

Interpretation of intrinsic and extrinsic structural relations by path analysis: theory and applications to assortative mating

BY C. ROBERT CLONINGER

Departments of Psychiatry and Genetics, Washington University School of Medicine, and the Jewish Hospital of St Louis, 4940 Audubon Ave., St Louis, Missouri 63110

(Received 7 April 1980)

SUMMARY

The theory of path analysis is extended by considering a multivariate system of correlations from a dual perspective. Intrinsic factors exert a unidirectional influence on both the variance and covariance of dependent factors. In contrast, extrinsic factors have a bidirectional influence on the covariance structure of both antecedent and dependent factors and do not influence intrinsic variability. The mathematical model assumes a formally complete linear system of unitary factors. A coefficient is defined to quantify the influence of adventitious associations and is called a copath. Copaths are compared to path coefficients and to correlations due to common antecedents. The chain properties of these coefficients are derived along with a general formula and computational algorithm. The method is illustrated for multifactorial inheritance in extended pedigrees in the presence of different types of assortative mating.

1. INTRODUCTION

Path analysis is applicable to a broad class of problems, but some unexpected limitations of the classical method (Wright, 1918 *et seq.*) have been encountered in recent work on assortative mating (Rice, Cloninger & Reichs, 1978; Cloninger, Rice & Reich, 1979*a, b*; Rao, Morton & Cloninger 1979). Extension of classical path analytic work (Wright, 1921, 1978; Reeve, 1953) to consider phenotypic homogamy in three or more generations required the use of a variety of new conventions that appear arbitrary or cumbersome. These included the use of reverse paths (Wright, 1978; Cloninger *et al.* 1979*a, b*), duplication of variables and brackets (Wright, 1978), and parallel paths (Rao *et al.* 1979). Further study has revealed that it is possible to extend the theory of path analysis to provide a simpler and more flexible approach to the fundamental problem that emerged in studying assortative mating. This theoretical extension is described here along with simple but general computational algorithm. The extended method is illustrated by application to several types of assortative mating.

2. THE DISTINCTION BETWEEN INTRINSIC AND EXTRINSIC STRUCTURAL RELATIONS

Path analysis was developed for the interpretation of multivariate systems of correlations (Wright, 1918, *et seq.*). In the classical method a particular model of the functional relationships among the variables is specified by a qualitative diagram in which every variable (whether measured or hypothetical) is represented either as additively and completely determined by certain antecedent factors or as an ultimate factor. The functional relationship between a dependent and an antecedent variable is quantified by a path coefficient which is a standardized partial regression coefficient and is represented in the path diagram by a unidirectional arrow (\rightarrow) pointing toward the dependent factor. A correlation between variables due to determination by a common antecedent factor is represented either by adjacent paths pointing away from the specified antecedent X ($\leftarrow X \rightarrow$), or by a bar with arrowheads at both ends (\leftrightarrow) when a common antecedent is implied but not specified.

In the original mathematical model correlations among dependent factors could only be represented as due to common antecedent factors which necessarily contribute to the variance of the dependent factors. Also correlations between factors were assumed to influence the relations of subsequent dependent factors but not antecedent factors. In this paper the mathematical model is extended to allow for correlations due to adventitious associations, which influence the covariance structure of both antecedent and subsequent factors but do not contribute to their variances. These extrinsic or adventitious correlations are represented by a headless bar (---) in path diagrams and distinguished from intrinsic correlations, which are represented by a double-headed arrow (\leftrightarrow) indicating common antecedents. Such distinctions permit the consideration of structural relations from the dual perspective of a system of intrinsic relationships, which are assumed to be essential in all populations, and extrinsic relationships, which may be absent or vary between populations.

The choice of intrinsic variables may be arbitrary in purely mathematical applications, but the most important application is in etiological evaluations involving both natural causation and adventitious associations. Natural causation is the prototype of intrinsic structural relations with particular properties including a specific direction in time and space. Specific examples of natural causes include transfers of momentum along a succession of impacts, transfers of energy by electromagnetic waves at velocities never exceeding that of light, chains of physiological reactions, transfers of genes along pedigree lines, chains of ecological events, ontogenetic successions, etc. In contrast, adventitious associations involve only non-random pairing of factors due to artificial control or chance drift in finite populations. Examples include phenotypic assortative mating, the correlation between linked loci of uniting gametes in a foundation stock (as produced by crossing homoallelic strains of $A_1A_1B_1B_1$ with $A_2A_2B_2B_2$), gametic disequilibrium, and most commonly, controlled experiments in which factors are selectively matched.

3. THE BASIC MATHEMATICAL MODEL

The properties of path and correlation coefficients are most easily deduced from consideration of linear functions of unitary factors. Let us consider the two unitary factors X and Y , shown in figure 1. Variables X and Y are completely determined by the sets of intrinsic antecedents $\{B, C, D\}$ and $\{F, G, H\}$ respectively. Unspecified common antecedents are shared by B and C and by F and G . The variables X and Y have no common antecedent but are correlated due to an extrinsic source of influence. The latter adventitious correlation is represented by the headless bar labelled a , but the source of the correlation, a hypothetical

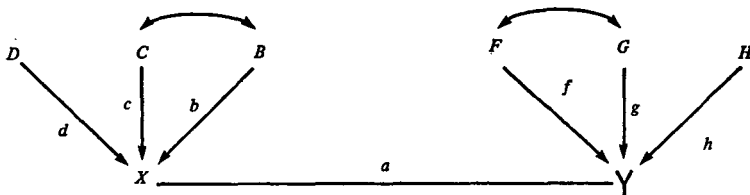


Fig. 1. Path diagram of two factors X and Y with only a direct extrinsic correlation.

variable A , is not shown in Fig. 1. In other words r_{XY} is an adventitious correlation such that in a subpopulation in which X and/or Y are (is) fixed at a particular value, neither of the sets $\{B, C, D\}$ and $\{F, G, H\}$ has any member that is correlated with a member of the other set. Without loss of generality for convenience we assume that each variable is standardized to have mean 0 and unit variance.

The mathematical model involves three basic assumptions: (i) *additive or linear structural relations among all factors* (whether intrinsic or extrinsic); (ii) *in all structural relations each factor acts as a unitary whole* rather than as a composite variable in which one part is more significant in one relationship than in another; and (iii) *complete determination of the variance of intrinsic factors by antecedent factors only*.

Assumptions (i) and (ii) are identical to those of classical path analysis and assumption (iii) differs only in that the possibility of extrinsic influences is admitted. Assumption (iii) provides the assumed fundamental distinction between intrinsic and extrinsic factors from which other differences in properties will now be deduced.

4. INTRINSIC FACTORS BUT NOT EXTRINSIC FACTORS DETERMINE THE VARIANCES.

From the assumption of additivity, the relations between the dependent variables X and Y and their antecedents may be specified in terms of the standardized linear regression equations

$$X = p_{XB}B + p_{XC}C + p_{XD}D = bB + cC + dD, \tag{1}$$

$$Y = p_{YF}F + p_{YG}G + p_{YH}H = fF + gG + hH. \tag{2}$$

From the assumption that X and Y are completely determined by their intrinsic

antecedents, it follows that their variances are unchanged in the absence of the adventitious correlation. This property of adventitious relations may be designated as $V_{X^*A} = V_X$ and is read as the variance of X conditional on the absence of the adventitious association A equals the (unconditional) variance of X . For example, when the choice of mates' phenotypes is randomized rather than assortative the phenotypic variances remain the same in mates (Reeve, 1953) even though the variances may change in subsequent generations. In other words, the relationship of a variable with itself involves only intrinsic relations whereas an adventitious correlation involves extrinsic influences by definition. This is equivalent to specifying that (1) and (2) are conditional on the absence of any adventitious correlation. Thus, from (1), the correlation between X and Q in the absence of the association A is

$$r_{XQ^*A} = br_{BQ^*A} + cr_{CQ^*A} + dr_{DQ^*A} = \sum p_{XI} r_{IQ^*A} \tag{3}$$

where $I = \{B, C, D\}$. An important special case of (3) occurs when $Q = X$:

$$r_{XX^*A} = \sum p_{XI} r_{IX^*A} = s^2_{X^*A} \equiv s^2_X \equiv 1, \tag{4}$$

where s^2 denotes the variance of the subscripted standardized variable. For example, in Fig. 1, $s^2_X = s^2_{X^*A} = b(b + cr_{BC}) + c(c + br_{BC}) + d^2 = 1$. This shows that the assumption that intrinsic antecedents completely determine a dependent factor necessarily implies that adventitious correlations change only the covariance structure of a system and not the variances.

5. EXTRINSIC FACTORS EXERT A BIDIRECTIONAL INFLUENCE ON COVARIANCE STRUCTURE

The influence of adventitious correlations on the covariances is easily deduced in the simple case shown in Fig. 1 where no pair of variables have both common antecedents and adventitious correlations. The validity of path analysis requires the assumption that composite variables act as unitary wholes in all their structural relations. It necessarily follows from this that extrinsic correlations influence the covariances of both antecedents and subsequent factors. From (1) and (2), denoting expectations by $E[]$,

$$r_{XY} = E[(bB + cC + dD)(fF + gG + hH)] \tag{5}$$

$$= b(fr_{BF} + gr_{BG} + hr_{BH}) + c(fr_{CF} + gr_{CG} + hr_{CH}) + d(fr_{DF} + gr_{DG} + hr_{DH}).$$

Since $r_{XY} = a$ by definition also, the correlations among the antecedents (components) of X and those of Y must be proportional to the adventitious correlation between X and Y and to the intrinsic correlations between the antecedent and dependent factors. Specifically the correlation between any antecedent of X , designated as Q where $Q = \{B, C, D\}$, and any antecedent of Y , designated as R where $R = \{F, G \text{ or } H\}$, is $r_{QX^*A} r_{XY} r_{RX^*A}$, or, in this special case where there are no common antecedents, simply $r_{QX^*A} r_{XY} r_{RY}$. In Fig. 1 this may be proven

from the assumption of the conditional independence of Q and R given fixed values of X and/or Y . It may be verified in Fig. 1 that

$$\begin{aligned} r_{BF} &= (b + cr_{BC}) a(f + gr_{FG}), & r_{BG} &= (b + cr_{BC}) a(g + fr_{FG}), & r_{BH} &= (b + cr_{BC}) ah, \\ r_{CF} &= (c + br_{BC}) a(f + gr_{FG}), & r_{CG} &= (c + br_{BC}) a(g + fr_{FG}), & r_{CH} &= (c + br_{BC}) ah, \\ r_{DF} &= da(f + gr_{FG}), & r_{DG} &= da(g + fr_{FG}), & r_{DH} &= dah. \end{aligned}$$

Recalling from (4) that $r_{XX} = b^2 + c^2 + 2bcr_{BC} + d^2$ and $r_{YY} = f^2 + g^2 + 2fgr_{FG} + h^2$ and that these equations of complete determination equal unity, substitution in (5) confirms that $r_{XY} = a$. On inspection of the path diagram in Fig. 1, the correlation between any two variables is readily seen to be the sum of all compound chains that may be traced without passing through adjacent arrowheads.

6. DEFINITION OF COPATHS IN COMPLEX ADDITIVE SYSTEMS

The system is said to be complex if the same pair of variables have both common antecedents and adventitious associations (direct and/or indirect). Nevertheless, the property of additivity of multiple compound chains may be preserved by the appropriate definition of an adventitious correlation to eliminate any overlap or redundancy with the contributions of intrinsic antecedents and indirect associations.

In Fig. 2 the system shown in Fig. 1 is extended to include a variable K , which is a common antecedent of components of X and Y . Also an extrinsic association m between components of X and Y is present due to the influence of a hypothetical extrinsic variable M not shown in the figure. The correlation between X and Y involves the correlation due to the common antecedent K (namely $bkk'f$), the correlation due to the indirect association (namely dmh), and the direct association a . Specifically, assuming additivity of separate compound chains,

$$r_{XY} = bkk'f + dmh + a, \tag{6a}$$

so that

$$a = r_{XY} - bkk'f - dmh \equiv \text{cov}_{XY \cdot K \cdot M} / \sigma_X \sigma_Y, \tag{6b}$$

where $\sigma_X \sigma_Y$ is the geometric mean of the total variances of the unstandardized variables. Similarly

$$r_{DH} = dah + m \text{ so that } m = r_{DH} - dah \equiv \text{cov}_{DH \cdot A} / \sigma_D \sigma_H, \tag{7}$$

where $\text{cov}_{DHA \cdot} = \text{cov}_{DH \cdot X} = \text{cov}_{DH \cdot Y}$ and $\sigma_D \sigma_H$ is the geometric mean of the total variances of the unstandardized variables. In other words (6) and (7) define a coefficient to quantify extrinsic influences so that the additivity assumption is preserved.

In general, assuming additivity, an adventitious association is defined as the fraction of the correlation coefficient due to direct association only. That is, an adventitious association between X and Y is the conditional correlation in which all correlated intrinsic antecedents are fixed and the influence of indirect associations is absent but the variance of X and of Y are as great as in the total

population. An adventitious association differs from an unconditional correlation in that the numerator is a partial covariance; it differs from a partial correlation in that the denominator is the product of the unconditional standard deviations. This definition insures that no correlation exceeds the possible range + 1 to - 1.

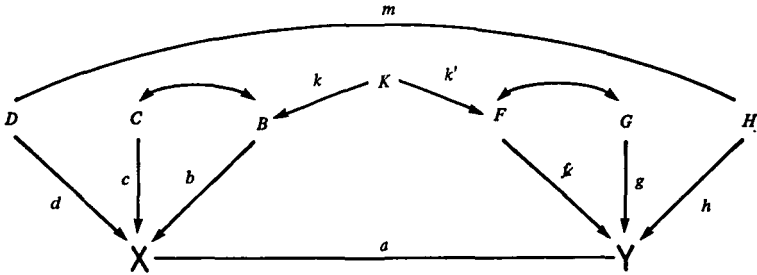


Fig. 2. Path diagram of two factors X and Y with both common antecedents and extrinsic correlations (direct and indirect).

7. CHAIN PROPERTIES OF CORRELATIONS AND COPATHS

The relationship between two factors A and B that are each correlated with a common factor C depends on the nature of the correlations r_{AC} and r_{BC} . For example in Fig. 1 we observed that $C \leftrightarrow X - Y$ implies a correlation $r_{CY} = r_{CX}r_{XY}$, where r_{XY} is an adventitious correlation; however, in contrast $C \leftrightarrow X \leftarrow D$ implies no correlation between C and D. In general the relationship of two factors each correlated with a common factor is uncertain until more is specified about the nature of the correlations.

This specification may be approached from the perspective of two alternate conventions depicted in Fig. 3. In Fig. 3 factors A and B are each correlated with factor C. In case I we assume unconditional independence ($r_{AB} = 0$) whereas in case II we assume conditional independence only ($r_{AB.C} = 0$). Thus, case I is convenient when the correlations are due to unspecified antecedents shared by A and C and shared by B and C, as assumed in classical path analysis. However, case II is convenient when at least one correlation is adventitious and thereby conditionally (but not unconditionally) independent of the other correlation. In case II we assume $r_{AB.C} = r_{AB} - r_{AC}r_{BC} = 0$ and so $r_{AB} = c_{AC}c_{BC}$ whereas in case I we assume $r_{AB} = 0$.

In the most general case where $r_{AB.C}$ is not negligible, we have two alternative conventions. Proceeding from case I, we let $r_{AB} \neq 0$ and may draw an additional two-headed arrow (\leftrightarrow) connecting A and B, as does Wright (1918, *et seq.*). Proceeding from case II, we may let $c_{AB} \neq 0$ and $r_{AB} = c_{AC}c_{BC} + c_{AB}$. Wright would prefer that the assumption of unconditional independence be maintained for all correlations, but this is not possible for chains of intrinsic and extrinsic correlations such as $B \leftrightarrow X - Y \leftrightarrow F$ in Fig. 1. Accordingly it is most convenient to reserve the extension of case I for correlations symbolized by a two-headed arrow (\leftrightarrow) and the extension of case II for correlations symbolized by a headless bar ($-$). Thereby the two-headed arrow symbolizes the most restricted type of correlation.

When the extension of case II is assumed as a general convention for chains of adventitious correlations these coefficients will be referred to as copaths and symbolized as c_{XY} . It should be noted that Wright has used the symbol c otherwise to denote concrete regressions, but such regressions are more commonly denoted by the symbol b . The term copath is chosen for this convention in view

(I) Unconditional independence

(II) Conditional independence



Fig. 3. Alternative conventions about the relations of two factors, each correlated with the same third factor.

of the continuity of the underlying mathematical model with classical path analysis and because the coefficient has properties of both correlations and paths in a unique combination that is equivalent to its being a symmetrical or reversible path. Under this convention a compound copath c_{XYZ} is assumed to be the product of the elementary copaths $c_{XY}c_{YZ}$. This provides a useful basis for denoting the source of an association as a hypothetical intermediate variable. Thus in Fig. 2, if M is the hypothetical source of the association between D and H , $m = c_{DHM} = c_{DM}c_{MH}$, where $c_{DM} = \sqrt{m} = c_{MH}$.

8. PHENOTYPIC ASSORTATIVE MATING

Compound chains involving copaths such as $\rightarrow - \leftarrow$ and $\leftrightarrow - \leftrightarrow$ have the novel features of multiple correlations in the same compound chain (which are separated by an intermediate copath) and opposing arrowheads (which are not adjacent). This is illustrated in Fig. 4, which depicts multifactorial inheritance with phenotypic assortative mating. For purposes of emphasis the relations between P , G and B are represented as correlations for the mother (M) and as the natural causal paths for the father (F), but both are equivalent.

Consider the correlation between the mother's phenotype P_M and the child's phenotype P_K . The correlation between the child's genic value G_K and the mother's phenotype may be decomposed into chains of paths and correlations ($G_K \leftarrow G_M \leftrightarrow P_M$) and compound chains of paths, correlations, and copaths ($G_K \leftarrow G_F \leftrightarrow P_F \leftarrow P_M$):

$$r_{G_K P_M} = \frac{1}{2}\gamma + \frac{1}{2}(h + wb) p = \frac{1}{2}\gamma(1 + p) \tag{9}$$

as obtained from Fig. 4 following the rule of summing all chains without passing through adjacent arrowheads. Similarly the correlation between the child's cultural value B_K and the mother's phenotype is

$$r_{B_K P_M} = \beta_M \phi + \beta_F(b + wh) p = \phi(\beta_M + p\beta_F). \tag{10}$$

From (9) and (10) the mother-child phenotypic correlation is simply

$$r_{P_M P_K} = hr_{G_K P_M} + br_{B_K P_M}. \tag{11}$$

The correlation between the grandchild's phenotype P_0 and the grandmother's phenotype P_M also may be decomposed into compound chains of paths and correlations (\leftrightarrow) and compound chains of intrinsic and adventitious correlations

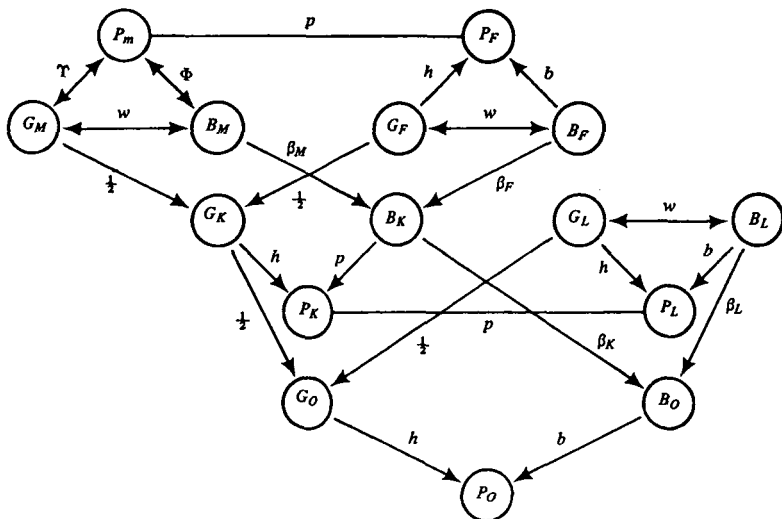


Fig. 4. Multifactorial inheritance with phenotypic assortative mating: phenotypes (P), genotypes (G) and heritable cultural factors (B) shown for the mother (M), father (F), child (K), child's mate (L), and grandchild (O). $\gamma = h + wb$, $\phi = b + wh$.

(\leftrightarrow — \leftrightarrow). The correlation between the grandchild's genic value G_0 and the grandmother's phenotype is the sum of the chain $G_0 \leftarrow G_K \leftrightarrow P_M$ and the chain $G_0 \leftarrow G_L \leftrightarrow P_L \leftrightarrow P_K \leftrightarrow P_M$:

$$r_{G_0 P_M} = \frac{1}{2}r_{G_K P_M} + \frac{1}{2}\gamma\phi r_{P_K P_M}. \tag{12}$$

The correlation between B_0 and P_M is

$$r_{B_0 P_M} = \beta_K r_{B_K P_M} + \beta_L \phi \rho r_{P_K P_M} \tag{13}$$

and between P_0 and P_M , using (12) and (13),

$$r_{P_0 P_M} = hr_{G_0 P_M} + br_{B_0 P_M}. \tag{14}$$

General expressions for vertical and collateral relatives of any degree of relationship are derived elsewhere for various family structures and types of assortative mating (Cloninger *et al.* 1979a, b; Rao *et al.* 1979).

Using copaths it is clear that phenotypic assortative mating induces a correlation between cognate relatives (R) of one spouse (X) and cognate relatives (S) of the other spouse (Y). Quantifying the extent of phenotypic assortment as the copath c_{XY} , the remote affinate correlation r_{RS} is $r_{RX}c_{XY}r_{YS}$. Prior formulations do not permit derivation of this correlation.

For pure phenotypic assortment the copath p between mates' phenotypes is also the correlation between mates. However, if the copath p were represented as a two-headed arrow instead of a headless bar, the contribution of the mates would be spuriously neglected unless tracing through adjacent arrowheads was permitted. Other spurious results occur if the association between mates is represented by simultaneous reciprocal causation with a path m from man to woman and f from woman to man: then the full covariance between mates is not taken into account in extended pedigrees and, if $m \neq f$, the grandparent/grandchild correlations depend on the sex of the intermediate offspring even for a purely polygenic trait.

9. SERIAL PHENOTYPIC MONOGAMY

Serial monogamy (repeated divorce and remarriage to another) illustrates the interpretation of a series of multiple associations based on phenotypes at different times. In Fig. 5, M_1 is married to F_1 and F_2 is married to M_2 at time t_1 . After divorce and subsequent change in phenotypes over time (denoted by paths m and f)

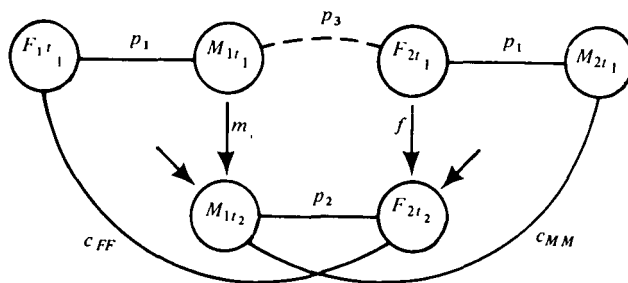


Fig. 5. Serial phenotypic monogamy: mates include M_1 with F_1 and F_2 with M_2 at time t_1 , then M_1 with F_2 at time t_2 after change in phenotype (paths m and f). Redundant copath p_3 in dashed line.

m and f), M_1 and F_2 marry. Since the phenotypes of prior mates may influence subsequent mate selection, the copath between different mates of the same individual may be non-zero. The extent of phenotypic assortment for first (unstable) marriages may differ from that for subsequent marriages (Cattell & Nesselrode, 1967) so $p_1 \neq p_2$ in general. Phenotypes may vary over time, and the correlation between future mates p_3 is mfp_2 , which is less than p_2 as expected. The correlation between two mates of the same man is $r_{FF} = p_1p_2m + c_{FF}$ whereas that between different mates of the same woman is $r_{MM} = p_1p_2f + c_{MM}$. Hence r_{FF} and r_{MM} are likely to be unequal unless $m = f$ and $c_{FF} = c_{MM}$.

10. MULTIPLE CORRELATED TRAITS

Assortative mating may be determined both by associations directly based on the marital phenotype (phenotypic or primary homogamy) and by correlations among the causes of the phenotype (secondary homogamy). Secondary homogamy

may be due to either associations among additional traits which are correlated to the primary phenotype under consideration or natural causal effects such as inbreeding and social stratification (Wright, 1978; Rao *et al.*, 1979) sometimes collectively called social homogamy. Wright proposed a model of social homogamy which assumed a unitary common cause of both the genic and cultural correlations between mates (Wright, 1978). The simultaneous combination of phenotypic homogamy and social homogamy is called mixed homogamy and has been extended to the treatment of multigeneration pedigrees by Rao, Morton, and Cloninger (1979). Although not initially obvious, it will be shown here that prior models of social homogamy are mathematically equivalent to phenotypic assortment for multiple correlated traits, as depicted in figure 6. The association directly between the phenotypes P_M and P_F is shown by a compound copath p involving the hypothetical source called primary homogamy (H_1). The causes of the primary phenotype are also correlated due to an association for another correlated trait S , such as social class or various personality factors. The trait S is correlated with P due to paths from G , B and E denoted as g/α , v/α , and f/α so that the compound copath $c_{S_M H_2 S_F} \equiv \alpha^2$ is incidental unless S is an observed trait. The residual factors of S may include either genetic or environmental factors uncorrelated with P . Hence the correlation between mates is $\mu = p + \mu^*$, where μ^* is the contribution of secondary homogamy, defined as $(\mu^* H_1) = (\mu | p = 0)$. From the basic algorithm that the correlation between any two variables is the sum of all compound chains that may be traced without passing through adjacent arrowheads, it may be seen that

$$\mu^* = h^2 m + b^2 u + e^2 f^2 + 2hbs,$$

where

$$m \equiv r_{G_M G_F * H_1} = (g + wv)^2$$

$$u \equiv r_{B_M B_F * H_1} = (v + wg)^2 \quad \text{and} \quad s \equiv \sqrt{mu}.$$

If multiple unobserved secondary phenotypes S_i with $i = 1, 2, \dots, n$ are associated, m and u are actually the sum of several chains $\Sigma(g_i + wv_i)^2$ respectively. Whether S is a unitary or composite variable, these expressions for m and u may be solved for g and v , showing the mathematical equivalence with other superficially different models (19):

$$v = (\sqrt{u} - w\sqrt{m}) / (1 - w^2) \quad \text{and} \quad g = (\sqrt{m} - w\sqrt{u}) / (1 - w^2).$$

The use of copaths obviates the need for special conventions or derivation of hypothetical marital path coefficients (Rao *et al.* 1979) because compound chains involving copaths are functionally equivalent to the relevant unidirectional path coefficients; specifically, $X \rightarrow \text{---} \leftrightarrow Y$ is equivalent to $X \rightarrow Y$. This may be seen in Fig. 5, where

$$m^* \equiv p_{G_F G_M} = p_{G_L G_K} = g(g + wv) = g\sqrt{m},$$

$$u^* \equiv p_{B_F B_M} = p_{B_L B_K} = v(v + wg) = v\sqrt{u},$$

$$s^* \equiv p_{B_F G_M} = p_{B_L G_K} = g(v + wg) = g\sqrt{u},$$

$$t^* \equiv p_{G_F B_M} = p_{G_L B_K} = v(g + wv) = v\sqrt{m},$$

agreeing with derivations by a different method (Rao *et al.* 1979).

From Fig. 6 the correlation between vertical relatives separated by n generations may be shown to be

$$r_{Vni} = \frac{1}{2}[\gamma(1+p) + hm + bs] K_n + [\phi(\beta_i + p\beta_j) + \beta_j(hs + bu)] \Lambda_n,$$

where for parent/child relations $K_1 = h$ and $\Lambda_1 = b$, and for $n > 1$,

$$K_n = \frac{1}{2}(1 + m^* + \gamma ph) K_{n-1} + \beta_L(s^* + \phi ph) \Lambda_{n-1},$$

$$\Lambda_n = [\beta_K + \beta_L(u^* + \phi pb)] \Lambda_{n-1} + \frac{1}{2}(t^* + \gamma pb) K_{n-1},$$

as previously shown by a more cumbersome method (19). Further extensions to consider cultural inheritance (e.g. paths from P_M or S_M to B_K) and multiple correlated traits are straightforward using copaths.

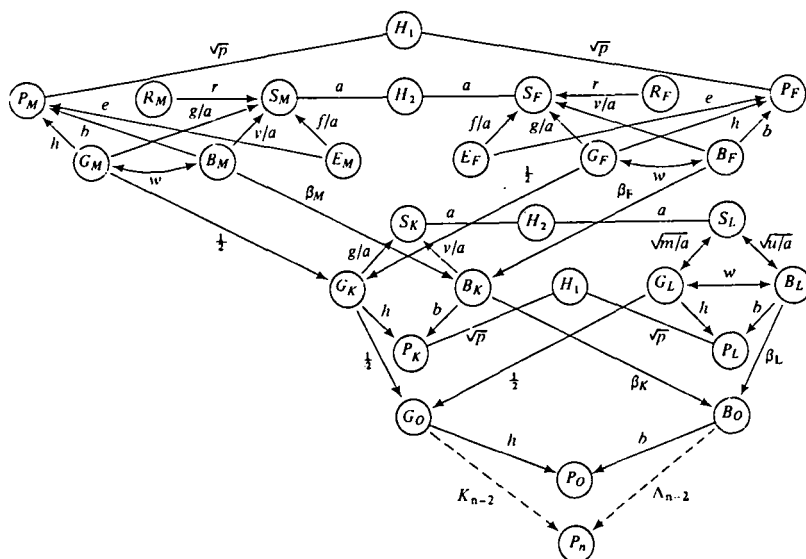


Fig. 6. Multifactorial inheritance with assortative mating for both the primary phenotype (P) and a correlated trait (S) such as social class.

11. GENERAL FORMULA AND ALGORITHM

Let us consider two unitary factors X and Y in which X has a set of component factors $I = \{1, 2, \dots, i, i+1, \dots, n\}$ and Y has a set $J = \{1, 2, \dots, j, j+1, \dots, m\}$. Variable X is fully determined by i antecedents and has $(n-i)$ dependent factors. Variable Y is fully determined by j antecedents and has $(m-j)$ dependent factors. As in (12), any correlation r_{XY} may be decomposed into compound chains of paths and correlations ($\leftarrow \leftrightarrow$) and compound chains of intrinsic and extrinsic correlations ($\leftrightarrow \text{---} \leftrightarrow$); that is,

$$r_{XY} = \sum_{I=1}^i p_{XI} r_{YI^*A} + \sum_{\substack{I=X,1 \\ X \neq Y}}^n \sum_{J=Y,1}^m r_{XI^*A} c_{IAJ} r_{YJ^*A}, \tag{15}$$

where c_{IAJ} denotes the matrix of associations between nm pairs of antecedent

and dependent factors of X and Y , n components of X itself, m components of Y with X itself, and directly between X and Y (when $X = I$ and $Y = J$). The conditions on the correlations (denoting absence of association by an asterisk) serve to eliminate any redundant chains involving adventitious associations. We define $c_{XX} \equiv 0 \equiv c_{II}$ since associations by definition involve extrinsic influences and the relationship of X or I with itself involves only intrinsic relations; thus letting $X = Y$ (and so $I = J$), (15) reduces to

$$r_{XX} = \sum p_{XI} r_{YI^*A} = 1, \quad (16)$$

which is also Wright's equation of complete determination in the absence of association since extrinsic factors do not contribute to variances ($V_{X^*A} = V_X$).

The residual correlations may themselves be decomposed by applying the general formula in a stepwise fashion. Such repeated applications lead to a simple algorithm for deriving correlations from inspection of a path diagram for any fully recursive system: *The correlation between any two variables may be obtained as the sum of all compound chains of paths, copaths, and correlations leading from one to the other without passing through adjacent arrowheads and without passing through the same variable twice in the same compound chain.* Simultaneous reciprocal interaction (Wright, 1960) may be analysed by repeated application of the general formula in a stepwise fashion just as a path diagram is inspected.

The simplicity of this algorithm strongly justifies the choice of a headless bar to represent extrinsic correlations. The rule against tracing 'first forward and then backward' for unidirectional determination is appropriate but superfluous (since this would require passing through adjacent arrowheads) and, in the presence of bidirectional association, would lead to spuriously neglecting some induced correlations.

12. DISCUSSION

The distinction between intrinsic and extrinsic factors allows for the interpretation of a multivariate system for two perspectives simultaneously. Intrinsic factors have a specific unidirectional influence and contribute to the variance of their dependent factors. In contrast extrinsic factors have a bidirectional influence on the covariance structure of both antecedent and subsequent factors but do not contribute to the intrinsic variability of any factor. These concepts underly the definitions and symbols described here for paths (\leftarrow), intrinsic correlations (\leftrightarrow), and copaths or extrinsic correlations (---). In etiological applications these distinctions have an obvious relevance to the differentiation of natural causation and adventitious associations, as illustrated here for multifactorial inheritance and assortative mating.

These distinctions greatly simplify the formulation of models involving correlations among dependent variables by eliminating the need for reverse paths, brackets, and duplication of variables to represent adventitious association. In view of confusion about the differences between causation and association by many statisticians and geneticists, it is hoped that the operational distinctions given here

will aid understanding. Since adventitious associations are likely to occur in the small local populations which are the subject of most observation and experimentation, the failure to allow for both intrinsic and extrinsic influences would needlessly limit the appropriate application of path analysis. Classical path analysis has already proven to be a powerful and flexible technique in genetics and other natural and social sciences (Li, 1975). The extensions described here should further increase its applicability by clarifying the distinction between intrinsic and extrinsic determination and by simplifying their simultaneous quantitative analysis.

Population Genetics Lab MS. no. 224. This report was initiated while Dr Cloninger was Visiting Investigator (1978–9) at the University of Hawaii at Manoa, supported in part by Research Scientist Development Award MH-00048 and USPHS grants MH-31302 and AA-03539.

I am particularly thankful to Sewall Wright for many constructive comments on several earlier drafts of this paper. The consultation of N. E. Morton, J. Rice, M. Taibleson, P. VanEerdwegh, R. Wette, and S. Yokoyama was also helpful. Without the advice and encouragement of D. C. Rao and T. Reich, I would not have persevered in my efforts to extend my early complicated results to the simple and general formulation presented here. Any errors or obscurity that remain are my sole responsibility.

REFERENCES

- CATTELL, R. B. & NESSELRODE, J. R. (1967). Likeness and completeness theories examined by sixteen personality factor measures on stably and unstably married couples. *Journal of Personality and Social Psychology* 7, 351–361.
- CLONINGER, C. R., RICE, J., & REICH, T. (1979a). Multifactorial inheritance with cultural transmission and assortative mating. II. A general model of combined polygenic and cultural inheritance. *American Journal of Human Genetics* 31, 176–189.
- CLONINGER, C. R., RICE, J. & REICH, T. (1979b). Multifactorial inheritance with cultural transmission and assortative mating. III. Family structure and the analysis of experiments. *American Journal of Human Genetics* 31, 366–388.
- LI, C. C. (1975). *Path Analysis – A Primer*, pp. 1–346. Pacific Grove, Calif: Boxwood Press.
- RAO, D. C., MORTON, N. E. & CLONINGER, C. R. (1979). Path analysis under generalized assortative mating. I. Theory. *Genetical Research* 33, 175–188.
- REEVE, E. C. R. (1953). Studies in Quantitative Inheritance. III. Heritability and genetic correlation in progeny tests using different mating systems. *Journal of Genetics* 51, 520–542.
- RICE, J., CLONINGER, C. R. & REICH, T. (1978). Multifactorial inheritance with cultural transmission and assortative mating. I. Description and basic properties of the unitary models. *American Journal of Human Genetics* 30, 618–643.
- WRIGHT, S. (1918). On the nature of size factors. *Genetics* 3, 367–374.
- WRIGHT, S. (1921). Correlation and causation. *Journal of Agricultural Research, Washington* 20, 557–585.
- WRIGHT, S. (1923). The theory of path coefficients – a reply to Niles' criticism. *Genetics* 8, 239–255.
- WRIGHT, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 5, 161–215.
- WRIGHT, S. (1954). The interpretation of multivariate systems. In *Statistics and Mathematics in Biology*, (ed. O. Kempthorne, T. A. Bancroft, J. W. Gowen and J. L. Lush), pp. 11–33.
- WRIGHT, S. (1960). The treatment of reciprocal interaction, with or without lag by path analysis. *Biometrics* 16, 423–445.
- WRIGHT, S. (1969). *Evolution and Genetics of Populations*, vol. 1, pp. 299–372. University of Chicago Press.
- WRIGHT, S. (1978). *Evolution and the Genetics of Populations*, vol. 4, pp. 366–438. University of Chicago Press.