# Type IV error in marketing research: The investigation of ANOVA interactions

**Article** *in* Journal of the Academy of Marketing Science · December 1995

DOI: 10.1007/BF02893934



**4 authors**, including:

U. N. Umesh
Washington State University
**33** PUBLICATIONS **417** CITATIONS
SEE PROFILE

Robert A. Peterson
University of Texas at Austin
**130** PUBLICATIONS **9,738** CITATIONS
SEE PROFILE

Rajiv Vaidyanathan
University of Minnesota Duluth
**26** PUBLICATIONS **419** CITATIONS
SEE PROFILE

# Type IV Error in Marketing Research: The Investigation of ANOVA Interactions

**U. N. Umesh**
*University of Washington–Vancouver*

**Robert A. Peterson**
*University of Texas at Austin*

**Michelle McCann-Nelson**
*University of Northern British Columbia*

**Rajiv Vaidyanathan**
*University of Minnesota at Duluth*

*Considerable attention is typically given to Type I and Type II errors when conducting empirical research. This article presents an error, often ignored in marketing and consumer behavior research, termed Type IV error. This error results from the improper investigation of interactions in an analysis of variance A review of research published in* Journal of Marketing Research *and* Journal of Consumer Research *found widespread occurrence of Type IV errors. Illustrative improper interpretations of interactions are discussed and approaches for properly investigating interactions are presented. Situations where interactions need to be tested and interpreted are noted. These situations are contrasted with those where it is more appropriate to examine cell mean differences. The correct use of graphs of cell means is also discussed. Guidelines are presented for avoiding Type IV errors*

In classical hypothesis testing in marketing and consumer behavior research, researchers have traditionally been concerned with two types of errors: Type I error, falsely rejecting $H_0$, and Type II error, not rejecting $H_0$ when it is actually false (cf. Sawyer and Ball 1981). These two errors are well established in the research literature and are addressed in virtually every marketing research text (e.g., Green, Tull, and Albaum 1988; Lehmann 1985; Peterson 1988).

Another error has been labeled Type III error. This error includes "the error committed by giving the right answer to the wrong problem" (Kimball 1957, p. 134), "correctly rejecting the null hypothesis for the wrong reason" (Mosteller 1948, p. 61), "deciding upon a difference in the wrong direction" (Kaiser 1960, p 163), or reaching the wrong conclusion because an inappropriate statistical technique was used (Carrier and Wallace 1989)

The focus of this article is Type IV error Type IV error is defined as the incorrect investigation of an interaction in an analysis of variance. Type IV error occurs when interactions are examined although the hypotheses require cell means to be analyzed or, conversely, cell means are examined when the hypotheses require that interactions be analyzed. As an extension, Type IV error occurs when cell means are graphed and the patterns in these graphs are used to interpret interactions. Incorrect interpretation of interactions can lead to erroneous inferences from the research and, consequently, incorrect decisions.

## INTERACTIONS AND TYPE IV ERRORS

An interaction can be viewed from two different perspectives (Sohn 1991). First, it can be viewed from a

**Journal of the Academy of Marketing Science.**
**Volume 24, No. 1, pages 17-26.**

conceptual perspective in which two or more independent variables (factors) influence each other in a number of different possible ways to yield an outcome For instance, a drug can interact with a medical condition to produce a positive or negative effect in a patient. This perspective is fairly broad and includes many different natural phenomena. Second, it can be viewed from a statistical perspective. Methodologists such as Rosenthal and Rosnow (1991) define interactions as residual effects remaining in an analysis of variance once the main effects of two or more variables have been removed. A slightly different, although related, definition is that of Keppel (1991) "An interaction is present when the main effect of an independent variable is not representative of the simple effects of that variable" (p 197). Similarly, Peterson (1988) stated that "[t]echnically, an interaction exists when the effect of one treatment factor varies across levels of one or more other treatment factors If no interaction exists, the effect of one factor is independent of that of another" (p 524). The definition of Type IV error used in this manuscript is based on the statistical concept of an interaction in an analysis of variance.

Type IV errors have been examined—and disputed—in the behavioral science literature for more than a quarter of a century. Marascuilo and Levin (1970, 1976; Levin and Marascuilo 1972, 1973) and Rosnow and Rosenthal (1989, 1991, Rosenthal and Rosnow 1991) are widely recognized as the principal individuals warning about the pitfalls associated with the primary genre of Type IV errors, that is, incorrectly interpreting a statistically significant interaction through a comparison of cell means. However, their views and positions have been questioned and alternative perspectives put forth.

In criticizing the views of Marascuilo and Levin, Games (1973, 1978) argued that cell means are often more easily interpretable than interaction terms and, therefore, researchers should be more concerned about producing interpretable results than employing a single underlying mathematical model. He called for sometimes changing the model during the course of analyzing a complex data set. This, Games argued, is part of the flexibility desired in good scientists and data analysts. In their reply to Games (1973), Levin and Marascuilo (1973) pointed out that such an approach violates a fundamental tenet of hypothesis testing—the existence of a priori significance levels. Further, they noted that it is quite possible to have a statistically significant interaction and no significant differences between cell means. In such a case, using Games' suggestions, the interpretation of an interaction would be unclear. Similarly, Meyer (1991) and Becker and Coolidge (1991) criticized Rosnow and Rosenthal's argument against the use of cell means to interpret interactions. Instead, they stated that much could be gained by examining cell means when interpreting interactions.

Despite the recognition of Type IV error in the behavioral science literature, its occurrence seems to be widespread For example, Kaufman, Dudley-Marling, and Serlin (1986) found that 79 percent of the articles they reviewed in the special education literature that reported statistically significant interaction effects committed a Type IV error, and Rosnow and Rosenthal (1989) found that 56 percent of the articles they reviewed in psychology demonstrated confusion about the interpretation of interactions. Moreover, Type IV errors are also routinely committed in analysis of variance texts and treatises.

Only one article in the marketing/consumer behavior literature was uncovered that addressed Type IV error, even though an extensive search was undertaken. Although they did not use the term "Type IV error," Ross and Creyer (1993) discussed various approaches to the analysis and interpretation of interactions. Their article, like most of the 30-odd papers addressing the interpretation of interactions that have appeared in the behavioral science literature, examines cell means and interactions The present article goes beyond Ross and Creyer's work in several significant ways Specifically, given the state of knowledge regarding Type IV error, the initial objective of the present article was to document its incidence in marketing and consumer behavior research. A second objective was to provide a series of recommendations on how to avoid Type IV errors.

## THE INCIDENCE OF TYPE IV ERROR

To establish the incidence rate of Type IV error in marketing and consumer behavior research, every article and research note published in the *Journal of Marketing Research* (*JMR*) and the *Journal of Consumer Research* (*JCR*) from 1987 to 1993 was examined These journals were deemed representative of the marketing/consumer behavior literature, the time period was sufficiently recent to be relevant yet sufficiently long for analysis purposes. In all, 610 articles and notes (282 in *JMR*; 328 in *JCR*) were examined. Of these, 149 reported a statistically significant ANOVA interaction (46 in *JMR*; 103 in *JCR*)

The occurrence of Type IV error can be studied by considering two common manifestations relating to the investigation of statistically significant interaction effects·

- Interpreting a significant interaction by comparing cell means
- After finding a significant interaction, interpreting graphed cell means without additional statistical analysis

Of the reviewed articles in *JMR* and *JCR* containing statistically significant interactions, 75 percent committed some form of Type IV error. Table 1 summarizes the results of the review. Although seemingly high, as previously noted, similar levels of Type IV error have also been observed in other disciplines

The most common Type IV error found in *JMR* and *JCR* was the investigation of interactions using cell mean differences. Ninety-two percent of the articles containing a Type IV error committed this manifestation of the error. The other manifestation of Type IV error was the investigation of an interaction using a graph of cell means Twenty-three percent of the articles with Type IV errors exhibited this manifestation of the error. The percentages

**TABLE 1**
**Incidence of Type IV Error in *JMR* and *JCR*, 1987-1993**

| | JMR | JCR | *Total* |
|---|---|---|---|
| Number of articles and notes published | 282 | 328 | 610 |
| Number of articles with statistically significant interaction effect | 46 | 103 | 149 |
| Number of articles with statistically significant interaction effect that have Type IV error | 32 | 80 | 112 |
| Percentage of articles with Type IV error | 70 | 78 | 75 |
| Error manifestation[a] | | | |
| Number of articles using cell mean differences | 31 | 72 | 103 |
| Number of articles using graph of means | 8 | 18 | 26 |

a Error manifestation does not equal the total number of articles with Type IV error because some articles have more than one manifestation of Type IV error

in Table 1 add up to more than 100 percent because some articles committed both manifestations of Type IV error.

## INVESTIGATING INTERACTION EFFECTS

To fully understand and appreciate Type IV error, it is necessary to review the basic main-effect/interaction-effect ANOVA model. The overwhelming majority of interaction effects reported in marketing and consumer behavior studies are two-factor interactions. Consequently, the two-way ANOVA is used as the vehicle to discuss interactions. In those instances where higher order interactions exist, typically researchers merely report their existence without attempting to interpret them. Higher order interactions, such as four-factor interactions, are extremely difficult to interpret, and consequently their potential for misinterpretation is of less interest because they are rarely, if ever, investigated by researchers. Having said this, however, it is important to note that interpreting interactions using cell mean differences and/or examining graphical patterns is just as problematic for higher order interactions as it is for two-factor interactions. Consequently, the discussion of improper investigations of interactions applies to higher order interactions as well.

### Interpreting Interaction Effects Using Cell Means

The most common Type IV error made when investigating interaction effects involves comparison of cell means when interpreting a statistically significant interaction. This results in part from an intuitive but incorrect notion of interaction that does not correspond to the conceptualization of interaction in the ANOVA statistical model (Rosenthal and Rosnow 1991).

Consider an ANOVA design with $y$ as the dependent variable. Let the cell means, $\bar{y}_{ij}$, represent unbiased estimators of the population means, $\mu_{ij}$, where $i$ designates the column number and $j$ the row number. Using common ANOVA notation,

$$\bar{y}_{ij} = \mu + (\alpha_i + \beta_j + \gamma_{ij}) \tag{1}$$

where

$\mu$ = overall mean of all cells
$\alpha_i$ = treatment effect for $i$th level of factor A
$\beta_j$ = treatment effect for $j$th level of factor B
$\gamma_{ij}$ = interaction effect of factors A and B.

Thus interpreting a cell mean in terms of the interaction would ignore the main effects and the grand mean. Rewriting equation (1) with the interaction term on the left side of the equation readily illustrates that it is a residual effect after removing the grand mean and the main effects.

$$\gamma_{ij} = \bar{y}_{ij} - \mu - (\alpha_i + \beta_j).$$

Assuming equal cell sizes, the difference between cell means $\bar{y}_{ij}$ and $\bar{y}_{i'j'}$ can be represented by

$$\begin{aligned}\bar{y}_{ij} - \bar{y}_{i'j'} &= (\mu + \alpha_i + \beta_j + \gamma_{ij}) - (\mu + \alpha_{i'} + \beta_{j'} + \gamma_{i'j'}) \\ &= (\alpha_i - \alpha_{i'}) + (\beta_j - \beta_{j'}) + (\gamma_{ij} - \gamma_{i'j'}).\end{aligned} \tag{2}$$

Thus when the two cell means, $\bar{y}_{ij}$ and $\bar{y}_{i'j'}$, are compared, the comparison is not only between the interaction effects $\gamma_{ij}$ and $\gamma_{i'j'}$. The comparison also incorporates the main effects $\alpha$ and $\beta$. Even if the difference between two cell means is zero, there is no guarantee that the interaction effects are zero, because the other terms in equation (2) may cancel out the interaction terms. Likewise, if the cell means are significantly different, that is $\bar{y}_{ij} - \bar{y}_{i'j'} \neq 0$, there is no way to know if this is due to the differences between $\alpha_i$ and $\alpha_{i'}$, $\beta_j$ and $\beta_{j'}$, or $\gamma_{ij}$ and $\gamma_{i'j'}$. That is, the interaction terms may well be nonsignificant. Equation (2) can be rewritten to show that the difference between the interaction terms in two cells equals not only the difference between the cell means but also a number of main effect terms.

$$\gamma_{ij} - \gamma_{i'j'} = (\bar{y}_{ij} - \bar{y}_{i'j'}) - (\alpha_i - \alpha_{i'}) - (\beta_j - \beta_{j'}).$$

In brief, the magnitude of any difference between cell means does not necessarily reflect the magnitude of interactions or vice versa.

In many applications, researchers compare cell values within a row (or within a column) when examining the effect of one factor at a time. For instance, within column 1, the difference in the cell means $\bar{y}_{1j}$ and $\bar{y}_{1j'}$ (where $j$ and $j'$ are two rows in column 1) can be written as.

$$\begin{aligned}\bar{y}_{1j} - \bar{y}_{1j'} &= (\mu + \alpha_1 + \beta_j + \gamma_{1j}) - (\mu + \alpha_1 + \beta_{j'} + \gamma_{1j'}) \\ &= (\beta_j - \beta_{j'}) + (\gamma_{1j} - \gamma_{1j'}).\end{aligned} \tag{3}$$

Thus, in general, a difference between cell means within a column equals a combination of differences between main effects ($\beta$) and interactions ($\gamma$) Therefore, a difference between the cell means does not reflect the magnitude of interactions or vice versa. The difference between cell means within a column (or row) cannot be used to interpret a finding of significant interaction. Consequently, as before, by itself a significant difference between cell means, within a column or across a row, cannot be interpreted as an indication of a significant interaction.

All of the arguments used in this article regarding the misinterpretation of interactions presume that the interaction is in fact statistically significant. Consequently, ANOVA statistics such as sum of squares, mean squares, and degrees of freedom are not presented here to demonstrate statistical significance testing. (Besides, information on basic ANOVA tests is available in most introductory statistics textbooks.)

In some situations, researchers explain phenomena by examining the difference between each of two pairs of cell means and then calculating the difference between the two differences. For example, a manager might be interested in the difference in the sales of products A and B, and how this difference is different for East Coast and West Coast markets. When the differences between differences of cell means are calculated, the only remaining terms are the interaction terms ($\gamma$). However, a test of these differences is not the same as the omnibus $F$ test for an interaction in the ANOVA model The difference between differences in cell means, such as

$$\begin{aligned}(\bar{y}_{ij} - \bar{y}_{ij'}) - (\bar{y}_{i'j} - \bar{y}_{i'j'}) = {} & [(\mu + \alpha_i + \beta_j + \gamma_{ij}) \\ & - (\mu + \alpha_i + \beta_{j'} + \gamma_{ij'})] \\ & - [(\mu + \alpha_{i'} + \beta_j + \gamma_{i'j}) \\ & - (\mu + \alpha_{i'} + \beta_{j'} + \gamma_{i'j'})] \\ = {} & (\gamma_{ij} - \gamma_{ij'}) - (\gamma_{i'j} - \gamma_{i'j'}),\end{aligned} \tag{4}$$

can be expressed purely with interaction terms. However, because the omnibus $F$ test for an interaction in the ANOVA model tests the null hypothesis $H_0$: $\gamma_{ij} = 0$ for all $ij$, the two tests are not equivalent This is because the former tests whether differences between or among the interaction terms are zero, whereas the latter tests if each of the interaction terms is zero

The consequences of a misinterpretation of an interaction can be illustrated with a hypothetical example Consider a $3 \times 3$ ANOVA where factor A represents price level and factor B represents type of promotion The dependent variable is a subject-rated perception of quality on a –10 to +10 scale. Panel A of Table 2 contains the mean perceived quality for each of the nine cells, together with the respective column and row means.

Because interaction effects are residual effects, in a two-way analysis of variance (A × B) they are the effects remaining after the row and column effects (the main effects of A and B, respectively) have been removed Interaction effects are extracted by subtracting the grand mean and the row and column effects from the observed cell means. (For those who prefer to work with row and column means rather than effects, the interaction terms are obtained by subtracting the row and column means and adding back the grand mean to the observed cell means )

**TABLE 2**
**Hypothetical Example Illustrating the Extraction of Interaction Effects**

| | *Type of Promotion* | | | | |
|---|---|---|---|---|---|
| *Price* | *Coupon (1)* | *FSI (2)* | *Rebate (3)* | *Row Mean* | *Row Main Effect* |
| A Mean perceived quality ratings with row and column effects shown | | | | | |
| High price (a) | 7 | 2 | 0 | 3 | 2 |
| Medium price (b) | 5 | 0 | 1 | 2 | 1 |
| Low price (c) | –6 | 4 | –4 | –2 | –3 |
| Column mean | 2 | 2 | –1 | 1 | |
| Column main effect | 1 | 1 | –2 | | |
| B Interaction effect of Price × Promotion on perceived quality[a] | | | | | |
| High price (a) | 3 | –2 | –1 | | |
| Medium price (b) | 2 | –3 | 1 | | |
| Low price (c) | –5 | 5 | 0 | | |

a Overall mean, row effects, and column effects have been removed

Specifically, the first step in isolating interaction effects is to obtain the row and column effects as shown in panel A of Table 2 These row and column effects are calculated by subtracting the grand mean from the respective row and column means. The next step is to subtract the corresponding row and column main effects from each cell Finally, after subtracting the grand mean from each cell, only the interaction effect remains (panel B of Table 2) It should be noted that negative interaction terms should not be interpreted as literal negative differences, but rather as indicating relatively less of an effect of the treatment in that cell (Becker and Coolidge 1991)

To examine interaction effects, it is necessary to go through this series of steps Most marketing studies do not go through these steps (for an exception see Yadav and Monroe 1993) Testing cell mean differences is not sufficient to draw inferences about interactions; it is obvious from Table 2 that there is no particular relationship between interaction terms and cell means Similarly, differences between cell means do not show any pattern of relationship with the interaction terms For instance, the difference between the first two terms in the first row of panel A equals 5 (i e , 7 – 2 = 5) This number, or similar differences between other pairs of cells, does not correspond to the interaction term. Using equation (3) it can be demonstrated that this number consists of the difference between the two corresponding column effects and the difference between the two interaction terms. Without specifically subtracting out the column effects, it is not possible to relate the cell means or their differences directly to the interaction terms Further, the magnitudes of some of the interaction terms exceed the row and column effects The pattern of numbers shows that examining cell means, or differences between cell means, does not provide any implications about the interaction between two factors

The above example can also be used to show that simply testing differences between differences in cell means is not equivalent to a test of interaction terms as shown in equation (4). Consider again the matrix of cell means in panel A of Table 2 The difference between the first two cells in column 1 (= 7 – 5) and that between the first two cells in column 2 (= 2 – 0) are both equal to 2, and consequently the difference between the differences in those cell means equals zero. The interaction terms in panel B, however, show that there is an interaction between the two factors

Equation (4) is noteworthy because even those textbooks and articles that are sympathetic to the viewpoint that interactions must be correctly interpreted make the mistake of assuming that the test of the interactions is the same as the test of the differences between differences of cell means. One explanation for this misinterpretation is that in the special case of a 2 × 2 analysis of variance, which is typically the case used in text illustrations, the interaction terms are equal in magnitude but opposite in sign. By definition, interaction terms sum to zero along each row and column Consequently, in a 2 × 2 matrix, the two cells in each row (or column) are equal but of opposite sign. All differences within a pair of rows (or a pair of columns) are of equal magnitude but of opposite signs. In this special case, the difference between differences is always nonzero except when all the interaction terms in the four cells are also zero (e g., if one difference equals $a$ and the other difference equals $-a$, their difference equals either $2a$ or $-2a$) In all other cases, such as ANOVAs of size 2 × 3, 3 × 3, 4 × 2, and so forth, the test of the difference between differences of cell means is not the same as the omnibus test for interactions

## Empirical Illustrations

The importance of separating interaction effects from main effects is illustrated by a reanalysis of experimental data reported by Oliver and DeSarbo (1988) in their exploration of the effects of expectation and performance on satisfaction. Oliver and DeSarbo hypothesized a high level of satisfaction in a low expectation, high performance treatment condition because of a "surprise effect" (i.e., surprise due to unexpectedly high performance). Their results indicated a significant Expectation × Performance interaction They interpreted this interaction in terms of cell means and stated that they did not find the hypothesized surprise effect because "[a]lthough the subjects responded favorably to this [low expectation, high performance] situation, responses to the high expectation, high performance condition show higher levels of satisfaction" (Oliver and DeSarbo 1988, p. 504). They interpreted the interaction as not directly supporting their hypothesis because of the large cell mean in the high expectation, high performance condition. However, it is necessary to isolate the interaction effects by decomposing the cell means and extracting the interaction effects before such an inference is warranted This has been done in panels A and B of Table 3 for the Oliver and DeSarbo data.

**TABLE 3**
**Reanalysis of Data From Oliver and DeSarbo (1988)**

| | *Expectation* | | | |
|---|---|---|---|---|
| *Performance* | *Low (1)* | *High (2)* | *Row Mean* | *Row Main Effect* |
| A Mean satisfaction ratings and row and column effects | | | | |
| High (a) | 4 5 | 4 9 | 4 7 | 0 35 |
| Low (b) | 3 7 | 4 3 | 4 0 | –0 35 |
| Column mean | 4 1 | 4 6 | 4 35 | |
| Column main effect | –0 25 | 0 25 | | |
| B Interaction effect of Performance × Expectation on satisfaction[a] | | | | |
| High (a) | 0 05 | –0 05 | | |
| Low (b) | –0 05 | 0 05 | | |

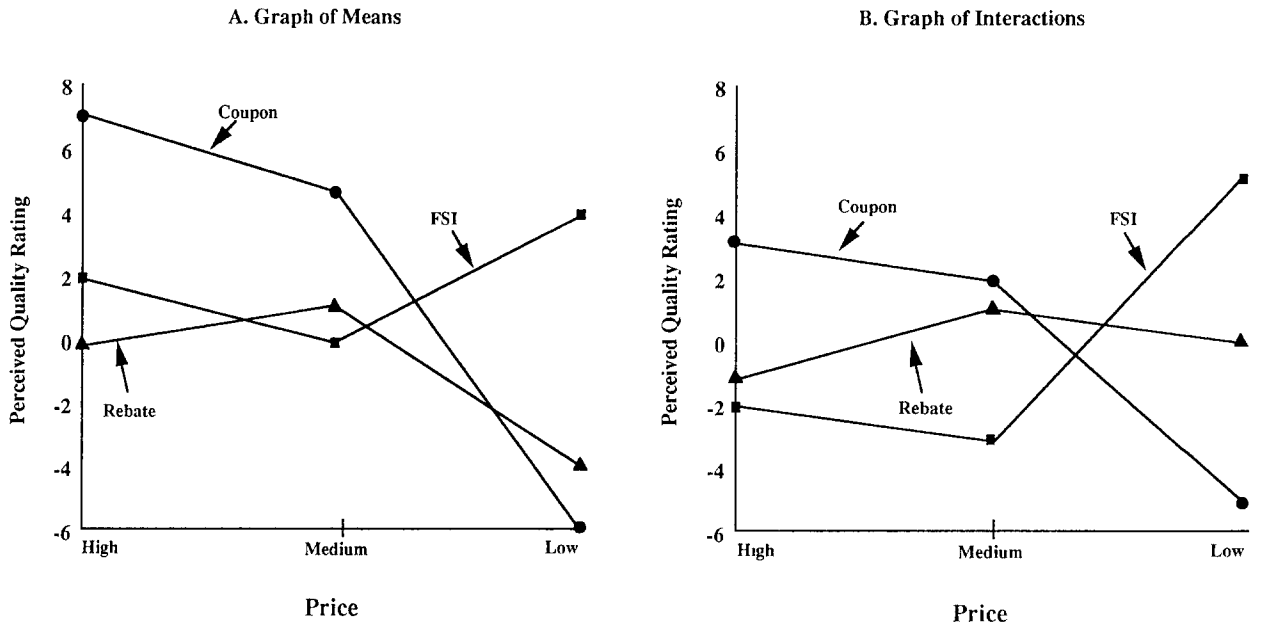
a Overall mean, row effects, and column effects have been removed

From Table 3, it is apparent that, contrary to Oliver and DeSarbo's interpretation, the combination of low expectations and high performance did result in a higher level of satisfaction after removing the individual main effects of expectation and performance. In fact, the negative interaction effect for the high expectation, high performance condition implies further that the large cell mean for that condition (which Oliver and DeSarbo [1988] used to interpret the interaction) was a result of strong positive main effects for expectation and performance and not the joint effect of high expectations and high performance. Therefore, Oliver and DeSarbo failed to identify the source of the large cell mean because it was not properly decomposed into main effects and interaction effects. If the surprise effect is the effect beyond the main effects of performance and expectation, then the combination of low expectation and high performance did produce a positive surprise effect (as shown in panel B of Table 3). This result was significant and consistent with their original hypothesis, but contrary to their interpretation.

In some instances, statistical tests are not conducted to evaluate mean differences. Instead, interpretations of significant interactions are simply based on observed mean differences. For example, in their investigation of the effect of argument strength and frequency of summarizing rhetorical questions on argument recall, Munch and Swasy (1988) found a significant Argument Strength × Summarization interaction and concluded that "[f]or strong arguments, rhetorical and declarative combined, acceptance decreased with increasing frequency (5.73, 5.58, 5.29). For the combined conditions of weak arguments, acceptance increased with increasing frequency (4.59, 4.66, 4.95)" (p. 73). Similarly, Hastak and Olson (1989) interpreted a Message Quality × Processing Goals interaction following a simple visual comparison of means within levels of message quality.

Even though there may be a statistically significant interaction, there is no reason to believe that cell mean differences exist or that they are statistically significant. Although simply observing cell mean differences (without using a statistical test) is a relatively common way of

**FIGURE 1**
**Graphs of Table 2 Data**




interpreting interactions in marketing and consumer research, such a procedure may lead to improper inferences regarding interactions. The point is not that researchers should test interactions for statistical significance; they almost always do Rather, the point is that after finding a statistically significant interaction, it may be misleading to simply visually inspect patterns of means as if the patterns themselves were statistically significant. In the Munch and Swasy (1988) study, the finding of significant interactions should not necessarily imply that a statistically significant trend effect exists in the cell means. Likewise, Hastak and Olson (1989) would have to conduct statistical tests of cell mean differences before reaching conclusions about the means.

## Interpreting Interactions Using Graphed Means

Frequently, graphs of cell means are used to interpret interactions. Although the underlying problem of doing so is similar to the misinterpretation of interactions using cell mean differences in that cell means are used in both situations, it differs in the sense that graphs of the means are used to interpret significant interactions rather than individual cell means.

To graph interaction effects, the effects need to be extracted from the observed cell means and then graphed. The example presented in Table 2 has been graphed in Figure 1. The magnitude of the interaction terms shown in panel B of Figure 1 cannot be easily determined by examining the graph of cell means (panel A of Figure 1) or vice versa For instance, the graph of perceived quality ratings for the high price and medium price conditions consists of parallel lines for coupons and free standing inserts However, the parallelism does not lead to zero interactions in the corresponding cells. Consequently, examining a graph of cell means is not recommended for interpreting interactions The crossed lines in a graph of cell means is not a graph of an interaction effect; rather, it is a graph of all the effects—main and interaction—in the ANOVA model

Wiener, LaForge, and Goolsby (1990) conducted an experiment to examine the effect of self-interest in a personal communication situation They hypothesized that when self-interest is high, an Argument Strength × Expertise interaction is present; that is, communication effectiveness is enhanced when high-expertise communicators use strong arguments and when low-expertise communicators use weak arguments. On the other hand, when self-interest is low, they hypothesized that no Argument Strength × Expertise interaction would be present; that is, communication effectiveness is enhanced when communicators use strong arguments regardless of the level of expertise. They then graphed the cell means and used these graphs to summarize the interactions

Earlier in the manuscript it was shown (in panels A and B of Table 2) how cell means can be decomposed into main effects and interaction effects The graph of interaction effects would be a graph of the effects obtained after the removal of the overall mean, row, and column effects as illustrated in Table 2. The magnitude of the residuals would provide an indication of the effect sizes of the interactions, particularly in comparison to row and column effects. Similar tables could be constructed for the Wiener et al analysis. If the researchers intended to graph the interaction effect, they should have extracted and graphed it

Their figure, labeled "Summary of Interactions," is really a graph of cell means. A graph of cell means should not be referred to as a graph of interaction effects, because these means are composed of both main effects and interaction effects. Despite the misnomer (and the possibility of confusion), the researchers reached the proper inference, at least in direction if not in magnitude, in this instance.

Because an interaction effect refers to the residual effect of two or more factors, it cannot be traced exclusively to a single factor. However, some studies attribute an interaction effect to a particular level of a factor. For example, Madden, Allen, and Twible (1988) investigated the effects of humorous and nonhumorous ad executions and disguised and nondisguised processing sets on attitudes toward ads. They presented a graph that displayed a statistically significant mean difference in the "disguised" level (within the "processing set" factor) and a nonsignificant mean difference in the "nondisguised" level of that factor. After finding a significant Ad Execution × Processing Set interaction, they interpreted the graph to mean that "the plotting of the means reveals that the obvious source of the effects for this variable is based exclusively in the disguised condition" (pp. 247-248). In other words, they concluded that the interaction was due to one level (i.e., disguised) of the processing set factor. This is incorrect By definition, the interaction effect includes the combined effect of at least two factors and cannot be traced exclusively to one factor. In fact, the main effect for processing set (which includes the disguised and nondisguised conditions) is not statistically significant. The correct interpretation of interactions based on the Madden et al. (1988) data is that the combined effect of ad execution and processing set exerted an influence on positive affect (the dependent variable) beyond the separate influence of the main effects. Madden et al.'s misinterpretation arises from only a visual inspection of cell means that have a large difference for the disguised condition and only a small difference for the nondisguised condition.

Although interactions cannot be attributed to a single factor, a cell mean can be decomposed into main effects and interaction effects as demonstrated in equation (1). The interaction term corresponding to each cell, as shown in panel B of Table 2, provides the relative magnitude of the effect of the interaction of the two factors. Further, the absolute value of the interaction terms can be compared to the absolute values of the corresponding main effects to reveal their relative importance. For the (1, 1) cell in panel A of Table 2, the row main effect for high price equals 2 and the column main effect for coupon equals 1, whereas the interaction of high price and coupon equals 3. In other words, the magnitude of the interaction effect is greater than either main effect and equals the sum of the two main effects Such comparisons of magnitude provide guidance when interpreting the relative importance of interactions and main effects in influencing cell mean magnitudes. In a 2 × 2 design, the interaction terms are all equal in (absolute) magnitude. As such, a table of interaction terms in the 2 × 2 case provides limited insight into the influence of factors by indicating the absolute magnitude of the interaction of two factors. In contrast, a table of interactions is particularly useful in 2 × 3 and larger designs. Such a table would not only indicate the absolute magnitude of the interaction effects but would also provide insights into the differential effects of interactions based on the relative magnitude of the interaction term for each cell.

#### Should Cell Means Be Graphed?

The problem of confusing cell mean differences with interactions must not be viewed as a rationale for not graphing cell means. In fact, it is recommended that cell means routinely be graphed. Doing so facilitates inferences about the effect of the factors. It is just that cell mean differences should not be confused with interactions.

Graphs of cell means are useful because they provide an intuitive understanding of the magnitude of the dependent variable The visual distance between two plotted means provides a clear and memorable picture about differences between the cells. This is particularly important when there are numerous cells in a multifactor design (as compared to a simple 2 × 2 design). When there is an $n \times n$ design, frequently it is not immediately apparent which cell mean differences are the largest, smallest, or so forth, without conducting a pairwise comparison of all cell means. For instance, with a 5 × 5 design, the process of evaluating 300 pairwise comparisons ($C_2^{25} = 300$) would be mentally taxing (see Monlezun [1979] and Stanley [1969] for suggested approaches for visually inspecting interaction graphs). Even with a 2 × 2 design, a graphical plot might provide a clearer perspective on the respective cell means than a simple numerical presentation of the means. Moreover, a graphical plot also provides a visual representation of how the different treatments affect the dependent variable. Such a graph, however, is not the same as a graph of the interaction effect as demonstrated in Figure 1

The parallelism or nonparallelism of lines in graphical plots of cell means should not be used to interpret interactions. Furthermore, as noted with regard to Figure 1, the magnitude of nonparallelism does not necessarily reflect the magnitude of the interaction Nonparallel lines, although suggestive of potential interaction, should not be used to interpret an interaction because, as shown by the equations in this manuscript, cell means reflect both interaction and main effects. To investigate and interpret interactions, it is necessary to calculate the interaction terms as noted in equations (1-4).

## When to Test Cell Means Rather Than Interactions

Examination of cell means is necessary in many marketing studies. At a fundamental level, cell means indicate the combined effects of two or more factors on the dependent variable. Frequently, a hypothesis simply states that one particular cell mean is greater than (or equal to) another cell mean In such a situation, a test of whether the cell mean difference is statistically significant is sufficient to

test the hypothesis. An interaction test is not meaningful here and should not be performed (Schoorman, Bobko, and Rentsch 1991)

A potential problem with testing an interaction, when only differences among cell means are of interest, is the possibility that a significant cell mean difference might exist even though a nonsignificant interaction effect is observed The $F$ ratio used in an ANOVA to test whether an interaction term is statistically significant consists of a ratio of two mean squares (i.e., $MS_{A\times B}/MS_E$) However, this test assumes that the sum of squares associated with $MS_{A\times B}$ is more or less equally distributed among all cells contributing to the A × B interaction (Marascuilo and Levin 1970) If an interaction term possesses many degrees of freedom, it might be found nonsignificant even if most of the sums of squares can be attributed to one outlier cell In other words, if one cell mean is significantly different from all other cell means, but all other cell means are approximately equal, then an overall test of the interaction may indicate a nonsignificant interaction even though significant cell mean differences exist. Consider the following hypothetical 2 × 2 ANOVA design with three observations in each cell.

| | *Factor A* | |
|---|---|---|
| *Factor B* | *A1* | *A2* |
| B1 | 140 | 80 |
| | 150 | 100 |
| | 160 | 120 |
| B2 | 73 | 45 |
| | 75 | 50 |
| | 77 | 55 |

In this example, multiple pairwise comparisons (Keppel 1991) to control the Type I experiment-wise error rate at .05 reveal that the mean for cell A1B1 is significantly different from all other cell means. Also, cell mean A2B1 is significantly different from cell mean A2B2. Testing the standard ANOVA hypotheses with the hypothetical data, there is a significant main effect for A ($p < .0005$) and a significant main effect for B ($p < .0001$). However, the A × B interaction is not statistically significant ($p = 0965$) The example illustrates the folly of following a policy of examining differences between cell means only if the interactions are significant. If only cell mean differences are of interest but the researcher fails to explore them because of a nonsignificant interaction effect, the effects of interest may be overlooked. If a researcher is only interested in cell mean differences, then the significance or nonsignificance of the interaction is irrelevant Tests of interactions should not be used as a device for detecting significant differences between cell means. Insignificant interactions can sometimes occur despite the existence of significant cell mean differences.

## CONCLUSIONS AND RECOMMENDATIONS

Based on the definition of Type IV error as the incorrect investigation of an interaction in the analysis of variance, *JMR* and *JCR* were examined for instances of this error, 75 percent of the articles reporting a statistically significant interaction effect in an ANOVA contained a Type IV error The most common manifestation of Type IV error was misinterpreting an interaction by using cell means and cell mean differences The other common manifestation of Type IV error was inappropriately examining the graphical patterns of cell means to interpret a statistically significant interaction. Given the equations that underlie interactions, such approaches involve comparisons of not just interaction effects, but (confounded) main and interaction effects

The principal focus of this article has been on the investigation of interactions in ANOVA designs. Investigation involves not only using the appropriate statistical tests, but also making the proper interpretations and inferences. Four recommendations for correctly investigating interactions and thus avoiding Type IV errors are presented below

Perhaps the most important and intuitively obvious recommendation is that researchers first decide on the precise hypothesis to be tested prior to any analysis The analysis should then be tailored to appropriately test this hypothesis For instance, researchers must decide in advance whether the hypothesis requires a comparison of cell means or an analysis of interaction effects. If the test used does not correspond to the hypothesis, researchers run the risk of reaching erroneous conclusions regarding the hypothesis. Researchers must also be aware of the importance of interpreting the results of the test in a manner consistent with the hypothesis and the analysis

Second, when researchers are interested in testing a specific interaction, the hypothesis should be set up to do so. An example of a specific interaction null hypothesis might be that the interaction term in cell (1,1) is not significantly different from zero. That is,

$$H_0\colon \gamma_{11} = 0.$$

The subsequent analysis and test should correspond to this hypothesis. Researchers often use an omnibus $F$ test to test for the existence of significant interactions. However, an omnibus $F$ test is not suitable for testing this hypothesis because it would test the null hypothesis that the interaction terms associated with all treatment cells are simultaneously zero (Keppel 1991). For a given two-way ANOVA design, the omnibus $F$ test would test the null hypothesis that all interaction terms are not significantly different from zero, or

$$H_0 \quad \gamma_{ij} = 0 \qquad \text{for all } ij.$$

According to Keppel (1991), researchers are typically not interested in the omnibus $F$ test because it is appropriate only in the absence of specific hypotheses. Cohen (1990) made a similar point; he recommended the use of "a few highly targeted variables and hypotheses" (p. 1305). Schoorman et al. (1991) provided empirical support for this viewpoint They found that testing a spe-

cific hypothesis using planned contrasts resulted in more powerful tests than the omnibus ANOVA interaction test

Another relevant point here is that the lack of significance of an interaction does not imply that there are no statistically significant differences between any of the cell means Failing to reject a null hypothesis of no interaction does not obviate the need to conduct tests of complex hypotheses based on theory.

When all main effects are zero, an interesting issue of interaction effect definition comes to light. The term "residual effects" that is used to describe interactions by some researchers is really a misnomer. In a study with zero main effects, the interaction effects are the only relevant effects of interest. In other studies, the interaction effects could well be more important than the main effects. The primary reason interaction effects are sometimes connoted as being residual is that they are calculated from cell means after main effects have been subtracted out. Researchers are cautioned not to consider interaction effects as being secondary or "leftover" effects.

Third, there is nothing wrong with comparing cell means if doing so is appropriate for the hypothesis being tested. In fact, as previously noted, tests of interactions are inappropriate when the hypotheses call for tests of cell mean differences. Toothaker (1991, p. 120) provided an example where the residual effects (i.e., interactions) remaining after the main effects had been removed were not really meaningful. In many situations, researchers are interested in all of the effects of the factors, main and interaction, and cell mean comparisons are best suited to study them. It is necessary to reiterate, however, that the difference between two cell means is only partially composed of interaction effects. In many cases, a cell mean difference may result more from a main effect than from an interaction. Keeping this in mind, differences between means can be examined to provide implications about the cell means when the hypotheses demand doing so.

Finally, finding a significant interaction in an ANOVA model and failing to investigate it further may also present a problem. By leaving a significant interaction uninterpreted, useful information may be lost because of the failure to investigate the interaction.

In brief, this article has pointed out the existence of Type IV error and its importance. A major implication of this article is that interaction effects should only be tested where the hypotheses clearly require doing so At the same time, differences in cell means should be tested only where hypotheses require their evaluation. Further, the interaction effect in the omnibus $F$ test need not be statistically significant to investigate hypothesized specific effects that may be of interest. A danger in searching for statistically significant interaction effects when other specific effects are of interest is that the nonsignificance of the interaction effect in an omnibus $F$ test may dissuade the researcher from examining the specific effects that are of real interest

Examples from the marketing/consumer behavior literature were provided to illustrate the widespread prevalence of Type IV error. The intent in doing so was not to be critical of any particular article or researcher but to illustrate the importance of following the identification of a statistically significant interaction with appropriate analysis and interpretation As such, the intent was to highlight Type IV error and suggest solutions for avoiding the error in the future

## ACKNOWLEDGMENTS

This research was supported in part by the IC$^2$ Institute, the University of Texas at Austin The opinions expressed in the article are those of the authors and do not necessarily reflect the views of the Institute

## REFERENCES

Becker, Lee A and Frederick L Coolidge 1991 "On the Proper Interpretation of Residualized Interaction Means in Analysis of Variance A Reply to Rosnow and Rosenthal" *Psychological Reports* 68 (April) 483-490

Carrier, Harold D and William A Wallace 1989 "An Epistemological View of Decision-Aid Technology With Emphasis on Expert Systems" *IEEE Transactions on Systems, Man, and Cybernetics* 19 (September/October) 1021-1029

Cohen, Jacob 1990 "Things I Have Learned (So Far)" *American Psychologist* 45 (December) 1304-1312

Games, Paul A 1973 "Type IV Errors Revisited" *Psychological Bulletin* 80 (October) 304-307

——— 1978 "Nesting, Crossing, Type IV Errors, and the Role of Statistical Models" *American Educational Research Journal* 15 (March) 253-258

Green, Paul E , Donald S Tull, and Gerald Albaum 1988 *Research for Marketing Decisions* 5th ed Englewood Cliffs, NJ Prentice Hall

Hastak, Manoj and Jerry C Olson 1989 "Assessing the Role of Brand-Related Cognitive Responses as Mediators of Communication Effects on Cognitive Structure" *Journal of Consumer Research* 15 (March) 444-456

Kaiser, Henry F 1960 "Directional Statistical Decisions" *Psychological Review* 67 (May) 160-167

Kaufman, Nancy J , Curtis Dudley-Marling, and Ronald L Serlin 1986 "An Examination of Statistical Interactions in the Special Education Literature" *Journal of Special Education* 20 (Spring) 31-42

Keppel, Geoffrey 1991 *Design and Analysis A Researcher's Handbook* 3rd ed Englewood Cliffs, NJ Prentice Hall

Kimball, A W 1957 "Errors of the Third Kind in Statistical Consulting" *Journal of the American Statistical Association* 52 (June) 133-142

Lehmann, Donald R 1985 *Market Research and Analysis* 2nd ed Homewood, IL Irwin

Levin, Joel R and Leonard A Marascuilo 1972 "Type IV Errors and Interactions" *Psychological Bulletin* 78 (November) 368-374

——— and ——— 1973 "Type IV Error and Games" *Psychological Bulletin* 80 (October) 308-309

Madden, Thomas J , Chris T Allen, and Jacquelyn L Twible 1988 "Attitude Toward the Ad An Assessment of Diverse Measurement Indices Under Different Processing 'Sets'" *Journal of Marketing Research* 25 (August) 242-252

Marascuilo, Leonard A and Joel R Levin 1970 "Appropriate Post Hoc Comparisons for Interaction and Nested Hypotheses in Analysis of Variance Designs The Elimination of Type IV Errors" *American Educational Research Journal* 7 (May) 397-421

——— and ——— 1976 "The Simultaneous Investigation of Interaction and Nested Hypotheses in Two-Factor Analysis of Variance Designs" *American Educational Research Journal* 13 (January) 61-65

Meyer, Donald L 1991 "Misinterpretation of Interaction Effects A Reply to Rosnow and Rosenthal" *Psychological Bulletin* 110 (November) 571-573

Monlezun, Charles J 1979 "Two-Dimensional Plots for Interpreting Interactions in the Three-Factor Analysis of Variance Model" *The American Statistician* 33 (May) 63-69

Mosteller, Frederick 1948 "A K-Sample Slippage Test for an Extreme Population" *Annals of Mathematical Statistics* 19 (March) 58-65

Munch, James M and John L Swasy 1988 "Rhetorical Question, Summarization Frequency, and Argument Strength Effects on Recall" *Journal of Consumer Research* 15 (June) 69-76

Oliver, Richard L and Wayne S DeSarbo 1988 "Response Determinants in Satisfaction Judgments" *Journal of Consumer Research* 14 (March) 495-507

Peterson, Robert A 1988 *Marketing Research* 2nd ed Plano, TX Business Publications

Rosenthal, Robert and Ralph L Rosnow 1991 *Essentials of Behavioral Research Methods and Data Analysis* New York McGraw-Hill

Rosnow, Ralph L and Robert Rosenthal 1989 "Definition and Interpretation of Interaction Effects" *Psychological Bulletin* 105 (January) 143-146

——— and ——— 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 (November) 574-576

Ross, William T, Jr and Elizabeth H Creyer 1993 "Interpreting Interactions Raw Means or Residual Means?" *Journal of Consumer Research* 20 (September) 330-338

Sawyer, Alan G and A Dwayne Ball 1981 "Statistical Power and Effect Size in Marketing Research" *Journal of Marketing Research* 18 (August) 275-290

Schoorman, F David, Philip Bobko, and Joan Rentsch 1991 "The Role of Theory in Testing Hypothesized Interactions An Example From the Research on Escalation of Commitment" *Journal of Applied Social Psychology* 21 (August) 1338-1355

Sohn, David 1991 "The Interaction Concept in Scientific Discourse and in the Analysis of Variance" *Journal of Psychology* 125 (November) 621-629

Stanley, Julian C 1969 "Plotting ANOVA Interactions for Ease of Visual Interpretation" *Educational and Psychological Measurement* 29 (Winter) 793-797

Toothaker, Larry E 1991 *Multiple Comparisons for Researchers* Newbury Park, CA Sage

Wiener, Joshua L, Raymond W LaForge, and Jerry R Goolsby 1990 "Personal Communications in Marketing An Examination of Self-Interest Contingency Relationships" *Journal of Marketing Research* 27 (May) 227-231

Yadav, Manjit S and Kent B Monroe 1993 "How Buyers Perceive Savings in a Bundle Price An Examination of a Bundle's Transaction Value" *Journal of Marketing Research* 30 (August) 350-358

## ABOUT THE AUTHORS

**U. N. Umesh** is an associate professor of marketing at Washington State University. He has previously been a visiting associate professor at the Wharton School and visiting research fellow at the $IC^2$ Institute, University of Texas at Austin, where he has also taught at the marketing department. He has published in the *Journal of Marketing Research*, *Journal of Marketing*, *Psychometrika*, *Marketing Letters*, *Sociological Methods & Research*, *Decision Sciences*, and other journals He has won a number of teaching awards

**Robert A. Peterson** holds the John T Stuart III Centennial Chair in Business Administration and the Charles E Hurwitz Fellowship at the University of Texas at Austin. A former editor of the *Journal of the Academy of Marketing Science* and the *Journal of Marketing Research*, he is currently working on a book on the direct selling industry

**Michelle McCann-Nelson** is an assistant professor at the University of Northern British Columbia She has published in the *Journal of Current Issues and Research in Advertising*, the *Journal of Professional Services Marketing*, and in many national and international conference proceedings

**Rajiv Vaidyanathan** is an assistant professor of marketing at the University of Minnesota at Duluth He received his Ph D from Washington State University His research interests include the examination of how consumers perceive prices and evaluate advertised deals. He has published in the proceedings of several national and international conferences