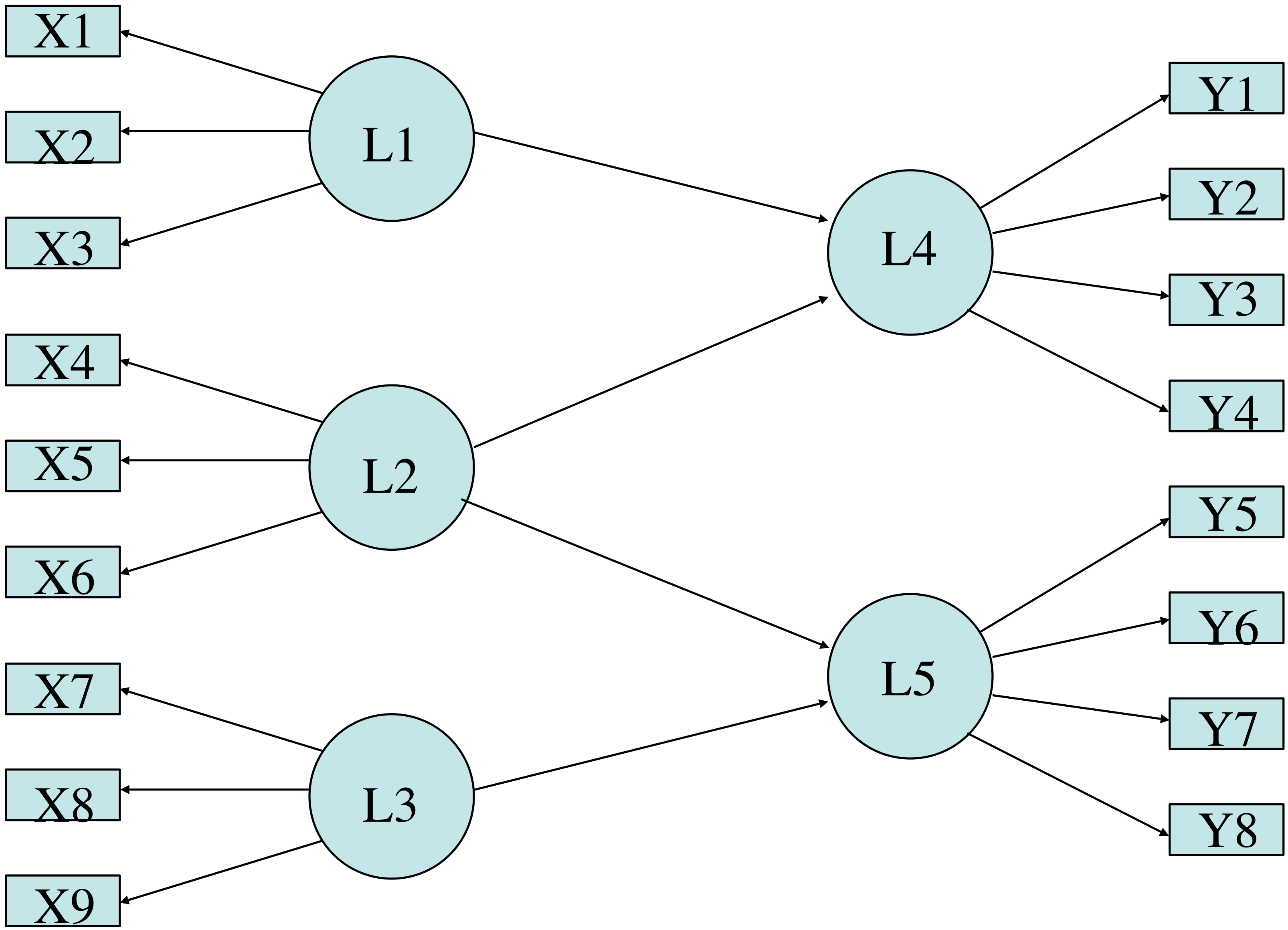
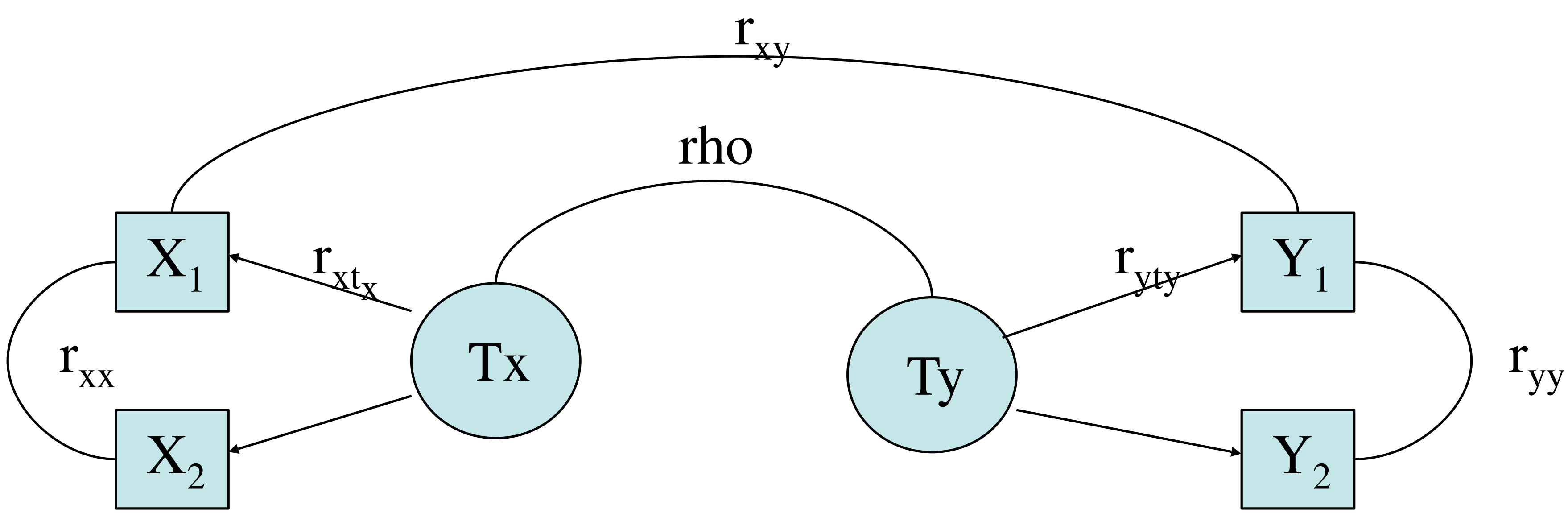
Validity

Face, Concurrent, Predictive,
Construct

Psychometric Theory: A conceptual Syllabus



Reliability- Correction for attenuation

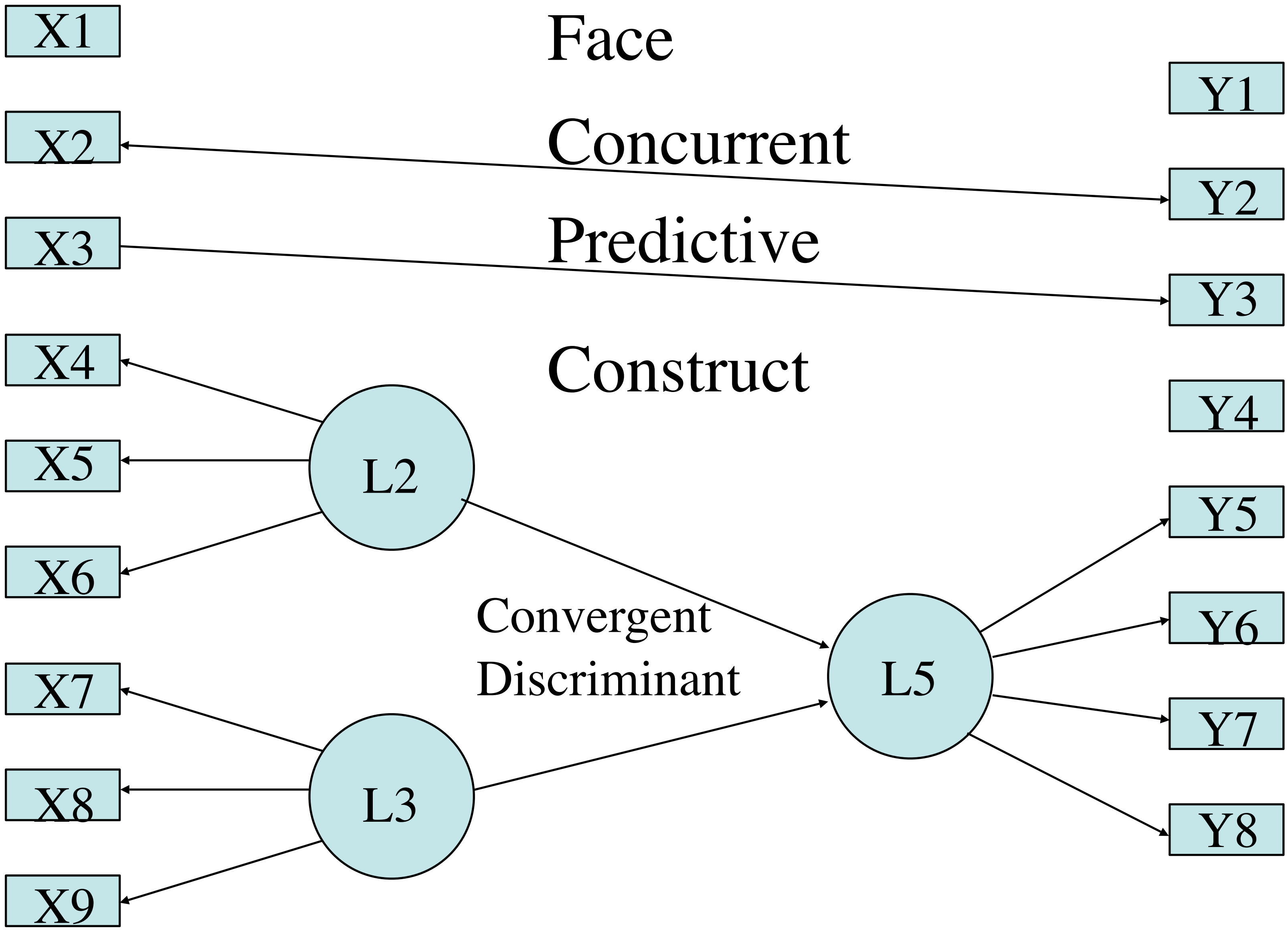


$r_{xt_x} = \text{sqrt}(r_{xx})$

$r_{yty} = \text{sqrt}(r_{yy})$

$$\text{Rho} = \text{rxy}/\text{sqrt}(r_{xx} * r_{yy})$$

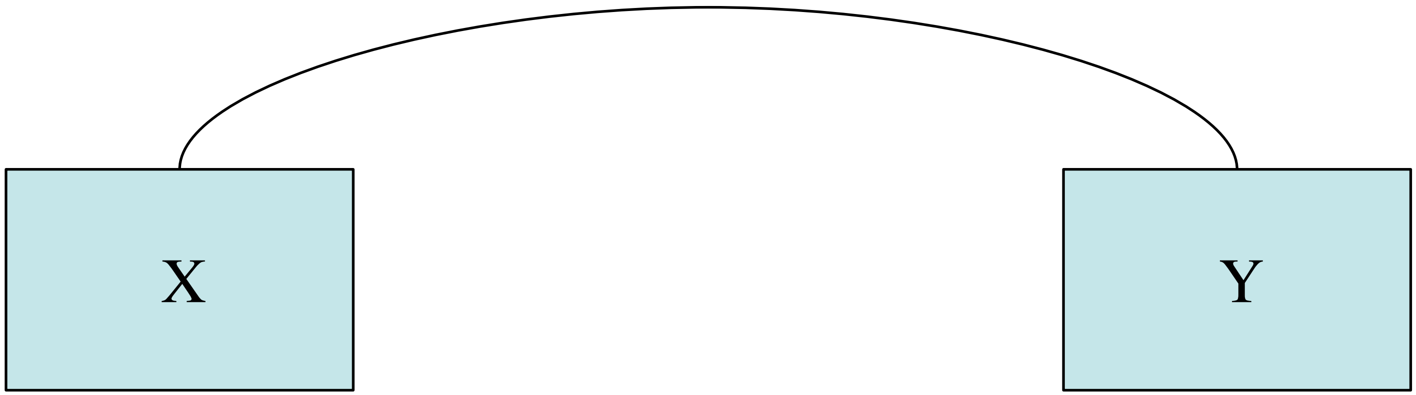
Types of Validity: What are we measuring



Face (Faith Validity)

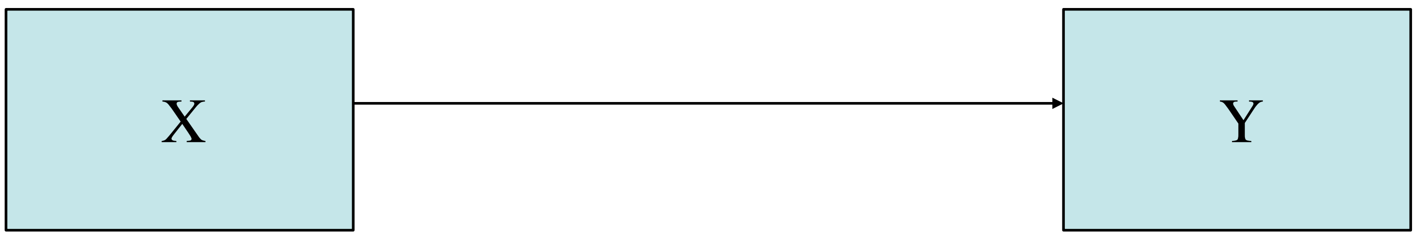
- Representative content
- Seeming relevance

Concurrent Validity



- Does a measure correlate with the criterion?
- Need to define the criterion.
- Assumes that what correlates now will have predictive value.

Predictive Validity

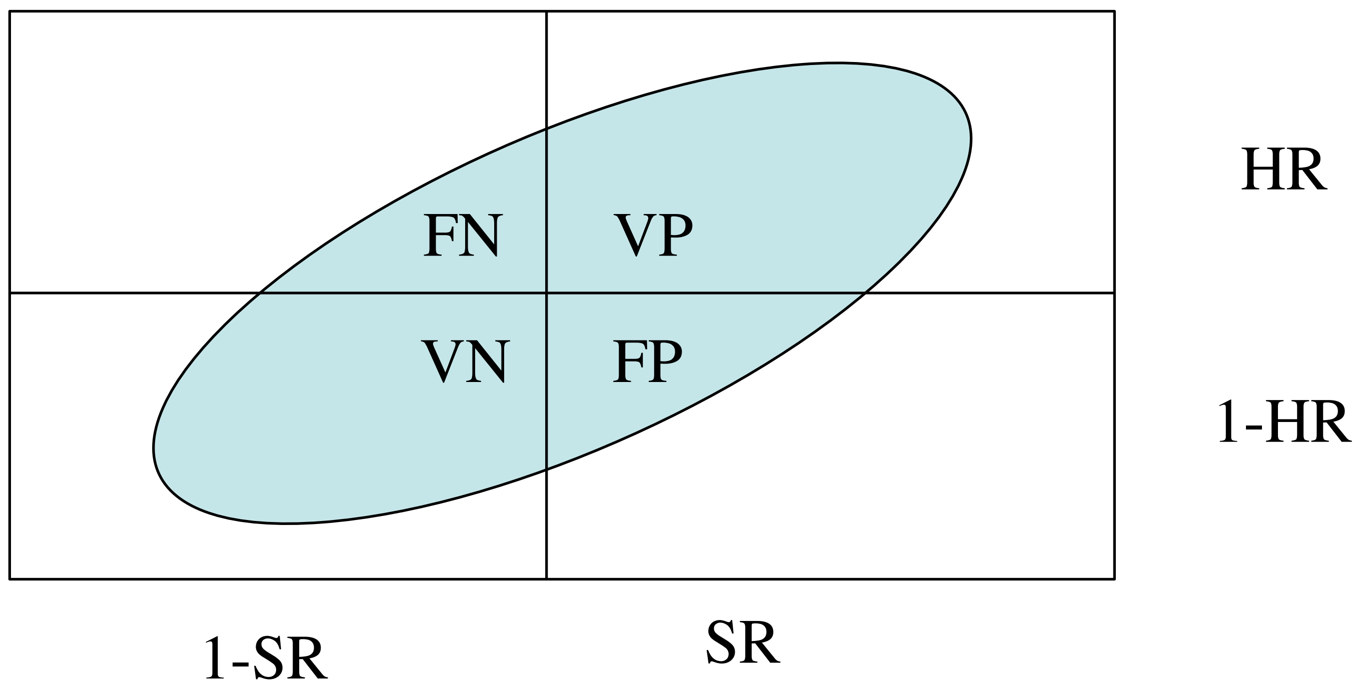


- Does a measure correlate with the criterion?
- Need to define the criterion.
- Requires waiting for time to pass.

Predictive and Concurrent Validity and Decision Making

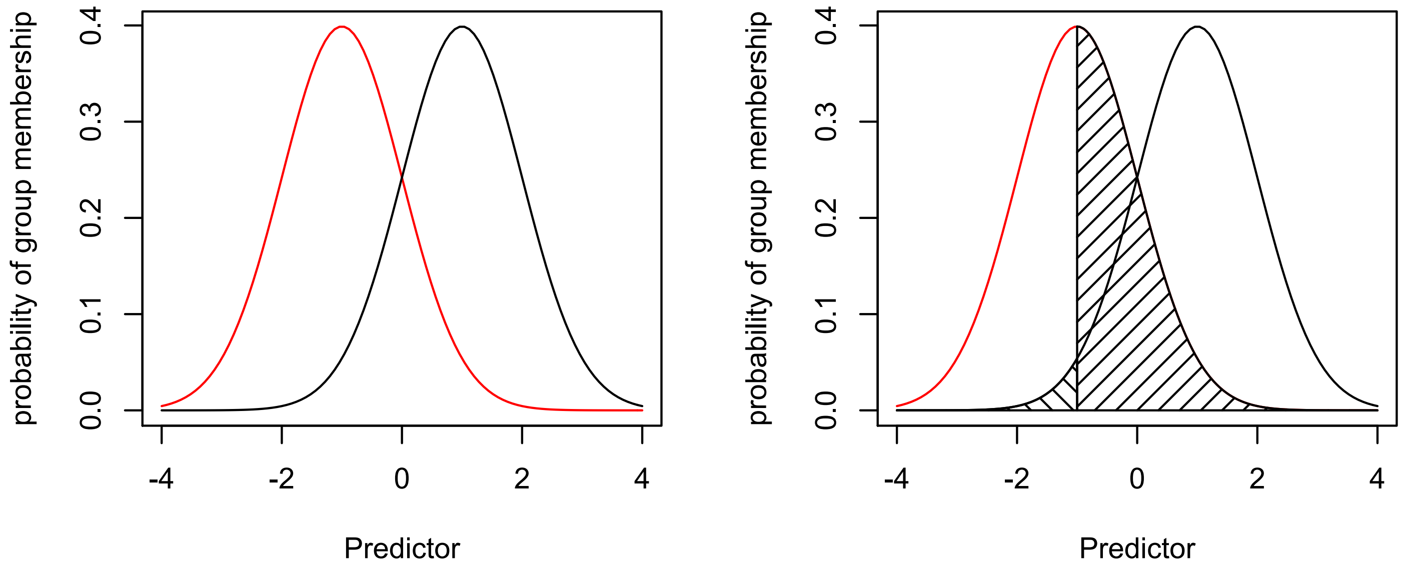
Hit Rate = Valid Positive + False Negative

Selection Ratio = Valid Positive + False Positive

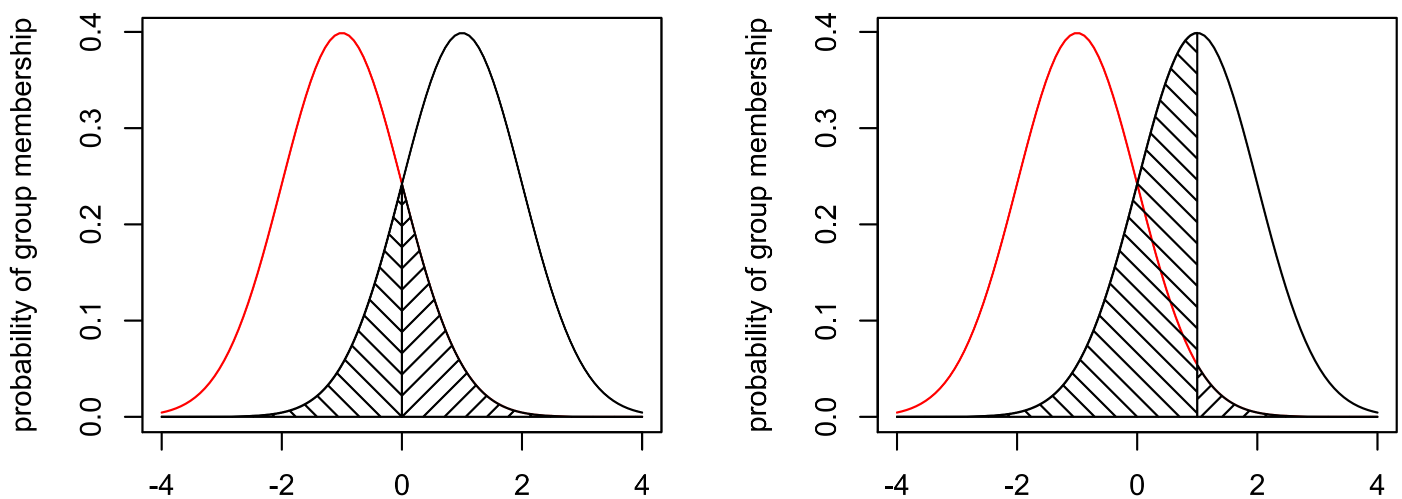


Phi =(VP - HR*SR) /sqrt(HR*(1-HR)*(SR)*(1-SR)

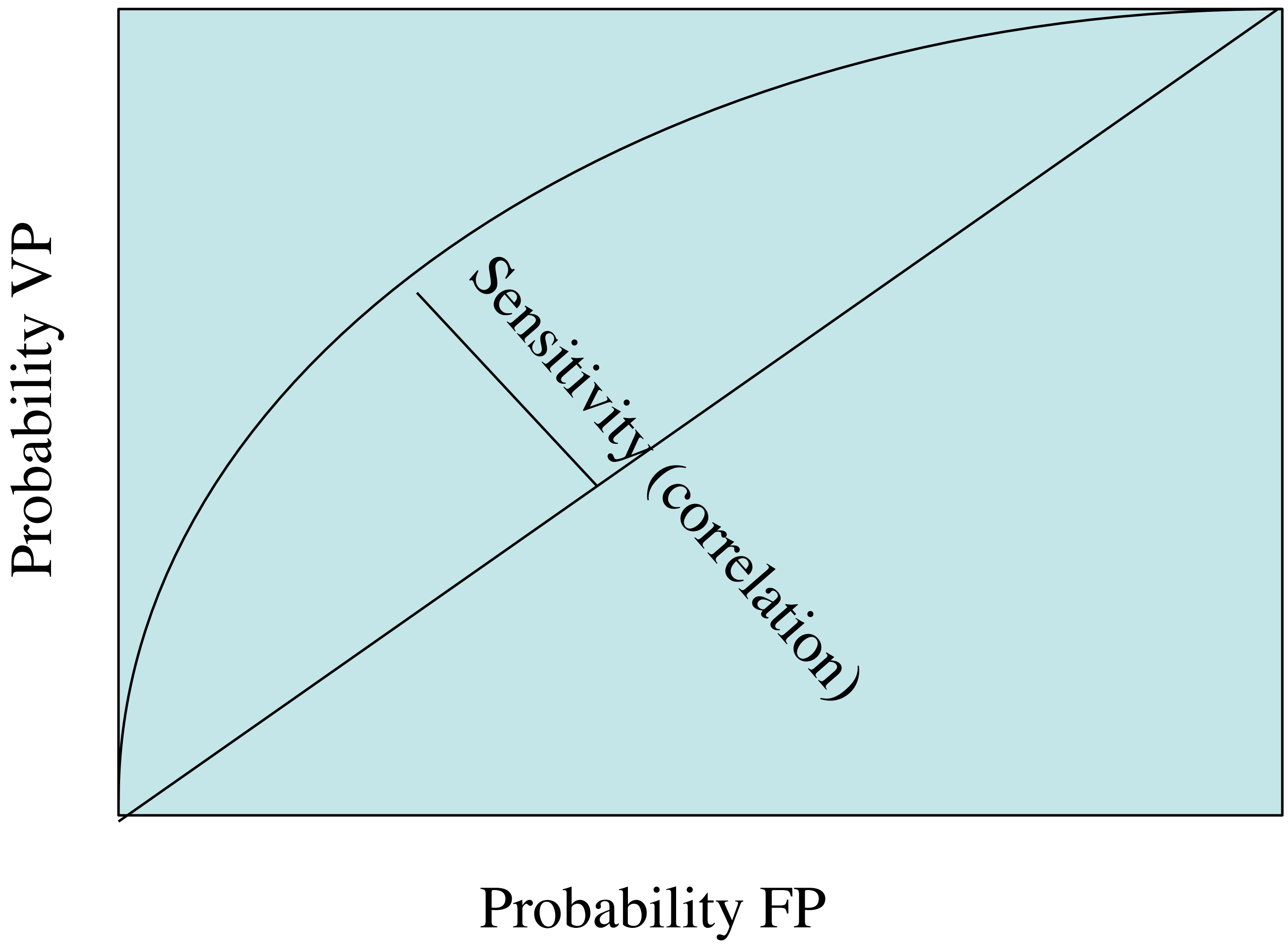
Validity as decision making



Trading off Valid positives for False Positives



Decision Theory and Signal Detection



Signal detection theory

- d prime and beta
  - d prime maps to the correlation
  - beta maps to selection ratio
- type I and type II error
  - Need to consider utility of types of error

Predictive Validity and Decision Theory

			State of world
	FN	VP	Hit rate
	VN	FP	1-HR
Decision	1-SR	Selection Ratio	

Predictive Validity, Utility and Decision

			State of world
	FN $*U_{FN}$	VP $*U_{VP}$	Hit rate
	VN $*U_{VN}$	FP* U_{FP}	1-HR
Decision	1-SR	Selection Ratio	

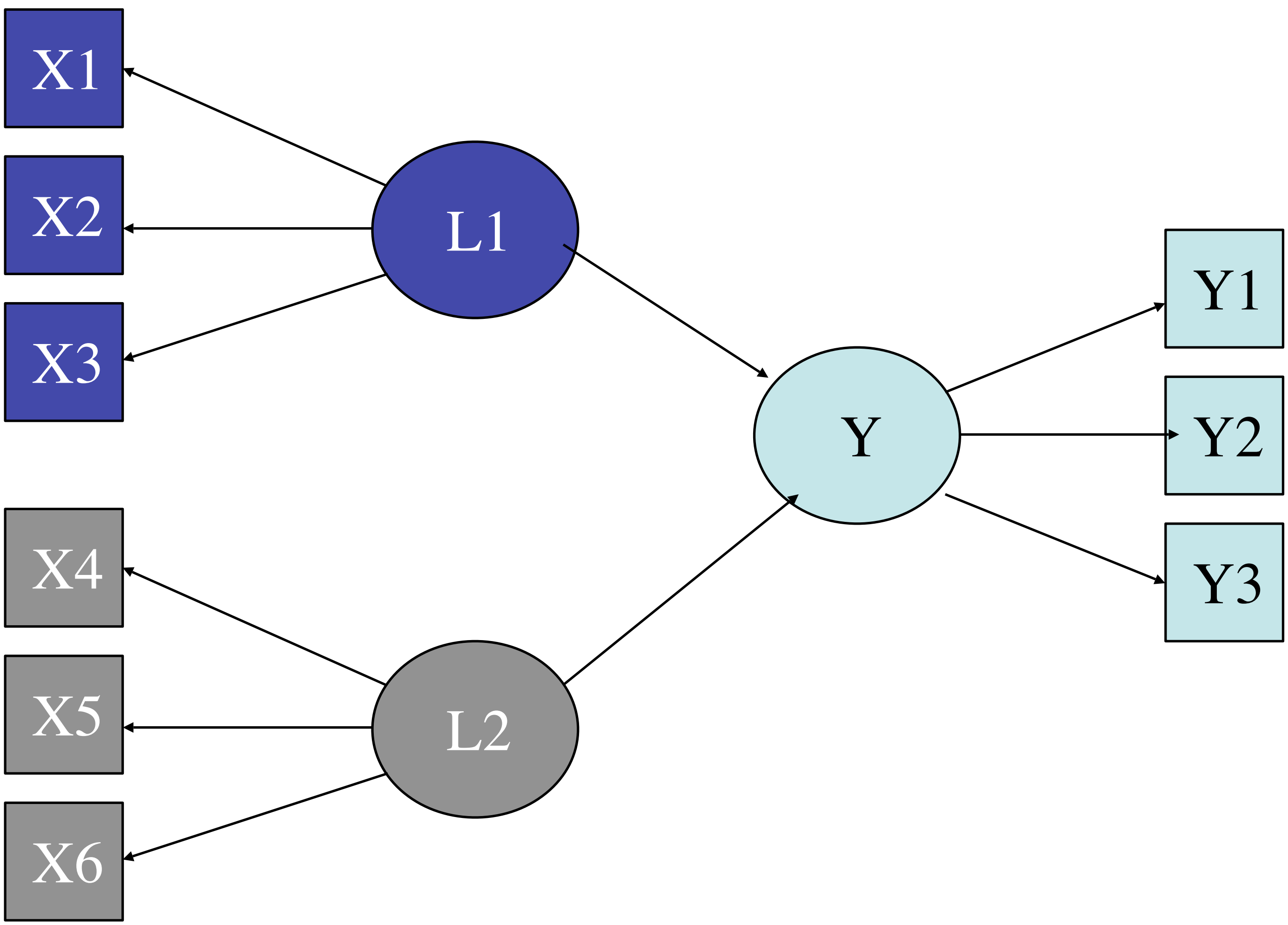
Utility of test = VP $*U_{VP}$ + VN $*U_{VN}$ + FN $*U_{FN}$ + FP* U_{FP} - Cost of test

Decisions for institutions, advice for individuals

			State of world
	FN $*U_{FN}$	VP $*U_{VP}$	Hit rate
	VN $*U_{VN}$	FP* U_{FP}	1-HR
Decision	1-SR	Selection Ratio	

Utility of test = VP $*U_{VP}$ + VN $*U_{VN}$ + FN $*U_{FN}$ + FP* U_{FP} - Cost of test

Construct Validity: Convergent, Discriminant, Incremental



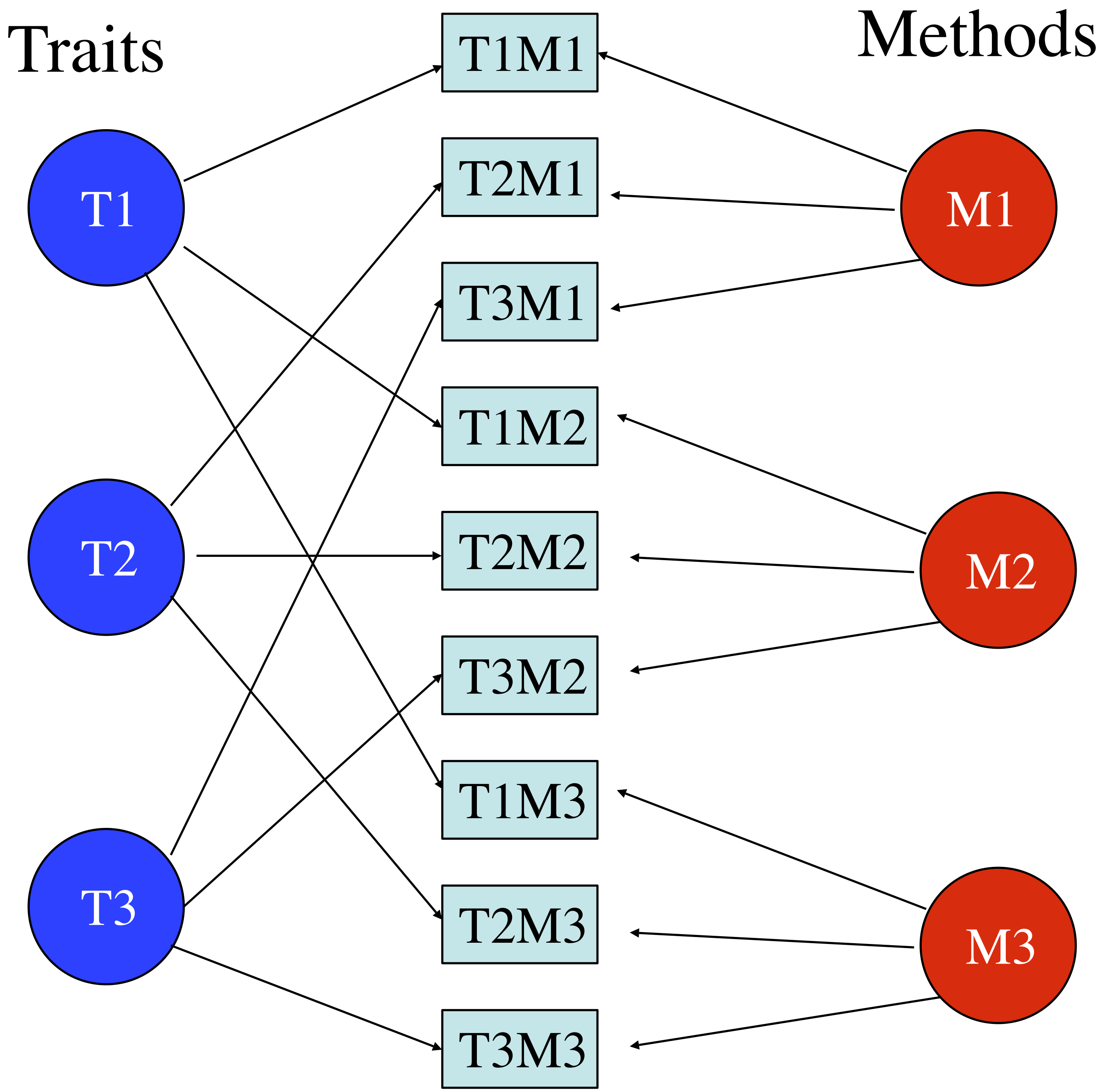
Multi-Trait, Multi-Method Matrix

	T1M1	T2M1	T3M1	T1M2	T2M2	T3M2	T1M3	T2M3	T3M3
T1M1	T1M1								
T2M1	M1	T2M1							
T3M1	M1	M1	T3M1						
T1M2	T1			T1M2					
T2M2		T2		M2	T2M2				
T3M2			T3	M2	M2	T3M2			
T1M3	T1			T1			T1M3		
T2M3		T2			T2		M3	T2M3	
T3M3			T3			T3	M3	M3	T3M3

Mono-Method, Mono trait = reliability

Hetero Method, Mono Trait = convergent validity

Hetero Method, Hetero Trait = discriminant validity



Traits
Methods
T1
T2
T3
M1
M2
M3
T1M1
T2M1
T3M1
T1M2
T2M2
T3M2
T1M3
T2M3
T3M3

Model Fitting

Structural Equation Models

Reliability + Validity

Basic concepts of iterative fit

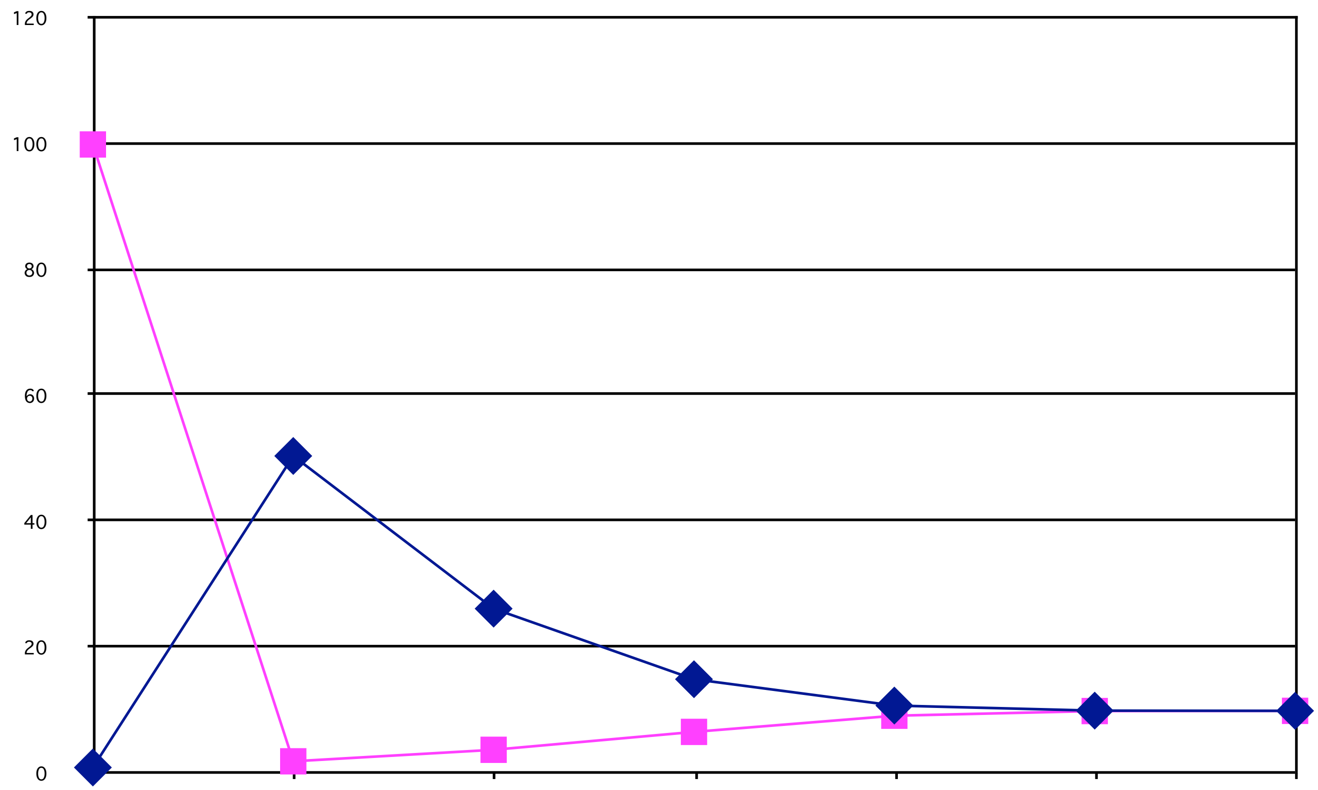
- Most classical statistics (e.g. means, variances, regression slopes) may be found by algebraic solutions of closed form expressions
- More recent statistics are the results of iteratively fitting a model until some criterion is either minimized or maximized.

Simple example: the square root

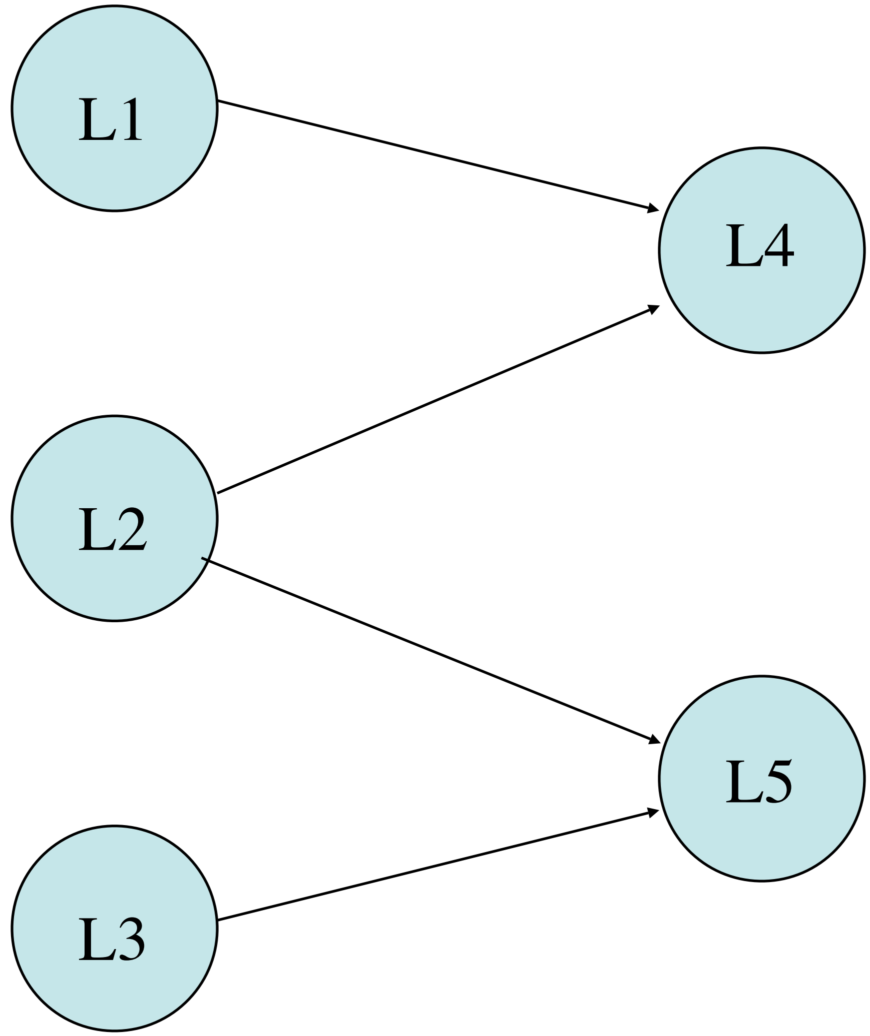
Target 100

Trial	guess	fit	diff
1	1.0	100.0	-99.0
2	50.5	2.0	48.5
3	26.2	3.8	22.4
4	15.0	6.7	8.4
5	10.8	9.2	1.6
6	10.0	10.0	0.1
7	10.0	10.0	0.0

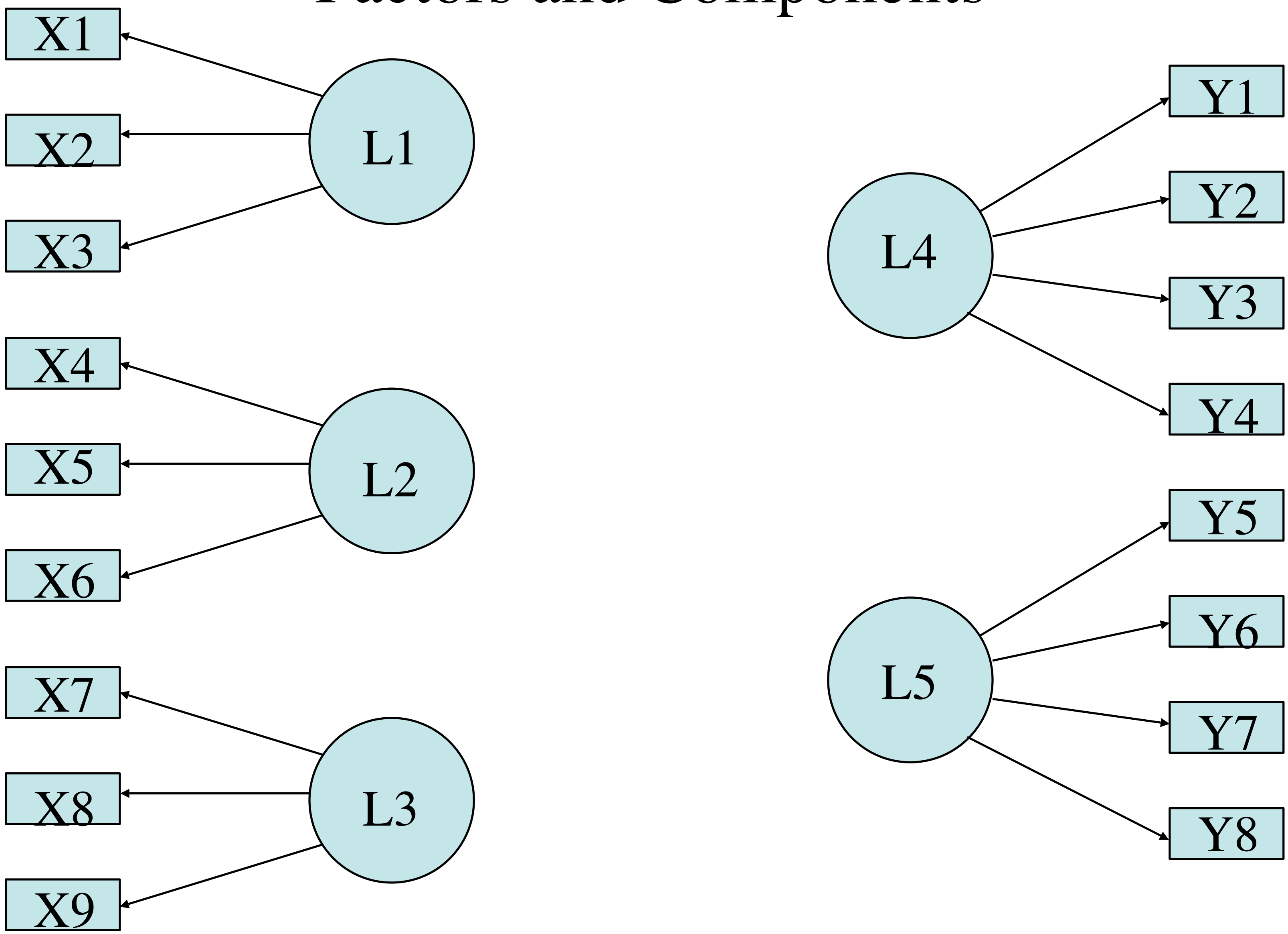
Iteratively estimating the square root of 100



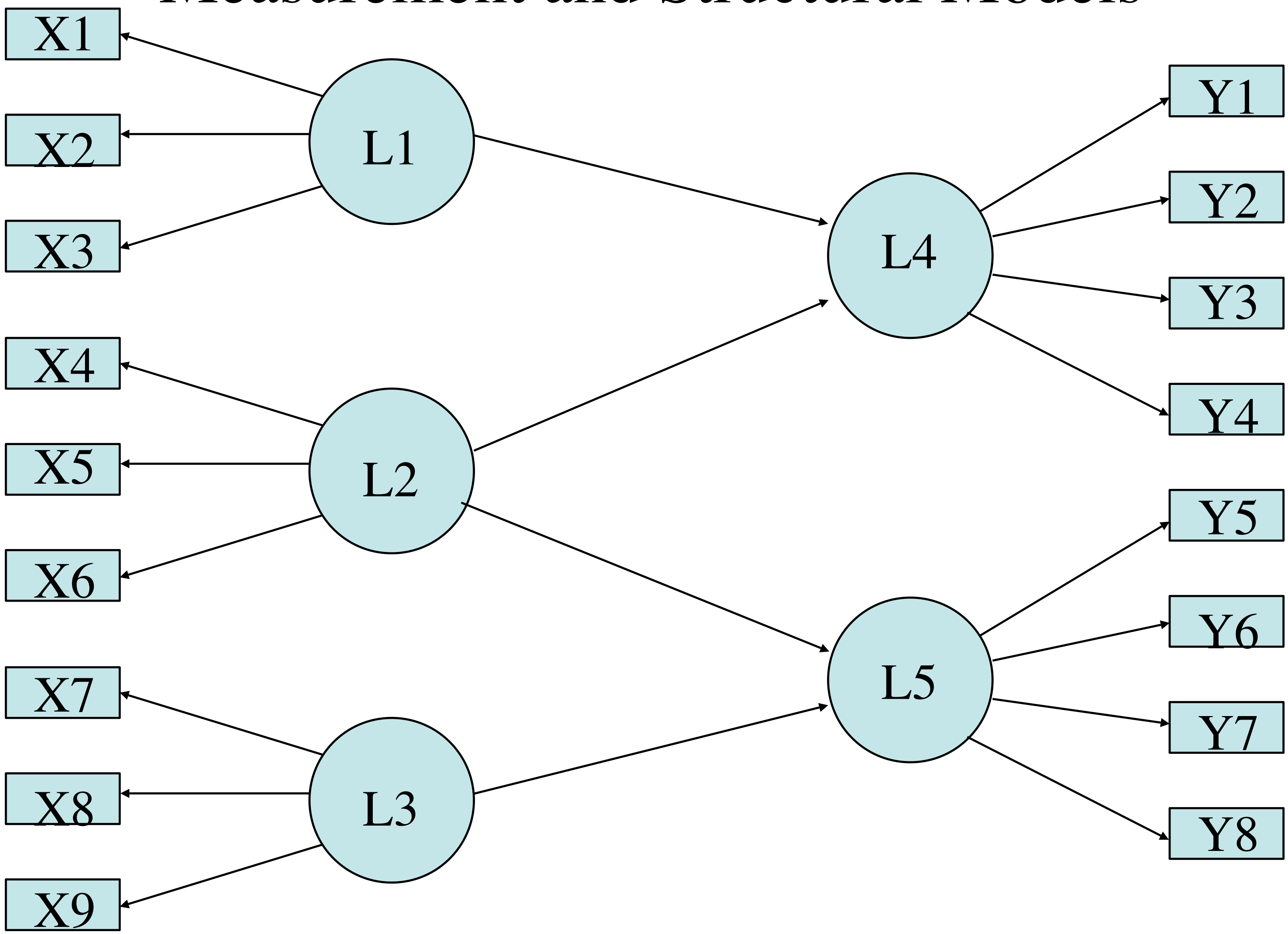
Theory as organization of constructs



Techniques of Data Reduction: Factors and Components



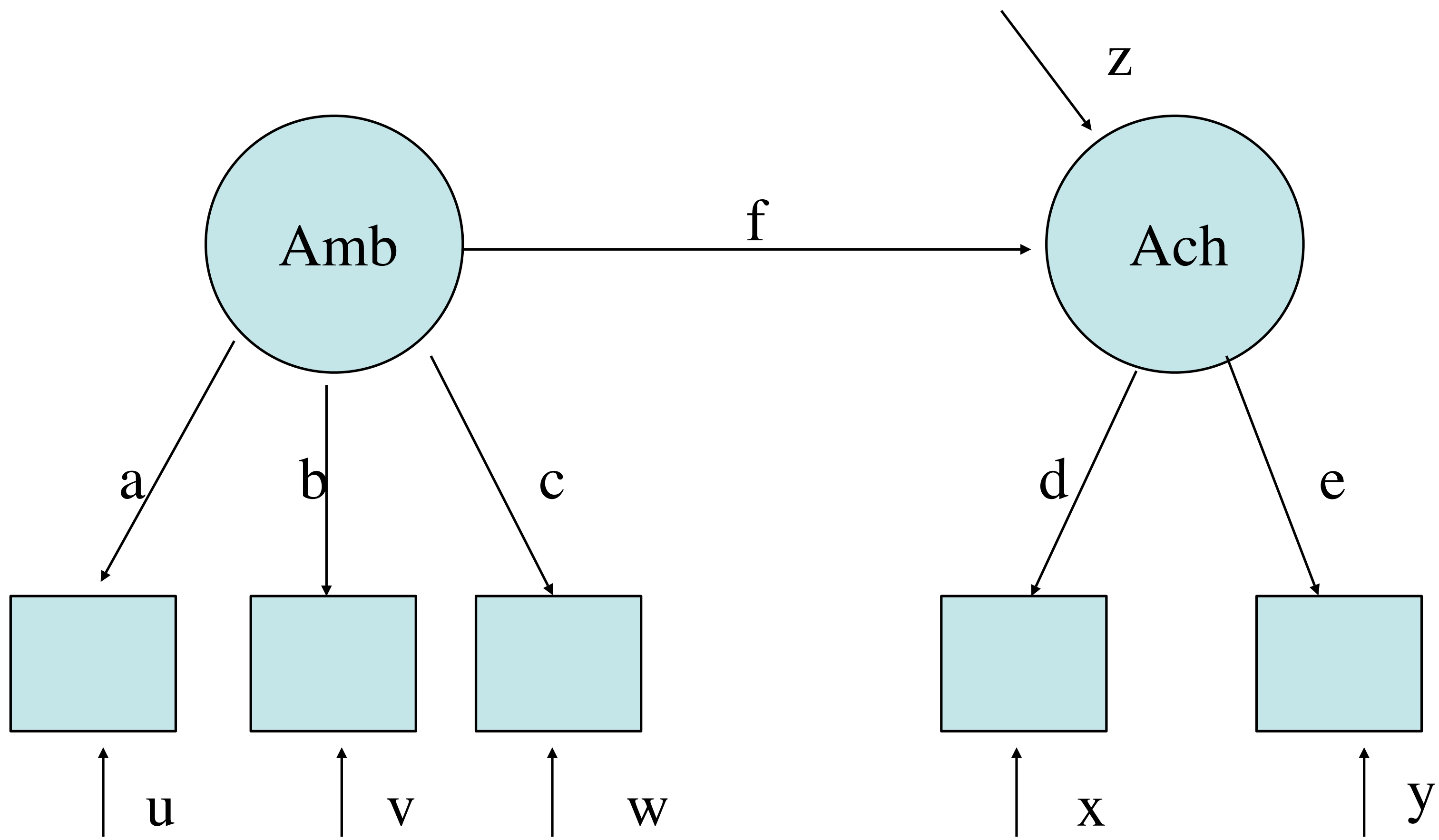
Structural Equation Modeling: Combining Measurement and Structural Models



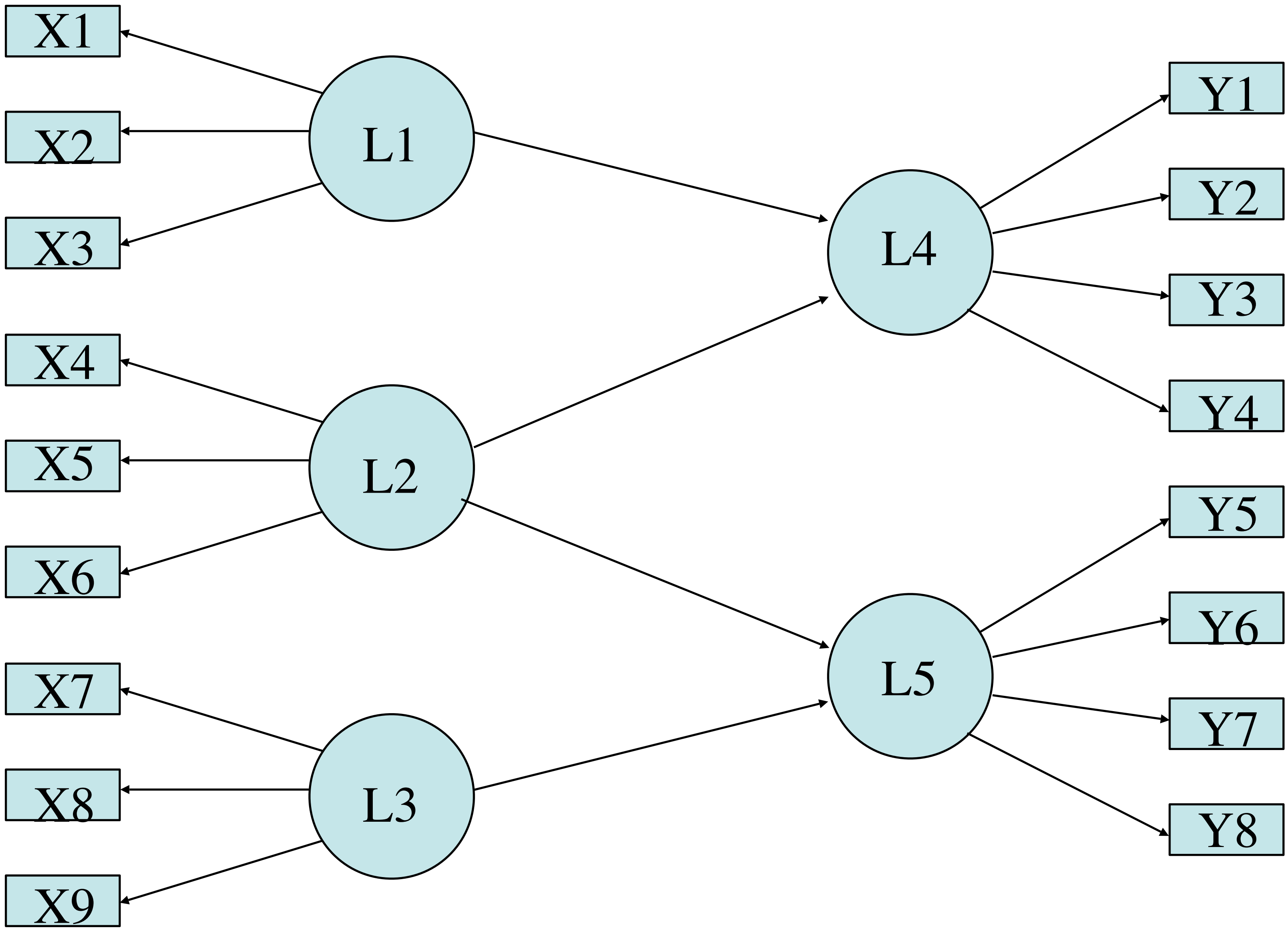
SEM problem (Loehlin 2.5)

Ach1	Ach2	Amb1	Amb2	Amb3
1	0.6	0.3	0.2	0.2
0.6	1	0.2	0.3	0.1
0.3	0.2	1	0.7	0.6
0.2	0.3	0.7	1	0.5
0.2	0.1	0.6	0.5	1

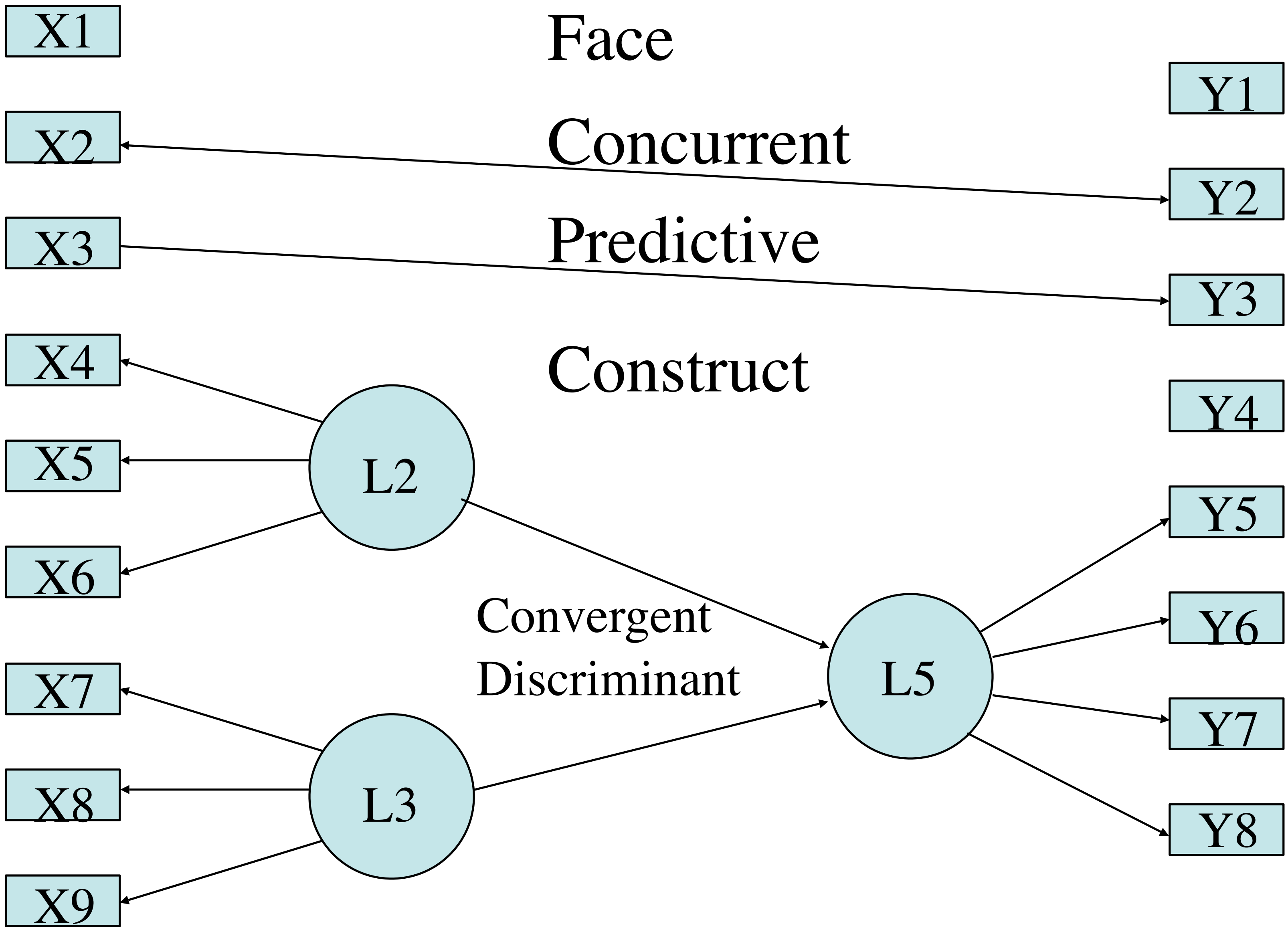
Ambition and Achievement



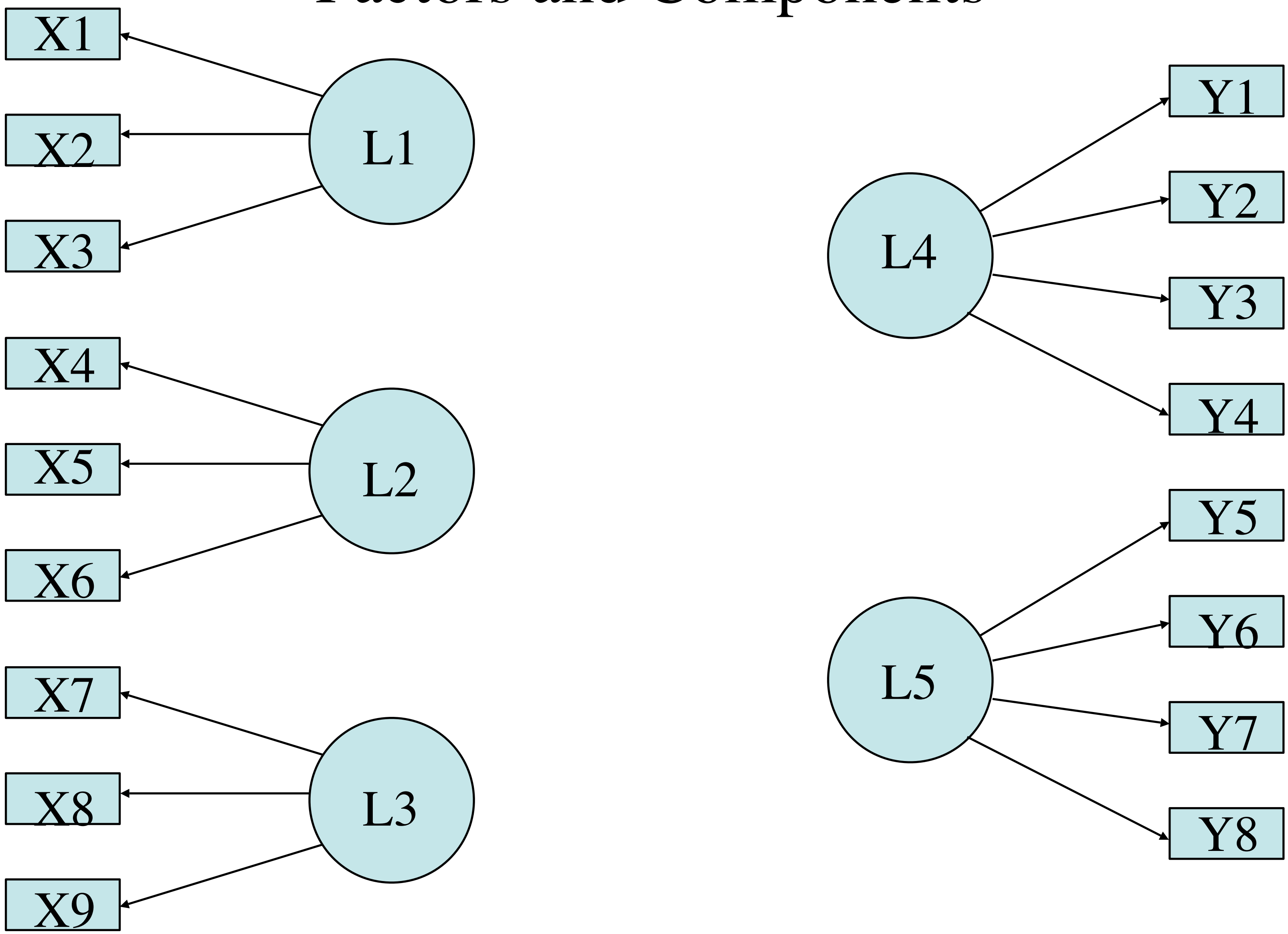
Psychometric Theory: A conceptual Syllabus



Types of Validity: What are we measuring



Techniques of Data Reduction: Factors and Components



Methods of Homogeneous Keying

Factor analysis, principal components analysis, and cluster analysis

Methods of Homogeneous Keying

- Factor Analysis
- Principal Components Analysis
- Cluster Analysis

Factor Analysis
Consider the following r matrix

	X_1	X_2	X_3	X_4	X_5
X_1	1.00				
X_2	0.72	1.00			
X_3	0.63	0.56	1.00		
X_4	0.54	0.48	0.42	1.00	
X_5	0.45	0.40	0.35	0.30	1.00

Create an example factor matrix

```
x <- c(.9,.8,.7,.6,.5,.4)
 r <- x %*% t(x)
 colnames(r) <- rownames(r) <- paste("V",1:6,sep="")
 diag(r) <- 1
 r
f1 <- factor.pa(r,1,TRUE,max.iter=1)
```

Observed Correlation matrix

	V1	V2	V3	V4	V5	V6
V1	1.00	0.72	0.63	0.54	0.45	0.36
V2	0.72	1.00	0.56	0.48	0.40	0.32
V3	0.63	0.56	1.00	0.42	0.35	0.28
V4	0.54	0.48	0.42	1.00	0.30	0.24
V5	0.45	0.40	0.35	0.30	1.00	0.20
V6	0.36	0.32	0.28	0.24	0.20	1.00

Find the characteristic roots

- Eigen vectors as characteristic roots of a matrix. (The vectors unchanged by the matrix.)
- Eigen values as a scaling of the roots
- Given a matrix R, the eigen vectors solve the equation
  - xR = vx or XR = vX where X is a matrix of eigen vectors and v is a vector of lengths of the X
  - XR - vXI = 0 <=> X(R-vI) = 0 and we solve for X

Eigen vectors and eigen values

$values

ev1	ev2	ev3	ev4	ev5	ev6
3.16	0.82	0.72	0.59	0.44	0.26

$vectors

	ev1	ev2	ev3	ev4	ev5	ev6
ev1	-0.50	0.061	0.092	0.14	0.238	0.816
ev2	-0.47	0.074	0.121	0.21	0.657	-0.533
ev3	-0.43	0.096	0.182	0.53	-0.675	-0.184
ev4	-0.39	0.142	0.414	-0.78	-0.201	-0.104
ev5	-0.34	0.299	-0.860	-0.20	-0.108	-0.067
ev6	-0.28	-0.934	-0.178	-0.10	-0.067	-0.045

Principal components

- Merely a transformation of original matrix to show the eigen vectors and eigen values
- The vectors provide the basis space of the original matrix
- The lengths of the vectors (the eigen values) reflect the importance of the vector
- Also known as singular value decomposition,

Principal components as rescaled eigen vectors

- If V = the matrix of eigen vectors and v = vector of eigen values
- Although I = VV' = V'V
- if we let C = sqrt(v)*V then
- R = CC'

Principal Components

```
e <- eigen(r)
c <- matrix(rep(sqrt(e$values),6),byrow=TRUE,ncol=6) * e
$vectors
```

	ev1	ev2	ev3	ev4	ev5	ev6
ev1	-0.88	0.055	0.078	0.107	0.158	0.417
ev2	-0.83	0.067	0.103	0.165	0.436	-0.273
ev3	-0.77	0.087	0.154	0.408	-0.448	-0.094
ev4	-0.69	0.128	0.351	-0.599	-0.133	-0.053
ev5	-0.60	0.271	-0.729	-0.151	-0.071	-0.034
ev6	-0.50	-0.846	-0.151	-0.077	-0.044	-0.023

Take just the first n principal components

- R = CC’ for all components
- R ≈ CC’ for first n components
- Evaluate R* = R - CC’
- pc <- principal(r,1,residuals=TRUE)

```
$loadings
   PC1
V1 0.88
V2 0.83
V3 0.77
V4 0.69
V5 0.60
V6 0.50
```

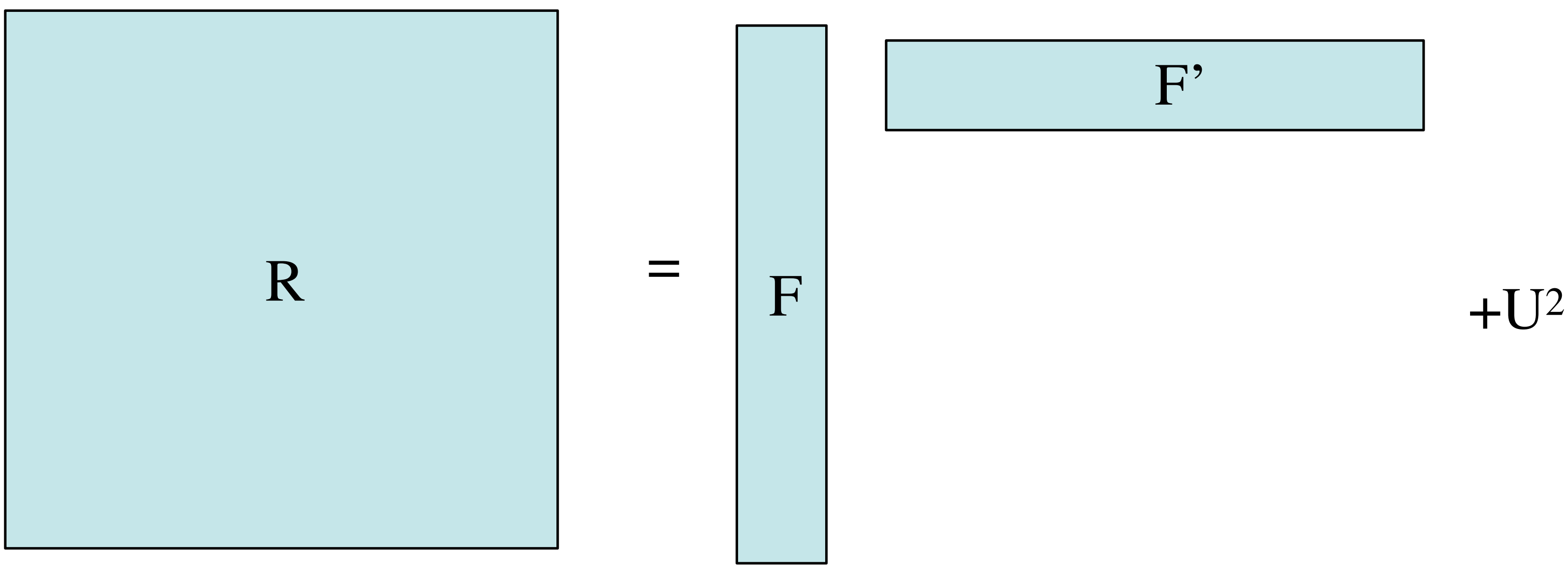
R - CC' for 1 component

$residual

	V1	V2	V3	V4	V5	V6
V1	0.220	-0.015	-0.050	-0.073	-0.084	-0.084
V2	-0.015	0.307	-0.081	-0.098	-0.103	-0.098
V3	-0.050	-0.081	0.408	-0.114	-0.115	-0.107
V4	-0.073	-0.098	-0.114	0.519	-0.119	-0.108
V5	-0.084	-0.103	-0.115	-0.119	0.635	-0.104
V6	-0.084	-0.098	-0.107	-0.108	-0.104	0.748

Fit = $1 - R^{*2}/R^2 = .85$

Factor analysis



Factor Analysis: the model

- $R \approx FF' + U^2$
- Residual Matrix R* = R - (FF'+ U2)
- Try to minimize the residual
- Variables are linear composites of unknown (latent) factors.
- Covariance structures of observables in terms of covariance of unobservables

Structure of mood - how not to display data

	AFRAID	AT_EASE	CALM	ENERGETI	HAPPY	LIVELY	SLEEPY	TENSE	TIRED
AFRAID	1.000								
AT_EASE	-0.209	1.000							
CALM	-0.157	0.586	1.000						
ENERGETI	0.019	0.230	0.056	1.000					
HAPPY	-0.070	0.452	0.294	0.595	1.000				
LIVELY	0.018	0.255	0.073	0.778	0.609	1.000			
SLEEPY	0.087	-0.112	0.031	-0.457	-0.264	-0.405	1.000		
TENSE	0.397	-0.337	-0.332	0.088	-0.103	0.084	0.044	1.000	
TIRED	0.082	-0.141	0.012	-0.484	-0.297	-0.439	0.808	0.044	1.000
UNHAPPY	0.350	-0.283	-0.187	-0.185	-0.314	-0.187	0.202	0.360	0.235

Structure of mood:
“Alabama need not come first”

	AFRAID	AT_EASE	CALM	ENERGETI	HAPPY	LIVELY	SLEEPY	TENSE	TIRED
AFRAID	1.0								
AT_EASE	-0.2	1.0							
CALM	-0.2	0.6	1.0						
ENERGETI	0.0	0.2	0.1	1.0					
HAPPY	-0.1	0.5	0.3	0.6	1.0				
LIVELY	0.0	0.3	0.1	0.8	0.6	1.0			
SLEEPY	0.1	-0.1	0.0	-0.5	-0.3	-0.4	1.0		
TENSE	0.4	-0.3	-0.3	0.1	-0.1	0.1	0.0	1.0	
TIRED	0.1	-0.1	0.0	-0.5	-0.3	-0.4	0.8	0.0	1.0
UNHAPPY	0.3	-0.3	-0.2	-0.2	-0.3	-0.2	0.2	0.4	0.2

Structure of mood data

	ENERGETI	LIVELY	TIRED	SLEEPY	AFRAID	TENSE	AT_EASE	CALM	HAPPY
ENERGETI	1								
LIVELY	0.8	1							
TIRED	-0.5	-0.4	1						
SLEEPY	-0.5	-0.4	0.8	1					
AFRAID	0	0	0.1	0.1	1				
TENSE	0.1	0.1	0	0	0.4	1			
AT_EASE	0.2	0.3	-0.1	-0.1	-0.2	-0.3	1		
CALM	0.1	0.1	0	0	-0.2	-0.3	0.6	1	
HAPPY	0.6	0.6	-0.3	-0.3	-0.1	-0.1	0.5	0.3	1
UNHAPPY	-0.2	-0.2	0.2	0.2	0.3	0.4	-0.3	-0.2	-0.3

NUMBER OF OBSERVATIONS: 3748

Correlation of mood data possible structure

	ENERGETI	LIVELY	TIRED	SLEEPY	AFRAID	TENSE	AT_EASE	CALM	HAPPY
ENERGETI	1.0								
LIVELY	0.8	1.0							
TIRED	-0.5	-0.4	1.0						
SLEEPY	-0.5	-0.4	0.8	1.0					
AFRAID	0.0	0.0	0.1	0.1	1.0				
TENSE	0.1	0.1	0.0	0.0	0.4	1.0			
AT_EASE	0.2	0.3	-0.1	-0.1	-0.2	-0.3	1.0		
CALM	0.1	0.1	0.0	0.0	-0.2	-0.3	0.6	1.0	
HAPPY	0.6	0.6	-0.3	-0.3	-0.1	-0.1	0.5	0.3	1.0
UNHAPPY	-0.2	-0.2	0.2	0.2	0.3	0.4	-0.3	-0.2	-0.3

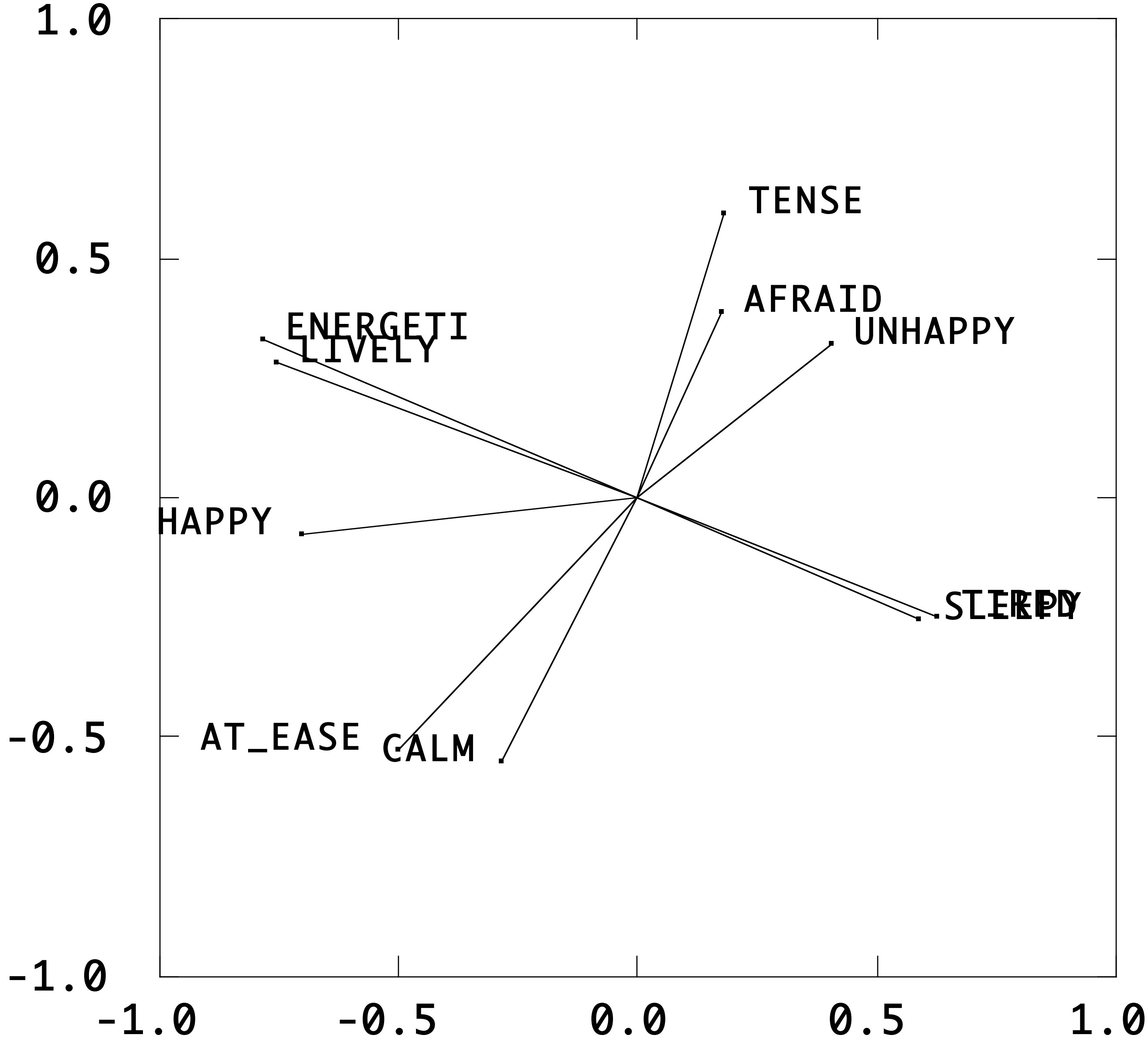
NUMBER OF OBSERVATIONS: 3748

Factor analysis 2 factor solution

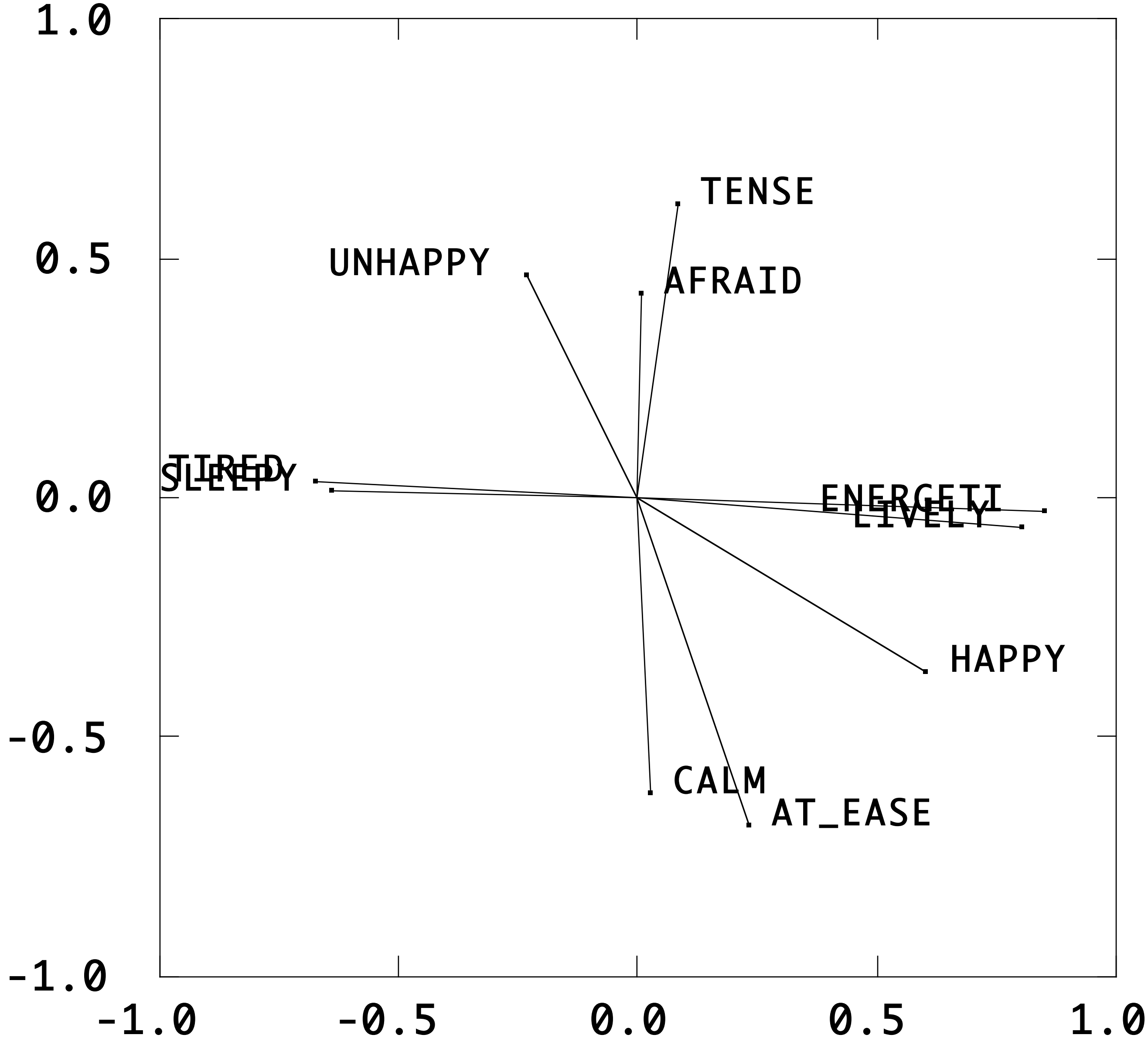
FACTOR PATTERN

	1	2	h2
ENERGETI	-0.8	0.3	0.73
LIVELY	-0.8	0.3	0.73
HAPPY	-0.7	-0.1	0.50
TIRED	0.6	-0.3	0.45
SLEEPY	0.6	-0.3	0.45
TENSE	0.2	0.6	0.40
CALM	-0.3	-0.6	0.45
AT_EASE	-0.5	-0.5	0.50
AFRAID	0.2	0.4	0.20
UNHAPPY	0.4	0.3	0.25
VARIANCE EXP	3.0	1.50	

2 factors of mood



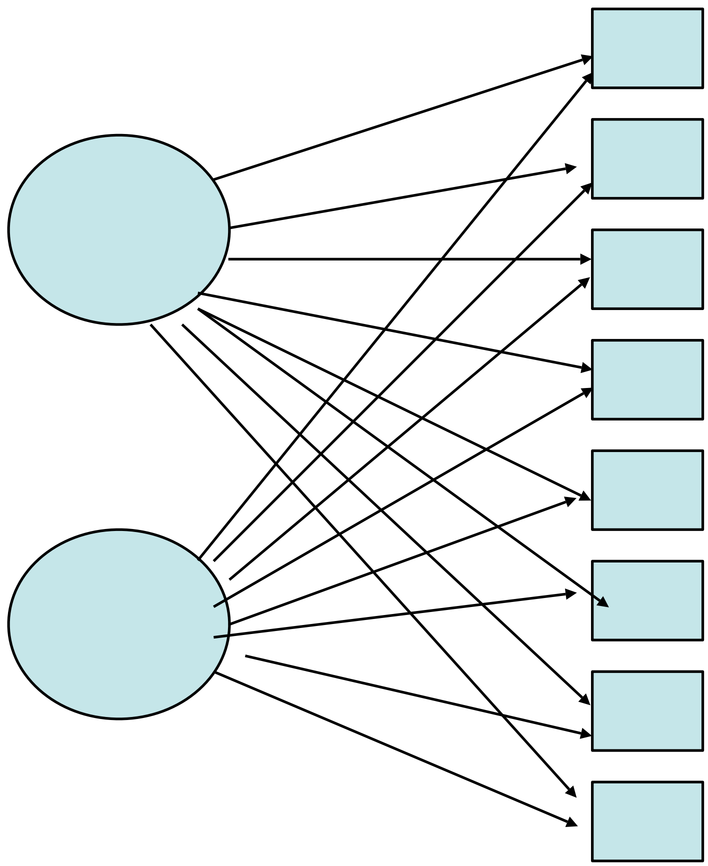
2 factors of mood (rotated)



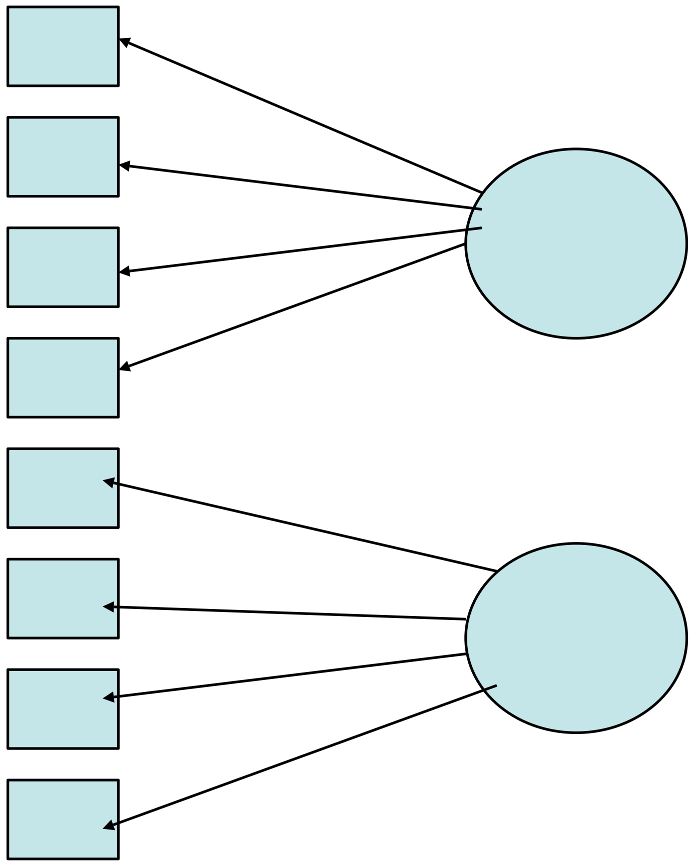
Rotation as orthogonal transformation

	F1	F2	F1'	F2'
ENERGETI	-0.8	0.3	0.8	0.0
LIVELY	-0.8	0.3	0.8	-0.1
HAPPY	-0.7	-0.1	0.6	-0.4
TIRED	0.6	-0.3	-0.7	0.0
SLEEPY	0.6	-0.3	-0.6	0.0
TENSE	0.2	0.6	0.1	0.6
CALM	-0.3	-0.6	0.0	-0.6
AT_EASE	-0.5	-0.5	0.2	-0.7
AFRAID	0.2	0.4	0.0	0.4
UNHAPPY	0.4	0.3	-0.2	0.5
eigen values	3	1.5	2.7	1.8

Rotation to simple structure Original

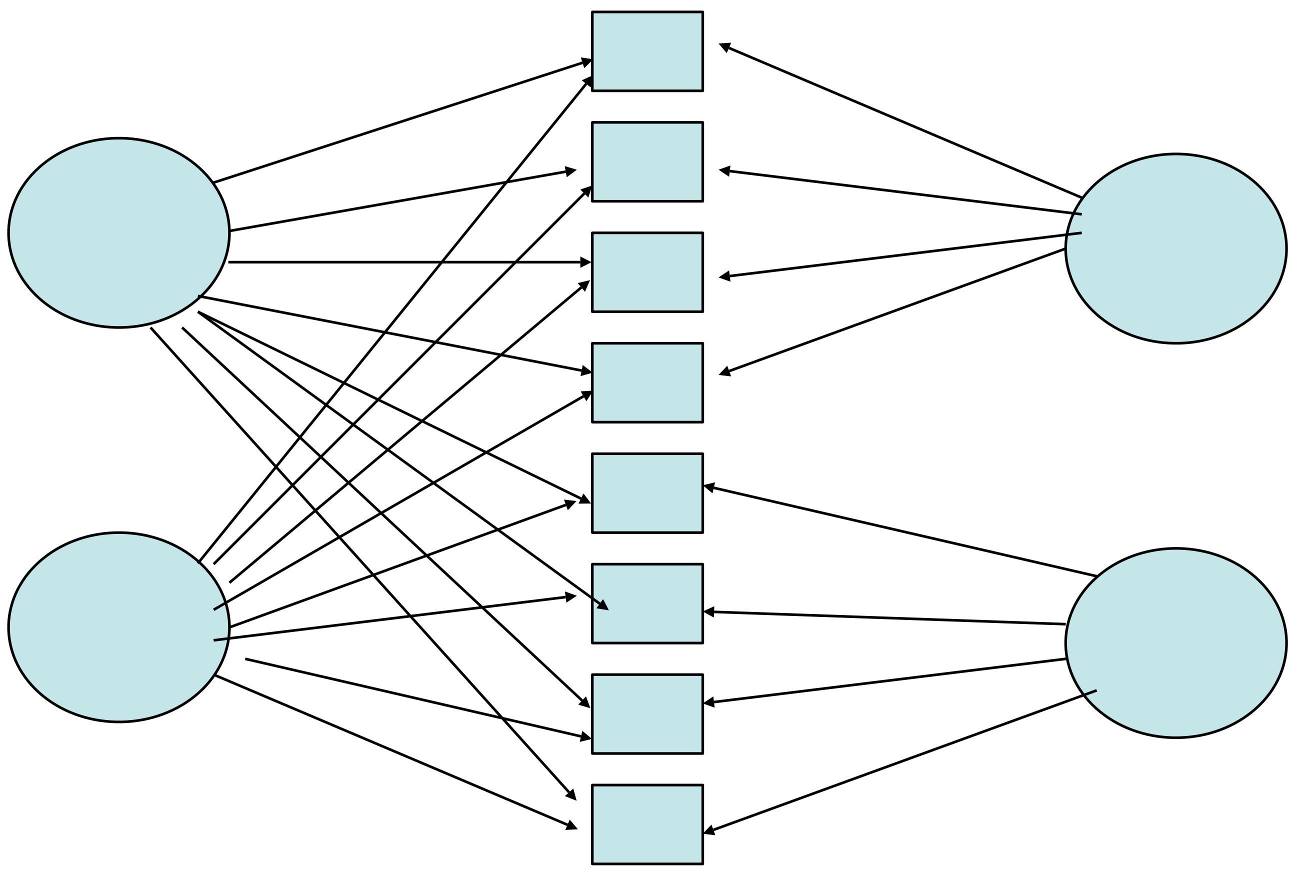


Rotation to simple structure
Simple Structure



Rotation to simple structure

Original Rotated



Principal components

- $R \approx CC'$
- Residual Matrix $R^* = R - (CC')$
- Try to minimize the residual
- Components are linear composites of known variables.
- Covariance structures of observables in terms of covariance of observables
- Components account for observed variance

2 Principal Components of mood

	C1	C2
ENERGETI	-0.8	-0.4
LIVELY	-0.8	-0.3
HAPPY	-0.8	0
TIRED	0.7	0.3
SLEEPY	0.7	0.3
AT_EASE	-0.6	0.5
TENSE	0.2	-0.7
CALM	-0.3	0.6
AFRAID	0.2	-0.5
UNHAPPY	0.5	-0.4
eigen values	3.4	2.1

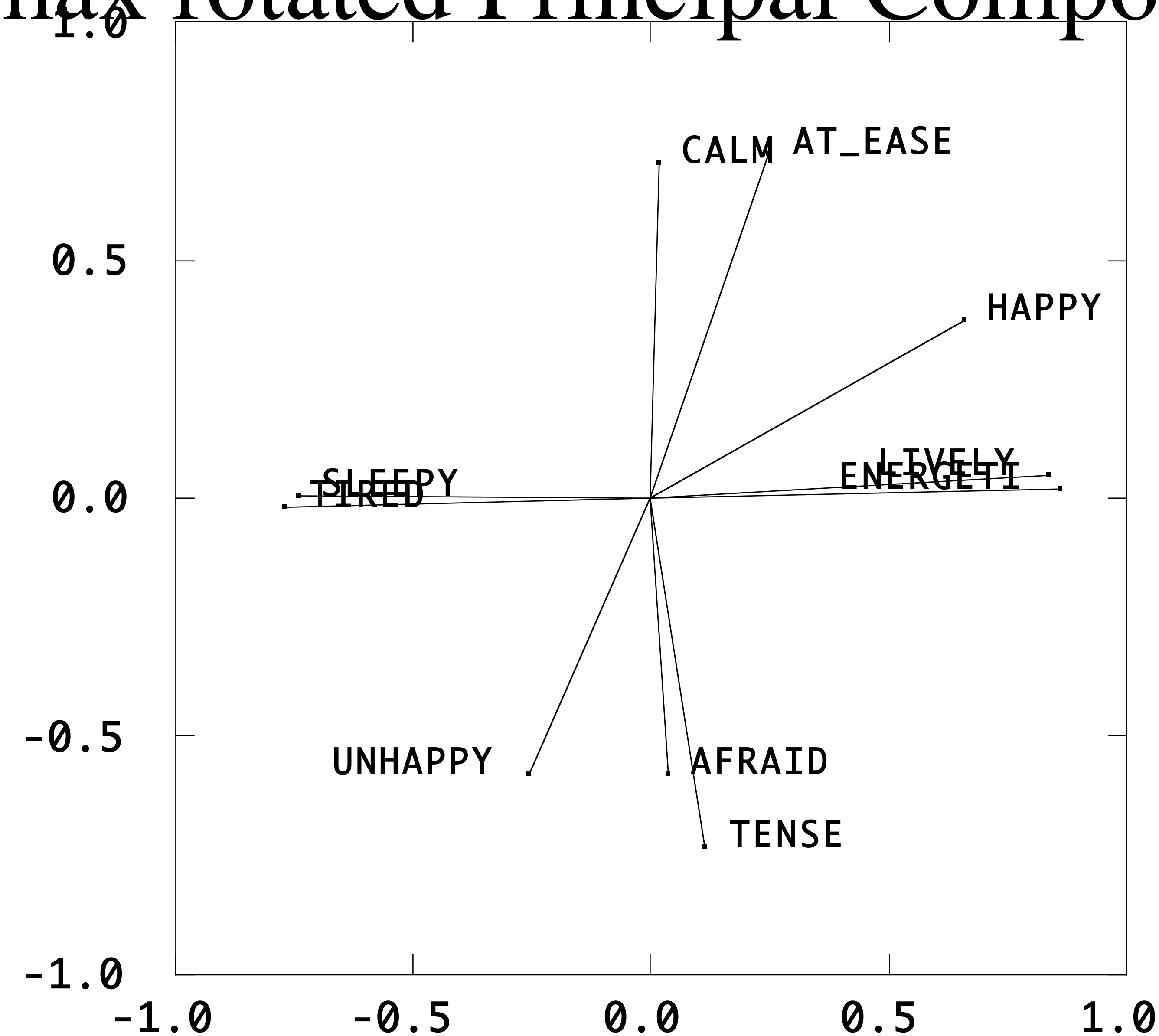
Unrotated and rotated PCs

	C1	C2	C1'	C2'
ENERGETI	-0.8	-0.4	0.9	0
LIVELY	-0.8	-0.3	0.8	0
HAPPY	-0.8	0	0.7	0.4
TIRED	0.7	0.3	-0.8	0
SLEEPY	0.7	0.3	-0.7	0
AT_EASE	-0.6	0.5	0.2	0.7
TENSE	0.2	-0.7	0.1	-0.7
CALM	-0.3	0.6	0	0.7
AFRAID	0.2	-0.5	0	-0.6
UNHAPPY	0.5	-0.4	-0.3	-0.6
eigen values	3.4	2.1	3.2	2.4

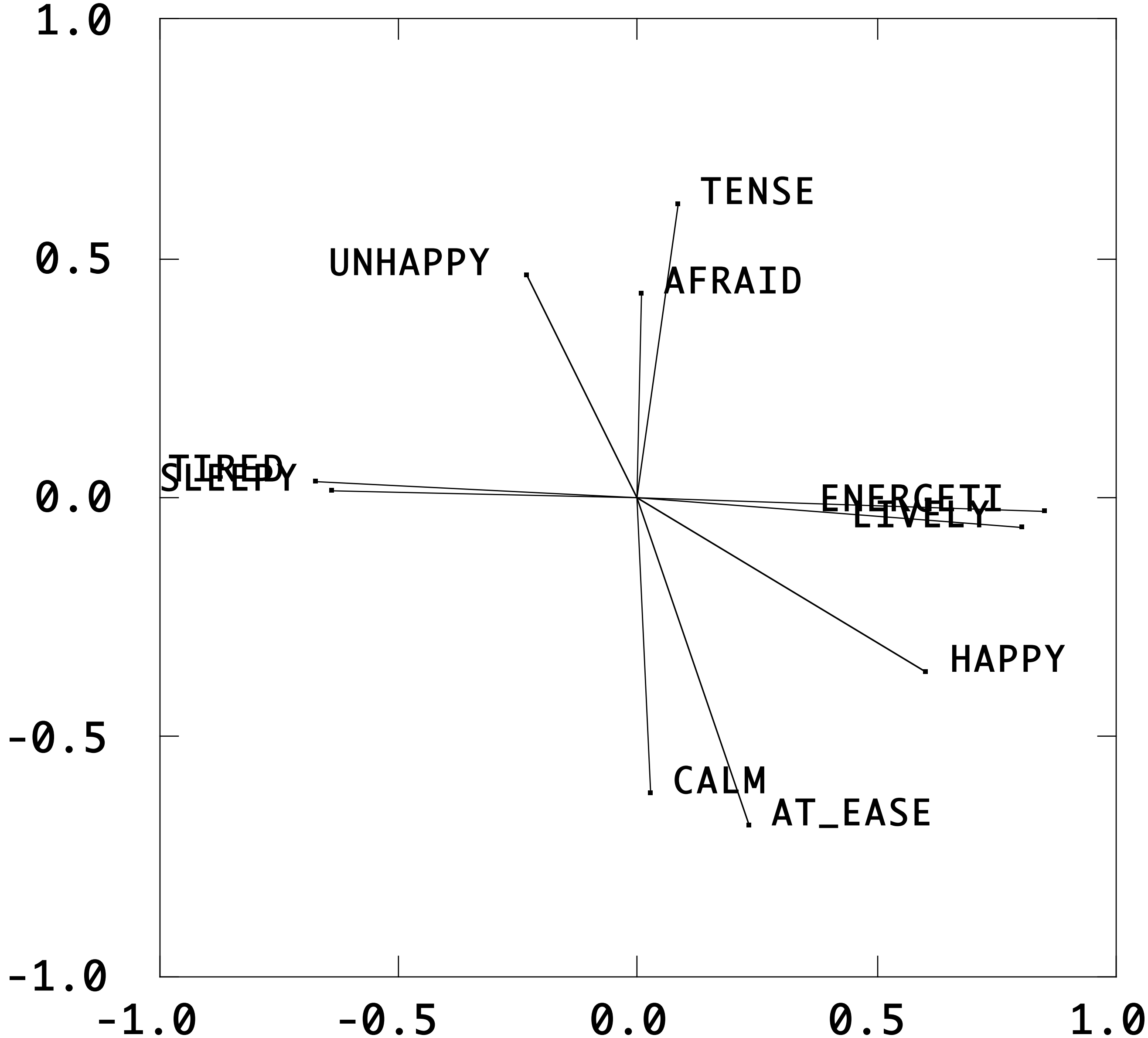
Varimax factors and components

	F1'	F2'	C1'	C2'
ENERGETI	0.8	0.0	0.9	0.0
LIVELY	0.8	-0.1	0.8	0.0
HAPPY	0.6	-0.4	0.7	0.4
TIRED	-0.7	0.0	-0.8	0.0
SLEEPY	-0.6	0.0	-0.7	0.0
AT_EASE	0.2	-0.7	0.2	0.7
TENSE	0.1	0.6	0.1	-0.7
CALM	0.0	-0.6	0.0	0.7
AFRAID	0.0	0.4	0.0	-0.6
UNHAPPY	-0.2	0.5	-0.3	-0.6
eigen values	2.7	1.8	3.2	2.4

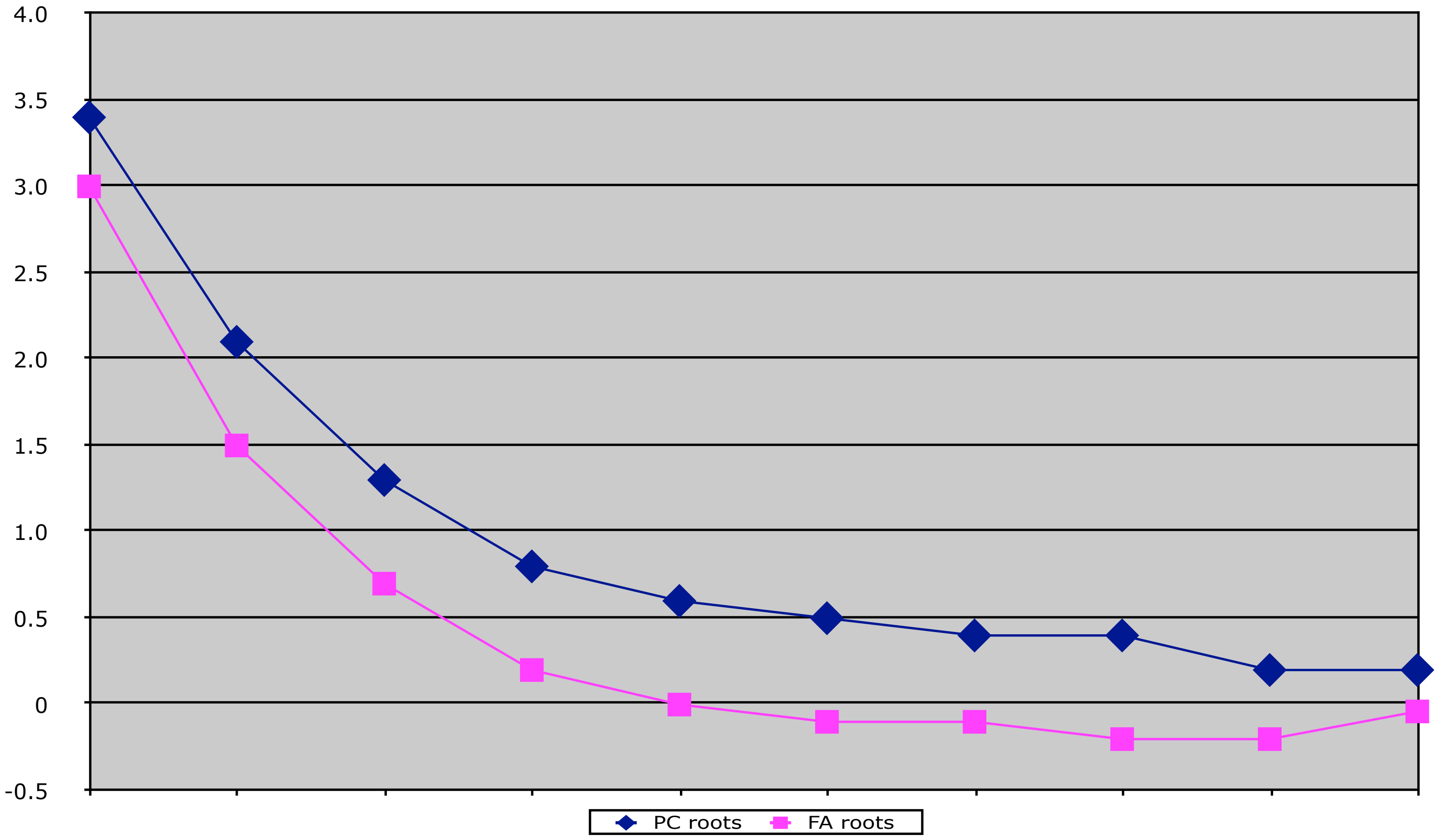
Varimax rotated Principal Components



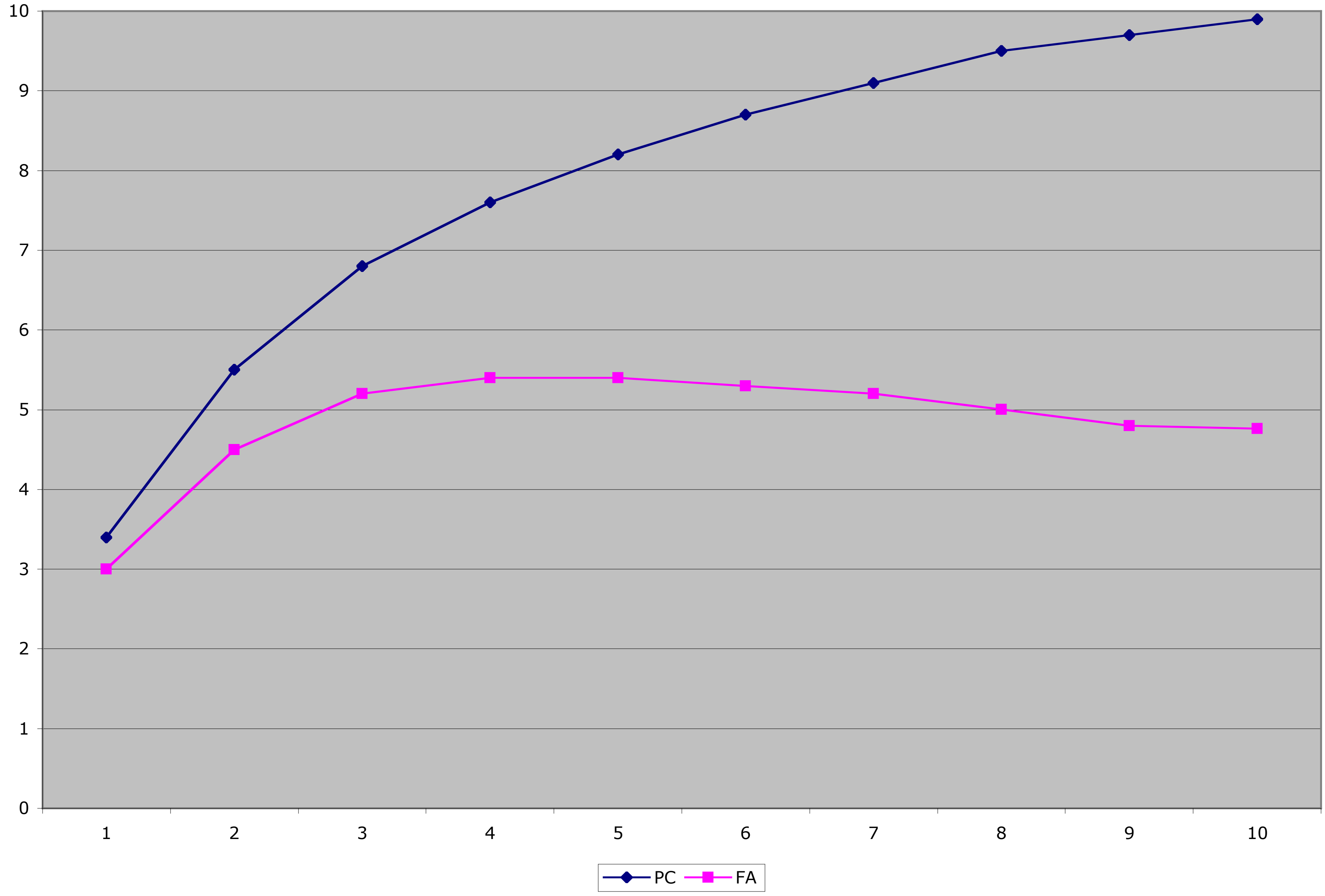
Varimax rotated factors of mood



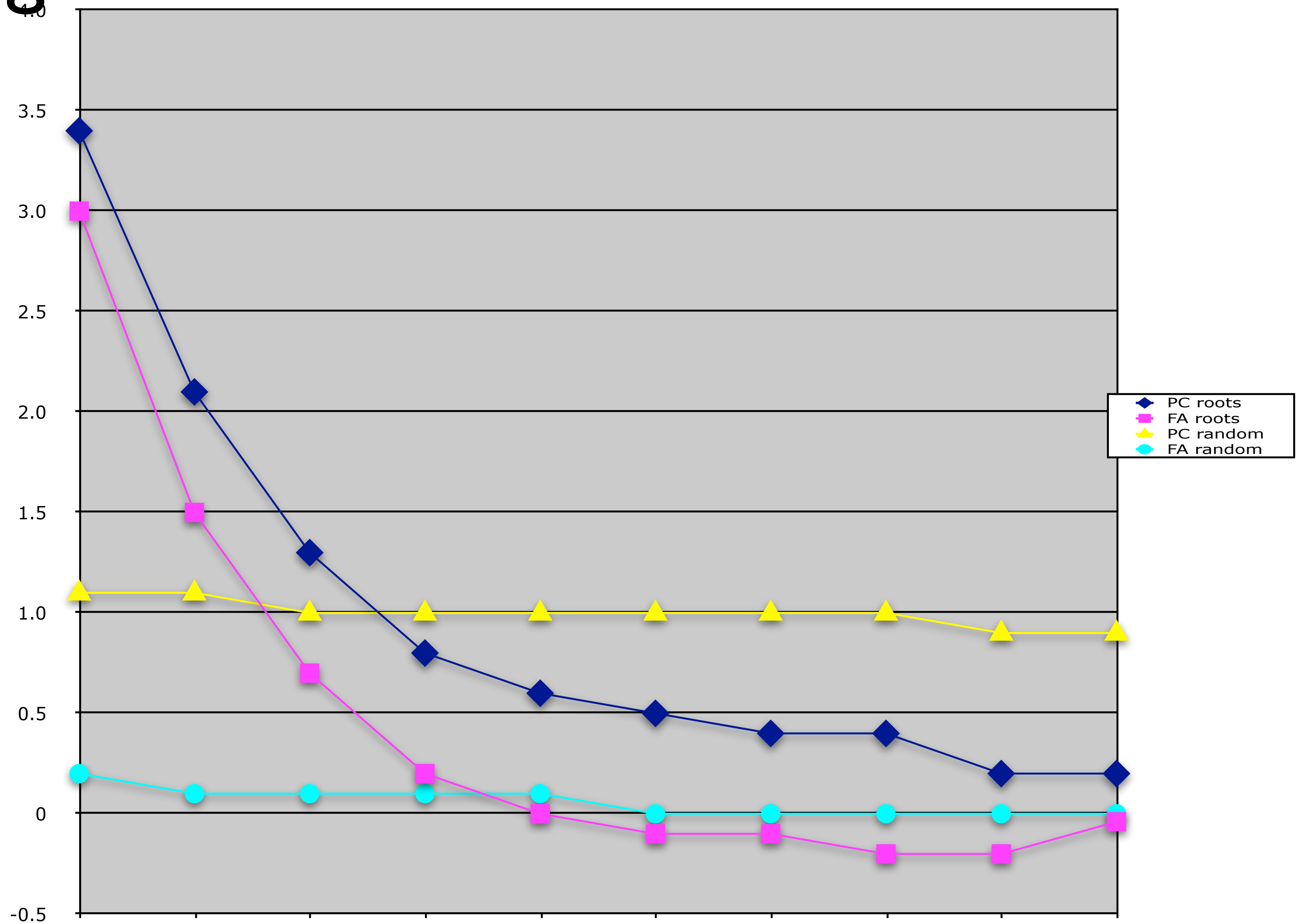
Eigenvalues and the scree test



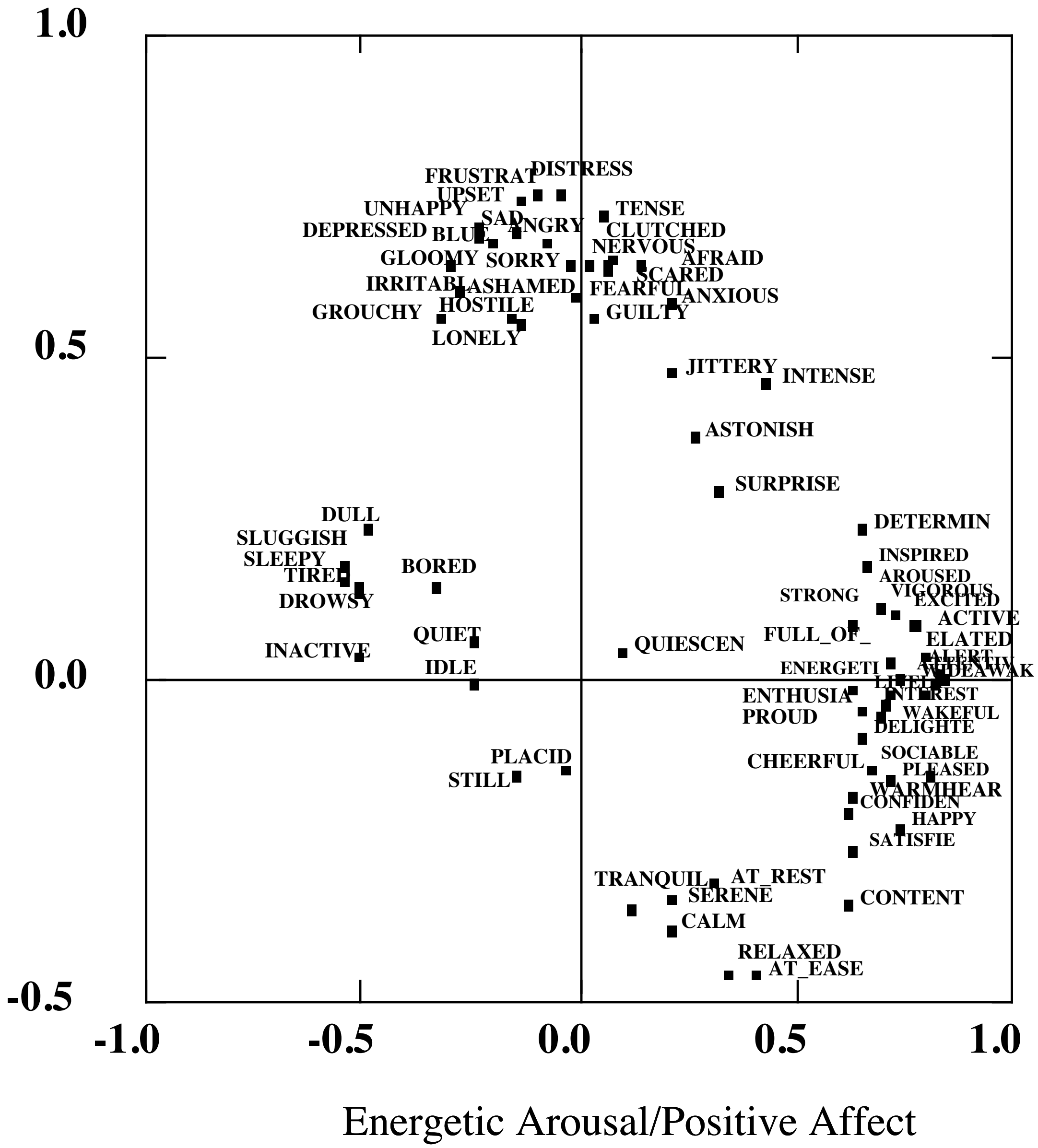
Cumulative variance explained



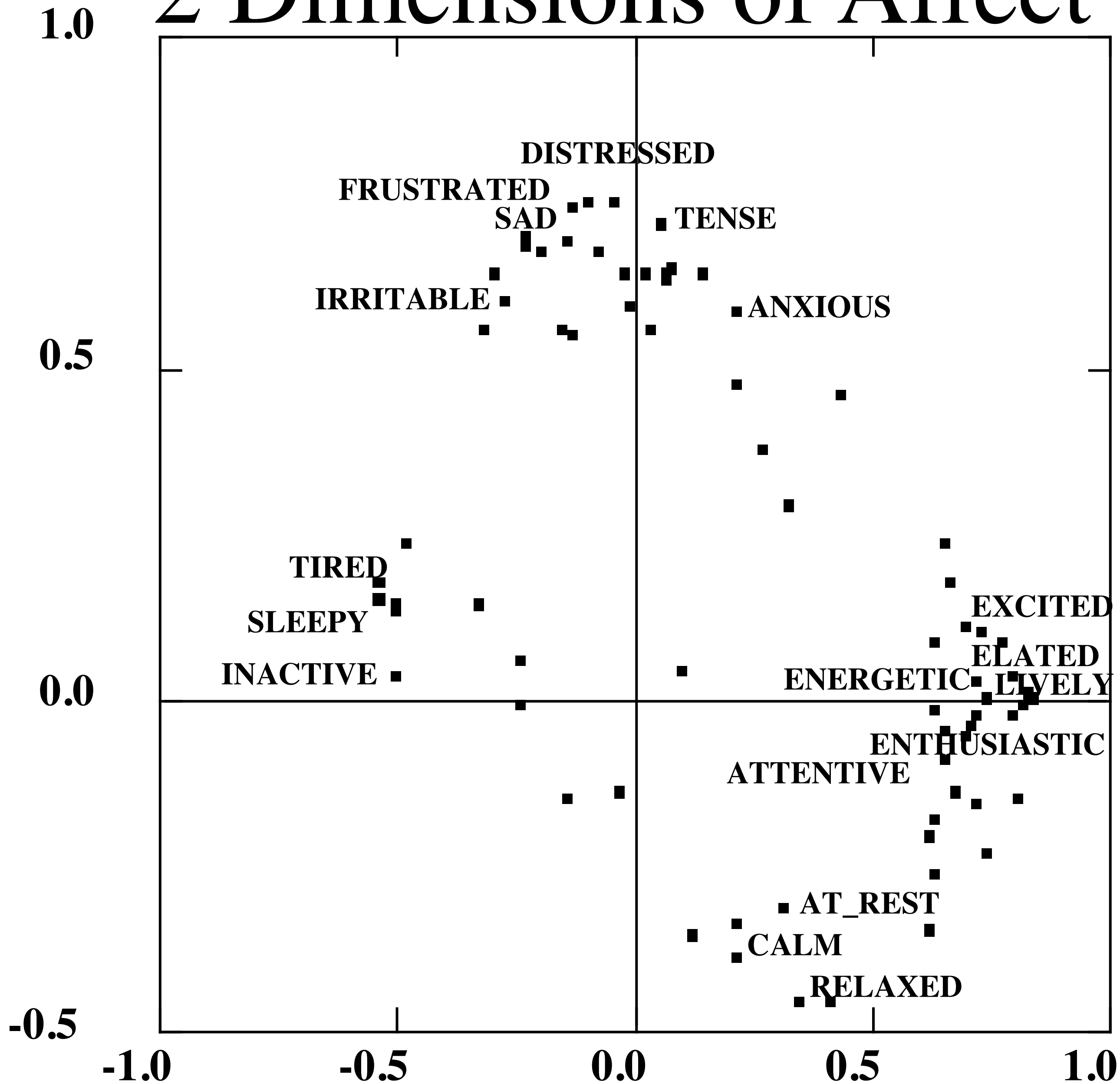
Eigen values for real vs. random data



2 Dimensions of Affect



2 Dimensions of Affect



Representative MSQ items (arranged by angular location)

Item	EA-PA	TA-NA	Angle
energetic	**0.8**	0.0	1
elated	**0.7**	0.0	2
excited	**0.8**	0.1	6
anxious	0.2	**0.6**	70
tense	0.1	**0.7**	85
distressed	0.0	**0.8**	93
frustrated	-0.1	**0.8**	98
sad	-0.1	**0.7**	101
irritable	-0.3	**0.6**	114
sleepy	**-0.5**	0.1	164
tired	**-0.5**	0.2	164
inactive	**-0.5**	0.0	177
calm	0.2	**-0.4**	298
relaxed	0.4	**-0.5**	307
at ease	0.4	**-0.5**	312
attentive	**0.7**	0.0	357
enthusiastic	**0.8**	0.0	358
lively	**0.9**	0.0	360

FA and PCA vocabulary

- Eigen values = $\sum(\text{loading}^2)$ across variables = amount of variance accounted for by factor
- Communality = $\sum(\text{loading}^2)$ across factors = amount of variance accounted for in a variable by all the factors
- Rotations versus Transformations
  - Rotations are orthogonal transformations
  - Oblique Transformations

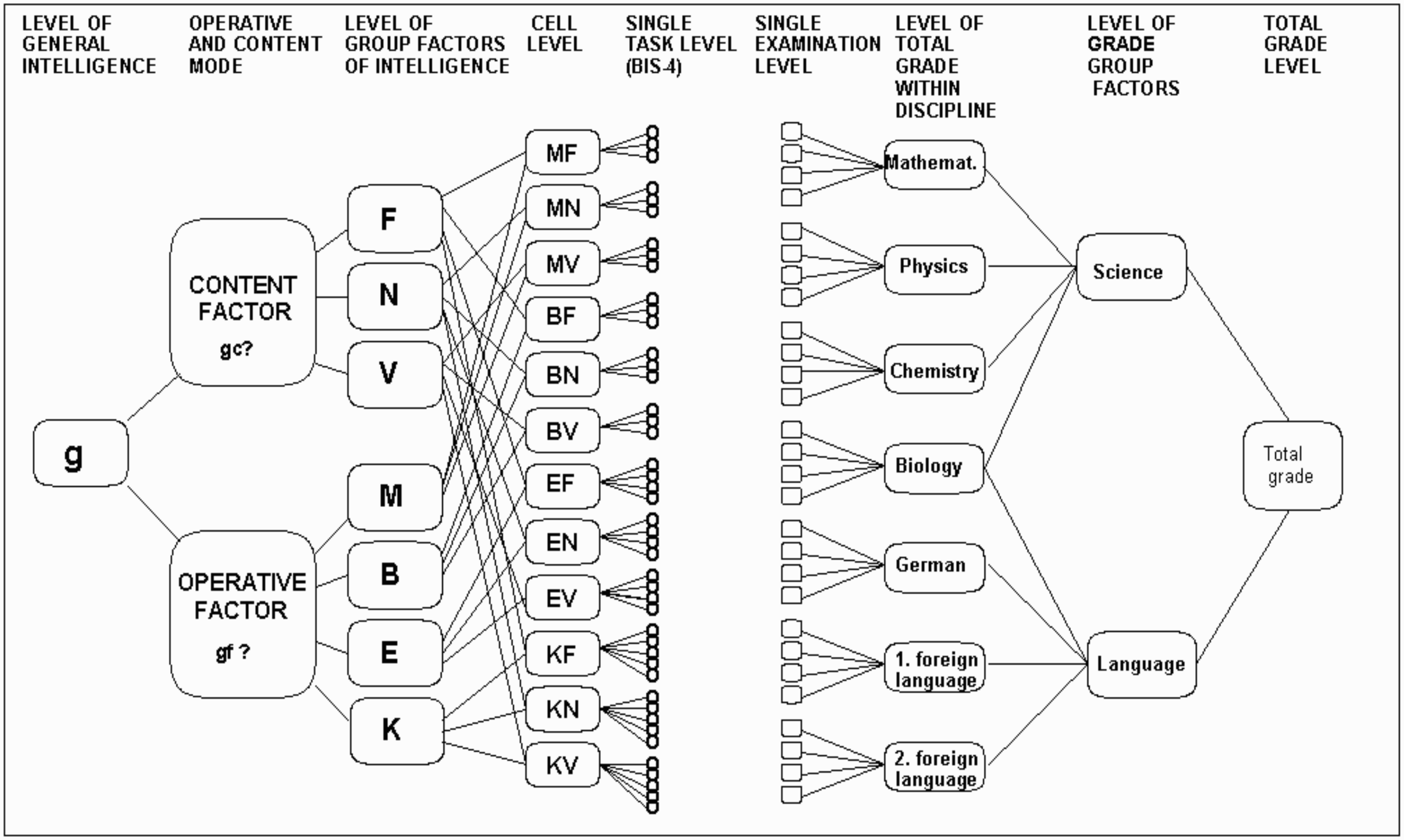
Rotations and transformations

- Simple structure as a criterion for rotation
- Simple structure in the eye of the beholder
- Simple factors (few high, many 0 loadings)
- Simple variables (few high, many 0 loadings)
- VARIMAX, Quartimax, Quartimin
- Procrustes

Rotations and transformations

- Orthogonal rotations
  - Factors are orthogonal, rotated to reduce (or maximize) particular definition of simple structure
- Oblique transformations and higher order factors
  - Allows factors to be correlated (and thus have higher order factors)

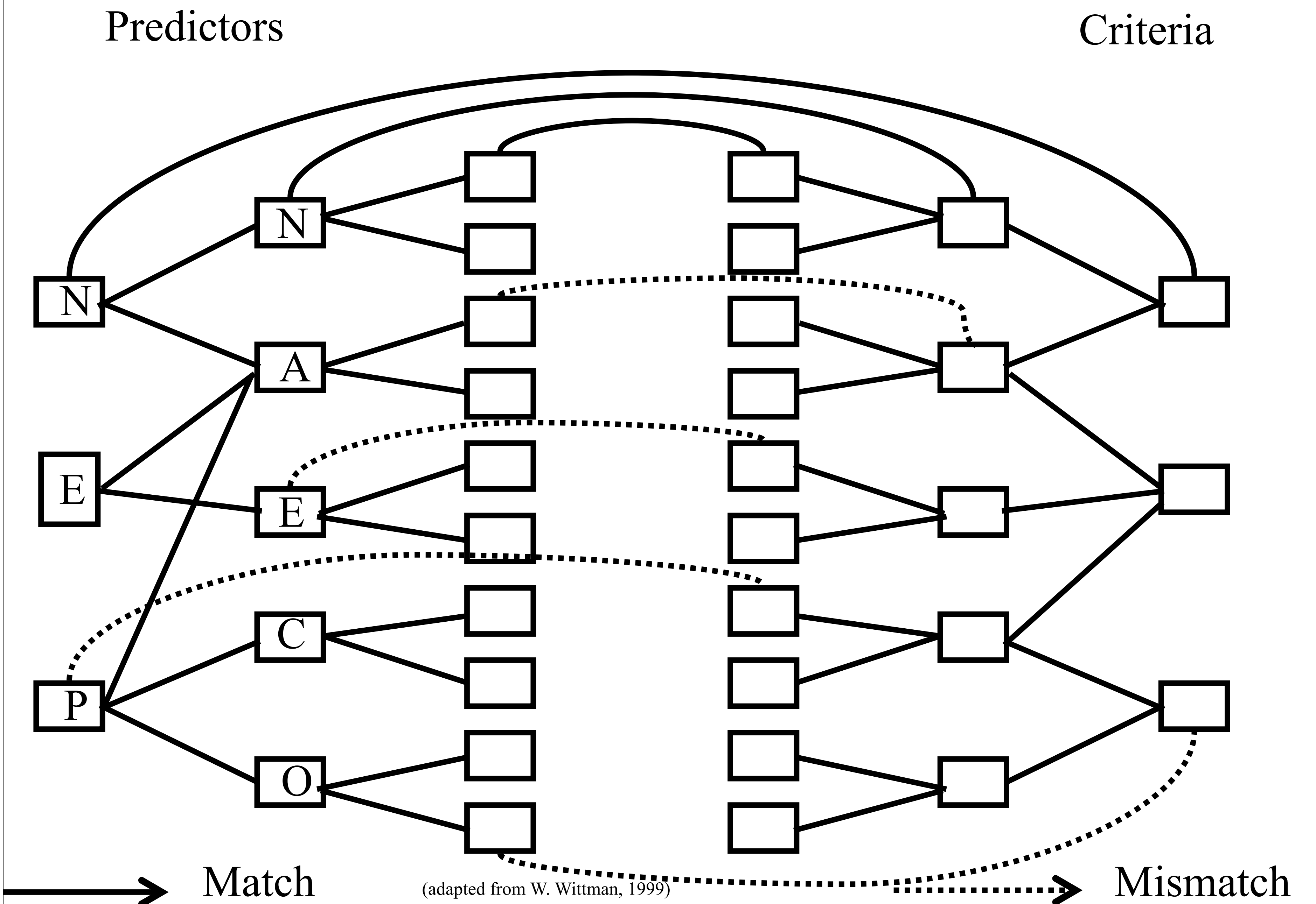
Fig. 9: **Hierarchical version of the Berlin model of intelligence and a grade hierarchy model**



K: Processing capacity for complex information, i.e. reasoning
E: Creativity
B: Speed on relatively simple tasks
M: Memory, i.e. storage capacity for information

F: figural Intelligence
N: numerical Intelligence
V: verbal intelligence

Specificity vs. generality --Matching predictors to outcomes



(adapted from W. Wittman, 1999)

Exploratory versus Confirmatory

- Exploratory:
  - How many factors and best rotation
  - Extraction
    - How many factors?
    - Algorithm for extraction
  - Rotation to simple structure -- what is best SS?
- Confirmatory: does a particular model fit?
  - Apply statistical test of fit
  - But larger N => less fit
  - Is the model the original one or has it been modified?

Components versus Factors

- Components are linear sums of variables and are thus defined at the data level
- Factors represent the covariances of variables and are only estimated (variables are linear sums of factors)
  - Model is undefined at data level, but is defined at structural level
  - Factor indeterminancy problem

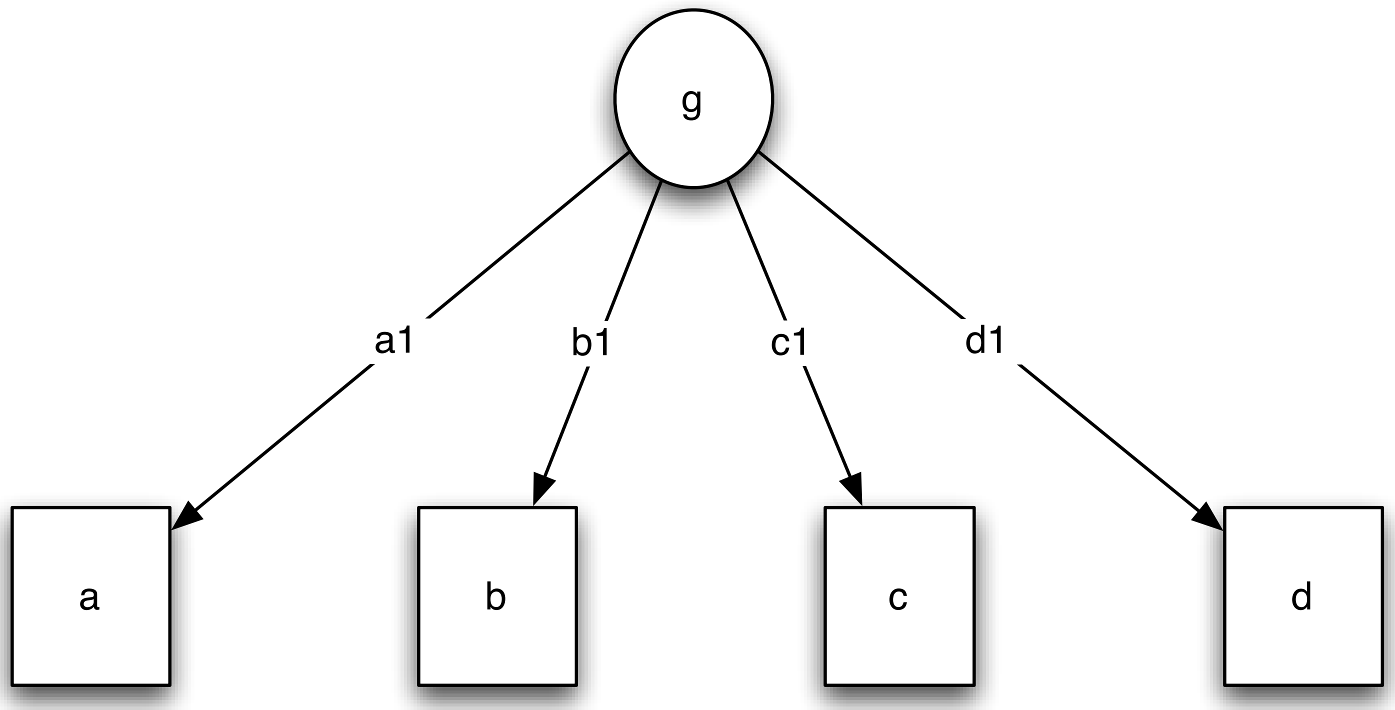
Structural Equation Models

- Estimating factor model of reliability (the measurement model
- Estimates of validity model corrected for attenuation (the structural model)
- Maximum likelihood solutions of covariance matrices

correlation matrix

	A	B	C	D
A	1.00	0.30	0.20	0.10
B	0.30	1.0	0.20	0.20
C	0.20	0.20	1.00	0.30
D	0.10	0.20	0.30	1.00

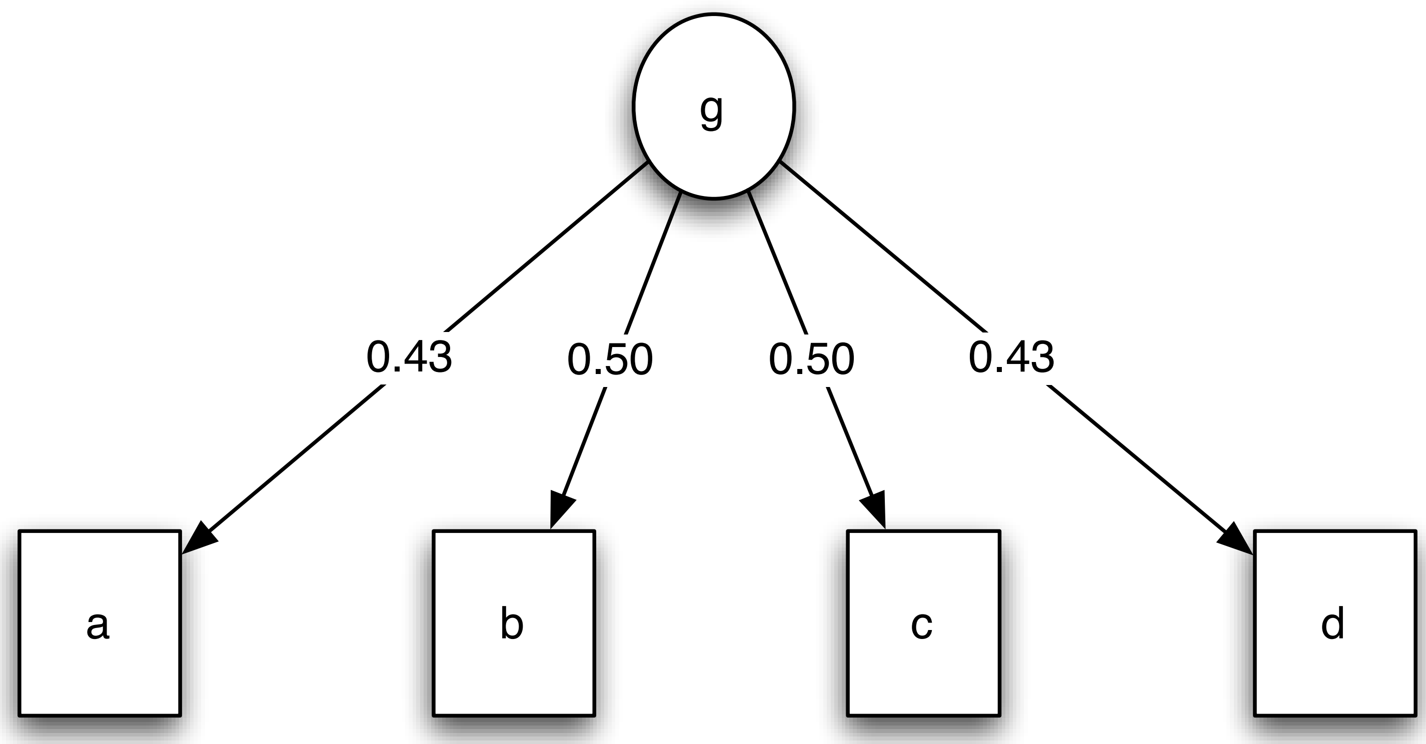
Hypothetical structure



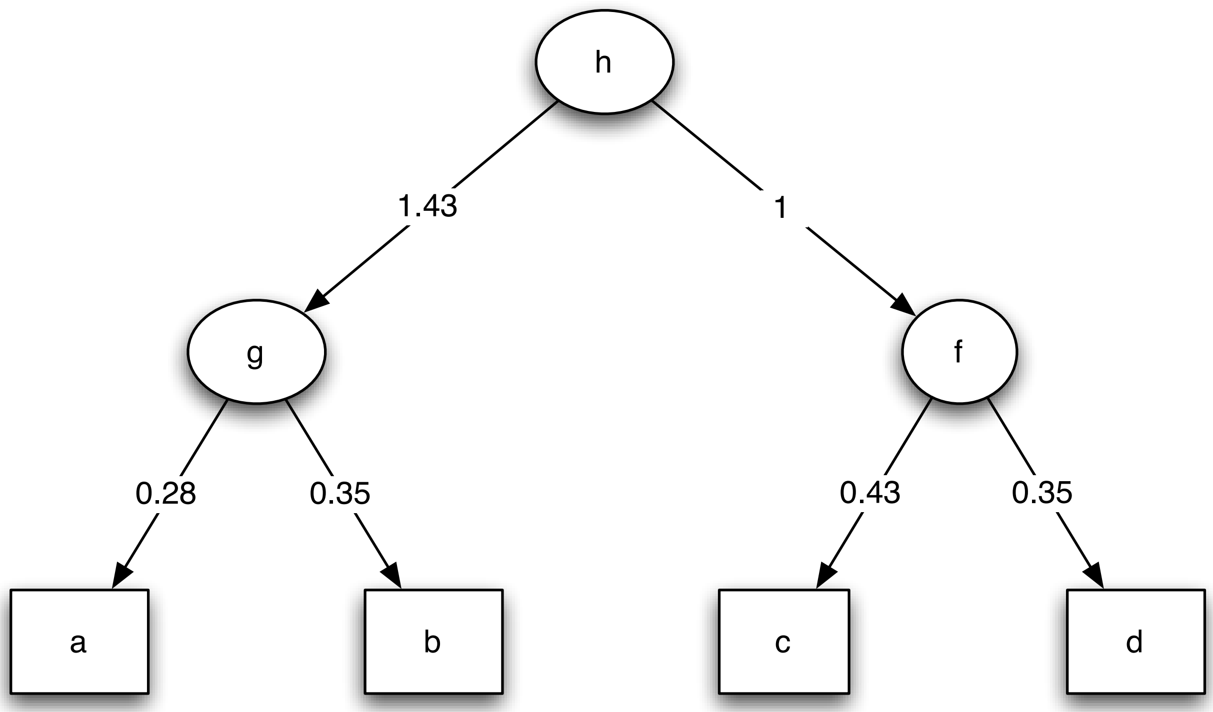
R for sem

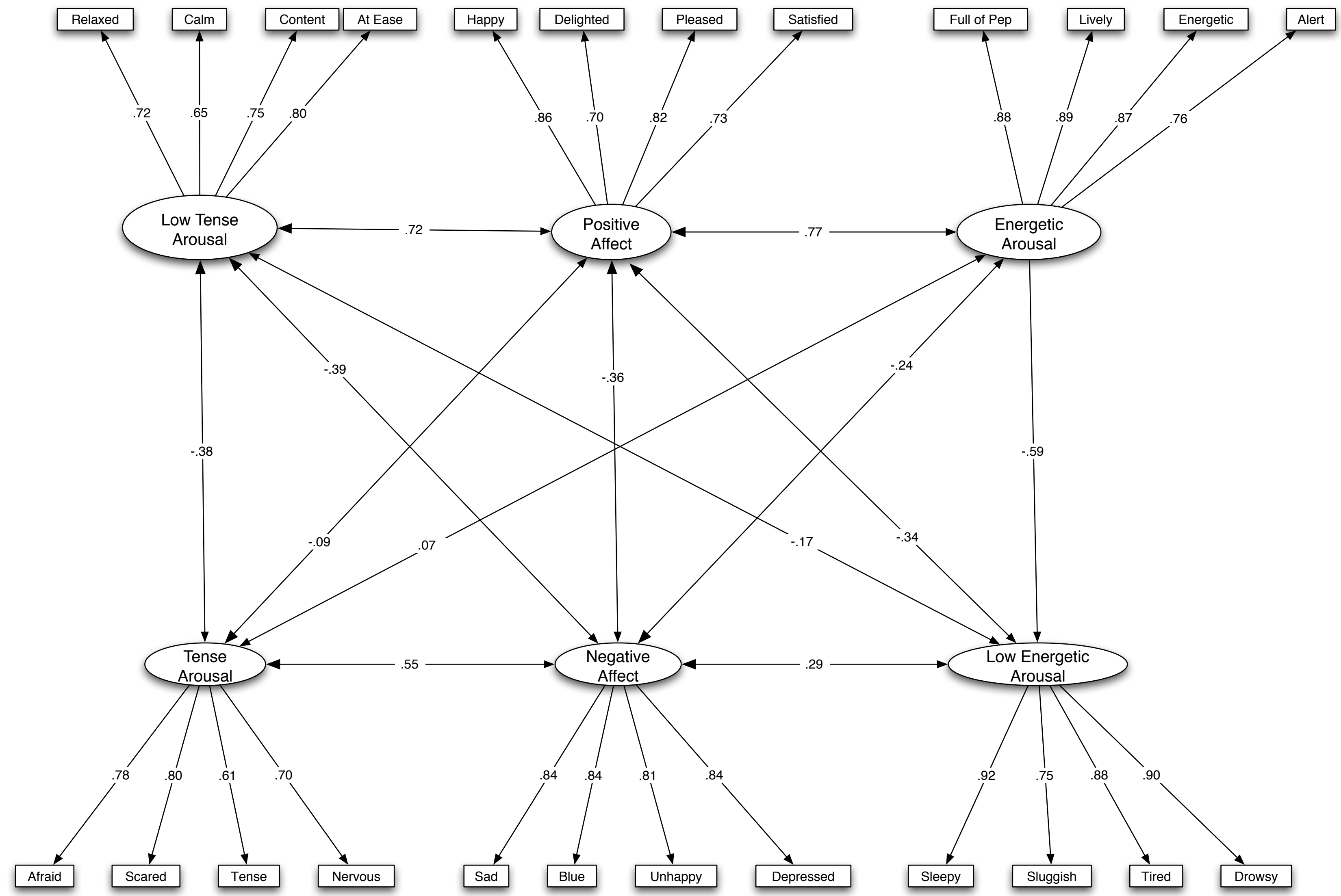
```
obs.var2.12 = c('a',  'b',  'c',  'd')
  R.prob2.12 = matrix(c(
  1.00 ,  .30,    .20,   .10,
   .30,  1.00,   .20,   .20,
   .20,   .20,  1.00,   .30,
   .10,   .20,   .30,  1.00),
   ncol=4,byrow=TRUE)

  model2.12a=matrix(c(
        'g ->  a',       'a1', NA,
        'g -> b' ,       'b1', NA,
        'g -> c' ,       'c1', NA,
        'g -> d',     'd1', NA,
        'a <-> a',       'e1', NA,
        'b <-> b',       'e2', NA,
        'c <-> c',       'e3', NA,
        'd <-> d',       'e4', NA,
        'g <-> g',        NA, 1),
        ncol=3, byrow=TRUE)
  sem2.12a= sem(model2.12a,R.prob2.12,120, obs.var2.12)
 summary(sem2.12a,digits=3)
```

SEM of Congeneric measures

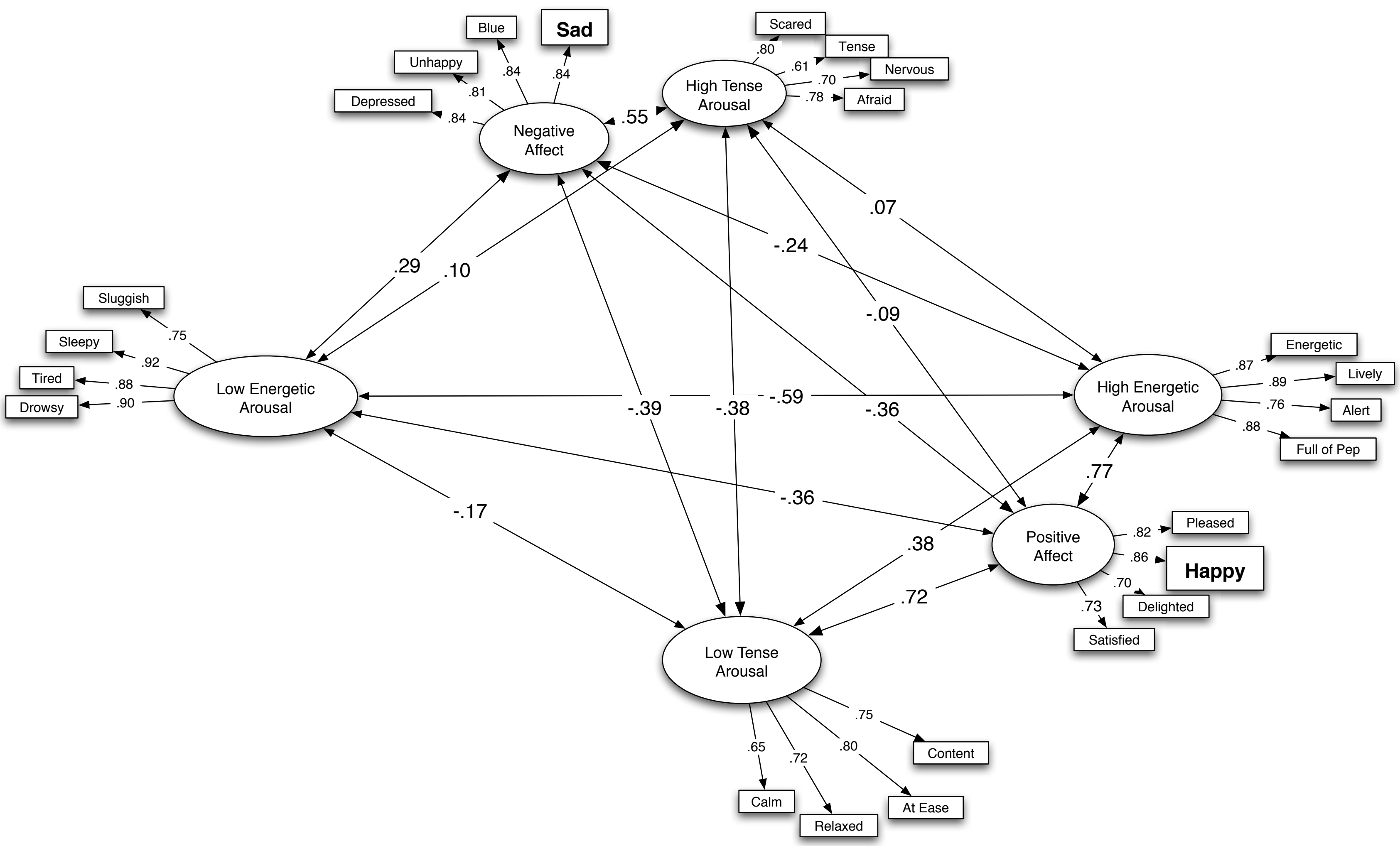


an alternative model




Relaxed
Calm
Content
At Ease
Happy
Delighted
Pleased
Satisfied
Full of Pep
Lively
Energetic
Alert
.72
.65
.75
.80
.86
.70
.82
.73
.88
.89
.87
.76
Low Tense Arousal
.72
Positive Affect
.77
Energetic Arousal
-.39
-.36
-.24
-.38
-.59
-.09
.07
-.17
-.34
Tense Arousal
.55
Negative Affect
.29
Low Energetic Arousal
.78
.80
.61
.70
.84
.84
.81
.84
.92
.75
.88
.90
Afraid
Scared
Tense
Nervous
Sad
Blue
Unhappy
Depressed
Sleepy
Sluggish
Tired
Drowsy

Structure of Affect



Methods of Scale Construction

- Empirical
  - MMPI, Strong
- Rational
  - CPI
- Theoretical
  - NAch
- Homogeneous

EPI, 16PF, NEO

Empirical Keying

- Ask items that discriminate known groups
  - People in general versus specific group
  - Choose items that are maximally independent and that have highest validities
- Example:
  - MMPI
  - Strong-Campbell
- Problem:
  - What is the meaning of the scale?
  - Need to develop new scale for every new group

Rational Keying

- Ask items with direct content relevance
- Example: California Psychological Inventory
- Problems
  - Not all items predict in obvious way
  - Need evidence for validity
  - Easy to fake

Theoretical Keying

- Ask items with theoretical relevance
- Example: Jackson Personality Research Form
- Problems:
  - Theoretical circularity
  - Need evidence for validity

Homogeneous Keying

- Select items to represent single domain
  - Exclude items based upon internal consistency
- Examples:
  - 16PF, EPI/EPQ, NEO
- Problems
  - Garbage In, Garbarge Out
  - Need evidence for validity

Methods of Homogeneous Keying

- Factor Analysis
- Principal Components Analysis
- Cluster Analysis