Difference score correlations in relationship research: A conceptual primer

DALE GRIFFIN,[a] SANDRA MURRAY,[b] AND RICHARD GONZALEZ[c]
[a]*University of Sussex, Brighton, UK;* [b]*State University of New York, Buffalo; and*
[c]*University of Michigan, Ann Arbor*

Abstract

The practice of computing correlations between "difference" or "discrepancy" scores and an outcome variable
is common in many areas of social science. Relationship researchers most commonly use difference scores to
index the (dis)similarity of members of two-person relationships. Using an intuitive, graphical approach—and
avoiding formulas and pointing fingers—we illustrate problems with using difference score correlations in
relationship research, suggest ways to ensure that difference score correlations are maximally informative, and
briefly review alternatives to difference score correlations in studying similarity, accuracy, and related
constructs.

Studies of interpersonal relationships often focus on the similarity, matching, or congruence of two individuals in a dyad. For example, many studies have investigated whether it is better for two partners to be similar or dissimilar in some domain of interest. Is it best for working women when they and their partners earn similar salaries? Is it best for a married woman to perceive her partner accurately, or is it better for a wife to idealize her husband? Is it best for a man when his ideal partner is similar to his actual partner or is it best to hold high ideals regardless of the real state of affairs? These kinds of questions—about similarity, understanding, congruency or accuracy—are often translated into statistical hypotheses involving difference score correlations. In the first example, a researcher might take the difference between the man's salary and the woman's salary for each of N heterosexual couples ($Sal_M - Sal_F$) and correlate that with the woman's level of relationship satisfaction. Or, in the second example, a researcher might take the difference between a woman's view of her partner and that partner's own self-perception ($WPer_M - MPer_M$) and correlate that with the woman's relationship satisfaction. Or, in the final example, a researcher might take the difference between a man's report of his ideal partner and his perception of his actual partner ($WIdeal_M - WPer_M$) and correlate that with his relationship satisfaction.

Most relationship researchers are aware that the practice of calculating such differ-

ence score correlations has been discussed and criticized by methodologists and psychometricians (e.g., Cohen & Cohen, 1983; Cronbach & Furby, 1970; Edwards, 1994a; Johns, 1981; Zimmerman, 1997). But it has been our experience, particularly as reviewers, that whereas most researchers acknowledge the complexities introduced by the well-advertised "unreliability" of difference scores, many do not acknowledge the more basic conceptual problems inherent in difference score correlations. Although a number of excellent discussions of difference score correlations are available in standard textbooks (for a particularly clear and insightful example, see Cohen & Cohen, 1983), our presentation here is more practical than most, focusing on graphs and applied examples rather than formulas, and emphasizing applications encountered in relationship research. Throughout, our focus is not on declaring what methods are "correct" versus "incorrect," but on illustrating the conceptual meaning of particular analytic methods. Our aim is to shed light on the issues, not to bury difference scores. (For a good selection of more contentious approaches on both sides, see the debate among Edwards, 1994b; Bedeian & Day, 1994; and Tisak & Smith, 1994, in the special issue of the *Journal of Management,* 1994, Vol. 20, No. 3, and the exchange between Zuckerman & Knee, 1996, and Colvin, Block, & Funder, 1996.)

Historically, three primary arguments have been presented in favor of using difference scores (see Tisak & Smith, 1994, for a detailed review of these arguments) in correlational research. First, difference scores $(X_1 - X_2)$ have face validity along with the intuitive appeal that people constantly think about differences in everyday life: If we discuss whether it is better for partners to be similar in height or physical attractiveness, we talk about *differences* in height and physical attractiveness. Thus, everyone knows what such a combination means, don't they? This represents what Johns (1981) has called the "seductive face validity" of difference scores for measuring dyadic similarity. As we show in Part 1 be-

low, although the meaning of a given difference score seems obvious, it is impossible to know what a *difference score correlation* means without looking at the relations between the components of the difference score and the outcome measure (e.g., without examining r_{x1y} and r_{x2y} as well as $r_{(x1-x2)y}$).

Second, proponents argue that difference scores capture a unique *combination* of the underlying components; that is, difference scores are more than a sum of their parts. Tisak and Smith (1994) assert that "Compared to their component variables, difference scores . . . certainly capture something conceptually different" (p. 679). The statistical logic of this point is hard to follow: Difference scores explain exactly as much variance in a regression equation as the two components entered individually and constrained to have regression weights of $+1$ and -1, respectively. Thus, when their use is appropriate, conclusions derived from difference score correlations or regressions can always be translated into conclusions about the individual measures.

However, the reverse is not true: Conclusions derived from entering individual components *may or may not* translate into conclusions about difference scores. Again, the fundamental problem is that significant difference score correlations are uninformative about which of a number of underlying models are true. The problem, as we document below, is that difference score correlations alone are *less* informative than the sum of their parts.

A third point raised in defense of difference scores is that difference scores are used, explicitly or implicitly, in repeated measures analysis of variance (ANOVA) and its variants, including mixed within- and between-subjects designs. If difference scores are acceptable within one standard statistical technique (ANOVA) why should they be scorned when used within another standard technique (correlations)? In repeated-measures ANOVAs, difference scores are used as *dependent variables,* rather than as *independent variables,* but the issues are the same. In fact, using differ-

ence score correlations alone is analogous to conducting a mixed-model ANOVA, reporting an overall *F*-statistic, and neglecting to consider the pattern of means and variances in one's interpretation! As we note below, the minimal requirement for informed use of difference score correlations is to report the underlying effects analogous to the means and variances in an ANOVA model. In order to interpret a significant interaction in an ANOVA, one must examine all cell means.

Our short review proceeds in three stages. In Part 1 we demonstrate the conceptual problems with difference score correlations using a set of contrived data and simplified questions. In Part 2 we illustrate the same problems using examples from actual data sets. In Part 3 we discuss some of the alternatives to difference score correlations and briefly review the pros and cons of various methods.

Part 1: The Basics

In this section we will consider the first example study listed above: How does the difference between the salary earned by a man and a woman affect that woman's relationship satisfaction? We use salaries earned by husbands and wives as the "difference" variable because such a quantity could be measured without error, and thus reliability issues do not contribute to the problems illustrated here. To simplify the illustration (but not to change the underlying algebra in any way) we dichotomize the salary variable and represent the continuous distribution of salaries by only two values: Salaries can be low (represented by 10k) or high (represented by 30k). Thus, couples can fall into one of four categories (both highly paid, the 30 30 group; high pay men and low pay women, the 30 10 group; low pay men and high pay women, the 10 30 group; and both poorly paid, the 10 10 group). Imagine that our survey indicates that there is some degree of "sorting" by salary, so that the two dichotomous variables correlate at .25. This dichotomization allows us to illustrate our points easily; even though the distribu-

tion is no longer multivariate normal, the underlying points remain the same.

Imagine that we surveyed 24 working couples, verified the annual salaries of each member of each couple (to ensure error-free data), and then proceeded to create difference scores in the direction of husband minus wife. Imagine that our theoretical model led us to predict that women who earn salaries similar to their partners will be happiest. Imagine further that we obtained a strong—and significant—negative correlation between the signed difference in salary and the satisfaction variable. Does this mean that small differences (i.e., similarity in salary) relate to high satisfaction? Should we jump up and prepare to publish this confirmatory finding?

On first thought, it might appear that a negative correlation between the difference in salary between men and women and women's satisfaction implies that high satisfaction is associated with similar salary levels. However, there are two problems with this inference. First, of course, difference scores do not represent "similarity"; they range between large negative values and large positive values, and similarity is thus in the middle of the scale near 0. This is a seemingly trivial point, but one that comes up repeatedly in reviews of relationship research. Second, as noted in the introduction, difference score correlations themselves tell little about the actual *meaning* of the result. For example, all four patterns of data presented in Figure 1 (panels A through D) are consistent with a negative difference score correlation. In fact, significant difference score correlations of $-.5$ can be found in each case. But each of these patterns has a very different interpretation!

Before reviewing the figures in detail, it is helpful to review some combination models that are of interest in relationship research: a *discrepancy* model, a *similarity* model, and a *superiority* model. A discrepancy, or signed difference, model is what we described in the introduction; this implies that the outcome will be greatest when X1 is highest and X2 is lowest; a similarity, or absolute difference, model implies that the
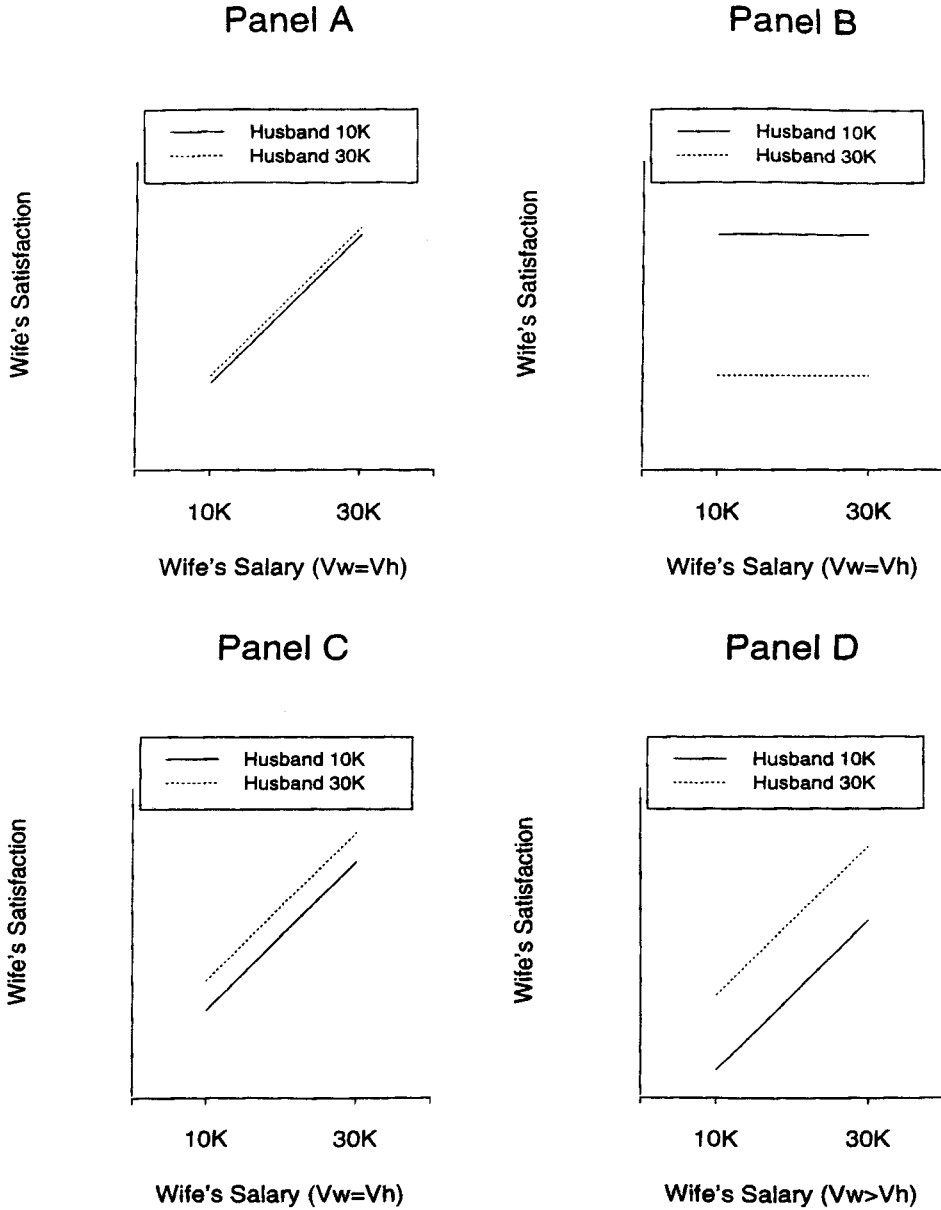
Panel A

Panel B

Panel C

Panel D

Figure 1. Four different patterns that give rise to the same difference score correlation. The example is illustrative so we do not scale the Y axis. Vw and Vh refer to the variance of wife variable and husband variable, respectively.

outcome will be greatest when X1 and X2 are closest together, and a superiority model implies that there is something special about X1 exceeding the level of X2. For the salary example, the discrepancy model implies that a woman's satisfaction will be greatest the more she earns and the less her partner earns, the similarity model implies that the woman's satisfaction will be greatest the smaller the absolute value of the difference between her own and her partner's salary, and the superiority model implies that the woman will be happiest when she earns *more* than her partner. Note that

each of these models takes into account the combination of scores as opposed to simple main effect models that imply that satisfaction will be increased the more either partner's salary increases.

Now let us turn back to Figure 1 and consider this distinction between combination models and main effect models. In panel A, it seems that a woman's level of satisfaction is virtually *unrelated* to her partner's salary, and instead depends only on her own salary: a positive main effect for own salary. Almost regardless of what her husband earns (the correlation between satisfaction and partner's salary is negligible),[1] a woman tends to be more satisfied the more money she earns (the correlation between satisfaction and own salary is positive and large)—yet despite the "main effect" interpretation, the correlation between the salary difference score and the woman's satisfaction is −.5. In panel B, in contrast, the interpretation is conceptually reversed—though the difference score correlation is identical! In this case, a woman's level of satisfaction is virtually *unrelated* to her own salary, and negatively related to her partner's salary (a negative main effect for partner's salary). In panel C, there is yet another interpretation for an identical difference score correlation. This time, a woman's relationship satisfaction is moderately related to her husband's salary in a positive direction, but is strongly and positively related to her own salary (two positive main effects, with the own salary effect stronger than partner salary). Note that for these three examples, we explicitly noted in Figure 1 that the variances of the men's salaries equalled the variances for the women's salaries. In such situations, the difference score correlation depends only on the difference between the relevant "main effect" correlations.[2]

In many cases, such as those illustrated in panels A through C of Figure 1, the variances of the two variables entering into the difference score are equal (or approximately equal), and one can then interpret the difference score correlation with reference to the two main effect correlations (r_{x1y} and r_{x2y}). However, unequal variances can also affect the difference score correlation and its interpretation. Panel D portrays such a circumstance. Again, the difference score correlation is significantly negative, but that is not because the two main effect correlations differ. In fact, despite the apparent mean differences, the correlation between satisfaction and partner's salary is identical to that between satisfaction and the woman's own salary—but the variability of women's salaries is much greater. That is, *taking the difference in variability into account* between the salaries paid to men and those paid to women, a woman's satisfaction is equally related to her own and to her partner's salary. The phenomenon to be explained here, then, is the difference in variability.[3] But the difference score does not know this. A given difference score does not know or care if its value is created by one positive main effect only, one negative main effect only, two main effects where one is larger than the other, or two main effects where one has greater variability than the other. But the re-

1. In Figure 1 the correlation between satisfaction and husband's salary is nonzero only because of the intercorrelation between partners' salaries.
2. A quick look at the formula for difference score correlations is revealing. When the difference score is created as $D = X_1 - X_2$ and correlated with Y, then the correlation between D and Y can be expressed

as a function of the correlations among, and standard deviations (denoted S) of, the original variables. The numerator of that equation, which controls the sign of the correlation, is $[r_{x1y}S_{x1} - r_{x2y}S_{x2}]S_y$. From this, one can see that the standard deviations of X_1 and X_2 *weight* the respective main-effect correlations. One obvious "solution" to the unequal variances problem is to standardize the variables so that both variables entering the difference score have a variance of 1. However, this practice throws away the original metric of the measures and so is irrelevant to questions of similarity.

3. Note that in a dichotomous design, variance is at the maximum when there are an equal number of observations at each dichotomous value. One plausible explanation for greater variability for men's salaries would be if most salaries were generally low and only a few "stars" made larger amounts of money.

searcher should know! And like the good reporter, the researcher must get the story right. Too often, stories are told about discrepancy, similarity, or superiority models when only simple main effect stories are there to be told.

The lesson so far is that difference score correlations must be interpreted in light of both the main effect correlations and the main effect variances. A particular difference score correlation does not have a unique interpretation. This problem is commonly known as the "confounding of difference scores with their constituents" (see, e.g., Edwards, 1994b; Johns, 1981). Only by examining the main effect constituent correlations can one identify and interpret the specific confound (or demonstrate that a given combination model is appropriate—but we return to this issue later). There are a number of ways of thinking about the problem with main effects confounding the difference scores. One way is to consider extreme examples. If one of the variables entering into the difference is a constant so that every score in the sample is the same, the difference score correlation will be exactly the same as that between the outcome and the other (nonconstant) constituent variable. Or if one of the variables entering into the difference is randomly generated (and therefore correlates near-zero with all other variables), the difference score will be virtually the same as that between the outcome and the other (nonrandomly generated) variable. Again, such circumstances can only be diagnosed by looking at the individual correlations that together make up the difference score correlation.

Multiple regression gives another explanatory perspective: If both main effects (constituents) are entered as predictors to explain an outcome variable, there is no additional variance that the raw *difference* between the two predictors can explain. The difference score is one possible weighted combination of the two original predictors that is entirely redundant with the main effects. The difference score weights (1, and −1) are but one possible

way of combining the two variables, but not necessarily the right way (Cohen & Cohen, 1983).

The way that main effects confound difference scores is most clearly seen in a simple bivariate scatterplot. Figure 2 presents a scatterplot of a hypothetical data set containing the annual salaries for husbands and wives in 24 couples. Within couples, salary level is correlated at $r = .35$. For simplicity, mean salary is 25K for both husbands and wives. The solid diagonal line on the graph represents the points of equality, where husbands and wives earn equal salary; the central square represents the mean salary point [25, 25] (and obscures one data point). Each point on the graph represents the two salaries for a given couple, with husbands plotted on the Y axis and wives plotted on the X axis. Triangles represent couples below the equality line (where men earn less than women—that is, those couples for whom the difference score is negative) and pluses represent couples above the equality line (where men earn more than women—that is, those couples for whom the difference score is positive). The difference score for each couple is simply the vertical distance between the solid line and the relevant point.

Consider the 13 pluses on the graph in Figure 2; how many of these positive difference scores include couples that have women earning more than average? Three pluses fall to the right of $M = 25$ on the X axis. But 10 pluses fall to the left of that average. This means that the majority of the difference scores that are positive (men earning more than women) include women who are "below average" in earning power, compared to the other women sampled. Similarly, seven triangles fall to the right of $M = 25$, three to the left. Thus, the majority of negative difference scores (cases where women earn more than men) include women who are "above average" earners. This is the most direct way to illustrate how the difference score is highly *confounded* with the main effect of wife's salary. (The same counting exercise on the Y axis will show that the difference score is just as
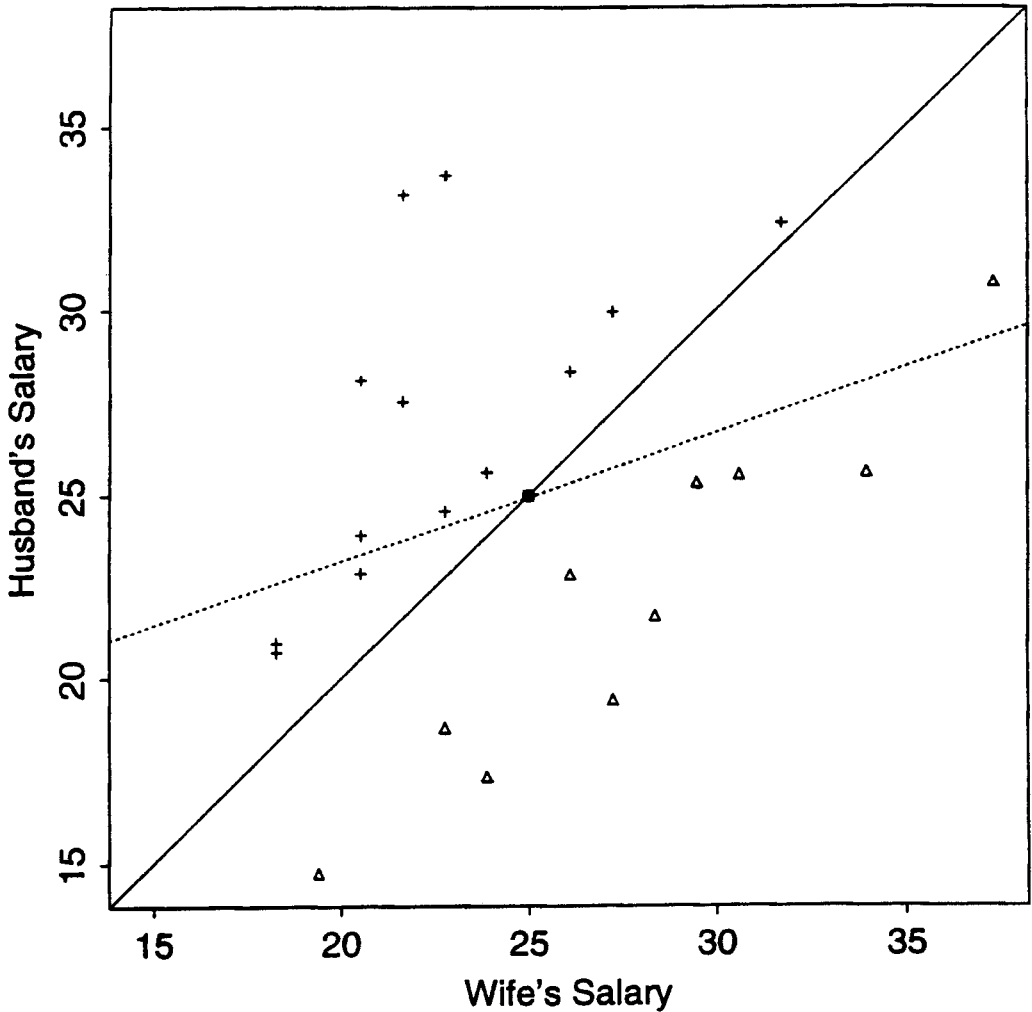
Figure 2. A bivariate scatterplot of women's salary (X) against men's salary (Y). Note that variables X and Y are intercorrelated at +.35. Triangles denote couples where women earn more than men, and pluses denote couples where men earn more than women.

confounded with the number of men who are low and high earners.)

This result is easiest to see when the two predictor variables have equal means, but it is true no matter what the mean structure. Imagine that we added 20k to each husband's salary, thus making the husband's mean salary equal to 45k. This would result in every point becoming a plus (i.e., every point falling above the equality line). In such a case, our demonstration would proceed by drawing a new line parallel to the line of equality, but raised upwards by 20

points. This line would serve to categorize the couples into those whose *differences* (always in favor of the husband's salary) were above the mean and those who were below the mean. We would then find that those couples who were above the mean in salary difference were also below the mean in wife's salary, and so forth. Thus, the problems of confounding would remain. It should also be clear from this example that *categorizing* the difference scores into "underachieving women" and "overachieving women" and then relating this binary vari-

able to some outcome will lead to the same kind of confounding with the main effects. Hence, the same problems occur when categories are created from difference scores and used in subsequent t-tests or ANOVAs.

The equal-mean case displayed in Figure 2 invites consideration of a related concept: absolute difference scores. When we have as many points below the line as above it, absolute difference scores seem like a very sensible way to capture similarity. Those points that are close to the equality line represent couples who are similar, regardless of who makes more money. And best of all, in this case the absolute difference score will be uncorrelated (approximately) with the signed difference score and both of the main effects. But consider the unequal mean case that we discussed above; if all husbands earned more money than their wives, what would the absolute difference score tell us? In this case, all *signed* difference scores are positive, and so an absolute value transformation changes nothing. In other words, for the second case, absolute difference scores would be perfectly correlated with the signed difference scores, and thus highly correlated with each of the main effects. Lesson 1: Absolute difference score correlations mean what you think they mean only in very special cases. Lesson 2: We need tests that directly target and discriminate among the combination hypotheses common in relationship research (i.e., at least the discrepancy, similarity, and superiority hypotheses) and avoid the confounding effects of difference score correlations. But first, should the reader not yet be convinced, let us examine the problems of difference score correlations in actual data.

Part 2: Two Paradigmatic "Real-life" Examples

It does not take contrived data sets to show the dramatic misunderstandings that can arise from analyzing difference scores without considering their underlying constituents. Consider the hypothesis that egocentric self-perception or self-enhancement is related to conflict in a romantic relationship. That is, conflict should increase to the extent that one member's (the actor's) self-perceptions are more positive than the partner's view of that actor. Reanalyzing data collected by Murray, Holmes, and Griffin (1996) on 121 heterosexual couples we can test this hypothesis using a difference score approach. By subtracting the positivity of a woman's view of her partner from the positivity of the man's self-perception, we created a men's self-enhancement score, and by a corresponding transformation, a women's self-enhancement score (the positivity measure was the within-subject correlation between trait ratings and the social desirability score of each trait, as used by Colvin, Block, & Funder, 1996). These difference scores were then correlated with the outcome variable of interest: the opposite-sex partner's report of relationship conflict. Thus, according to a simplistic "congruence," "accuracy," "similarity" hypothesis, men's self-enhancement, which is not "congruent" with their partners' perception, will be positively related to women's reports of conflict, and, by the same logic, women's self-enhancement will be positively related to men's reports of conflict.

The results of these difference score analyses seem to offer clear support for the notion that self-enhancement leads to a higher level of conflict. The self-enhancement difference score for men correlated .29 with women's report of conflict, and the difference score for women correlated .28 with men's report of conflict. Both correlations are highly significant. But what do the "main effect" correlations tell us? The positivity of women's self-perceptions are correlated *negatively* $(-.35)$ with their partner's reports of conflict, and men's perceptions of their partner are even more highly *negatively* correlated $(-.57)$ with those reports of conflict. Thus, the significantly positive difference score correlation comes from two (unequal) main effects: Men's reports of conflict are more closely tied to their perceptions of their partners than to their partners' self-perceptions. Although the difference score correlations for

women's conflict reports are the same as for men's reports, the underlying main effects are different. In this case, the positivity of men's self-perceptions are virtually unrelated (−.06) to their partner's reports of conflict, whereas women's perceptions of their partner are strongly related (−.45) with their own reports of conflict. In this case, then, the significantly positive difference score reflects only the women's tendency to report more conflict if they view their partner negatively.

Consider another hypothesis that can be tested on the Murray et al. (1996) data. Is it better to have high romantic ideals or more realistic romantic ideals? Perhaps having an unrealistically high ideal partner makes one more disappointed in one's actual partner. One way to test this hypothesis, it would seem, is to take the difference between how people rate their ideal partner and how they rate their actual partner, and compare this difference to how much they love their partner. Measures of how positively men and women rate their ideal partner, how positively they rate their actual partner, and how much they love their actual partner are available in the Murray et al. (1996) data set. As operationalized by difference scores between ideal and actual partner, the degrees of unrealistic optimism in people's ideals were correlated *negatively* with love for one's partner for both men (−.28) and women (−.21). Apparently, high ideals—at least relative to one's own partner—are bad for romance.

An examination of the relevant main effect correlations leaves us with a puzzle. For both men and women, the correlation between ideal ratings and love for partner was almost exactly equal to the correlation between ratings of partner and love for partner. And all these correlations were positive! For men, ideal ratings correlated .49 with love for partner whereas actual partner ratings correlated .53 with love for partner. For women, ideal ratings correlated .39 with love for partner and actual partner ratings correlated .41 with love for partner. How, then, did the negative difference score correlations come about? An inspection of

the relevant main effect variances helps to clear up this mystery: For both men and women, variances for the actual partner ratings were about three times greater than for the ideal partner ratings. Intuitively, it is easy to see why this should be so. People tend to have very similar (and high) ideals, yielding little between-person variance, but there will be substantial variation in people's perceptions of their actual partners. It is this unremarkable variance effect that gives rise to the negative difference score. On the evidence we have so far, it is those people who express the highest ideals who tend to express the greatest love for their partners, in clear contradiction of what the difference score correlations, when examined in isolation, seemed to imply.

Part 3: Beyond Difference Score Correlations

Any researcher who uses a difference score correlation must *at minimum* describe the underlying correlations and variances that control the interpretation of the overall result. With this information in hand, the reader is then able to determine whether the researcher's preferred interpretation actually fits the pattern of data observed. However, researchers usually want to go beyond this level of description for two reasons. First, the two main effects that make up the difference are often correlated—in other words, the two main effects are nonorthogonal. A researcher often will desire to know the effect of one of the main effect variables on the outcome above and beyond the effect of the second constituent variable. Second, the descriptive correlations (and, by extension, the difference score correlation itself) are generally uninformative about the questions of similarity, understanding, and accuracy with which we began. That is, they do not help test different combination models.

Standard multiple regression approaches

It is well known that multiple regression analysis can be used to "unconfound" the

effects of two predictor variables. The dotted line in Figure 2 represents the regression of husband's salary on wife's salary. The vertical distance between the dotted line and any given point is the *residual* salary for the wife controlling for the husband's salary. When we then relate marital satisfaction to such residual scores, we can ask the *conditional* question "for women who have identical salaries, what effect does their husband's salary have on their marital satisfaction?" Note that the regression line is drawn in such a way that there are about an equal number[4] of positive and negative residuals regardless of whether the wife's salary is high, medium, or low. Thus, counting will show that of the positive residualized scores for husband's salary, seven are found above the mean on X, and eight below. Of the negative residualized scores for husband's salary, four are above, and six below. With larger samples under a bivariate normal distribution, these values would become even closer. Thus, the problem of confounding is solved by conducting a simultaneous multiple regression on both constituent variables. Unfortunately, this is not a complete solution to the more general problems of testing combination models, because the multiple regression procedure in its simple form is designed to test main effects models.

As noted, the multiple regression approach has the advantage of producing results that are more clearly interpretable because they are "cleaner": What is the net effect of one predictor controlling for the other? Looking back to the real example of self-enhancement and relationship conflict, we find that multiple regression does produce a useful answer. In contrast to the difference score result (significantly positive correlations between self-enhancement difference score and conflict for both men and women), the multiple regression approach reveals that relatively higher self-images for women were associated with their part-ners reporting relatively less conflict (standardized $B = -.19$, $p < .05$) when their male partners' views were held constant, whereas there was virtually no relationship between the self-images of men and their partners' reports of conflict (standardized $B = .05$), when their female partners' views were held constant. Thus, on average, it seems it is better for a relationship when the woman partner has a higher, rather than a lower, opinion of herself. However, this "more is better" result is averaged over the whole range of the woman's self-perceptions, and does not directly address whether having a higher—or lower—perception of oneself *relative* to the partner's perception is bad or good. This is a critical problem.

There are other limitations to the multiple regression approach as well. It does not solve the "reliability" problem, for example.[5] The residualized scores used in such multiple regression analyses are just as unreliable as are difference scores when the two predictors are highly correlated (see Judd & Kenny, 1981). Unreliability is a fact of life when one wishes to separate the effects of two highly correlated variables. In addition, the interpretation of multiple regression results can be complex when the two predictors are the same variable measured over time (e.g., a pre-test and a post-test). In such cases, whereas difference scores do measure "change" (even though the difference score correlations will still need to be interpreted in light of their main effects), the residualized scores do not. The time-2 residualized scores represent "deviations from expectation" based on time 1, and can be negative when the change is positive and vice versa. There are even special cases of modeling change where the difference score approach is most appropriate to the model being tested (see Judd & Kenny, 1981, ch. 6).

However, the most serious problem with multiple regression in assessing hypotheses

---

4. More precisely, the regression line is drawn so that the residuals on Y are uncorrelated with the original scores on X.

5. The "reliability problem" is that as two predictors become more highly positively correlated, the difference between them becomes less reliable.

about agreement, understanding, or accuracy is that standard multiple regression is a very general approach, whereas specific models such as "similar is better" or "self-enhancing is better" are not directly addressed by the regression coefficients. In this area, like others, specific focused questions require a specific focused method of analysis (Kenny, 1996). One common solution is to use a multiplicative interaction term, a "cross-product" term, combining the two predictors. Unfortunately, this approach is still not precisely focused on a *specific* model of interest. The multiplicative interaction term, when entered after the two main effect predictors, does indicate whether or not the effect of one predictor depends on the level of the other predictor. So, in the salary example, a significant interaction would mean that the effect of the husband's salary on the wife's satisfaction would vary according to the level of the wife's salary. In general, the specific form of the interaction can be understood by plotting the regression of satisfaction on husband's salary separately for high, medium, and low levels of wife's salary (see Cohen & Cohen, 1983). Graphical approaches to understanding interaction terms in multiple regression are presented in Judd and McClelland (1989). Such plots can be informative about the general type of interaction between the two predictors, but they are not precisely targeted at determining whether similarity is good or bad.

Dichotomized variables approach

A more precise, but statistically questionable, technique is to categorize the continuous variables (e.g., high, medium, and low) and then examine the resulting 3 × 3 table. Murray et al. (1996) used this approach (among others) to examine whether the congruency between a partner's perception and one's own self-perception was related to relationship satisfaction. The congruency model predicted that the diagonal or agreement cells (high-high, medium-medium, and low-low) where partner perception

mirrored self-perception should all be marked by high satisfaction (see Abelson & Prentice, 1997, for a more formal discussion of such tests). Murray and colleagues contrasted this with a "partner-enhancement" model which predicted that satisfaction would increase with the positivity of partner's impressions, regardless of the congruency between self and partner. In general, the enhancement model received support, but this kind of analysis is always best viewed as descriptive, given the statistical problems associated with categorizing continuous variables in the context of multiple regression (Maxwell & Delaney, 1993).

It is also possible to combine a multiple regression approach (to control for confounding) with a categorical approach that offers a more focused test. If one is only interested in the effects of absolute similarity, then one may enter the absolute value term after the two main effect terms have been entered in a multiple regression equation. This method (recommended by Kenny, 1988) will test whether similarity is good or bad, after correcting for the main effects. However, it should be noted that the absolute similarity score still suffers from confounding when there are unequal variances among the two components. The other disadvantage with this method is that it is focused on only one model (absolute similarity) and cannot diagnose the correct combination model in a given set of data.

A model-based approach

Ideally, a method should be able to distinguish among the three basic combination models that might underlie the dyadic data structures we have discussed (as well as identify even more complex models that were not discussed in this paper). Recall that in the salary example, the discrepancy model implied that a woman's satisfaction would be greatest the more she earned and the less her partner earned; the similarity model implied that the woman's satisfaction would be greatest the smaller the absolute value of the difference between her

own and her partner's salary; and the supe-
riority model implied that the woman
would be happiest when she earned *more*
than her partner.

Recently, Edwards (1994a) explained
how these models can be distinguished us-
ing a *piece-wise* linear regression that in es-
sence computes two separate regression
equations simultaneously. For the salary ex-
ample presented in Figure 2, two regres-
sions would be computed: one for those
couples below the equality line and one for
those above the equality line (any couples
exactly on the equality line would be as-
signed to one of the other groups unless
there were enough to be treated as a sepa-
rate group). Each regression analysis would
enter men's and women's salary with the
women's level of satisfaction as the depend-
ent variable. According to the discrepancy
hypothesis, both regressions should yield a
negative coefficient for husband's salary
and a positive coefficient for wife's salary
(maximizing the difference).

The combination models imply different
patterns for the regression coefficients. Ac-
cording to the similarity hypothesis, the
analysis on scores below the equality line
(where wives' salaries exceed husbands'
salaries) should yield a positive coefficient
for the husbands' salaries and a negative
coefficient for wives' salaries because in
this region higher husbands' salaries iden-
tify couples who are closer to equality. In
contrast, the analysis on scores above the
equality line (where husbands' salaries ex-
ceed wives' salaries) should yield a negative
coefficient for the husbands' salaries and a
positive coefficient for wives' salaries be-
cause in this region higher wives' salaries
identify couples who are closer to equality.
Finally, according to the superiority hy-
pothesis, satisfaction should be particularly
high for those women who are below the
equality line because they earn more than
their partners, and this should translate into
either a larger positive coefficient for
wives' salaries for those below the equality
line or a larger positive intercept term for
those below the equality line. Thus, each
hypothesis or model implies a qualitatively

different pattern of regression weights
across the two regressions.[6]

When the piecewise method is used to
examine the link between self-enhance-
ment and conflict in our ongoing example, it
reveals a complicated, but informative,
story. In brief, none of the basic models fit
the data. Consider women's self-ratings
first. In this example, self-enhancement is
above the equality line (women's self-rat-
ings were higher than partner's ratings of
them in 59 cases) and self-derogation is be-
low the equality line (self-ratings were
lower than partner's ratings of them in 62
cases). When women were self-derogating,
the positivity of their self-ratings strongly
and negatively related to their partners' re-
ports of conflict (and their partners' ratings
were marginally negatively related to con-
flict), but when women were self-enhancing,
the positivity of their self-reports was unre-
lated to partner reports of conflict. In other
words, none of the three "simple" models fit
these data: For modest women (or women
with flattering partners) relationships
worked best the higher their self-image, but
for self-enhancing women (or women with
critical partners) self-image was unrelated
to how smoothly their relationships ran.

The pattern for men was quite different.
First, compared to women, a larger number
of men (89) were self-enhancing, and a
smaller number (32) were self-derogating.
Second, men's self-ratings were unrelated
to their partners' reports of conflict,
whether or not the men were self-enhanc-
ing. Third, women's ratings of their partners
were significantly and negatively related to
their own reports of conflict—this relation-
ship was moderately, but not significantly,
greater when men were self-enhancing.
Thus, in neither case was there evidence
that similarity, modesty, or superiority was
beneficial, and for women only, a hint that
a modest self-image was bad for the rela-
tionship. Naturally, all of these conclusions

---

6. Edwards's method does not actually calculate two
   separate regression equations, but uses dummy
   variables to estimate simultaneously the different
   "pieces" of the piecewise regression equation.

need to be qualified by the usual disclaimers that follow from analyzing nonexperimental data.

Although we believe that the capacity for testing among different models of similarity or accuracy makes the piecewise approach the method of choice for such questions, the approach is not without limitations. First, it is complicated; it involves several steps and several tests for the equality of dependent regression coefficients. Thus, it can be most easily carried out using Structural Equation Modeling software. Second, there are patterns of data in which it is impossible to distinguish these three models by empirical means. For example, when all points are on one side or the other of the equality line, there is no way of distinguishing whether the discrepancy, similarity, or superiority model is best. The piecewise method cannot be used unless there are discrepancies of both signs. Third, the practice of dividing points at the equality line depends on having measures that are reliable enough to categorize most points correctly. This issue has not been studied, to our knowledge. But the model certainly deserves attention—from metho-

dologists and from practically oriented relationship researchers. (The same general approach can be used in place of difference scores as *dependent variables*; Edwards, 1995.)

Conclusion

The practice of using difference score correlations alone to draw inferences about the benefits or costs of similarity or accuracy is like characterizing the movie *Casablanca* as a story about a man, a woman, and an airplane. The results of a difference score correlation are just as "correct" as the foregoing summary of the movie, but in both cases the overall summaries leave the audience in the dark. Relationship researchers should tell the whole story and, like a successful movie director, try to squeeze all the information out of their source. We have briefly reviewed some methods to guide true stories about similarity, accuracy, and understanding. The relationships area is full of good real-life stories that require precise methods because, unlike a movie director, the relationship researcher is constrained by the data.

References

Abelson, R. P., & Prentice, D. A. (1997). Contrast tests of interaction hypotheses. *Psychological Methods, 2,* 315–328.

Bedeian, A. G., & Day, D. V. (1994). Difference scores—rationale, formulation, and interpretation. *Journal of Management, 20,* 673–674.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Colvin, C. R., Block, J., & Funder, D. C. (1996). Psychometric truths in the absence of psychological meaning: A reply to Zuckerman and Knee. *Journal of Personality and Social Psychology, 70,* 1252–1255.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change"—or should we? *Psychological Bulletin, 74,* 414–417.

Edwards, J. R. (1994a). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes, 58,* 51–100.

Edwards, J. R. (1994b). Regression analysis as an alternative to difference scores. *Journal of Management, 20,* 683–689.

Edwards, J. R. (1995). Alternatives to difference scores as dependent variables in the study of congruence

in organizational research. *Organizational Behavior and Human Decision Processes, 64,* 307–324.

Johns, G. (1981). Difference score measures of organizational behavior variables: A critique. *Organizational Behavior and Human Performance, 27,* 443–463.

Judd, C. M., & Kenny, D. A. (1981). *Estimating the effects of social interventions.* Cambridge: Cambridge University Press.

Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model-comparison approach.* San Diego: Harcourt Brace Jovanovich.

Kenny, D. A. (1988). The analysis of two-person relationships. In S. Duck (Ed.), *Handbook of personal relationships* (pp. 57–77). London: Wiley.

Kenny, D. A. (1996). Models of nonindependence in dyadic research. *Journal of Social and Personal Relationships, 13,* 279–294.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113,* 181–190.

Murray, S. L., Holmes, J. G., & Griffin, D. W. (1996). The benefits of positive illusions: Idealization and the construction of satisfaction in close relationships. *Journal of Personality and Social Psychology, 70,* 79–98.

Tisak, J., & Smith, C. S. (1994). Defending and extending difference scores. *Journal of Management, 20,* 675–682.

Zimmerman, D. (1997). A geometric interpretation of the validity and reliability of difference scores. *Brit-*

ish Journal of Mathematical and Statistical Psychology, 50, 73–80.

Zuckerman, M., & Knee, C. R. (1996). A comment on Colvin, Block, and Funder (1995). *Journal of Personality and Social Psychology, 70,* 1250–1251.