

ASSORTATIVE MATING IN THE JEWEL WASP: 2. SEQUENTIAL CANONICAL ANALYSIS AS AN EXPLORATORY FORM OF PATH ANALYSIS

AURELIO JOSÉ FIGUEREDO, Department of Psychology, University of Arizona, Tucson, AZ 85721; and
RICHARD L. GORSUCH, Graduate School of Psychology, Fuller Theological Seminary, Pasadena, CA
91182

ABSTRACT

This paper discusses the conceptual and mathematical relationships between three statistical models: 1) traditional, or “Simultaneous”, Canonical Analysis, 2) hierarchical, or “Sequential”, Canonical Analysis, and 3) Structural Equations Modeling, or confirmatory path analysis. The advantages of Sequential over Simultaneous Canonical Analysis are reviewed. The relationships between Sequential Canonical Analysis and Path Analysis are explored. The need for a legitimate exploratory form of path analysis, analogous to existing exploratory forms of both multiple regression and factor analysis, is discussed. A logical extension of Sequential Canonical Analysis is proposed as adequately serving the function of an exploratory path analysis. Empirical data from psychological research is used to illustrate and qualitatively compare and contrast the results of these three approaches.

INTRODUCTION

It is often the case that the multiple outcomes of an experimental treatment or manipulation need to be assessed. Moreover, these dependent variables are often selected to measure the impact of the treatment on several conceptually distinct outcomes, rather than converge upon a single construct. Thus, multivariate data reduction methods, such as common factor modeling, may not be appropriate. On the other hand, because the intervention may exert a common causal influence on these outcomes, the dependent variables are likely to be at least spuriously correlated with each other. In addition, the multiple outcomes may also subsequently exert various causal influences on each other. Thus, separate causal analyses for each of these dependent variables may also not be appropriate.

Fortunately, there exist several statistical procedures for the analysis of multiple correlated dependent variables. One of these is structural equations modeling (SEM), or confirmatory path analysis, in which the hypothesized causal network between outcomes can be fully specified, estimated, and tested (Bentler 1989). This method, however, requires the guidance of a strong causal theory which is often not available. Another multivariate method is simultaneous canonical analysis (SIMCA), which requires little theory. This method, however, produces empirically-derived linear composites of dependent variables, or “canonical variates”, which are often difficult to interpret pragmatically in terms of the concrete outcomes that we are interested in. The linear composites are statistically defined, but are dependent upon the exact data set and so may not be replicated if the next study uses a slightly different set of variables.

A third alternative is sequential canonical analysis (SEQCA), which combines some of the advantages of both (Gorsuch 1991, Gorsuch and Figueredo 1991). Rather than combining all dependent variables into uninterpretable linear composites, as a simultaneous regression does for independent variables, it partitions their covariance sequentially, as a hierarchical regression does, while maintaining their separate identity. This method isolates the direct effects of the independent variables, or interventions, sequentially on each of the dependent variables, or outcomes, controlling for all indirect effects through the prior dependent variables, or outcomes. Because the only theoretical guidance required is a tentative specification of a meaningful order among the dependent variables, we propose that this sequential method can be developed into an exploratory form of path analysis.

The purpose of this paper is to explore the properties of sequential canonical analysis both in relation to those of confirmatory path analysis and simultaneous canonical analysis, and as a potential model for exploratory path analysis. To do this, we will use an example from insect psychobiology which has the advantage of clearly illustrating these principles with very concrete, “brass tacks”, numbers.

ASSORTATIVE MATING IN THE JEWEL WASP

The natural history of the Jewel wasp and the psychobiological rationale for this study is described in Figueredo and Sage (2007). Basically, this experiment was designed to test the effects of juvenile feeding experience on adult mate choice in the Jewel wasp. To

accurately determine the female choice of male wasps as sires for their offspring, genetically marked wasps were used. Two recessive mutant genotypes were used, one for white eyes and one for red eyes, which, when crossed, produce hybrids with brown eyes. Each female was given a choice of two males, one of each eye-color genotype (which were counterbalanced for the particular prey species, either housefly or blowfly, that they had been fed on as larvae). Females of independently crossed eye-color genotype and juvenile feeding experience were also used. Because these different genotypes might themselves influence mate choice, the experiment was designed to test the relative influence of two factors, 1) inherited eye-color genotype, and 2) conditioned foraging phenotype, on assortative mating in the Jewel wasp.

We wound up with a slight, though nonsignificant, negative correlation between the independent variables, female genotype (FG) and female juvenile host (FJ), due to postoperative differential subject mortality. Such correlated independent variables are easily modeled by hierarchical partitioning of variance. To be conservative in our testing of the effects of female host, we assigned causal priority to female genotype. Moreover, we did not hypothesize an interaction between these two factors.

The dependent variables, however, presented problems that could not be handled by unaided multiple regression. First, female Jewel wasps that feed on housefly hosts are significantly less fertile than those that feed on blowfly hosts (Figueredo 1987, 1989). This means that we could expect significantly different total numbers of offspring (TOT), regardless of paternity, from subjects reared on different juvenile hosts. Second, a variable number of Jewel wasp offspring remain immature in a dormant condition, called diapause, for up to six months past the normal fourteen days (and, thus, the end of our experiment). The total number of such dormant immature offspring (TIO) is known to vary with certain environmental factors, such as ambient temperatures and photoperiods, but might also be affected in some unknown way by the local mating conditions experienced by the mother. Because these dormant larvae had not yet developed eyes of either color, it was not possible to determine their paternity, perhaps producing nonrandom missing data. Third, male Jewel wasps are always produced asexually by the mother and, therefore, have no biological father at all. Thus, a female Jewel wasp that does not mate with either of the males provided might produce a clutch of all male offspring. The total number of male offspring (TMO) could not be simply ignored in this study because it might have indicated female rejection of both males, and perhaps constituted more nonrandom missing data.

Finally, the critical information in this study was provided by the paternity of the female offspring, but had to be controlled for all of the above causally prior offspring outcomes. Thus, the two dependent variables of principal interest were the total number of female

offspring sired by red-eyed, rather than white-eyed, males (TRFO), and the total number of female offspring sired by blowfly-reared, rather than housefly-reared, males (TBFO). In the interests of brevity, these female offspring were, respectively, referred to as “red-sired female offspring”, “white-sired female offspring”, “blowfly-sired female offspring”, and “housefly-sired female offspring”. Of course, the specified eye colors and juvenile hosts referred to are those of the fathers, rather than those of the daughters, because all genetic hybrids had brown eyes, and all posttest offspring were reared on the larger blowfly hosts for convenience. Because paternal genotypes and juvenile hosts were also slightly correlated, by the postoperative mortality noted above, eye color was modeled as causally prior to juvenile host in these two dependent variables, for the same reasons that they were in the two corresponding independent variables.

ALTERNATIVE METHODS OF ANALYSIS

The question then becomes, precisely how do we accomplish the requisite feats of multivariate statistical control? To expedite our discussion, we report the results of the most uninformative statistical model first, namely, simultaneous canonical analysis (SIMCA). This analysis was performed using PROC CANCELL in SAS (SAS Institute 1989). SIMCA constructed two pairs of canonical variates, V1 and W1, and V2 and W2, to represent our manifest variables. Pillai's Trace for the whole model was .21, $F(10,158) = 1.856$, $p = .055$. Furthermore, V1 was correlated to W1 by a canonical correlation of .42, which was statistically significant ($F(10,156) = 1.89$, $p = .05$); V2, however, was correlated to W2 by a canonical correlation of .17, which was statistically nonsignificant ($F(4,79) = 0.61$, $p = .66$). Presumably, we could therefore proceed to interpret V1 and W1, and forget about V2 and W2. Fortunately, V1 and W1 were defined by the following linear equations, using the standardized canonical coefficients, as seen in Table 1 above.

Using SIMCA, we were also helpfully provided with the resulting indirect correlations between W1 and the independent manifest variables (FG and FJ) and between V1 and the dependent manifest variables (TOT through TBFO), as well as all the corresponding estimates for the nonsignificant pair of canonical variates, V2 and W2, in both their raw and standardized forms. Unfortunately, in spite of this apparent wealth of information, we still had some trouble understanding what any of this meant for the results of our experiment. How did some linear combination of female genotype and female juvenile host somehow produce some other linear combination of the total numbers of various types of offspring?

A more informative approach was structural equations modeling (SEM), or confirmatory path analysis. This analysis was done using EQS (Bentler 1989). Be-

<i>Table 1. Linear equations for dependent canonical variates V1 and W1.</i>				
V1 =				0.864*FG - 0.413*FJ
W1 =	- 1.015*TOT	- 0.082*TIO	+ .046*TMO	+ .919*TRFO + 0.274*TBFO

<i>Table 2. Standardized path coefficients for saturated structural model.</i>				
TOT =				- .105*FG +.218*FJ
TIO =	.579*TOT			- .117*FG +.004*FJ
TMO =	.760*TOT	- .022*TIO		- .054*FG +.012*FJ
TRFO =	.856*TOT	- .255*TIO	-.241*TMO	+ .263*FG - .090*FJ
TBFO =	.753*TOT	- .103*TIO	-.307*TMO	-.083TRFO + .123*FG +.004*FJ

<i>Table 3. Standardized path coefficients for restricted model with non-significant pathways omitted.</i>				
TOT =				+ .230*FJ
TIO =	.595*TOT			
TMO =	.756*TOT			
TRFO =	.646*TOT	-.247*TIO		+ .281*FG
TBFO =	.630*TOT		-.296*TMO	

cause we had no strong causal theory to specify the structural relations between the multiple dependent variables, a fully saturated structural model was run, freely estimating every causal pathway possible between every dependent variable, specifying only the presumed causal priority. This method has also been referred to as the *cascade model* in cognitive psychology (e.g., Mouyi 2006, Demetriou et al. 2002). This specification produced a saturated structural model perfectly reproducing the data, with a residual chi-squared of zero, based on zero degrees of freedom. The standardized path coefficients are summarized in Table 2 above.

Using the statistical tests reported in the EQS output, we tested each of these path coefficients against zero. We then eliminated nonsignificant causal pathways to develop a more restricted model in which all the remaining pathways were statistically significant. Not surprisingly, this model was statistically acceptable: the chi-squared was 11.06 based on 14 degrees of freedom ($p = .68$). The standardized path coefficients for this model are described in Table 3.

Such brazenly exploratory use of SEM, however, is technically inappropriate. The probability of the model chi-squared reported, for example, no longer represented a prior probability under the null hypothesis, due to the many empirical model respecifications. Another disadvantage was that SEM did not provide for performing hierarchical tests of significance, except through such nested model comparisons, which required running multiple causal models. Of course, such results could be obtained by running a series of separate hierarchical regressions, as specified by the structural equations written above. This piecemeal method had the disadvantage of requiring a separate regression model for each of the dependent variables and, thus, still providing no valid overall protective test of significance. This alternative procedure was performed first using PROC GLM in SAS (SAS Institute 1989), and then using UNIMULT (Gorsuch 1991, Gorsuch and Figueredo 1991), and, in spite of the algorithmic differences in parameter estimation, both produced nearly identical results to each other. Furthermore, this alternative procedure has the practical advan-

tage of not requiring any specialized software to perform. Any standard software package (such as SAS or SPSS) that can estimate hierarchical multiple regressions (using Type I sums of squares) can be used.

In spite of all these limitations, at least SEM provided us with some clue as to what is going on. Controlling for all significant prior dependent variables, it appeared that there were only two significant causal effects of our two independent variables. As indicated by the positive effect of FJ on TOT, blowfly-reared females produced more total offspring than housefly-reared females, indicating their generally higher fertility, as was expected. As indicated by the positive effect of FG on TRFO, red-eyed females produced more red-sired female offspring than white-eyed females, indicating assortative mating by eye color. Contrary to the principal experimental hypothesis, however, blowfly-reared females apparently produced no more blowfly-sired female offspring than housefly-reared females, indicating no assortative mating by juvenile host. This was indicated by the lack of a significant causal effect of FJ on TBFO.

The third multivariate method used on this data was sequential canonical analysis (SEQCA). This analysis was performed using UNIMULT (Gorsuch 1991). All that was required for running this model was the specification of two hierarchies of causal priority, one for the two independent variables and another for the five dependent variables. The Pillai-Bartlett *V* for the whole model was .21, $F(10,158) = 1.86$, $p = .05$, which was the same, within rounding error, as the corresponding SIMCA results. In addition, we also obtained overall tests of the proportions of the variance of each of the five dependent variables accounted for by the linear combinations of the two independent variables, which are shown in Table 4.

SEQCA also provided the following hierarchical significance tests for the separate effects of each of the two independent variables on each of the five dependent variables and these are shown in Table 5.

Table 4. Multiple Correlations and Associated Tests of Significance for Each Dependent Variable in the Multivariate Model.

Variable	Effect Size	PBV	$F(2,158)$	<i>p</i>
TOT	R = .25	0.06	2.68	0.07
TIO	R = .14	0.03	1.31	0.3
TMO	R = .08	0.02	0.65	n.s.
TRFO	R = .32	0.15	7.06	< 0.005
TBFO	R = .12	0.02	0.73	n.s.

Table 5. Semipartial Correlations and Associated Tests of Significance for Each Independent Variable in the Multivariate Model.

Variable	Effect Size	PBV	$F(1,158)$	<i>p</i>
Dependent variable: TOT				
FG	r = -.13	0.02	1.43	0.2
FJ	r = .22	0.05	3.93	0.05
Dependent variable: TIO				
FG	r = -.14	0.03	2.62	0.1
FJ	r = .00	0	0	n.s.
Dependent variable: TMO				
FG	r = -.08	0.02	1.25	0.3
FJ	r = .02	0	0.05	n.s.
Dependent variable: TRFO				
FG	r = .31	0.14	13	<.001
FJ	r = -.09	0.01	1.12	0.3
Dependent variable: TBFO				
FG	r = .12	0.02	1.44	0.2
FJ	r = .01	0	0.02	n.s.

Note that the two statistically significant effects of the independent variables on the dependent variables were the same as those identified by SEM. Controlling for all the prior dependent variables, the standardized regression coefficients obtained by SEQCA were also very similar to those produced by SEM, but tended to become somewhat higher as one stepped sequentially down the causal hierarchy of dependent variables. These slightly higher parameter estimates were a systematic property of the current UNIMULT implementation of SEQCA, not an artifact of slightly different estimation algorithms. Recall that when the separate multiple regressions for the structural equations model were performed by UNIMULT, nearly identical parameter estimates were obtained. The SEQCA estimates were systematically higher because UNIMULT currently uses what Cohen and Cohen (1983) referred to as the “partial” rather than the “semipartial” residual correlations used by SEM. In the current version of UNIMULT, SEQCA residualized the denominator, or total variance to be explained (including the error), as well as the numerator, or portion of variance actually explained, of each of the sequential dependent variables on all the prior ones. This affected significance testing, as well as parameter estimation, by increasing the power of the tests. For com-

Table 6. Standardized Regression Coefficients for Each Independent Variable in the Multivariate Model.

TOT =	-.105*FG	+ .217*FJ
TIO =	-.144*FG	+ .004*FJ
TMO =	-.080*FG	+ .016*FJ
TRFO =	.300*FG	-.092*FJ
TBFO =	.121*FG	+ .012*FJ

parison with those obtained by SEM, the SEQCA standardized estimates are reported in Table 6.

The UNIMULT implementation of SEQCA does not explicitly provide estimates of the sequential effects of the dependent variables upon each other. Doing so would provide an alternative implementation of SEQCA that would be fully equivalent to an exploratory path analysis by explicitly including estimates of sequential effects between dependent variables. Such a model would restore the status of hierarchically residualized correlations as the “semipartial” (Cohen and Cohen 1983) typically estimated in SEM. This new model would have the added advantages of both hierarchical partitioning of variance and protective overall tests of significance, such as the Pillai-Bartlett V statistic, so important to exploratory data analysis. It would also help legitimize an alternative path analytic model for empirically-assisted theory development, instead of perpetuating the widespread abuse of available structural models that are clearly intended exclusively for theory confirmation. As it stands, as illustrated by our Jewel wasp example, SEQCA represents a useful diagnostic tool for isolating and identifying the direct effects of independent variables on a multiplicity of correlated dependent variables, if not for estimating the magnitudes of these effects in the conventional way. In either case, parameter estimation should never be based on the initially saturated model used for data exploration because the probable inclusion of nonsignificant variables substantially reduces the efficiency of estimation.

Superior estimates can be obtained by running the various final structural equations, as respecified by protected and hierarchical significance testing, as separate multiple regressions after the initial exploration of the data, as was briefly described above.

One potential disadvantage of SEQCA in relation to the alternative methods is that it requires the specialized UNIMULT software (Gorsuch 1991) to perform. Armed with the basic theory behind SEQCA, however, one may use the more commonly available SAS software (SAS Institute 1989), or even SPSS, to perform a series of hierarchical regressions in which multiple dependent

criterion variables are analyzed sequentially according to a hypothesized causal order. These dependent criterion variables can be entered sequentially into a system of multiple regression equations with each hierarchically prior criterion variable entered as the first predictor for the next, as we did in the SEM cascade model presented above. Each successive dependent variable can be predicted from an initial set of ordered predictor variables, each time entering the immediately preceding dependent variable hierarchically as the first predictor, then entering all the ordered predictors from the previous regression equation. Thus, each successive regression enters all of the preceding dependent variables in reverse causal order to statistically control for any indirect effects that might be transmitted through them. Within this analytical scheme, as with SEQCA, the estimated effect of each predictor is limited to its direct effect on each of the successive dependent variables. The general format for this system of hierarchical multiple regressions is as shown in Table 7 below.

Table 7. General format for multiple dependent criterion variables analyzed sequentially according to a hypothesized causal order.

$Y_4 =$		$\beta_1 X_1$	$+\beta_2 X_2$	$+\beta_3 X_3$	
$Y_5 =$	$\beta_4 Y_4$	$+\beta_1 X_1$	$+\beta_2 X_2$	$+\beta_3 X_3$	
$Y_6 =$	$\beta_5 Y_5$	$+\beta_4 Y_4$	$+\beta_1 X_1$	$+\beta_2 X_2$	$+\beta_3 X_3$

Where X_1 , X_2 and X_3 are the ordered predictor variables and Y_4 , Y_5 , and Y_6 are the ordered criterion variables, numbered consecutively after the predictors to avoid confusion among the subscripts. What is lost by using this method, as opposed to SEQCA, is the protective overall test of significance. What is gained by this method, as opposed to SEQCA, is obtaining estimates of the sequential effects among dependent variables. Thus, this is superior to merely including all the prior dependent variables as “covariates”. What is gained as opposed SEM is the ability to perform hierarchical partitioning of variance and hypothesis testing, as in SEQCA. In addition, one avoids compromising the purely confirmatory nature of the SEM tests of whole-model goodness-of-fit.

SUBSTANTIVE AND METHODOLOGICAL CONCLUSIONS

To finish our story, our study revealed no evidence of assortative mating by juvenile feeding experience in the Jewel wasp. It is therefore unlikely that differential juvenile feeding experience can lead to sympatric speciation in this particular species by either of the mecha-

nisms that have been proposed. The requisite reproductive isolation can apparently not be produced by differential juvenile feeding experience either indirectly by permanently differentiated foraging behavior or directly by differentially imprinted sexual preferences. Instead, assortative mating was shown to be greatly influenced by genetic markers, such as eye color. Converging evidence for these conclusions was variously obtained by hierarchical multiple regression, conventional path analysis, and sequential canonical analysis. Finally, the methodological implications of these findings for research are as follows. It was shown how the separate direct effects of the experimental manipulations on a set of seemingly hopelessly interdependent outcomes could be readily discriminated by sequential canonical analysis, producing results very similar to those obtainable by confirmatory path analysis without requiring the stronger theoretical assumptions of that model. This was done directly on the variables of practical interest, without altering the basic nature of the research question by constructing either inappropriate common factors or uninterpretable canonical variates. The statistical results were readily interpretable and directly relevant to the experimental hypotheses that motivated the study.

SAS INSTITUTE, INC. 1989. *SAS Language and Procedures: Usage, Version 6, First Edition*. SAS Institute, Cary, NC.

LITERATURE CITED

- BENTLER, P. M. 1989. *EQS, Structural Equations, Program Manual, Program Version 3.0*. BMDP Statistical Software, Inc., Los Angeles, CA.
- COHEN, J., AND P. COHEN. 1983. *Applied multiple regression/correlation analysis for the behavioral sciences*. Lawrence Erlbaum, Hillsdale, NJ.
- DEMETRIOU, A., C. CHRISTOU, G. SPANOUDIS, AND M. PLATSIDOU. 2002. The development of mental processing: Efficiency, working memory, and thinking. *Monographs of the Society of Research in Child Development* 67(1, Serial No. 268):1-154.
- FIGUEREDO, A. J. 1987. *The statistical measurement, developmental mechanisms, and adaptive ecological functions of conditioned host selection in the parasitoid Jewel wasp*. Dissertation, University of California, Riverside, CA.
- FIGUEREDO, A. J. 1989. Host-cue information processing by foraging Jewel wasps: Response-bias tracking of expected host-species encounter rates, not modified profitabilities. *Psychobiology* 17:435-444.
- FIGUEREDO, A. J., AND R. S. S. SAGE. 2007. Assortative mating in the Jewel wasp: 1. Female matching of eye-color genotype, not host-feeding phenotype. *Journal of the Arizona-Nevada Academy of Science*, 39:51-57.
- GORSUCH, R. L. 1991. *UniMult: For univariate and multivariate data analysis*. UniMult, Altadena, CA.
- GORSUCH, R. L., AND A. J. FIGUEREDO. 1991. Sequential canonical analysis as an exploratory form of path analysis. Paper. American Evaluation Association Conference, Chicago, Illinois.
- MOUYI, A. 2006. Untangling the cognitive processes web. Paper. Seventh Annual Conference of the International Society for Intelligence Research, San Francisco, California.