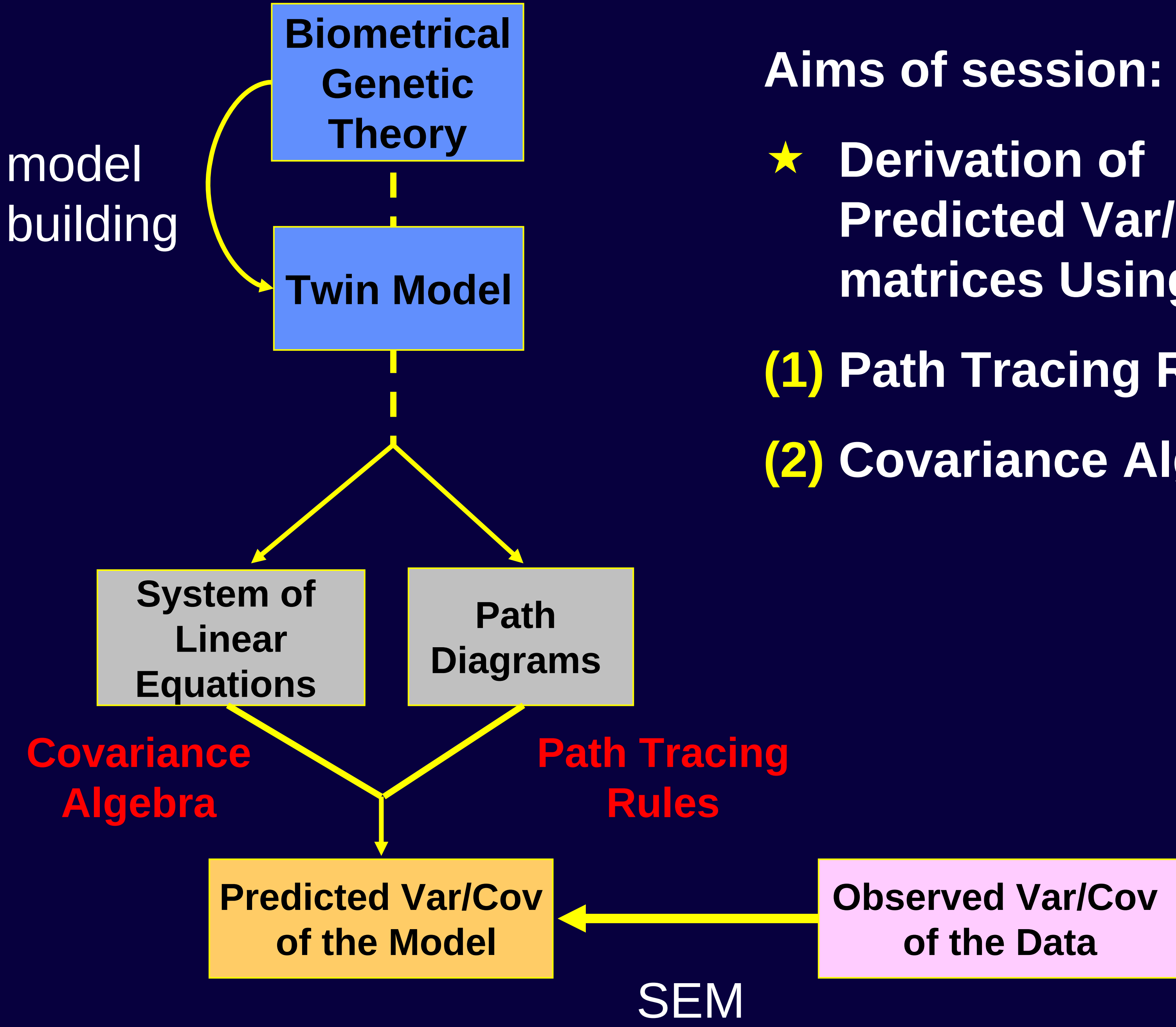
# Path Analysis

Frühling Rijsdijk




**Aims of session:**

★ **Derivation of Predicted Var/Cov matrices Using:**

**(1) Path Tracing Rules**

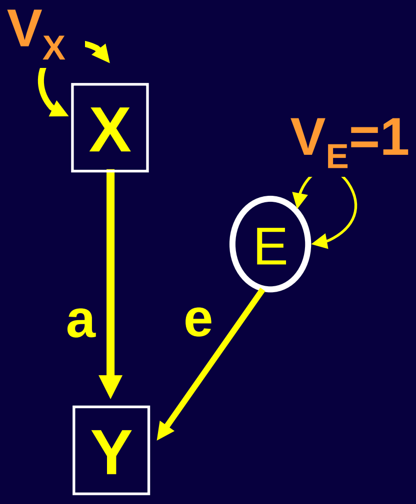
**(2) Covariance Algebra**

# Method of Path Analysis

- Allows us to represent linear models for the relationship between variables in diagrammatic form, e.g. a genetic model; a factor model; a regression model
- Makes it easy to derive expectations for the variances and covariances of variables in terms of the parameters of the proposed linear model
- Permits easy translation into matrix formulation as used by programs such as Mx, OpenMx.

# Conventions of Path Analysis I

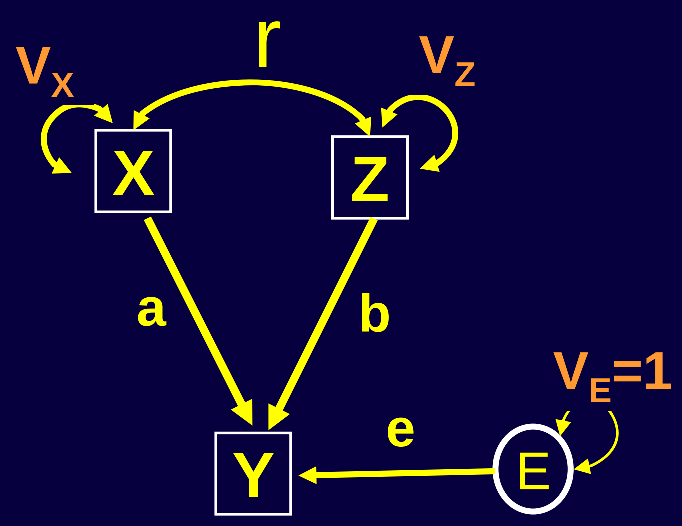
- Squares or rectangles denote observed variables
- Circles or ellipses denote latent (unmeasured) variables
- Upper-case letters are used to denote variables
- Lower-case letters (or numeric values) are used to denote covariances or path coefficients
- Single-headed arrows or paths (–>) represent hypothesized causal relationships
  - where the variable at the tail is hypothesized to have a direct causal influence on the variable at the head




Y = aX + eE

# Conventions of Path Analysis II

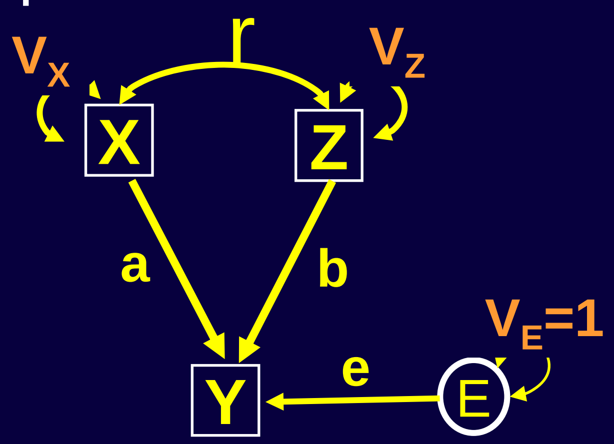
- Double-headed arrows (<–>) are used to represent a covariance between two variables, which may arise through common causes not represented in the model.
- Double-headed arrows may also be used to represent the variance of a variable.




$$Y = aX + bZ + eE$$

# Conventions of Path Analysis III

- Variables that do not receive causal input from any one variable in the diagram are referred to as ***independent,*** or ***predictor*** or **exogenous** variables.
- Variables that do, are referred to as ***dependent*** or ***endogenous*** variables.
- Only independent variables are connected by double-headed arrows.
- Single-headed arrows may be drawn from independent to dependent variables or from dependent variables to other dependent variables.



# Conventions of Path Analysis IV

- Omission of a two-headed arrow between two independent variables implies the assumption that the covariance of those variables is zero
- Omission of a direct path from an independent (or dependent) variable to a dependent variable implies that there is no direct causal effect of the former on the latter variable

# 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!

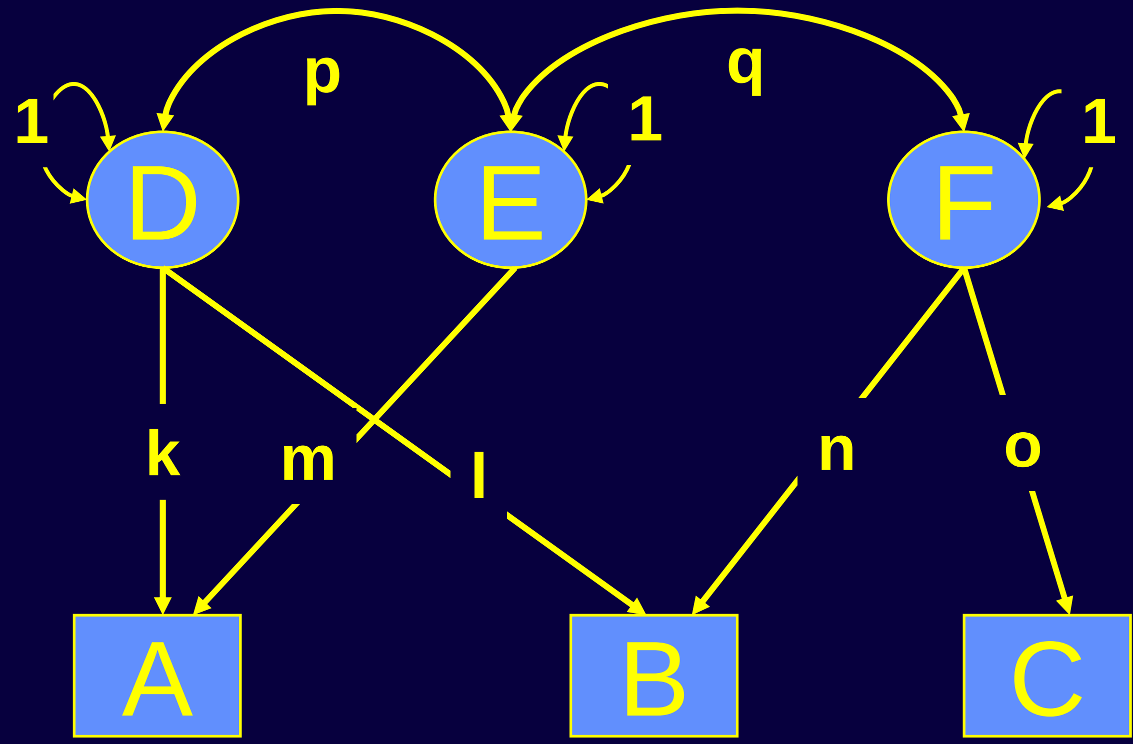
Include double-headed arrows from the independent variables to itself. These variances will be **1** for **latent variables**

(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 the path tracing rules**

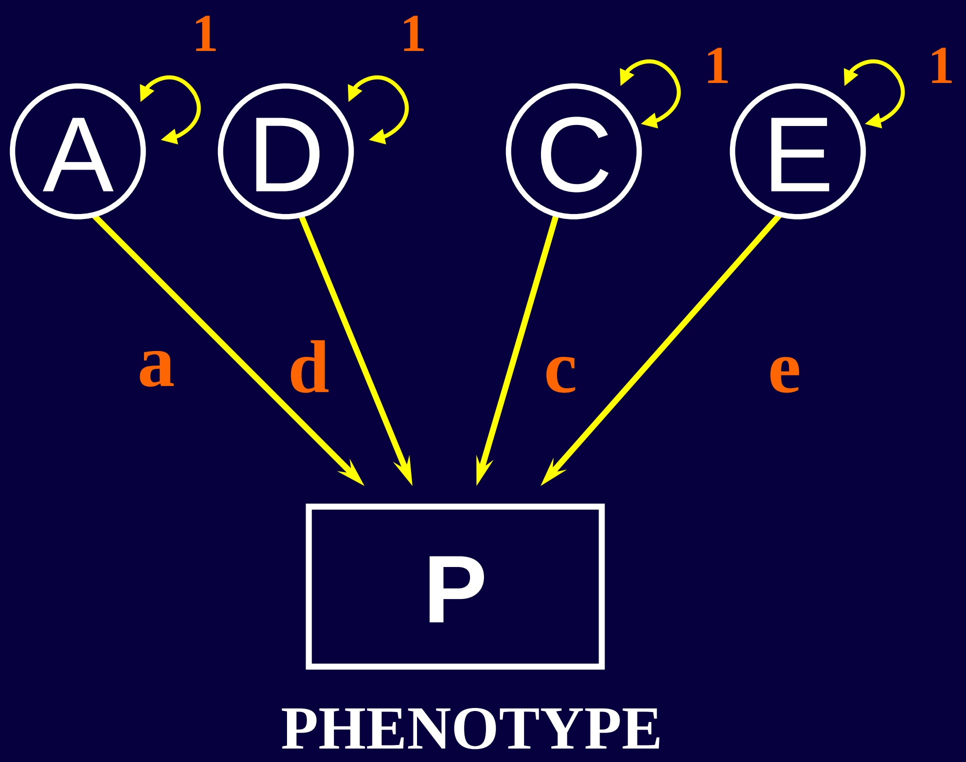



- Cov AB = kl + mqn + mpl
- Cov BC = no
- Cov AC = mqo
- Var A = $k^2 + m^2 + 2\ kpm$
- Var B = $l^2 + n^2$
- Var C = $o^2$

# Path Diagrams for the Classical Twin Model

# Quantitative Genetic Theory

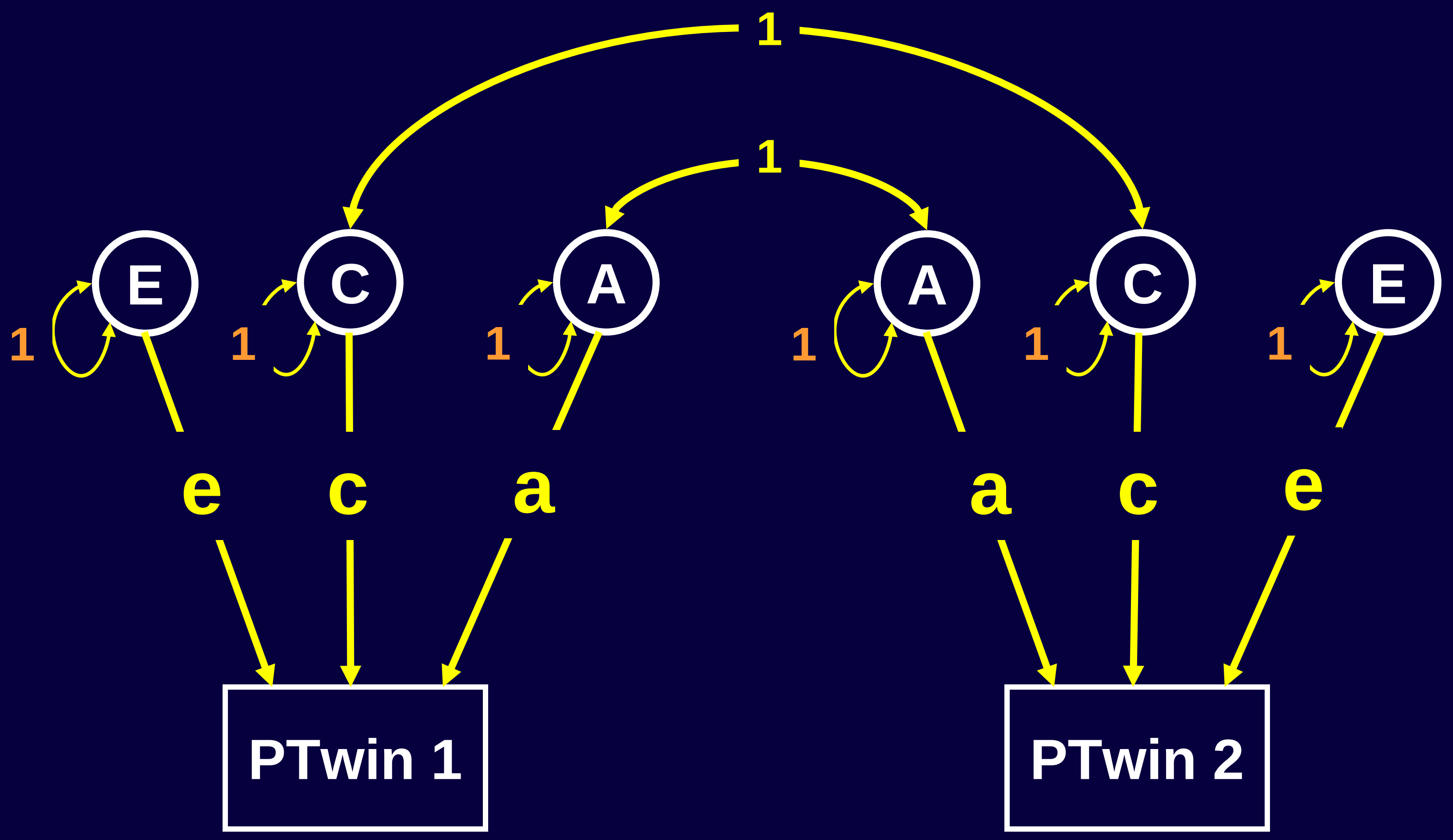
- There are two sources of Genetic influences: ***Additive (A)*** and ***non-additive or Dominance (D)***
- There are two sources of environmental influences: ***Common or shared (C)*** and ***non-shared or unique (E)***



# In the preceding diagram…

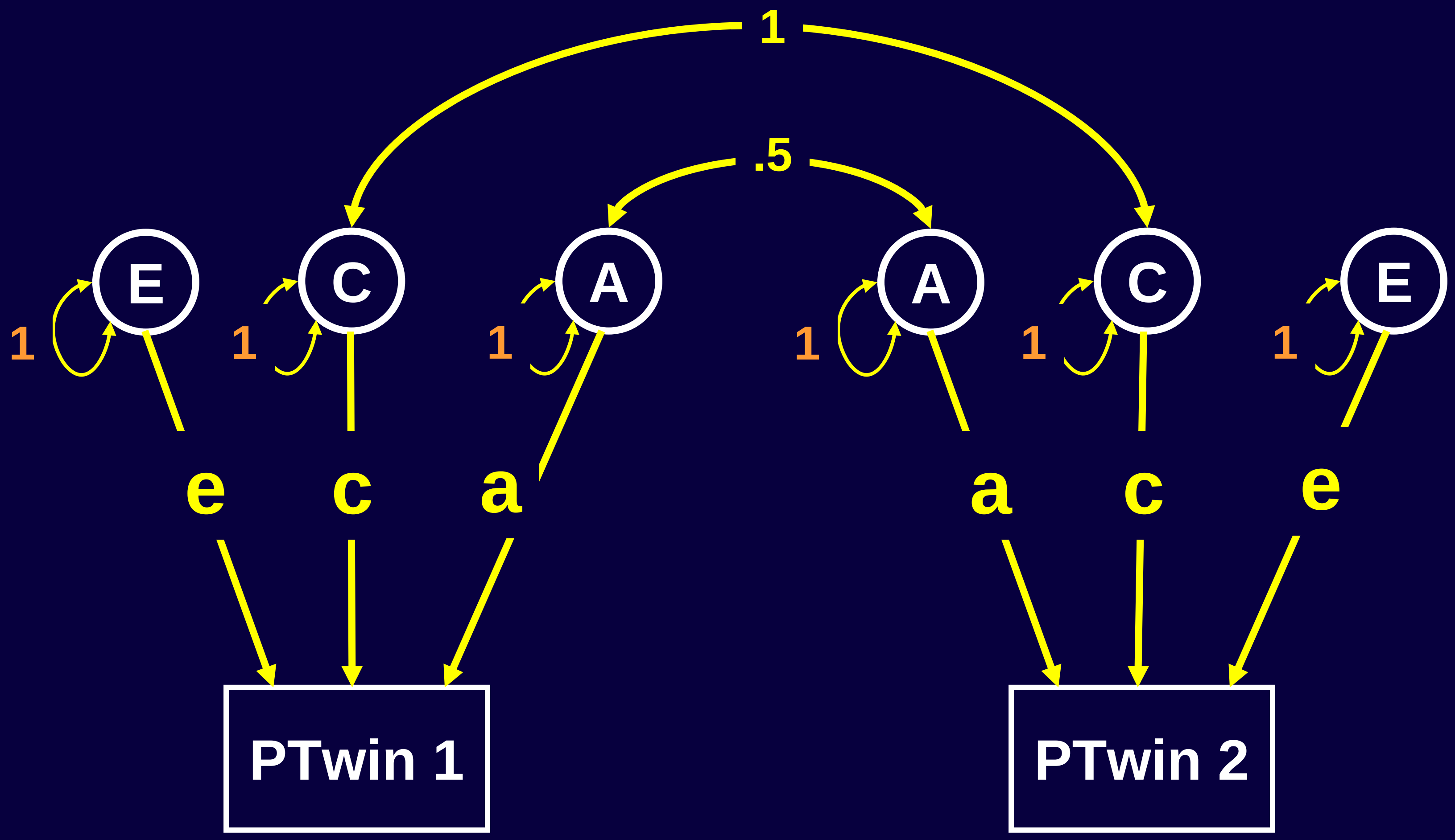
- A, D, C, E are *independent* variables
  - A = Additive genetic influences
  - D = Non-additive genetic influences (i.e., dominance)
  - C = Shared environmental influences
  - E = Non-shared environmental influences
  - A, D, C, E have variances of 1
- Phenotype is a dependent variable
  - P = phenotype; the measured variable
- a, d, c, e are parameter estimates

# Model for MZ Pairs Reared Together



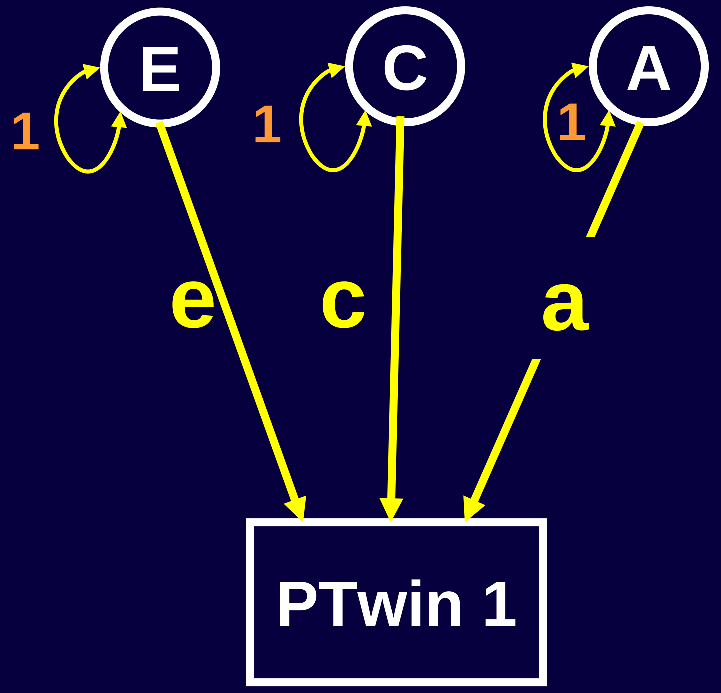

Note: a, c and e are the same cross twins

# Model for DZ Pairs Reared Together

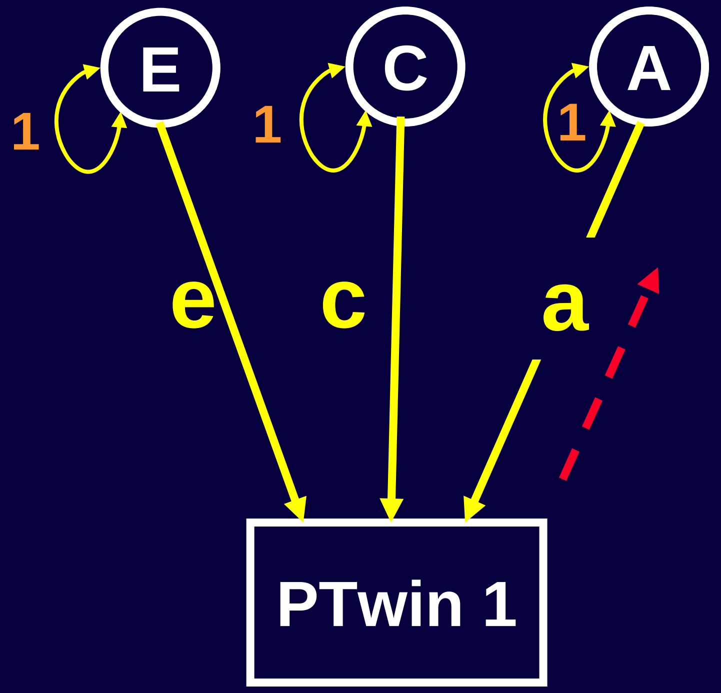



Note: a, c and e are also the same cross groups

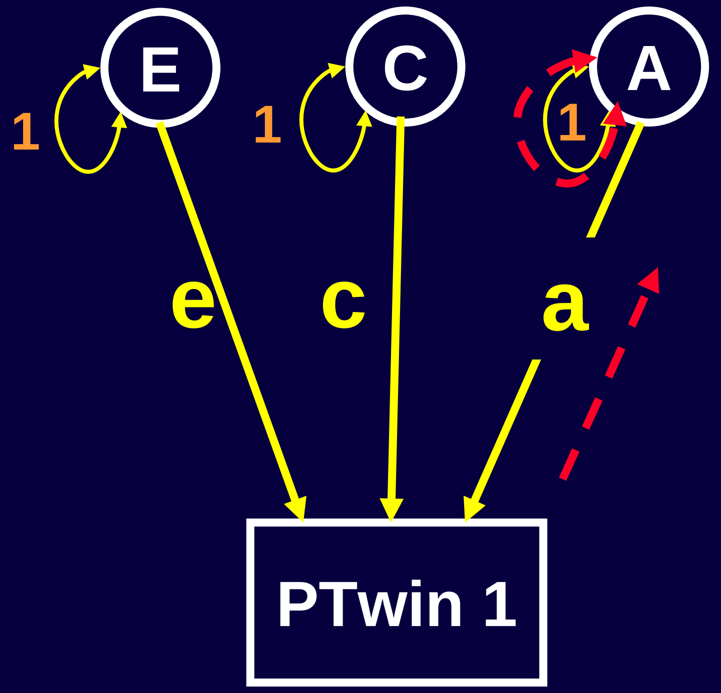
# Variance of Twin 1 AND Twin 2 (for MZ and DZ pairs)



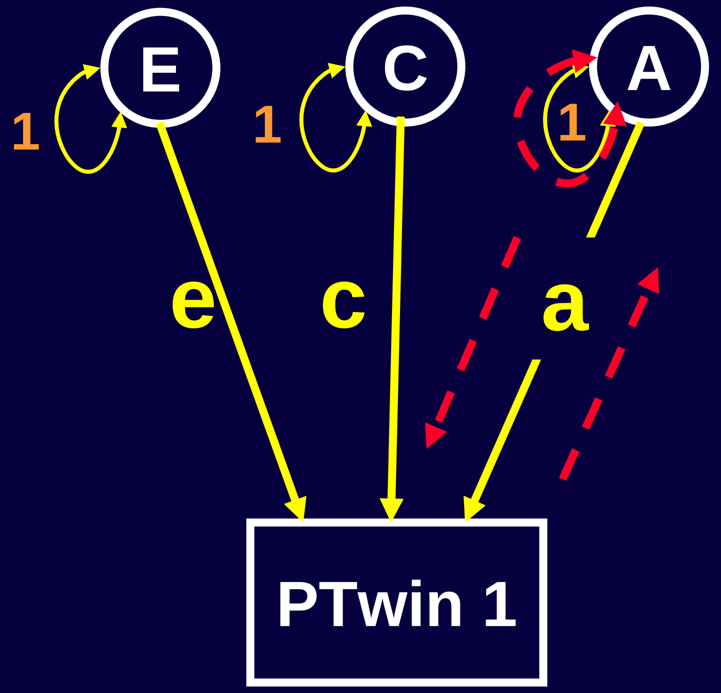
# Variance of Twin 1 AND Twin 2 (for MZ and DZ pairs)



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



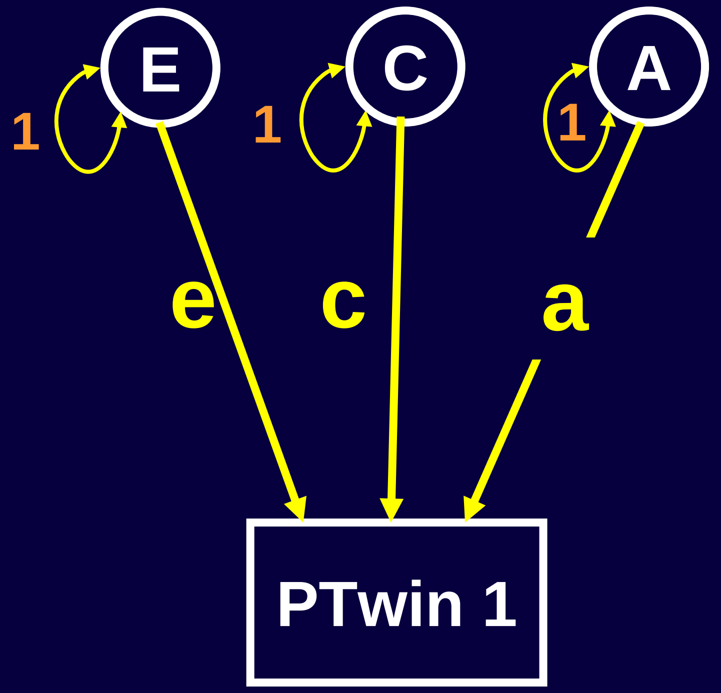
# Variance of Twin 1 AND Twin 2 (for MZ and DZ pairs)




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

+

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




$a*1*a = a^2$

$+$

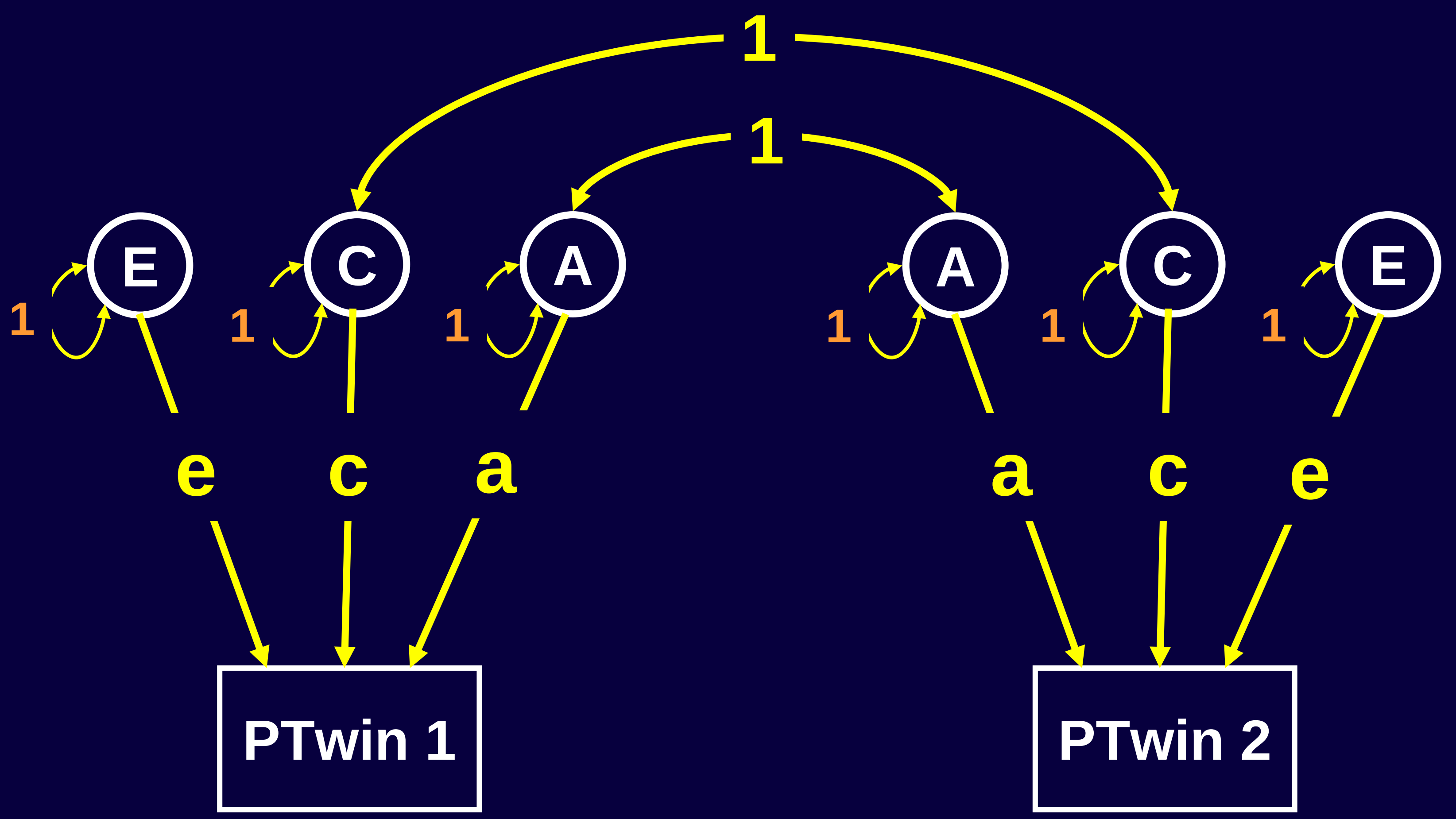
$c*1*c = c^2$

$+$

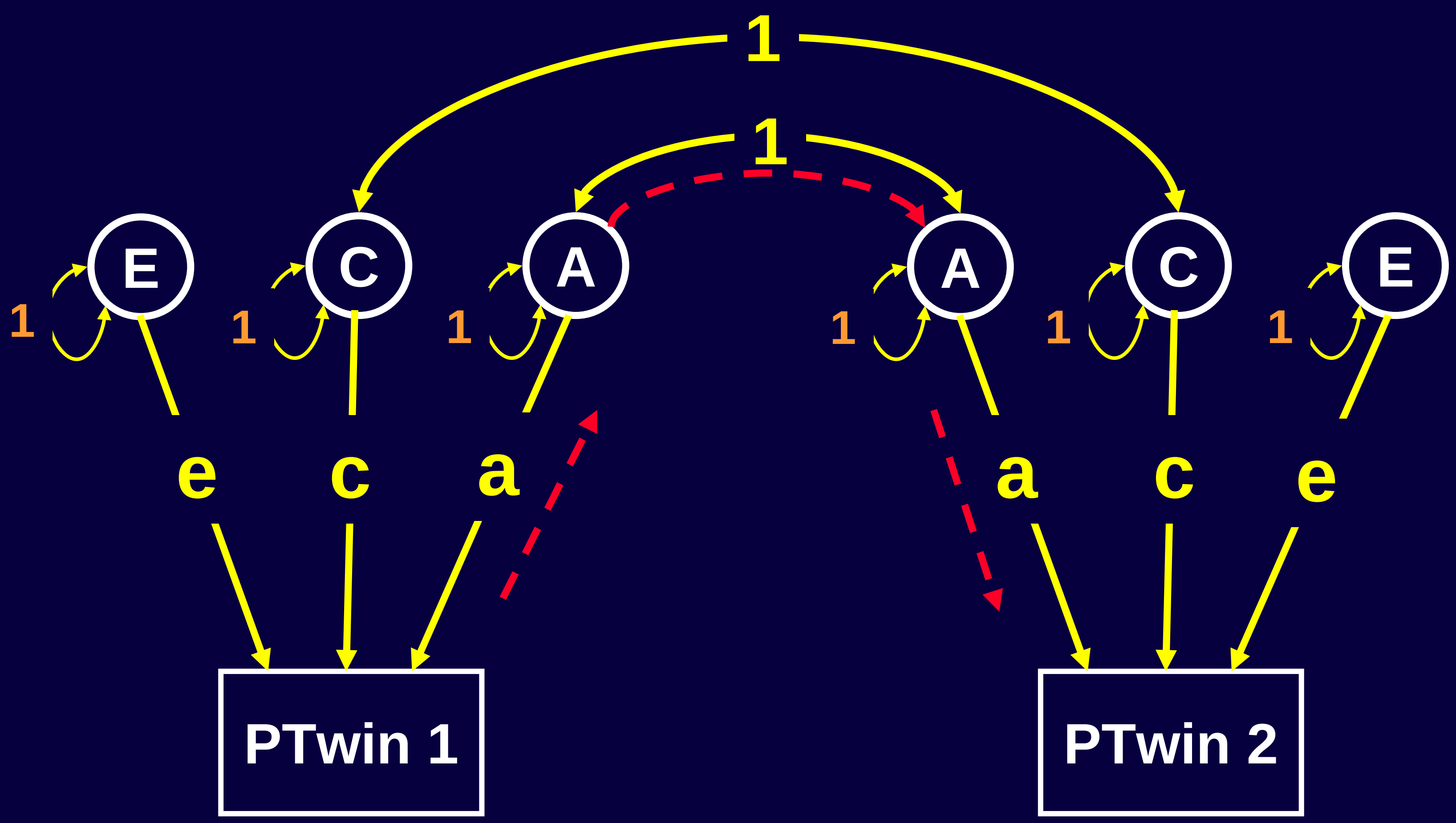
$e*1*e = e^2$

Total Variance = $a^2 + c^2 + e^2$

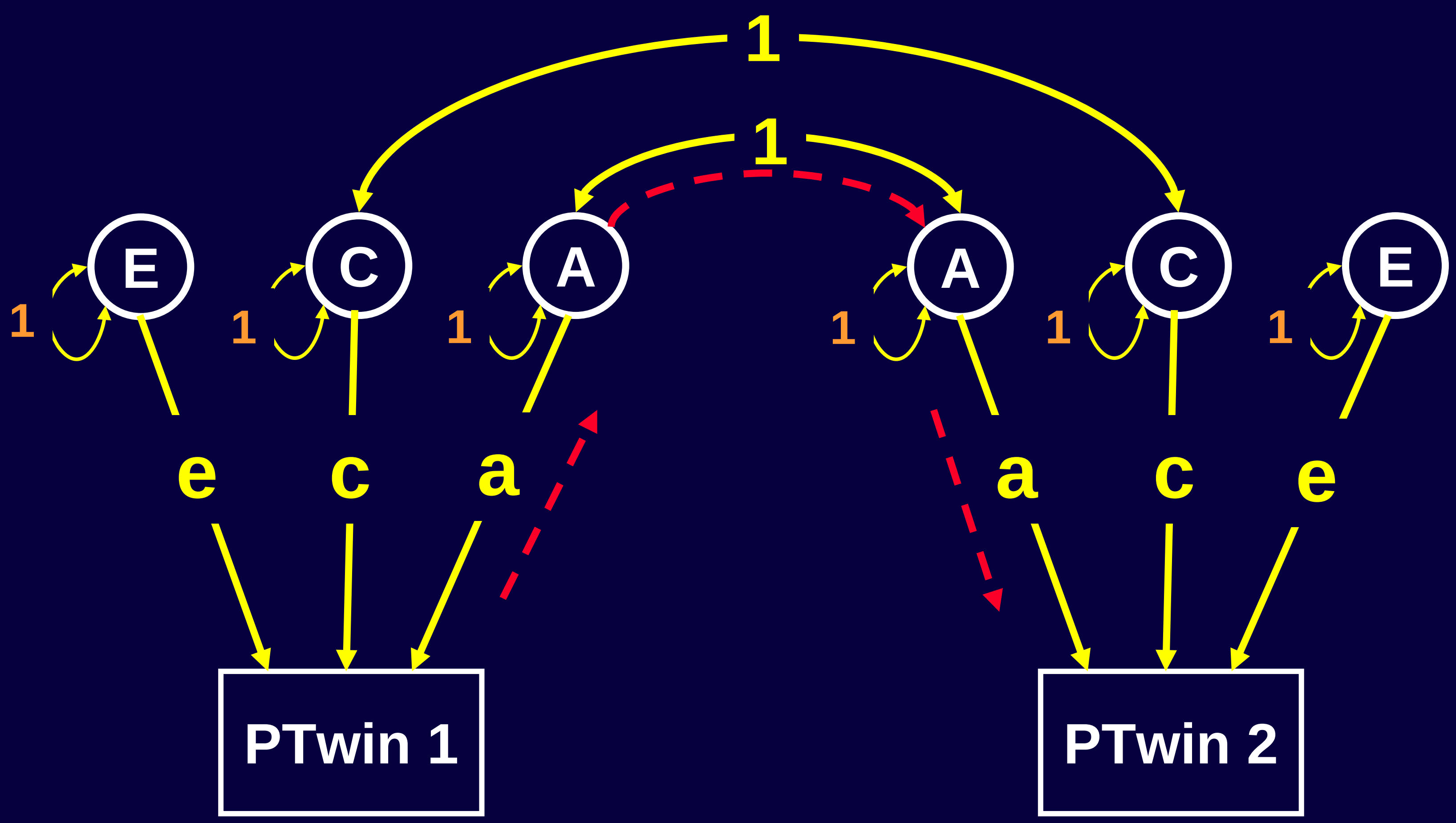
# Covariance Twin 1-2: MZ pairs



# Covariance Twin 1-2: MZ pairs

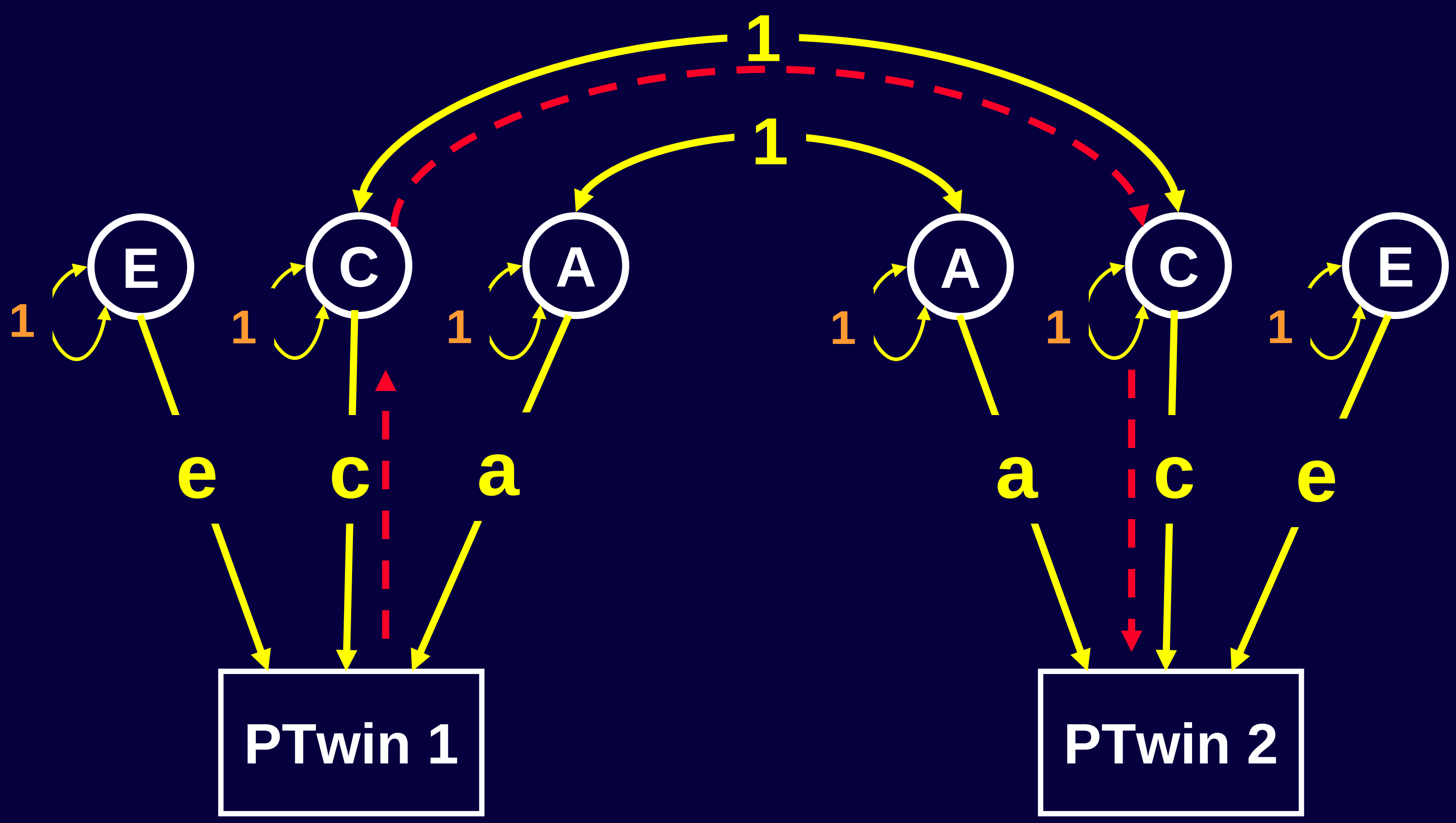


# Covariance Twin 1-2: MZ pairs



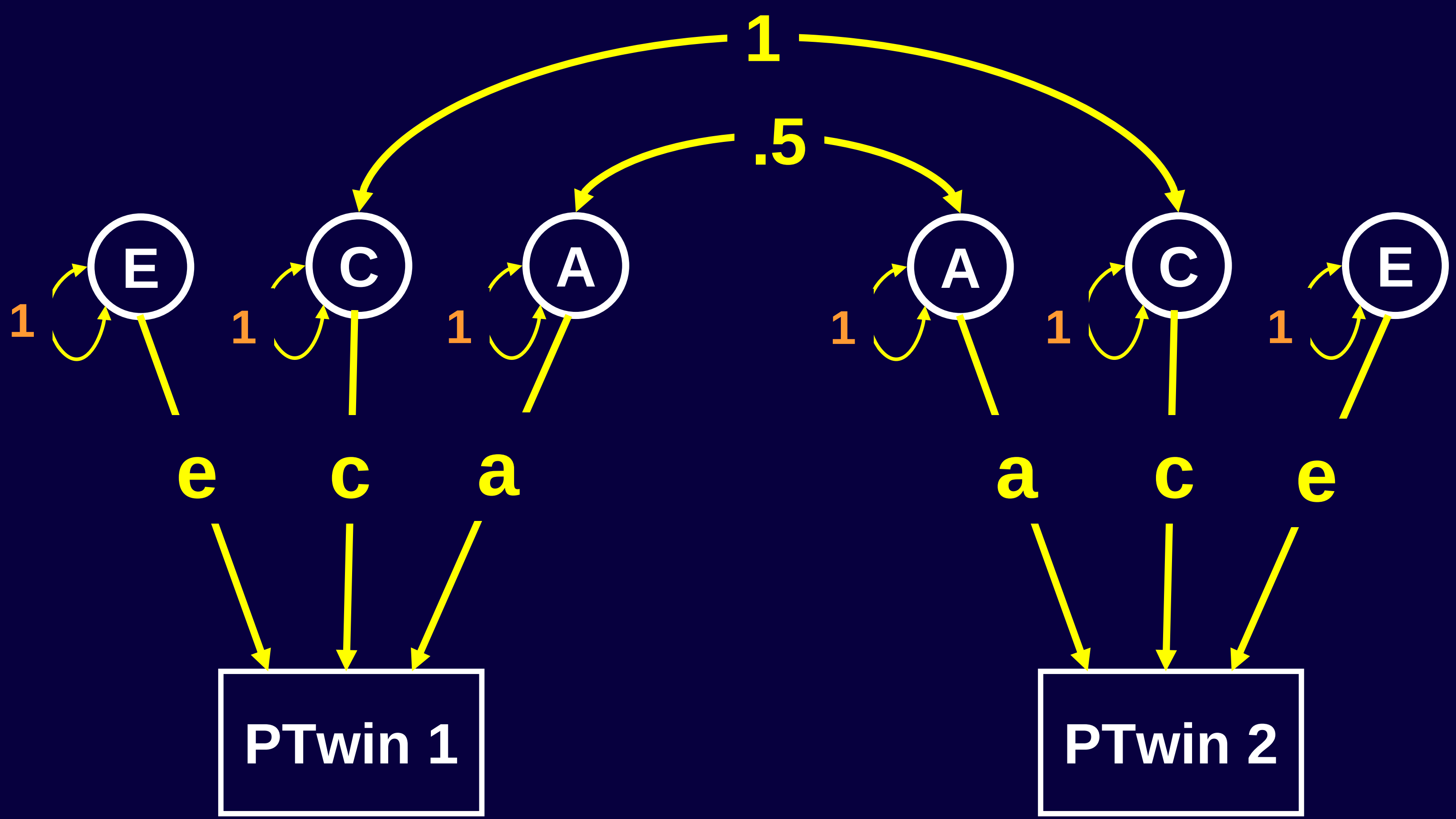

Total Covariance = $a^2$ +

# Covariance Twin 1-2: MZ pairs




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

# Covariance Twin 1-2: DZ pairs



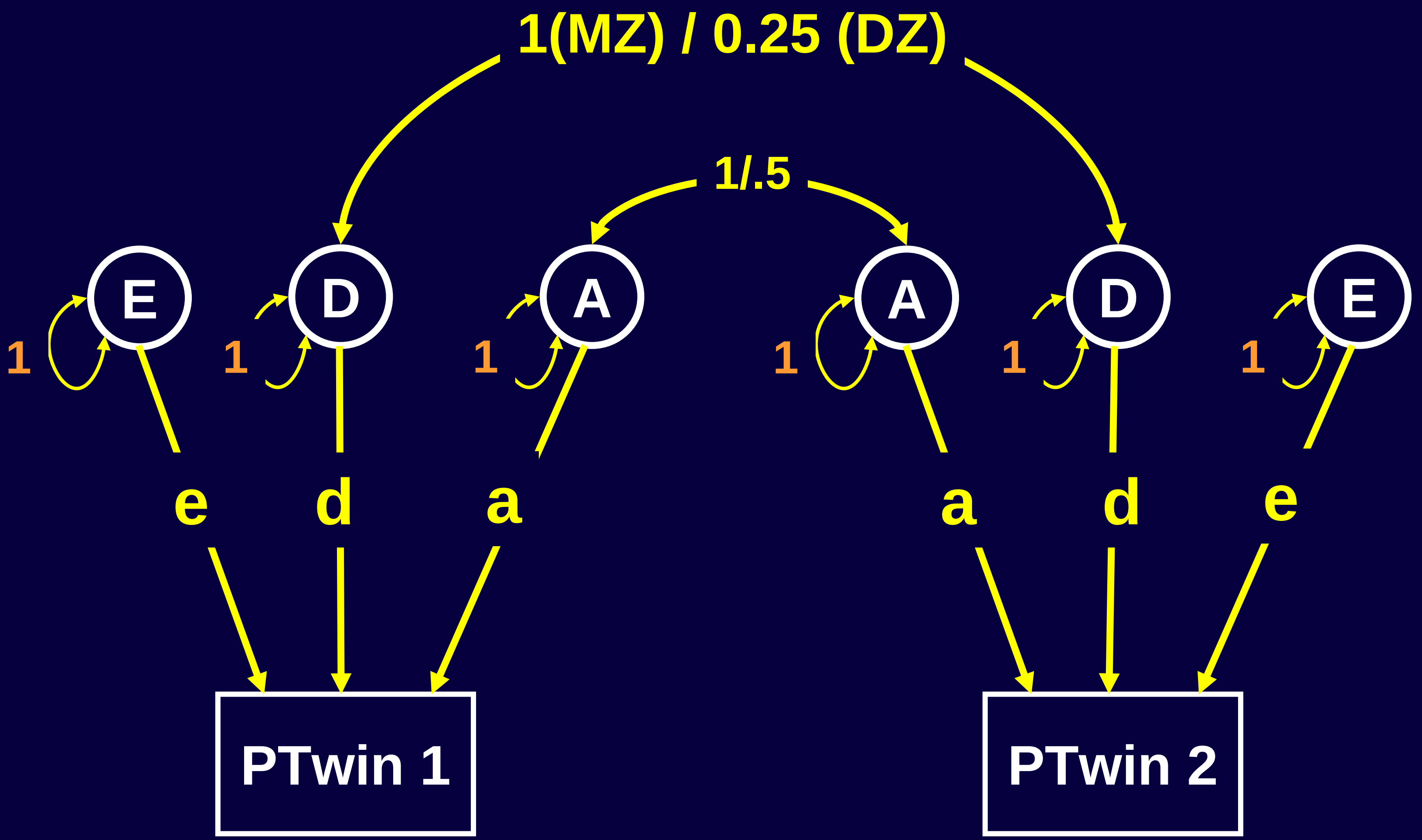

Total Covariance = $.5a^2 + c^2$

# 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 predictive 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 fit the data:**

$\text{Cor}(mz) = a^2 + c^2$ or $a^2 + d^2$

$\text{Cor}(dz) = \frac{1}{2} a^2 + c^2$ 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 predictive statistics:

Cov MZ, Cov DZ, Cov adopSibs, Vp

this model is **just identified**

# Path Tracing Rules are based on Covariance Algebra

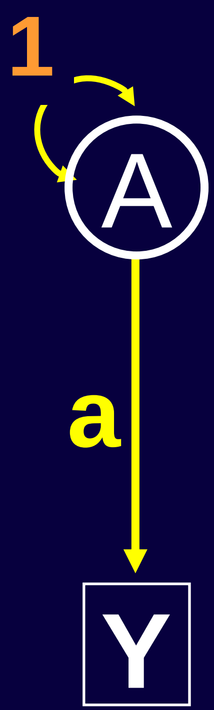
# Three Fundamental Covariance Algebra Rules

$$Var(X) = Cov(X,X)$$

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

$$Cov(X,Y+Z) = Cov(X,Y) + Cov(X,Z)$$

# Example 1

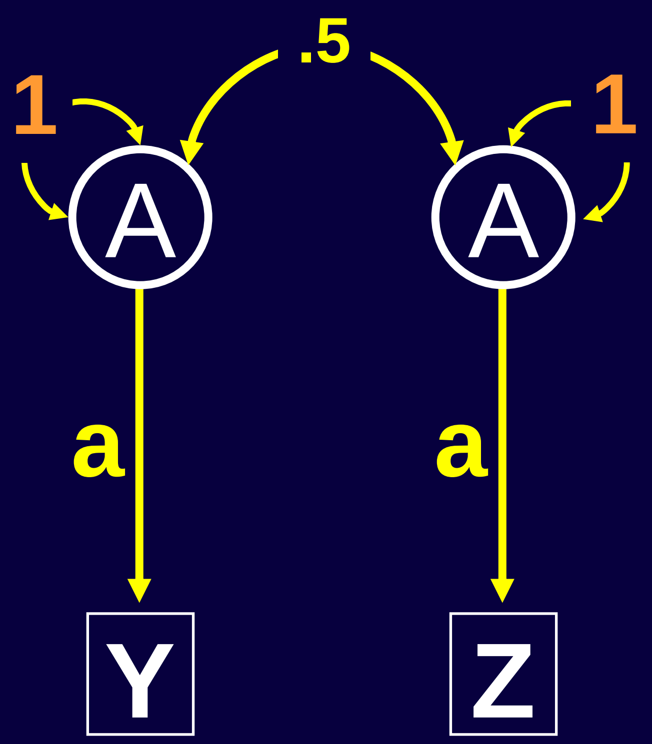



Y = aA

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

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 the specified model

- The Ultimate Goal is to fit Predicted Variances / Covariances to observed Variances / Covariances of the data in order to estimate model parameters :
  - regression coefficients, correlations