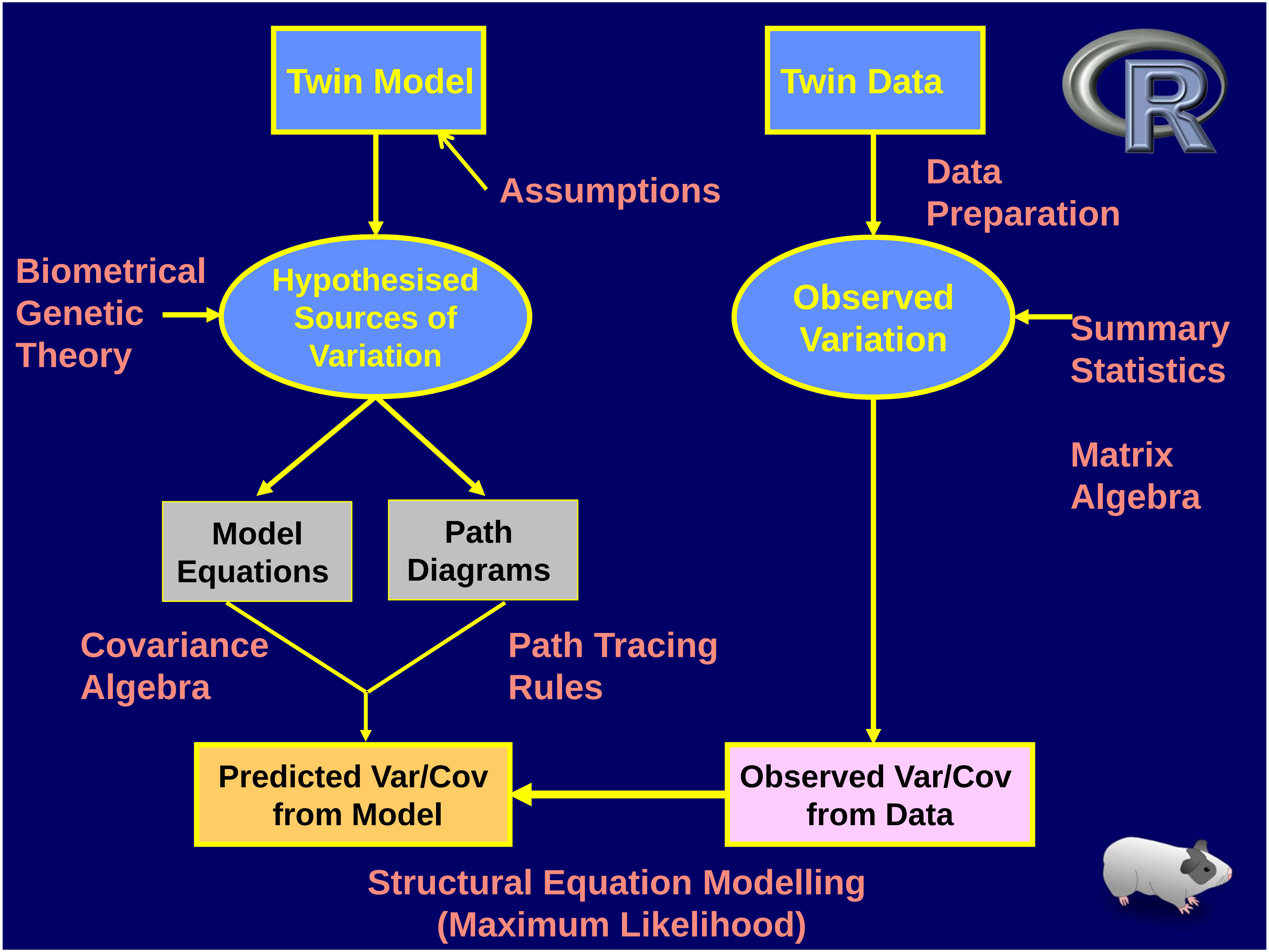
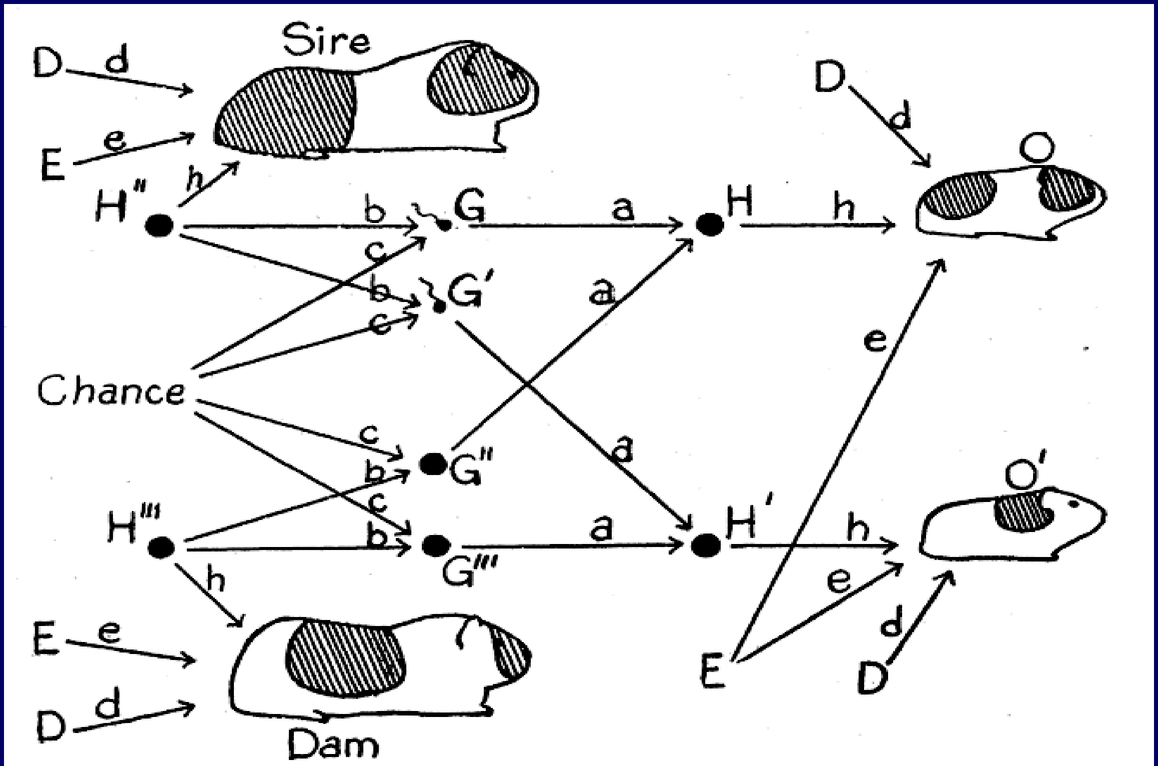
Path Analysis

Frühling Rijsdijk

SGDP Centre
Institute of Psychiatry
King's College London, UK


Twin Model
Twin Data
Assumptions
Data Preparation
Biometrical Genetic Theory
Hypothesised Sources of Variation
Observed Variation
Summary Statistics
Matrix Algebra
Model Equations
Path Diagrams
Covariance Algebra
Path Tracing Rules
Predicted Var/Cov from Model
Observed Var/Cov from Data
Structural Equation Modelling (Maximum Likelihood)

Path Analysis



Developed by the geneticist Sewall Wright (1920)

Now widely applied to problems in genetics and the behavioural sciences.

Path Analysis

This technique allows us to present linear relationships between variables in diagrams and to derive predictions for the variances and covariances of the variables under the specified model.

The relationships can also be represented as structural equations and covariance matrices
All three forms are mathematically complete, it is possible to translate from one to the other.

Structural equation modelling (SEM) represents a unified platform for path analytic and variance components models.

In (twin) models, expected relationships between observed variables are expressed by:

- **A system of linear model equations**

  or

- **Path diagrams which allow the model to be represented in schematic form**

Both allow us to derive predictions for the variances and covariances of the variables under the specified model

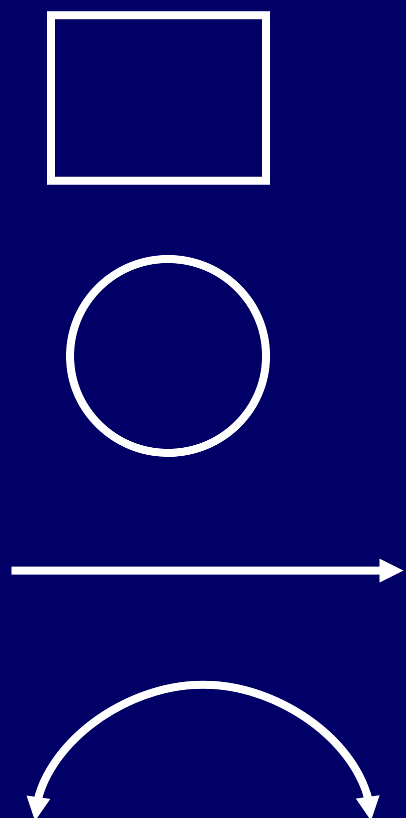
Aims of this Session

Derivation of Predicted Var-Cov matrices of a model using:

(1) Path Tracing

(2) Covariance Algebra

Path Diagram Conventions



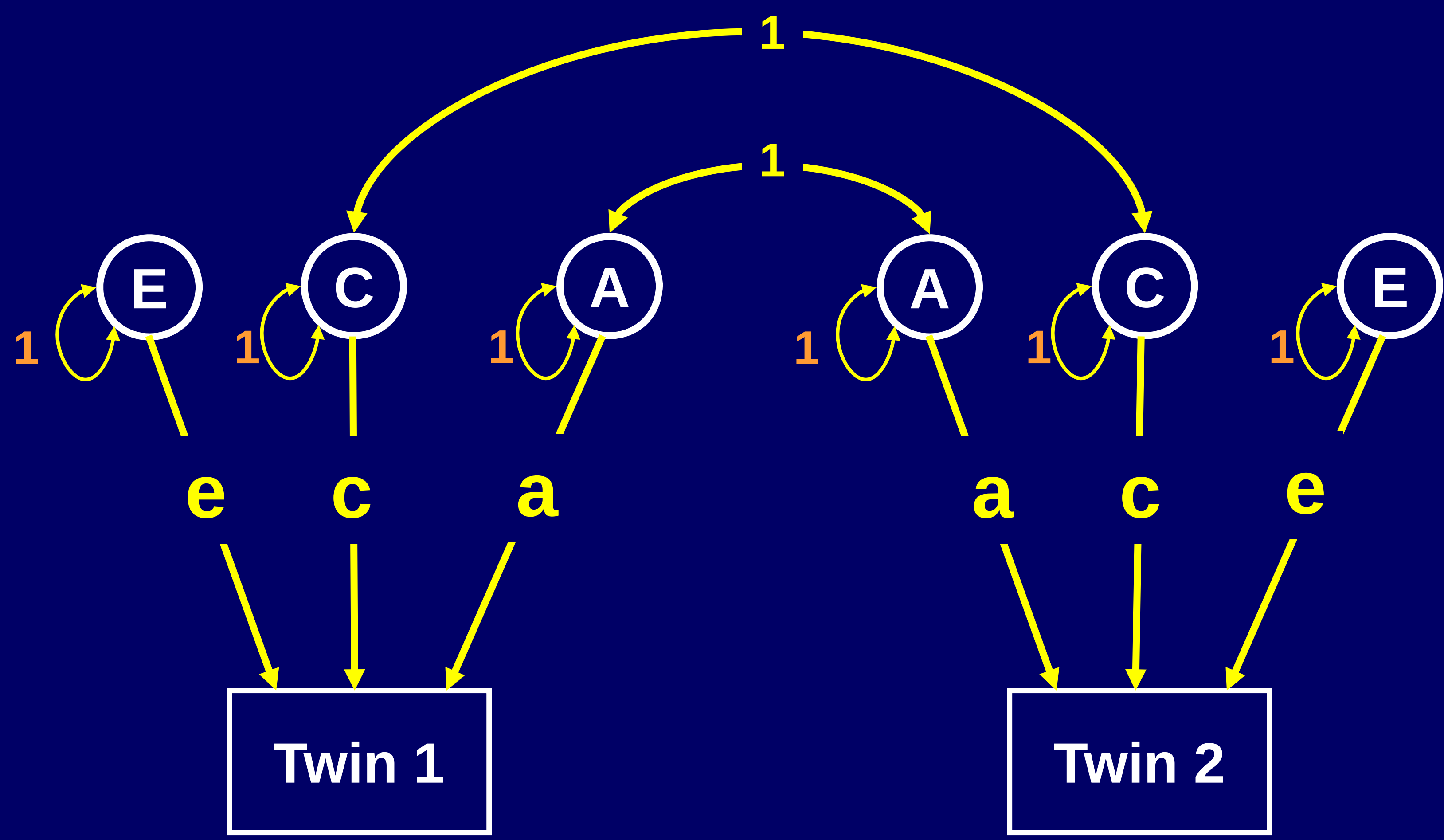
Observed Variable

Latent Variable

Causal Path

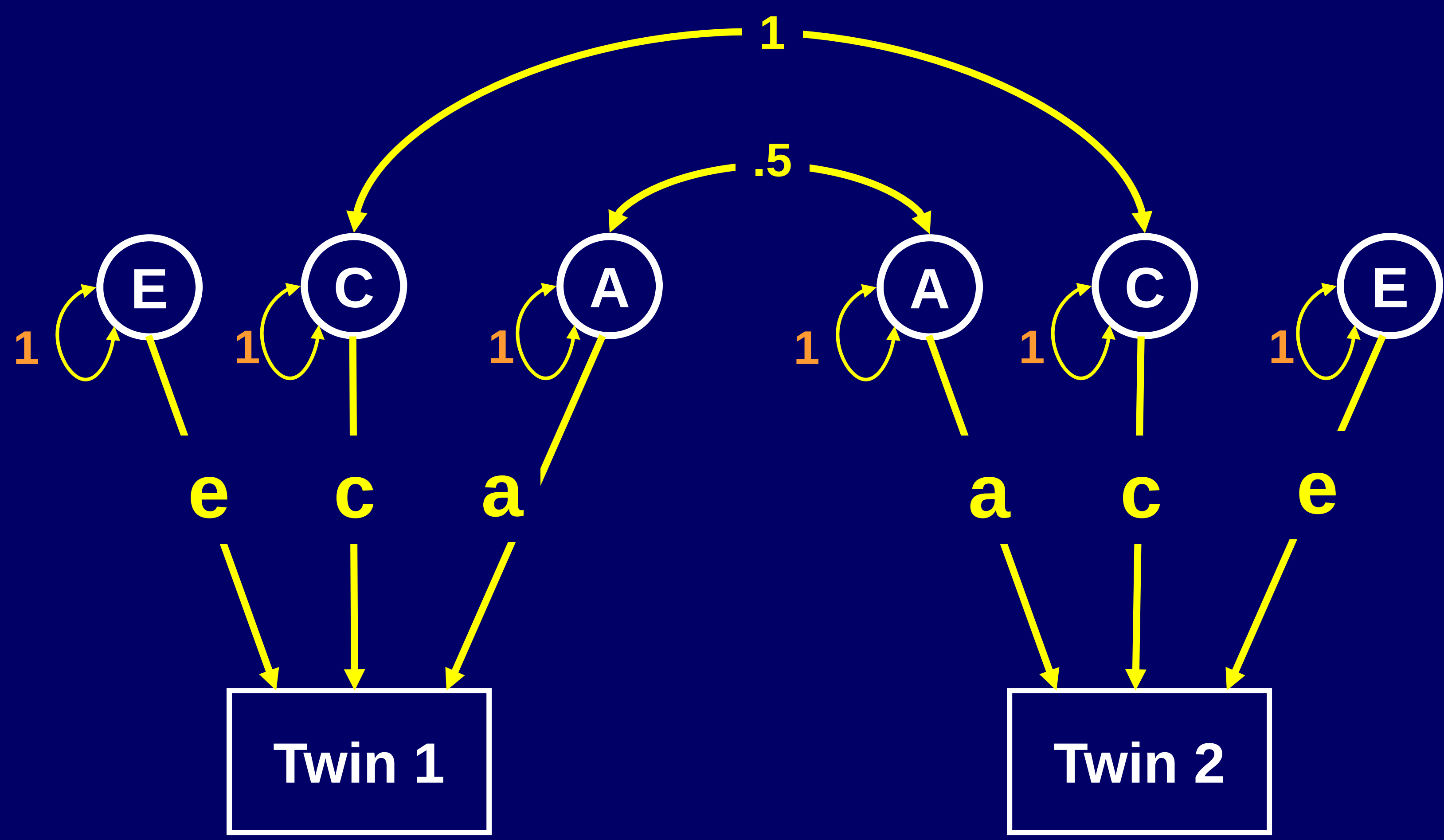
Covariance Path

Path Diagrams for the Classical Twin Model



Model for an MZ PAIR

Note: a, c and e are the same cross twins



Model for a DZ PAIR

Note: a, c and e are also the same cross groups

Path Tracing

The covariance between any two variables is the sum of all legitimate chains connecting the variables

The numerical value of a chain is the product of all traced path coefficients in it

A legitimate chain

is a path along arrows that follow 3 rules:

(i) Trace backward, then forward, or simply forward from one variable to another.
NEVER forward then backward!
Include double-headed arrows from the independent variables to itself. These variances will be **1** for **latent variables**

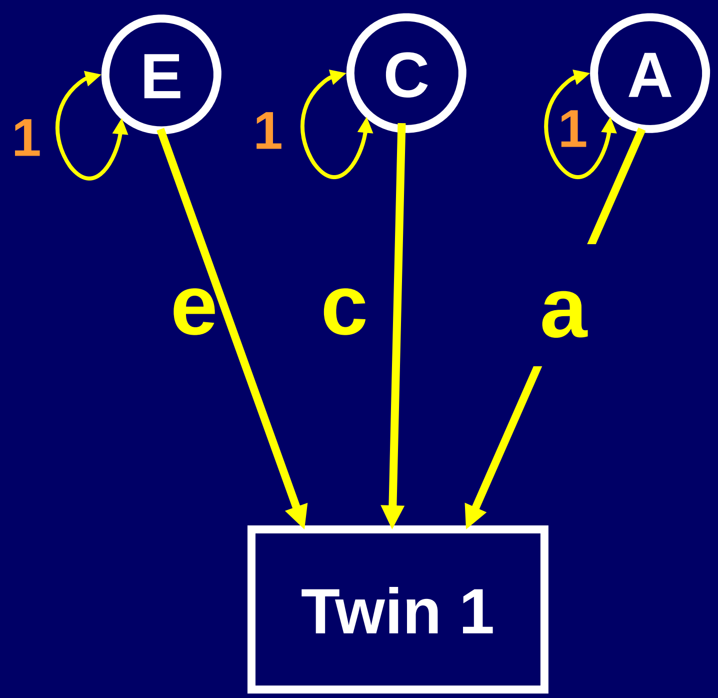
(ii) Loops are not allowed, i.e. we can not trace twice through the same variable

(iii) There is a maximum of **one curved** arrow per path.
So, the double-headed arrow from the independent variable to itself is included, unless the chain includes another double-headed arrow (e.g. a correlation path)

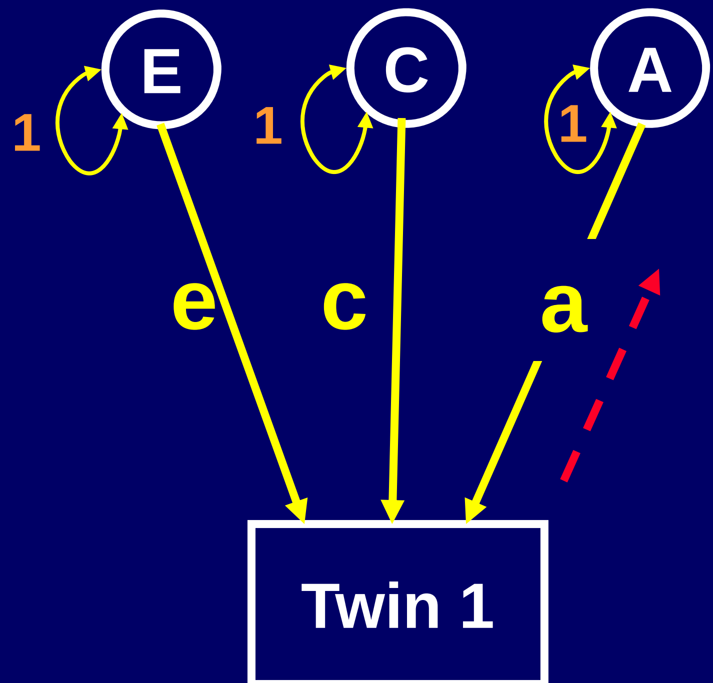
The Variance

Since the variance of a variable is the covariance of the variable with itself, the expected variance will be the sum of all paths from the variable to itself, which follow Wright's rules

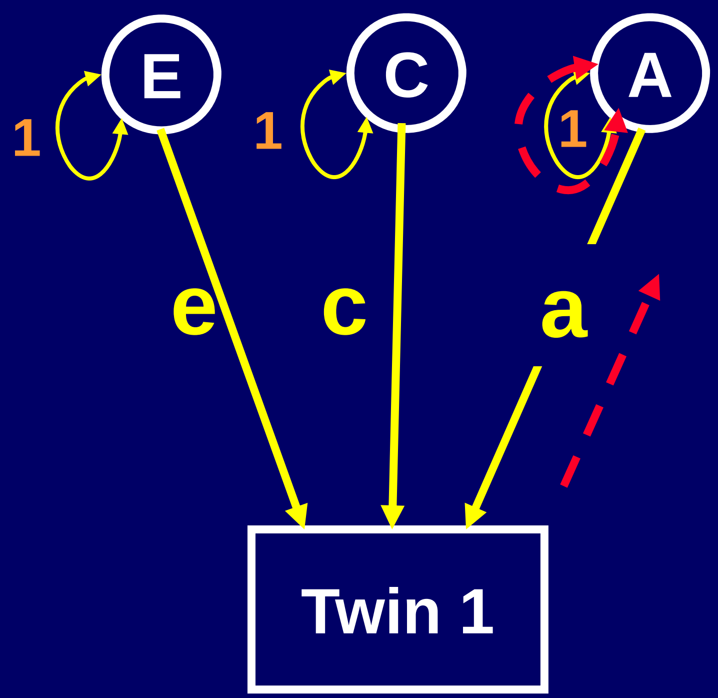
Variance of Twin 1 AND Twin 2 (for MZ and DZ pairs)



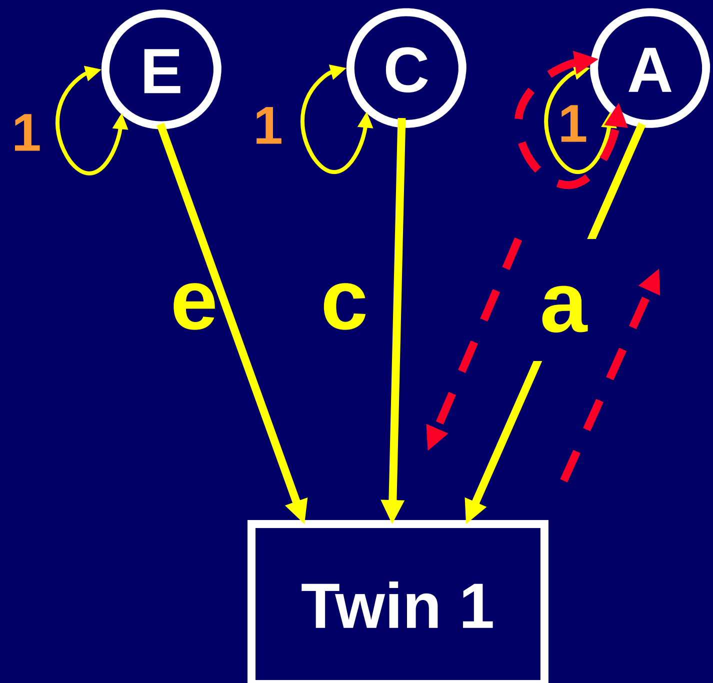
Variance of Twin 1 AND Twin 2 (for MZ and DZ pairs)



Variance of Twin 1 AND Twin 2 (for MZ and DZ pairs)



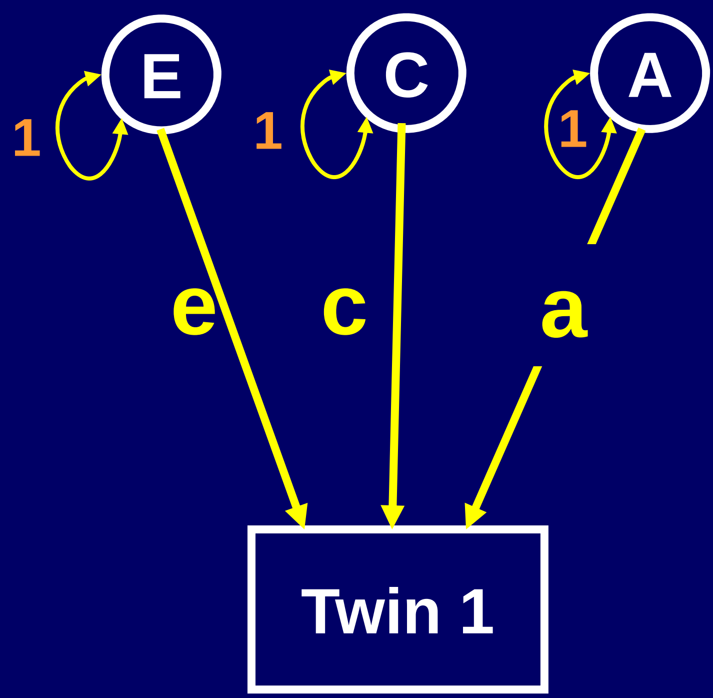
Variance of Twin 1 AND Twin 2 (for MZ and DZ pairs)



$a*1*a = a^2$

+

Variance of Twin 1 AND Twin 2 (for MZ and DZ pairs)



$$a*1*a = a^2$$

$$+$$

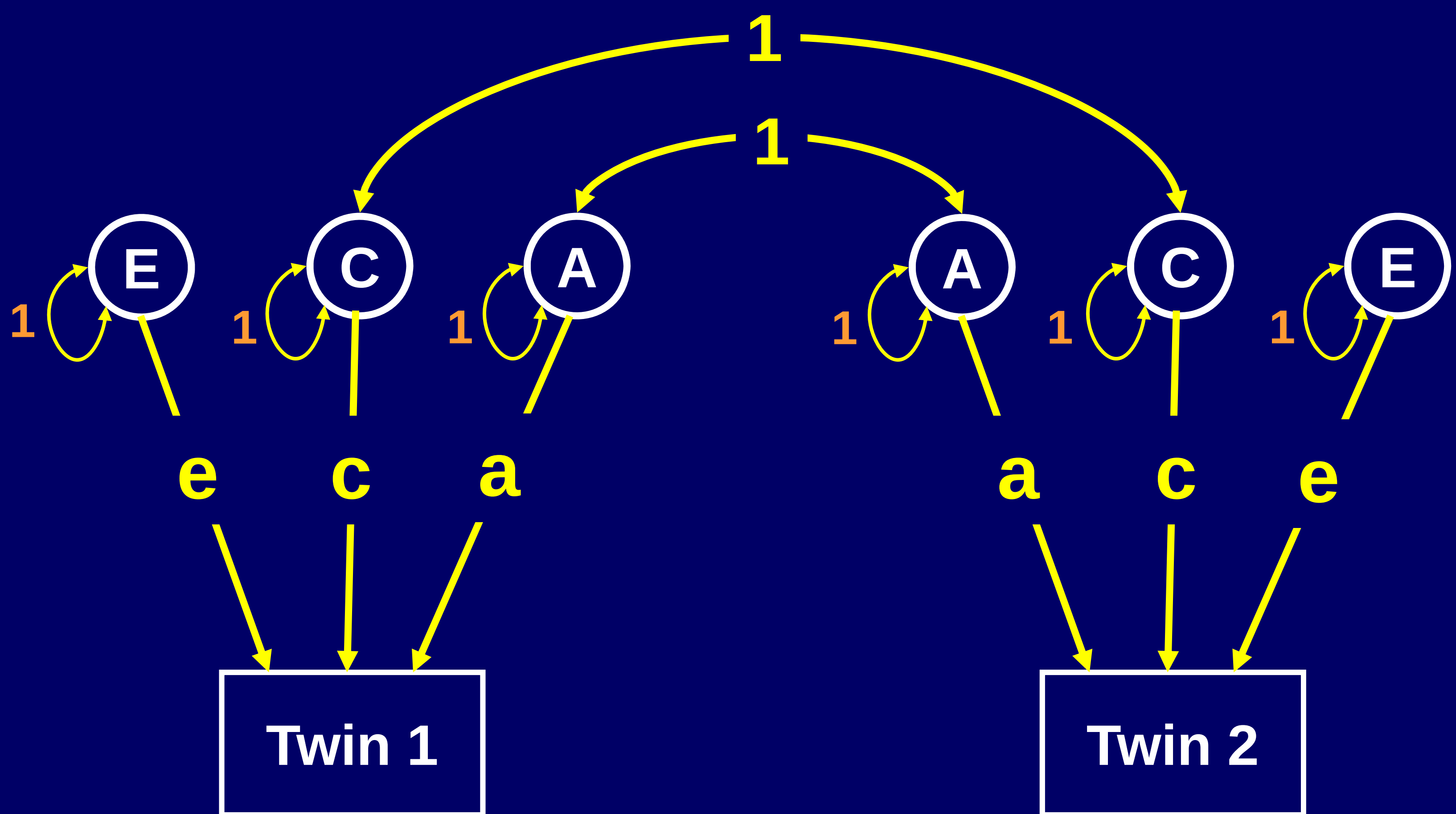
$$c*1*c = c^2$$

$$+$$

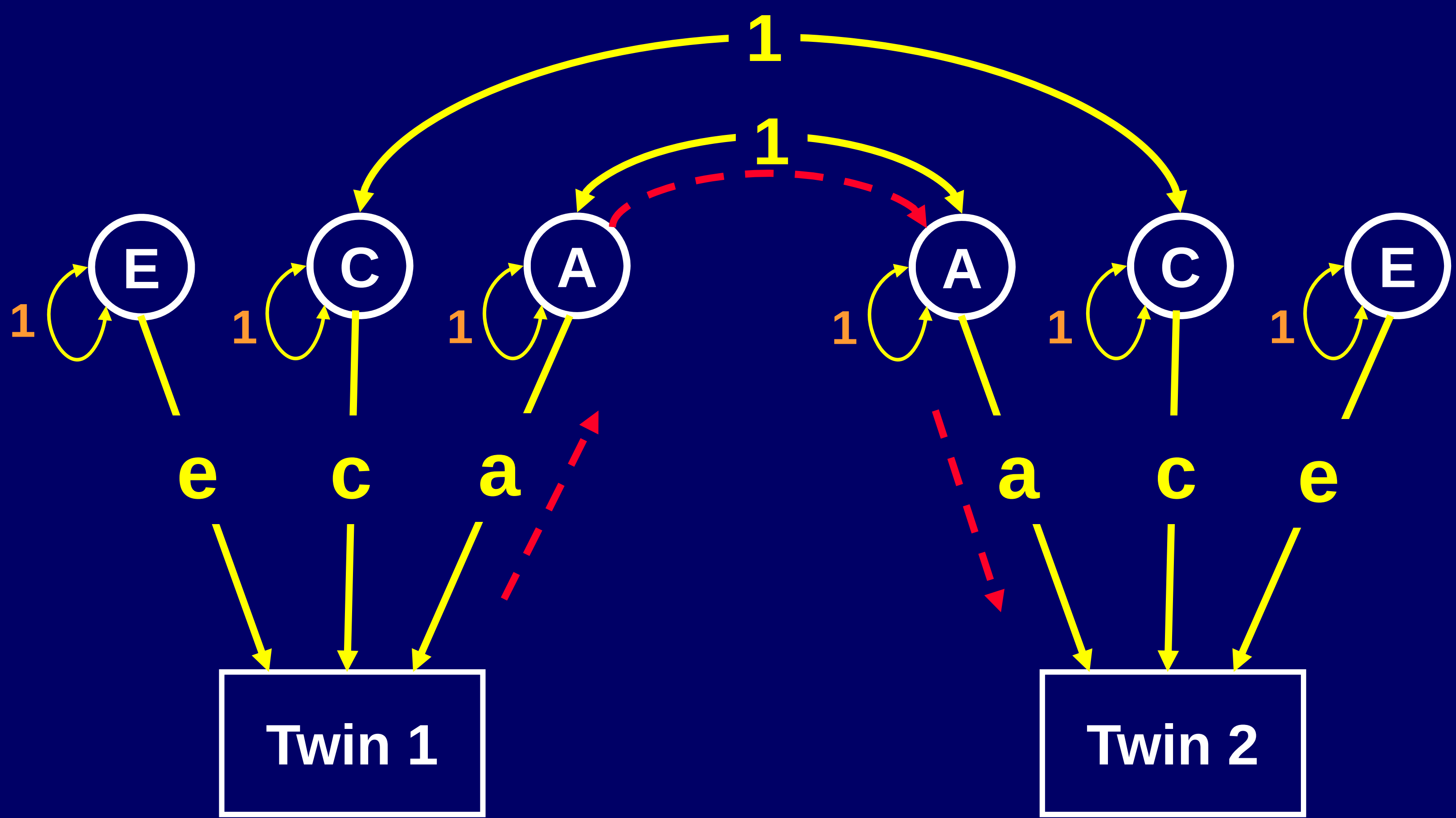
$$e*1*e = e^2$$

$$\text{Total Variance} = a^2 + c^2 + e^2$$

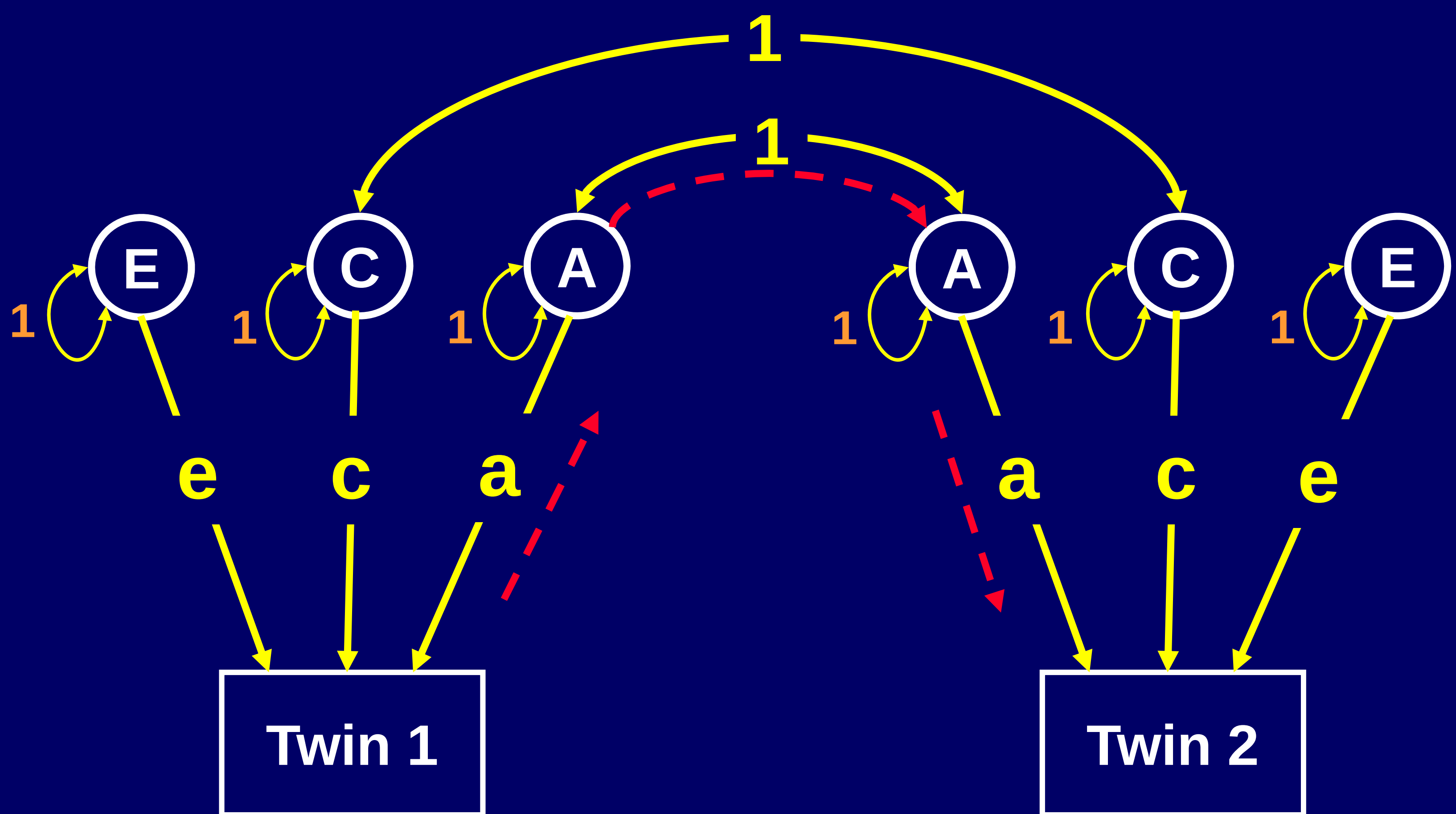
Covariance Twin 1-2: MZ pairs



Covariance Twin 1-2: MZ pairs

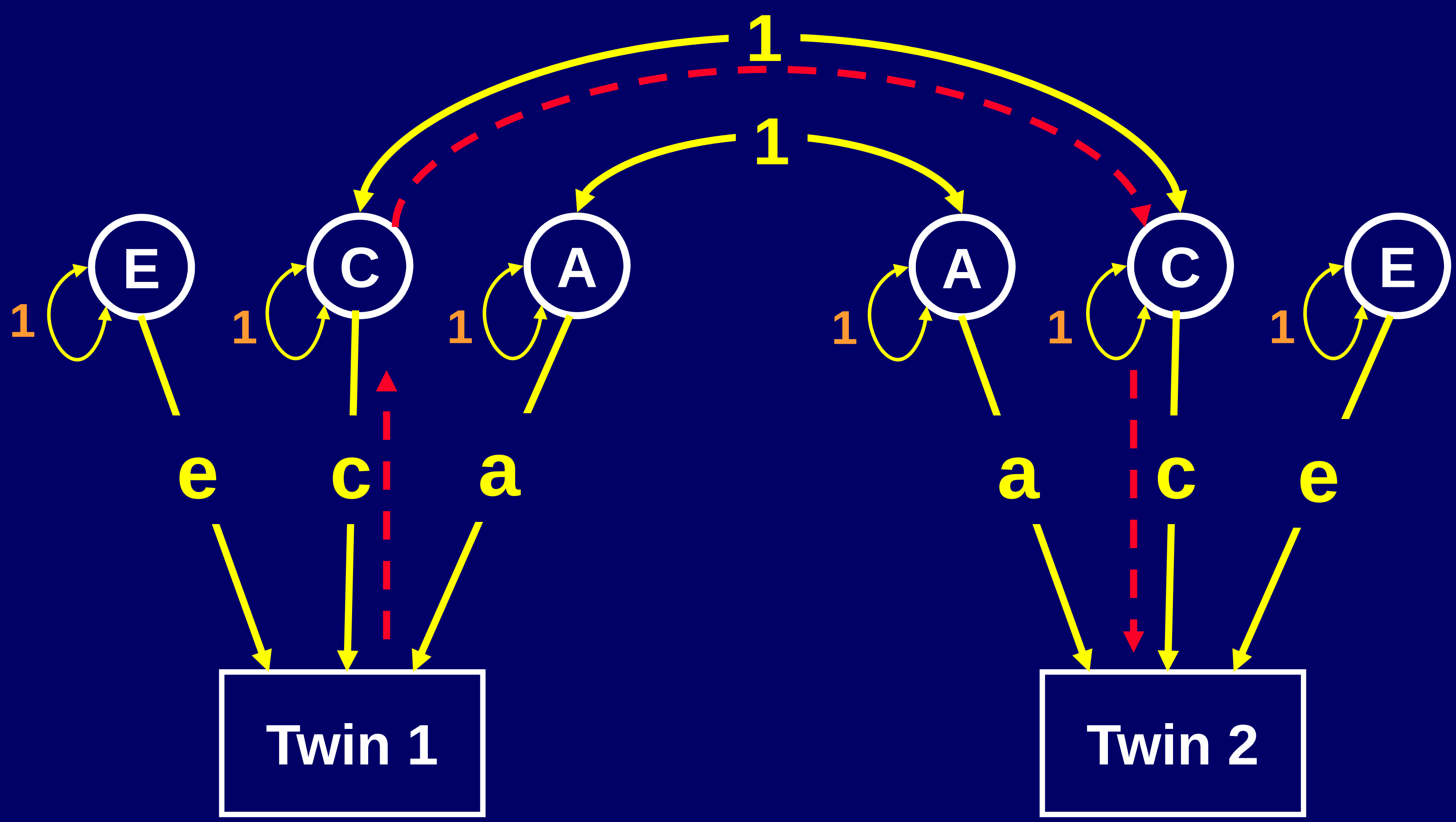


Covariance Twin 1-2: MZ pairs



Total Covariance = a^2 +

Covariance Twin 1-2: MZ pairs



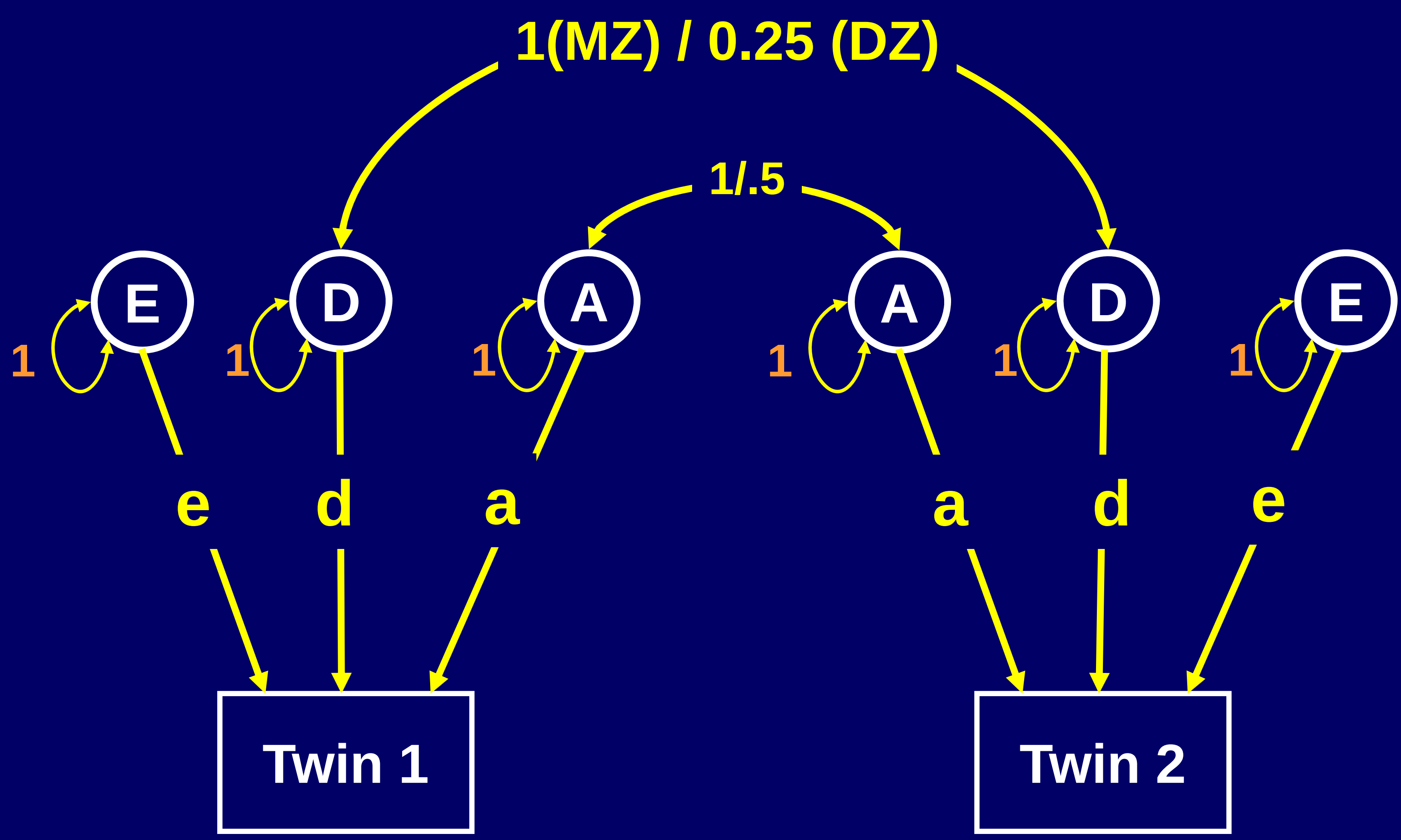
Total Covariance = $a^2 + c^2$

Predicted Var-Cov Matrices

$$
Cov\,MZ = \begin{array}{c} \\ \text{Tw1} \\ \text{Tw2} \end{array}
\begin{array}{c} \begin{array}{cc} \text{Tw1} & \text{Tw2} \end{array} \\
\begin{bmatrix} a^2 + c^2 + e^2 & a^2 + c^2 \\ a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix} \end{array}
$$

$$
Cov\,DZ = \begin{array}{c} \\ \text{Tw1} \\ \text{Tw2} \end{array}
\begin{array}{c} \begin{array}{cc} \text{Tw1} & \text{Tw2} \end{array} \\
\begin{bmatrix} a^2 + c^2 + e^2 & \frac{1}{2}a^2 + c^2 \\ \frac{1}{2}a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix} \end{array}
$$

ADE Model



Predicted Var-Cov Matrices

$$Cov\,MZ = \begin{array}{c} \\ Tw1 \\ Tw2 \end{array} \begin{array}{c} \begin{array}{cc} Tw1 & Tw2 \end{array} \\ \begin{bmatrix} a^2 + d^2 + e^2 & a^2 + d^2 \\ a^2 + d^2 & a^2 + d^2 + e^2 \end{bmatrix} \end{array}$$

$$Cov\,DZ = \begin{array}{c} \\ Tw1 \\ Tw2 \end{array} \begin{array}{c} \begin{array}{cc} Tw1 & Tw2 \end{array} \\ \begin{bmatrix} a^2 + d^2 + e^2 & \frac{1}{2}a^2 + \frac{1}{4}d^2 \\ \frac{1}{2}a^2 + \frac{1}{4}d^2 & a^2 + d^2 + e^2 \end{bmatrix} \end{array}$$

ACE or ADE

Cov(mz) = $a^2 + c^2$ or $a^2 + d^2$

Cov(dz) = $½ a^2 + c^2$ or $½ a^2 + ¼ d^2$

$V_P = a^2 + c^2 + e^2$ or $a^2 + d^2 + e^2$

3 unknown parameters (a, c, e or a, d, e), and only **3** distinct predicted statistics:

Cov MZ, Cov DZ, Vp)

this model is **just identified**

Effects of C and D are confounded

The twin correlations indicate which of the two components is more likely to be present:

Cor(mz) = $a^2 + c^2$ or $a^2 + d^2$

Cor(dz) = $\frac{1}{2} a^2 + c^2$ or $\frac{1}{2} a^2 + \frac{1}{4} d^2$

If $a^2 = .40$, $c^2 = .20$

$r_{mz} = 0.60$

$r_{dz} = 0.40$

ACE

If $a^2 = .40$, $d^2 = .20$

$r_{mz} = 0.60$

$r_{dz} = 0.25$

ADE

ADCE: classical twin design + adoption data

$Cov(mz) = a^2 + d^2 + c^2$

$Cov(dz) = \frac{1}{2} a^2 + \frac{1}{4} d^2 + c^2$

$Cov(adopSibs) = c^2$

$V_P = a^2 + d^2 + c^2 + e^2$

4 unknown parameters (a, c, d, e), and 4 distinct predicted statistics:

Cov MZ, Cov DZ, Cov adopSibs, Vp)

this model is **just identified**

Path Tracing Rules are based on Covariance Algebra

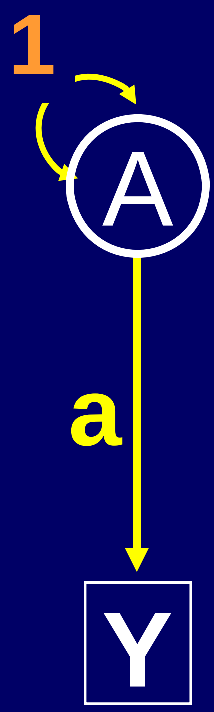
Three Fundamental Covariance Algebra Rules

$$\text{Var}(X) = \text{Cov}(X,X)$$

$$\text{Cov}(aX,bY) = ab\,\text{Cov}(X,Y)$$

$$\text{Cov}(X,Y+Z) = \text{Cov}(X,Y) + \text{Cov}(X,Z)$$

Example 1

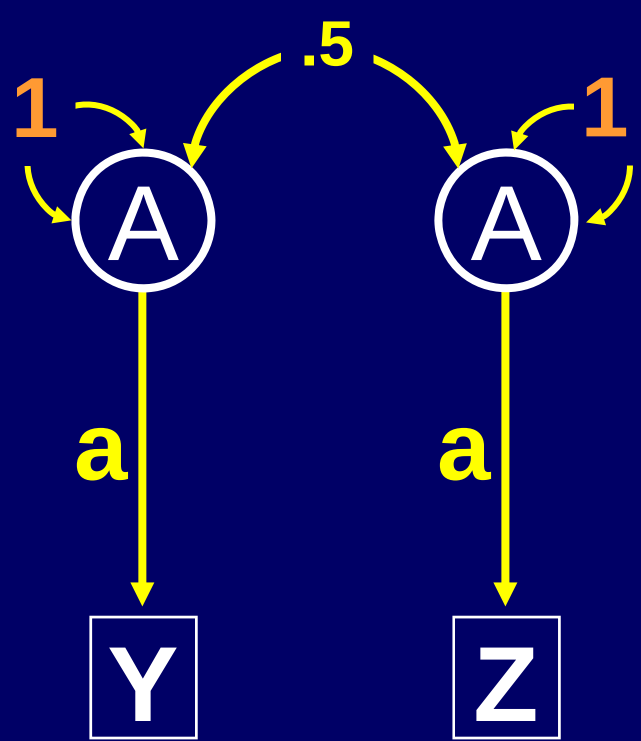


Y = aA

$$
\begin{aligned}
Var(Y) &= Var(aA) \\
&= Cov(aA, aA) \\
&= a^2 Cov(A, A) \\
&= a^2 Var(A) \\
&= a^2 * 1 \\
&= a^2
\end{aligned}
$$

The variance of a dependent variable (Y) caused by independent variable A, is the squared regression coefficient multiplied by the variance of the independent variable

Example 2



Y = aA   Z = aA

$$Cov(Y,Z) = Cov(aA, aA)$$
$$= a^2 Cov(A,A)$$
$$= a^2 * .5$$

Summary

Path Tracing and Covariance Algebra have the same aim :

to work out the predicted Variances and Covariances of variables, given a specified model