JOURNAL OF PERSONALITY ASSESSMENT, *84*(1), 37–48

STATISTICAL DEVELOPMENTS AND APPLICATIONS

Conducting and Interpreting Canonical Correlation Analysis in Personality Research: A User-Friendly Primer

Alissa Sherry
Counseling Psychology Program
University of Texas at Austin

Robin K. Henson
Department of Technology and Cognition
University of North Texas

The purpose of this article is to reduce potential statistical barriers and open doors to canonical correlation analysis (CCA) for applied behavioral scientists and personality researchers. CCA was selected for discussion, as it represents the highest level of the general linear model (GLM) and can be rather easily conceptualized as a method closely linked with the more widely understood Pearson *r* correlation coefficient. An understanding of CCA can lead to a more global appreciation of other univariate and multivariate methods in the GLM. We attempt to demonstrate CCA with basic language, using technical terminology only when necessary for understanding and use of the method. We present an entire example of a CCA analysis using SPSS (Version 11.0) with personality data.

Many applied behavioral researchers are not aware that there is a general linear model (GLM) that governs most classical univariate (e.g., analysis of variance [ANOVA], regression) and multivariate (e.g., multivariate ANOVA [MANOVA], descriptive discriminant analysis) statistical methods. Accordingly, many persons view these statistical methods as separate entities rather than conceptualizing their distinct similarities within the GLM. For example, because all classical parametric analyses are part of the GLM, all of these analyses have certain things in common, including the facts that they (a) are ultimately correlational in nature, (b) yield r^2-type effect sizes, (c) maximize shared variance between variables or between sets of variables, and (d) apply weights to observed variables to create synthetic (i.e., unobserved, latent) variables that often become the focus of the analysis (cf. Bagozzi, Fornell, & Larcker, 1981; Cohen, 1968; Henson, 2000; Knapp, 1978; Thompson, 1991).

Knowledge of the commonalities among statistical analyses is in stark contrast to the often compartmentalized statistical education that many graduate students and faculty have received. Unfortunately, this compartmentalization can lead to rigidity of thought concerning the methods as opposed to a fluid understanding of their purpose and utility, thereby hindering appropriate methodological applications in applied psychological research. Indeed, at least partially because of this educational paradigm, it not uncommon to see some graduate students physically shudder at the thought of enduring advanced methodological coursework. It should not be surprising, then, to find some graduate students taking great lengths to desperately avoid methodology curricula and, in extreme cases, seeking psychotherapy to reduce the systemic anxiety these courses seem to invoke!

Statistics anxiety notwithstanding, the GLM provides a framework for understanding all classical analyses in terms of the simple Pearson *r* correlation coefficient. We demonstrate later, for example, the interpretation of a canonical correlation analysis (CCA), which has as its foundation the Pearson *r* correlation. The GLM can also be conceptualized as a hierarchal family, with CCA serving as the parent analysis. Contrary to the compartmentalized understanding of statistical methods held by many researchers, CCA subsumes both univariate and multivariate methods as special cases

(Fan, 1996, 1997; Henson, 2000; Thompson 2000). Actually, structural equation modeling represents the highest level of the GLM. However, structural equation modeling explicitly includes measurement error as part of the analysis, whereas other classical statistical procedures do not. Knowledge of the inner workings of CCA can inform researchers regarding the application of GLM concepts across analyses and extension of these concepts to vital multivariate methods (Fish, 1988).

In theory, CCA has been available to researchers since Hotelling (1935, 1936) initially developed the method's analytic framework. More recently, however, CCA has become practically available due to the advent of statistical software programs. Nevertheless, some researchers continue to use univariate statistical analyses (i.e., one dependent variable), such as multiple regression and ANOVA, to analyze data that might better be analyzed using a multivariate technique (i.e., more than one dependent variable) such as CCA.

PURPOSE

The purpose of this article is to reduce potential statistical barriers and open doors to CCA for applied behavioral scientists and personality researchers. CCA was selected for discussion, as it represents the highest level of the GLM and can be rather easily conceptualized as a method closely linked with the more widely understood Pearson *r* correlation coefficient. In addition, an understanding of CCA can lead to a more global appreciation of other univariate and multivariate methods in the GLM. We attempt to demonstrate CCA with basic language, using technical terminology only when necessary for understanding and use of the method. Readers interested in more technical, theoretical discussions of CCA are referred to Stevens (2002), Tabachnick and Fidell (1996), and Thompson (1984). We present an entire example of a CCA analysis using SPSS (Version 11.0) with personality assessment data.

ADVANTAGES OF CCA (AND OTHER MULTIVARIATE METHODS)

There are several advantages to CCA, many of which are due to the fact that CCA is a multivariate technique. First, multivariate techniques such as CCA limit the probability of committing Type I error anywhere within the study (Thompson, 1991). Risk of Type I error within a study is sometimes referred to as "experimentwise error" and relates to the likelihood of finding a statistically significant result when one should not have (e.g., finding a difference, effect, or relationship when it really does not exist in the population). Increased risk of this error occurs when too many statistical tests are performed on the same variables in a data set, with each test having its own risk of Type I error (often set by tradition at $\alpha = .05$ and sometimes called "testwise error"). Multivariate techniques minimize this because they allow for simultaneous comparisons among the variables rather than requiring many statistical tests be conducted.

For example, if a researcher wants to see if four attachment style variables can predict 10 personality disorder variables, then a series of 10 multiple regressions are required to examine each criterion variable separately. As each additional regression is run and the multiple *R* tested for statistical significance, then the experimentwise (EW) Type I error rate would increase. Assuming each hypothesis were independent and a traditional testwise (TW) error rate of .05, then the experimentwise error rate could be estimated as $\alpha_{EW} = 1 - (1 - \alpha_{TW})^k = 1 - (1 - .05)^{10} = .40$, which would be considered quite substantial even by those most tolerant of Type I errors. What's more, if a Type I error did occur, the researcher cannot identify which of the statistically significant results are errors and which reflect true relationships between the variables, thereby potentially invalidating the entire study! However, using a multivariate technique such as CCA, the relationships between the four attachment variables and the 10 personality variables could be examined simultaneously. Because only one test was performed, the risk of committing a Type I error is minimized. Of course, even with one statistical significance test at $\alpha = .05$, one still does not know for sure whether one has committed a Type I error. Nevertheless, as the experimentwise error increases, so does our confidence that a Type I error may have been committed somewhere in the study.

An extremely important second advantage of multivariate techniques such as CCA is that they may best honor the reality of psychological research. Most human behavior research typically investigates variables that possibly have multiple causes and multiple effects. Determining outcomes based on research that separately examines singular causes and effects may distort the complex reality of human behavior and cognition. Therefore, it is important to not only choose a statistical technique that is technically able to analyze the data but also a technique that is theoretically consistent with the purpose of the research. This congruence between the nature of the problem and the choice of statistical methods is particularly salient in personality research given the complexity of the constructs examined. Fish (1988) demonstrated, for example, how important multivariate relationships can be missed when data are studied with univariate methods.

Finally, and more specific to CCA, this technique can be used instead of other parametric tests in many instances, making it not only an important technique to learn but a comprehensive technique as well. As has been demonstrated by Henson (2000), Knapp (1978), and Thompson (1991), virtually all of the parametric tests most often used by behavioral scientists (e.g., ANOVA, MANOVA, multiple regression, Pearson correlation, *t* test, point-biserial correlation, discriminant analysis) can be subsumed by CCA as special cases in the GLM. This is not to say that CCA should always

be used instead of these other methods because, in many cases, this may be a long, tedious way to conduct an otherwise simple analysis. However, there are two important implications here. First, it is important to note that there are special circumstances in which CCA may be more appropriate than some of these other analytical techniques. Second, and more important, understanding that these techniques are intricately related and fundamentally the same in many respects may help facilitate conceptual understanding of statistical methods throughout the GLM.

APPROPRIATE USES AND GENERAL OVERVIEW OF CCA

CCA is most appropriate when a researcher desires to examine the relationship between two variable sets. For CCA to make theoretical sense as a multivariate analysis, there should be some rationale for why the variables are being treated together in variable sets. For example, a researcher may have four different measures of intelligence in the predictor variable set and three different measures of creativity in the criterion variable set. The research question of interest, then, would be whether there is a relationship between intelligence and creativity as multioperationalized in the variable sets. In contrast, if the researcher only had one criterion measure of creativity, then multiple regression would be conducted. If only one variable set were available (e.g., many indexes of intelligence), then the researcher may choose to conduct some sort of factor analysis to synthesize the variables. If more than one variable exists in both sets, then CCA may be the analysis needed.

Although one variable set is often identified as the predictor set and the other as the criterion set based on a researcher's expectations about predictive causality, the nature of CCA as a correlational method makes the declaration ultimately arbitrary and suggests the researcher should be cautious of making causal inferences outside of an experimental design. Because CCA examines the correlation between a synthetic criterion and synthetic predictor variable that are weighted based on the relationships between the variables within the sets, CCA can be conceptualized as a simple bivariate correlation (Pearson *r*) between the two synthetic variables.

Figure 1 illustrates the variable relationships in a hypothetical CCA with three predictor and two criterion variables. To evaluate the simultaneous relationship between several predictor and several criterion variables, the observed variables in each set must somehow be combined together into one synthetic (also called unobserved or latent) variable. These synthetic variables are created in CCA by applying a linear equation to the observed predictor variables to create a single synthetic predictor variable and another linear equation to the observed dependent variables to create a single synthetic criterion variable.

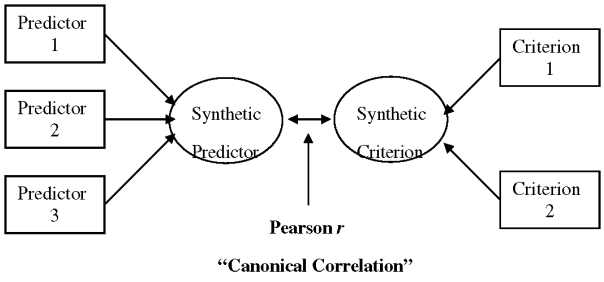

FIGURE 1 Illustration of the first function in a canonical correlation analysis with three predictors and two criterion variables. The canonical correlation is the simple Pearson *r* between the two synthetic variables, which were linearly combined from the observed variables.

This use of linear equations is directly analogous to the use of linear equations in the more familiar multiple regression in which beta (β) weights are multiplied with observed scores (in *Z* score form) and then summed to yield synthetic predicted scores (i.e., $Y' = \beta_1 X_1 + \beta_2 X_2$). Through the use of standardized weights analogous to beta weights, CCA creates two linear equations, one for the predictor variables and one for the criterion variables. These equations then yield the two synthetic variables illustrated in Figure 1.

It is important to note, however, that these two equations are generated to yield the largest possible correlation between the two synthetic variables. That is, the variance in the observed predictor variables is combined to maximally correlate with the combined variance in the observed criterion variable set. It is at this point that one can now see the most foundational component of all GLM analyses. The most central statistic in a CCA is the canonical correlation between the two synthetic variables, and this statistic is nothing more or less than a Pearson *r* (see Figure 1). Everything that occurs in the CCA is designed to maximize this simple correlation. The reader is referred to Henson (2002) and Thompson (1984) for accessible demonstrations of the equations to create the synthetic variables; and we therefore assume here that the reader is familiar with the role of linear equations as is performed by the standardized regression equation. Together, this set of equations is called a *canonical function* (or *variate*).

Furthermore, in a CCA, there will be as many canonical functions as there are variables in the smaller of the two variable sets (e.g., two functions for the example in Figure 1). As discussed, the first function creates the two synthetic variables so that they are as strongly correlated as possible given the scores on the observed variables. Unless the canonical correlation from this first function is a perfect 1.00 (which of course is not very realistic), then there will be residual variance left over in the two variable sets that cannot be explained. The second function creates two more synthetic variables that are as strongly correlated as possible given the residual variance left over after the first function and given

the condition that these new synthetic variables are perfectly uncorrelated with both of the synthetic variables in the first function. This condition is sometimes called *double orthogonality* because both synthetic variables in subsequent functions must be uncorrelated with both synthetic variables in all preceding functions. Analogous to a principal component analysis, this process repeats until either all the variance is explained from the original variables or until there are as many functions as there are variables in the smaller variable set. As we note later, however, only those functions that are able to explain a reasonable amount of the relationship between the original variable sets are considered for interpretation. This decision is analogous to a researcher only defining and interpreting the strongest factors in a factor analysis.

SOME BASIC ASSUMPTIONS OF THE CCA PROCEDURE

As with all analyses, appropriate use of CCA comes with some basic assumptions. In the interests of brevity and journal space, these assumptions will not be extensively addressed here. Tabachnick and Fidell (1996) and Thompson (1984) have provided sufficient reviews. Among CCA assumptions (e.g., sample size issues, linearity), perhaps the most important one is multivariate normality. This assumption is the multivariate analog of the univariate normality assumption and, put simply, requires that all variables and all linear combinations of variables are normally distributed. However, the evaluation of multivariate normality can be difficult. Mardia (1985) presented a statistical approach and Henson (1999) demonstrated a graphical method.

SOME IMPORTANT CCA TERMS

CCA language is important to learn and understand to interpret a CCA and subsequently write a concise results section for manuscripts. Many of the statistics in a CCA have univariate analogs, and it would be helpful if similar statistics would have similar names across analyses. Unfortunately, this is often not the case (which contributes to the compartmentalized knowledge of many regarding classical statistical methods). If it were, then graduate students would be much less confused, and we methodology professors would appear much less intelligent because others besides ourselves would happen to know the lingo! At the risk of establishing some commonalities with other analyses such as multiple regression, we present the following brief definitions of the most relevant CCA statistics. In isolation, these terms probably have limited utility; nevertheless, it is hoped that this list will help inform the CCA example to follow.

The *canonical correlation coefficient* (R_c) is the Pearson *r* relationship between the two synthetic variables on a given canonical function (see Figure 1). Because of the scaling created by the standardized weights in the linear equations, this value cannot be negative and only ranges from 0 to 1. The R_c is directly analogous to the multiple *R* in regression.

The *squared canonical correlation* (R_c^2) is the simple square of the canonical correlation. It represents the proportion of variance (i.e., variance-accounted-for effect size) shared by the two synthetic variables. Because the synthetic variables represent the observed predictor and criterion variables, the R_c^2 indicates the amount of shared variance between the variable sets. It is directly analogous to the R^2 effect in multiple regression.

A *canonical function* (or *variate*) is a set of standardized canonical function coefficients (from two linear equations) for the observed predictor and criterion variable sets. There will be as many functions as there are variables in the smaller variable set. Each function is orthogonal to every other function, which means that each set of synthetic predictor and criterion variables will be perfectly uncorrelated with all other synthetic predictor and criterion variables from other functions. Because of this orthogonality, the functions are analogous to components in a principal component analysis. A single function would be comparable to the set of standardized weights found in multiple regression (albeit only for the predictor variables). This orthogonality is convenient because it allows one to separately interpret each function.

Standardized canonical function coefficients are the standardized coefficients used in the linear equations discussed previously to combine the observed predictor and criterion variables into two respective synthetic variables. These weights are applied to the observed scores in *Z*-score form (thus the standardized name) to yield the synthetic scores, which are then in turn correlated to yield the canonical correlation. The weights are derived to maximize this canonical correlation, and they are directly analogous to beta weights in regression.

A *structure coefficient* (r_s) is the bivariate correlation between an observed variable and a synthetic variable. In CCA, it is the Pearson *r* between an observed variable (e.g., a predictor variable) and the canonical function scores for the variable's set (e.g., the synthetic variable created from all the predictor variables via the linear equation). Because structure coefficients are simply Pearson *r* statistics, they may range from –1 to +1, inclusive. They inform interpretation by helping to define the structure of the synthetic variable, that is, what observed variables can be useful in creating the synthetic variable and therefore may be useful in the model. These coefficients are analogous to those structure coefficients found in a factor analysis structure matrix or in a multiple regression as the correlation between a predictor and the predicted Y' scores (Courville & Thompson, 2001; Henson, 2002).

Squared canonical structure coefficients (r_s^2) are the square of the structure coefficients. This statistic is analogous to any other r^2-type effect size and indicates the proportion of variance an observed variable linearly shares with the synthetic variable generated from the observed variable's set.

A *canonical communality coefficient* (h^2) is the proportion of variance in each variable that is explained by the complete canonical solution or at least across all the canonical

functions that are interpreted. It is computed simply as the sum of the r_s^2 across all functions that are interpreted for a given analysis. This statistic informs one about how useful the observed variable was for the entire analysis.

EXAMPLE AND (STEP-BY-STEP) INTERPRETATION OF CCA

In this section, we detail how to run and interpret a CCA from example SPSS output (other program outputs would be similar). Our style is purposefully highly applied, and our intent is to provide a step-by-step guide for researchers and others seeking an initial exposure to the method for use. Hopefully, this context will allow the reader to gain increased understanding of the statistics discussed previously and a framework for more theoretical study.

The data used here were taken from Sherry, Lyddon, and Henson's (2004) study of the relationship between adult attachment variables and adult personality style. The basic question of this study asked whether adult attachment variables (theoretically presumed to be formed in the very early years of life) are predictive of certain personality styles (theoretically presumed to lie on a continuum as opposed to a purely diagnostic perspective). The predictor variable set contained four measures representing the dimensions of Bartholomew's adult attachment theory as assessed by the Relationship Scales Questionnaire (RSQ; 30 items on a 5-point scale; cf. Griffin & Bartholomew, 1994). These predictor variables were secure, dismissing, fearful, and preoccupied attachment. The criterion variable set contained personality variables as measured by the Millon Clinical Multiaxial Inventory–III (MCMI–III; Millon, Davis, & Millon, 1997). The scales that relate to the 10 personality disorders recognized in the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; American Psychiatric Association, 1994) were used (raw scores), which included the Schizoid, Avoidant, Dependent, Histrionic, Narcissistic, Antisocial, Compulsive, Schizotypal, Borderline, and Paranoid personality scales. The participants included 269 undergraduate students recruited from three different universities located in the South central, Southeastern, and Pacific Northwestern regions of the United States.

Unfortunately, there is no "point-and-click" option in SPSS for CCA. However, creating some short computer commands (syntax) allows one to easily conduct the analysis. Simply click the File, New, Syntax sequence and then type the following syntax in the window provided.

```
MANOVA
schizdr avoidr dependr histrior narcisr antisocr compulsr
schtypr borderr parandr WITH secure dismiss fearful
preoccup
/PRINT=SIGNIF(MULTIV EIGEN DIMENR)
/DISCRIM=(STAN ESTIM COR ALPHA(.999)).
```

This syntax will remain the same for any CCA except of course for changing the variable names to match other data. The criterion set of variables are listed before the WITH and the predictor variables are listed afterward. The commands can be implemented using the RUN menu on the toolbar. An abbreviated output is presented in Appendix A.

A General Framework for Interpreting GLM Analyses

In many explanatory research contexts, it is often important to identify variables that contribute to the model being tested. For example, in this example, we not only care about whether there is a relationship between the predictor and criterion variable sets, but we also want to know what attachment and personality variables are more or less useful in the model and whether they relate to each other in expected directions. Identification of variable importance, then, is fundamental to many of the analyses we conduct.

Furthermore, within the GLM, all analyses yield r^2-type effect sizes that must be considered prior to evaluating what variables contributed to this effect. It makes no sense, for example, to have a minuscule (and uninterpretable) effect size and yet try to identify variables that contributed to that effect! Accordingly, Thompson (1997) articulated a two-stage hierarchal decision strategy that can be used to interpret any GLM analysis:

> All analyses are part of one general linear model. … When interpreting results in the context of this model, researchers should generally approach the analysis hierarchically, by asking two questions:
>
> Do I have anything? (Researchers decide this question by looking at some combination of statistical significance tests, effect sizes … and replicability evidence.)
>
> If I have something, where do my effects originate? (Researchers often consult both the standardized weights implicit in all analyses and structure coefficients to decide this question.). (p. 31)

Once notable effects have been isolated, then (and only then) interpretation shifts to the identification of what variables in the model may have contributed to that effect. The weights (often standardized) present in all GLM analyses are typically examined to judge the contribution of a variable to the effect observed. For example, within regression, many researchers may discount the value of a variable with a small or near-zero β weight. This hierarchal strategy is employed in the following to help frame the interpretation of the CCA.

Do You Have Anything?

The researcher initially needs to determine whether the canonical model sufficiently captures the relationship between the predictor and criterion variable sets to warrant interpretation. That is, does a noteworthy relationship between the variables exist? As noted previously, there are several ways

to evaluate a possible answer to this question, but we delimit ourselves to the most common approaches of statistical significance testing and effect size interpretation.

Step 1. The initial consideration is to evaluate the full canonical model. The first portion of Appendix A presents four ways to evaluate for statistical significance with multivariate tests. These test statistics are for the full model, which means they evaluate the shared variance between the predictor and criterion variables across all of the canonical functions. Each test statistic can be converted to the more familiar *F* statistic, which can then be evaluated for statistical significance. Of importance, because each of the four methods is based on somewhat different theoretical frameworks, each can lead to different conclusions. The astute reader will note, for example, that the approximate *F* statistics in Appendix A were all slightly different. Furthermore, in this particular case, one of the methods (Roy's) did not even yield a result due to some limits to this approach.

Nevertheless, by far the most common method used is Wilks's lambda (λ), as it tends to have the most general applicability. In our example, the full model was statistically significant, with a Wilks's λ of .439, $F(40, 968.79) = 5.870$, $p < .001$. (Note that the column in Appendix A labeled "Significance of *F*" presents the *p* value associated with the probability of the sample results assuming the null hypothesis is exactly true in the population given the sample size. Because the *p* value is rounded to three decimal places, we can only note that $p < .001$ in this case.) Accordingly, we can reject the null hypothesis that there was no relationship between the variable sets (i.e., reject $R_c = 0$) and conclude that there probably was a relationship.

Of course, this statistical significance test tells us absolutely nothing about the magnitude of the relationship, which is one limitation of such tests about which increasing numbers of researchers are becoming aware (Wilkinson & APA Task Force on Statistical Inference, 1999). As a bit of a caveat, statistical significance tests are impacted rather heavily by sample size, and it is very possible, with large enough sample sizes, to get statistically significant outcomes for very small, unimportant effects. Therefore, it is important to interpret effect size indexes (and perhaps other information, such as confidence intervals) alongside *p* values to determine the practical significance of study outcomes. The interested reader is referred to Harlow, Mulaik, and Steiger (1997) for discussion of the debate surrounding statistical significance tests.

Conveniently, Wilks's λ has a useful property that helps inform this issue because it represents something of an inverse effect size or the amount of variance not shared between the variable sets. Therefore, by taking $1 - \lambda$, we found an overall effect of $1 - .439 = .561 = R_c^2$ for the full model. This effect statistic can be interpreted just like the multiple R^2 in regression as the proportion of variance shared between the variable sets across all functions. Thus far, then, we have noted that the full model was both statistically significant and had what may be considered a large effect size.

Step 2. Of course, it would be too easy if we only had to evaluate the full canonical model to decide if we had anything. Instead, we need to dig a bit deeper and evaluate each canonical function. Remember that there will be as many functions (i.e., variates) as there are variables in the smaller set, which in this case is four (the predictor set). Each function must be evaluated because some of them may not explain enough of the relationship between the variable sets to warrant interpretation, much like a weak or poorly defined factor would be discarded. Furthermore, it is possible that the full model appears noteworthy at the cumulative level, but examination of each function reveals each of them to be weak and not interpretable in and of themselves. For example, each function may not contribute much to the total solution, but the cumulative total solution may be statistically significant and perhaps noteworthy. In such cases, interpretation of each function separately would be questionable.

The next section of the Appendix A output lists each function separately along with its canonical correlation. (Note that the term *root* is equivalent to *function* in this output.) Recall that the first function will be created to maximize the Pearson *r* (canonical correlation) between the two synthetic variables. Then, using the remaining variance in the observed variables, the next function will be created to maximize another Pearson *r* (the second canonical correlation) between two other synthetic variables under the condition that these new synthetic variables are perfectly uncorrelated with all others preceding them. For this example, this continued until four orthogonal (i.e., uncorrelated) functions were created.

The CCA researcher should only interpret those functions that explain a reasonable amount of variance between the variable sets or risk interpreting an effect that may not be noteworthy or replicable in future studies. In our example, we chose to interpret the first two functions, as they explained 38.1% and 20.0% of the variance within their functions, respectively. Note that these numbers are the squared canonical correlations in Appendix A. Note as well that this means we have decided that the third and fourth functions, which each explained less than 10% of the variance in their functions (9.6% and 1.9%, respectively), were sufficiently weak so as to not warrant interpretation.

The highly observant reader may notice that the sum of the squared canonical correlations (38.1% + 20.0% = 58.1%) for just the first two functions was larger than the overall effect size we found from the Wilks's λ (56.1%). This, of course, begs the question of how the variance explained by the full model can be less than that explained by its parts! The answer to this question lies in the orthogonal nature of the functions. Recall that the second function is created after the first has explained as much of the variability in the observed variable sets as possible. Also recall that the second function must be

orthogonal to the first function. This means that the second function is not explaining the original observed variance. Instead, it is explaining what is left over, and it may explain a fairly large amount (say, 20.0% as in our example) of this left over variance. Thus, the sum of R_c^2 the effect sizes for each function will often be larger than the full model effect.

Step 3. Those readers with a penchant for using statistical significance tests to evaluate results may be wondering why we did not just test each function's canonical correlation for statistical significance to decide whether the function should be interpreted. There are two reasons this may be problematic as a sole consideration. First, the dependent relationship between statistical significance tests and sample size has been well documented, and increasing numbers of researchers are realizing that even small, nonmeaningful effects can be statistically significant at some sufficiently large sample size (see, e.g., Cohen, 1994; Henson & Smith, 2000; Thompson, 1996; Wainer & Robinson, 2003; Wilkinson & APA Task Force on Statistical Inference, 1999). Given that multivariate analyses such as CCA are generally large sample techniques, one must be careful of not overinterpreting results that may be statistically but not practically significant.

Second, and more important, there is no easy way to directly test each function separately for statistical significance. Instead, the functions are tested in hierarchal fashion in which the full model (Functions 1 to 4) is tested first, then Functions 2 to 4 are tested and so forth until only the last function is tested by itself. Because the final functions in a CCA are often weak and uninterpretable anyway, the statistical significance test of the final function is often uninformative. (Of course, if the last function were statistically significant, then one could infer that all functions preceding it were as well.) The third section of Appendix A lists the dimension reduction analysis in which these hierarchal statistical significance tests are presented. Unfortunately, it is a common error in reports of CCA to assume that the 1 to 4 test evaluates the first function, the 2 to 4 test evaluates the second function, and so forth.

For example, Sciarra and Gushue (2003) conducted a CCA between six racial attitude variables and four religious orientation variables. Sciarra and Gushue (2003) reported that

> Assumptions regarding multivariate normality were met, and four pairs of variates [i.e., functions] were generated from the data. A dimension reduction analysis showed the first three of these to be [statistically] significant, with Wilks's lambdas of .69 ($p < .01$), .84 ($p < .01$), and .92 ($p < .02$), respectively. The canonical correlations for the three pairs were .43, .28, and .24, respectively. (p. 478)

Note that this quote implies that all three functions are statistically significant in and of themselves. However, it is entirely possible that the third function is not, given that the Wilks's lambda presented (.92, $p < .02$) is actually a cumulative effect from Functions 3 and 4. If 3 were able to be isolated, the effect would be likely be smaller, and the p value would be larger, perhaps even larger than a traditional $\alpha = .05$.

Returning now to our example, Appendix A presents the dimension reduction analysis in which the hierarchal statistical significance tests are presented. Here we see that the full model was statistically significant (but we already knew that) as well as the cumulative effects of Functions 2 to 4 and 3 to 4. Function 4 was not statistically significant in isolation. Even though functions 3 to 4 were cumulatively statistically significant, we have chosen not to interpret either one, as they only explained 9.6% and 1.9%, respectively, of the variance by themselves (see R_c^2 for each function). When one considers that these R_c^2 actually represent less than 10% of the remaining variance after that explained by Functions 1 and 2, then the effect sizes of Functions 3 and 4 become even a bit less impressive.

Summary. In this example then, we have thus far concluded that there indeed was a noteworthy relationship between our variables sets by evidence of statistical significance and effect sizes. Furthermore, this relationship was largely captured by the first two functions in the canonical model.

Where Does the Effect Come From?

Because we have established that we have something, we can turn now to the second question in our interpretation strategy. That is, what variables are contributing to this relationship between the variables sets across the two functions? Identification of the contributing variables can be critical to informing theory. In our example, we want to know (in terms of degree and directionality) what attachment variables were related to what personality variables in this multivariate analysis.

Traditionally, researchers have examined the weights inherent in all GLM analyses to help answer this second question. In regression, beta weights are often consulted. Beta weights reflect the relative contribution of one predictor to the criterion given the contribution of other predictors. Unfortunately, researchers have less often consulted structure coefficients, which reflect the direct contribution of one predictor to the predictor criterion variable regardless of other predictors. This neglect occurs in spite of the fact that these coefficients can be critical in the presence of multicollinearity, which is jargon for when you have correlated predictor variables in a regression analysis (Courville & Thompson, 2001). In multivariate analyses, structure coefficients are more often consulted, such as when a factor analyst reports the structure matrix for correlated factors.

Indeed, structure coefficients increase in importance when the observed variables in the model increase in their correlation with each other. Because multivariate researchers can purposefully use variables that are related (we did, after

all, select variables that can be logically grouped into sets for our CCA), structure coefficients are critical for deciding what variables are useful for the model. (Readers unfamiliar with structure coefficients are strongly encouraged to review Courville & Thompson, 2001, for a demonstration of structure coefficients in the context of regression.) We therefore assume that interpretation of both standardized weights and structure coefficients are necessary for understanding variable importance in a CCA.

Step 4. We first examine the standardized weights and structure coefficients to interpret the first function. Appendix A presents the weights and structure coefficients for the criterion (called "Dependent") and predictor (called "Covariates") variables for all four functions. Of course we are only concerned with the first two functions and will ignore the last two.

At this point, it is quite useful to create a table of these coefficients to help us understand the patterns among our variables. Table 1 represents our recommended method for reporting CCA results and, for this example, presents the standardized canonical function coefficients (i.e., the weights) and structure coefficients for all variables across both functions. The squared structure coefficients (r_s^2) are also given, which represent the percentage of shared variance between the observed variable and the synthetic variable created from the observed variable's set. The last column lists the communality coefficients (h^2), which represent the amount of variance in the observed variable that was reproducible across the functions. Note that these are simply the sum of the variable's r_s^2s . The communalities are analogous to communality coefficients in factor analysis and can be viewed as an indication of how useful the variable was for the solution. For emphasis, structure coefficients above .45 are underlined in Table 1 (following a convention in many factor analyses). Communalities above 45% are also underlined to show the variables with the highest level of usefulness in the model.

Looking at the Function 1 coefficients, we see that relevant criterion variables were primarily avoidant, dependent, borderline, and paranoid, with histrionic, schizotypal, and schizoid having made secondary contributions to the synthetic criterion variable. This conclusion was supported mainly by the squared structure coefficients, which indicated the amount of variance the observed variable can contribute to the synthetic criterion variable. The canonical function coefficients were also consulted, and these personality styles tended to have the larger coefficients. A slight exception involves the borderline and paranoid personality styles, which had modest function coefficients but large structure coefficients. This result is due to the multicollinearity that these two variables had with the other criterion variables. In essence, the linear equation that used the standardized coefficients to combine the criterion variables (on Function 1) only modestly incorporated the variance of the borderline and paranoid variables when, in fact, these variables could have contributed substantially to the created synthetic variable (as shown by the r_s and r_s^2). Notice as well that with the exception of histrionic, all of these variables' structure coefficients had the same sign, indicating that they were all positively related. Histrionic was inversely related to the other personality styles.

The other side of the equation on Function 1 involves the predictor set. The Table 1 results inform us that the secure and preoccupied attachment variables were the primary contributors to the predictor synthetic variable, with a secondary contribution by fearful. Because the structure coefficient for secure was positive, it was negatively related to all of the personality styles except for histrionic. Preoccupied and fearful

TABLE 1
Canonical Solution for Attachment Predicting Personality for Functions 1 and 2

	Function 1			*Function 2*			
Variable	*Coef*	r_s	r_s^2 (%)	*Coef*	r_s	r_s^2 (%)	h^2 (%)
Schizoid	.427	–.454	20.61	.627	.713	50.84	71.45
Avoidant	–.467	–.806	64.96	.074	.356	12.67	77.63
Dependent	–.442	–.782	61.15	–.797	–.394	15.52	76.67
Histrionic	.494	.583	33.99	–.136	–.574	32.95	66.94
Narcissistic	–.298	.294	8.64	–.081	–.104	1.08	9.72
Antisocial	–.070	–.280	7.84	–.193	–.129	1.66	9.50
Compulsive	–.163	.061	0.37	.224	.332	11.02	11.39
Schizotypal	.224	–.542	29.38	.082	.272	7.40	36.78
Borderline	–.340	–.677	45.83	.297	.022	0.05	45.88
Paranoid	–.234	–.651	42.38	–.004	.290	8.41	50.79
R_c^2			38.10			20.00	
Secure	.578	.720	51.84	–.208	–.356	12.67	64.51
Dismissing	–.163	–.100	1.00	.550	.801	64.16	65.16
Fearful	–.226	–.477	22.75	.353	.396	15.68	38.43
Preoccupied	–.664	–.693	48.02	–.538	–.642	41.22	89.24

Note. Structure coefficients (r_s) greater than |.45| are underlined. Communality coefficients (h^2) greater than 45% are underlined. Coef = standardized canonical function coefficient; r_s = structure coefficient; r_s^2 = squared structure coefficient; h^2 = communality coefficient.

attachment were positively related to the personality disorders, again except for histrionic.

These results are generally supportive of the theoretically expected relationships between adaptive and maladaptive adult attachment and personality disorders. Note that the relevant personality disorders tended to involve social apprehension and negative symptomology at a general level, with the exception of histrionic. Because the histrionic personality disorder is marked with excessive emotionality and attention seeking, it seems theoretically consistent that it should have been negatively related to the other relevant disorders in this function. Therefore, this function seems to capture theoretically consistent relationships that we may collectively call "attachment and social apprehension." Note that this process for interpreting a function is directly analogous to identifying the useful predictors in a regression or interpreting and naming a factor, with the exception that the CCA has two equations that one must consider.

Step 5. Moving on to Function 2, the coefficients in Table 1 suggest that the only criterion variables of relevance were schizoid and histrionic, albeit less so for the latter. These personality styles were inversely related on this function. As for attachment, dismissing was the dominant predictor, along with preoccupied again. These attachment variables were also inversely related. Looking at the structure coefficients for the entire function, we see that dismissing was positively related to schizoid and negatively related to histrionic. Preoccupied attachment had the opposite pattern. Given that the dismissing and preoccupied predictors and schizoid criterion variable were the dominant contributors, we collectively label this function as "social detachment," given the nature of these variables. In cases in which the researcher has additional noteworthy functions, the previous process would simply be repeated.

Summary

The complexity of a CCA analysis is perhaps justified given the richness of the relationships it intends to model. In this example, the first function demonstrated theoretically consistent relationships among all of the variables that contributed to the function. The Function 1 results also point to a need for further study regarding the histrionic variable. For example, it may be important to examine the various defense mechanisms used in the presentation of this style versus other styles. Perhaps the histrionic personality style is so dominated by defense mechanisms that on measures such as the RSQ, which primarily rely on self-report of one's internal affective experience, people with histrionic personality features report as securely attached.

The second function also yielded theoretically expected relationships; however, this function capitalized on variance in the dismissing predictor that was not useful in the first function. Therefore, not only do we learn about relationships between attachment and personality, we also learn that dismissing attachment is something of a different animal than the other attachment variables. Additional work is needed to further explicate this possibility.

We also learn a good deal from the variables not (or only moderately) useful in the model. For example, the fearful predictor only made a marginal contribution as a predictor (see the fearful h^2 in Table 1), thereby suggesting that it may not have been strongly related to personality style. Furthermore, the narcissistic, antisocial, and compulsive personality styles did not appear to be related to attachment (see the h^2 statistics in Table 1). This is informative, particularly given the general disregard for some social norms that these disorders typify and the general sense of social apprehension that characterizes the other disorders (again, with the exception of histrionic, which represents something of the opposite of social apprehension).

Writing Up the Results

Perhaps one of the most challenging aspects of employing a newly learned method in research is actually writing up the results in a format appropriate for the journal article or dissertation. In light of this, we present in Appendix B a brief sample write-up of these findings. This narrative may serve as a guide for others seeking to use CCA in their research, although it is recognized that other writing styles are certainly possible and that other researchers may choose to emphasize differing elements of the findings.

CONCLUSIONS

This article was meant to be a practical introduction to CCA. However, it should be noted that our brief discussion is not meant as a detour around the needed quantitative foundations of human behavior research for full understanding of CCA and related analyses. Advanced quantitative coursework notwithstanding, it is our social cognitive theory position that learning requires some sense of self-efficacy in one's ability to acquire and utilize new information. This self-efficacy is often best developed with mastery experiences occurring within reach of one's already possessed skill set. Compartmentalized statistical education that does not seek to establish links and conceptual understanding among analyses is unfortunately not conducive to this goal. As such, we applaud the *Journal of Personality Assessment*'s creation of the "Statistical Developments and Applications" section in which methodological issues can be addressed from a practical and comprehensible manner for graduate students and applied researchers. As many readers know, other journals have created similar sections with outstanding results.

It is hoped that this article demonstrates the utility of CCA for some personality research. Our example was drawn from a substantive study, but CCA's flexibility in the GLM allows

it to be employed in a variety of applications such as, for example, multivariate, criterion-related validity studies. Furthermore, like all GLM analyses, the nature of CCA as a fundamentally correlational technique enhances its accessibility. Almost all of the previous discussion hinges on Pearson r or r^2-type statistics; what changes from analysis to analysis are the variables being related and the language used to discuss it all.

REFERENCES

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Bagozzi, R. P., Fornell, C., & Larcker, D. F. (1981). Canonical correlation analysis as a special case of a structural relations model. *Multivariate Behavioral Research, 16,* 437–454.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin, 70,* 426–443.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49,* 997–1003.

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement, 61,* 229–248.

Fan, X. (1996). Canonical correlation analysis as a general analytic model. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 4, pp. 71–94). Greenwich, CT: JAI.

Fan, X. (1997). Canonical correlation analysis and structural equation modeling: What do they have in common? *Structural Equation Modeling, 4,* 65–79.

Fish, L. (1988). Why multivariate methods are usually vital. *Measurement and Evaluation in Counseling and Development, 21,* 130–137.

Griffin, D., & Bartholomew, K. (1994). Models of the self and other: Fundamental dimensions underlying measures of adult attachment. *Journal of Personality and Social Psychology, 67,* 430–445.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Henson, R. K. (1999). Multivariate normality: What is it and how is it assessed? In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 193–211). Stamford, CT: JAI.

Henson, R. K. (2000). Demystifying parametric analyses: Illustrating canonical correlation as the multivariate general linear model. *Multiple Linear Regression Viewpoints, 26*(1), 11–19.

Henson, R. K. (2002, April). *The logic and interpretation of structure coefficients in multivariate general linear model analyses.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Henson, R. K., & Smith, A. D. (2000). State of the art in statistical significance and effect size reporting: A review of the APA Task Force report and current trends. *Journal of Research and Development in Education, 33,* 284–294.

Hotelling, H. (1935). The most predictable criterion. *Journal of Educational Psychology, 26,* 139–142.

Hotelling, H. (1936). Relations between two sets of variables. *Biometrika, 28,* 321–377.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. *Psychological Bulletin, 85,* 410–416.

Mardia, K. V. (1985). Mardia's test of multinormality. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (Vol. 5, pp. 217–221). New York: Wiley.

Millon, T., Davis, R., & Millon, C. (1997). *MCMI–III manual* (2nd ed.). Minneapolis, MN: National Computer Systems.

Sciarra, D. T., & Gushue, G. V. (2003). White racial identity development and religious orientation. *Journal of Counseling and Development, 81,* 473–482.

Sherry, A., Lyddon, W. J., & Henson, R. K. (2004). *Adult attachment and developmental personality styles: An empirical study.* Manuscript submitted for publication.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins.

Thompson, B. (1984). *Canonical correlation analysis: Uses and interpretation.* Newbury Park, CA: Sage.

Thompson, B. (1991). A primer on the logic and use of canonical correlation analysis. *Measurement and Evaluation in Counseling and Development, 24,* 80–95.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher, 25*(2), 26–30.

Thompson, B. (1997). Editorial policies regarding statistical significance tests: Further comments. *Educational Researcher, 26*(5), 29–32.

Thompson, B. (2000). Canonical correlation analysis. In L. Grimm & P. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 207–226). Washington, DC: American Psychological Association.

Wainer, H., & Robinson, D. H. (2003). Shaping up the practice of null hypothesis significance testing. *Educational Researcher 32*(7), 22–30.

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594–604.

APPENDIX A

This appendix includes an abbreviated SPSS output for the CCA example. Entries in the following prefaced with "Note" were added to help clarify the portions of the output being referenced in the article discussion. For the sake of brevity, elements of the original output that were not specifically salient to interpreting the CCA were deleted, such as univariate results for each dependent variable.

Note: Statistical Significance Tests for the Full CCA Model
Effect … Within Cells Regression Multivariate Tests of Significance ($S = 4$, $M = 2½$, $N = 126½$)

Test Name	*Value*	*Approximate F*	*Hypothesis DF*	*Error DF*	*Significance of F*
Pillais's	.69574	5.43238	40.00	1032.00	.000
Hotelling's	.99148	6.28351	40.00	1014.00	.000
Wilks's	.43916	5.86990	40.00	968.79	.000
Roy's	.38145				

Note: Canonical Correlations for Each Function Separately
Eigenvalues and Canonical Correlations

Root No.	*Eigenvalue*	*%*	*Cumulative %*	*Canonical Correlation*	*Squared Correlation*
1	.617	62.198	62.198	.618	.381
2	.250	25.225	87.423	.447	.200
3	.106	10.673	98.096	.309	.096
4	.019	1.904	100.000	.136	.019

Note: Hierarchal Statistical Significance Tests In Which Only the Last Canonical Function Is Tested Separately
Dimension Reduction Analysis

Roots	*Wilks λ*	*F*	*Hypothesis DF*	*Error DF*	*Significance of F*
1 to 4	.43916	5.86990	40.00	968.79	.000
2 to 4	.70998	3.44861	27.00	748.29	.000
3 to 4	.88755	1.97446	16.00	514.00	.013
4 to 4	.98147	0.69595	7.00	258.00	.675

Note: Standardized Weights for All Functions for the Criterion Variable Set
Standardized Canonical Coefficients for Dependent Variables

	Function No.			
Variable	*1*	*2*	*3*	*4*
SCHIZDR	.427	.627	0.983	.929
AVOIDR	–.467	.074	0.371	.591
DEPENDR	–.442	–.797	0.071	.463
HISTRIOR	.494	–.136	1.160	.484
NARCISR	–.298	–.081	0.133	.015
ANTISOCR	–.070	–.193	0.154	–.634
COMPULSR	–.163	.224	0.553	–.384
SCHTYPR	.224	.082	–0.529	–.380
BORDERR	–.340	.297	0.821	–.044
PARANDR	–.234	–.004	–0.559	–.897

Note: Structure Coefficients for All Functions for the Criterion Variable Set
Correlations Between Dependent and Canonical Variables

	Function No.			
Variable	*1*	*2*	*3*	*4*
SCHIZDR	–.454	.713	.112	.146
AVOIDR	–.806	.356	–.174	.218
DEPENDR	–.782	–.394	.042	.271
HISTRIOR	.583	–.574	.400	–.295
NARCISR	.294	–.104	.448	–.495
ANTISOCR	–.280	–.129	.303	–.460
COMPULSR	.061	.332	.081	.096
SCHTYPR	–.542	.272	–.054	–.202
BORDERR	–.677	.022	.375	–.226
PARANDR	–.651	.290	–.002	–.467

Note: Standardized Weights for All Functions for the Predictor Variable Set
Standardized Canonical Coefficients for Covariates

	Canonical Variable			
Covariate	*1*	*2*	*3*	*4*
Secure	.578	–.208	.822	.240
Dismissing	–.163	.550	.713	–.595
Fearful	–.226	.353	.244	.927
Preoccupied	–.664	–.538	.584	–.303

Note: Structure Coefficients for All Functions for the Predictor Variable Set
Correlations Between Covariates and Canonical Variables

	Canonical Variable			
Covariate	*1*	*2*	*3*	*4*
Secure	.720	–.356	.573	.160
Dismissing	–.100	.801	.399	–.435
Fearful	–.477	.396	.215	.755
Preoccupied	–.693	–.642	.328	–.011

APPENDIX B

Sample Write-Up of the Results

A canonical correlation analysis was conducted using the four attachment variables as predictors of the 10 personality variables to evaluate the multivariate shared relationship between the two variable sets (i.e., adult attachment and personality). The analysis yielded four functions with squared canonical correlations (R_c^2) of .381, .200, .096, and .019 for each successive function. Collectively, the full model across all functions was statistically significant using the Wilks's $\lambda = .439$ criterion, $F(40, 968.79) = 5.870$, $p < .001$. Because Wilks's λ represents the variance unexplained by the model, $1 - \lambda$ yields the full model effect size in an r^2 metric. Thus, for the set of four canonical functions, the r^2 type effect size was .561, which indicates that the full model explained a substantial portion, about 56%, of the variance shared between the variable sets.

The dimension reduction analysis allows the researcher to test the hierarchal arrangement of functions for statistical significance. As noted, the full model (Functions 1 to 4) was statistically significant. Functions 2 to 4 and 3 to 4 were also statistically significant, $F(27, 748.29) = 3.449$, $p < .001$, and $F(16, 514) = 1.974$, $p = .013$, respectively. Function 4 (which is the only function that was tested in isolation) did not explain a statistically significant amount of shared variance between the variable sets, $F(7, 258) = .696$, $p = .675$.

Given the R_c^2 effects for each function, only the first two functions were considered noteworthy in the context of this study (38.1% and 20% of shared variance, respectively). The last two functions only explained 9.6% and 1.9%, respectively, of the remaining variance in the variable sets after the extraction of the prior functions.

Table 1 presents the standardized canonical function coefficients and structure coefficients for Functions 1 and 2. The squared structure coefficients are also given as well as the communalities (h^2) across the two functions for each variable. Looking at the Function 1 coefficients, one sees that relevant criterion variables were primarily avoidant, dependent, borderline, and paranoid, with histrionic, schizotypal, and schizoid making secondary contributions to the synthetic criterion variable. This conclusion was supported by the squared structure coefficients. These personality styles also tended to have the larger canonical function coefficients. A slight exception involved the borderline and paranoid personality styles, which had modest function coefficients but large structure coefficients. This result was due to the multicollinearity that these two variables had with the other criterion variables. Furthermore, with the exception of histrionic, all of these variables' structure coefficients had the same sign, indicating that they were all positively related. Histrionic was inversely related to the other personality styles.

Regarding the predictor variable set in Function 1, secure and preoccupied attachment variables were the primary contributors to the predictor synthetic variable, with a secondary contribution by Fearful. Because the structure coefficient for secure was positive, it was negatively related to all of the personality styles except for histrionic. Preoccupied and fearful attachment were positively related to the personality disorders, again except for histrionic. These results were generally supportive of the theoretically expected relationships between adaptive and maladaptive adult attachment and personality disorders, and we labeled Function 1 as "attachment and social apprehension" (for rationale, see Discussion section).

Moving to Function 2, the coefficients in Table 1 suggest that the only criterion variables of relevance were schizoid and histrionic, albeit less so for the latter. These personality styles were inversely related on this function. As for attachment, dismissing was now the dominant predictor, along with Preoccupied again. These attachment variables were also inversely related. Looking at the structure coefficients for the entire function, we see that dismissing was positively related to schizoid and negatively related to histrionic. Preoccupied attachment had the opposite pattern. Given the nature of these variables, we labeled this function as "social detachment" (for rationale, see Discussion section).

Robin K. Henson
University of North Texas
Department of Technology and Cognition
P.O. Box 311335
Denton, TX 76203
Email at rhenson@unt.edu

Received February 20, 2004
Revised April 21, 2004