

Technical Commentary

UNDERSTANDING DATA WHEN INTERACTIONS ARE PRESENT OR HYPOTHESIZED

Richard E. Petty,¹ Leandre R. Fabrigar,² Duane T. Wegener,³ and Joseph R. Priester⁴

¹Ohio State University, ²Queen's University, ³Yale University, and ⁴University of Michigan

Abstract—*The traditional approach to interpreting data when an interaction is present is to interpret cell means as reflecting a difference between differences. An alternative is to interpret any main effects and separately interpret interaction residuals. Interpretation of interaction residuals can sometimes lead to nonsensical conclusions. Thus, in deciding between the approaches, researchers should consider (a) the conceptual nature of the variables involved, (b) relevant theories, and (c) the additional data that might be necessary to distinguish among competing plausible representations. Issues have also arisen regarding how to best test a hypothesis that involves interaction variance. Some researchers have argued that the use of a focused contrast designed to test a specific ordering of cell means (e.g., a linear contrast) is often the best strategy. We note potential problems with the use of such contrasts and discourage researchers from blanket use of contrasts that combine main effect and interaction variance.*

Statistical interactions often play a crucial role in testing psychological theory. Yet recent articles have raised important issues regarding how researchers should interpret their data when a statistical interaction is present, and how they should best test an interaction hypothesis. In this article, we discuss conceptual issues regarding understanding data from factorial designs (especially the common 2×2) when an interaction is present or hypothesized. We focus on the 2×2 in part because of the ubiquity of this design, and because appropriate procedures for testing and interpreting the data when interactions are obtained in this design have been the subject of considerable recent attention and controversy (e.g., see Bobko, 1986, Meyer, 1991, Rosnow & Rosenthal, 1989a, 1989b, 1991, 1995, Ross & Creyer, 1993, Zuckerman, Hodgins, Zuckerman, & Rosenthal, 1993).

UNDERSTANDING THE DATA WHEN AN INTERACTION IS PRESENT

In order to make psychological sense of data, one must have some agreed-upon method of representing the pattern of results obtained. This is particularly important when an interaction is present because there are multiple ways to represent or depict

interactions (cf. Judd, McClelland, & Culhane, 1995). For example, Rosnow and Rosenthal (1989a, 1989b, 1991, 1995) have suggested that understanding the data is fostered by representing and interpreting the interaction as the residual remaining after main effects have been removed. We argue that a blanket use of this approach rather than the more traditional approach of representing and interpreting an interaction as a difference between differences in cell means (e.g., Keppel, 1991) can lead to misleading conclusions about the psychological processes underlying the results. It is important to note that representation of interactions is not a statistical issue per se. The two ways of representing interactions are mathematically equivalent (i.e., interaction variance is the same, whether calculated through residuals or differences between differences, see Guilford & Fruchter, 1978, Keppel, 1991, Rosnow & Rosenthal, 1989b). However, the different representations can suggest different (and conflicting) substantive conclusions.

Traditional Versus Residual Approach to Interactions

Following the finding of a significant interaction in an analysis of variance (ANOVA), researchers have traditionally described the interaction effect through comparisons of original cell means (e.g., through simple effects tests, Keppel, 1991). Take, for instance, a hypothetical example presented by Rosenthal and Rosnow (1985). In a 2×2 design, one of two versions of a political advertisement was shown to either liberals or conservatives. A traditional ANOVA indicates a significant main effect of advertisement, no main effect of political ideology, and a significant interaction between the two independent variables (see Fig. 1a). According to the difference-between-differences approach (e.g., see Keppel, 1991), one acceptable interpretation of these data is that there is a greater effect of the differing ad types for liberals than for conservatives. Keppel (1991) recommended analyses of the two simple effects (of ad type, in this example) as one method for examining the meaning of the interaction (pp. 236–245, see also Myers, 1979, Pedhazur, 1982). This approach to interpretation resonates with the views espoused by Fisher (1947, pp. 92–93) and other researchers (e.g., Hoaglin, Mosteller, & Tukey, 1991).

An alternative view has been espoused by proponents of the residual approach. Rosenthal, Rosnow, and their colleagues (e.g., Rosenthal & Rosnow, 1985, 1991, Zuckerman et al., 1993) have noted correctly that cell means reflect not only the influence of the interaction term in an ANOVA, but also the influence of all lower order effects. Therefore, according to Rosenthal and Rosnow, if one is interested in understanding the psychological meaning of an interaction in a 2×2 design, one

Address correspondence to Richard E. Petty, Department of Psychology, Ohio State University, 1885 Neil Avenue Mall, Columbus, OH 43210, e-mail: petty.1@osu.edu

Interactions

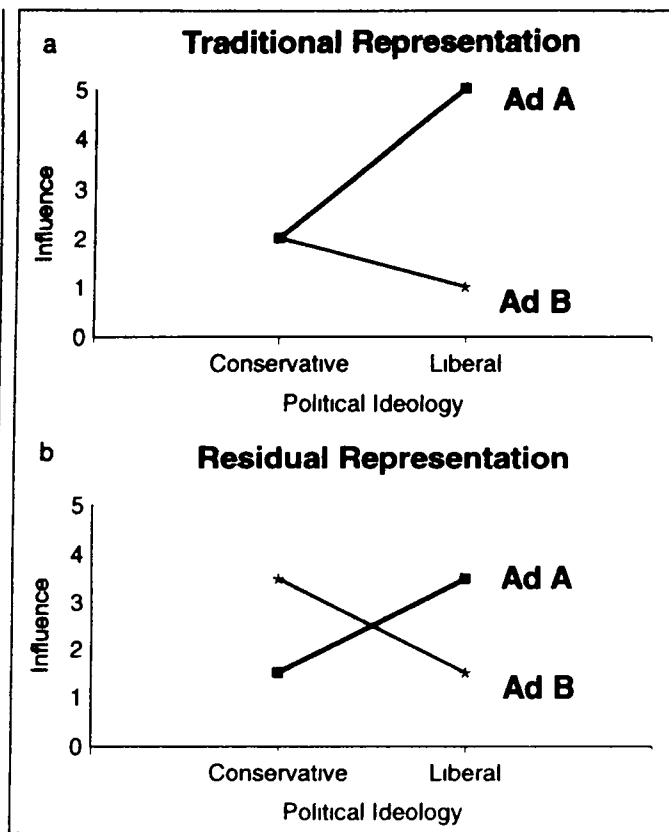


Fig 1 Traditional representation of the difference between differences (a) and residual representation of the interaction (b) in a 2 (political ideology) \times 2 (ad version) factorial

should not do so by examining a representation that also reflects main effects. Rather, one must remove the influence of the main effects from the cell means prior to imparting meaning to the interaction.

Consider the Ad Version \times Political Ideology data in Figure 1. According to Rosenthal and Rosnow (1985), a researcher following the traditional approach might be tempted to interpret the presence of an interaction in the data as indicating that liberals are more strongly influenced by version A than version B, whereas conservatives are equally influenced by both advertisements. Rosenthal and Rosnow (1985) asserted, however, that such a conclusion "would be wrong" (p. 5). Instead, they prefer to interpret the representation of the interaction as depicted in Figure 1b. Consistent with this representation, Rosenthal and Rosnow stated that "the interaction actually shows that conservatives and liberals reacted in exactly opposite ways to the two types of propaganda" (p. 8, see Rosenthal & Rosnow, 1985, for calculation of residuals).

Obviously, this residual representation suggests something very different about the behavior of conservatives than does the representation following from the traditional approach. That is, a difference-between-differences approach encourages one to develop theory that explains why liberals are affected differently by the two ads but conservatives are not, whereas the residual approach encourages one to explain why liberals and

conservatives are influenced in opposite ways by the two ads.¹ The nature of the variables in this example does not preclude generating a meaningful psychological interpretation of the interaction residuals. Therefore, the representations encouraged by the traditional and the residual views are both potentially accurate.

Comparing the Approaches

Rosenthal and Rosnow (1991) asserted that "before an interaction effect can be understood, the residuals defining the interaction must be displayed" (p. 367), and Zuckerman et al. (1993) stated that "interaction effects cannot be interpreted on the basis of comparisons between cell means" (p. 53). If these scholars meant to suggest only that researchers who describe cell means are not describing solely interaction variance, this is correct. However, if they meant that researchers should necessarily impart psychological meaning to interaction residuals (as suggested by their assertion about the "reactions" of liberals and conservatives depicted in Fig. 1), then we disagree. In particular, it is not clear that authors who wish to understand data containing an interaction should invariably devote journal space to displaying and interpreting interaction residuals instead of interpreting cell means. In fact, interpretation of residuals can sometimes lead to nonsensical conclusions because the interaction per se is a statistical entity that might or might not directly correspond to any meaningful underlying psychological process.

Consider a pattern of data similar to that in Figure 1a, but with different variables. Specifically, imagine an experiment investigating the influence of exposure duration (subliminal [5 ms] vs. supraliminal [500 ms]) and actual letter height ($\frac{1}{4}$ in. vs. 9 in.) on subjects' verbal estimates of the height of letters. See Figure 2a for the expected cell means.

A 2 \times 2 ANOVA shows a main effect of letter size and an Exposure Duration \times Letter Size interaction. The difference-between-differences approach suggests that the interaction exists because there is a greater impact of letter size on perceptions of height when the letters are presented supraliminally than when the letters are presented subliminally. Advocates of the residual approach would presumably suggest that this characterization is wrong. That is, they would assert that the meaning of the interaction can be understood only by interpreting the residuals represented in Figure 2b. According to the residual approach, then, one would interpret the presence of the interaction as indicating that supraliminal- and subliminal-exposure subjects reacted to the presented letters in exactly opposite ways. That is, supraliminal-exposure subjects reacted to the letters by perceiving 9-in. letters as taller than $\frac{1}{4}$ -in. letters, but

1 In the traditional approach, the description and interpretation of the interaction (i.e., difference between differences) often encompasses a description and interpretation of any main effects. Thus, main effects often receive little attention when an interaction is present because the interaction shows that these effects are not uniform across levels of at least one other factor (e.g., Kirk, 1968). Representing interaction variance separately from main effects in the residual approach encourages people to interpret main effect and interaction variance separately, however.

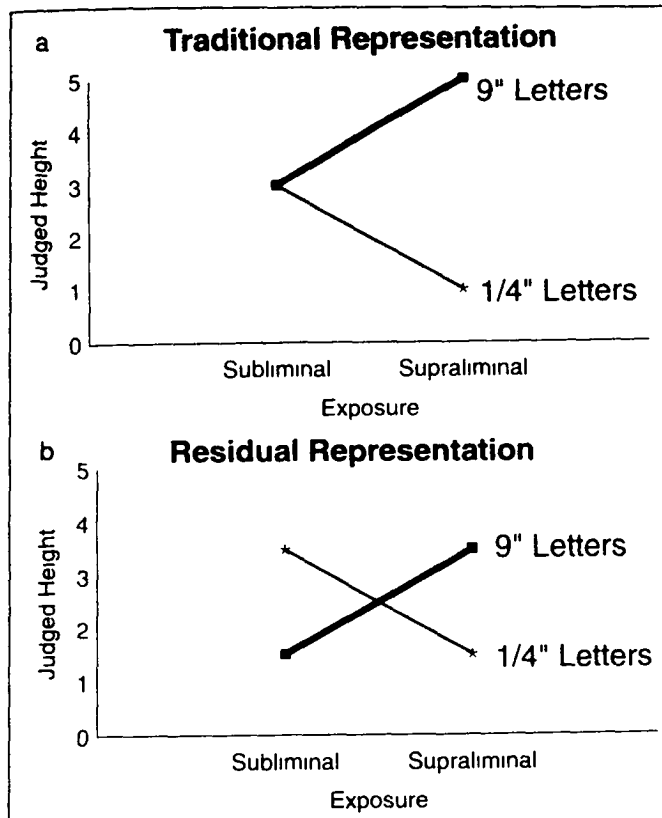


Fig 2 Traditional representation of the difference between differences (a) and residual representation of the interaction (b) in a 2 (exposure) \times 2 (letter height) factorial

subliminal-exposure subjects reacted to the letters by perceiving 1/4-in letters as taller than 9-in letters! Because people who participated in the subliminal conditions could not even see the letters, such a representation of how the subjects reacted seems absurd. In fact, in psychological research, there are many variables that do not lend themselves well to the removal of main effects when interpreting the meaning of interactions.

What Should Researchers Do?

As a matter of practice, we believe that researchers should not confuse statistical sources of variance with underlying psychological processes when interpreting their data. For example, three sources of variance in a 2 \times 2 ANOVA (two main effects and one interaction) could result from three psychological processes, or two, or one, or even four or more! Thus, the residual approach—which encourages interpretation of interaction variance separate from main effect variance—should not be viewed as the default way to impart psychological meaning to data when an interaction is present. For example, it would be inappropriate to take results from a study that has produced an interaction in which the cell means do not show a crossover pattern, remove the main effects from the original cell means, and then argue that the pattern of residuals necessarily provides evidence for a theory that predicts a crossover pattern. Com-

puting residuals for a significant interaction in the 2 \times 2 design will always produce a crossover pattern. This does not imply that a crossover process is responsible for the effect or that a crossover interpretation should be offered.²

Just because an interaction is obtained does not mean that one should necessarily impart psychological meaning to the interaction residuals. If the residual approach leads to nonsensical conclusions given the variables under study (as in the letter-size example), the residual approach can be discarded. Nevertheless, the residual approach can be useful if one has a theory that accounts separately for the main effect and interaction variance. As noted, however, if a researcher prefers a crossover (i.e., residual-based) interpretation of the interaction, he or she must also develop an explanation of the removed main effect (or effects). If two or more plausible interpretations of the interaction still exist, it might then be possible to distinguish among the alternatives using additional data (e.g., measures of the proposed underlying mechanisms, or a new study in which the "main effect" process is removed or inhibited so a proposed crossover process can emerge in the cell means).

USING CONTRASTS IN PLACE OF INTERACTION TESTS

Given the controversy over how to interpret data when an interaction is present, it is not surprising that some controversy has also arisen over how to provide the best statistical test when a pattern that includes interaction variance is hypothesized. In particular, some scholars have recommended that because interaction predictions often suggest a specific pattern of cell means, such predictions should be tested using focused contrasts that conform to the hypothesized pattern (e.g., see Rosenthal & Rosnow, 1985, 1991, Rosnow & Rosenthal, 1995). In this section, we examine the utility of using focused contrasts to test hypothesized patterns that include interactions.

Hypotheses That Imply an Ordering of Cell Means

Contrasts are comparisons of two or more cell means. Contrasts are often used to test the predicted pattern of means by

2 Of course, neither the residual nor the traditional approach necessitates a particular interpretation of the psychological processes underlying the interaction. A researcher using the residual approach could determine that one process accounts for both main effect and interaction variance. If one were to regard both main effect and interaction variance as accounted for by one process, however, there would seem to be little reason to represent and interpret the interaction variance separately from the main effect (or effects). Maintaining that different sources of variance must be separated from each other to be interpreted suggests that the sources of variance have some psychological meaning separate from one another. Although one can imagine some cases in which separable processes are responsible for main effects versus interactions, there are many cases in which moderation of the same process creates both main effect and interaction variance. Therefore, use of residuals as the default method for understanding interactions appears unwise.

Interactions

assigning to each cell a weight that corresponds to the predicted pattern. For instance, if a researcher expects that the means of four cells in a study will be ordered in a decreasing linear fashion with equal spacing between cells, then contrast weights assigned to test that prediction would be +3, +1, -1, and -3, respectively. Any number of possible combinations of weights is possible as long as the weights sum to zero (see Rosenthal & Rosnow, 1985, 1991). The primary purported advantage of contrast analyses is the focused nature of the research questions that can be addressed. For instance, in a one-way design with four cells, investigating a linear trend of cell means by using the contrast weights just listed is more focused on the hypothesis than an omnibus F test in the traditional one-way ANOVA. In fact, ordering contrasts such as the linear contrast described have even been advocated to replace the single- df main effect and interaction tests in a 2×2 design when hypotheses imply an ordering of cell means (see Rosnow & Rosenthal, 1991, 1995).

For instance, consider a 2×2 design discussed by Rosnow and Rosenthal (1989a, 1991) and Meyer (1991). The investigator hypothesizes that the health status of one's child (healthy vs unhealthy) has a larger impact on parental grief when the child dies when the child is male rather than female (see Fig 3). Note that this hypothesis implies the following ordering of cell means on the measure of grief: B (healthy male) > A (healthy female) > C (unhealthy female) > D (unhealthy male). Because contrasts specify particular orders of cell means, Rosnow and Rosenthal (1991) stated that "contrasts sharpen our conception of the hypothesized trends" and "our preference to test [the predicted pattern] would be a contrast with weights +3, +1, -1, -3" (p 574, see Rosnow & Rosenthal, 1995, for a similar recommendation).

One reason for favoring the contrast approach is that a difference-between-differences interaction test ($df = 1$) could show significant results in a 2×2 design even if the ordering of means were quite different from the hypothesized ordering. Therefore, the linear contrast is viewed as answering a more focused question than the interaction test. A perusal of the journals on our shelves turned up many instances of the use of

contrasts to examine hypothesized patterns that included interactions.³

Potential Problems With the Contrast Approach

In order to demonstrate some potential problems with using certain contrasts to replace the traditional interaction test when a difference between differences is hypothesized, we reanalyzed the data from one published study that used such a contrast. The researchers used a Variable X (high vs low) \times Variable Y (high vs low) design, and hypothesized a "specific interaction" such that Y would have a greater impact on the dependent measure when X was high than when X was low.⁴ The results obtained in their study are graphed in Figure 4a.

Because the specific interaction implied that the Y -high/ X -high and Y -low/ X -high cells would be most extreme (with the Y -high/ X -low and Y -low/ X -low cells between the extremes), the authors decided to test the ordering of means using the linear contrast recommended by Rosnow and Rosenthal (1991). That is, the authors weighted the Y -high/ X -high cell as +3, Y -high/ X -low cell as +1, Y -low/ X -low cell as -1, and Y -low/ X -high cell as -3. The linear contrast on the data yielded a highly significant result ($p < .001$), and the authors concluded that their hypothesis was supported. Yet when the data are examined using a traditional ANOVA, there is a strong main effect for variable Y ($p < .001$), but no significant X -by- Y interaction ($p > .25$). That is, Y does not differentially influence the dependent measure across levels of X . The discrepancy between these tests lies in the fact that the linear contrast combines both main effect and interaction variance, whereas the ANOVA interaction reflects the unique variance accounted for by the difference between differences. It would thus appear that the linear contrast used to analyze a 2×2 design can be highly significant when only the main effect portion of the contrast is truly present in the data. This property of linear contrasts implies that a variety of unanticipated patterns of cell means could produce a significant contrast effect.

For example, imagine that the effect of Y was exactly the same across levels of X in the data (see Fig 4b). The same linear contrast supposedly testing a hypothesized specific interaction would produce an even more significant result than for the original data ($p < .00006$), despite the fact that the interaction sum of squares for these data would be zero! Finally, the linear contrast could be significant even if there were a significant difference between differences in the opposite direction of the hypothesized pattern ($p < .023$, see Fig 4c).¹ Given that the chosen contrast is designed to test a specific linear ordering of means, it might surprise some readers to learn that cell means that deviate dramatically from the hypothesized ordering can produce even larger test statistics than cell means that conform to the hypothesis.

3 Rosnow and Rosenthal might not have intended the focused contrast to replace the difference-between-differences approach, but, nonetheless, their comments have apparently been interpreted by some researchers as suggesting this.

4 We do not identify the study because we are simply using these data for convenience. That is, we could have made up some hypothetical data to make the same points.

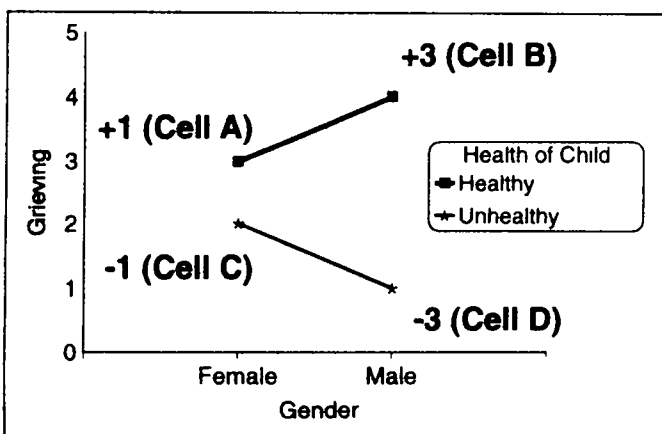


Fig 3 Hypothesized ordering and corresponding contrast weights in a 2 (health of child) \times 2 (gender of child) factorial design

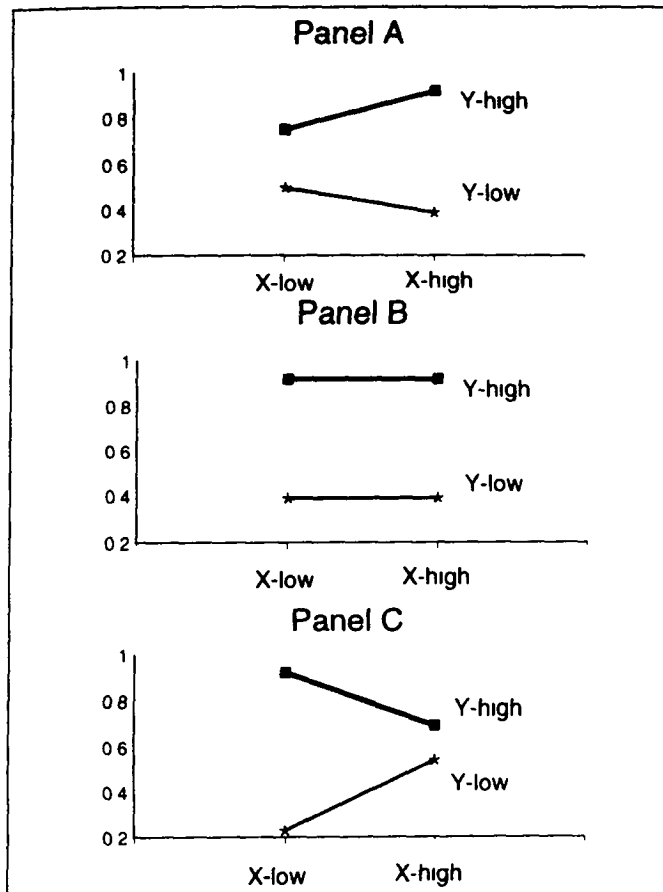


Fig 4 Results subjected to the same linear contrast as in Figure 3. The graphs show (a) actual results from a study hypothesizing an interaction between variables X and Y , (b) results consisting of only a main effect of variable Y , and (c) results including a significant interaction opposite to that hypothesized.

Two Problems With "Focused" Contrasts

This example illustrates two potential problems with the use of contrasts to test a specific ordering of cell means implied by a hypothesis that includes both main effect and interaction variance. The first problem concerns the match between the research hypotheses and the statistical tests. When researchers propose "differential impact" hypotheses, using statistical techniques that confound "differential impact" variance (i.e., interaction variance) with "overall impact" variance (i.e., main effect variance) is inappropriate. Using a test that combines interaction variance with main effect variance provides no diagnostic information about the differential impact portion (or the main effect portion) of the hypothesis. If a researcher hypothesizes a pattern of results that should produce both main effect and interaction variance, the hypothesis is fully supported only if both sources of variance are present in the data. Unfortunately, the use of ordering contrasts in such cases provides information only that the combination of both sources of variance is significant.

The second problem is that a significant focused contrast

does not preclude the possibility that other contrasts account for the data as well as or better than the specified contrast. This difficulty is illustrated in Figure 4b. The linear contrast is significant (and the residual left after the linear contrast is nonsignificant) even though there is no interaction variance in the data. Even though a researcher using such a linear contrast might profess that he or she is not interested in main effects or interactions, it is impossible for the data to entirely match the proposed linear ordering (i.e., $B > A > C > D$, as in Fig. 3) unless both main effect and interaction variance are present in the data (i.e., if one or the other is absent, then at least some of the proposed ordering will fail to occur). Thus, if the data can be completely characterized by a main effect, then a main effect interpretation is a more accurate (and more conceptually parsimonious) characterization of the data than one that implies both main effect and interaction variance should be present.

What Should Researchers Do?

There are at least two possible approaches to address the problem of confounding sources of variance in a contrast: creating contrast weights that orthogonalize the relevant sources of variance and simultaneously testing multiple nonorthogonal contrasts in regression analyses. In the presented example, the authors' predictions included a main effect of Y and the differential impact of Y at different levels of X (which is orthogonal to the main effect). Thus, the statistical tests should have examined the significance of each source of variance, not some combination of the two. Thus, we believe that authors with such hypotheses would be better served by using a traditional ANOVA (which tests orthogonal main effects and interactions) than by using the confounded linear contrast.

An alternative strategy would be to specify several contrasts, each of which might theoretically account for the pattern of data observed. These contrasts could then be compared through the use of simultaneous regression analyses in order to determine which contrast accounts for the most unique variance (i.e., partialing for the effects of the other contrasts being examined, cf. Rosenthal & Rosnow, 1991, pp. 477-478). If the authors of our example study had tested the linear contrast controlling for the main effect of Y , they would have found that the linear contrast becomes nonsignificant ($p > .25$). Of course, computing a contrast that combines main effect and interaction variance while controlling for another contrast that includes only the main effect leaves one with a test that is basically the ANOVA test of the interaction.⁵

GENERAL CONCLUSIONS

When researchers obtain a statistical interaction, there are different approaches to understanding its psychological mean-

⁵ In the preceding sections, we have outlined potential pitfalls in the use of contrasts to test specific orderings of cell means, especially within factorial designs. We do not wish to imply by this critique that contrasts have no place in testing or articulating psychological theory. Contrasts can be especially useful in designs with more than two levels of one or more factors (see Abelson & Prentice, 1995) or as a technique for articulating results (see Abelson, 1995).

ing Although some scholars have suggested that the psychological meaning of interactions in data can be understood only by removing variance due to lower order effects, we believe that there is no canned way to interpret data when an interaction is present. Researchers should not necessarily assume that the significant sources of variance in an ANOVA separately represent some meaningful psychological process (or processes). We believe that in representing and interpreting data containing interactions, researchers should consider the conceptual nature of the variables under study (which can often rule out certain representations of the interaction as implausible), the relevant psychological theories (which can direct researchers toward an appropriate understanding of the interaction), and any additional data necessary to distinguish among plausible interaction interpretations. It is only after these considerations that researchers can make an informed decision regarding how an interaction can best be understood. When researchers use focused contrasts as tests of trends in data, they should be aware that a single contrast sometimes confounds multiple sources of meaningful variance (e.g., the linear contrast we described earlier included both main effect and interaction variance). In such cases, the contrast is not diagnostic of the existence of either effect included (i.e., confounded) in the test. Thus, if one's hypothesis depends on a specific source of variance being present, one cannot adequately test for the existence of that effect using a contrast that combines that variance with other sources of variance.

Acknowledgments—We are grateful to Robert Abelson, Anthony Greenwald, Geoffrey Keppel, Robert C. MacCallum, Gifford Weary, and Timothy Wilson for comments on the July 1993 version of this article. Preparation of this article was supported by National Science Foundation Grants BNS-9021647 and SBR-9520854 to R.E.P., a National Science Foundation predoctoral fellowship to L.R.F., and National Institute of Mental Health traineeships (T32 MH19728) to L.R.F., D.T.W., and J.R.P.

REFERENCES

- Abelson R P (1995) *Statistics as principled argument*. Hillsdale, NJ: Erlbaum.
- Abelson R P & Prentice D A (1995) *Contrast tests of interaction hypotheses*. Unpublished manuscript, Yale University, New Haven, CT.
- Bobko P (1986) A solution to some dilemmas when testing hypotheses about ordinal interactions. *Journal of Applied Psychology* 71, 323-326.
- Fisher R A (1947) *The design of experiments* (4th ed.). Edinburgh: Oliver and Boyd.
- Guilford J P & Fruchter B (1978) *Fundamental statistics in psychology and education* (6th ed.). New York: McGraw-Hill.
- Hoaglin D C, Mosteller F & Tukey J W (1991) *Fundamentals of exploratory analysis of variance*. New York: Wiley.
- Judd C M, McClelland G H & Culhane S E (1995) Data analysis: Continuing issues in the everyday analysis of psychological data. *Annual Review of Psychology* 46, 433-465.
- Keppel G (1991) *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Kirk R E (1968) *Experimental design: Procedures for the behavioral sciences*. Belmont, CA: Brooks-Cole.
- Meyer D L (1991) Misinterpretation of interaction effects: A reply to Rosnow and Rosenthal. *Psychological Bulletin* 110, 571-573.
- Myers J L (1979) *Fundamentals of experimental design* (3rd ed.). Boston: Allyn and Bacon.
- Pedhazur E J (1982) *Multiple regression in behavioral research: Explanation and prediction* (2nd ed.). New York: Holt, Rinehart & Winston.
- Rosenthal R & Rosnow R L (1985) *Contrast analysis: Focused comparisons in the analysis of variance*. New York: 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 (1989a) Definition and interpretation of interaction effects. *Psychological Bulletin* 105, 143-146.
- Rosnow R L & Rosenthal R (1989b) Statistical procedures and the justification of knowledge in psychological science. *American Psychologist* 44, 1276-1284.
- 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 & Rosenthal R (1995) Some things you learn aren't so Cohen's paradox, Asch's paradigm, and the interpretation of interaction. *Psychological Science* 6, 3-9.
- Ross W T & Creyer E H (1993) Interpreting interactions: Raw means or residual means? *Journal of Consumer Research* 20, 330-338.
- Zuckerman M, Hodgins H S, Zuckerman A & Rosenthal R (1993) Contemporary issues in the analysis of data. *Psychological Science* 4, 49-53.

(RECEIVED 3/23/95, ACCEPTED 7/13/95)

