International Statistical Genetics Workshop • Boulder, Colorado • IBG

Conor Dolan
Biological psychology, VU,
Amsterdam

Dirk Pelt
Biological psychology, VU,
Amsterdam

PPT duration: 14 slides

Multivariate twin models

- Section 1: from **uni**variate to **bi**variate
- Section 2: from **bi**variate to **multi**variate
- Section 3: independent pathway model (IPM) a.k.a. biometric model
  - Recap multivariate model
  - The common factor model
- Section 4: the independent pathway model (continued)
- Section 5: the common pathway model

From bivariate to multivariate

From p=2 to P>2 .... Just larger matrices…

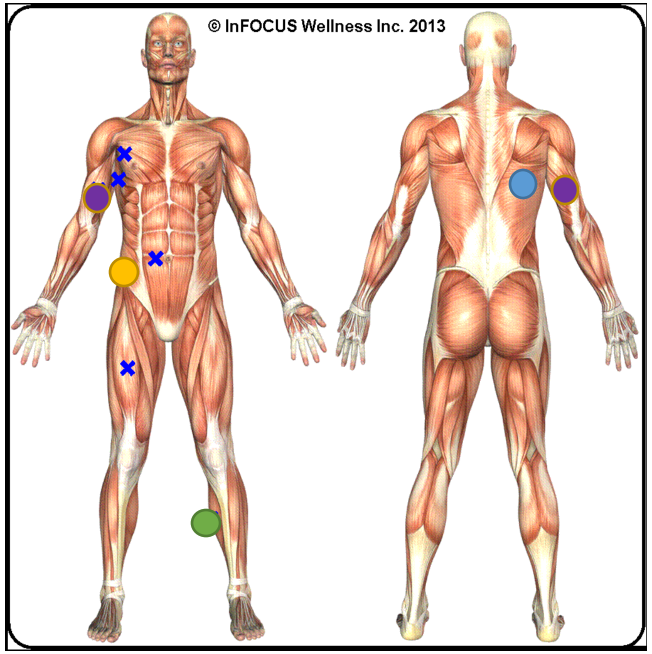
Refresher on the notation

Notation	
Phenotype (variable)	Ph (Ph_1, Ph_2) or sensible ("IQ")
Latent variable - additive genetic	A
Latent variable - shared environmental	C
Latent variable - dominance	D
Latent variable - unshared environmental	E
variance of phenotype	var(Ph) or σ_{Ph}^2
variance of A (D, C,E)	var(A) or σ_A^2 (σ_D^2 σ_C^2 σ_E^2)
standard deviation of A (D, C,E)	std(A) or σ_A (σ_D σ_C σ_E)
Covariance matrix of p phenotypes (pxp symmetric matrix) - Phenotypic	Σ_{Ph} containing p variances and ((p-1)*p)/2 covariances
Covariance matrix of p phenotypes (pxp symmetric matrix) - A (D, C, E)	Σ_A (Σ_D, Σ_C, Σ_E)
Covariance and correlation between X and Y (off-diagonal element in Σ_{Ph})	cov(X,Y) or $\sigma_{X,Y}$ or $s_{X,Y}$ cor(X,Y) or $\rho_{X,Y}$ or $r_{X,Y}$

The generalization from p=1 (univariate) to p=2 (bivariate) to p>2 (multivariate).

Σ_{Ph} MZ	p phenotypes	p phenotypes
	$\Sigma_A + \Sigma_D + \Sigma_E$	$\Sigma_A + \Sigma_D$
	$\Sigma_A + \Sigma_D$	$\Sigma_A + \Sigma_D + \Sigma_E$

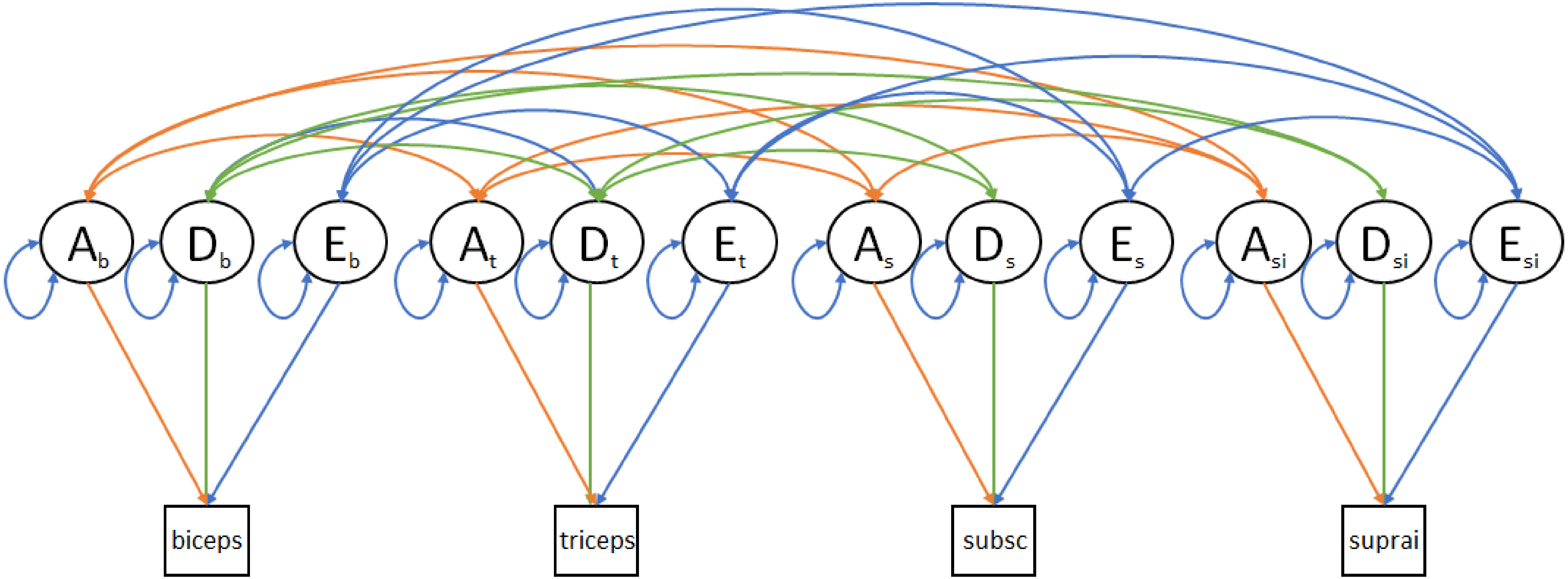
Σ_{Ph} DZ	p phenotypes	p phenotypes
	$\Sigma_A + \Sigma_D + \Sigma_E$	$\frac{1}{2}\Sigma_A + \frac{1}{4}\Sigma_D$
	$\frac{1}{2}\Sigma_A + \frac{1}{4}\Sigma_D$	$\Sigma_A + \Sigma_D + \Sigma_E$



Biceps and Triceps (purple ●), ~~Calf (green ●),~~

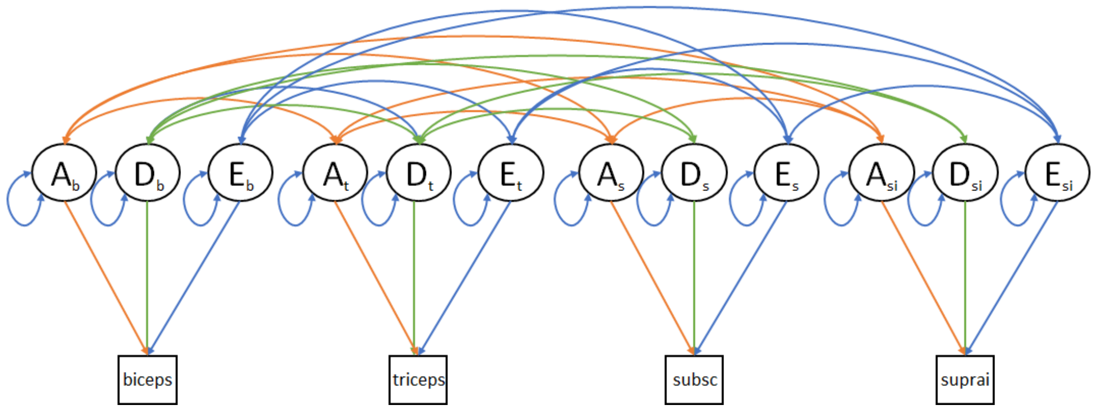
Subscapular (blue ●), Suprailiacal (orange ●).

Spaghetti

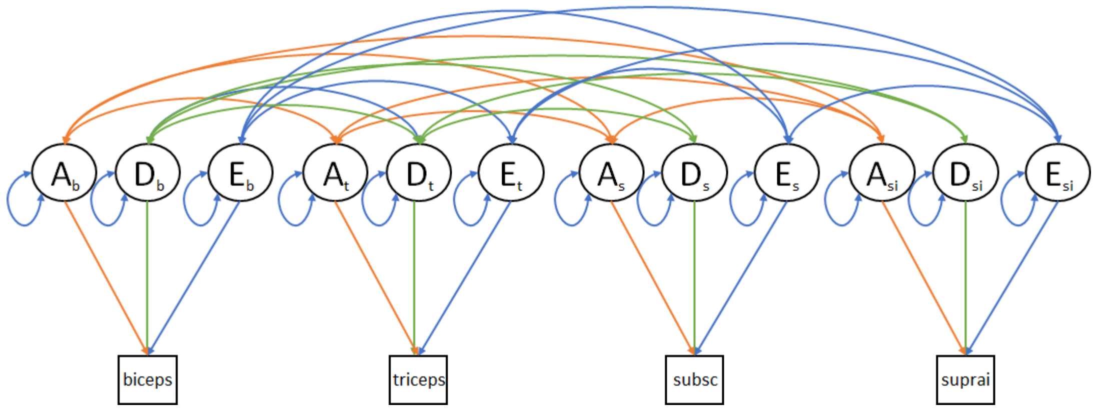

A_b
D_b
E_b
A_t
D_t
E_t
A_s
D_s
E_s
A_{si}
D_{si}
E_{si}
biceps
triceps
subsc
suprai

Spaghetti^2 = Spaghetti * Spaghetti

Connect the latent variable A, D, E according to the classical twin ADE model ....



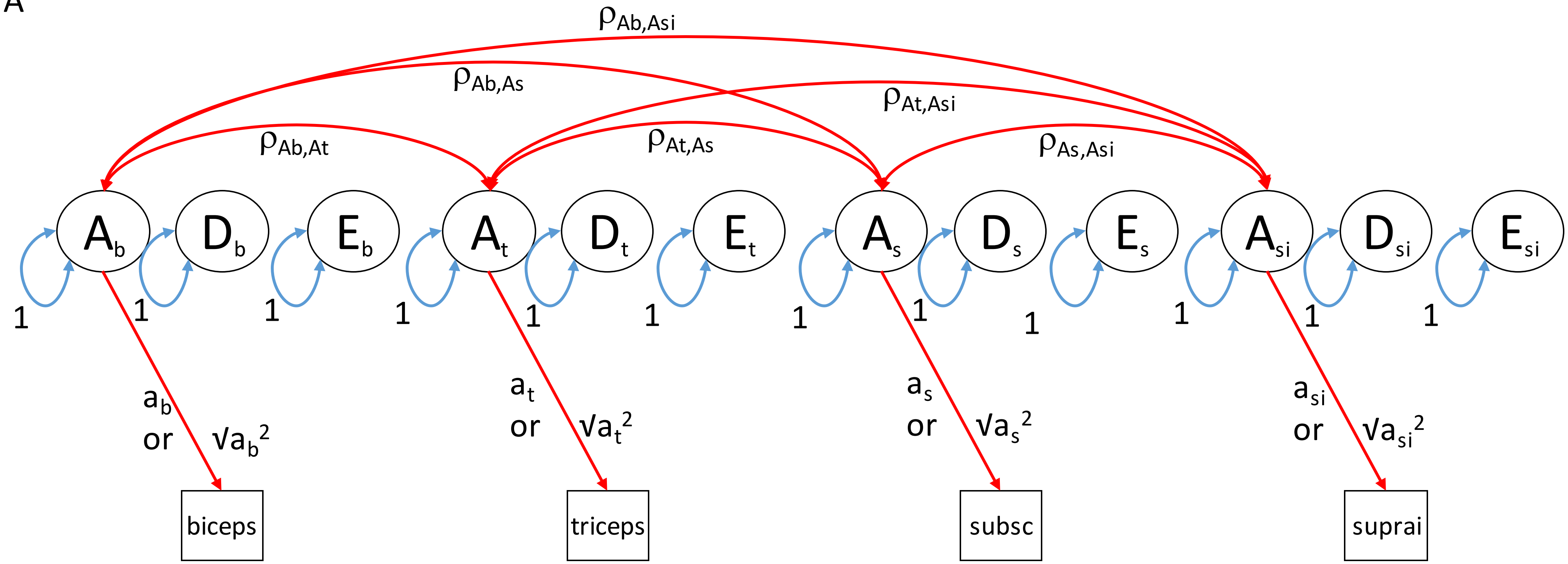
twin 1



twin 2

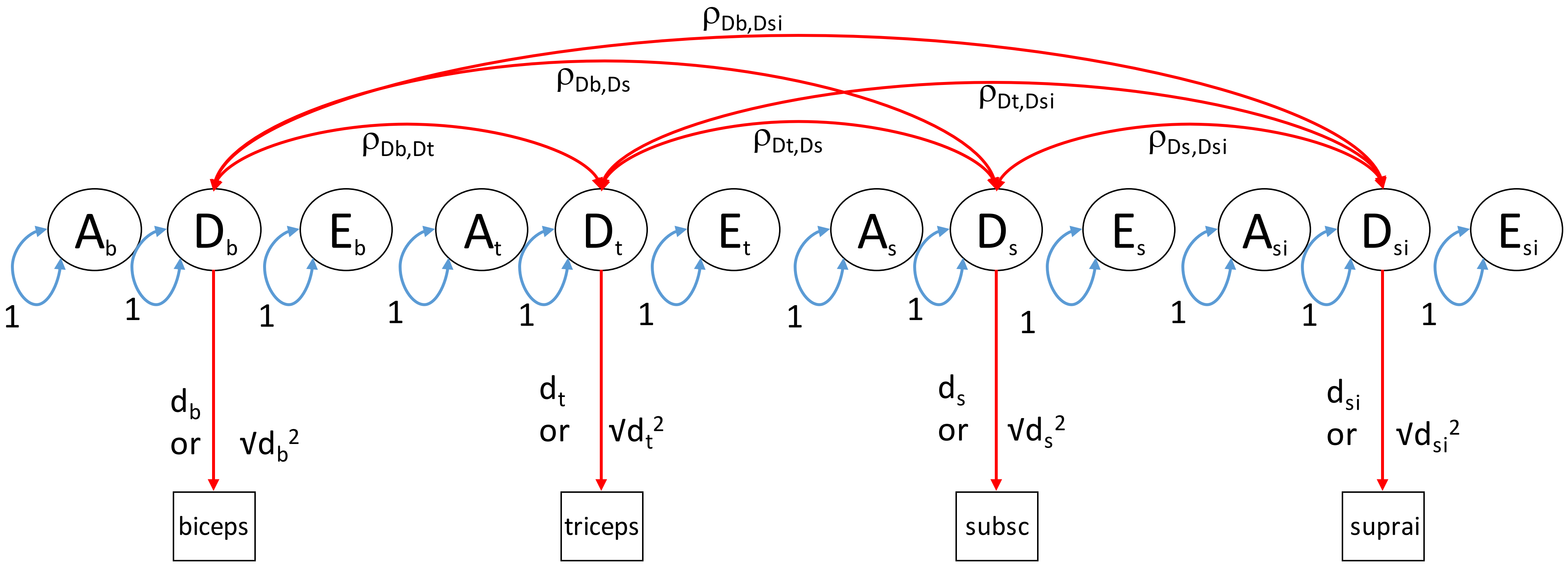
Good luck!

Σ_A



Number of estimated A parameters:
$(p*(p-1))/2 = (4*3)/2 = 6$ covariances / correlations
$p = 4$ variances

Σ_D

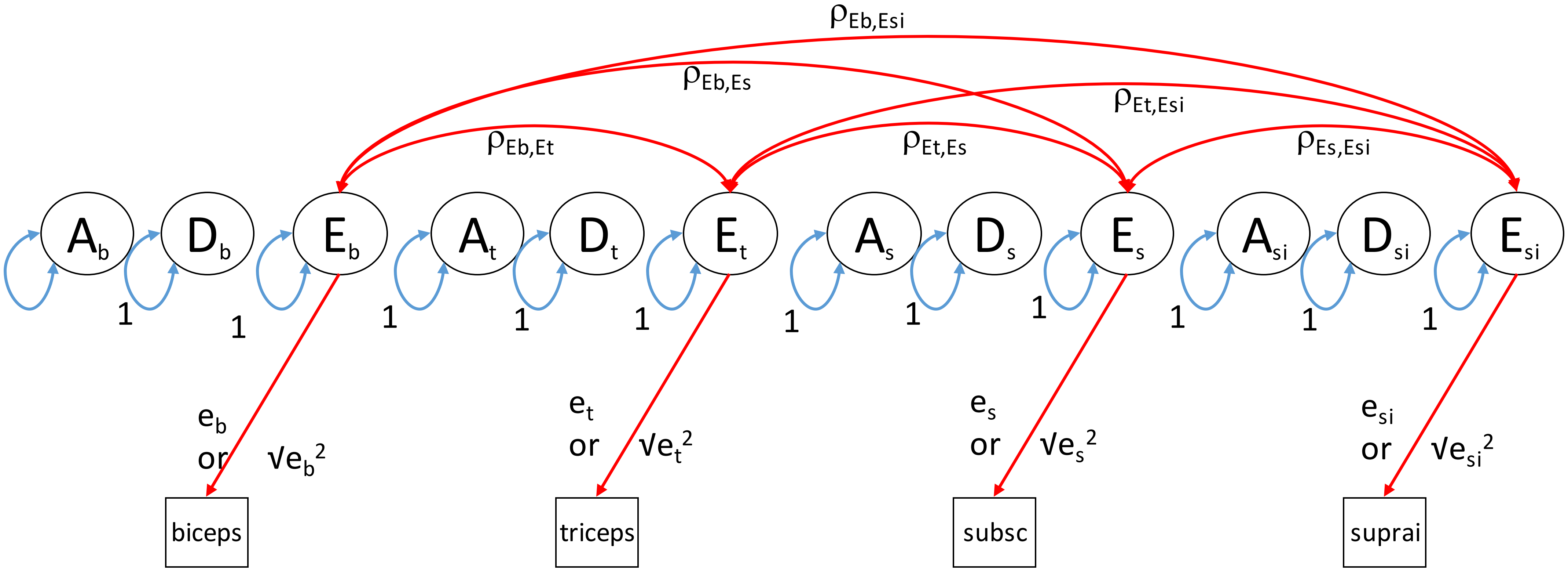


Number of estimated D parameters:

$(p*(p-1))/2 = (4*3)/2 = 6$ covariances / correlations

$p = 4$ variances

Σ_E



Number of estimated E parameters:
$(p*(p-1))/2 = (4*3)/2 = 6$ covariances / correlations
$p = 4$ variances

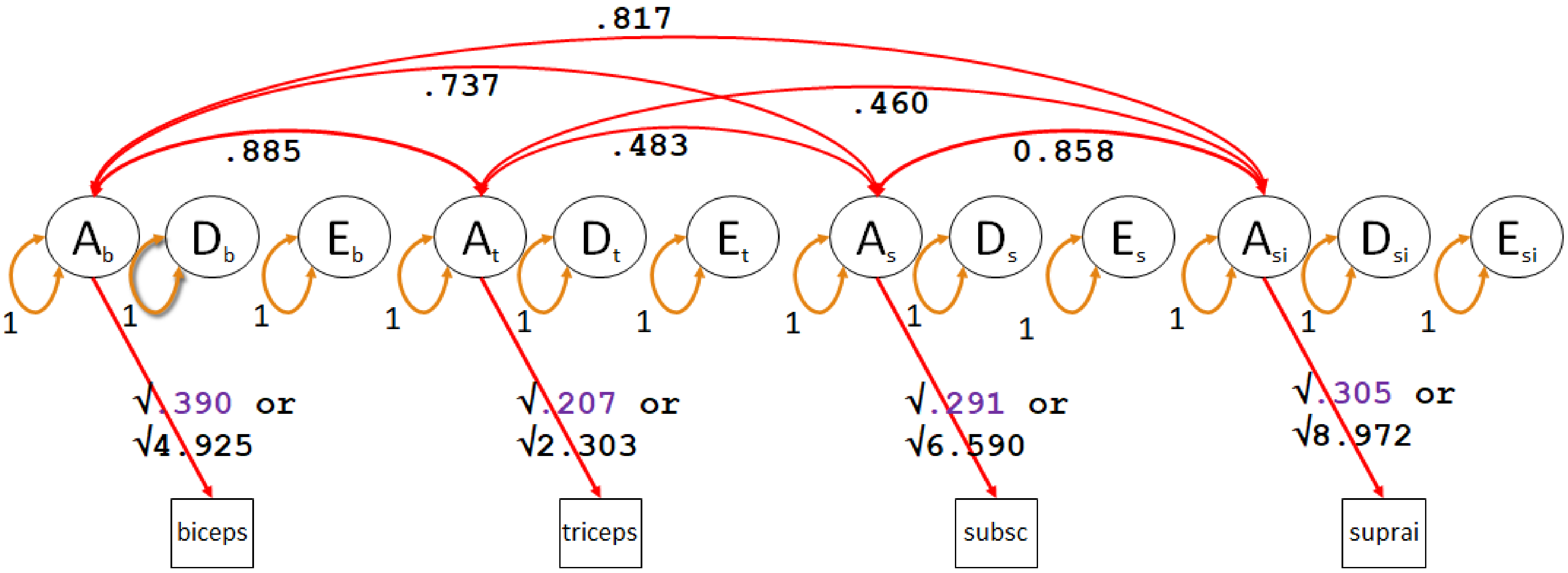
```
R MZ
        bic_T1 tri_T1 ssc_T1 sil_T1 bic_T2 tri_T2 ssc_T2 sil_T2
bic_T1   1.000
tri_T1   0.902  1.000
ssc_T1   0.816  0.809  1.000
sil_T1   0.859  0.837  0.924  1.000
bic_T2   0.787  0.723  0.695  0.735  1.000
tri_T2   0.752  0.808  0.680  0.725  0.889  1.000
ssc_T2   0.642  0.629  0.783  0.744  0.828  0.806  1.000
sil_T2   0.681  0.668  0.757  0.823  0.854  0.835  0.908  1.000
```

```
R DZ
        bic_T1 tri_T1 ssc_T1 sil_T1 bic_T2 tri_T2 ssc_T2 sil_T2
bic_T1   1.000
tri_T1   0.910  1.000
ssc_T1   0.881  0.893  1.000
sil_T1   0.892  0.881  0.935  1.000
bic_T2   0.324  0.280  0.267  0.276  1.000
tri_T2   0.264  0.272  0.231  0.237  0.890  1.000
ssc_T2   0.269  0.244  0.290  0.298  0.858  0.867  1.000
sil_T2   0.255  0.217  0.241  0.288  0.845  0.864  0.934  1.000
```

R_A	bic	tri	s	si
bic	1.000	0.885	0.737	0.817
tri	0.885	1.000	0.483	0.460
s	0.737	0.483	1.000	0.858
si	0.817	0.460	0.858	1.000
	bic	tri	s	si
σ_A^2	4.925	2.303	6.590	8.972
h^2	0.390	0.207	0.291	0.305

Additive genetic covariance matrix (4x4) Σ_A

$\text{cov2cor}(\Sigma_A) = R_A$



R_A	bic	tri	s	si
bic	1.000	0.885	0.737	0.817
tri	0.885	1.000	0.483	0.460
s	0.737	0.483	1.000	0.858
si	0.817	0.460	0.858	1.000
	bic	tri	s	si
σ_A^2	4.925	2.303	6.590	8.972
h^2	0.390	0.207	0.291	0.305

R_E	bic	tri	s	si
bic	1.000	0.789	0.726	0.777
tri	0.789	1.000	0.761	0.765
s	0.726	0.761	1.000	0.836
si	0.777	0.765	0.836	1.000
	bic	tri	s	si
σ_E^2	2.725	2.115	4.732	5.252
	0.216	0.190	0.209	0.179

R_D	bic	tri	s	si
bic	1.000	0.989	0.983	0.941
tri	0.989	1.000	1.041	1.069
s	0.983	1.041	1.000	0.999
si	0.941	1.069	0.999	1.000
	bic	tri	s	si
σ_D^2	4.975	6.700	11.29	15.18
	0.394	0.603	0.499	0.516

Conclusion 1: h_b^2 are high: .78, .81, .79, .82

Conclusion 2: the A, D, E correlations are high

Conclusion 3: A+D account for<80% of the covariance

Conclusion 4: There appear to be a single set of genes that contribution dominance effects

Conclusion 3: A+D account for > 80% of the covariance

$$\Sigma_{Ph} = \Sigma_A + \Sigma_D + \Sigma_E$$

Σ_A / Σ_{Ph} } diagonals narrow sense heritabilities

$(\Sigma_A + \Sigma_D) / \Sigma_{Ph}$ } diagonals broad sense heritabilities
off diagonals > .80

Conclusion 3: *A+D account for<80% of the covariance*

$$\Sigma_{Ph} = \Sigma_A + \Sigma_D + \Sigma_E$$

$$[\Sigma_A + \Sigma_D]/\Sigma_{Ph}$$

	bic	tri	s	si
bic	0.784	0.821	0.816	0.822
tri	0.821	0.810	0.820	0.835
s	0.816	0.820	0.791	0.825
si	0.822	0.835	0.825	0.821

Conclusion 4: *There appear to be a single set of genes that contribution dominance effects*

Conclusion based on what?

Eyeballing the D correlation matrix and Independent Pathway Modeling (IPM)!

The next PPT presentation is about the IPM