General Article

# "SOME THINGS YOU LEARN AREN'T SO": Cohen's Paradox, Asch's Paradigm, and the Interpretation of Interaction

By Ralph L. Rosnow[1] and Robert Rosenthal[2]

[1]Temple University and [2]Harvard University

*When interpreting an interaction in the analysis of variance (ANOVA), many active researchers (and, in turn, students) often ignore the residuals defining the interaction. Although this problem has been noted previously, it appears that many users of ANOVA remain uncertain about the proper understanding of interaction effects. To clear up this problem, we review the way in which the ANOVA model enables us to take apart a table of group means or the individual measurements contributing to the means to reveal the underlying components. We also show how (using only published data) to compute a contrast on the question that may be of primary interest and illustrate strategies for interpreting tables of residuals. We conclude with an exercise to check on students' understanding of ANOVA and to encourage increased precision in the specification of research results.*

Not long ago, 551 active psychological researchers were surveyed by Zuckerman, Hodgins, Zuckerman, and Rosenthal (1993). The researchers were asked to respond to a set of data-analytic problems, one of which concerned the proper way to understand interaction effects. The desired answer was to examine the leftover effects after freeing the cell means of everything but the interaction. However, Zuckerman et al. reported that about a third of the respondents answered incorrectly, often accompanying their answer by the logic of simple effects or by stating, "It's what I was taught" (p. 51). Such revelations are reminiscent of the quote appearing in the title, which we borrowed from Cohen's (1990, p. 1304) distillation of the wisdom of a lifetime of data analysis. One of his points was that some of what we all learned in graduate school about the analysis of data can be questioned.

Cohen's paradox also helps to explain our earlier finding based on a systematic review of articles published in leading research journals in psychology (Rosnow & Rosenthal, 1989). In a substantial number of cases, there were clear-cut indications of confusion in thinking about interactions, and none of the journals surveyed was immune from this problem. To be sure, the mathematical meaning of interaction effects is unambiguous and is routinely included in textbooks of mathematical and psychological statistics. However, the problem is that many textbooks, when explaining how to *interpret* interactions, ignore the implications of the additive model on which the analysis of variance (ANOVA) is based.

For example, such explanations may take the form of a set of diagrams of crossed and uncrossed combinations of group means in two-way designs. It is pointed out that when the lines cross, this implies that an interaction is present; if the lines do not cross, but remain parallel to one another, this implies that there is no interaction. There is nothing wrong with this idea, although a problem arises when the pattern of the interaction is interpreted solely on the basis of the configuration of group means. Many textbooks go so far as to label the plot of group means the "interaction," in effect creating a non sequitur by ignoring the premise of the additive model. As a result, several generations of teachers and researchers have become unwitting players in an Asch-type drama with real-world consequences. That is to say, when authorities insist that the sum of the parts is synonymous with one of the parts, it is not surprising that impressionable students fail to comprehend a basic distinction.

In this article, we try once more to break this unfortunate chain of conformity. We begin by reviewing the additive model's implicit prescription for decomposing cell means and individual measurements. We also illustrate a handy procedure for computing contrasts on other people's published data to address the question or hy-

Address correspondence to Ralph L. Rosnow, Department of Psychology, Temple University, Philadelphia, PA 19122.

pothesis that may be of primary interest. We then sample some strategies for imposing substantive meaning on interactions. Finally, we present a simple exercise to serve as a check on students' understanding of the additive model and the mathematical meaning of interaction.

## ILLUSTRATION OF THE PROBLEM

Suppose a team of clinical investigators performed a study using the 2 × 2 layout in Table 1, which represents a between-subjects design in which male and female therapists are assigned at random to male or female patients. The dependent variable is the average rating by judges of the effectiveness of the therapy on each patient. The researchers hypothesize an ordering of the means such that $A > B > C = D$. They call this predicted outcome the "patient-by-therapist interaction" and decide that they will test it by examining the appropriate $F$ in a two-way ANOVA. They perform such an analysis and report that, having found $F$ for the 2 × 2 interaction to be significant at $p < .05$, they interpreted this result by inspecting the group means. The reader's attention is directed to a figure (labeled "interaction effect") that plots the four group means and, the researchers confidently assert, shows the pattern of interaction they predicted.

Although there is nothing unusual about such a strategy of data analysis, it is nevertheless based on a misconception about both the mathematical meaning of interaction and the additive model. According to this model, interactions are defined, and therefore completely described, by the leftover effects (or *residuals*) after removing other effects contributing to the group or condition means. In the case of a two-way interaction, we need to remove the row effects (defined for each row as the mean of that row minus the grand mean) and the column effects (defined for each column as the mean of that column minus the grand mean) from the overall effects.[1] It is not absolutely necessary to remove the grand mean from the overall effects, but it may be advantageous to do so. The reason is that the grand mean adds a constant value to the residuals, and freeing them of this constant makes it easier to compare the absolute values of the interaction with the absolute values of the row and column variables. Thus, if adding up the grand mean, row effects for each condition, and column effects for each condition reproduces the obtained group means, we may infer that the interaction is zero. Alternatively, if we are unable to reconstruct the group means from these estimates, we have an interaction to consider.

**Table 1.** *2 × 2 factorial*

| | Therapist sex | |
|---|---|---|
| Patient sex | Female | Male |
| Female | A | B |
| Male | C | D |

Although the researchers slept soundly, secure in the belief of a job well done, they created a cloud of confusion for students as well as research consumers. Where exactly did the researchers go wrong? First, they claimed to have predicted a 2 × 2 interaction when all they really predicted was the rank ordering of the means. They need not have computed a 2 × 2 ANOVA or even invoked the concept of an interaction, but simply computed a contrast on the four group means. Second, they waited to inspect the group means until they felt prompted to do so by the level of significance associated with an $F$ test for interaction in a 2 × 2 ANOVA. But the analysis of group means is not a "Simon says" game in which one must first ask permission of the $p$ value for an interaction $F$ whether it is all right to proceed. Third, they referred to the plot of the group means as "the interaction," when it was actually a plot of the overall effects (i.e., interaction *and* the main effects and grand mean that contributed to the means). The additive model teaches us that plotting the means in a two-way design never plots only the interaction unless the main effects contributing to the group means are exactly zero (e.g., Rosnow & Rosenthal, 1991).

## BREAKING GROUP MEANS INTO COMPONENTS

To flesh out our hypothetical example, imagine that the researchers had accumulated composite bipolar ratings of 20 male and female patients, each of whom was independently treated by either a male or a female therapist. The composite scores for the 5 patients in each condition (cell) were as follows: A—1, 1, 2, 3, 3; B—−1, −1, 0, 1, 1; C—0, 0, −1, −2, −2; D—0, 0, −1, −2, −2. Table 2 summarizes the two-way ANOVA computed by the researchers, in which we see the $F$ test for the interaction component, which they believed was a signal to examine the group means. It is convenient to think of contrasts as "wired-in" (i.e., inherent) in a 2 × 2 factorial; that is, in this case, we would have a top-versus-

1. In higher order designs, we remove not only the row and column effects but also other relevant effects as specified by the additive model. In a three-way factorial, for example, each of the three two-way interactions is defined by the leftover effects (i.e., residuals) after removal of the main effects of the two factors contributing to the particular interaction. The three-way interaction is the set of residuals after removal of the three main effects and the three two-way interactions. In a 2 × 2 × 2 × 2 factorial, the four-way interaction is the set of residuals after subtracting the four main effects, six two-way interactions, and four three-way interactions from the total of all between-conditions effects.

**Table 2.** *Summary of factorial analysis of variance*

| Source | *SS* | *df* | *MS* | *F* | *p* | *r* |
|---|---|---|---|---|---|---|
| Patient sex | 20 | 1 | 20 | 20 | .0004 | .75 |
| Therapist sex | 5 | 1 | 5 | 5 | .04 | .49 |
| Patient × therapist interaction | 5 | 1 | 5 | 5 | .04 | .49 |
| Error term | 16 | 16 | 1 | | | |

*Note*. When working with $t$ tests, the product-moment effect size may be computed by $r = t/\sqrt{t^2 + df}$, where $df$ = degrees of freedom for the two samples compared by $t$. Because $t^2 = F$, it follows that the product-moment effect size of $F$ with numerator $df = 1$ may be computed as $r = \sqrt{F/(F + df\ \text{error})}$, where $df$ error = degrees of freedom for the error term as shown in this table.

bottom-row effect; a left-versus-right-column effect; and a row × column effect. Such contrasts are generally useful, but casting these data into a 2 × 2 factorial would seem to be a poor data-analytic procedure if all that was of interest was the predicted ordering of group means and not the wired-in contrasts.

Table 3 (modeled after Mosteller, Fienberg, & Rourke, 1983) demonstrates, in a concrete way, why interactions are defined by a table of residuals and also why 2 × 2 designs are said to embody three contrasts (i.e., row, column, and interaction). Listed in Part A are the group means, the means of the rows and columns, the grand mean (*G*), and the row and column effects (i.e., differences between the grand mean and the row and column means). Because the grand mean is zero in this example, the corresponding means and effects in the rows and columns are equivalent. Summing the grand mean plus the relevant row and column effects for each cell produces the table of estimates (i.e., estimated overall effects). Had these been exact estimates of the observed group means (i.e., no leftover effects), it would tell us that the interaction component is zero. But this is not the case, and so we turn to the table of residuals (i.e., residual = group mean − estimate) in order to lay open the pattern of the interaction.

**Table 3.** *Decomposition of table of means*

A. Table of means

| | Therapist sex | | | |
|---|---|---|---|---|
| Patient sex | Female | Male | Row mean | Row effect |
| Female | +2 | 0 | +1 | +1 |
| Male | −1 | −1 | −1 | −1 |
| Column mean | +0.5 | −0.5 | 0 = *G* | 0 |
| Column effect | +0.5 | −0.5 | 0 | |

B. Table of estimates

| | |
|---|---|
| +1.5 | +0.5 |
| −0.5 | −1.5 |

C. Table of residuals

| | |
|---|---|
| +0.5 | −0.5 |
| −0.5 | +0.5 |

*Note*. Estimate = grand mean (*G*) + row effect + column effect; residual = group mean − estimate.

What else can we learn from Table 3? First, we are reminded that plotting the group or condition means is a plot not only of the interaction but also of other components contributing to the overall effects. That is, group mean = grand mean + row effect + column effect + residual (i.e., the additive model). Second, we see why row, column, and interaction effects in a 2 × 2 design may be said to constitute three wired-in contrasts. Contrasts are defined by fixed weights (lambdas), with the stipulation that they must sum to zero (i.e., $\Sigma\lambda = 0$). Clearly, the absolute values of the row, column, and interaction effects in Table 3 possess this characteristic. Third, we see that the table of residuals implies an X-shaped plot. Such a plot is characteristic of any nonzero 2 × 2 interaction when (a) there are equal *n*s per cell or (b) the row and column means have not been weighted by the unequal *n*s per cell.[2] Other patterns are possible in more complex designs (e.g., Rosenthal & Rosnow, 1991), but it is always true that the 2 × 2 interaction will be X-shaped.

And finally, we see why examining the residuals uninflated by the grand mean makes it easier to compare them with the absolute values of the row and column effects. We will have more to say about interpreting interactions, but perhaps the most striking feature in Table

2. The usual statistical reason for weighting is that the corresponding *n*s per cell are known to be unequal in the population and we want our sample values to reflect the naturally occurring pattern. However, if we believe our unequal *n*s to be an "accident" of sampling, then it would distort the population estimates if we weighted the row and column means. When in doubt about whether to weight or not to weight, it generally makes sense not to weight—which is to say to treat the cells and marginals *as if* there were equal *n*s per cell.

3 is that each level of analysis tells a different story. It is not that one story invalidates, or necessarily supersedes, another story. It is instead that, as if looking at the world through a prism that disperses light in different directions, we can interpret the data from different perspectives. The situation is reminiscent of Tukey's (1969) observation that a "body of data can—and usually should—be analyzed in more than one way" (p. 83). It is also instructive to note that in absolute values, the column factor and interaction each contributed half as much as the row factor to the group means. This finding is illuminating because it runs counter to the widely accepted notion that once researchers obtain an interaction, they should then regard the main effects as meaningless. These absolute values reveal that it is quite possible to have a meaningful main effect in the presence of interaction.

## BREAKING MEASUREMENTS INTO COMPONENTS

It is also instructive to consider how the additive model provides a template for the decomposition of individual scores into their components. Because each score can be viewed as varying around (or deviating from) its group mean (i.e., score = group mean + deviation), it is convenient to think of these deviations as signifying the degree of accuracy with which individual scores may be predicted from a knowledge of group membership. This relationship explains why deviations from the mean are referred to as "error" (i.e., error = score − group mean), in that the magnitude of the deviations signals how poorly one would do in predicting individual scores from a knowledge of group membership. Given that group mean = grand mean + row effect + column effect + interaction residual, it follows that each measurement can be rewritten as score = grand mean + row effect + column effect + interaction residual + error.

Using this schema as a conceptual stepping-stone, Table 4 lists for individual sampling units the grand mean, row and column effects, and interaction residual as constituted in Table 3. Error, as just noted, is defined as the individual score minus the mean of the group in which it is located. Beneath each column is shown the sum of the squared values, and we see immediately how the *SS* in Table 2 originated.

## TESTING THE PREDICTED PATTERN OF MEANS

We have still not addressed the researchers' prediction of an ordering of group means, but that can be accomplished with just the raw ingredients at hand. In research journals that insist that authors comply strictly with the American Psychological Association's (1983) publication

**Table 4.** *Decomposition of individual scores*

| Group | Subject | Score | = | Grand mean | + | Row effect | + | Column effect | + | Interaction effect | + | Error |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 1 | +1 | = | 0 | + | 1 | + | 0.5 | + | 0.5 | + | (−1) |
| | 2 | +1 | = | 0 | + | 1 | + | 0.5 | + | 0.5 | + | (−1) |
| | 3 | +2 | = | 0 | + | 1 | + | 0.5 | + | 0.5 | + | 0 |
| | 4 | +3 | = | 0 | + | 1 | + | 0.5 | + | 0.5 | + | 1 |
| | 5 | +3 | = | 0 | + | 1 | + | 0.5 | + | 0.5 | + | 1 |
| B | 6 | −1 | = | 0 | + | 1 | + | (−0.5) | + | (−0.5) | + | (−1) |
| | 7 | −1 | = | 0 | + | 1 | + | (−0.5) | + | (−0.5) | + | (−1) |
| | 8 | 0 | = | 0 | + | 1 | + | (−0.5) | + | (−0.5) | + | 0 |
| | 9 | +1 | = | 0 | + | 1 | + | (−0.5) | + | (−0.5) | + | 1 |
| | 10 | +1 | = | 0 | + | 1 | + | (−0.5) | + | (−0.5) | + | 1 |
| C | 11 | 0 | = | 0 | + | (−1) | + | 0.5 | + | (−0.5) | + | 1 |
| | 12 | 0 | = | 0 | + | (−1) | + | 0.5 | + | (−0.5) | + | 1 |
| | 13 | −1 | = | 0 | + | (−1) | + | 0.5 | + | (−0.5) | + | 0 |
| | 14 | −2 | = | 0 | + | (−1) | + | 0.5 | + | (−0.5) | + | (−1) |
| | 15 | −2 | = | 0 | + | (−1) | + | 0.5 | + | (−0.5) | + | (−1) |
| D | 16 | 0 | = | 0 | + | (−1) | + | (−0.5) | + | 0.5 | + | 1 |
| | 17 | 0 | = | 0 | + | (−1) | + | (−0.5) | + | 0.5 | + | 1 |
| | 18 | −1 | = | 0 | + | (−1) | + | (−0.5) | + | 0.5 | + | 0 |
| | 19 | −2 | = | 0 | + | (−1) | + | (−0.5) | + | 0.5 | + | (−1) |
| | 20 | −2 | = | 0 | + | (−1) | + | (−0.5) | + | 0.5 | + | (−1) |
| $\Sigma X^2$ | | 46 | = | 0 | + | 20 | + | 5 | + | 5 | + | 16 |

**Table 5.** *One-way analysis of variance reconstituted from Table 2*

| Source | *SS* | *df* | *MS* | *F* | *p* |
|---|---|---|---|---|---|
| Between conditions | 30 | 3 | 10 | 10 | .0006 |
| Within conditions | 16 | 16 | 1 | | |

manual, we should be given the condition means, the value of the statistical test employed, and the degrees of freedom (*df*) associated with that test. The procedure we now describe simply carves a contrast out of the overall $F$ for the four group means as a one-way design. Table 5 displays the overall ANOVA reconstituted from the results in Table 2, and we can now compute a contrast on the simple effects in three easy steps:

Step 1 is to obtain the maximum possible contrast $F$ (MPC-$F$), which represents the largest possible value of $F$ that can be achieved by a contrast carved out of the between-conditions $SS$ for the numerator of $F$. It could achieve this value only if *all* the variation among the means tested by the overall $F$ were associated with the contrast computed, with nothing left over. To obtain this value, we multiply the overall $F = 10$ by its numerator $df = 3$ and find MPC-$F = 30$.

Step 2 is to compute the aggregate $r^2$, which we find by correlating the obtained group means with the lambda weights we create to represent our prediction. In this case, we might choose $+3$, $+1$, $-2$, $-2$ as weights to represent the prediction that $A > B > C = D$. Correlating these values with the obtained group means ($+2$, $0$, $-1$, $-1$) yields $r = .9623$, and the aggregate $r^2 = .926$.

Step 3 is to multiply the results of Steps 1 and 2 to obtain the contrast $F$, which gives us $F(1, 16) = 30 \times .926 = 27.78$; the $p$ associated with this $F$ is less than .00008. The effect-size $r$, computed as $\sqrt{F/(F + df\,\text{error})}$, is .80.[3] This jumbo-sized effect is decisive, and the associated $p$ suggests that it should not be dismissed as merely a chance event.

## OTHER STATEMENTS ABOUT INTERACTIONS

If we are actually interested in the interaction component, there are different forms of statements that impose substantive meaning on the residuals. For example, it may be possible to clarify or accentuate the pattern of the residuals by cutting away some level of a factor. There are several strategies that produce this kind of Occam's razor simplification of tables of residuals.

One approach, characterized as the *method of meaningful differences* (Rosenthal, 1987), consists of subtracting one level of a factor from the other level of that factor for any two-level factor for which the difference between levels can be seen as conceptually meaningful. In Table 3 (Part C), by subtracting the residual in Cell B ($-0.5$) from that in Cell A ($+0.5$), and also subtracting the residual in Cell D ($+0.5$) from that in Cell C ($-0.5$), we recast the two-way interaction as a change in a main effect due to the introduction of a second independent variable. We thus form a new measure, which may be described as the "advantage" of one level of the column factor over the other level, in this case, the advantage to patients of having female rather than male therapists. The form of our statement about the residuals is that the advantage of having a female therapist is greater for female patients ($+1.0$) than it is for male patients ($-1.0$).

An alternative approach, the *method of meaningful diagonals* (Rosenthal, 1987), focuses on the diagonals of the table of residuals if a suitable concept can be found to describe each diagonal. With the residuals in Table 3, the concept would be same-sex dyad (upper left to lower right diagonal) and opposite-sex dyad (lower left to upper right diagonal). The form of our statement about the interaction is given a slightly different twist by this new measure. That is, dyads consisting of therapists and patients of the same sex are more apt to lead to a favorable therapeutic outcome ($+0.5$) than are opposite-sex dyads ($-0.5$). Once again, we have boiled down a two-dimensional table into a one-dimensional table.

It is, of course, possible to improvise variations on these simplifications for use in more complex tables of residuals. For example, another illustrative strategy uses *the difference between differences,* which produces the linguistically most economical statement about residuals. Although this approach is of more limited applicability (e.g., Rosnow & Suls, 1970), it has the interesting feature that it makes no difference whether we work with the residuals or with the group means as surrogates of the residuals. In both cases, the value of the two-way interaction equals $(A - B) - (C - D)$. In the case of a triple interaction, the most succinct form of statement is simply the difference between the two difference values.

## CHECKING ON STUDENTS' UNDERSTANDING OF ANOVA

As a check on students' understanding of interaction and the additive model, we may ask them to construct a table of predicted values. If they can generate effects in the rows, columns, and interactions of a table of data they have created, we can be sure that they understand

3. This effect-size $r$ is the partial correlation between the scores and their contrast weights controlling for two contrasts orthogonal to the one computed.

the basic distinction between the differences between means and the residuals. To illustrate, suppose we asked them to construct a 3 × 4 matrix of predicted values. They would proceed in four steps:

1. *Select an average value.* They would begin by assigning some mean value for each cell to reflect the underlying metric they have chosen. For example, they might assign the value 5 to each cell, as shown in Part A of Table 6.

2. *Select a row effect.* Assume a quadratic trend in the

**Table 6.** *Constructing a data table*

A. Data showing the same mean for each cell

| | Condition | | | | |
|---|---|---|---|---|---|
| Condition | $b_1$ | $b_2$ | $b_3$ | $b_4$ | Total |
| $a_1$ | 5 | 5 | 5 | 5 | 20 |
| $a_2$ | 5 | 5 | 5 | 5 | 20 |
| $a_3$ | 5 | 5 | 5 | 5 | 20 |
| Total | 15 | 15 | 15 | 15 | 60 |

B. Data of Part A after introducing a quadratic row effect

| | Condition | | | | |
|---|---|---|---|---|---|
| Condition | $b_1$ | $b_2$ | $b_3$ | $b_4$ | Total |
| $a_1$ | 4 | 4 | 4 | 4 | 16 |
| $a_2$ | 7 | 7 | 7 | 7 | 28 |
| $a_3$ | 4 | 4 | 4 | 4 | 16 |
| Total | 15 | 15 | 15 | 15 | 60 |

C. Data of Part B after introducing a column effect such that $b_1 = b_2 > b_4 > b_3$

| | Condition | | | | |
|---|---|---|---|---|---|
| Condition | $b_1$ | $b_2$ | $b_3$ | $b_4$ | Total |
| $a_1$ | 6 | 6 | 1 | 3 | 16 |
| $a_2$ | 9 | 9 | 4 | 6 | 28 |
| $a_3$ | 6 | 6 | 1 | 3 | 16 |
| Total | 21 | 21 | 6 | 12 | 60 |

D. Weights for linear and quadratic elements in an interaction effect

| | Condition | | | | |
|---|---|---|---|---|---|
| Condition | $b_1$ | $b_2$ | $b_3$ | $b_4$ | Total |
| $a_1$ | −1 | +1 | +1 | −1 | 0 |
| $a_2$ | 0 | −2 | 0 | +2 | 0 |
| $a_3$ | +1 | +1 | −1 | −1 | 0 |
| Total | 0 | 0 | 0 | 0 | 0 |

E. Data of Part C after introducing the interaction effects of Part D

| | Condition | | | | |
|---|---|---|---|---|---|
| Condition | $b_1$ | $b_2$ | $b_3$ | $b_4$ | Total |
| $a_1$ | 5 | 7 | 2 | 2 | 16 |
| $a_2$ | 9 | 7 | 4 | 8 | 28 |
| $a_3$ | 7 | 7 | 0 | 2 | 16 |
| Total | 21 | 21 | 6 | 12 | 60 |

row totals is predicted. If the students selected weights of $-1$, $+2$, and $-1$, they would subtract 1 from every entry of $a_1$, add 2 to every entry of $a_2$, and subtract 1 from every entry of $a_3$. This second step is shown in Part B of Table 6; notice that the column totals and grand total are unchanged from Part A.

3. *Select a column effect.* Assume we also predicted column effects such that $b_1$ and $b_2$ are equal to each other and 3 units greater than $b_4$, which, in turn, is 2 units greater than $b_3$. A set of weights satisfying these requirements is $+2$, $+2$, $-3$, and $-1$ for $b_1$, $b_2$, $b_3$, and $b_4$, respectively. Therefore, the students would add 2 to every entry of $b_1$ and $b_2$, subtract 3 from every entry of $b_3$, and subtract 1 from every entry of $b_4$. This step is shown in Part C of Table 6; notice that the row totals and grand total are unchanged from Part B.

4. *Select an interaction effect.* Suppose we also predicted interaction effects such that conditions $b_1$ and $b_3$ show linear trends in the row effects that are in opposite directions to each other, whereas conditions $b_2$ and $b_4$ show quadratic trends in the row effects that are in opposite directions to each other. The weights representing these effects are shown in Part D, and adding them to the effects built up in Part C gives us Part E. Notice that the row, column, and grand totals remain unchanged from Part C.

## ENCOURAGING PRECISION IN SPECIFICATION

We began by making the point that many researchers—and, in turn, students—perennially confuse the overall effects with the residuals when interpreting an obtained interaction in ANOVA. However, this article should not be viewed as an argument for encouraging researchers and students either (a) to focus on the residuals and to ignore the differences between means or (b) to focus on the differences between means and to ignore the residuals. If one claims an interaction, then in almost all cases one is obliged to interpret the residuals and not adopt the traditional mind-set of inspecting only the condition means (i.e., overall effects) generated by a computer. On the other hand, if all one is interested in are the overall effects, one can cut to the chase and analyze them using one or more contrasts (e.g., Rosenthal & Rosnow, 1985). Of course, it is usually prudent to accept the wisdom of Tukey's advice and subject the data to more than one analysis, because this approach is cost-effective and may spawn new insights for further testing. However, whatever strategy one adopts, this article is basically a plea for increased precision in the specification of research results.

**Acknowledgments**—We wish to acknowledge the funding received by the first author from Temple University's Thaddeus L. Bolton Endowment, and that received by the second author from the Spencer Foundation. The content of this article is solely the responsibility of the authors.

## REFERENCES

American Psychological Association. (1983). *Publication manual of the American Psychological Association* (3rd ed.). Washington, DC: Author.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45,* 1304–1312.

Mosteller, F., Fienberg, S.E., & Rourke, R.E.K. (1983). *Beginning statistics with data analysis.* Reading, MA: Addison-Wesley.

Rosenthal, R. (1987). *Judgment studies: Design, analysis, and meta-analysis.* Cambridge, England: Cambridge University Press.

Rosenthal, R., & Rosnow, R.L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance.* Cambridge, England: Cambridge University Press.

Rosenthal, R., & Rosnow, R.L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Rosnow, R.L., & Rosenthal, R. (1989). Definition and interpretation of interaction effects. *Psychological Bulletin, 105,* 143–146.

Rosnow, R.L., & Rosenthal, R. (1991). If you're looking at the cell means, you're not looking at *only* the interaction (unless all main effects are zero). *Psychological Bulletin, 110,* 574–576.

Rosnow, R.L., & Suls, J.M. (1970). Reactive effects of pretesting in attitude research. *Journal of Personality and Social Psychology, 15,* 338–343.

Tukey, J.W. (1969). Analyzing data: Sanctification or detective work? *American Psychologist, 24,* 83–91.

Zuckerman, M., Hodgins, H.S., Zuckerman, A., & Rosenthal, R. (1993). Contemporary issues in the analysis of data: A survey of 551 psychologists. *Psychological Science, 4,* 49–53.

(RECEIVED 2/1/94; ACCEPTED 4/24/94)