be attributed to group membership.

Figure 4 illustrates how the scaling of worth of qualifications can override relative comparisons of the groups. Figure 4 depicts a situation that would be described by all versions of Birnbaum's model (including both forward and reverse regression) as biased against women. Women are paid less than men with the same qualifications, and they are more qualified than men with the same salary. The fact that salaries are perfectly predictable from qualifications within each sex, but women are paid exactly $7,500 less than equally qualified men, appears to be evidence of some sort of bias. However, Gollob's difference index would describe this situation as biased *against* men, because men are again more underpaid on the average than women, given the scaling of qualifications. By changing the multiplier of either the salary or qualifications scale, again, these conclusions could be reversed. McFatter's model could also accept this case as unbiased (Birnbaum, 1982; McFatter, 1982).

Difference of Standard Differences

Suppose that salary and measured qualifications are equally correlated with true quality of work in Figure 1 ($s = m$). It follows that:

$$\rho_{MX} = mx \tag{4}$$

$$\rho_{X\$} = mx + b. \tag{5}$$

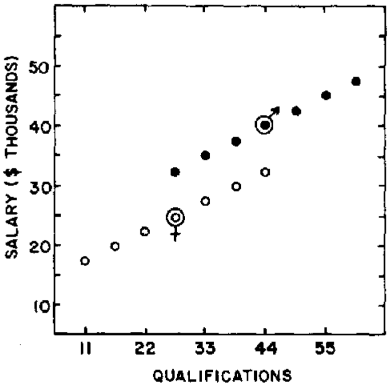
Therefore $b = \rho_{\$X} - \rho_{MX}$. The last expression can be rewritten, using the point-biserial formula for correlations, as follows:

$$b = \sqrt{pq}\left[\frac{(\bar{S}_2 - \bar{S}_1)}{SD_\$} - \frac{(\bar{M}_2 - \bar{M}_1)}{SD_M}\right]. \tag{6}$$

Where p = 1 − q = proportion female; $\bar{S}_2$ and $\bar{S}_1$ are mean salaries for males and females, respectively; $SD_\$$ = standard deviation of salaries; and $\bar{M}_2$, $\bar{M}_1$, and SD_M are the mean merits for men and women and the standard deviation of measured qualifications, respectively.

Although Gollob's (1984) difference index has difficulties produced by the scaling of worth of qualifications and the unit value of money, a special case of Gollob's (1984) difference index agrees (under fixed conditions) with the difference of standard differences model. Equation 6, which follows from the assumption that $m = s$, gives an expression for b that is proportional to a difference of differences in standard units, when the proportions are fixed. This index (Equation 6) would be compatible with Gollob's when the standard deviation of actual salaries equaled the standard deviation of worth of qualifications and at least one group is not underpaid (see Gollob, 1983, equation 7). From the general viewpoint of Gollob's approach, this limitation might seem a disadvantage because it restricts the definition of the worth scale. However, as an index of group

Figure 4
An Illustration of Opposite Conclusions Based on Alternative Models



Note. Gollob's difference index indicates bias against men even though women are paid less than equally qualified men and they are more qualified than equally paid men. In this case, Birnbaum's test and both regression definitions agree that the bias is against women; McFatter's model could accommodate this pattern as unbiased.

bias, it seems preferable to use a relative measure to avoid the difficulties noted in the analysis of Figures 2 and 3.

Testing for Nonlinearity

If salaries are a nonlinear monotonic function of the merit measures, or if this relationship differs for different groups, then the models of bias become more complex. The general nonlinear form of the model (Figure 1) was presented by Birnbaum (1979b). The linearity and parallelism properties of the special cases of the model can be checked by examination of graphs as in Figures 2, 3, and 4. If these properties are not satisfied, it may be necessary to fit the functional relationships with nonlinear or nonadditive models. The discussion that follows assumes that linearity and parallelism have been checked and are satisfied so that the correlation coefficients contain all of the relevant information concerning the variables.

Correlational Analysis

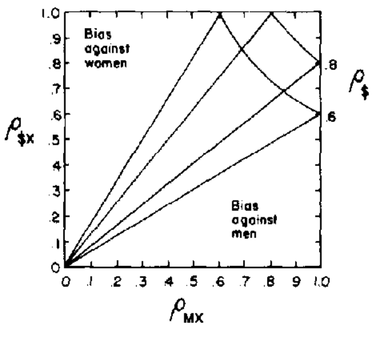
The three correlations can be used to assess the special cases of the model of Figure 1. Figure 5 shows a graphical representation of the requirement that $b = 0$, for the general model. According to Equations 1, 2, and 3, the ratio of the sex–salary correlation to the sex–merit correlation must fall between the salary–merit correlation and the reciprocal of that correlation. Graphically, that means that the point corresponding to $\rho_{\$X}$ and ρ_{MX} must fall between the straight lines for the corresponding value of $\rho_{M\$}$. Note that as $\rho_{M\$}$ increases, the region in which b might equal zero decreases. Thus, the higher the value of $\rho_{M\$}$, the greater the constraint imposed by the model. The additional requirement that $m = s$ would imply that $\rho_{\$X} = \rho_{MX}$.

Comparison of Models

Figure 6 shows regions in which the models agree in their conclusions concerning bias, or may disagree, depending on assumptions concerning parameters of the models. The ellipse in each panel represents the within-group correlation between salary and qualifications for the higher group (usually males).

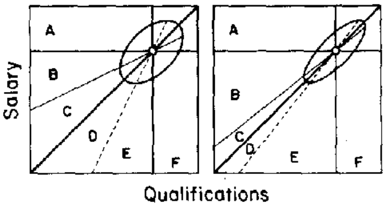
The question is, where can the centroid for the other ellipse be placed so that the situation will be fair? Each model specifies a region of acceptability for the hypothesis that the situation is not biased against either group. Birnbaum's general model requires that the centroid for the lower group fall in regions C or D. If it is also assumed that $m = s$, then both

Figure 5
Region of Acceptability of One-Mediator Model



Note. The region in which b might equal zero is shown as a function of correlations: sex and salary ($\rho_{\$X}$), sex and merit ($\rho_{MX}$), and salary and merit ($\rho_{\$M}$). If $b = 0$, then $\rho_{\$M} \le \frac{\rho_{\$X}}{\rho_{MX}} \le \frac{1}{\rho_{\$M}}$, so the point corresponding to the sex difference in salary and merit should fall inside the region bounded by the correlation between salary and merit.

Figure 6
Comparison of Models of Sex Bias



Note. Ellipses represent correlation between salary and merit for the higher group. If the other group falls in region A or F, all models agree. Birnbaum's general model, with $b = 0$, requires that centroid for the lower group fall in regions C or D. In these regions, b may be positive or negative, depending on assumptions concerning m and s.

centroids should fall on the identity line. If $m = 1$, then the centroid should fall on the "forward" regression line, predicting salary from merit. If $s = 1$, then the centroid should fall on the "reverse" regression line, predicting merit from salary (dashed line). If the centroid falls in regions C or D, the value of b could be positive, negative, or zero, depending on the assumptions concerning m and s. In regions A and B, all versions of Birnbaum's model (Figure 1) imply that $b < 0$. In regions E and F, $b > 0$, respectively. Note that the region in which b might be 0 (C and D) is smaller for larger within-group correlations between merit and salary.

Gollob's difference of differences model requires a scaling of worth of qualifications. However, depending on this scaling of value, the difference index may be zero, positive, or negative, whenever the lower group has its centroid in regions B, C, D, or E. However, in region A, both b and the difference index are less than zero; and in region F, both b and the difference index are greater than zero. McFatter's (1982) model shares the same regions as Gollob's, though the conclusions in that model depend on different parameters. Therefore, these models can disagree with the one-mediator model in regions B and E.

Role of Judgment

Many different definitions of sex bias have been proposed and defended. Choice among definitions is to some degree arbitrary, because definitions are neither true nor false. Philosophers explore definitions by forcing consideration of implications. They might say, "If you define an act as 'moral' whenever it is approved by the majority, then you must accept the consequence that the same act will change from 'moral' to 'immoral' when majority approval changes." Much of the argumentation over the definitions of sex bias in salaries has been based on exploration of the implications of the definitions. For example, the present comment points out that the difference index has the unfortunate property of changing from bias against one group to bias against another under multiplication of all of the salaries by a constant. Thus, fluctuations in the value of the dollar due to inflation or to valuation against gold or other currencies would reverse the direction of sex bias, unless the worth scale were similarly rescaled. As Gollob (1983, 1984) has emphasized, use of the difference index requires that the worth of qualifications be determined anew whenever the amount of money to be distributed changes. Figure 4 shows how the relative standings of groups can be overridden by this scaling factor. Thus, the difference index declares Figure 4 a case of bias against men, even though men are paid more than equally qualified women and equally paid women have more qualifications than corresponding men.

However, a test of usefulness of definitions for communication is one involving judgment of the implications. If a proposed definition of *chair* correctly identifies the things people call *chairs* without misnaming other things, the definition is consistent with normal usage. In the present discussion, we can examine cases such as those shown in Figures 2, 3, and 4 and ask people to judge the direction and extent of sex bias in each case. Birnbaum and Hynan (in press) carried out such an experiment using academics as judges and found that judgments of bias were reasonably consistent with Equation 6, which is the special case of the model of Figure 1 in which $m = s$. Birnbaum and Hynan's (in press) experiment was not designed to test Gollob's (1984) difference index; nevertheless, the reader should judge the direction and extent of bias in Figure 2, 3, 4 and decide which model best represents his or her own judgments.

Concluding Comments

No statistical procedure has yet been proposed that will allow a computer to draw sound inferences from data unaided by human thought. It would not be reasonable to propose that the model of Figure 1 be applied mechanically to draw extreme conclusions regarding sex bias. However, it can be argued that society should examine salary structures for individual and possible group-related inequities, and a good case can be made that the model of Figure 1 provides a useful framework for the analysis of group differences.

It is not possible to resolve paradoxical group differences in both salary and merit (as in Figure 2) without dealing with individual inequities (Birnbaum, 1979b). Furthermore, if the issue of individual inequities is solved, any group inequities will automatically be resolved. Considering these facts, it is surprising that there has been so much discussion of indices for the detection of sex bias compared with the attention given to methods for the correction of individual inequities (Birnbaum, 1979b; 1983).

REFERENCES

Birnbaum, M. H. (1979a). Is there sex bias in salaries of psychologists? *American Psychologist, 34,* 719–720.

Birnbaum, M. H. (1979b). Procedures for the detection and correction of salary inequities. In T. R. Pezzullo & B. E. Brittingham (Eds.), *Salary equity: Detecting sex bias in salaries among college and university professors* (pp. 121–144). Lexington, MA: Lexington Books.

Birnbaum, M. H. (1981). Reply to McLaughlin: Proper path models for theoretical partialing. *American Psychologist, 36,* 1193–1195.

Birnbaum, M. H. (1982). On the one-mediator null hypothesis of salary equity. *American Psychologist, 37,* 1146–1147.

Birnbaum, M. H. (1983). Perceived equity of salary policies. *Journal of Applied Psychology, 68,* 49–59.

Birnbaum, M. H., & Hynan, L. G. (in press). Judgments of salary bias and test bias from statistical evidence. *Organizational Behavior and Human Decision Processes.*

Gollob, H. F. (1983). An index of salary equity for groups. In *Proceedings of the Social Statistics Section of the American Statistical Association: 1983* (470–473). Washington, DC: American Statistical Association.

Gollob, H. F. (1984). Detecting sex bias in salaries. *American Psychologist, 39,* 448–451.

Humphreys, L. G. (1981). Theoretical partialing requires a defensible theory. *American Psychologist, 36,* 1192–1193.

McFatter, R. M. (1982). On detecting sex bias in salaries. *American Psychologist, 37,* 1144–1146.

McLaughlin, S. D. (1980). Atheoretical partialing in survey research. *American Psychologist, 35,* 851.

McLaughlin, S. D. (1982). Reply to Birnbaum. *American Psychologist, 37,* 1143–1144.

Roberts, H. V. (1980). Statistical biases in the measurement of employment discrimination. In E. R. Livernash (Ed.), *Comparable worth: Issues and alternatives* (pp. 173–195). Washington, DC: Equal Employment Advisory Council.

Veit, C. T. (1981). When not to comment. *American Psychologist, 36,* 1195.

Thanks are due Harry F. Gollob for helpful suggestions and enjoyable discussions of these ideas.

Correspondence should be sent to Michael H. Birnbaum, Department of Psychology, University of Illinois, 603 E. Daniel, Champaign, IL 61820.

Carelessness, Computers, and Critical Review

David J. Berndt
University of Chicago

As a frequent reviewer for American Psychological Association (APA) journals, I would like to express my dismay at an emerging phenomenon, before it increases exponentially. I have kept track now of the manuscripts I have reviewed in the years 1983 and 1984: a total of 39 manuscripts from seven different journals. In 1983, one fifth of the manuscripts I reviewed were printed on dot matrix printers and photocopied with various degrees of legibility. In 1984, just short of half of the manuscripts I reviewed were printed on poor-quality dot matrix printers. The most recent edition of the APA *Publication Manual* (APA, 1983) warns against the use of copy from a dot matrix printer and approves of it "only if it is clear and legible" (p. 137).

I myself use a word processor and find it an invaluable timesaver. I am comfortable with computers, and I encourage prospective authors to use them. Unfortunately, the marriage between the computer and dot matrix printers is an unsuccessful, unhappy union in most cases in which the product is for public consumption.