Methodika
1989, Vol. III, pp. 74–93

Moderating Effects in Multivariate Normal Distributions*)

Nanny Wermuth

Psychological Institute, University of Mainz

Abstract: For jointly normal variables, conditional independence graphs are used to present necessary and sufficient conditions for the lack of, what has been termed in social science literature as, a moderating effect. Variables have no moderating effect on a given measure of association, dependency, or variability, if this measure remains unchanged after marginalising over or after conditioning on these variables in a stepwise fashion. The results are applied to studying the association structure of certain personality characteristics, performance, and socioeconomic background of preschool children. In addition, it is shown that with the so-called moderated regression equations, a frequently recommended technique, it is impossible to deduce the lack or the presence of a moderating effect of a quantitative variable on a regression coefficient.

Key words: conditional independence graphs, moderating effects, parametric collapsibility.

1. Introduction

The concept of a moderating variable has received an immense attention in the social science literature as documented in a survey by Zedeck (1971) and in recent articles (see, e.g., Baron and Kenny (1986), Dalbert and Schmitt (1986)): however, its operationalisation has remained unsatisfactory until today.

Interest in this concept is strongest in nonexperimental research situations. For example, anxiety as a personality characteristic is known to be more likely to develop, the more a child perceives the parents as behaving inconsistently (Krohne, Kohlmann and Leidig (1982)). One then wants to determine conditions, under which this known relationship is intensified, reduced or, more generally, will change. This is an example, in which one wishes to establish *moderating effects* of further variables and to understand when such effects cannot occur.

---

*) I want to thank Carl-Walter Kohlmann, whose scepticism towards the moderated regression technique has led me to construct counterexamples which show the technique to be inappropriate for quantitative moderator variables.

Author's address: Nanny Wermuth, Psychologisches Institut, Universität Mainz, Postfach 39 80, D-6500 Mainz.

Related statistical tasks include the following: to estimate the bias introduced on a measure of association if a moderator variable has been left out of a model; and to state conditions, under which a measure remains unchanged, or, to put it differently, under which a moderating effect is lacking. Even though these aspects are well understood for linear regression coefficients (see, e.g., Goldberger (1964), chapter 10), they have, so far, not been discussed in connection with the concept of moderator variables. Furthermore, inappropriate techniques for establishing the lack of a moderating effect (Zedeck (1971), equations (1) to (3) and their interpretation on page 304) continue to be recommended in statistical textbooks (see, e.g., Cohen and Cohen (1983), chapter 10), and in recent articles (Cleary and Kessler (1982), Baron and Kenny (1986), Roos and Cohen (1987) and others).

In this paper we will summarise facts for various measures of associations in a joint normal distribution (section 2) which permit an easy derivation of moderating effects and of necessary and sufficient conditions for the lack thereoff. The techniques needed in that situation differ considerably from those used for only qualitative variables (Bishop (1971), Whittemore (1978), Wermuth (1987)), or for mixed variables (Wermuth (1989); the resulting conditions, however, are analogous and expressible with the help of conditional independencies.

We will define and discuss moderating effects on regression coefficients (section 3), on precisions and concentrations (section 4), on variances and covariances (section 5), and on standardised measures (section 6). Conditions for the lack of moderating effects are expressed as restrictions on parameters and as conditional independence statements. To simplify the communication of the obtained results, these are summarised, whenever it is possible, in terms of recursive conditional independence graphs. The latter are known to characterise associations structures of recursively factorised distributions and have been used previously for models with conditional Gaussian distributions (Lauritzen and Wermuth (1984), (1989), Wermuth and Lauritzen (1983), (1989), Edwards and Kreiner (1983)).

A set of data is presented (section 7) to illustrate the results. Finally, it is proven (section 8) that a zero regression coefficient of the constructed variable in a moderated regression is neither a necessary nor a sufficient condition for the lack of a moderating effect on a regression coefficient.

2. Notation and facts

In nondegenerate joint normal distributions of p variables with Σ as *covariance matrix*, the inverse Σ^{-1} is called the *concentration matrix* (Dempster (1969), p. 126). The elements σ_{ij} of Σ are *covariances* ($i \neq j$)

and variances; the elements of Σ^{-1} are *concentrations* ($i \neq j$) and *precisions*. Covariances and concentrations are measures of association; variances and precisions are measures of variability. Measures of linear dependencies are regression coefficients in linear regressions of response variables (regressands) $X_i (i \in a)$ on the influencing variables (regressors) $X_j (j \in b)$, where a and b partition $\{1, 2, \ldots, p\}$, the index set of the variables. The symmetric matrices Σ, Σ^{-1} may be partitioned accordingly as

$$\Sigma = \left[\begin{array}{c|c} \Sigma_{aa} & \Sigma_{ab} \\ \hline \Sigma_{ba} & \Sigma_{bb} \end{array}\right] \quad \Sigma^{-1} = \left[\begin{array}{c|c} \Sigma^{aa} & \Sigma^{ab} \\ \hline \Sigma^{ba} & \Sigma^{bb} \end{array}\right].$$

The *regression coefficient matrix* in a regression of X_a on X_b is $\Pi_{a|b} = \Sigma_{ab}\Sigma_{bb}^{-1}$, consisting of *regression vectors* $\beta_{i|b}^T$ for $i = 1, \ldots, |a|$.

Three basic facts from section 4.3 of Dempster (1969) must be extracted to derive moderating effects. Two are properties of the sweep-operator; the last is the special form of the triangular decomposition of a positive definite matrix in the case the latter is a concentration matrix.

Fact 1: SWP[b] Σ = RSW[a] $(-\Sigma^{-1})$ for disjoint a, b which partition $\{1, \ldots, p\}$.

This says that sweeping Σ on all rows and columns $i \in b$ and resweeping $(-\Sigma^{-1})$ on all rows and columns $i \in a$ yield identical results shown with the following two symmetric matrices:

$$\left[\begin{array}{c|c} \Sigma_{aa.b} & \Sigma_{ab}\Sigma_{bb}^{-1} \\ \hline & -\Sigma_{bb}^{-1} \end{array}\right] = \left[\begin{array}{c|c} (\Sigma^{aa})^{-1} & -(\Sigma^{aa})^{-1}\Sigma^{ab} \\ \hline & -\Sigma^{bb.a} \end{array}\right]. \tag{1.1}$$

This identity may also be proven by standard results on inverting partitioned matrices and it immediately gives different representations of the regression coefficient matrix ($\Pi_{a|b}$), as well as of the *partial covariance matrix* ($\Sigma_{aa.b}$), and of the partial concentration matrix ($\Sigma^{bb.a}$)

$$\begin{aligned} &\text{(i)} \quad \Pi_{a|b} = \Sigma_{ab}\Sigma_{bb}^{-1} = -(\Sigma^{aa})^{-1}\Sigma^{ab} \\ &\text{(ii)} \quad \Sigma_{aa.b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ab}^{T} = (\Sigma^{aa})^{-1} \\ &\text{(iii)} \quad \Sigma^{bb.a} = \Sigma^{bb} - (\Sigma^{ab})^{T}(\Sigma^{aa})^{-1}\Sigma^{ab} = \Sigma_{bb}^{-1}. \end{aligned} \tag{1.2}$$

Note that $\Sigma_{aa.b}$ and Σ^{aa} are the covariance matrix and concentration matrix in the conditional distribution of X_a given $X_b = x_b$, while Σ_{bb} and $\Sigma^{bb.a}$ are the corresponding matrices in the marginal distribution of X_b.

Fact 2: SWP[c, d] Σ = SWP[c] (SWP[d] Σ) = SWP[d] (SWP[c] Σ) for c, d any disjoint nonempty subsets of $\{1, \ldots, p\}$.

This commutativity property gives – for b = (c, d) – the following representation of the regression coefficient matrix:

$$\Pi_{a|b} = [\Sigma_{ac.d}\Sigma_{cc.d}^{-1} \;\vert\; \Sigma_{ad.c}\Sigma_{dd.c}^{-1}] \tag{1.3}$$

with e.g., $\Sigma_{ac.d}$, $\Sigma_{cc.d}$ being submatrices of the conditional covariance matrix ($\Sigma_{[a,c].d}$) of X_a and X_c given X_d.

To make this evident, the two sweeping steps are written out:

$$\mathrm{SWP}[d]\,\Sigma = \left[\begin{array}{c|c|c} \Sigma_{aa.d} & \Sigma_{ac.d} & \Pi_{a|d} \\ \hline & \Sigma_{cc.d} & \Pi_{c|d} \\ \hline & & -\Sigma_{dd}^{-1} \end{array}\right],$$

$$\mathrm{SWP}[c]\,\Sigma = \left[\begin{array}{c|c|c} \Sigma_{aa.c} & \Pi_{a|c} & \Sigma_{ad.c} \\ \hline & -\Sigma_{cc}^{-1} & \Pi_{d|c} \\ \hline & & \Sigma_{dd.c}^{-1} \end{array}\right],$$

where e.g., $\Pi_{a|d} = \Sigma_{ad}\Sigma_{dd}^{-1}$.

By letting a = (g, i), b = (j, d) and (a, b) = (1, ... p), one obtains from (1.3) and (1.2) the coefficient of X_j in the regression of X_a on X_b as

$$\beta_{ij.d} = \sigma_{jj.d}(\sigma_{ij.d})^{-1} = -\sigma^{ij.g}(\sigma^{ii.g})^{-1}. \tag{1.4}$$

It is identical to the coefficient of X_j in the regression of X_i on X_b alone and called a *partial regression coefficient* of X_j.

With d = (k, l) in addition to a = (g, i), b = (j, d) the partial regression coefficient $\beta_{ij.d}$ is – with the help of Fact 2 – seen to be related to other partial regression coefficients as

$$\beta_{ij.d} = \beta_{ij.kl} = \beta_{ij.l} - \beta_{ik.jl}\beta_{kj.l}, \tag{1.5}$$

where, for instance, $\beta_{ij.l}$ and $\beta_{kj.l}$ are both coefficients of X_j in regressions with regressors X_j, X_l, but with different response variables: X_i and X_k, respectively.

Fact 3: If $A^T D^{-1} A = \Sigma^{-1}$ is the upper triangular decomposition of Σ^{-1}, then A is an upper triangular matrix containing ones on the diagonal, and $-\beta_{i|d(i)}$ with $d(i) = \{i+1, \ldots p\}$, is the right off-diagonal part of row i. Furthermore, D^{-1} is a diagonal matrix with reciprocal values of partial variances as elements: $d_{ii} = \sigma_{ii.d(i)}^{-1}$.

This says, for example for $\Sigma^{-1} = \Sigma^{xyzu}$, that A contains in line 2 negative values of the regression coefficients in the linear regression of Y on Z and U; the matrix D contains precisions such as $d_{22} = \sigma^{yy.x} = 1/\sigma_{yy.zu}$.

By using Fact 3, one obtains the following simple expression for the determinants of Σ and Σ^{-1}, which by the positive definiteness of Σ are known to be positive. Since $|\Sigma^{-1}| = |A^T||D^{-1}||A|$ and $|A| = 1$, it follows that $|\Sigma^{-1}| = |D^{-1}|$ is a product of partial precisions, and $|\Sigma| = 1/|\Sigma^{-1}| = |D|$ is a product of partial variances. For instance we see that

$$\begin{aligned} &\text{(i)} \quad |\Sigma^{xyuz}| = \sigma^{xx}\sigma^{yy.x}\sigma^{zz.xy}\sigma^{uu.xyz} > 0\,, \\ &\text{(ii)} \quad |\Sigma_{xyzu}| = \sigma_{xx.yzu}\sigma_{yy.zu}\sigma_{zz.u}\sigma_{uu} > 0\,. \end{aligned} \tag{1.6}$$

Another well-known fact then becomes obvious. For positive definite Σ, all (partial) precisions and all (partial) variances are positive, since the order of variables in (1.6) can be permuted.

The definition of a recursive conditional independence graph and two more facts are needed for deriving and visualising conditions for the lack of moderating effects.

Fact 4: Given a joint normal distribution, one stays in the family of normal distributions after marginalising over a subset of variables X_a, and after conditioning on X_a (compare e.g., Anderson (1958), section 2).

A conditional independence graph provides the information for the picture of an association or dependence structure defined with the help of conditional independencies of variables pairs. A *recursive conditional independence graph* (G^r) has been used for a more general family of distributions by Lauritzen and Wermuth (1984), (1989). The graph consists of p vertices, and at most one edge for each pair $\{i, j\}$ of distinct vertices. This graph is an undirected one if it is to characterise an association structure without response variables. Otherwise the type and the direction of all edges are determined by a dependence chain $\mathscr{C} = (C_1, \ldots, C_T)$, with $T > 1$. This chain partitions the vertex set and defines T sets of so-called concurrent variables. The sets are $C_1 \cup \ldots \cup C_T$, $C_2 \cup \ldots \cup C_T$, $C_3 \cup \ldots \cup C_T, \ldots, C_T$. Concurrent means that the variables are to be analysed simultaneously. Exactly one of three possibilities holds for each pair of vertices.

For a pair $\{i, j\}$ with $i \in C_t$, $j \in C_l$ one has:

(i) the edge is missing, or
(ii) the edge is a line if i and j are in the same chain element ($t = l$), or
(iii) the edge is an arrow pointing from j to i if X_j is an influence to response $X_i (t < l)$.

A missing edge implies that X_i is to be conditionally independent of X_j given the remaining concurrent variables, the variables X_d with $d = \{C_t, \ldots, C_T\} \setminus \{i, j\}$. A chain element ($C_t$) is drawn as a box. A missing edge means an only indirect relation while present edges denote direct relations.

Fact 5: If the distribution of a system of variables given by a conditional independence graph G^r is jointly normal, then a missing edge for $\{i, j\}$ with $i \in C_t$, $j \in C_l$, $t < l$, is equivalent to a zero partial concentration of X_i and X_j partialled over the nonconcurrent variables, over X_k with $k \in \{C_1, \ldots, C_{t-1}\}$.

To make this evident, note (e.g. from Anderson (1958), section 2.5) that in an unconditional joint normal distribution of variables (X_i, X_j, X_d), X_i is conditionally independent of X_j given the remaining variables X_d $(X_i \perp\!\!\!\perp X_j \mid X_d)$, if and only if $\sigma_{ij.d} = 0$. By using (1.4) and (1.5) it follows that the following statements are equivalent:

$$\begin{aligned} &\text{(i)} && X_i \perp\!\!\!\perp X_j \mid X_d\,, \\ &\text{(ii)} && \sigma^{ij.g} = 0\,, \\ &\text{(iii)} && \sigma_{ij.d} = 0\,, \\ &\text{(iv)} && \beta_{ij.d} = 0\,, \end{aligned} \tag{1.7}$$

where $g = \{C_1, \ldots, C_{t-1}\}$ denotes variables not concurrent to X_i, X_j and $d = \{C_t \ldots, C_T\} \setminus \{i, j\}$ denotes the remaining variables concurrent to X_i, X_j.

We are now equipped to derive and describe the moderating effects for different parameters. In order to keep the notation simple, we restrict our arguments to an unconditional joint distribution of four variables. This situation is complex enough to see the possible extension to cases that are more general.

3. Moderating effects on regression coefficients

Given a joint normal distribution containing variables X, Y, U, Z, the moderating effects of variable U alone and of variables U and Z jointly on the regression coefficient of response Y on the influencing variable X are the changes introduced by modifying the regression from one without U and Z as regressors, to one with U alone added, and to one with both U and Z included.

For an unconditional joint density f_{XYZU}, this means to move from $f_{X|Y}$ to $f_{X|YU}$ and to $f_{X|YZU}$, and to register the changes in the first element of the regression vectors $\beta_{x|y}$, $\beta_{x|yu}$, and $\beta_{x|yzu}$: in β_{xy}, $\beta_{xy.u}$, and $\beta_{xy.zu}$.

Result 1.1: The moderating effects are of

(i) U alone on β_{xy}: $-\beta_{xu.y}\beta_{uy}$,

(ii) U and Z on β_{xy}: $-(\beta_{xu.y}\beta_{uy} + \beta_{xz.yu}\beta_{zy.u})$.

Result 1.2: There is no moderating effect of

(i) U alone on β_{xy} if and only if $X \perp\!\!\!\perp U \mid Y$ or $Y \perp\!\!\!\perp U$,

(ii) U nor of U and Z on β_{xy} if and only if $[X \perp\!\!\!\perp U \mid Y$ or $Y \perp\!\!\!\perp U]$ and $[X \perp\!\!\!\perp Z \mid (Y, U)$ or $Y \perp\!\!\!\perp Z \mid U]$.

Proof: With i and j as indices of X and Y, Result 1.1 (i) follows from (1.5) with k as index of U and $\{l\} = \emptyset$; Result 1.1 (ii) follows from (1.5) with k as index of Z, 1 as index of U and by use of Result 1.1 (i). Result 1.2 is obtained with (1.7) from Result 1.1.

Proposition 1: Let X, Y, Z, U have a nondegenerate joint normal distribution then $\beta_{xy} = \beta_{xy.u} = \beta_{xy.uz}$ if and only if the joint density is such that (i) *and* (ii) are satisfied:

(i) f_{XYZU} can be characterised as in Figure 1 by a conditional independence graph with chain $\mathscr{C} = (\{X\}, \{Z\}, \{U, Y\})$ having at least one of the marked arrows ("/") missing;

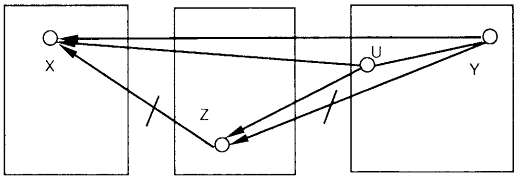


Figure 1: For $\beta_{xy} = \beta_{xy.u} = \beta_{xy.uz}$ one of the two marked relations has to be indirect (condition (i))

(ii) the marginal density f_{XYU} can be characterised as in Figure 2 with a least one of the marked arrows missing.

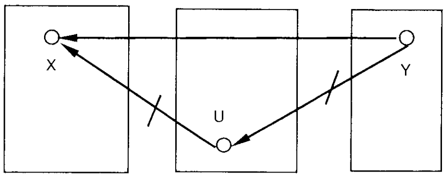


Figure 2: For $\beta_{xy} = \beta_{xy.u} = \beta_{xy.uz}$ one of the two marked relations has to be indirect (condition (ii))

Proposition 1 follows with Result 1.2 and Fact 4 from the definition of the recursive conditional independence graph.

4. Moderating effects on precisions and concentrations

In a joint normal distribution containing variables X, Y, U, Z, the moderating effects of variable U alone, and of variables U and Z jointly on the precision of X, or on the concentration of X, Y are the changes introduced by marginalising over U alone, and over both of U and Z.

For an unconditional joint distribution with density f_{XYZU}, this means to move from f_{XYZU} to the marginal distributions f_{XYZ} and f_{XY}, and to look at the relevant submatrices of Σ^{xyzu}, $\Sigma^{xyz.u}$, and $\Sigma^{xy.zu}$. Since precisions and concentrations do not change by conditioning (see (1.2 (ii)), one compares at the same time the concentration matrices in $f_{XY|ZU}$, $f_{XY|Z}$, and f_{XY}, which are $(\Sigma_{xy.zu})^{-1}$, $(\Sigma_{xy.z})^{-1}$, and $(\Sigma_{xy})^{-1}$. The conditional independence graphs in Figure 3 depict these situations.

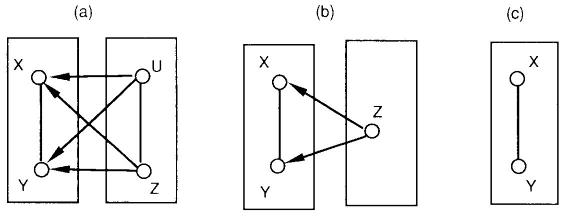


Figure 3: Dependence structure with the meaning of the edge for (X, Y) being σ^{xy} in (a), $\sigma^{xy.u}$ in (b) and $\sigma^{xy.uz}$ in (c)

Result 2.1: The moderating effects are of

(i) U alone on σ^{xy}: $\beta_{ux.yz}\sigma^{yu}$,

(ii) U and Z on σ^{xy}: $\beta_{ux.yz}\sigma^{yu} + \beta_{zx.y}\sigma^{yz.u}$.

Results 2.2: There is no moderating effect

(i) of U alone on σ^{xy} if and only if $X \perp\!\!\!\perp U|(Y, Z)$, or $Y \perp\!\!\!\perp U|(X, Z)$,

(ii) of U alone nor of U and Z on σ^{xy} if and only if $[X \perp\!\!\!\perp U|(Y, Z)$ or $Y \perp\!\!\!\perp U|(X, Z)]$, and $[X \perp\!\!\!\perp Z|Y$ or $Y \perp\!\!\!\perp Z|X]$.

For the proof of Result 2.1 note from Fact 3 and (1.4) the form of the triangular decomposition of $\Sigma^{-1} = A^T D^{-1} A$ as

$$A^T D^{-1} = \begin{bmatrix} \sigma^{11} & 0 & 0 & 0 \\ \sigma^{12} & \sigma^{22.1} & 0 & 0 \\ \sigma^{13} & \sigma^{23.1} & \sigma^{33.12} & 0 \\ \sigma^{14} & \sigma^{24.1} & \sigma^{34.12} & \sigma^{44.123} \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & -\beta_{12.34} & -\beta_{13.24} & -\beta_{14.23} \\ 0 & 1 & -\beta_{23.4} & -\beta_{24.3} \\ 0 & 0 & 0 & -\beta_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Next, write out the matrix product $[A^T D^{-1}]A$, fix the order of the variables to give Σ_{uxyz}, then (i) can be read off from position (3,2), while (ii) follows from position (4,3) in Σ_{uzxy}. Result 2.2 follows with (1.7) from Result 2.1.

Proposition 2: Let X, Y, Z, U have a nondegenerate joint normal distribution, then $\sigma^{xy} = \sigma^{xy.u} = \sigma^{xy.zu}$ if and only if f_{XYZU} can be characterised as in Figure 4 with a conditional independence graph having chain $\mathscr{C} = (\{U\}, \{Z\}, \{X, Y\})$ and at least one arrow marked "/" *and* another marked "//" missing.

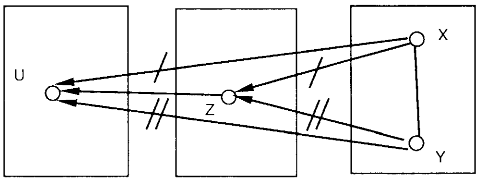


Figure 4: For $\sigma^{xy} = \sigma^{xy.z} = \sigma^{xy.zu}$ one of the relations marked / and one of the relations marked // have to be indirect

Proof: Proposition 2 results from Result 2.2 and Fact 4.

Result 3.1: The moderating effects are

(i) of U alone on σ^{xx}: $\beta^2_{xu.yz}\sigma^{xx}$,
(ii) of U and Z on σ^{xx}: $\beta^2_{xu.yz}\sigma^{xx} + \beta^2_{xz.y}\sigma^{xx.u}$.

Result 3.2: There is no moderating effect

(i) of U alone on σ^{xx} if and only if $X \perp\!\!\!\perp U | (Y, Z)$,
(ii) of U alone nor of U and Z on σ^{xx} if and only if $X \perp\!\!\!\perp U | (Y, Z)$ and $X \perp\!\!\!\perp Z | Y$.

Proposition 3: Let X, Y, Z, U have a nondegenerate joint normal distribution. Then $\sigma^{xx} = \sigma^{xx.u} = \sigma^{xx.uz}$ if and only if f_{XYZU} can be characterised by a conditional independence graph as in Figure 5 with at least the two marked lines missing

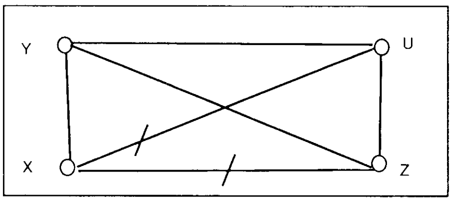


Figure 5: For $\sigma^{xx} = \sigma^{xx.u} = \sigma^{xx.uz}$ both of the marked relations have to be indirect

Proofs of the Results are analogous to those of Results 2.1 and 2.2 and are left to the reader. Proposition 2 follows from Result 3.2, and Fact 4 and after noting that $X \perp\!\!\!\perp U | (Y, Z)$ and $X \perp\!\!\!\perp Z | Y$ taken together imply $X \perp\!\!\!\perp (U, Z) | Y$.

5. Moderating effects on variances and covariances

In a joint normal distribution containing variables X, Y, U, Z, the moderating effects of variable U alone and, of variables U and Z jointly on the variance of X or on the covariance of X, Y are the changes introduced by conditioning on U alone, and on both of U and Z.

For an unconditional joint distribution with density f_{XYZU}, this means to move from f_{XYZU} to the conditional distributions with densities $f_{XYZ|U}$ and $f_{XY|ZU}$, and to look at the relevant submatrices of Σ_{xyzu}, $\Sigma_{xyz.u}$, and $\Sigma_{xy.zu}$. Since variances and covariances do not change by marginalising (see (1.2)(iii)), one simultaneously compares the covariance matrices Σ_{xy}, $\Sigma_{xy.u}$, and $\Sigma_{xy.uz}$ in f_{xy}, $f_{xy|u}$, and $f_{xy|uz}$, as well as the variances in f_x, $f_{x|u}$ and $f_{x|uz}$.

Result 4.1: The moderating effects are

(i) of U alone on σ_{xy}: $-\beta_{xu}\sigma_{yu}$,
(ii) of U and Z on σ_{xy}: $-(\beta_{xu}\sigma_{yu} + \beta_{xz.u}\sigma_{yz.u})$.

Result 4.2: There is no moderating effect

(i) of U alone on σ_{xy} if and only if $X \perp\!\!\!\perp U$ or $Y \perp\!\!\!\perp U$,
(ii) of U nor of U and Z on σ_{xy} if and only if [$X \perp\!\!\!\perp U$ or $Y \perp\!\!\!\perp U$ and $X \perp\!\!\!\perp Z|U$ or $Y \perp\!\!\!\perp Z|U$].

For the proof of Result 4.1, note from Fact 3 and (1.4) the simple form of the triangular decomposition of $\Sigma = A^{-1}D(A^T)^{-1}$ with

$$A^{-1}D = \begin{bmatrix} \sigma_{11.234} & \sigma_{12.34} & \sigma_{13.4} & \sigma_{14} \\ 0 & \sigma_{22.34} & \sigma_{23.4} & \sigma_{24} \\ 0 & 0 & \sigma_{33.4} & \sigma_{34} \\ 0 & 0 & 0 & \sigma_{44} \end{bmatrix}$$

$$(A^{-1})^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \beta_{12.34} & 1 & 0 & 0 \\ \beta_{13.4} & \beta_{23.4} & 1 & 0 \\ \beta_{14} & \beta_{24} & \beta_{34} & 1 \end{bmatrix}.$$

Next, write out the matrix product $[A^{-1}D]\,(A^T)^{-1}$, fix the order of the variables to give Σ_{zyxu}, then (i) can be read off from position (2, 3), while (ii) follows from position (1, 2) of Σ_{yxzu}. Result 4.2 follows with (1.7) from Result 4.1.

Proposition 4: Let X, Y, Z, U have a nondegenerate joint normal distribution. Then $\sigma_{xy} = \sigma_{xy.u} = \sigma_{xy.zu}$ if and only if the joint density is such that (i) or (ii) are satisfied:

(i) marginalising over $V \in (X, Y)$ leads to the density f_{WZU} with $W = \{X, Y\} \setminus V$, which can be characterised by a conditional independence graph as in Figure 6 having at least the two marked vertices missing

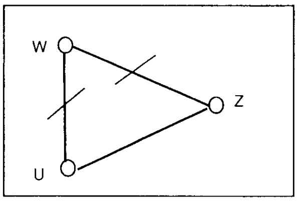


Figure 6: If the two marked relations are indirect then $\sigma_{xy} = \sigma_{xy.u} = \sigma_{xy.zu}$ (condition (i))

(ii) marginalising over $V \in \{X, Y\}$ and marginalising over Z and $W = \{X, Y\} \setminus V$ leads to the densities f_{WZU} and f_{VU}, which can be characterised by conditional independence graphs as in Figure 7 having at least the marked vertices missing

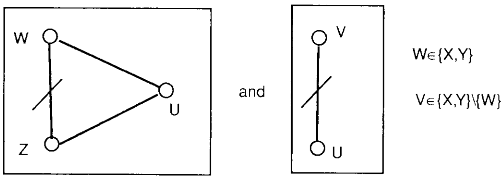


Figure 7: If the two marked relations are indirect then $\sigma_{xy} = \sigma_{xy.u} = \sigma_{xy.uz}$ (condition (ii))

Proof: After noting that $W \perp\!\!\!\perp U$ and $W \perp\!\!\!\perp Z | U$ implies $W \perp\!\!\!\perp (Z, U)$ and that the remaining possibilities in Result 4.2 (ii) can be represented as $W \perp\!\!\!\perp Z | U$ and $V \perp\!\!\!\perp U$ for $W \in \{X, Y\}$ and $V \in \{X, Y\} \setminus \{W\}$, Proposition 4 follows from Result 4.2 and Fact 4.

Result 5.1: The moderating effects are

(i) of U alone on σ_{xx}: $-\beta^2_{xu}\sigma_{uu}$,
(ii) of U and Z on σ_{xx}: $-(\beta^2_{xu}\sigma_{uu} + \beta^2_{xz.u}\sigma_{zz.u})$.

Result 5.2: There is no moderating effect

(i) of U alone on σ_{xx} if and only if $X \perp\!\!\!\perp U$,
(ii) of U nor of U and Z on σ_{xx} if and only if $X \perp\!\!\!\perp U$ and $X \perp\!\!\!\perp Z | U$.

Proposition 5: Let X, Y, Z, U have a nondegenerate joint normal distribution. Then $\sigma_{xx} = \sigma_{xx.u} = \sigma_{xx.zu}$ if and only if the marginal density f_{XZU} can be characterised by a conditional independence graph as in Figure 8 having at least the two marked vertices missing

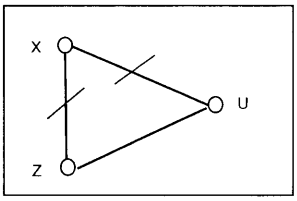


Figure 8: For $\sigma_{xx} = \sigma_{xx.u} = \sigma_{xx.zu}$ the two marked relations have to be indirect

Proofs of the Results are analogous to those of Results 4.1 and 4.2 and are left to the reader. Proposition 5 follows from Result 5.2 and Fact 4 after noting that $X \perp\!\!\!\perp Z|U$ and $X \perp\!\!\!\perp U$ implies $X \perp\!\!\!\perp (Z|U)$.

6. Moderating effects on the correlation coefficient and on standardised regression coefficients

Correlation coefficients are standardised measures of linear associations. Let T_{aa}, $T_{aa.d}$ be both diagonal matrices, the first with elements $(\sigma_{ii})^{-1/2}$ and the second with diagonal elements $(\sigma_{ii.d})^{-1/2}$, then $P_{aa} = T_{aa}\Sigma_{aa}T_{aa}$ and $P_{aa.d} = T_{aa.d}\Sigma_{aa.d}T_{aa.d}$ denote the *matrices of marginal* and of *partial correlation coefficients*, respectively. The *partial* correlation coefficient $\varrho_{xy.u}$ does not relate in a simple way to ϱ_{xy}, the *marginal*, or simple one:

$$\varrho_{xy.u} = (\varrho_{xy} - \varrho_{xu}\varrho_{yu})/[(1 - \varrho_{xu}^2)(1 - \varrho_{yu}^2)]^{-1/2}.$$

They coincide in the trivial case $(X, Y) \perp\!\!\!\perp U$, but a moderating effect of U on ϱ_{xy} can be lacking under conditions that are unrelated to independencies. This fact is illustrated with the following correlation matrix in which $\varrho_{12} = \varrho_{12.3}$, but no variable pair is marginally or conditionally independent:

$$\begin{bmatrix} 1 & 0.66 & 0.5 \\ & 1 & 0.2 \\ & & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 2.3624 & -1.3780 & -0.9055 \\ & 1.8455 & 0.3199 \\ & & 1.3888 \end{bmatrix}$$

Standardised regression coefficients or beta coefficients are standardised measures of linear dependencies. They are obtained as common regression coefficients by starting with the correlation matrix P instead of the covariance matrix Σ. They have traditionally been used in the social sciences as an aid in determining the relative importance of the different regressor or influencing variables. It is known that this is done only at the cost of ignoring possible differences in variability (see e.g., Weisberg (1980), p. 168).

Let T_{bb} be a diagonal matrix with diagonal elements $(\sigma_{ii})^{-1/2}$ for $i \in b$. Then the standardised regression coefficients matrix $\Pi^*_{a|b}$ and the residual matrix $\Sigma^*_{aa|b}$ relates to the nonstandardised one as

$$\Pi^*_{a|b} = T_{aa}\,\Pi_{a|b}\,T_{bb}^{-1}, \quad \Sigma^*_{aa.b} = T_{aa}\,\Sigma_{aa.b}\,T_{aa}\,.$$

This leads to well-known relations like

$$\beta^*_{xy} = \beta_{xy}\left(\frac{\sigma_{xx}}{\sigma_{yy}}\right)^{1/2}, \beta^*_{xy.u} = \beta_{xy.u}\left(\frac{\sigma_{xx}}{\sigma_{yy}}\right)^{1/2}, \beta^*_{xy.uz} = \beta_{xy.uz}\left(\frac{\sigma_{xx}}{\sigma_{yy}}\right)^{1/2},$$

from which it follows, in turn, that the conditions for the lack of a moderating effect have to coincide for standardised and unstandardised regression coefficients.

7. Judging moderating effects on a set of data

We illustrate judgements on moderating effects on a set of data taken from Hodapp (1984), p. 72). Self-concept (C) is regarded as a response to the potential influences intelligence (I), performance (P), and socioeconomic status (S). Available are data for 303 boys attending classes in kindergarten. The observed correlation matrix shows an excellent fit to the one estimated under a hypothesis characterised here with the conditional independence graph in Figure 9.

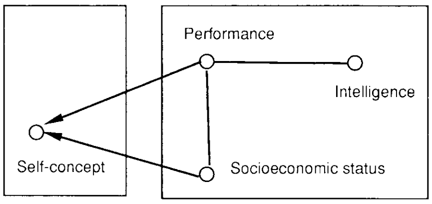
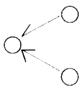


Figure 9: Well-fitting dependence structure to self-concept data

Expressed in words: self-concept depends directly on performance and socioeconomic status, but only indirectly on intelligence. The association between intelligence and socioeconomic status disappears if one controls for the relationship of performance to either variable.

Since the regression structure in Figure 9 does not contain a configuration of the type , it is equivalent to the following symmetric association structure (Wermuth and Lauritzen, (1989))

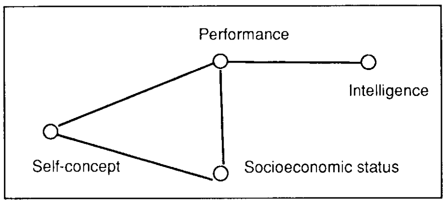


Figure 10: Well-fitting association structure to self-concept data

We know for this type of structure, how to read directly off the graph all independencies; in particular $I \perp\!\!\!\perp (C, S) | P$ as most condensed description of this structure (for proofs and motivations, see, Kiiveri, Speed and Carlin (1984), or Lauritzen and Wermuth (1984), (1989)).

The oberserved marginal correlations are displayed in Table 1 (a), as well as the corresponding precisions and concentrations, while the estimated marginal correlations $\hat{\varrho}_{ij}$, the estimated precisions $\hat{\varrho}^{ii}$, and concentrations $\hat{\varrho}^{ij}$ are in Table 1 (b). The likelihood-ratio chi-square statistic for the hypothesis in Figure 9 against the saturated, i.e. the observed association structure is

$$n \log(\det(\hat{P})/\det(R)) = 55 \log(0.1675/0.1673) = 0.30,$$

on two degrees of freedom. It is evident from the test results given in Table 2 that no additional independencies may be assumed.

Figure 10 and Proposition 3 indicate that C and S have no moderating effect on the precision of I if the hypothesis in Figure 9 is satisfied.

Figure 9 and the first picture of Proposition 1 when used repeatedly show that under the hypothesis $\beta_{CP.IS} = \beta_{CP.S}$ and $\beta_{CS.IP} = \beta_{CS.P}$. This means that up to random variation I has no moderating effect on the regression coefficients of C on P and S as estimated from the observed correlations:

$$\begin{aligned} \beta^*_{CP.IS} &= 0.753, \quad \beta^*_{CP.S} = 0.777 \\ \beta^*_{CS.PI} &= 0.479, \quad \beta^*_{CS.P} = 0.479. \end{aligned}$$

Similarly, one finds that the moderating effect of S on the regression coefficient $\beta^*_{CP.I}$ is not lacking. From (1.5) it is: $-\beta^*_{CS.PI} \cdot \beta^*_{SP.I}$ and is estimated from the observed correlations as $(-0.470)(-0.423) = 0.20$.

Table 1: Marginal correlations (upper half), precisions (diagonal), and concentrations (lower half) for the self-concept data as observed (a) and as estimated under the hypothesis of Figure 9(b).

(a) Observed values

Variables	C	P	S	I
C	2.104	0.578	0.157	0.454
P	− 1.584	3.759	− 0.415	− 0.758
S	− 1.009	1.270	1.692	− 0.310
I	− 0.068	− 1.736	− 0.020	2.353

(b) Values estimated under the hypothesis of Figure 9

Variables	C	P	S	I
C	2.102	0.578	0.157	0.438
P	− 1.634	3.828	− 0.415	0.758
S	− 1.008	1.285	1.691	− 0.315
I	0.000	1.782	0.000	2.351

Table 2: Liklihood-ratio test results

Conditional independence of pair	Concurrent variables	Chi-square value	Degrees of freedom
(C, P)	CPIS	115.54	1
(C, I)	CPIS	0.28	1
(C, S)	CPIS	101.97	1
(I, P)	IPS	228.36	1
(I, S)	IPS	0.02	1
(P, S)	IPS	26.68	1

8. Discussion

The results in this paper make it plain that moderator effects depend strongly on the type of the relationship investigated between two variables. A variable can have no moderating effect on a measure of dependency, but at the same time have such an effect on a symmetric measure of association and vice versa.

A moderating effect also need not be symmetric in the following sense: if Z has no moderating effect on the regression coefficient of Y on X, it is still possible that X has a moderating effect on the regression coefficient of Y on Z. This is, in fact, the case if X and Z are marginally dependent and $Y \perp\!\!\!\perp Z|X$, but not $Y \perp\!\!\!\perp X|Z$ (compare Proposition 1).

If one knows the necessary and sufficient conditions for the lack of moderating effects summarised in this paper, it is easy to perform a goodness-of-fit test to decide on the lack or presence of moderating effects. In what follows, it is shown that the technique of so-called moderated regressions is not suitable for this purpose. Following Saunders (1956) Zedeck had recommended the following procedure (Zedeck (1971), p. 304). Compute

"three regression equations

[1] $y = a + bx$,

[2] $y = a + b_1 x + b_2 z$,

where z is the potential moderator but is treated as an independent predictor, and

[3] $y = a + b_1 x + b_2 z + b_3 x \cdot z$

(moderated regression equation). If Equations 2 and 3 are significantly different from Equation 1, but not from each other, then the variable is an independent predictor and moderator variable."

This claim is proven false with the set of variables Y, Z, X in Table 3 (a), which have the covariance and concentration matrices, including the variable $W = X \cdot Z$ displayed in Table 4.

The regression coefficients for the above equations are derived from these as

[1] $y = (-0.101) + (0.862)\,x;\quad R^2 = 0.489,$
[2] $y = (0.000) + (0.600)\,x + (0.600)\,z;\quad R^2 = 0.616,$
[3] $y = (0.000) + (0.600)\,x + (0.600)\,z + (0.000)\,x \cdot z;\quad R^2 = 0.616.$

Thus, $b_3 = 0$, but Z has nevertheless a moderating effect on $b(=\beta_{yx})$. This effect is $-\beta_{yz.x}\beta_{zx} = 0.262$ (compare Result 1.1 (i)). This shows that the vanishing of the regression coefficient of the constructed variable $W = X \cdot Z$ in [3] is not a sufficient condition for the lack of a moderating effect.

Conversely, it has been claimed that the regression coefficient of the constructed variable (b_3 in [3]) has to be zero if Z has no moderating effect on the regression coefficient in the linear regression of Y on X (b in [1]) (Steyer (1983), Cohen and Cohen (1983), Borkenau (1985), Dalbert and Schmitt (1986), Baron and Kenny (1986), Roos and Cohen (1987)). A counterexample is given with the variables in Table 3 (b). The covariance and concentration matrices of these variables and $W = X \cdot Z$ are in Table 5. The regression coefficients and the coefficients of determination for the three regression equations are obtained as

[1] $y = (-0.029) + (1.152)\,x;\quad R^2 = 0.292,$
[2] $y = (0.089) + (1.152)\,x + (1.075)\,z;\quad R^2 = 0.511,$
[3] $y = (0.005) + (1.481)\,x + (0.532) + (1.889)\,x \cdot z;\quad R^2 = 0.828.$

Table 3: Two sets of variables, which give counterexamples to the moderated regression technique

(a) Counter example to sufficiency			(b) Counter example to necessity		
Y	X	Z	Y	X	Z
− 2.9482	− 1.7203	− 0.6522	2.0864	0.6357	0.2910
− 0.4704	− 0.2976	− 0.9510	− 0.9641	− 1.0646	− 0.4447
0.4827	− 0.0411	0.5867	− 2.2103	0.6225	− 1.6365
1.2208	1.3734	− 0.3243	− 0.7476	− 0.1389	− 1.1891
− 0.4263	− 1.4257	− 0.5514	1.4917	− 0.4424	0.6972
− 0.1742	2.0462	− 0.0564	− 0.7415	− 0.2823	− 0.1918
− 1.9735	− 0.8563	− 2.3751	1.7414	0.8237	0.6519
− 2.2837	− 1.3858	− 1.0777	− 3.7389	− 1.4988	0.3146
0.7962	− 0.0168	− 0.0134	0.2537	− 0.6213	− 0.4409
0.5078	− 0.1891	0.1523	0.3474	2.3227	− 0.6539
1.2930	1.7987	0.1447	− 1.5766	0.6171	− 1.1882
− 0.5129	− 0.8934	− 0.7651	− 2.5020	− 1.1147	0.3594
− 1.5520	0.0293	0.8194	0.1239	− 0.7265	− 1.5912
2.6078	1.4496	1.7027	− 0.6660	0.6225	− 0.4128

Table 3: (continued)

(a) Counter example to sufficiency			(b) Counter example to necessity		
Y	X	Z	Y	X	Z
0.6155	− 0.5828	0.3903	4.0091	0.7163	1.5748
− 0.0309	0.2132	0.3068	1.8579	0.3618	− 0.3873
− 0.3413	− 0.5022	− 0.6057	4.9534	1.6456	0.5523
− 1.3402	− 0.9463	− 1.7889	1.0457	0.2957	1.6746
− 0.0330	− 0.8767	0.2110	0.2615	1.5857	− 0.5250
0.1638	0.0743	0.2651	− 0.9870	− 0.3600	0.3656

Here, Z has no moderating effect on $b(=\beta_{yx})$ since $\sigma_{xz} = 0$ (compare Proposition 1 (ii)), but the regression coefficient of the constructed variable is nonzero. This shows that the vanishing of the regression coefficient of $W = X \cdot Z$ in [3] is also not a necessary condition for the lack of a moderating effect.

The moderated regression equations seem to stem from a false analogy to a result in analyses of variance. For these analyses it is known (compare Snedecor and Cochran (1967), chapter 16) that a vanishing interaction effect is a sufficient condition for the collapsibility of a main effect in two-way classifications, provided the numbers of observations in subclasses are equal or at least proportional.

References

Anderson, T. W. (1958) *An introduction to multivariate statistical analysis*. New York: Wiley.

Baron, R. M. and Kenny, D. A. (1986) The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Bishop, Y. M. M. (1971) Effects of collapsing multidimensional contingency tables. *Biometrics, 27*, 545–562.

Borkenau, P. (1985) Vergleiche einiger Verfahren zum Nachweis von Moderatoreffekten. *Z. Differentielle und Diagnostische Psychologie, 6*, 79–87.

Cleary, P. D. and Kessler, R. C. (1982) The estimation and interpretation of modified effects. *Journal of Health and Social Behaviour, 23*, 159–169.

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

Dalbert, C. and Schmitt, M. (1986) Einige Anmerkungen und Beispiele zur Formulierung und Prüfung von Moderatorhypothesen. *Z. Diff. Diag. Psych., 7*, 29–43.

Dempster, A. P. (1969) *Elements of continuous multivariate analysis*. Reading: Addison-Wesley.

Edwards, D. and Kreiner, S. (1983) The analysis of contingency tables by graphical models. *Biometrika, 70*, 553–563.

Goldberger, A. S. (1964) *Econometric Theory*. New York: Wiley.

Hodapp, V. (1984) *Analyse linearer Kausalmodell*. Bern: Huber.

Kiiveri, H., Speed, T. P. and Carlin, J. B. (1984) Recursive causal models. *J. Austral. Math. Soc.* (Series A), *36*, 30–52.

Table 4: Upper halves of the covariance matrix and of the concentration matrix of the variables Y, X, Z in Table 3a and of $W = X \cdot Z$

Covariance matrix					Concentration matrix				
	Variables					Variables			
Variables	Y	X	Z	W	Variables	Y	X	Z	W
Y	1.6674	0.9447	0.7665	− 0.2812	Y	1.5609	− 0.9365	− 0.9365	0.0000
X		1.0955	0.4790	− 0.2196	X		1.8128	− 0.1407	0.1535
Z			0.7984	− 0.2490	Z			2.3791	0.4588
W				0.6500	W				1.7660

Table 5: Upper halves of the covariance matrix and of the concentration matrix of the variables Y, X, Z in Table 3b and of $W = X \cdot Z$

Covariance matrix					Concentration matrix				
	Variables					Variables			
Variables	Y	X	Z	W	Variables	Y	X	Z	W
Y	4.2643	1.0798	0.8709	0.7776	Y	1.3639	− 2.0195	− 0.7255	− 2.5765
X		0.9374	0.0000	− 0.1631	X		4.1371	0.9420	4.2747
Z			0.8099	0.2330	Z			1.8394	0.6104
W				0.4739	W				7.5090

Krohne, H. W. Kohlmann, C. W. and Leidig, S. (1986) Erziehungsstildeterminanten kindlicher Ängstlichkeit, Kompetenzerwartungen und Kompetenzen. *Z. f. Entwicklungspsychologie und Pädagogische Psychologie, 18*, 70–88.
Lauritzen, S. L. and Wermuth, N. (1984) *Mixed interaction models*. Research Report, Institute of Electronic Systems. Aalborg: Aalborg University.
Lauritzen, S. L. and Wermuth, N. (1989) Graphical models for associations between variables, some of which are qualitative and some quantitative. *Annals of Statistics, 17*, 31–57.
Roos, P. E. and Cohen, L. H. (1987) Sex roles and social support as moderators of life stress adjustment. *J. of Personality and Social Psychology, 52*, 576–585.
Saunders, D. R. (1956) Moderator variables in prediction. *Educational and Psychological Measurement, 16*, 209–222.
Snedecor, G. W. and Cochran, W. G. (1967) *Statistical methods*. (6th ed.) Ames: The Iowa State University Press.
Steyer, R. (1983) Modelle zur kausalen Erklärung statistischer Zusammenhänge. In Bredenkamp, J. and Feger, H. (eds.). *Enzyklopädie der Psychologie, Forschungsmethoden der Psychologie*. Band 4, Göttingen: Hogrefe.
Weisberg, S. (1980) *Applied linear regression*. New York: Wiley.
Wermuth, N. (1987) Parametric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable. *J. Roy. Statist. Soc. B, 49*, 353–364.
Wermuth, N. (1989) Moderating effects of subgroups in linear models. *Biometrika, 76*, 81–92.
Wermuth, N. and Lauritzen, S. L. (1983) Graphical and recursive models for contingency tables. *Biometrika, 70*, 537–552.
Wermuth, N. and Lauritzen, S. L. On substantive research hypotheses, conditional independence graphs, and graphical chain models (with discussion). *J. Roy. Statist. Soc. B, 51* (to appear).
Whittemore, A. S. (1978) Collapsibility of multidimensional contingency tables. *J. Roy. Statist. Soc. B, 40*, 328–340.
Zedeck, S. (1971) Problems with the use of "moderator" variables. *Psychological Bulletin, 76*, 295–310.