Exploratory Factor Analysis

- Exploratory Factor Analysis: Why and When?
- Underlying Conceptual/Mathematical Model
- Running an EFA

What is Factor Analysis?

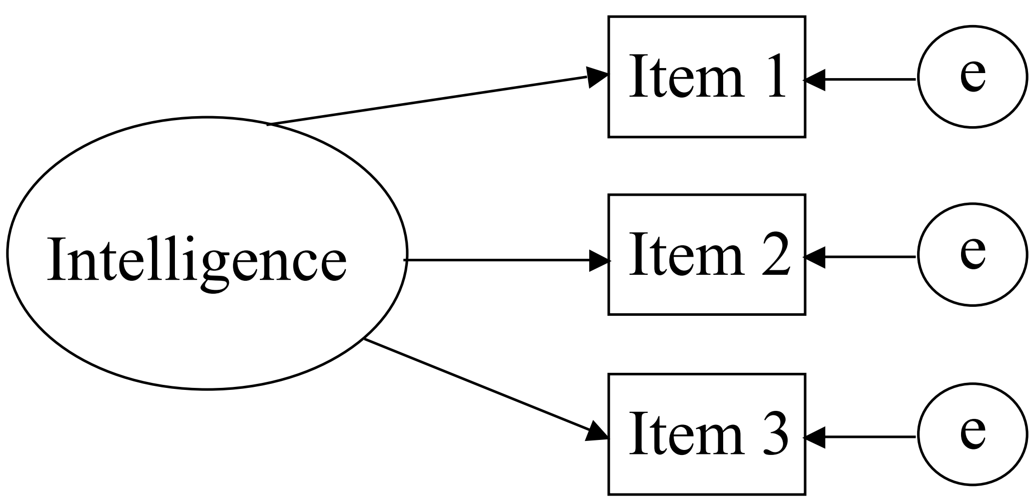
- Set of related techniques
  - » principal components analysis
  - » exploratory factor analysis
  - » confirmatory factor analysis
- Common objective: identify factors (new, hypothetical variables) or components that represent relationships among sets of variables
- Examples
  - » Personality/psychopathology (MMPI: 550 items represented as 10 scales)
  - » Social (RMA: 19 items, 1 factor)
  - » Developmental (MIDI: comprehension, language, fine motor, gross motor, personal-social)

Goals of Factor Analysis

- Data reduction: represent most of the variance in a set of variables using a smaller number of (hypothetical) variables
- Analyze associations (see which variables "hang together")
- Test hypotheses
  - » about dimensionality (e.g., are masculinity/femininity two constructs or two poles of one construct?)
  - » about measurement invariance (e.g., are sub-types of depression the same in different cultures?)
- Scale/test construction

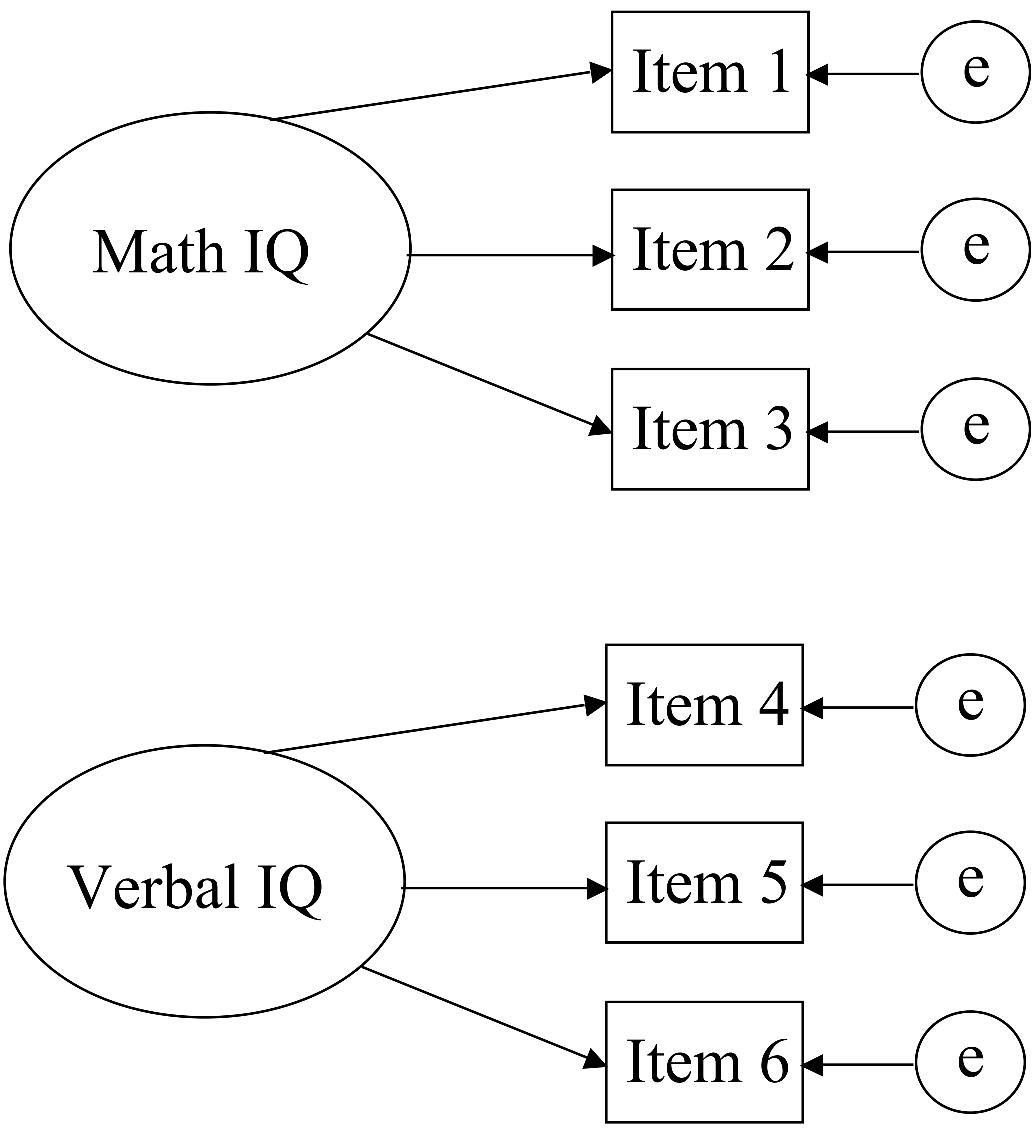
Conceptual Model

- Psychometric theory developed for research on intelligence testing
- "Intelligence" is the variable of interest, but it can't be measured directly
  - » "latent" or "unobserved" or "unmeasured"
- Responses on intelligence test (e.g., SAT) are "indicators" of intelligence
  - » "manifest" or "measured" variables
- Called the "common factor model"



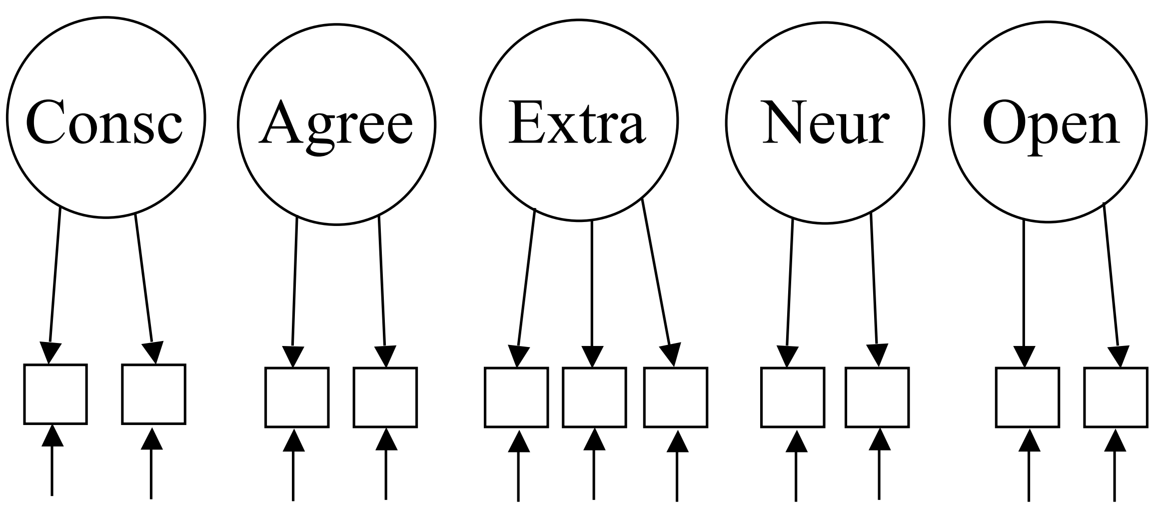
Multi-Factor Models

- Can easily generalize to more than one factor



Exploratory FA

- In exploratory FA, we typically don't know how many factors, or which items are indicators for which factor
- Example: trait theories of personality
  - » factor analysis of all adjectives in the lexicon that describe personality
- But, our underlying assumption is still that the factors cause the indicators to take on certain values



Example: Emotions

- 37 emotion adjectives
  - » "How much of this feeling are you experiencing right now?"
  - » 1-7 scale
- Don't want to have 37 IVs (or DVs)
- Can we create a smaller set of new variables that will capture most of the information in these 37 variables?

Correlation matrix

- We may be able to -- if there is some structure in the correlation matrix
- Sets of variables that correlate highly with each other, but much less so with other variables

Correlations

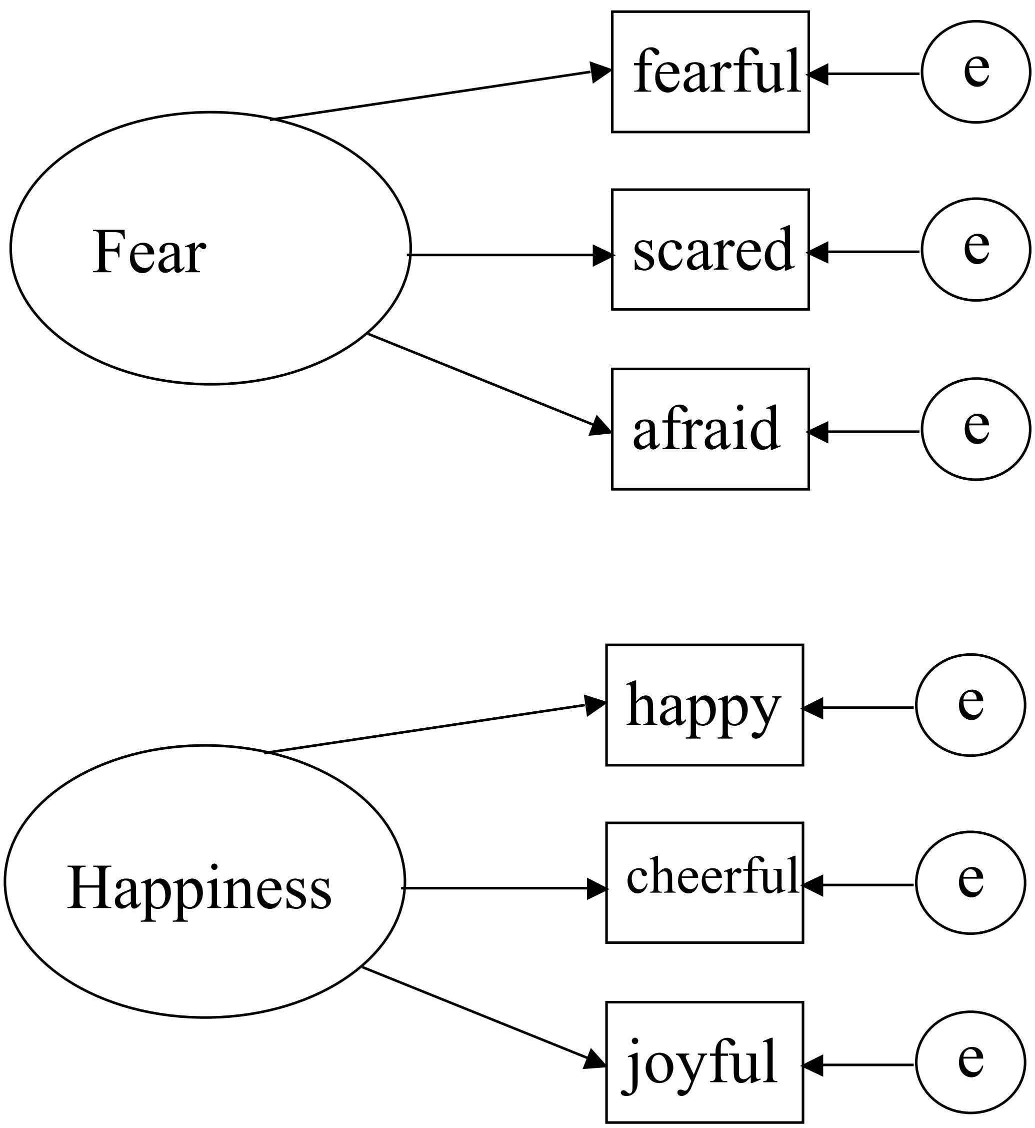
		POWEM9 fearful	POWEM28 scared	POWEM32 afraid	POWEM31 happy	POWEM34 cheerful	POWEM37 joyful
fearful	Pearson Correlation	1	.750**	.726**	-.238*	-.142	-.130
	Sig. (2-tailed)	.	.000	.000	.027	.192	.233
	N	87	87	87	86	86	86
scared	Pearson Correlation	.750**	1	.840**	-.196	-.184	-.026
	Sig. (2-tailed)	.000	.	.000	.070	.090	.811
	N	87	87	87	86	86	86
afraid	Pearson Correlation	.726**	.840**	1	-.167	-.109	-.056
	Sig. (2-tailed)	.000	.000	.	.125	.319	.611
	N	87	87	87	86	86	86
happy	Pearson Correlation	-.238*	-.196	-.167	1	.815**	.733**
	Sig. (2-tailed)	.027	.070	.125	.	.000	.000
	N	86	86	86	86	86	86
cheerful	Pearson Correlation	-.142	-.184	-.109	.815**	1	.690**
	Sig. (2-tailed)	.192	.090	.319	.000	.	.000
	N	86	86	86	86	86	86
joyful	Pearson Correlation	-.130	-.026	-.056	.733**	.690**	1
	Sig. (2-tailed)	.233	.811	.611	.000	.000	.
	N	86	86	86	86	86	86

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Common Factor Model of Emotions

- Underlying structure assumed



Steps in EFA

- Selecting variables/items
- Preparing/checking correlation matrix
- **Extracting factors**
- **Determining the number of factors**
- **Rotating factors**
- **Interpreting results**
- Verify structure by establishing construct validity

Extracting Factors

- Variable is a linear combination of factors
  - » e.g., fearful = B1*fear + B2*Happiness + $U_{fearful}$
- Want linear combinations that will account for as much of the variance in sample as possible
  - » in output, everything is standardized
  - » variance of each variable = 1
  - » so total variance = number of variables
  - » here, total variance = 6.0

Extracting Factors

- Goal of factor extraction is to determine the factors
- Factors are estimated as linear combinations of variables
  - » e.g. Fear = B1*fearful + B2*scared + B3*afraid + B4*happy + B5*cheerful + B6*joyful
  - » hopefully, only a few of these coefficients will be large
  - » e.g., B1, B2, B3 large; B4, B5, B6 close to zero
- Variety of methods for estimation
- Several of the most popular try to maximize the variance explained at each step

Principal Components Analysis

- First factor extracted in such a way as to explain the maximum amount of variance
- Second factor explains the maximum amount of the variance that is left
  - » must be orthogonal to first factor because it's trying to explain the residual variance -- what doesn't overlap with the just-extracted Factor 1
- Linear function (or principal component) is represented as an *eigenvector*
  - » vector of numbers; numbers = coefficients in the linear equation
- Variance explained by that linear combination is the *eigenvalue*

PCA of Emotions: 1st component

Component Matrix[a]

	Component					
	1	2	3	4	5	6
POWEM9 fearful	.727	.516	.294	.335	5.319E-02	6.707E-02
POWEM28 scared	.733	.589	-.189	-6.609E-02	.113	-.250
POWEM32 afraid	.707	.605	-4.427E-02	-.281	-.134	.188
POWEM31 happy	-.739	.574	8.067E-02	-.130	.307	8.564E-02
POWEM34 cheerful	-.686	.607	.304	-7.668E-02	-.204	-.145
POWEM37 joyful	-.601	.655	-.373	.240	-9.954E-02	5.792E-02

Extraction Method: Principal Component Analysis.

a. 6 components extracted.

- PC1 = .727*fearful + .733*scared + .707*afraid - .739*happy - .686*cheeful - .601*joyful
- Variance explained = $.727^2 + .733^2 + .707^2 + (-.739)^2 + (-.686)^2 + (-.601)^2 = 2.94$
- % variance explained = 2.94/6.0 = 49.1%

PCA of Emotions: 2nd component

Component Matrix[a]

	Component					
	1	2	3	4	5	6
POWEM9 fearful	.727	.516	.294	.335	5.319E-02	6.707E-02
POWEM28 scared	.733	.589	-.189	-6.609E-02	.113	-.250
POWEM32 afraid	.707	.605	-4.427E-02	-.281	-.134	.188
POWEM31 happy	-.739	.574	8.067E-02	-.130	.307	8.564E-02
POWEM34 cheerful	-.686	.607	.304	-7.668E-02	-.204	-.145
POWEM37 joyful	-.601	.655	-.373	.240	-9.954E-02	5.792E-02

Extraction Method: Principal Component Analysis.

a. 6 components extracted.

- PC2 = .516*fearful + .589*scared + .605*afraid + .574*happy + .607*cheeful + .655*joyful
- Variance explained = $.516^2 + .589^2 + .605^2 + .574^2 + .607^2 + .655^2 = 2.106$
- % variance explained = 2.106/6.0 = 35.1%

Table of Eigenvalues

- cf. SPSS summary of eigenvalues

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.942	49.033	49.033	2.942	49.033	49.033
2	2.106	35.102	84.135	2.106	35.102	84.135
3	.362	6.035	90.170	.362	6.035	90.170
4	.276	4.599	94.768	.276	4.599	94.768
5	.179	2.991	97.759	.179	2.991	97.759
6	.134	2.241	100.000	.134	2.241	100.000

Extraction Method: Principal Component Analysis.

Extraction and Parsimony

- Note that if we continue to extract components, we will eventually explain all of the variance
- However, we will have gained no parsimony
  - » we now have 6 components instead of 6 variables
- But the six components are uncorrelated
  - » this is sometimes useful
  - » eliminate multicollinearity, confounding
- Usually, though we want to reduce the number of variables
  - » SPSS menu label for factor is "Data Reduction"

Number of Components

- Recall that each variable has a variance of 1.0
- Thus, explaining one unit of variance doesn't "buy" us anything -- we could do this well just by using a variable
- May be reasonable to extract only those components that do better than this, in explaining variance
  - » "Kaiser method"
- Here, 2 components do so

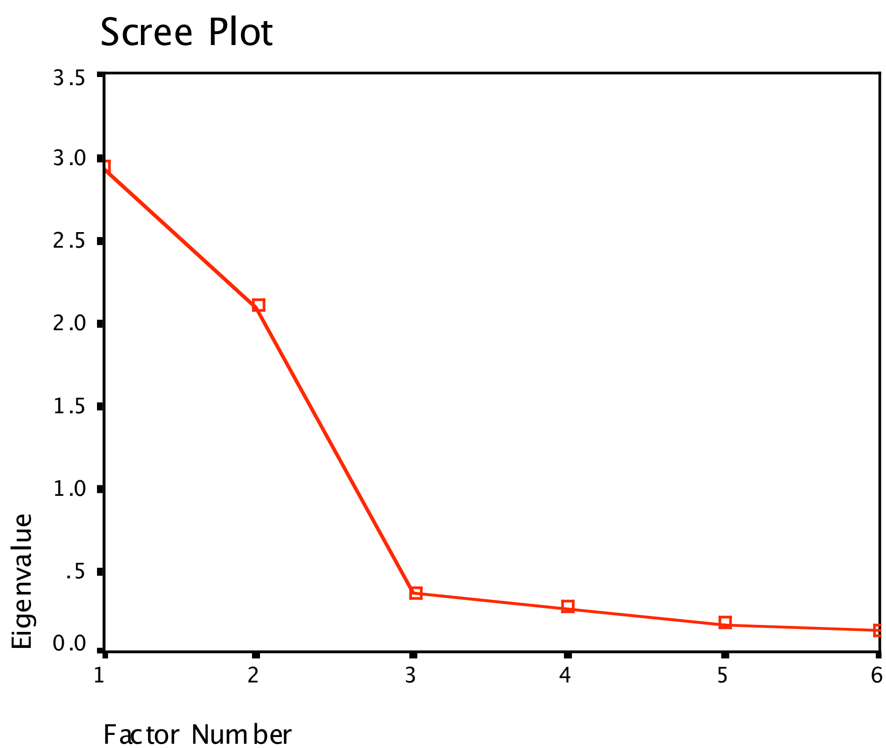
Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.942	49.033	49.033	2.942	49.033	49.033
2	2.106	35.102	84.135	2.106	35.102	84.135
3	.362	6.035	90.170	.362	6.035	90.170
4	.276	4.599	94.768	.276	4.599	94.768
5	.179	2.991	97.759	.179	2.991	97.759
6	.134	2.241	100.000	.134	2.241	100.000

Extraction Method: Principal Component Analysis.

Scree Plot

- Another method (from Cattell) is to look for the bend in a "scree plot"
- Plots eigenvalues on Y axis, from biggest to smallest



Output for 2 Components

- Request extraction of 2 components
  - » Output very similar, but now we can only approximate scores on the variables, we cannot reproduce them exactly
  - » That's ok -- we're still explaining 84% of the variance, and more parsimoniously

Component Matrix[a]

	Component	
	1	2
POWEM9 fearful	.727	.516
POWEM28 scared	.733	.589
POWEM32 afraid	.707	.605
POWEM31 happy	-.739	.574
POWEM34 cheerful	-.686	.607
POWEM37 joyful	-.601	.655

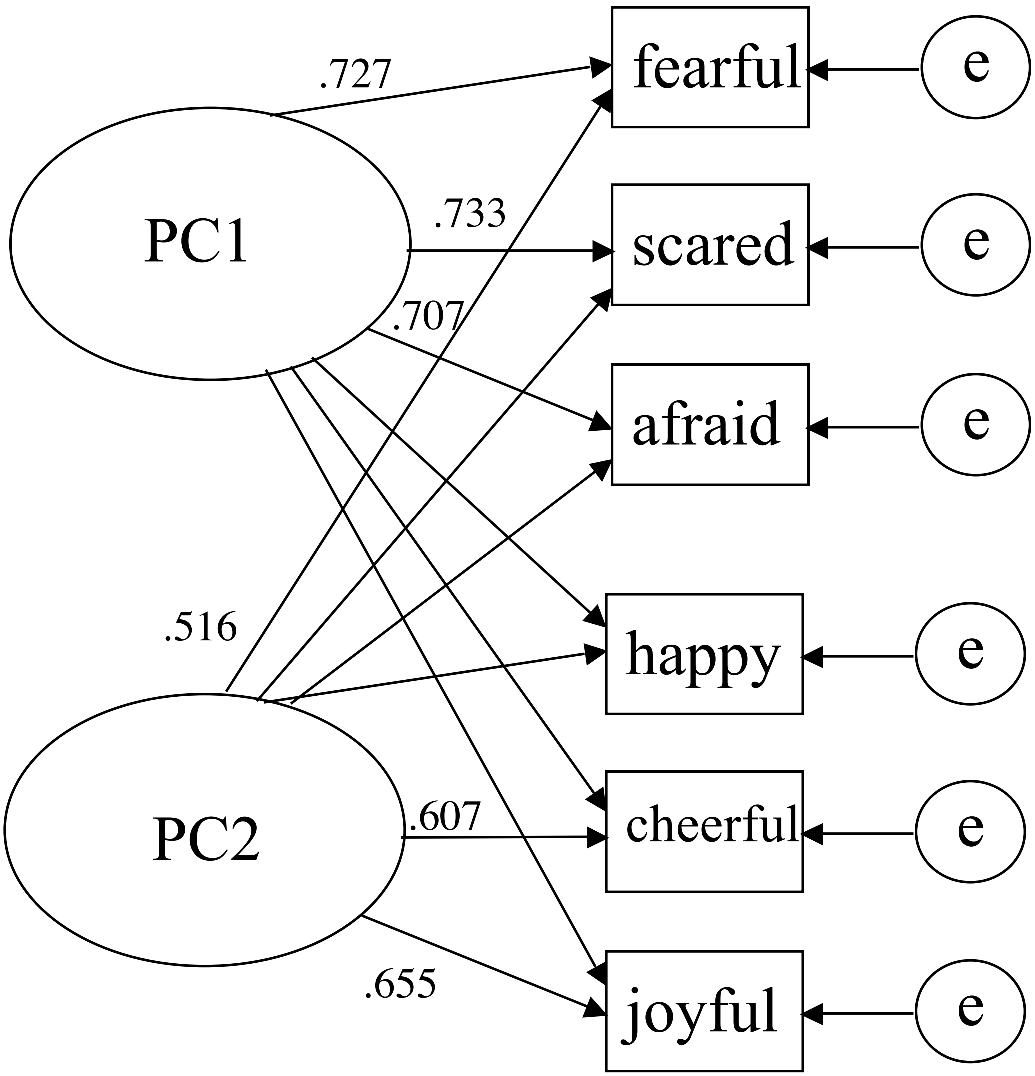
Extraction Method: Principal Component Analysis.
a. 2 components extracted.

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.942	49.033	49.033	2.942	49.033	49.033
2	2.106	35.102	84.135	2.106	35.102	84.135
3	.362	6.035	90.170			
4	.276	4.599	94.768			
5	.179	2.991	97.759			
6	.134	2.241	100.000			

Extraction Method: Principal Component Analysis.

Interpreting Components



Component Matrix[a]

	Component	
	1	2
POWEM9 fearful	.727	.516
POWEM28 scared	.733	.589
POWEM32 afraid	.707	.605
POWEM31 happy	-.739	.574
POWEM34 cheerful	-.686	.607
POWEM37 joyful	-.601	.655

Extraction Method: Principal Component Analysis.
a. 2 components extracted.

What do the components mean?

- If we tried to interpret the PC eigenvectors, we might say
  - » PC1 is negative affect
    - because high weights for negative emotions and low weights for positive emotions
  - » PC2 is general emotionaly
    - because moderately high weights for everything

Component Matrix[a]

	Component	
	1	2
POWEM9 fearful	.727	.516
POWEM28 scared	.733	.589
POWEM32 afraid	.707	.605
POWEM31 happy	-.739	.574
POWEM34 cheerful	-.686	.607
POWEM37 joyful	-.601	.655

Extraction Method: Principal Component Analysis.

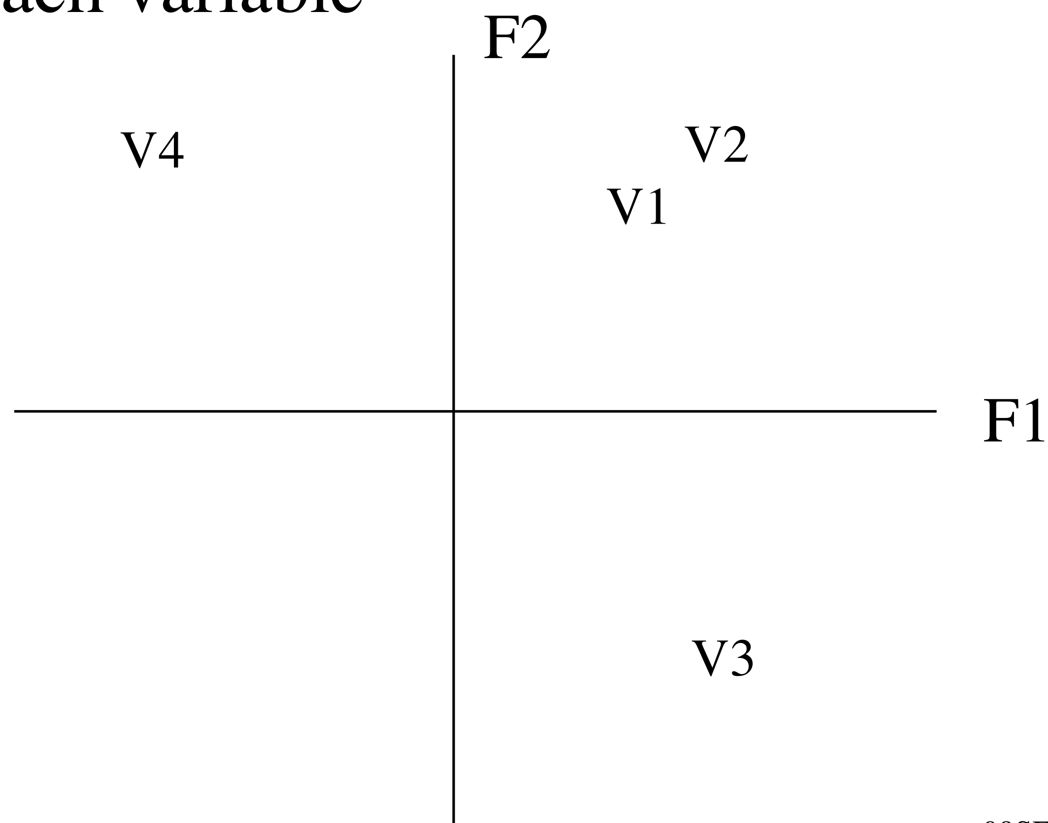
a. 2 components extracted.

What do the components mean?

- But it doesn't make sense to interpret these components
- Infinite number of equivalent sets of eigenvectors
- These particular ones are a result of our extraction strategy (i.e., maximize variance explained) and are in some sense arbitrary
- More meaningful alternatives exist
- These can be found via rotation

Rotation

- General idea: Make factors more interpretable
- Ideal: Each variable has high loading on one factor; negligible loadings on other factors
  - » "simple structure"
- To visualize, plot factor loadings for each variable



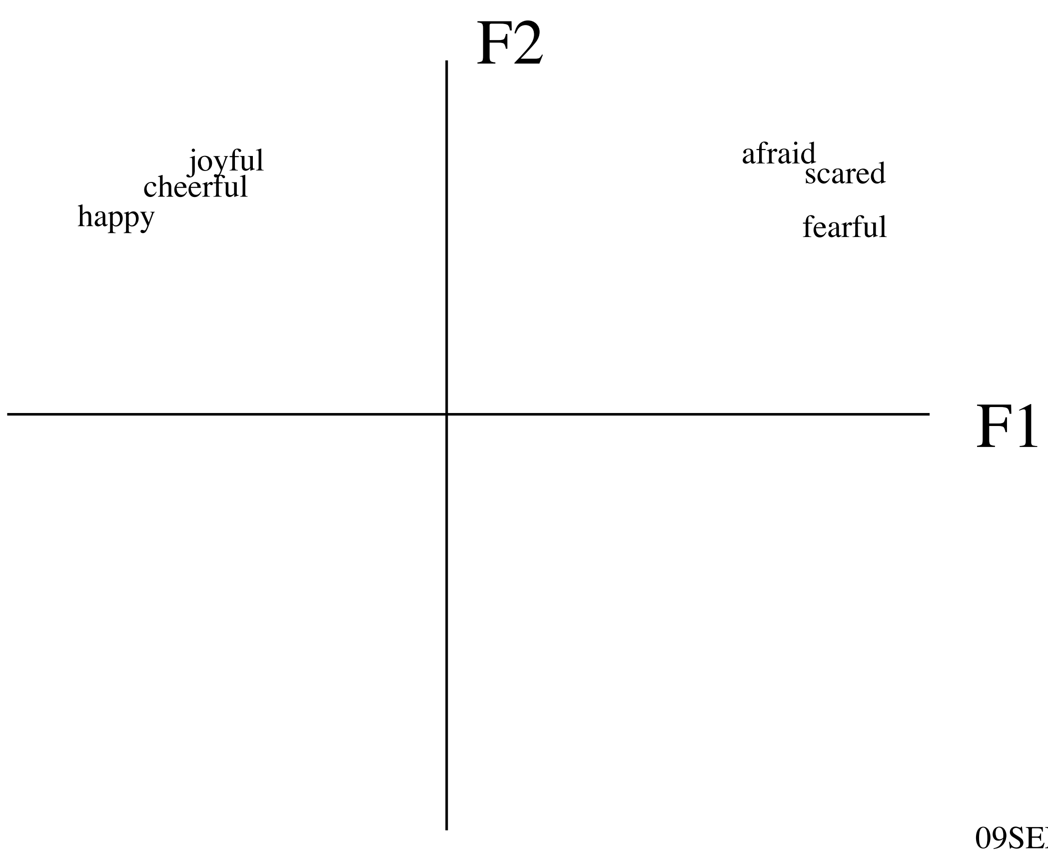
Rotation - Emotions

- We don't have simple structure

Component Matrix[a]

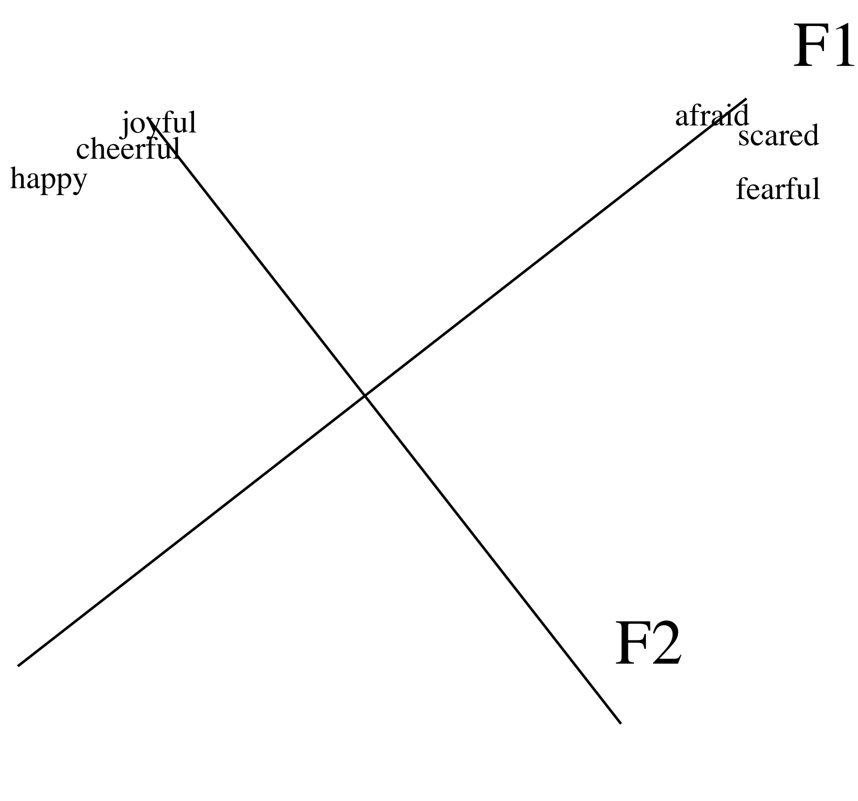
	Component	
	1	2
POWEM9 fearful	.727	.516
POWEM28 scared	.733	.589
POWEM32 afraid	.707	.605
POWEM31 happy	-.739	.574
POWEM34 cheerful	-.686	.607
POWEM37 joyful	-.601	.655

Extraction Method: Principal Component Analysis.
a. 2 components extracted.



Rotation - Emotions

- But we can obtain it by rotating the axes
- Now, F1 = fear and F2 = happiness
- We have simple structure
- Factors are interpretable



SPSS output

Rotated Factor Matrix[a]

	Factor	
	1	2
POWEM9 fearful	.796	-.121
POWEM28 scared	.931	-6.938E-02
POWEM32 afraid	.899	-4.178E-02
POWEM31 happy	-.147	.922
POWEM34 cheerful	-9.116E-02	.867
POWEM37 joyful	-8.549E-03	.795

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.