The Causes of Variation

2016 International Workshop on Statistical Genetic Methods

for Human Complex Traits

Boulder, CO.

Lindon Eaves,

VIPBG, Richmond VA.

March 2016

GOALS – “Orientation”

- To outline some of the basic issues that we will address this week
- To set some of the issues in their historical context in the story of genetics and its application to human variation.

“Genetics”

The Study of
Variation and Heredity

“Variation”

“Why aren’t we all the same?”

“Heredity”

“Why do things run in families?”

“VARIATION”

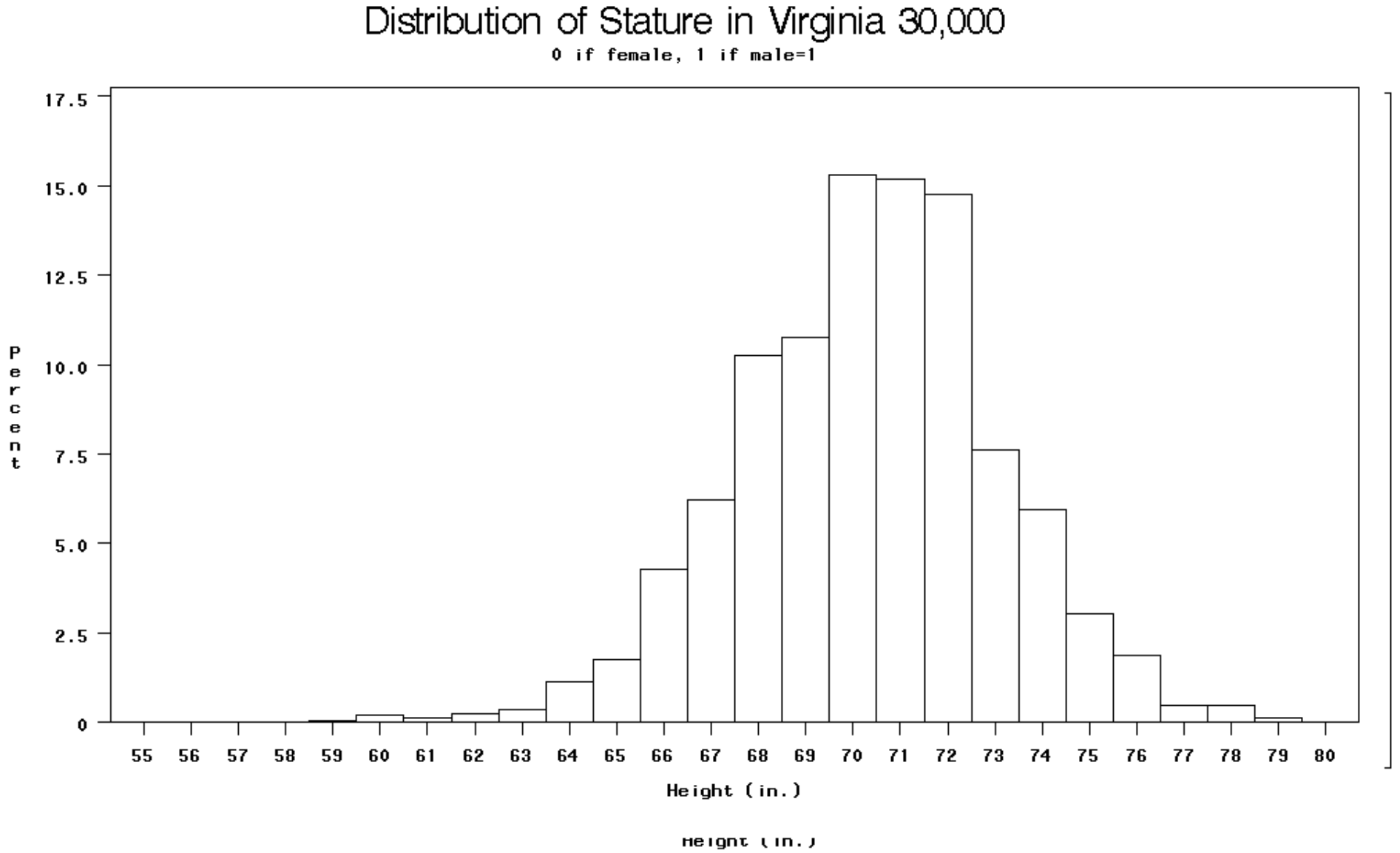
Gregor Mendel (1822-1885)

1865: Experiments in Plant Hybridisation

Mendel

- Binary traits (“White vs Pink”)
- Particulate (“Mendelian”) Inheritance
- Simple mathematical and statistical laws
- Corresponded to behavior of chromosomes
- Explained inborn errors of metabolism ->Medical genetics, “gene-hunting”
- Provided Mechanism for Natural Selection
- Illuminated by structure of DNA

Continuous variation


Distribution of Stature in Virginia 30,000
0 if female, 1 if male=1
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
Percent
55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
Height (in.)
Height (in.)

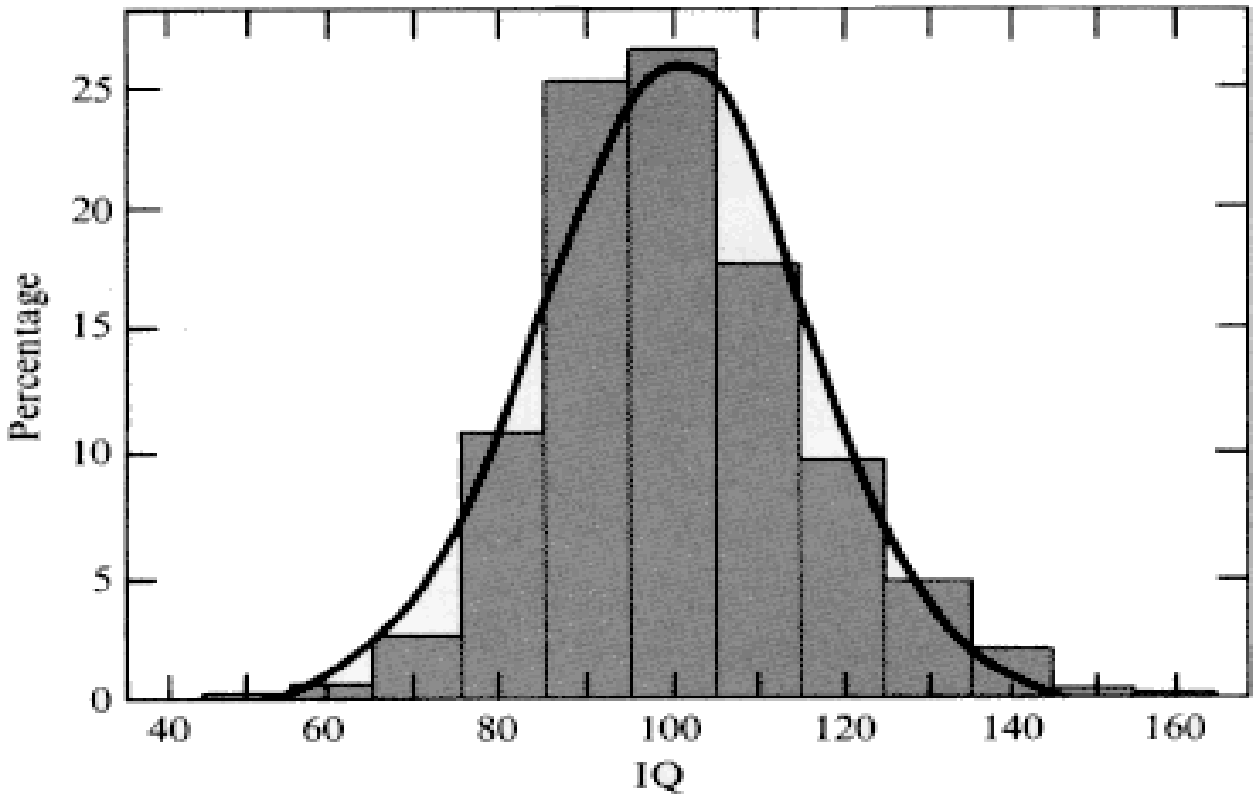
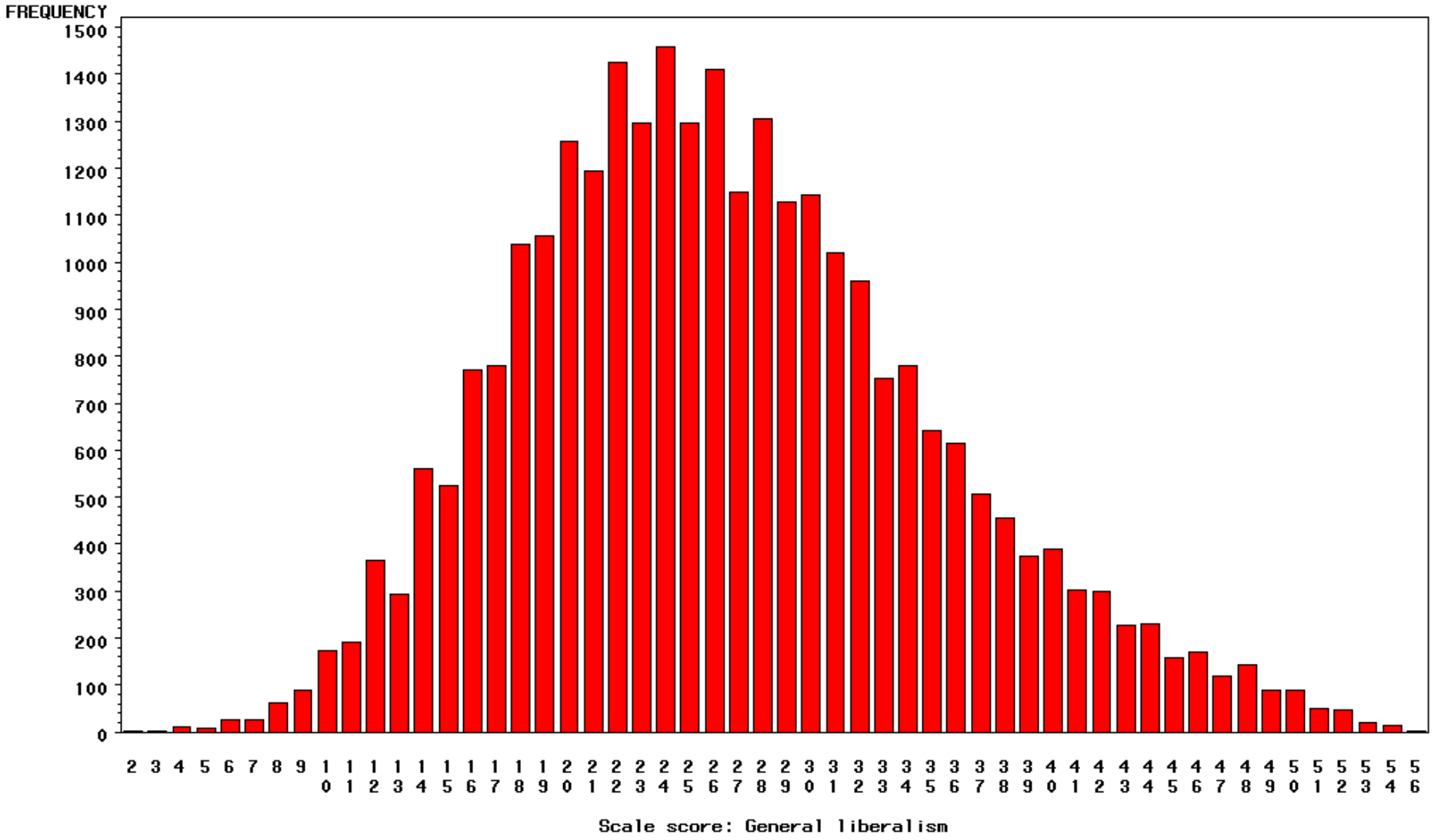


FIGURE 9.4
The distribution of IQ among the 14,963 children born in Scotland on February 1, May 1, August 1, and November 1, 1926. The shaded histogram shows the percentages of the group with IQ's in various ranges of 10 points. This grouping is artificial and is done solely for ease of representation: it does not imply any discontinuity in the values of IQ that children can show. The continuous curve shows the ideal distribution calculated from the observations and representing the statistical population of which the children actually observed are regarded as forming a sample. (Data from MacMeekan; from Mather 1964.)

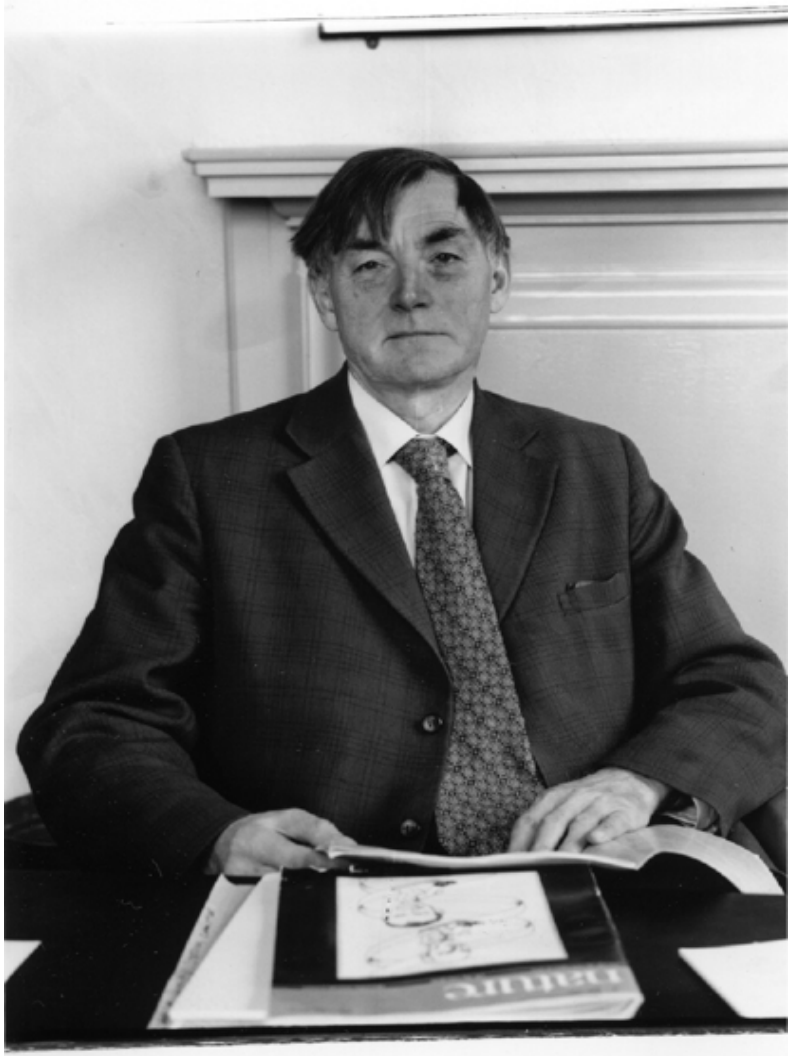
“Liberalism”


FREQUENCY
1500
1400
1300
1200
1100
1000
900
800
700
600
500
400
300
200
100
0
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 56
Scale score: General liberalism

Categorical Outcomes

Often called “threshold traits” because people “affected” if they fall above some level (“threshold”) of a measured or hypothesized continuous trait.

Douglas Scott Falconer, FRS, FRSE (1913-2004)



1965 Inheritance of liability to certain diseases estimated from incidence among relatives., *Ann. Hum. Genet.***29:**51ff.

1960 *Introduction to Quantitative Genetics.* Edinburgh: Oliver and Boyd.

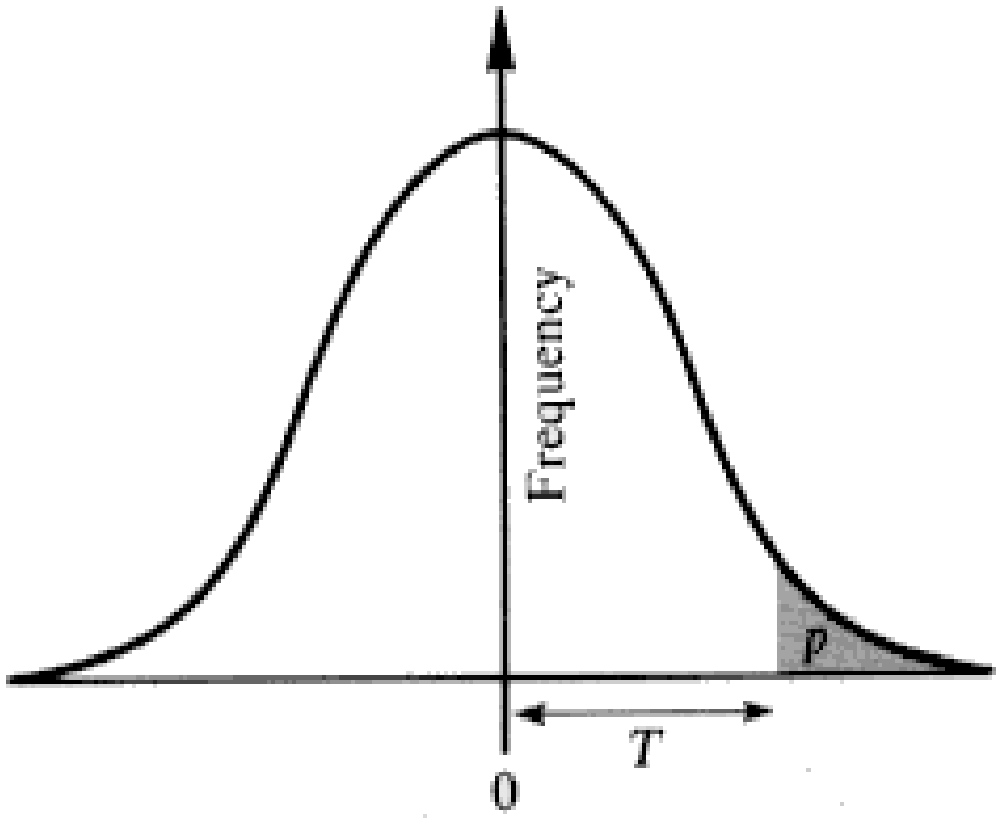
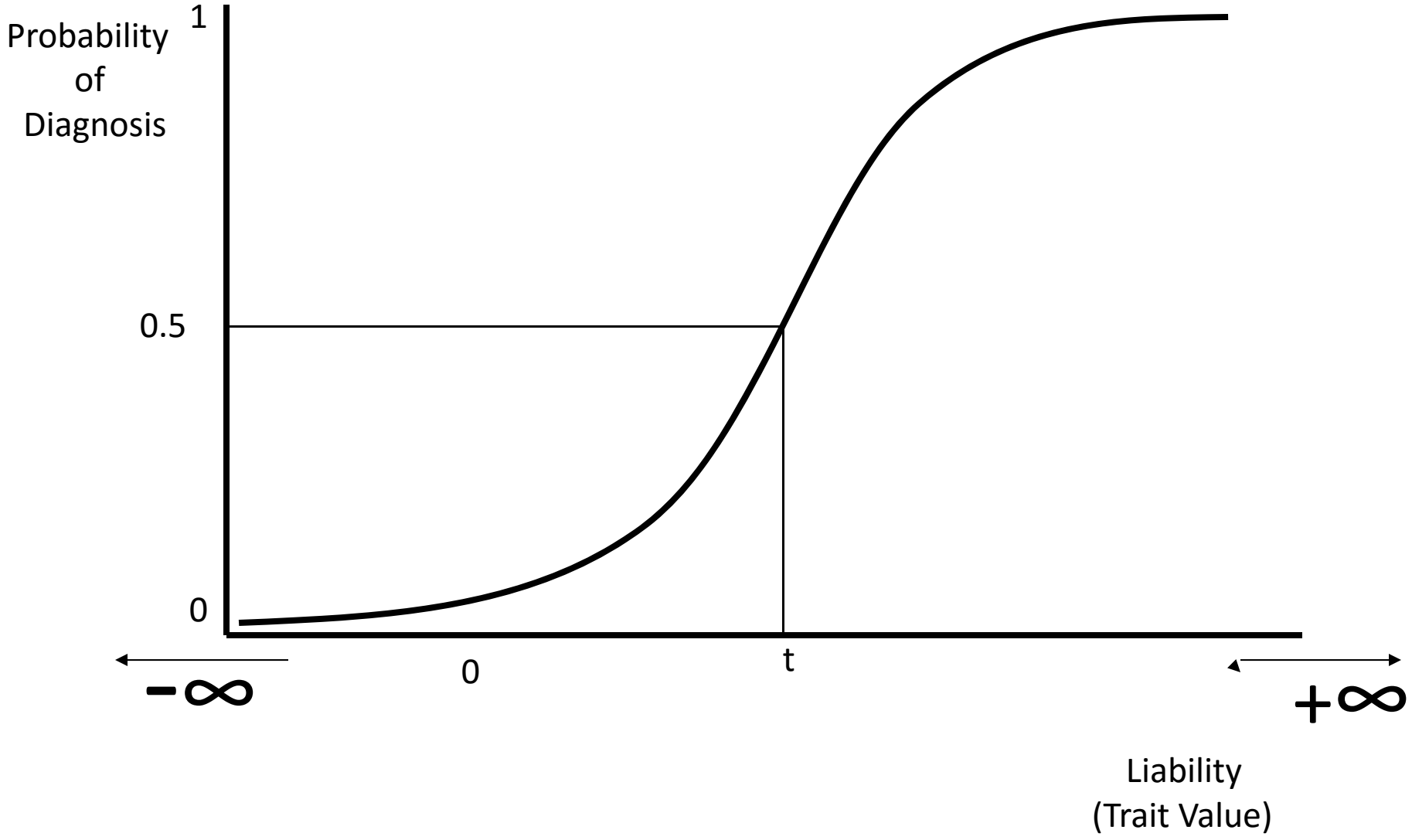


FIGURE 9.5
Threshold model. All individuals with a value of x greater than T are affected. The proportion of affected individuals is the area under the distribution curve beyond T.

∞

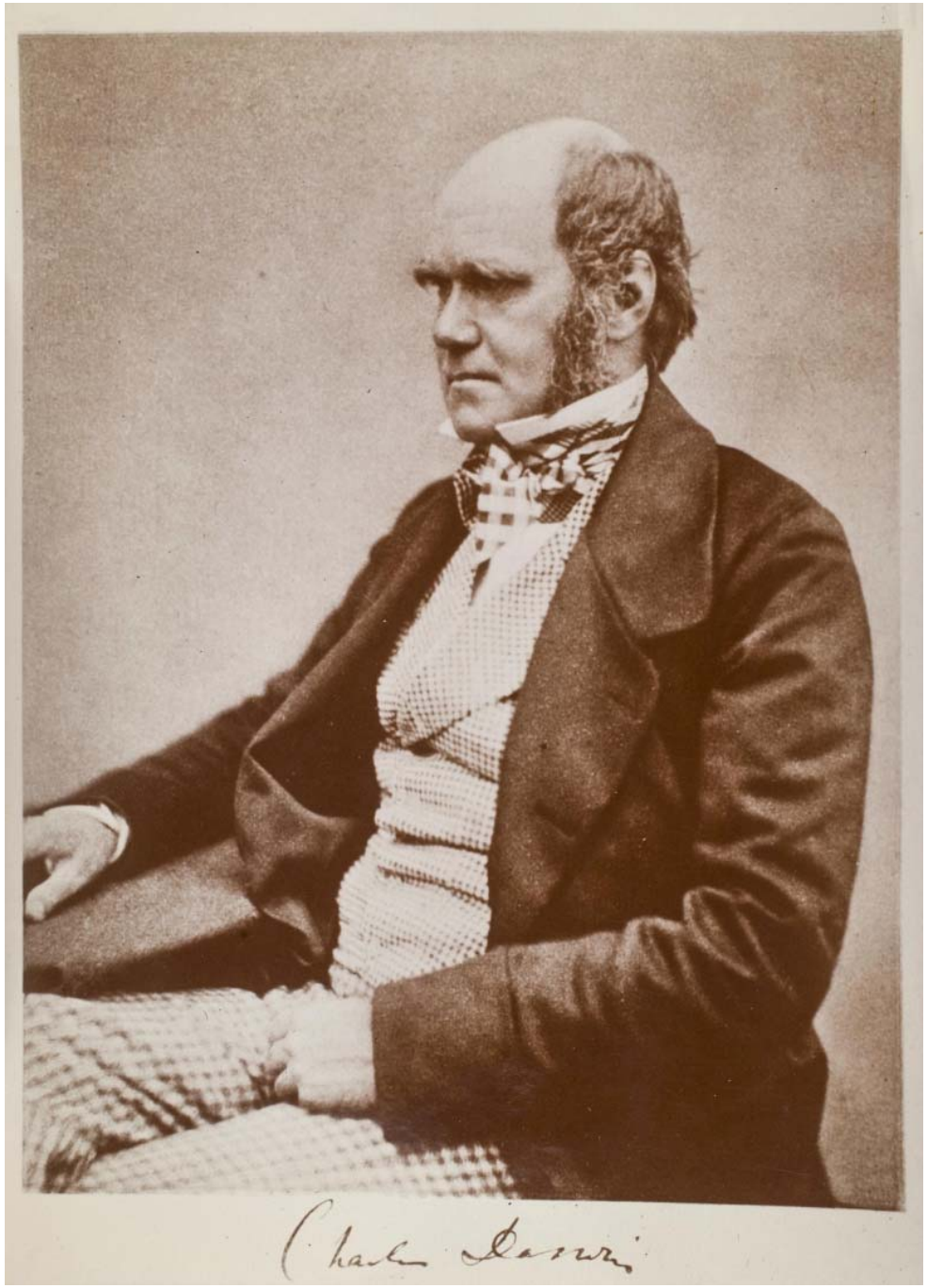
Relationship between continuous normal “liability” and risk of “diagnosis” (see I.R.T.)



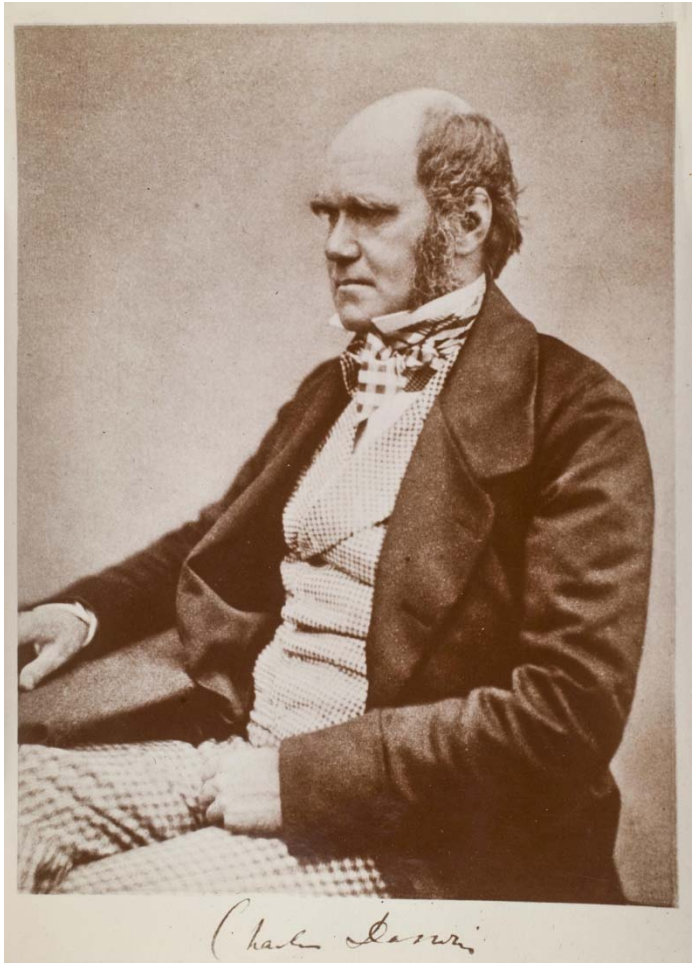
Question that bugs me:

How do you get from “liability”
to “catastrophe”?

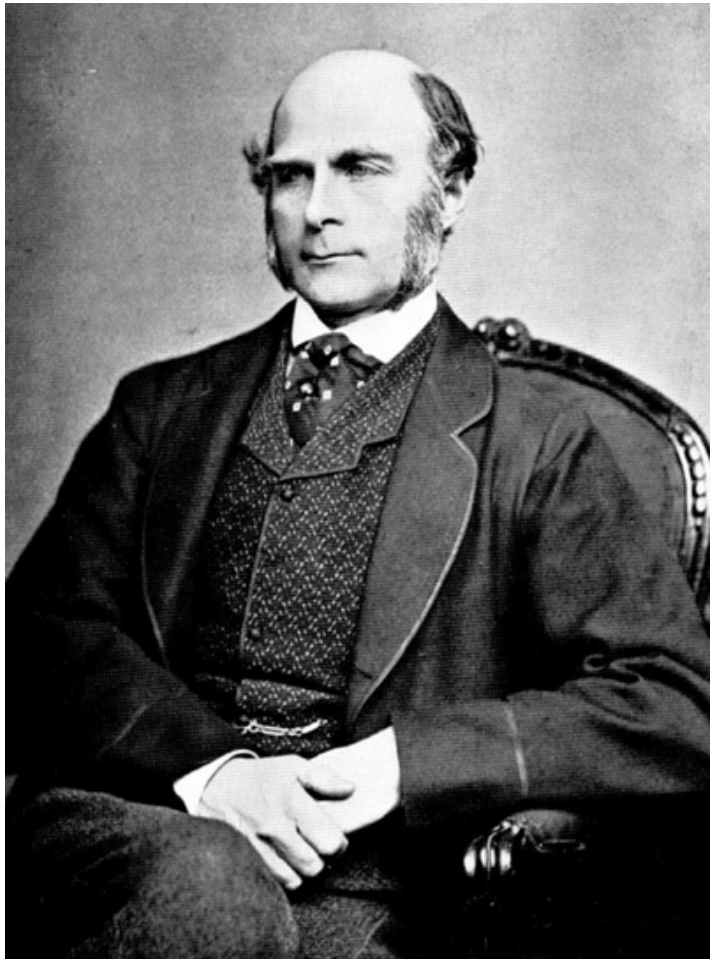
“HEREDITY”


Charles Darwin

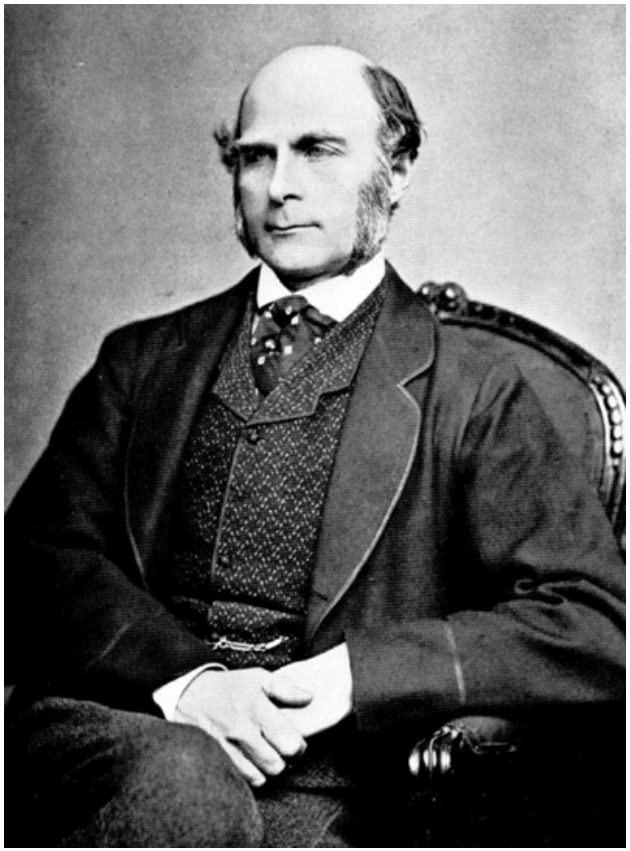
Charles Darwin (1809-1882)

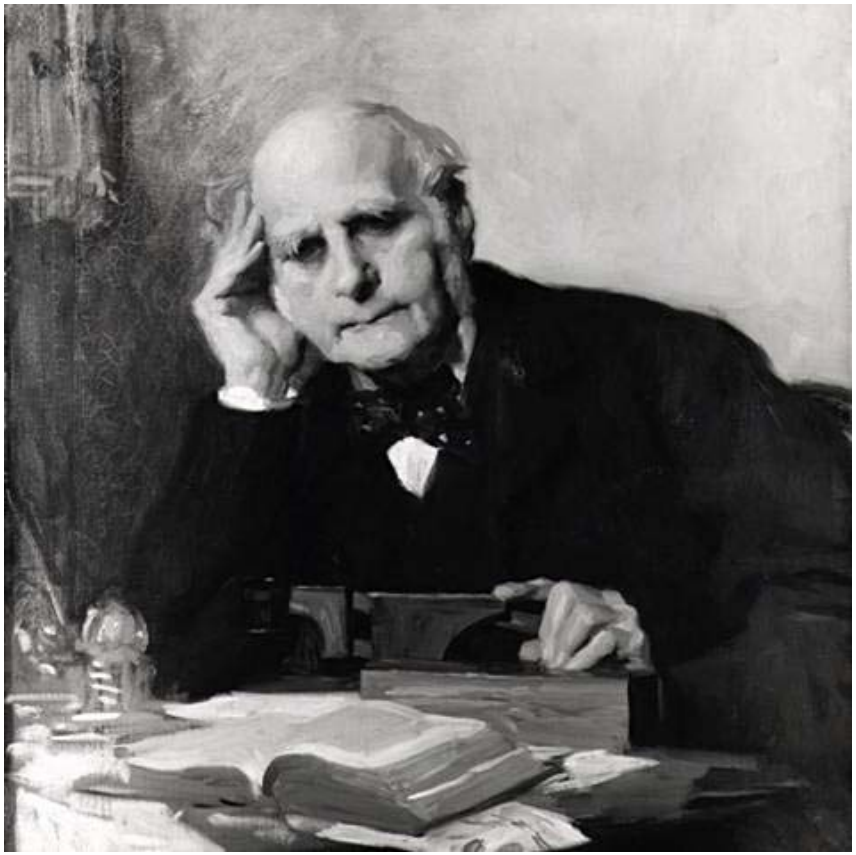


1865: On the Origin of Species



Francis Galton (1822-1911)



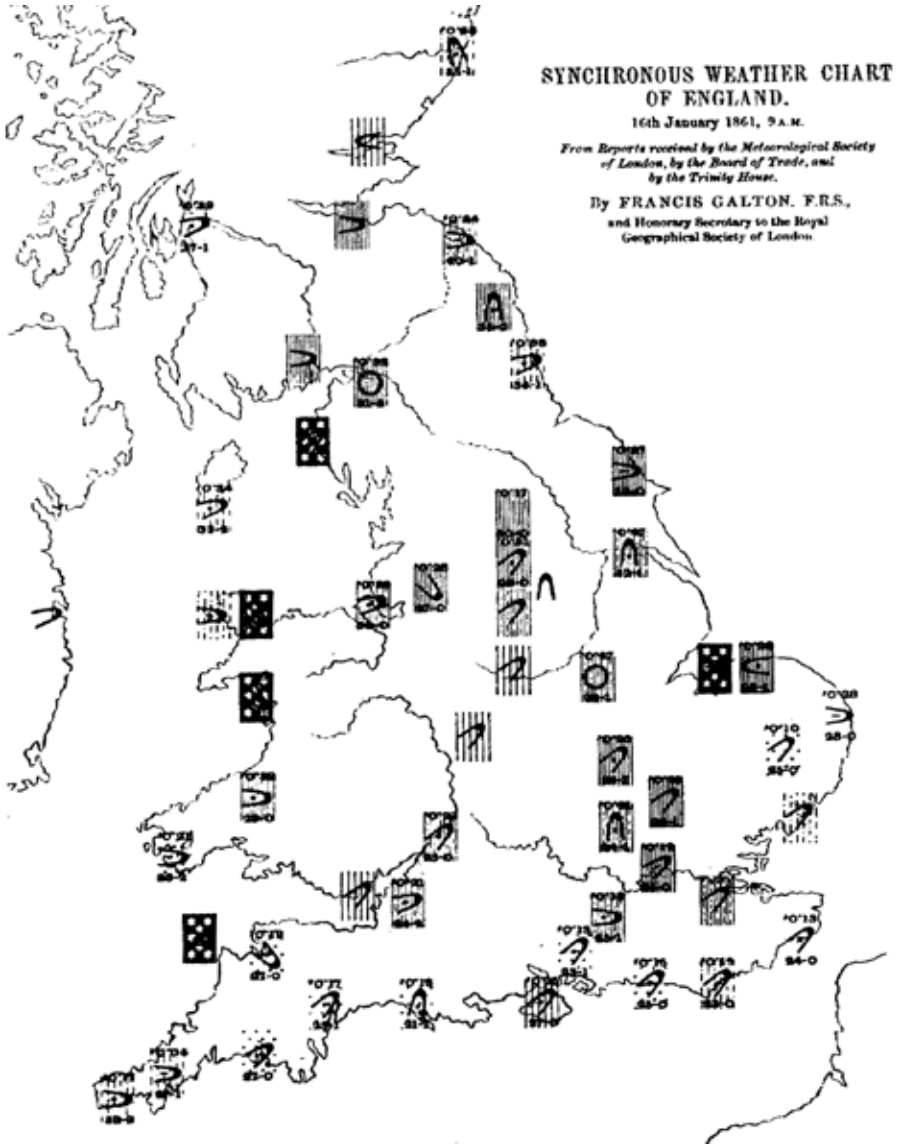


1869: Hereditary Genius

1883: Inquiries into Human Faculty and its Development

1884-5: Anthropometic Laboratory at “National Health Exhibition”

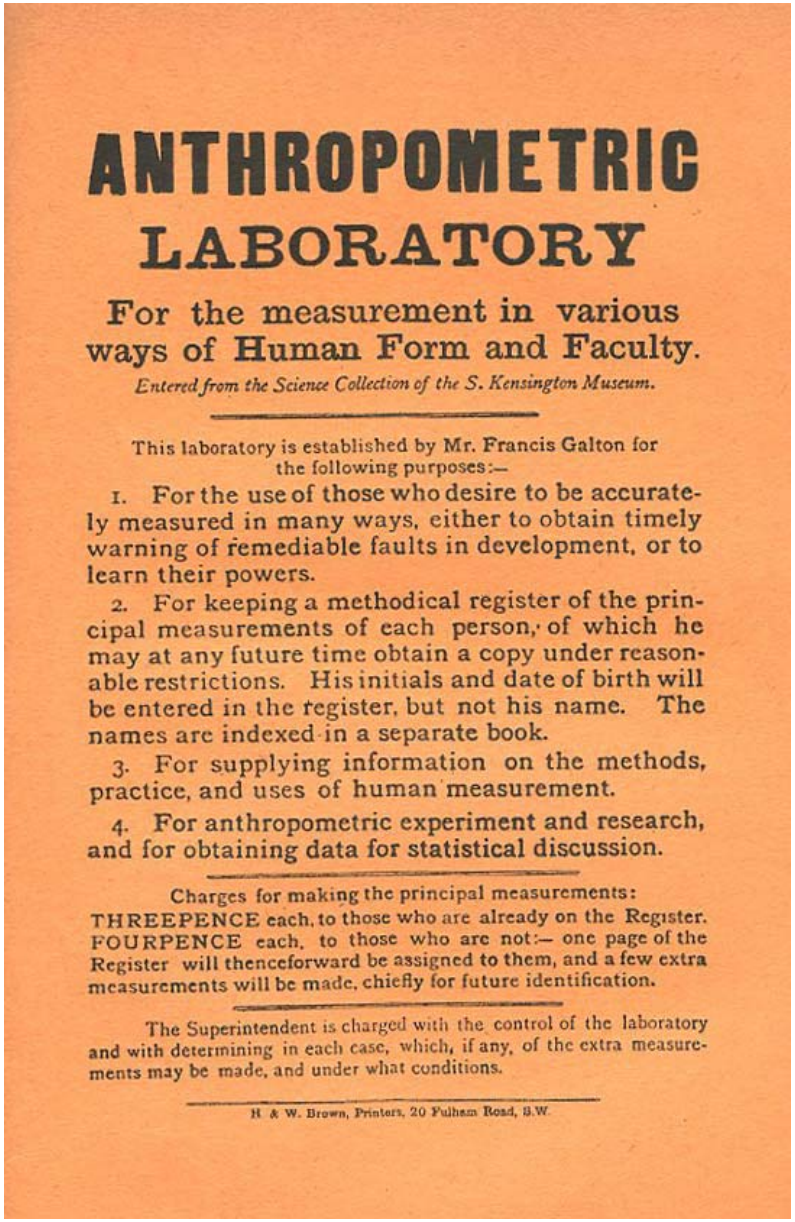
Galton's Other Work e.g. Meteorology



Hereditary Genius (1869, p 317)

	Judges, p. 61.	Statesmen, p. 109.	Commanders, p. 148.	Literary, p. 171.	Scientific, p. 195.	Poets, p. 227.	Artists, pp. 238 and 249.	Divines, p. 275.	Illustrious and Eminent Men of all Classes.		
	B.	B.	B.	B.	B.	B.	B.	B.	B.	C.	D.
Father	26	33	47	48	26	20	32	28	31	100	31
Brother	35	39	50	42	47	40	50	36	41	150	27
Son	36	49	31	51	60	45	89	40	48	100	48
Grandfather	15	28	16	24	14	5	7	20	17	200	8
Uncle	18	18	8	24	16	5	14	40	18	400	5
Nephew	19	18	35	24	23	50	18	4	22	400	5
Grandson	19	10	12	9	14	5	18	16	14	200	7
Great-grandfather	2	8	8	3	0	0	0	4	3	400	1
Great-uncle	4	5	8	6	5	5	7	4	5	800	1
First cousin	11	21	20	18	16	0	1	8	13	800	2
Great-nephew	17	5	8	6	16	10	0	0	10	800	1
Great-grandson	6	0	0	3	7	0	0	0	3	400	1
All more remote	14	37	44	15	23	5	18	16	31	?	...

Galton's Anthropometric Laboratory (1884-1885)



ANTHROPOMETRIC LABORATORY

For the measurement in various ways of Human Form and Faculty.

Entered from the Science Collection of the S. Kensington Museum.

This laboratory is established by Mr. Francis Galton for the following purposes:—

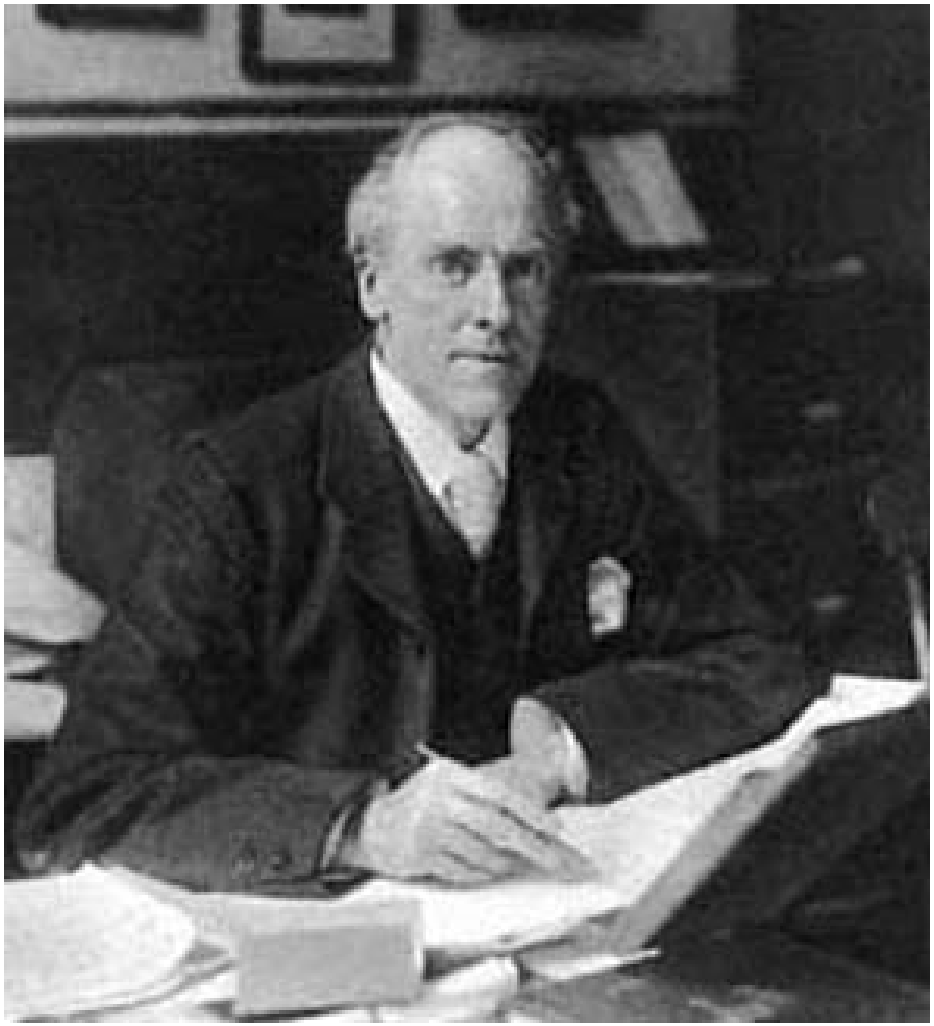
1. For the use of those who desire to be accurately measured in many ways, either to obtain timely warning of remediable faults in development, or to learn their powers.
2. For keeping a methodical register of the principal measurements of each person, of which he may at any future time obtain a copy under reasonable restrictions. His initials and date of birth will be entered in the register, but not his name. The names are indexed in a separate book.
3. For supplying information on the methods, practice, and uses of human measurement.
4. For anthropometric experiment and research, and for obtaining data for statistical discussion.

Charges for making the principal measurements:
THREEPENCE each, to those who are already on the Register.
FOURPENCE each, to those who are not:— one page of the Register will thenceforward be assigned to them, and a few extra measurements will be made, chiefly for future identification.

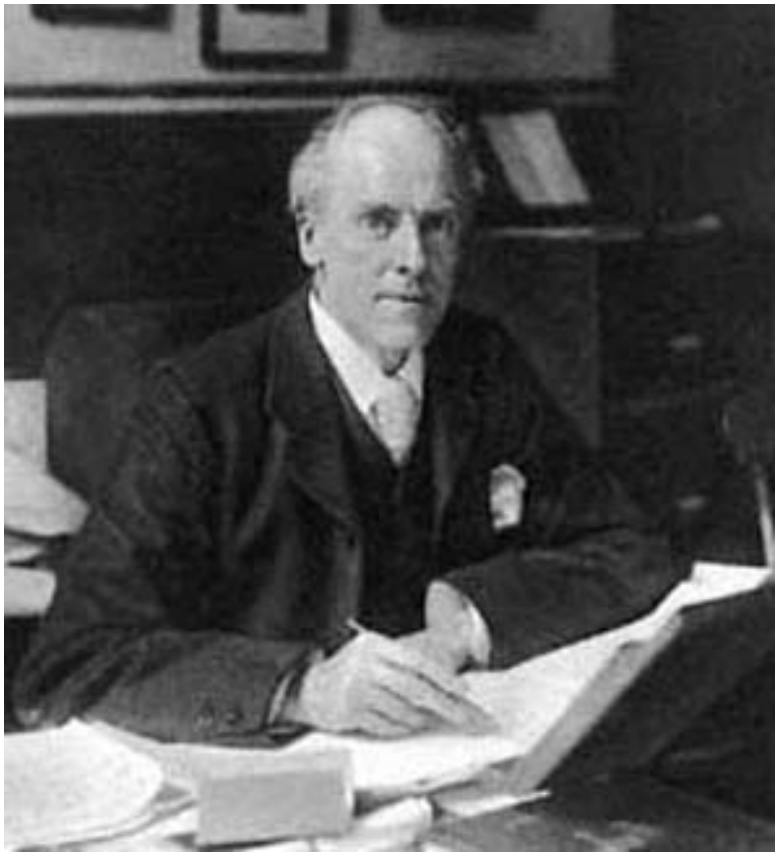
The Superintendent is charged with the control of the laboratory and with determining in each case, which, if any, of the extra measurements may be made, and under what conditions.

H. & W. Brown, Printers, 20 Fulham Road, S.W.

Francis Galton's First Anthropometric Laboratory at the International Health Exhibition, South Kensington, 1884–5.



Karl Pearson (1857-1936)



1903: On the Laws of Inheritance in Man: I Physical Characteristics (with Alice Lee)
1904: II Mental and Moral Characteristics
1914: The Life, Letters and Labours of Francis Galton

FAMILY MEASUREMENTS.

Professor KARL PEARSON, of University College, London, would esteem it a great favour if any persons in a position to do so, would assist him by making one set (or if possible several sets) of anthropometric measurements on their own family, or on families with whom they are acquainted. The measurements are to be made use of for testing theories of heredity, no names, except that of the recorder, are required, but the Professor trusts to the *bona fides* of each recorder to send only correct results.

Each family should consist of a father, mother, and at least one son or daughter, not necessarily the eldest. The sons or daughters are to be at least 18 years of age, and measurements are to be made on not more than two sons and two daughters of the same family. If more than two sons or two daughters are easily accessible, then not the tallest but the eldest of those accessible should be selected.

To be of real service the whole series ought to contain 1000—2000 families, and therefore the Professor will be only too grateful if anyone will undertake several families for him.

Copies of this paper, together with cards for recording data, may be obtained from

.....

or from the above-named Professor.

Pearson and Lee's diagram for measurement of "span" (finger-tip to finger-tip distance)

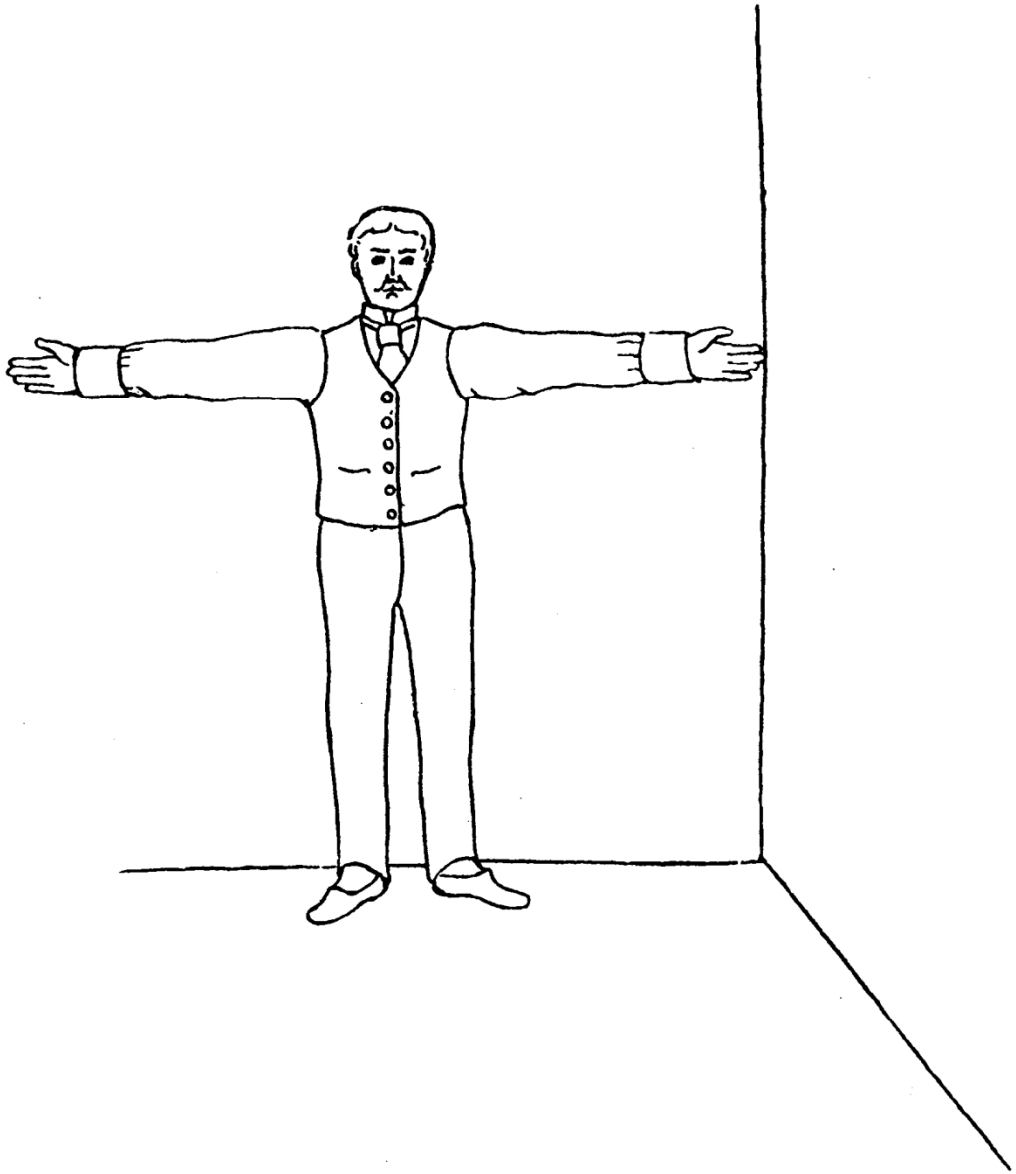


TABLE IV.

Coefficients of Heredity. Parents and Offspring.

Character	Father and		Mother and	
	Son	Daughter	Son	Daughter
Stature	·514 ± ·015	·510 ± ·013	·494 ± ·016	·507 ± ·014
Span	·454 ± ·016	·454 ± ·014	·457 ± ·016	·452 ± ·015
Forearm	·421 ± ·017	·422 ± ·015	·406 ± ·017	·421 ± ·015

From Pearson and Lee (1903) p.378

Correlation Coefficients for Direct Fraternal Heredity.

Character	Brother and Brother	Sister and Sister	Brother and Sister	Mean
Stature	·511 ± ·028	·537 ± ·022	·553 ± ·013	·534
Span	·549 ± ·026	·555 ± ·021	·525 ± ·013	·543
Forearm	·491 ± ·029	·507 ± ·023	·440 ± ·015	·479
Mean	·517	·533	·506	·519
Eye Colour*	·517 ± ·020	·446 ± ·023	·462 ± ·022	·475
Total mean	·517	·511	·495	·508

From Pearson and Lee (1903) p.387

Assortative Mating. Based on 1000 *to* 1050 *Cases of Husband and Wife.*

	Husband's Character	Wife's Character	Correlation and Probable Error	Symbol
Direct	Stature	Stature	·2804 ± ·0189	r_{12}
	Span	Span	·1989 ± ·0204	r_{34}
	Forearm	Forearm	·1977 ± ·0205	r_{56}

From Pearson and Lee (1903) p. 373

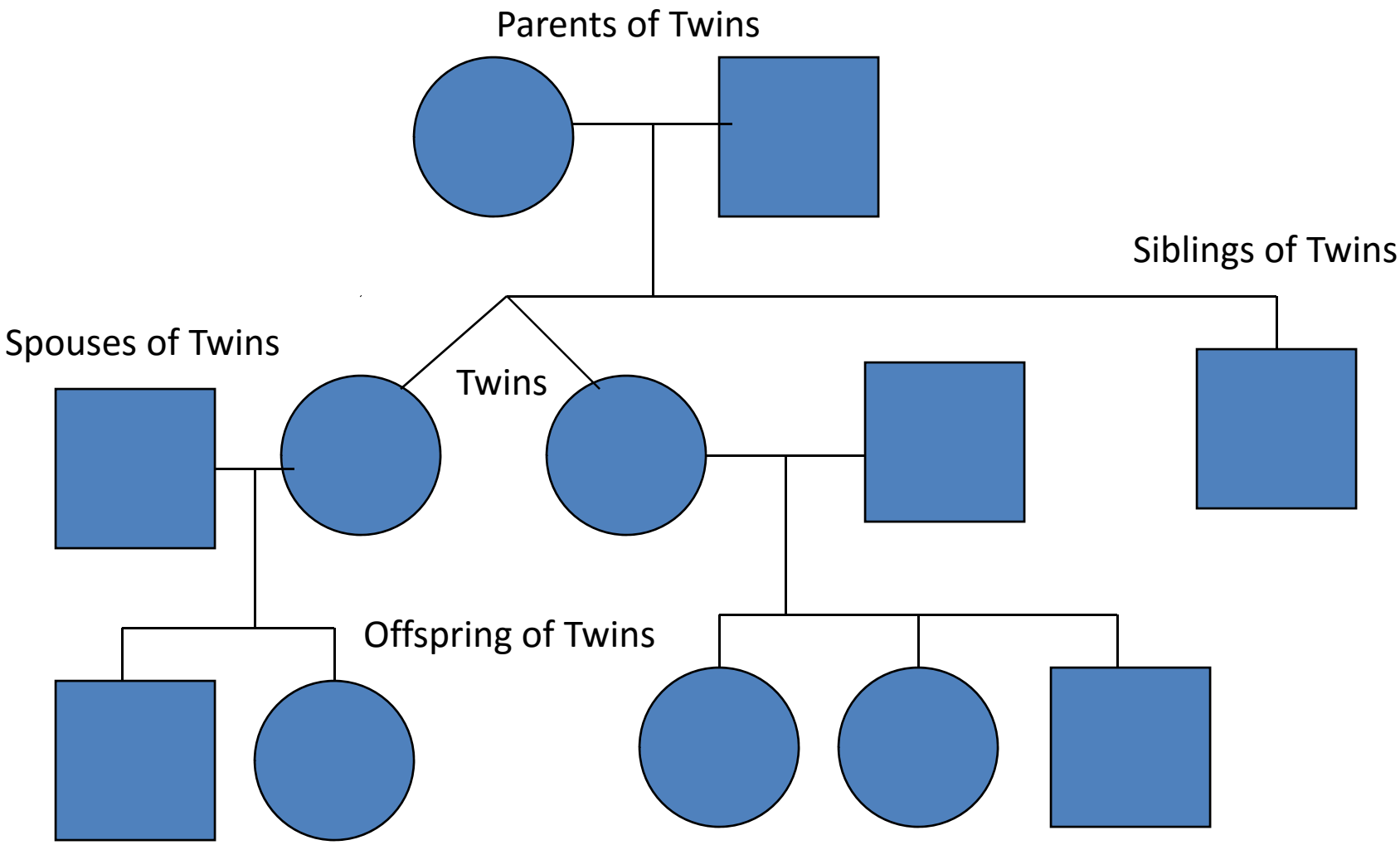
Modern Data

The Virginia 30,000
(N=29691)

The Australia 22,000

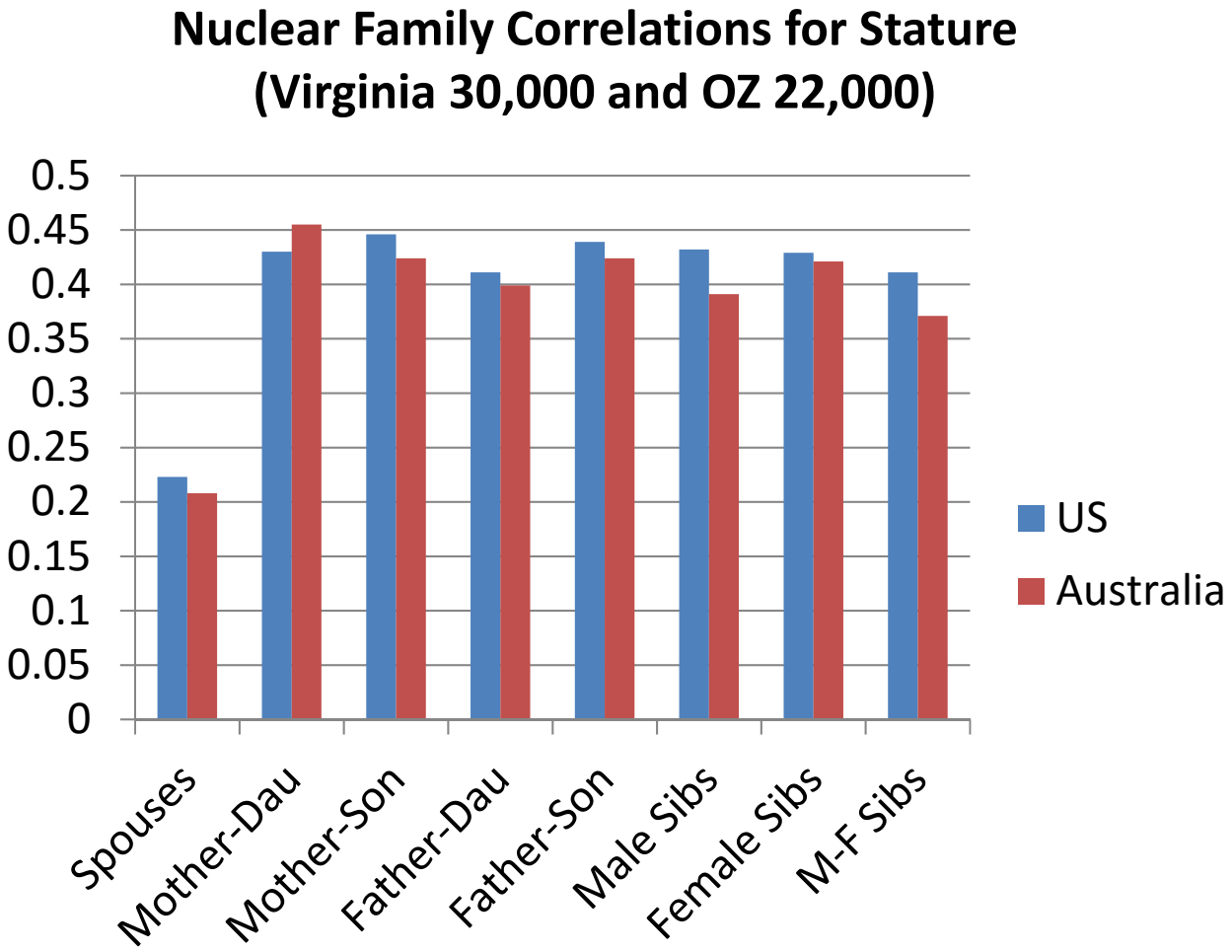
(N=20480)

ANZUS 50K: Extended Kinships of Twins

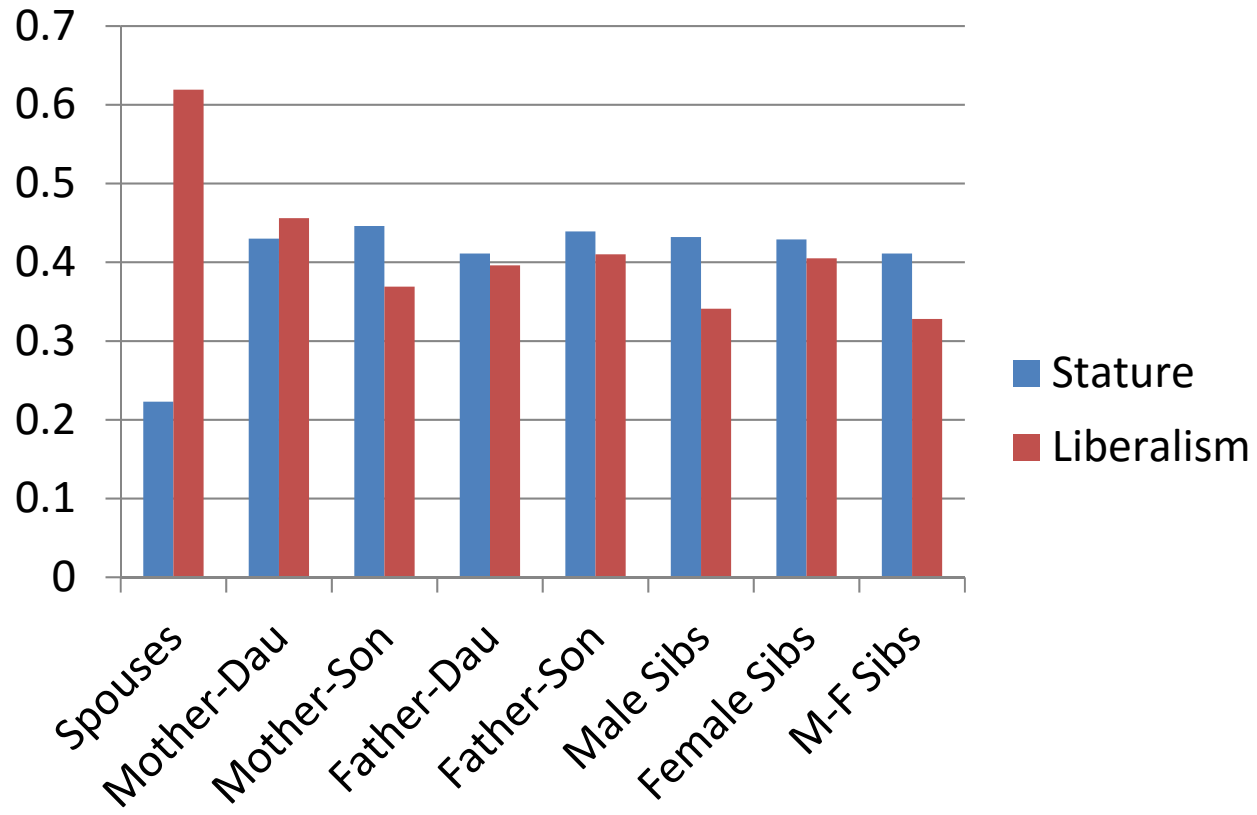


Overall sample sizes

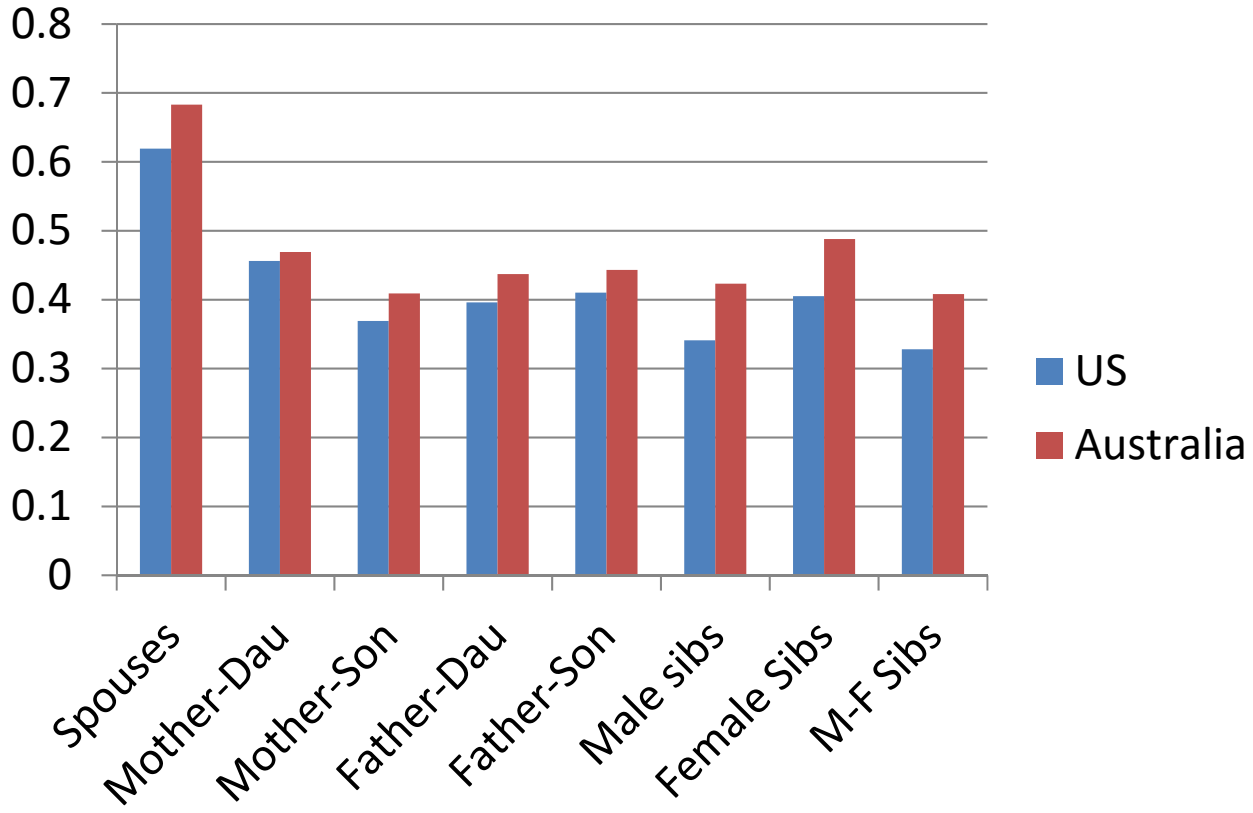
Relationship	# of pairs
Parent-offspring	**25018**
Siblings	**18697**
Spouses	**8287**
DZ Twins	**5120**
MZ Twins	**4623**


Nuclear Family Correlations for Stature
(Virginia 30,000 and OZ 22,000)
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Spouses
Mother-Dau
Mother-Son
Father-Dau
Father-Son
Male Sibs
Female Sibs
M-F Sibs
US
Australia

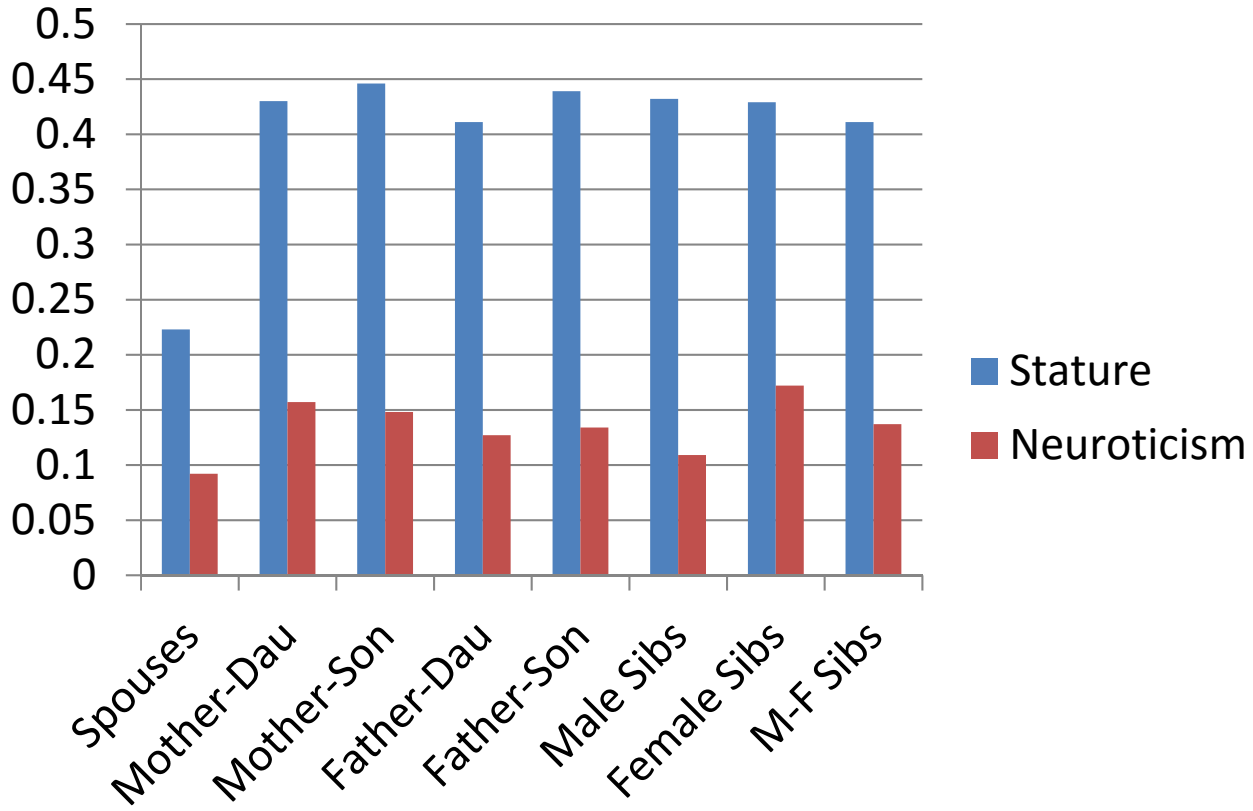
Nuclear Family Correlations for Stature and Liberalism/Conservatism (Virginia 30,000)



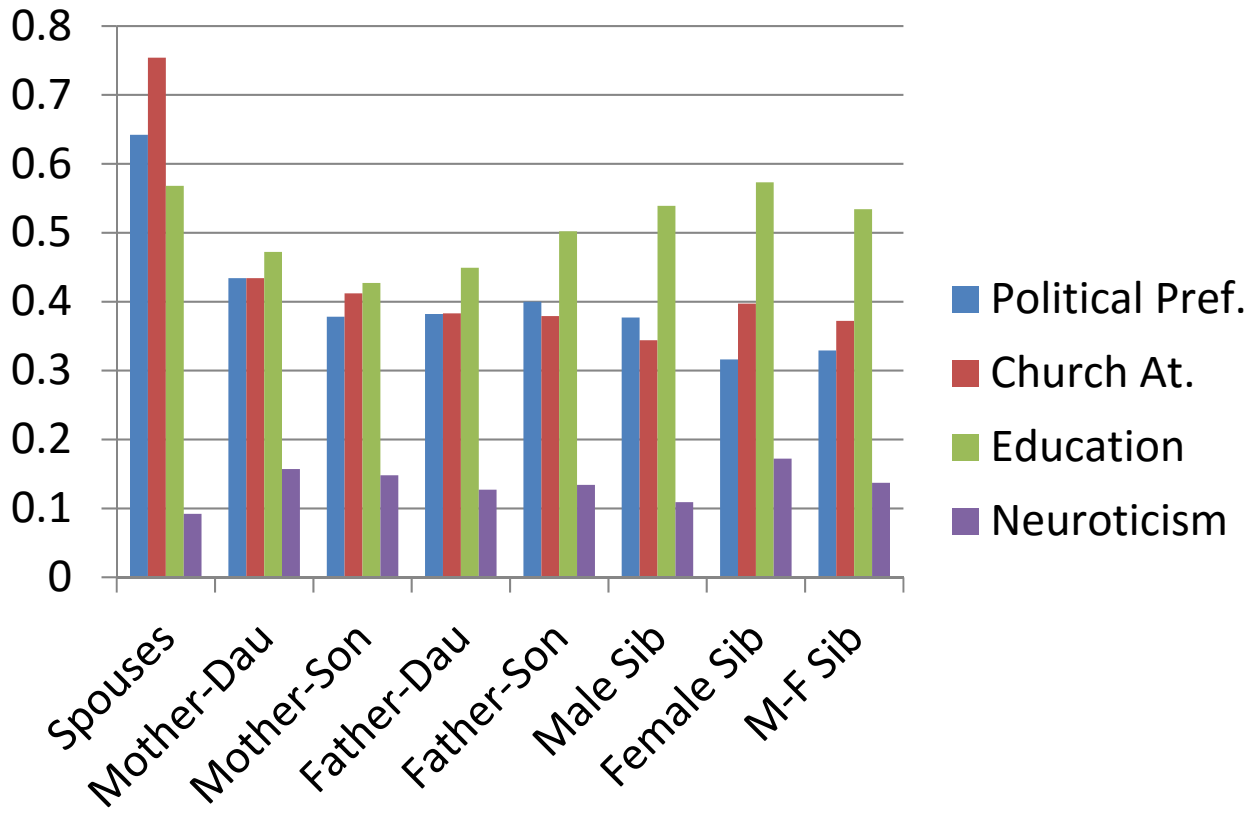
Nuclear Family Correlations for Liberalism/Conservatism
(Virginia 30,000 and Australia 22,000)



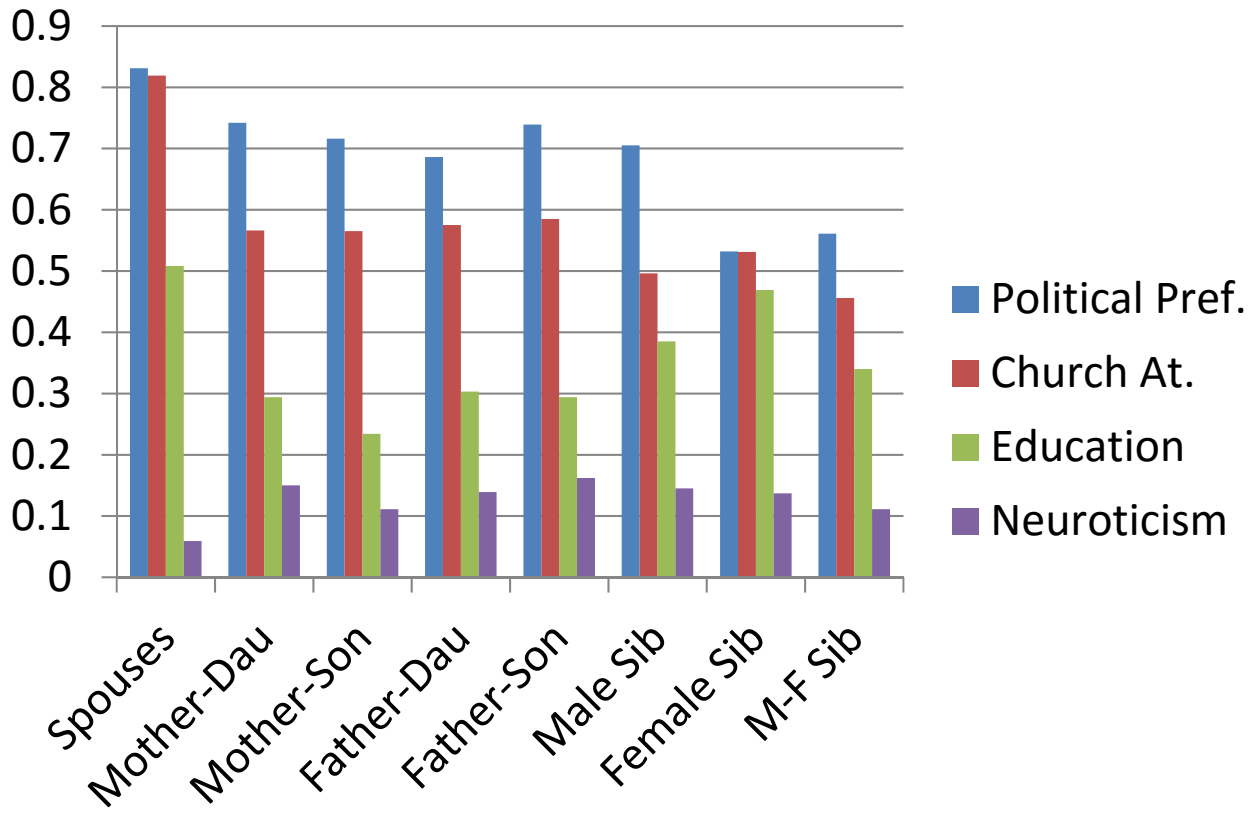
Nuclear Family Correlations for Stature and EPQ Neuroticism (Virginia 30,000)



Nuclear Family Correlations for Socially Significant Variables (Virginia 30,000)



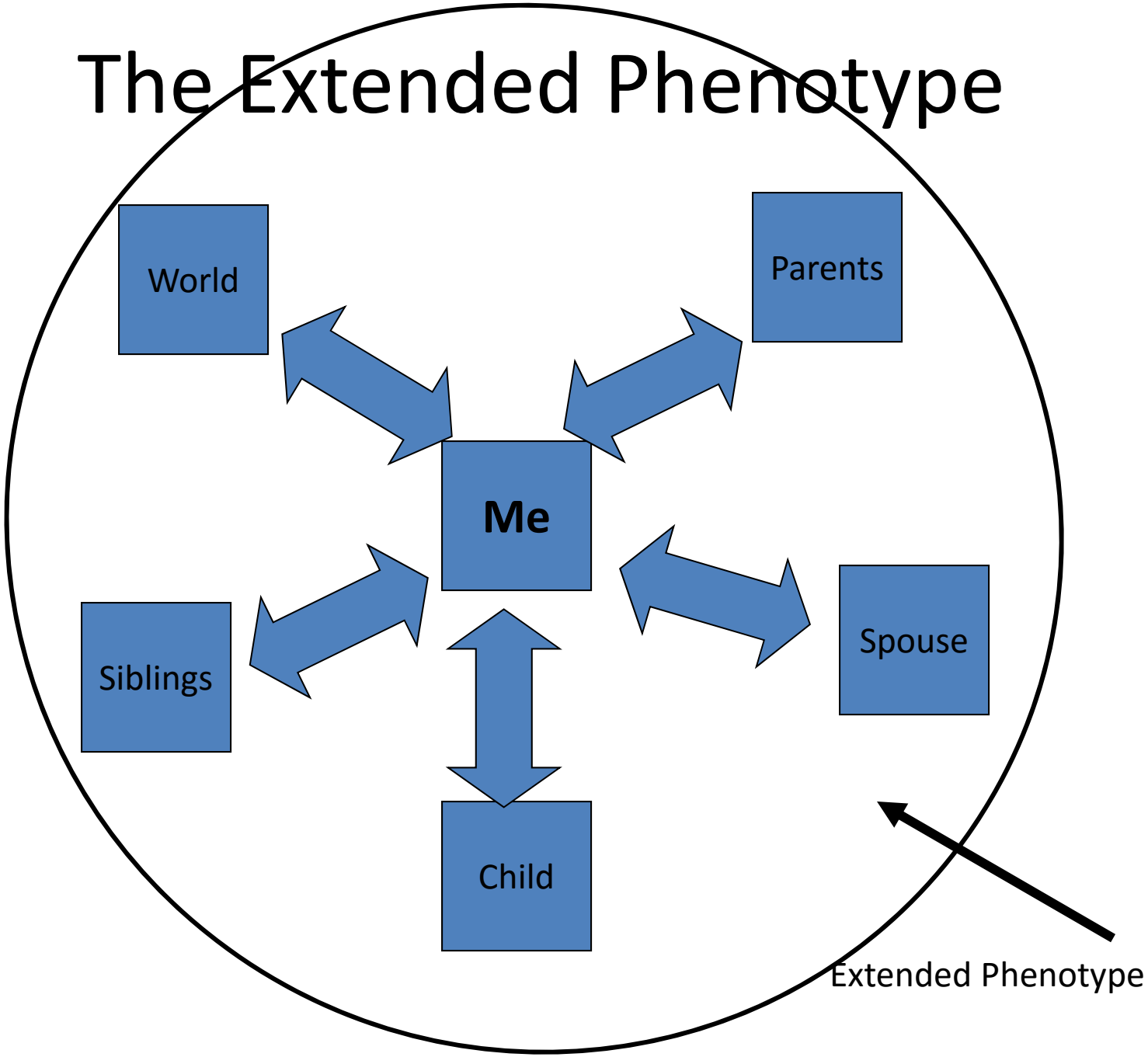
Nuclear Family Correlations for Socially Significant Variables (Australia 22K)



The (Really!) BIG Problem

Families are a mixture of genetic and social factors

The Extended Phenotype



A Basic Model

Phenotype=Genotype+Environment

P=G+E {+f(G,E)}

f(G,E) = G-E “Interplay”

i.e. Genotype-environment interaction (GxE) and G-E correlation (rGE)

GxE Interaction and Correlation

- GxE: SENSITIVITY to E controlled by G

- rGE: EXPOSURE to E correlated with ("depends on") G

Lots of good plant and animal models for both

Sources of rGE

- Environment is “caused by” (“selected by”) genetic characteristics of subject (“active/evocative” e.g. “niche selection”)
- Environment is “affected by” genetic characteristics of relatives (mothers, fathers, siblings, “passive”)

- Both are (may be) dynamic, temporal, developmental

Galton's Solution:

Twins

(Though Augustine may have got there first – 5th cent.)

One (?ideal) solution

Twins separated at birth

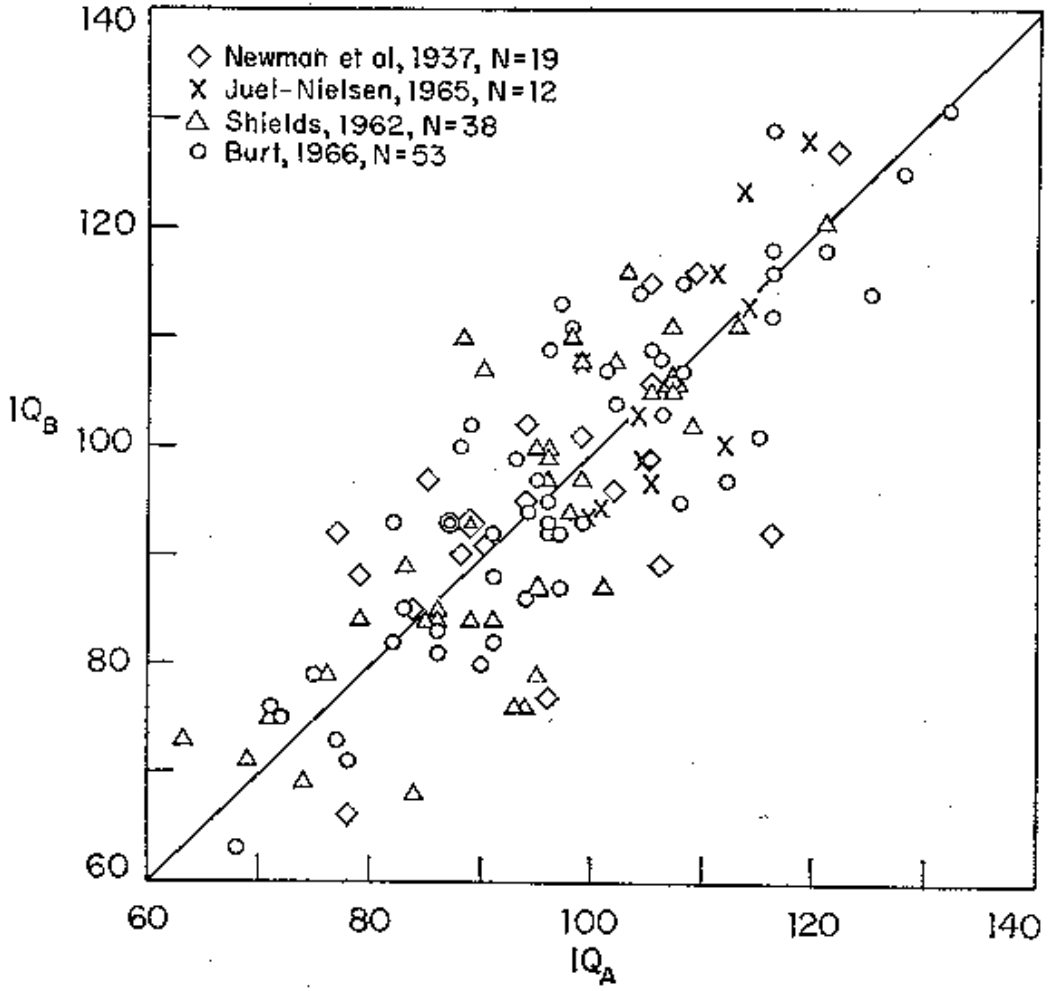


FIGURE 2. *Scatter diagram showing correlation between IQs of 122 sets of co-twins (A and B assigned at random). The obtained intra-class correlation (r_i) is 0·82. The diagonal line represents perfect correlation ($r_i = 1·00$).*

But separated MZs are rare

An easier alternative:

Identical and non-identical twins reared together:

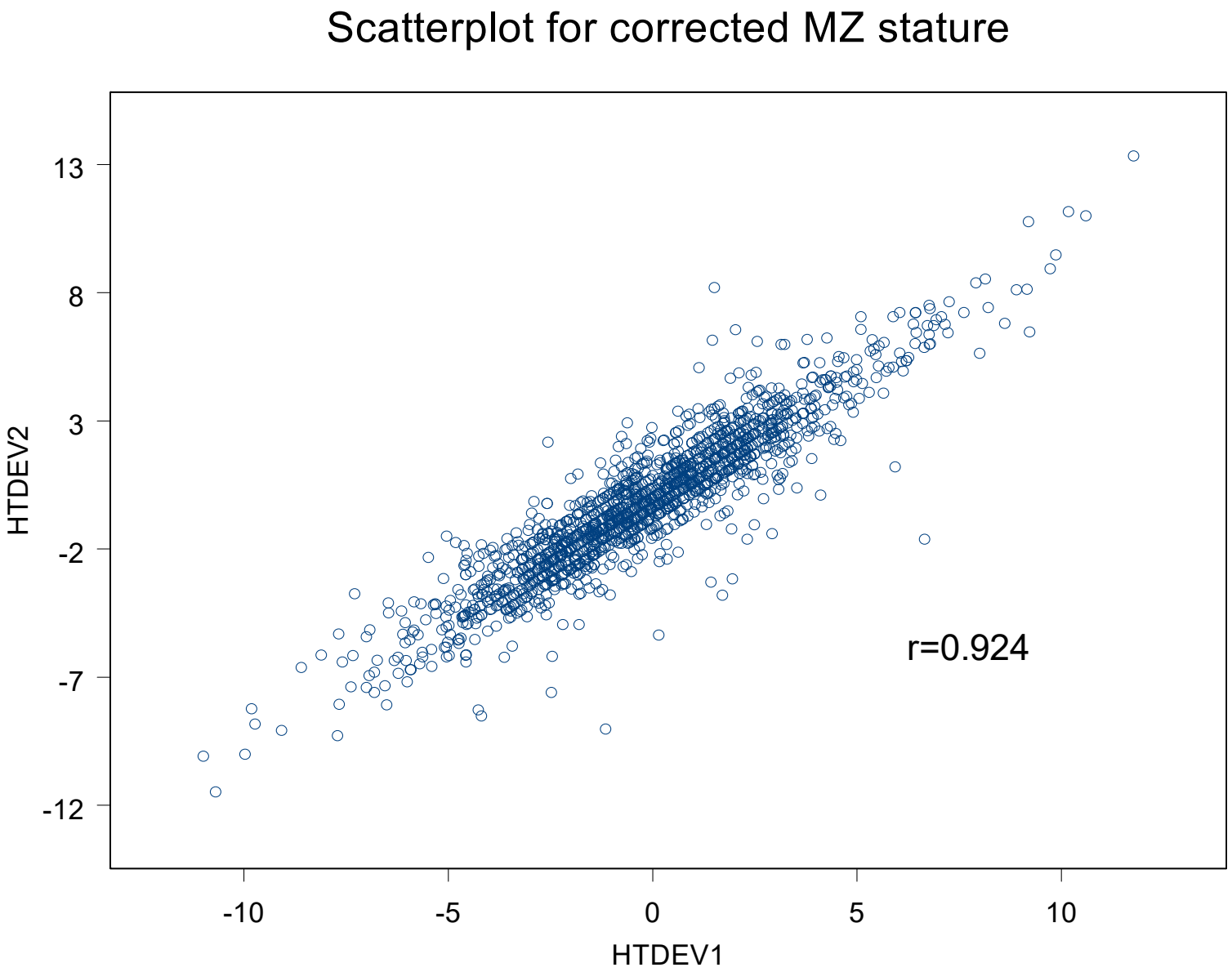
Galton (Again!)

IDENTICAL TWINS

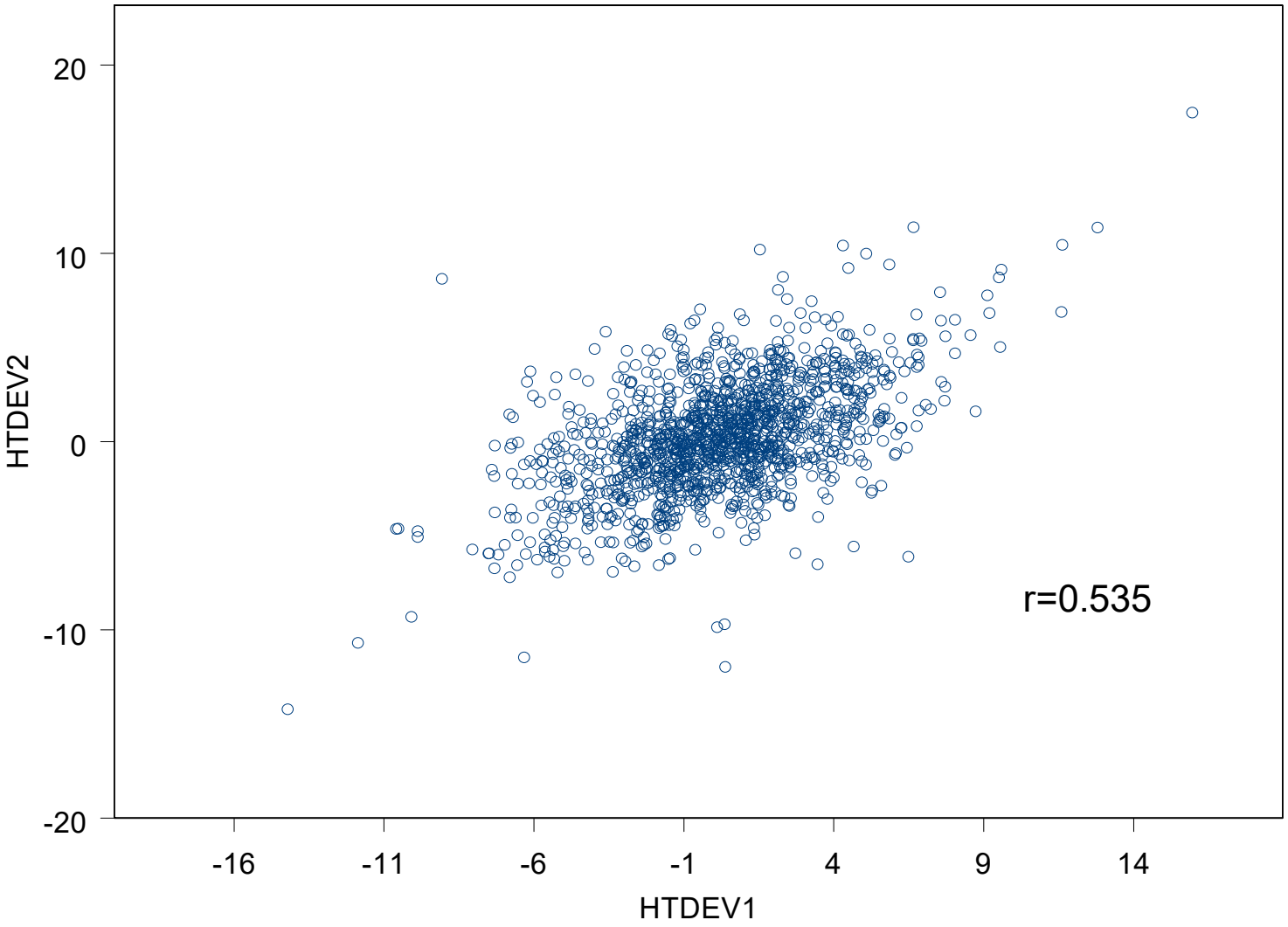
- MONOZYGOTIC: Have IDENTICAL genes (G)
- Come from the same family (C)
- Have unique experiences during life (E)

FRATERNAL TWINS

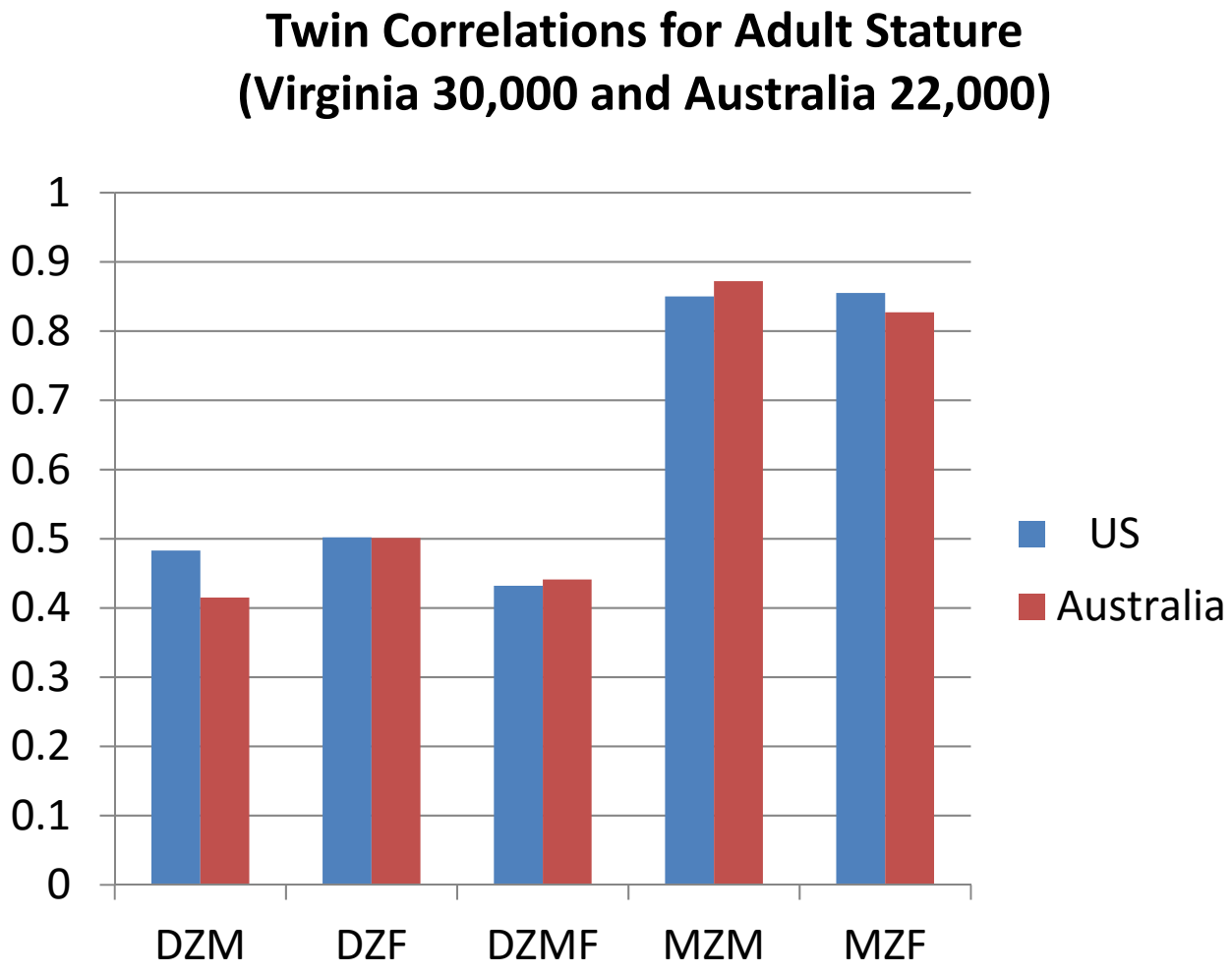
- DIZYGOTIC: Have DIFFERENT genes (G)
- Come from the same family (C)
- Have unique experiences during life (E)


Scatterplot for corrected MZ stature
HTDEV2
13
8
3
-2
-7
-12
-10
-5
0
5
10
HTDEV1
r=0.924
Data from the Virginia Twin Study of Adolescent Behavioral Development

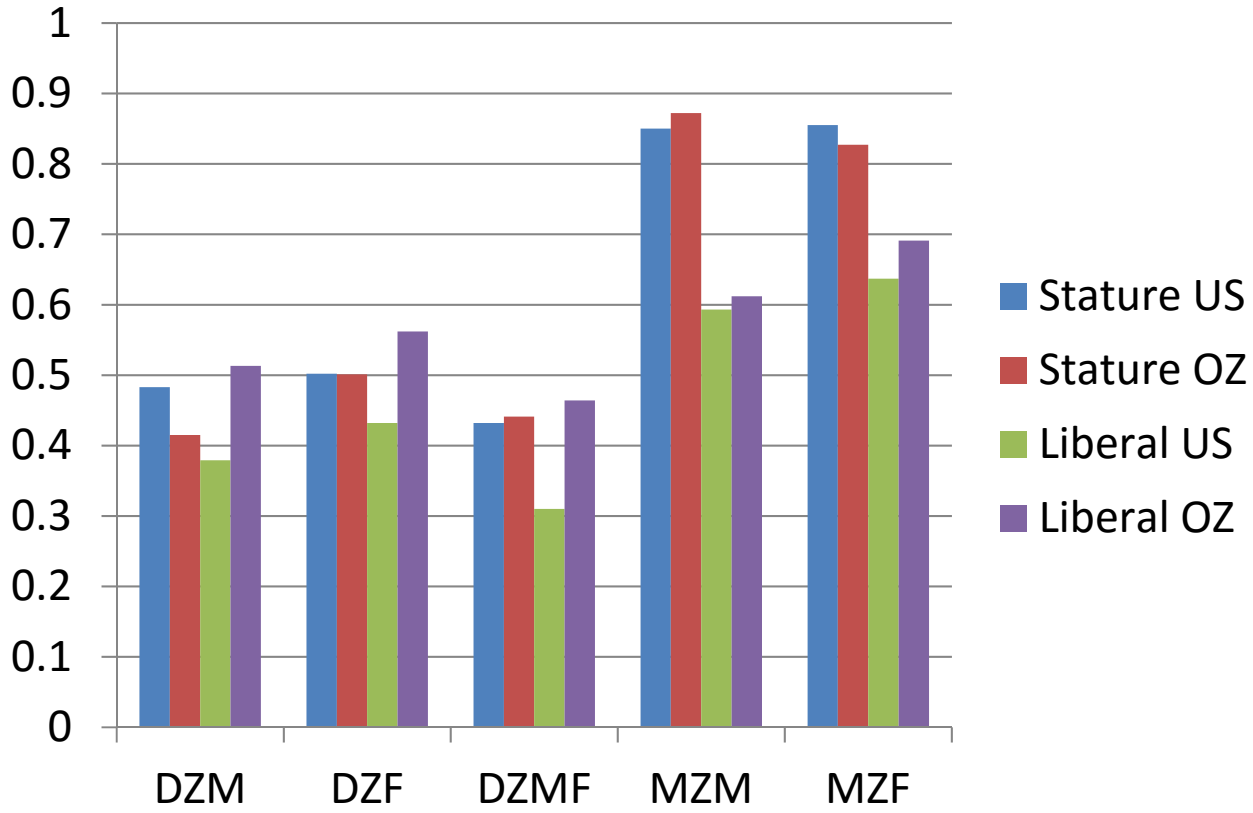
Scatterplot for age and sex corrected stature in DZ twins



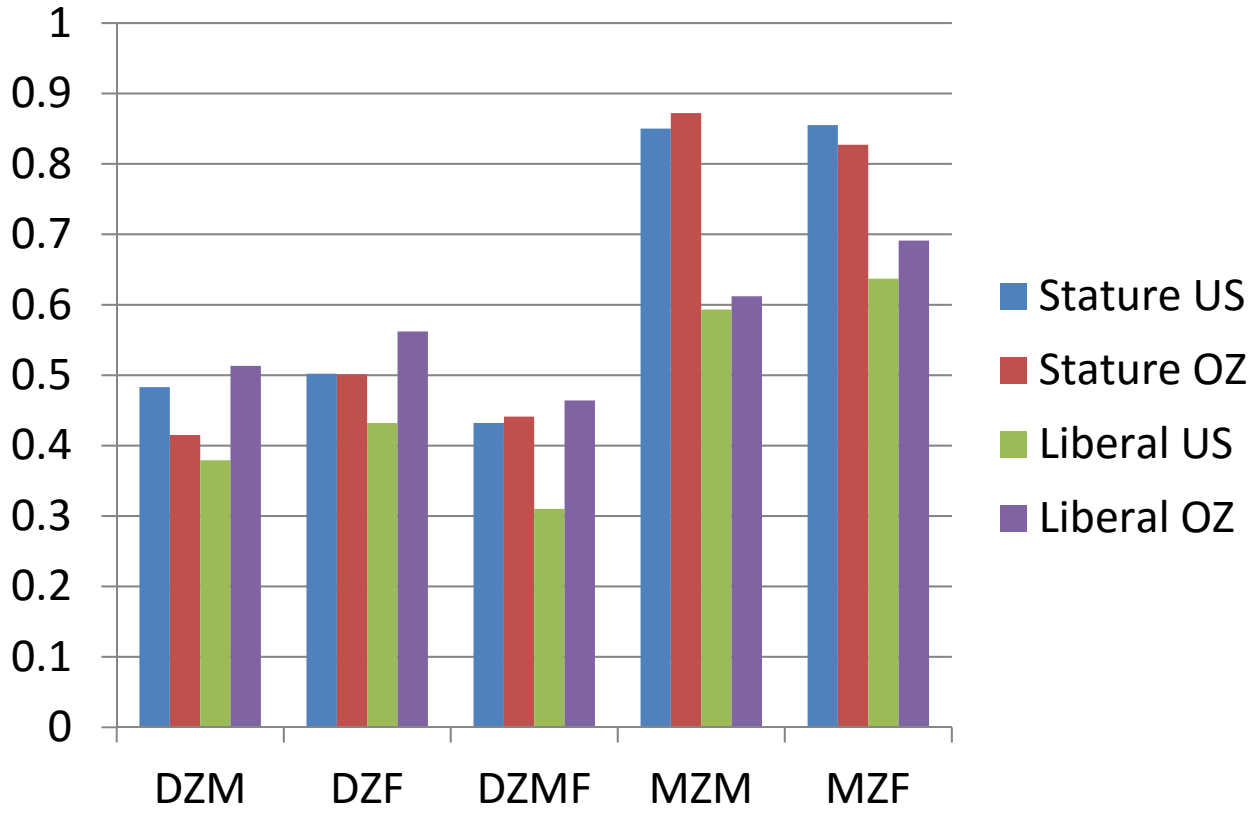
Data from the Virginia Twin Study of Adolescent Behavioral Development


Twin Correlations for Adult Stature
(Virginia 30,000 and Australia 22,000)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
DZM
DZF
DZMF
MZM
MZF
US
Australia

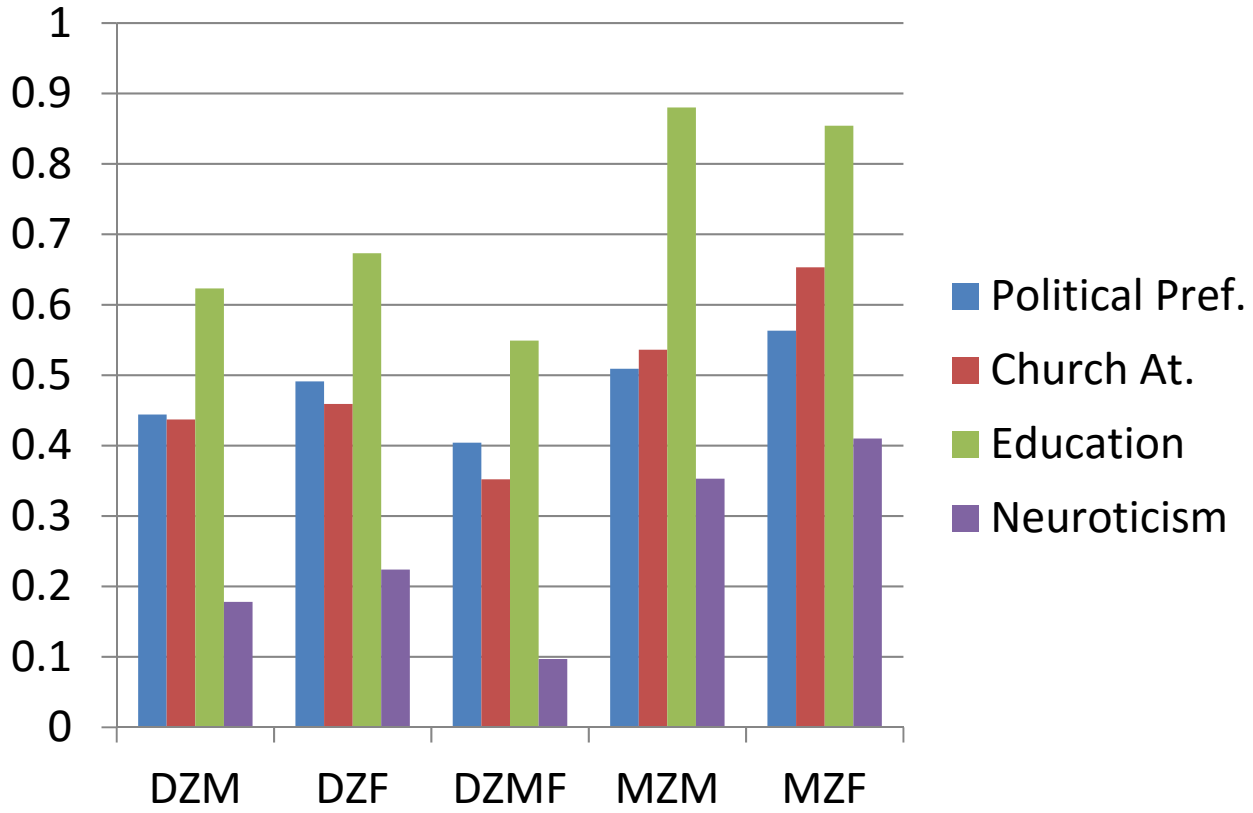
Twin Correlations for Stature and Liberalism (Virginia 30,000 and Australia 22,000)



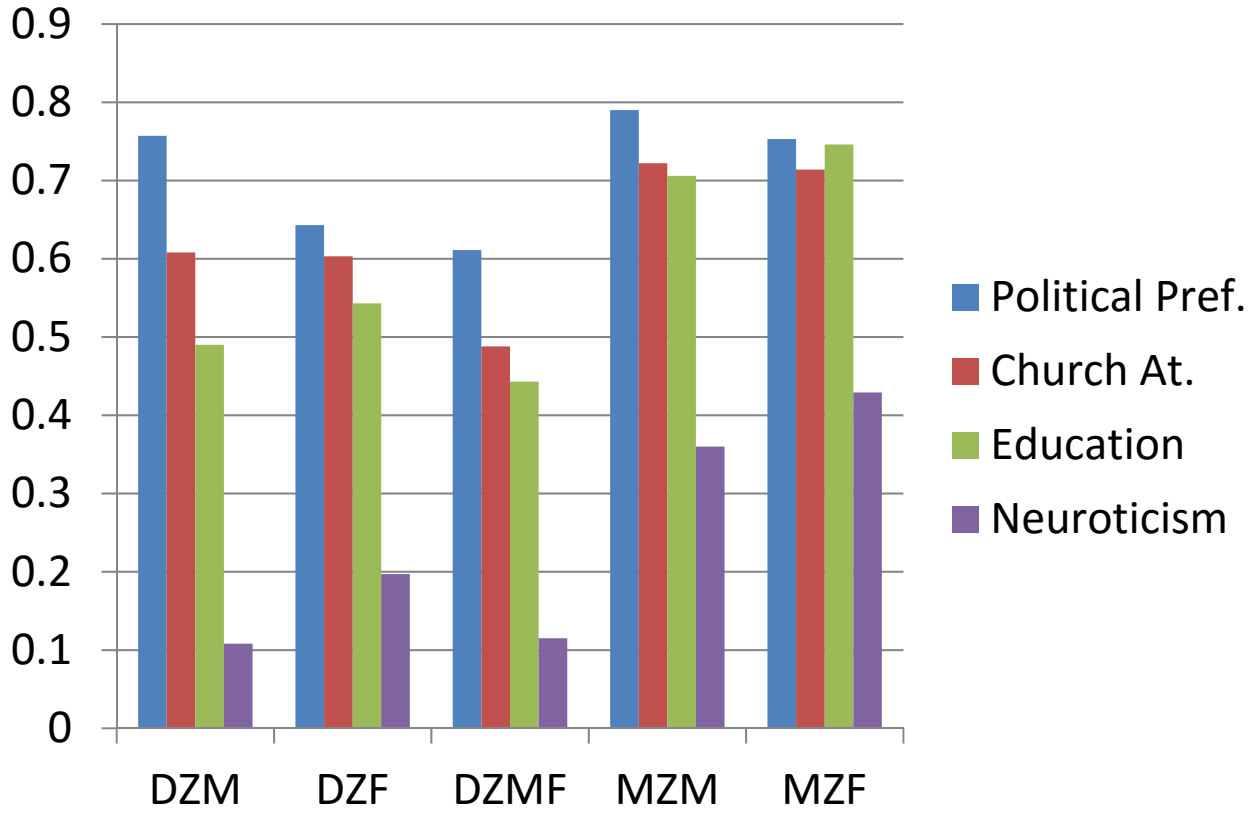
Twin Correlations for Stature and Liberalism (Virginia 30,000 and Australia 22,000)



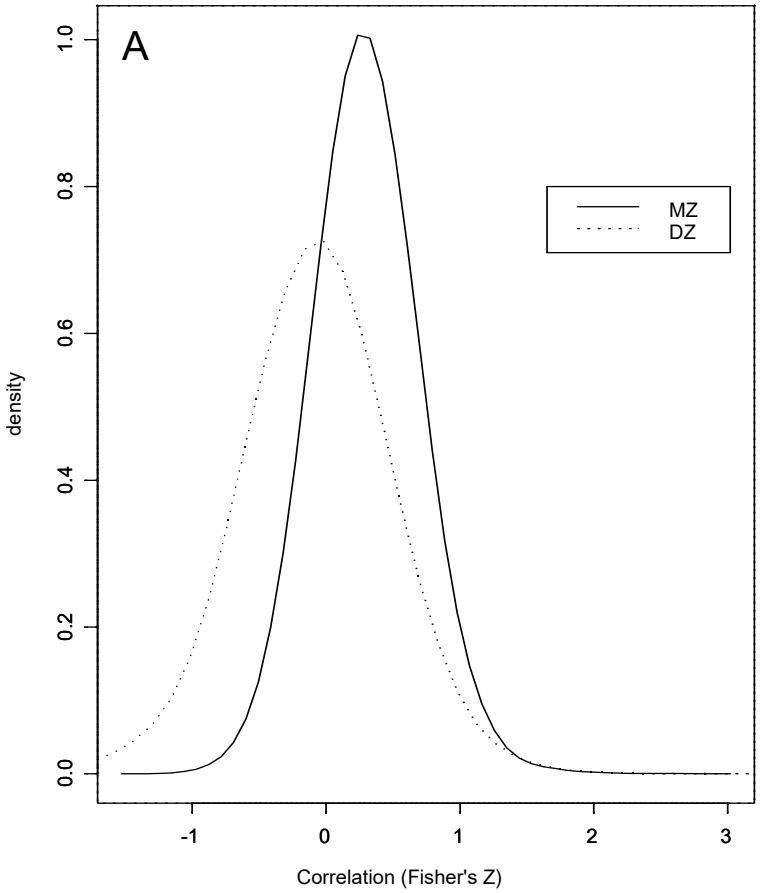
Twin Correlations for Socially Significant Variables (Virginia 30,000)

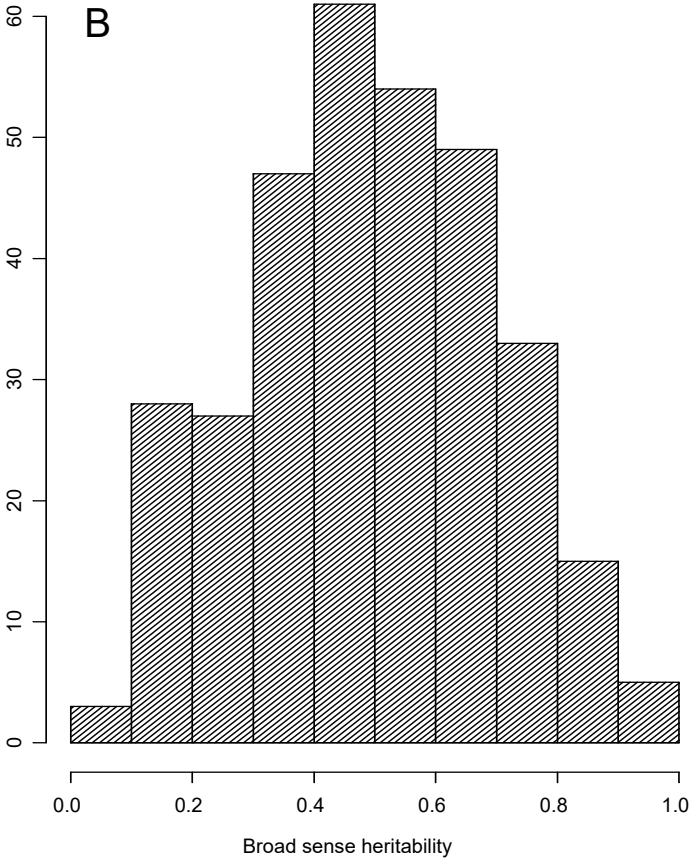


Twin Correlations for Socially Significant Variables (Australia 22,000)



Twin correlations for gene expression

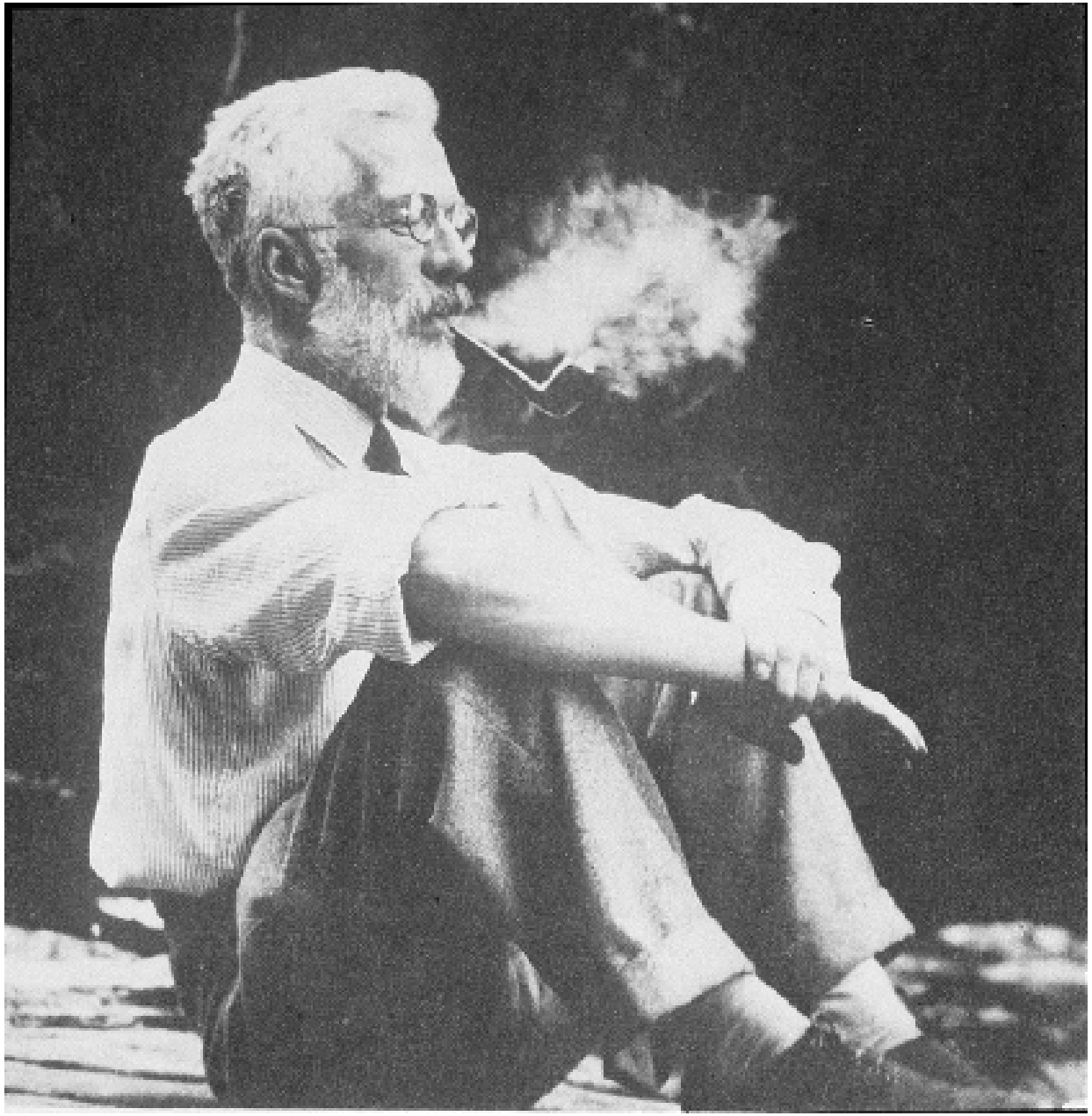




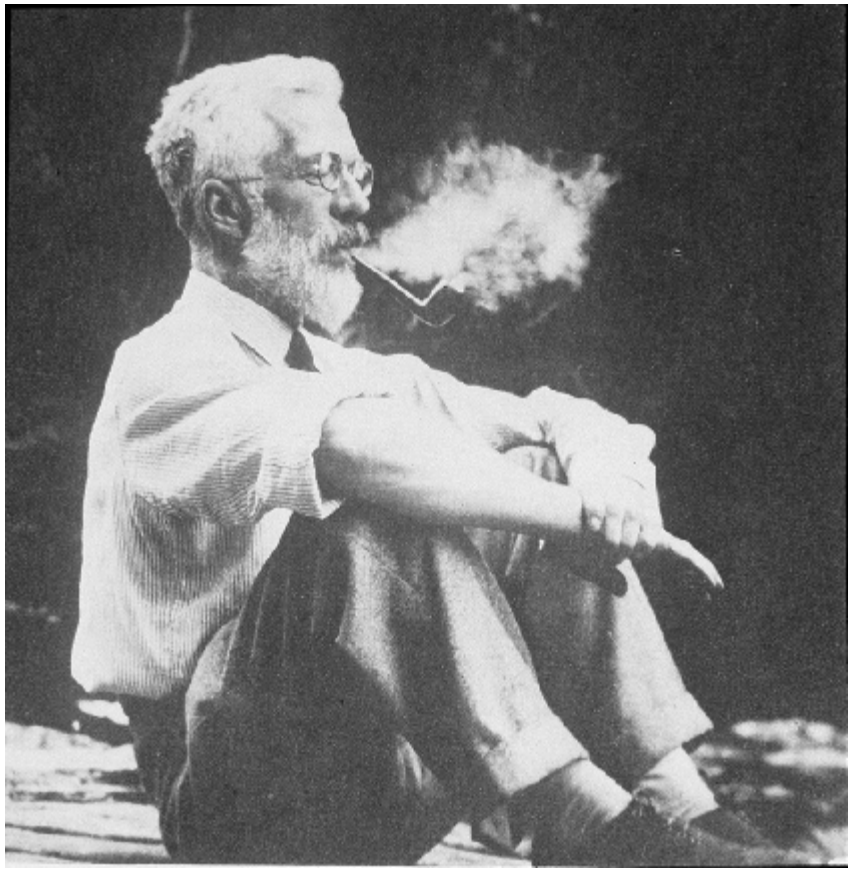
York et al.

Twin correlations for attitudes to gun control

Type	N	r	s.e.
MZM	147	0.594	0.085
MZF	630	0.383	0.059
DZM	65	0.119	0.187
DZF	315	0.366	0.083
DZMF	215	0.137	0.105



Ronald Fisher (1890-1962)



1918: On the Correlation Between Relatives on the Supposition of Mendelian Inheritance
1921: Introduced concept of “likelihood”
1930: The Genetical Theory of Natural Selection
1935: The Design of Experiments

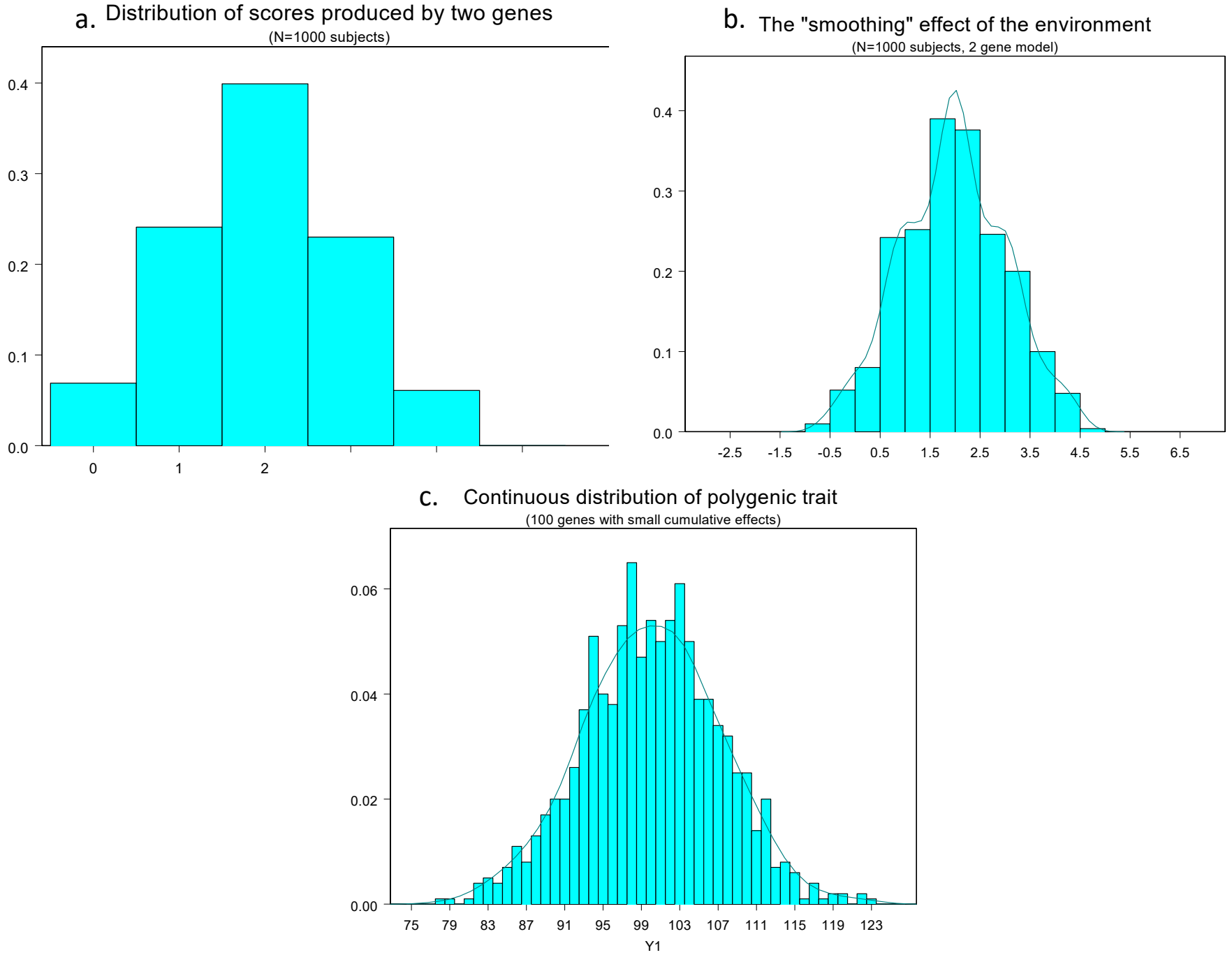
Fisher developed mathematical theory that reconciled Mendel's work with Galton and Pearson's correlations

XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance.

By **R. A. Fisher**, B.A. *Communicated by* Professor J. ARTHUR THOMSON. (With Four Figures in Text.)

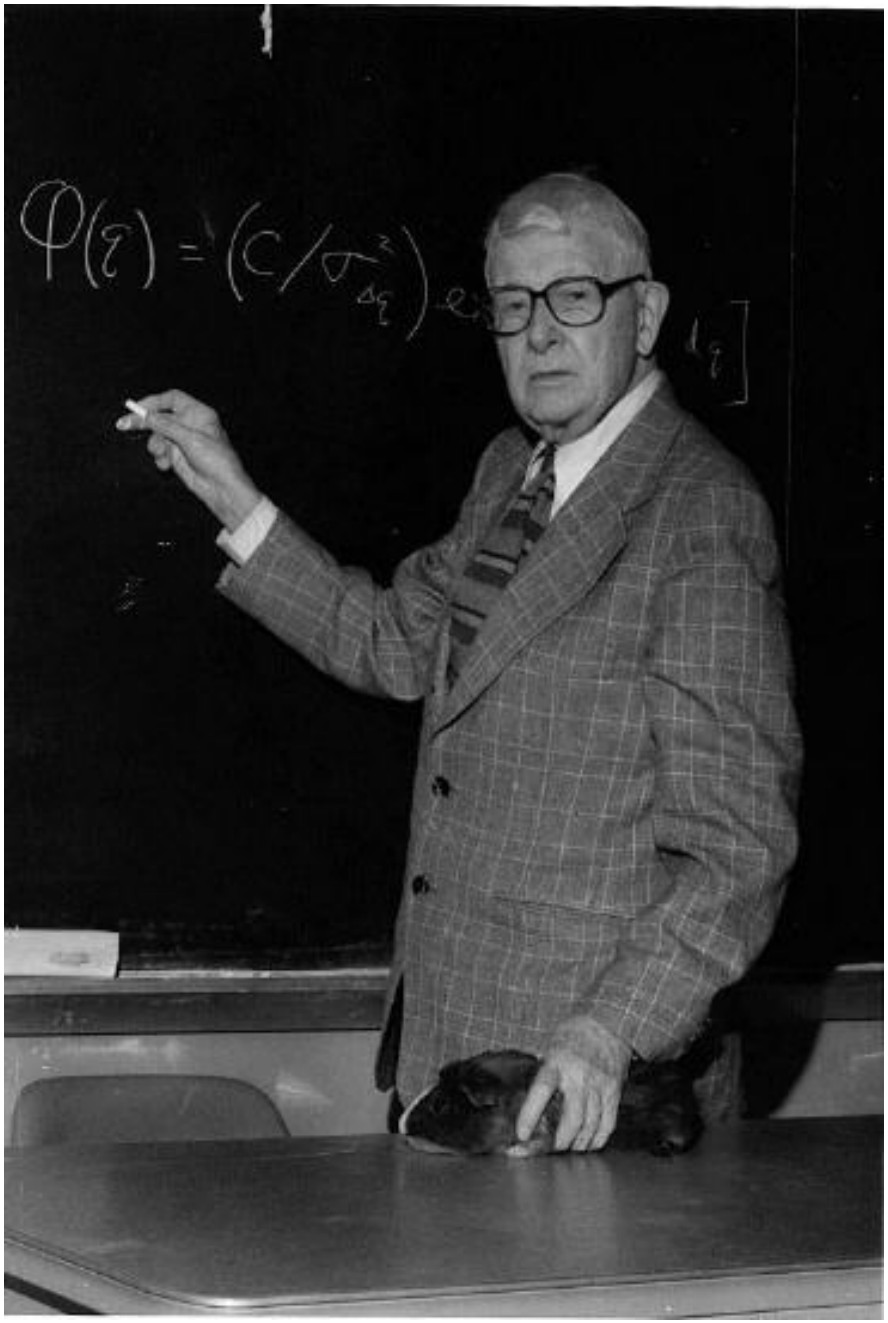
(MS. received June 15, 1918. Read July 8, 1918. Issued separately October 1, 1918.)

CONTENTS.

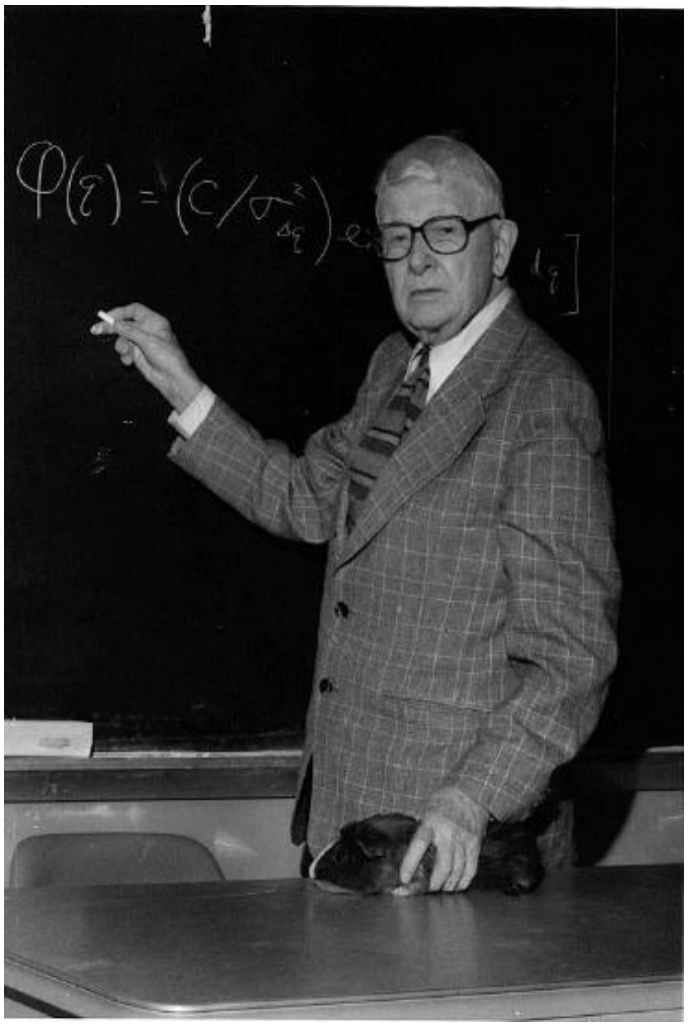

a. Distribution of scores produced by two genes
(N=1000 subjects)
0.4
0.3
0.2
0.1
0.0
0
1
2
b. The "smoothing" effect of the environment
(N=1000 subjects, 2 gene model)
0.4
0.3
0.2
0.1
0.0
-2.5
-1.5
-0.5
0.5
1.5
2.5
3.5
4.5
5.5
6.5
c. Continuous distribution of polygenic trait
(100 genes with small cumulative effects)
0.06
0.04
0.02
0.00
75
79
83
87
91
95
99
103
107
111
115
119
123
Y1

Fisher (1918): Basic Ideas

- Continuous variation caused by lots of genes ("polygenic inheritance")
- Each gene followed Mendel's laws
- Environment smoothed out genetic differences
- Genes may show different degrees of "dominance"
- Genes may have many forms ("mutliple alleles")
- Mating may not be random ("assortative mating")
- Showed that correlations obtained by e.g. Pearson and Lee were explained well by polygenic inheritance
- Let to "Biometrical Genetics" (Mather, Jinks etc.)


$\varphi(\xi) = (C/\sigma_{\Delta\xi}^{3})$

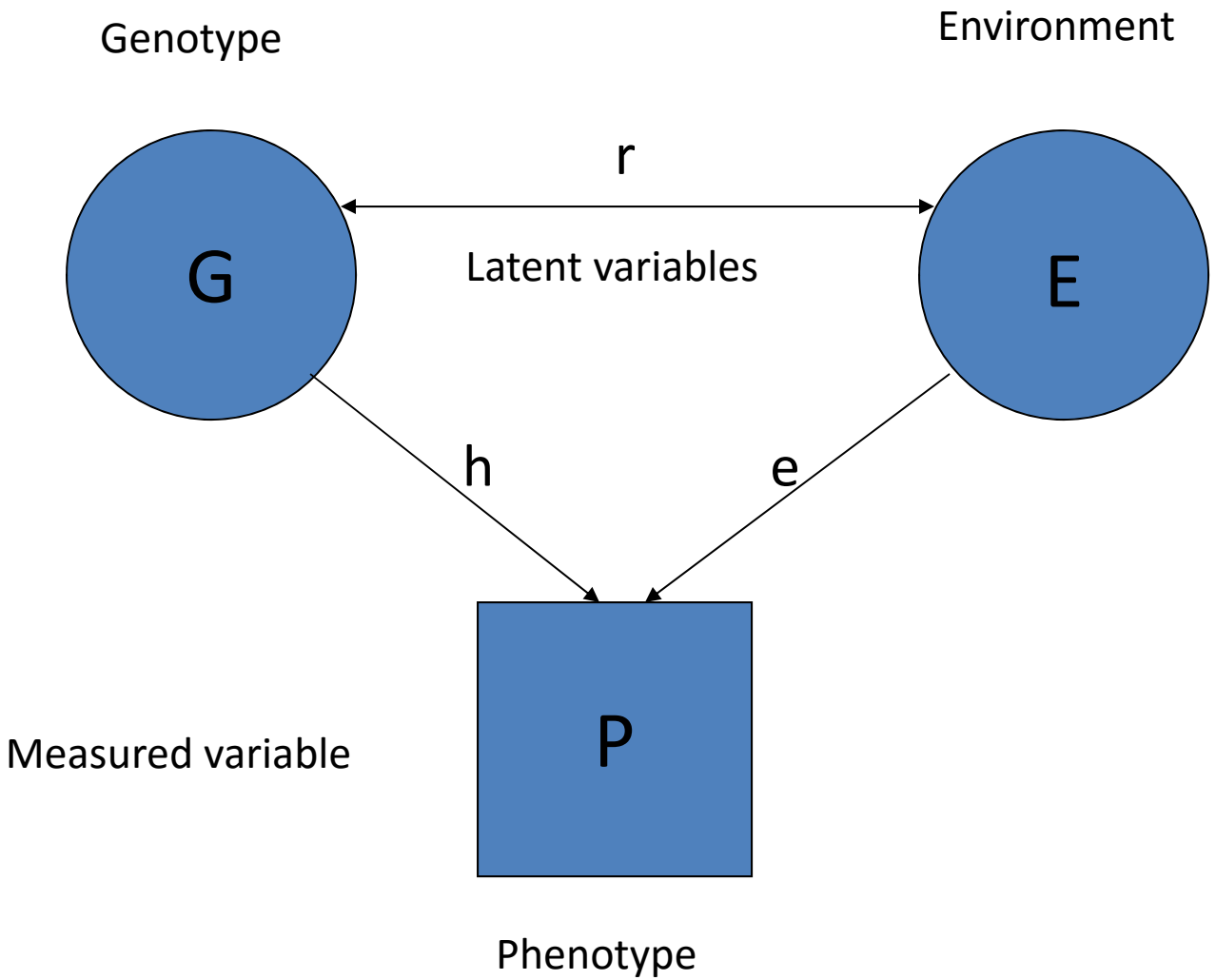
Sewall Wright (1889-1988



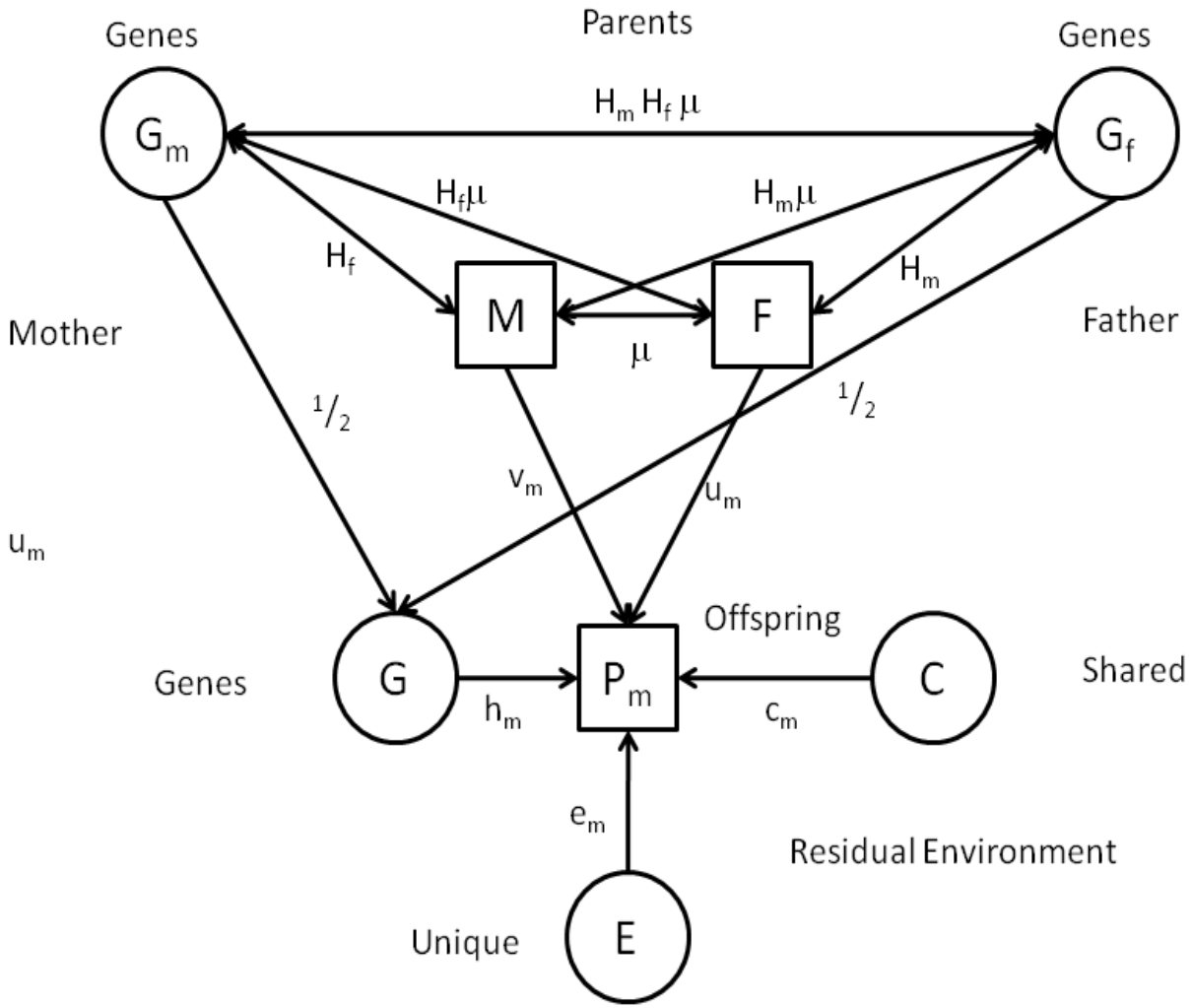
1984. Evolution and the Genetics of Populations:
Genetics and Biometric Foundations (4 vols.)

1934. The method of path coefficients. *Ann. Math. Staiist.* **5:** 161-215

Path diagram for the effects of genes and environment on phenotype

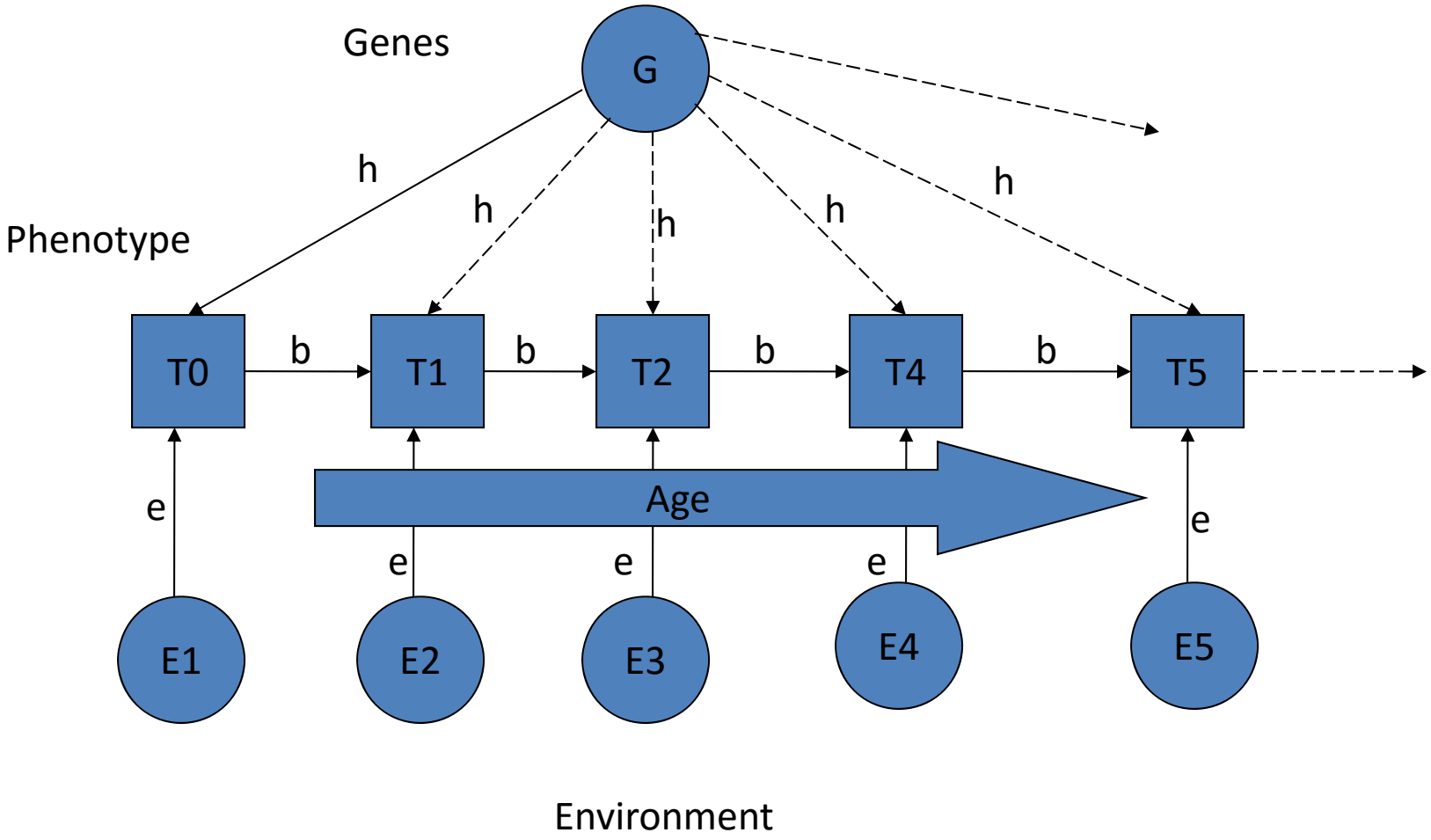


Genetic AND Cultural inheritance?

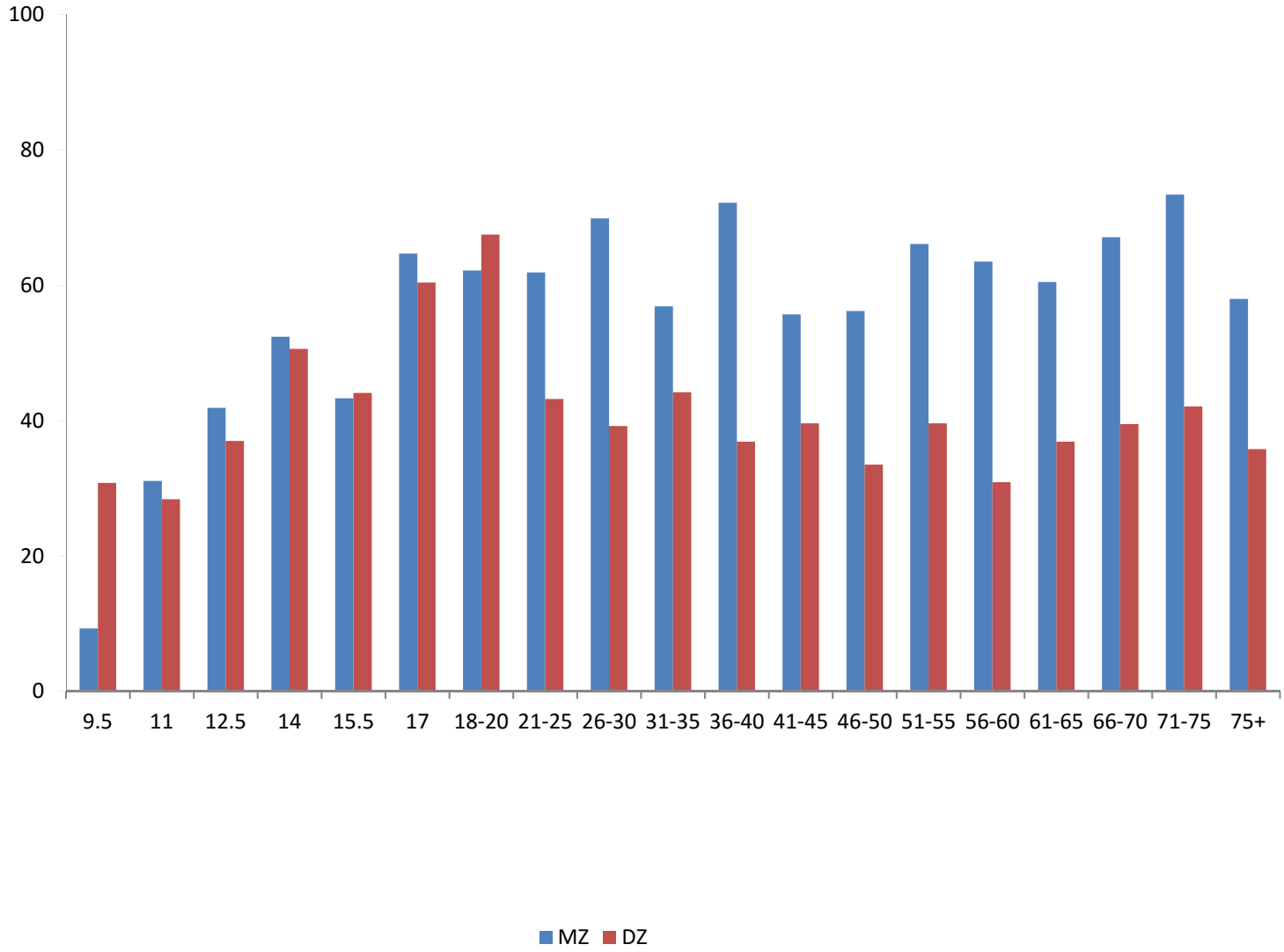


Development

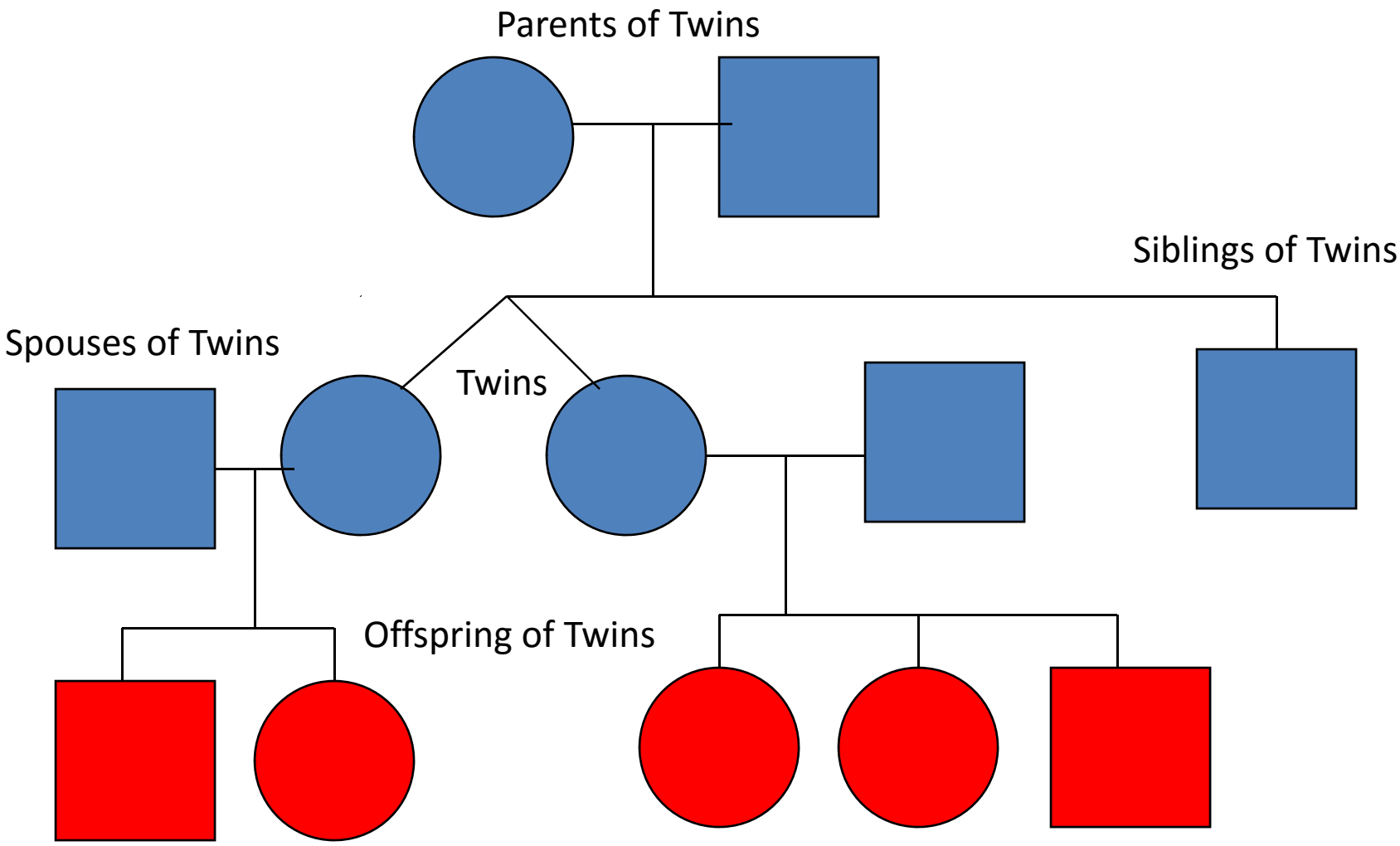
a. Genetic variation in developmental change: time series with common genes and time-specific environmental “innovations”



Attitudes over the life-span



Children of Twins ("COT")



Gestational Age

Racial Differences in Genetic and Environmental Risk to Preterm Birth

Timothy P. York, Jerome F. Strauss,
Michael C. Neale, Lindon J. Eaves

PLoS One. 2010 Aug 25;5(8):e12391.

Table 2. Sample frequencies by parental relationship and race.

Parental relationship	European American		African American	
	N. Families	**N. Births**	**N. Families**	**N. Births**
Sibship	284,446	575,709	66,983	119,791
Maternal half-sibship	6,736	12,269	2,431	4,515
Paternal half-sibship	5,419	9,800	2,839	5,292
MZ male twin	595	1,092	69	99
MZ female twin	618	1,212	98	144
DZ male twin	393	700	52	77
DZ female twin	368	696	72	119
DZ male-female twin	936	1,614	139	210
Total	299,511	603,092	72,683	130,247

Table 1. Expected covariance of gestational age expressed as variance components between pregnancy outcomes as a function of relationship between offspring.

Parental relationship	**Fetal relationship**	**Expected covariance**
MZ female twins	Half-sibling	$\frac{1}{4} f^2 + m^2$
DZ female twins	Cousin	$\frac{1}{8} f^2 + \frac{1}{2} m^2$
MZ male twins	Half-sibling	$\frac{1}{4} f^2$
DZ male twins	Cousin	$\frac{1}{8} f^2$
DZ male-female twins	Cousin	$\frac{1}{8} f^2$
Sibship	Sibling	$\frac{1}{2} f^2 + m^2 + c^2$
Maternal half-sibship	Half-sibling	$\frac{1}{4} f^2 + m^2 + hc^2$
Paternal half-sibship	Half-sibling	$\frac{1}{4} f^2 + hc^2$

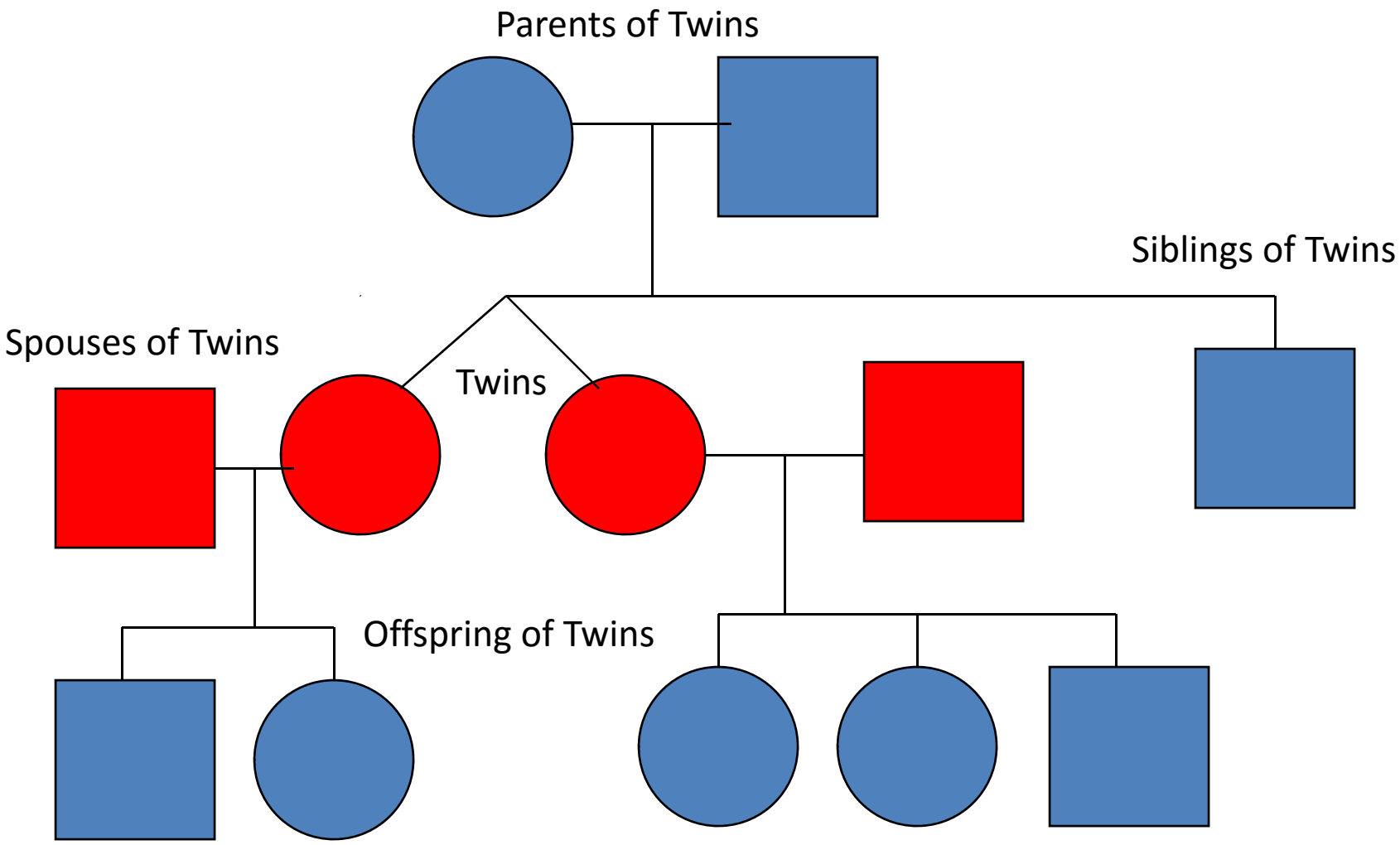
f^2 =fetal genetic, m^2 =maternal genetic, c^2 =shared familial environment
h = parameter to allow for differences in half-sibling versus full-sibling shared environment (“fudge factor”)

Table 4. Estimated variance components from model 2 with empirically derived 95% bootstrap confidence intervals adjusted for covariates (birth order, maternal age, fetal sex, source of care, smoking, maternal education).

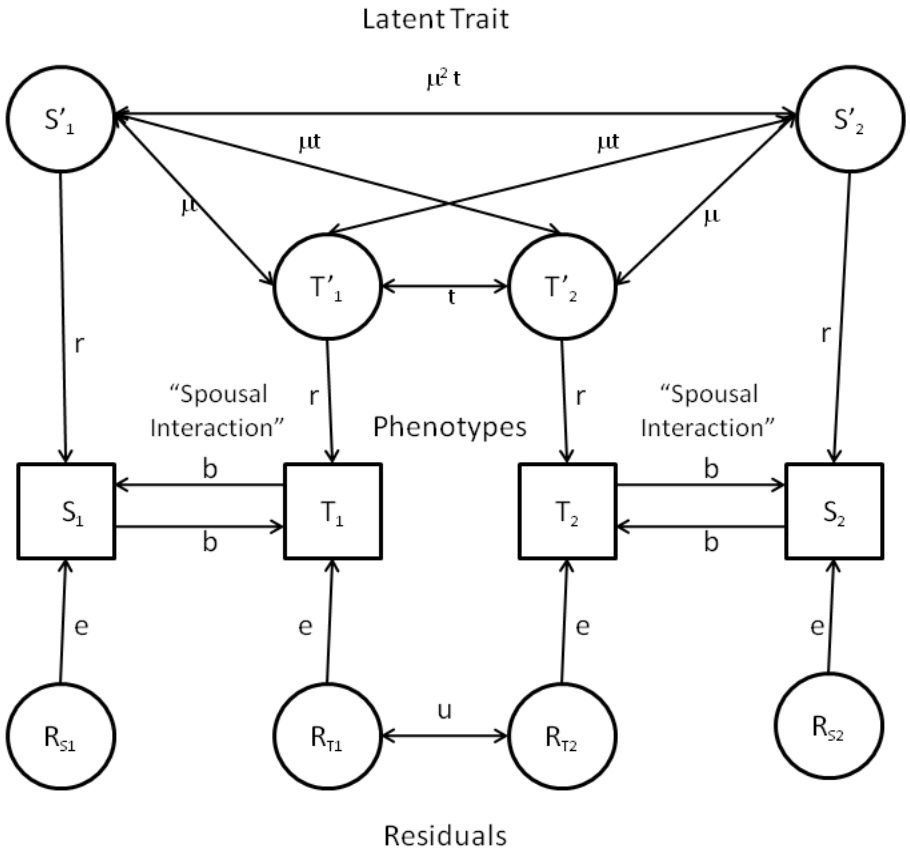
	Full Genetic Model (Model 2)			Reduced Genetic Model (Model 8)		
Source	**Estimate**	**95% CI**	**Percentage**	**Estimate**	**95% CI**	**Percentage**
African American						
Fetal genetic	0.264	(0.0, 2.302)	3.7	-	-	-
Maternal genetic	0.976	(0.274, 1.357)	13.8	1.040	(0.531, 1.445)	14.7
Shared environment	1.215	(0.499, 1.666)	17.1	1.281	(0.872, 1.781)	18.0
Unique environment	4.642	(3.559, 4.899)	65.4	4.777	(4.625, 4.927)	67.3
European American						
Fetal genetic	1.325	(0.640, 1.927)	35.2	1.325	(0.695, 1.964)	35.2
Maternal genetic	0.503	(0.263, 0.767)	13.4	0.503	(0.235, 0.758)	13.4
Shared environment	0.263	(0.006, 0.537)	7.0	0.264	(0.027, 0.537)	7.0
Unique environment	1.673	(1.355, 2.024)	44.4	1.674	(1.355, 1.990)	44.5

“Mating”

Spouses of Twins ("SPOT")



“Twins and Spouses”



f(G,E)

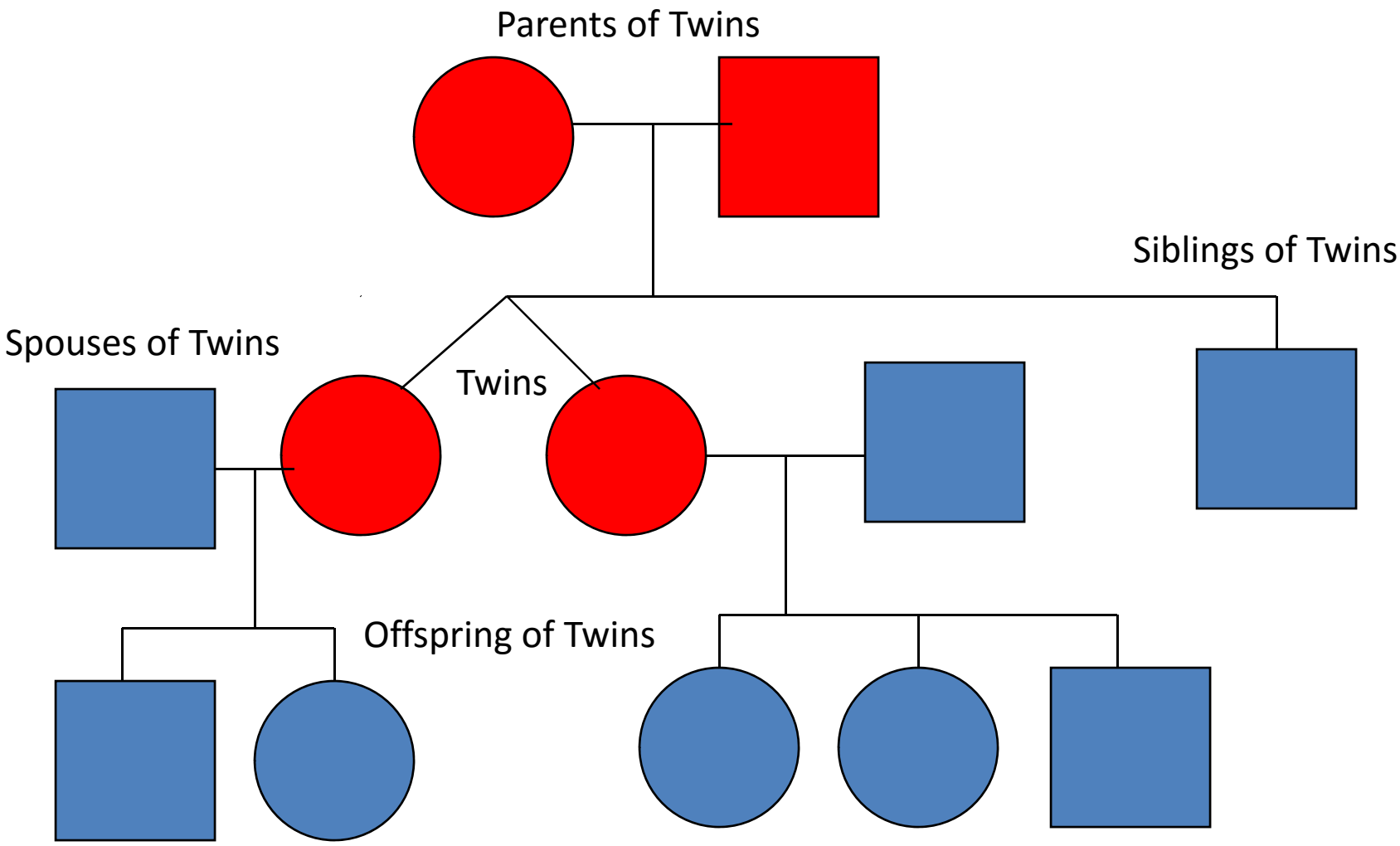
Genotype x Environment Interaction
("GxE")

Genotype-Environment Correlation
("rGE")

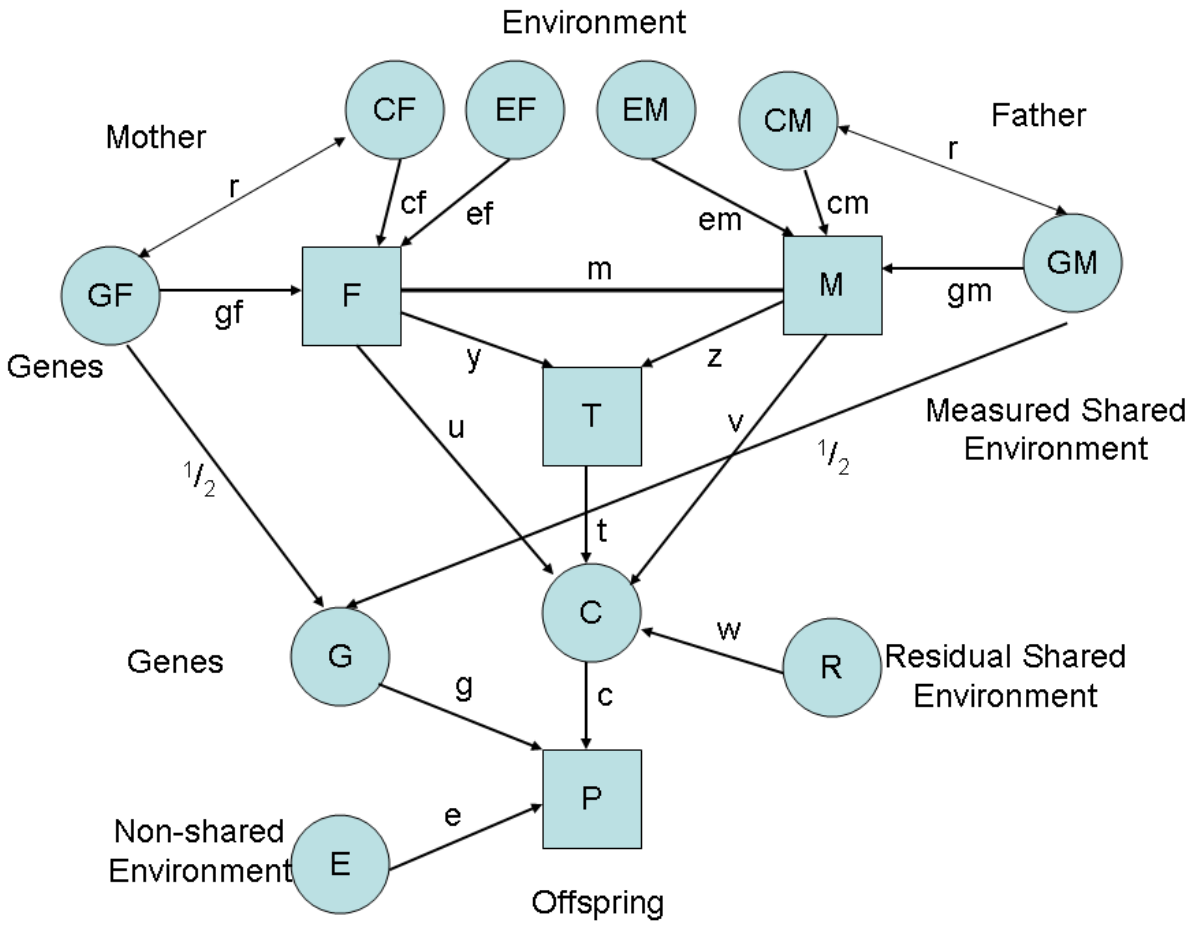
(“Passive”) rGE

Twins and Parents

Twins and Parents ("TAP")

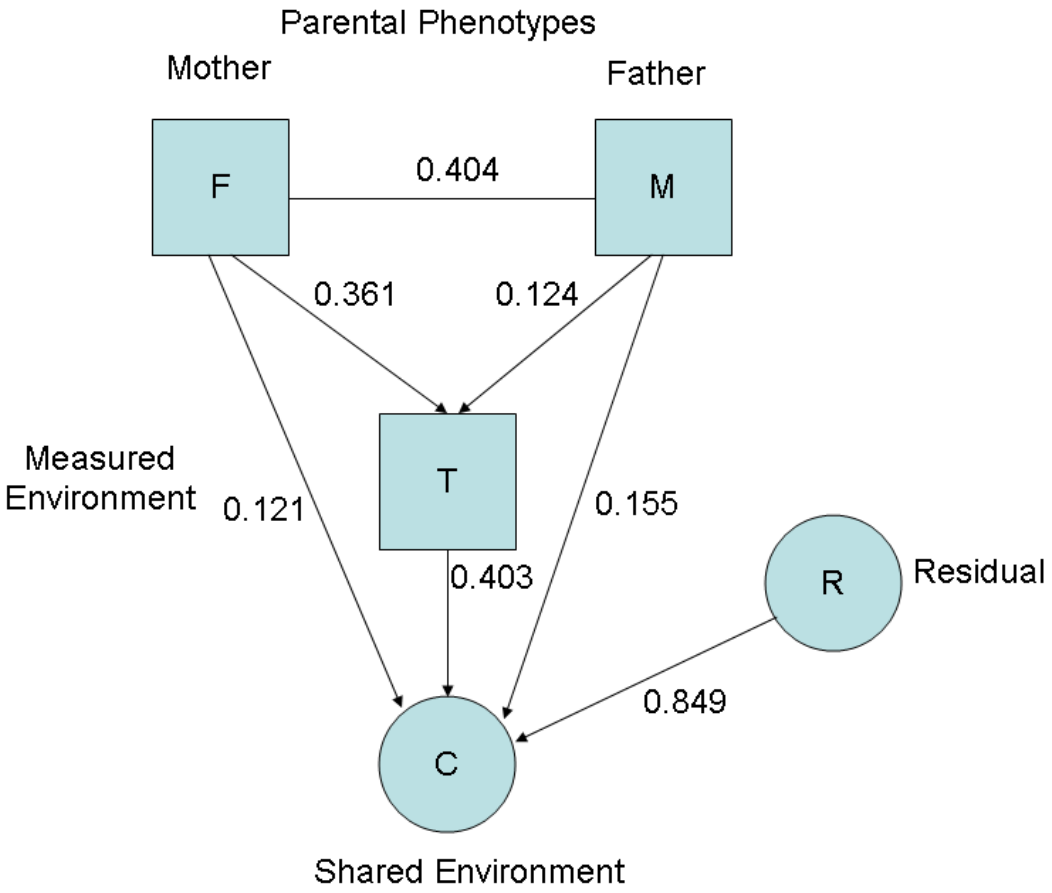


Parental Neglect and Anti-Social Behavior



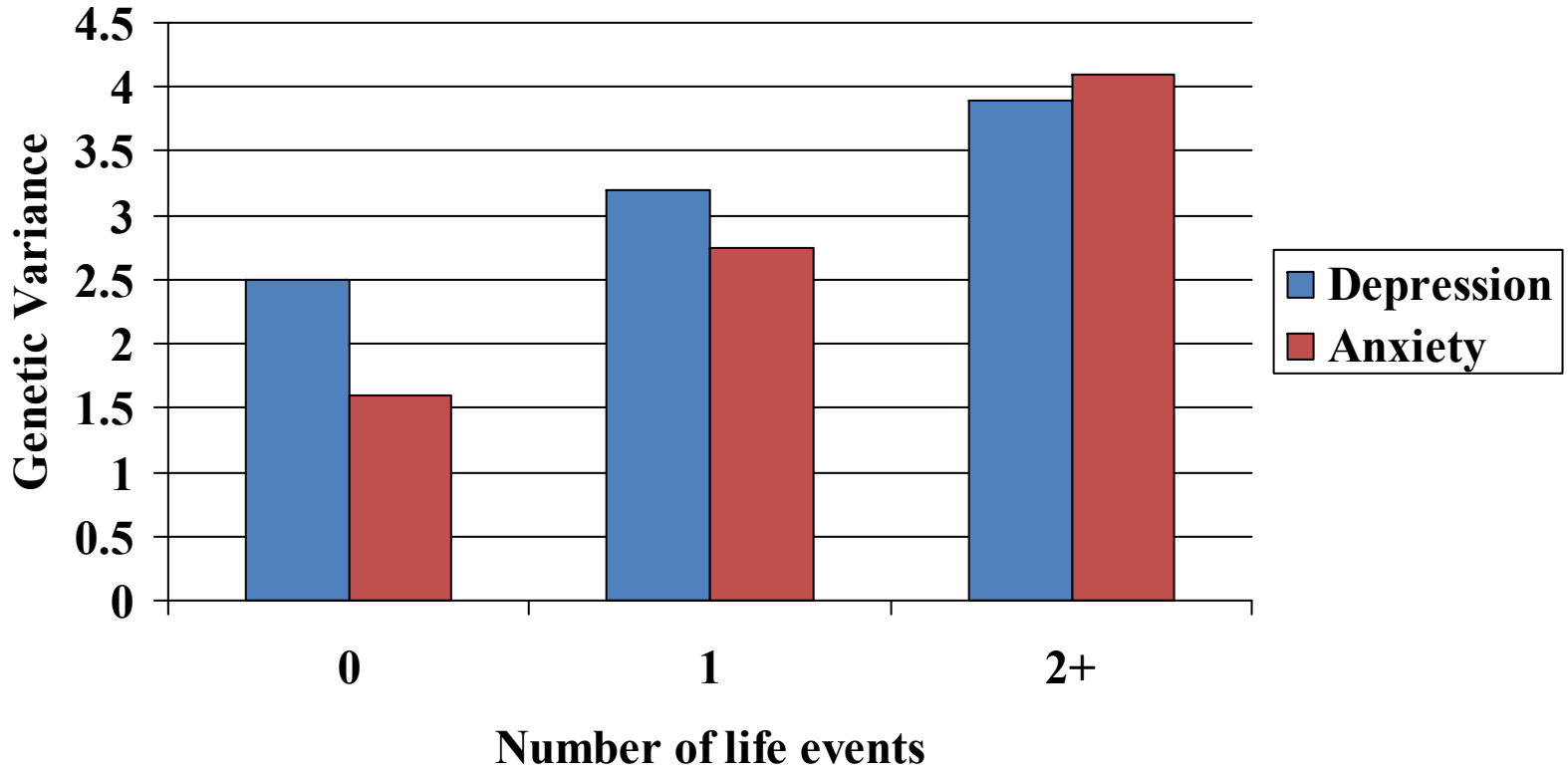
Eaves et al., 2010

Environmental pathways



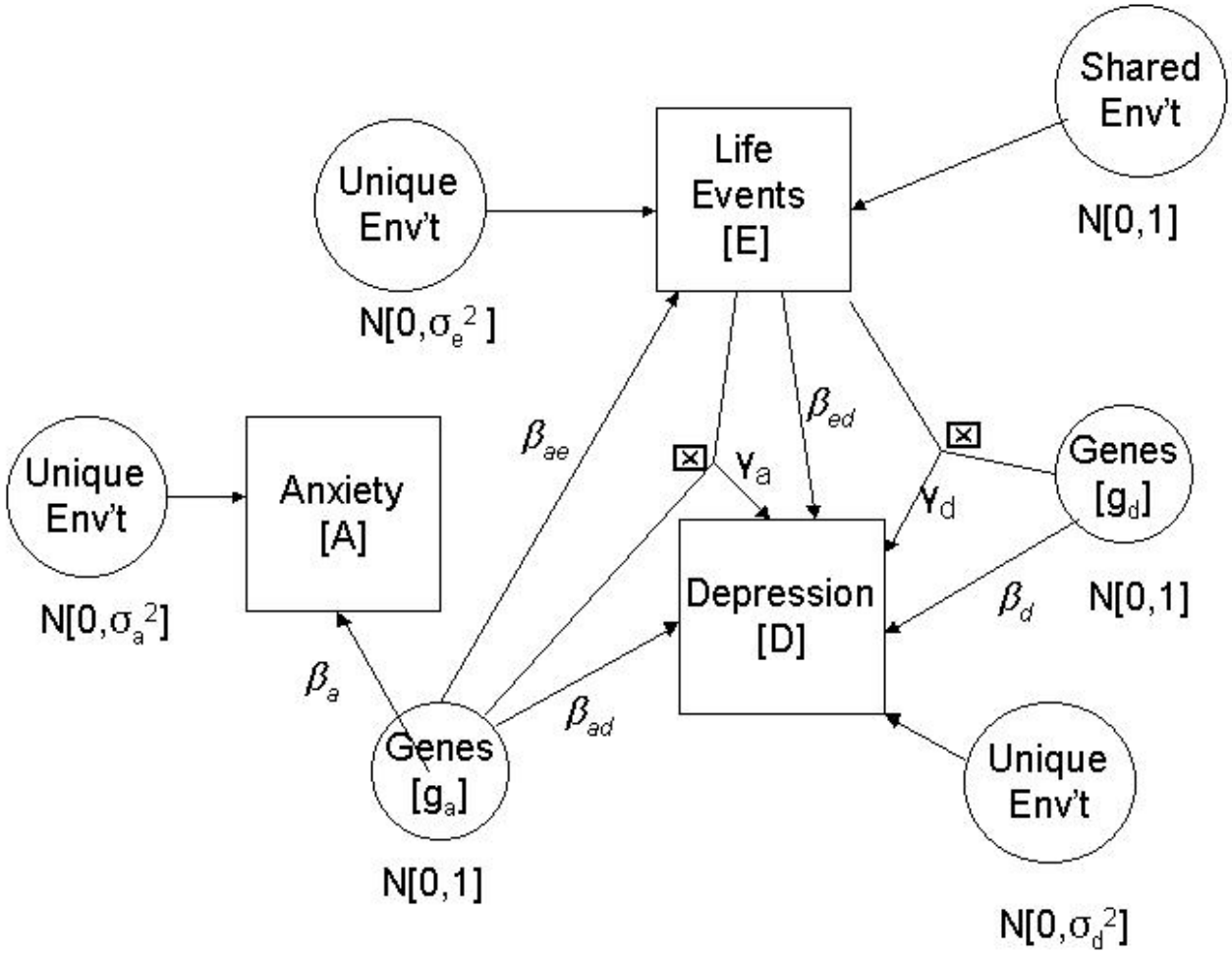
GxE

Genetic Variance and Shared Life Events in Adolescent Females



Putting it all together?

Multiple Genetic Pathways to Depression



Eaves et al. 2003