Consider an investigation into the nature of intelligence.

ability to recite song lyrics from memory

ability to hold two conversations at once

speed at completing crosswords

ability to assemble something from IKEA

ability to use a street directory

speed at completing jigsaw puzzles

What might be the ‘underlying factors’?

- Consider six variables that are intercorrelated.
  - Some more than others...
- The aim is to simplify our description of the information provided by the variables.
- A further aim may be to define the constructs which the variables describe.

$$\mathbf{X} =$$

Participant						
1	-0.14	-2.33	-2.75	0.71	1.87	-0.32
2	0.42	-0.35	0.38	-0.16	-0.62	0.52
3	-1.18	-0.53	-2.74	0.87	-0.61	0.51
4	0.67	1.44	-1.02	-1.11	-1.87	-0.54
5	0.27	0.26	1.66	0.81	0.58	1.29
6	0.11	0.25	-0.85	1.26	1.17	2.52
7	1.11	0.97	-1.66	1.30	0.79	2.60
8	0.54	0.43	-1.10	-1.51	-0.46	-1.96
9	0.50	1.72	2.25	-1.02	-0.97	-0.08
10	0.86	2.07	-0.40	-0.02	-0.30	-0.34
11	0.64	0.40	-0.56	1.42	1.31	1.95
12	-0.51	0.25	-0.05	0.39	0.25	-0.28
13	0.48	0.92	-0.95	-0.59	0.31	-2.62
14	0.54	1.13	0.11	-0.07	-0.74	0.12
15	-0.39	-0.92	-1.41	1.54	1.34	2.61
16	0.87	2.16	0.70	-0.83	-2.75	-1.41
17	-0.62	-1.88	-0.18	-0.32	-0.80	-1.00
18	1.80	0.44	1.68	0.90	1.38	0.08
19	0.65	-1.01	1.70	2.55	2.75	4.25
20	-1.42	-0.22	-2.00	-0.11	0.69	-1.00
⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	0.29	0.87	-1.99	0.24	-0.37	-0.28

$$Z =$$

Participant						
1	-0.10	-1.62	-1.78	0.78	1.45	-0.16
2	0.50	-0.19	0.31	-0.07	-0.28	0.37
3	-1.22	-0.32	-1.77	0.93	-0.28	0.37
4	0.77	1.10	-0.62	-0.99	-1.15	-0.30
5	0.34	0.25	1.16	0.87	0.55	0.86
6	0.17	0.24	-0.51	1.31	0.96	1.65
7	1.24	0.76	-1.05	1.35	0.70	1.70
8	0.63	0.37	-0.68	-1.38	-0.17	-1.21
9	0.59	1.30	1.56	-0.90	-0.53	-0.01
10	0.97	1.55	-0.21	0.07	-0.06	-0.18
11	0.74	0.35	-0.32	1.47	1.06	1.28
12	-0.50	0.24	0.02	0.47	0.32	-0.14
13	0.56	0.72	-0.58	-0.49	0.36	-1.63
14	0.63	0.87	0.13	0.02	-0.37	0.12
15	-0.37	-0.60	-0.88	1.58	1.08	1.70
16	0.98	1.62	0.52	-0.72	-1.77	-0.86
17	-0.62	-1.29	-0.06	-0.22	-0.41	-0.60
18	1.98	0.38	1.18	0.96	1.11	0.09
19	0.75	-0.67	1.19	2.56	2.06	2.75
20	-1.47	-0.10	-1.28	-0.02	0.63	-0.60
⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	0.36	0.69	-1.27	0.32	-0.11	-0.14

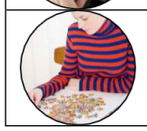
$\mathbf{R} =$

1	0.64	0.65	0.15	0.40	0.14
0.64	1	0.49	-0.04	0.19	-0.01
0.65	0.49	1	-0.13	0.15	-0.04
0.15	-0.04	-0.13	1	0.71	0.70
0.40	0.19	0.15	0.71	1	0.47
0.14	-0.01	-0.04	0.70	0.47	1

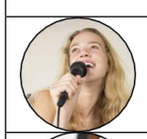
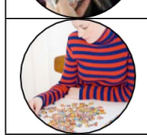
$\mathbf{R} =$

1	0.64	0.65	0.15	0.40	0.14
0.64	1	0.49	-0.04	0.19	-0.01
0.65	0.49	1	-0.13	0.15	-0.04
0.15	-0.04	-0.13	1	0.71	0.70
0.40	0.19	0.15	0.71	1	0.47
0.14	-0.01	-0.04	0.70	0.47	1

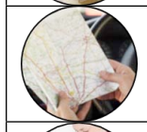
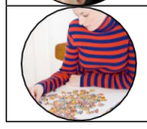
$\mathbf{L} =$

	1	2	3	4	5	6
	2.55	0	0	0	0	0
	0	2.02	0	0	0	0
	0	0	0.51	0	0	0
	0	0	0	0.49	0	0
	0	0	0	0	0.25	0
	0	0	0	0	0	0.18

$\mathbf{V} =$

	1	2	3	4	5	6
	0.48	-0.35	0.02	-0.07	-0.80	0.02
	0.35	-0.43	-0.59	0.47	0.34	-0.06
	0.32	-0.48	0.65	-0.18	0.43	-0.17
	0.41	0.48	-0.08	-0.07	0.02	-0.77
	0.50	0.24	-0.25	-0.58	0.24	0.49
	0.37	0.42	0.39	0.64	0.00	0.36

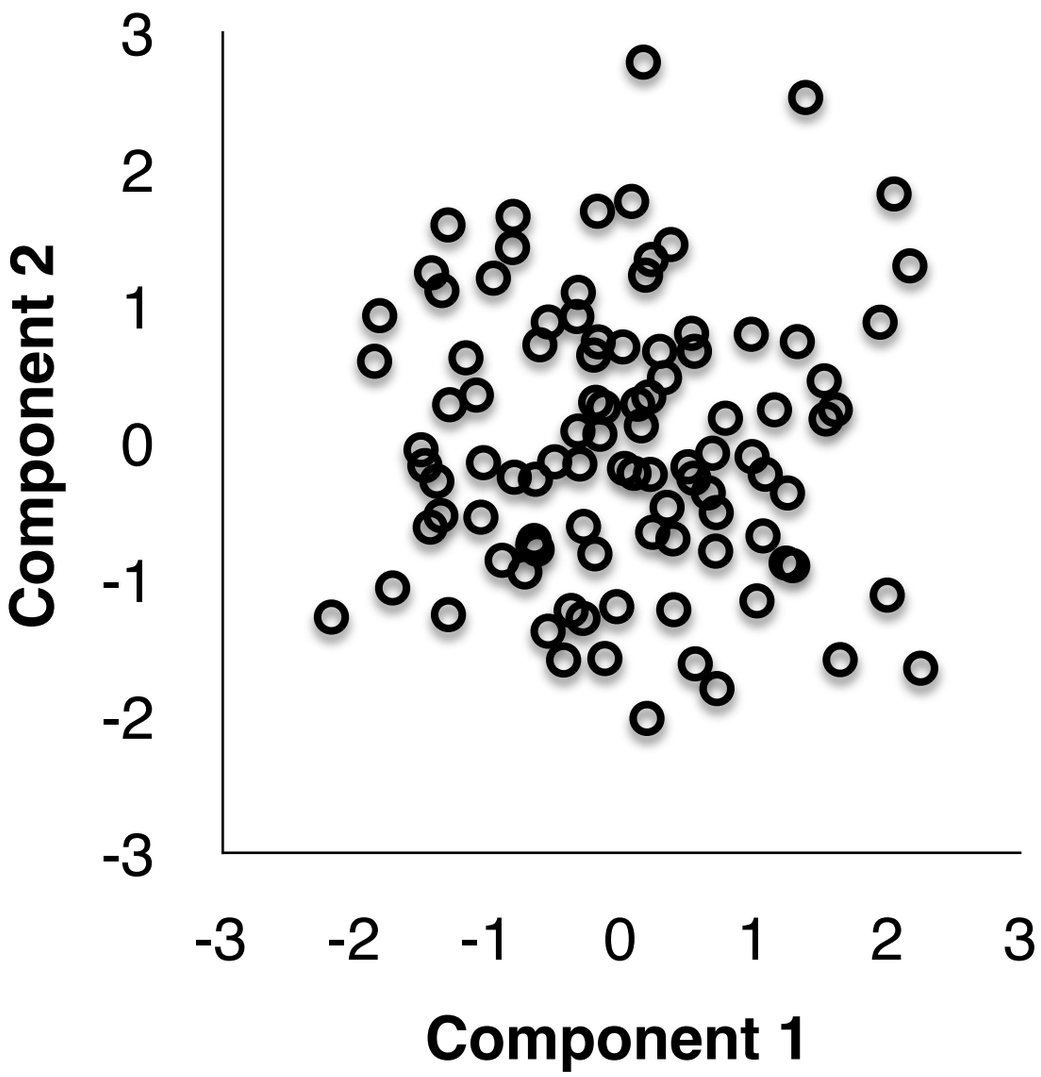
$\mathbf{A} =$

	1	2	3	4	5	6
	0.76	0.50	0.02	-0.05	-0.40	-0.01
	0.56	0.61	-0.42	0.33	0.17	0.02
	0.50	0.68	0.47	-0.12	0.22	0.07
	0.65	-0.68	-0.06	-0.05	0.01	0.33
	0.79	-0.34	-0.18	-0.40	0.12	-0.21
	0.59	-0.59	0.28	0.45	0.00	-0.15

These loadings represent correlations between factors and variables (i.e., between P and X)

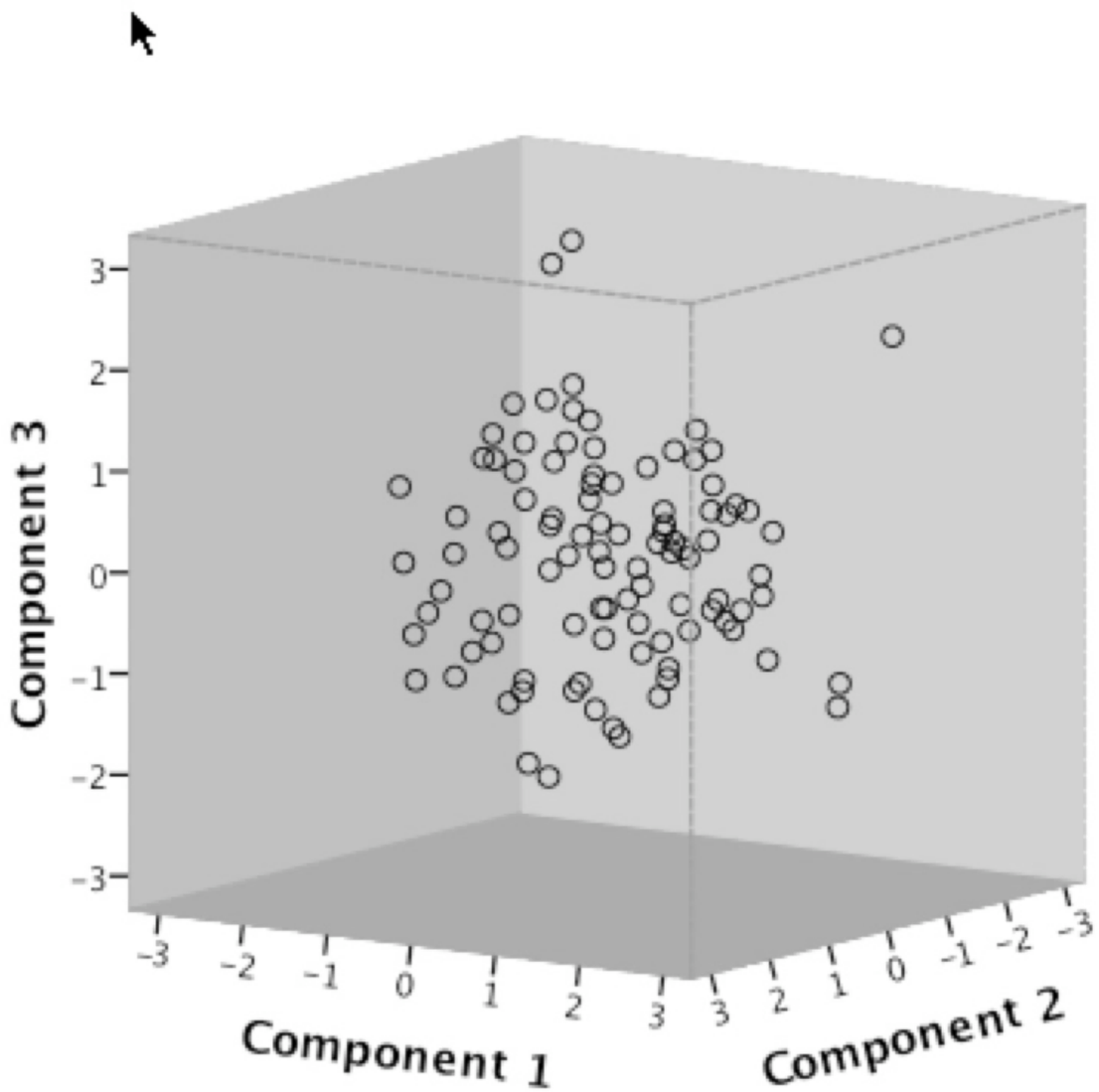
$\mathbf{P} =$

1	2	3	4	5	6
-0.12	-1.57	-0.96	-2.06	-1.73	-1.06
0.15	0.13	0.77	0.32	-0.80	-0.03
-0.55	-1.37	-1.20	0.81	0.11	0.98
-0.33	0.93	-1.12	1.59	-1.61	-0.34
0.98	-0.09	1.06	0.09	0.92	0.70
1.02	-1.15	-0.23	0.86	-0.02	-0.32
1.28	-0.89	-1.01	1.50	-1.96	-0.18
-0.55	0.89	-1.36	-0.47	-1.47	-1.52
0.37	1.46	0.64	0.94	0.99	-0.24
0.55	0.68	-1.53	0.89	-0.71	0.42
1.23	-0.87	-0.38	0.40	-0.64	0.22
0.09	-0.22	-0.43	-0.23	1.15	0.65
-0.17	0.75	-2.07	-1.16	-0.76	-0.08
0.32	0.48	-0.40	0.91	-0.47	0.50
0.72	-1.79	0.07	0.36	0.01	-0.26
-0.18	1.70	-0.61	1.61	-0.91	1.86
-0.80	-0.24	0.84	-0.99	-0.16	0.40
1.52	0.46	0.39	-1.15	-1.33	0.83
2.24	-1.64	2.18	-0.26	0.48	0.28
-0.66	-0.75	-1.68	-0.67	1.49	-0.73
⋮	⋮	⋮	⋮	⋮	⋮
0.02	-0.18	-1.79	0.68	-1.24	0.38



$$\mathbf{P} =$$

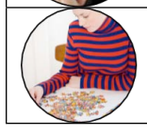
1	2	3	4	5	6
-0.12	-1.57	-0.96	-2.06	-1.73	-1.06
0.15	0.13	0.77	0.32	-0.80	-0.03
-0.55	-1.37	-1.20	0.81	0.11	0.98
-0.33	0.93	-1.12	1.59	-1.61	-0.34
0.98	-0.09	1.06	0.09	0.92	0.70
1.02	-1.15	-0.23	0.86	-0.02	-0.32
1.28	-0.89	-1.01	1.50	-1.96	-0.18
-0.55	0.89	-1.36	-0.47	-1.47	-1.52
0.37	1.46	0.64	0.94	0.99	-0.24
0.55	0.68	-1.53	0.89	-0.71	0.42
1.23	-0.87	-0.38	0.40	-0.64	0.22
0.09	-0.22	-0.43	-0.23	1.15	0.65
-0.17	0.75	-2.07	-1.16	-0.76	-0.08
0.32	0.48	-0.40	0.91	-0.47	0.50
0.72	-1.79	0.07	0.36	0.01	-0.26
-0.18	1.70	-0.61	1.61	-0.91	1.86
-0.80	-0.24	0.84	-0.99	-0.16	0.40
1.52	0.46	0.39	-1.15	-1.33	0.83
2.24	-1.64	2.18	-0.26	0.48	0.28
-0.66	-0.75	-1.68	-0.67	1.49	-0.73
⋮	⋮	⋮	⋮	⋮	⋮
0.02	-0.18	-1.79	0.68	-1.24	0.38

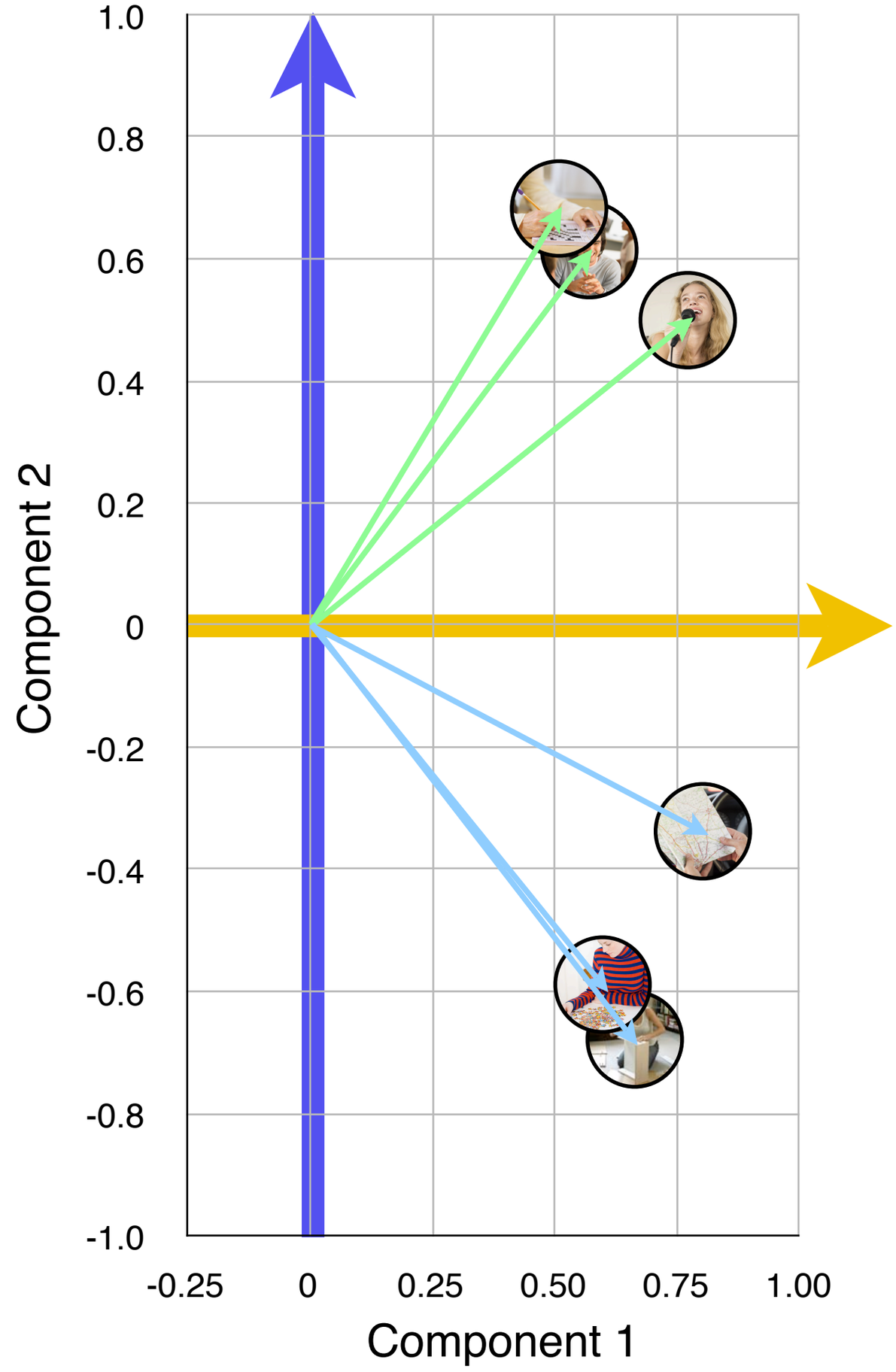


$$\mathbf{P} =$$

1	2	3	4	5	6
-0.12	-1.57	-0.96	-2.06	-1.73	-1.06
0.15	0.13	0.77	0.32	-0.80	-0.03
-0.55	-1.37	-1.20	0.81	0.11	0.98
-0.33	0.93	-1.12	1.59	-1.61	-0.34
0.98	-0.09	1.06	0.09	0.92	0.70
1.02	-1.15	-0.23	0.86	-0.02	-0.32
1.28	-0.89	-1.01	1.50	-1.96	-0.18
-0.55	0.89	-1.36	-0.47	-1.47	-1.52
0.37	1.46	0.64	0.94	0.99	-0.24
0.55	0.68	-1.53	0.89	-0.71	0.42
1.23	-0.87	-0.38	0.40	-0.64	0.22
0.09	-0.22	-0.43	-0.23	1.15	0.65
-0.17	0.75	-2.07	-1.16	-0.76	-0.08
0.32	0.48	-0.40	0.91	-0.47	0.50
0.72	-1.79	0.07	0.36	0.01	-0.26
-0.18	1.70	-0.61	1.61	-0.91	1.86
-0.80	-0.24	0.84	-0.99	-0.16	0.40
1.52	0.46	0.39	-1.15	-1.33	0.83
2.24	-1.64	2.18	-0.26	0.48	0.28
-0.66	-0.75	-1.68	-0.67	1.49	-0.73
⋮	⋮	⋮	⋮	⋮	⋮
0.02	-0.18	-1.79	0.68	-1.24	0.38

$\mathbf{A} =$

	1	2	3	4	5	6
	0.76	0.50	0.02	-0.05	-0.40	-0.01
	0.56	0.61	-0.42	0.33	0.17	0.02
	0.50	0.68	0.47	-0.12	0.22	0.07
	0.65	-0.68	-0.06	-0.05	0.01	0.33
	0.79	-0.34	-0.18	-0.40	0.12	-0.21
	0.59	-0.59	0.28	0.45	0.00	-0.15

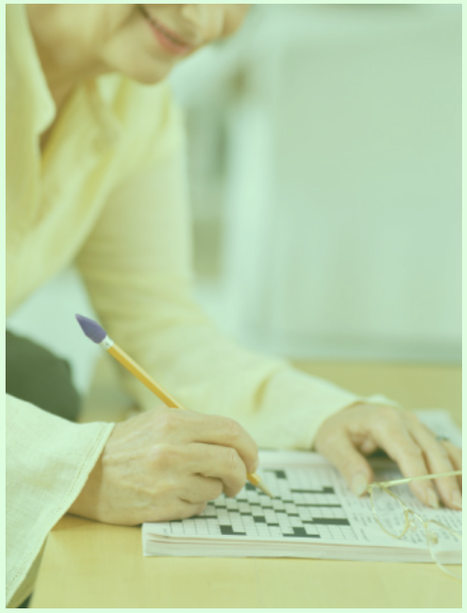


Another motivational example

Consider an investigation into the nature of intelligence. Data on six measures are collected:

ability to recite song lyrics from memory

ability to hold two conversations at once



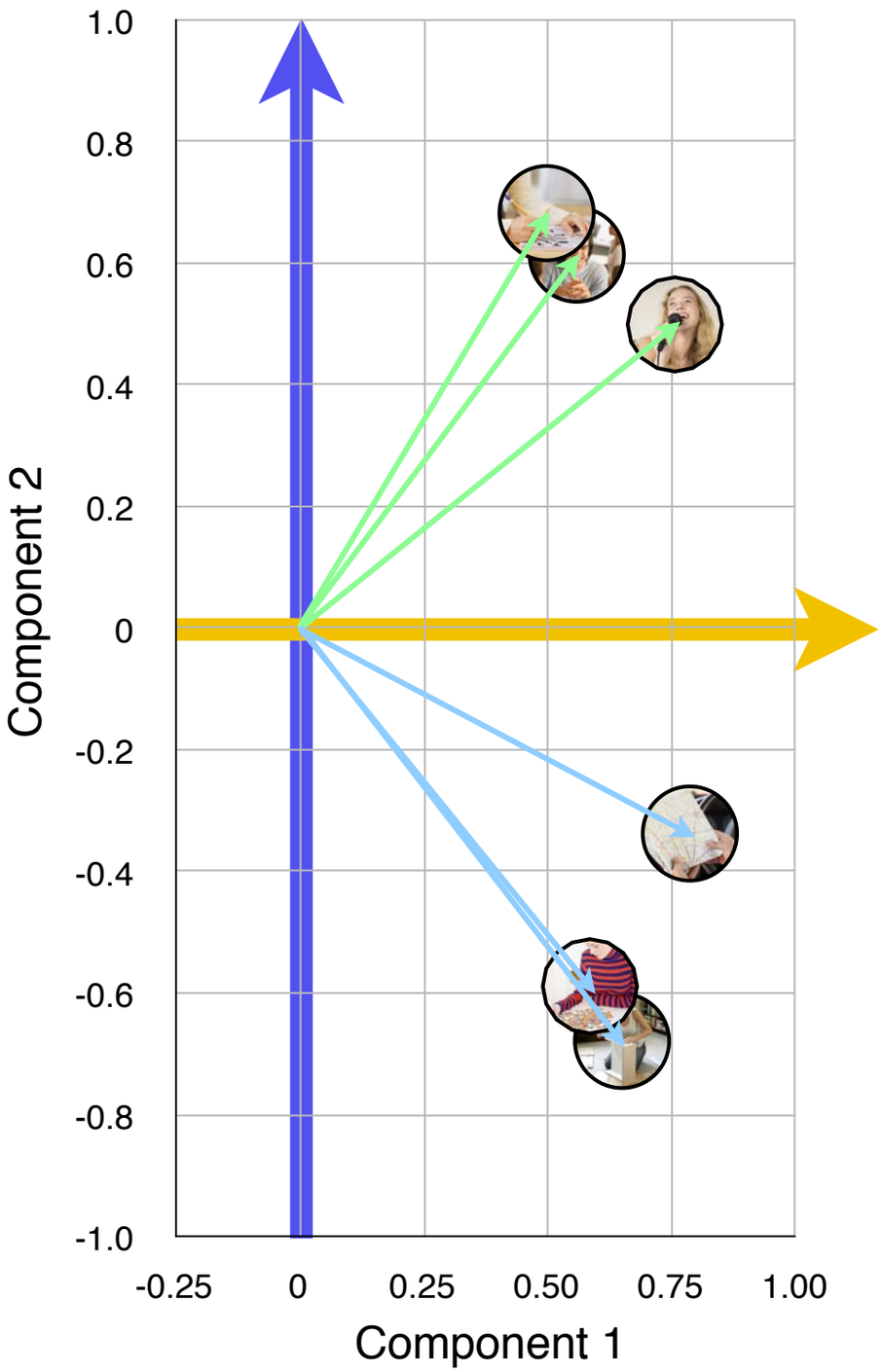
speed at completing crosswords

ability to assemble something from IKEA

ability to use a street directory

speed at completing jigsaw puzzles

What might be the ‘underlying factors’?



- The first principal component finds the direction in which all the variables seem to be pointing.
- However, the principal components don't seem to be pointing in the 'right' direction.
  - If we want a principal component (or factor) to be the thing that our variables have in common, we would like the principal components to 'line up' with our groups of variables.
- Two problems with a PCA solution:
  1. (above) 'lining up' our components with our groups of variables.
  2. These dimensions are plotted in 2D space. How do we know how many dimensions are needed to 'capture' the structure of the underlying variables?
- How do we assign meaning to components?
- How do we decide how many components can adequately account for the variation in the data?

Factor Analysis via PCA

- Overview
- The ‘number of factors’ problem
- The rotation problem
- Modelling data in factor analysis
- Schematic representation of factor analysis via PCA

Overview

Methods of Factor Analysis

Factor Analysis via PCA

- Based on PCA.
- Called “principal components analysis” by T&F.
  - involves performing a principal components analysis to repartition the entire variance.
  - choosing only *some* factors (and discarding the rest).
  - rotating the factors to make them easier to label.

Common Factor Analysis

- Not based on PCA.
- Called “factor analysis” by T&F.

“choosing only *some* factors (and discarding the rest).”

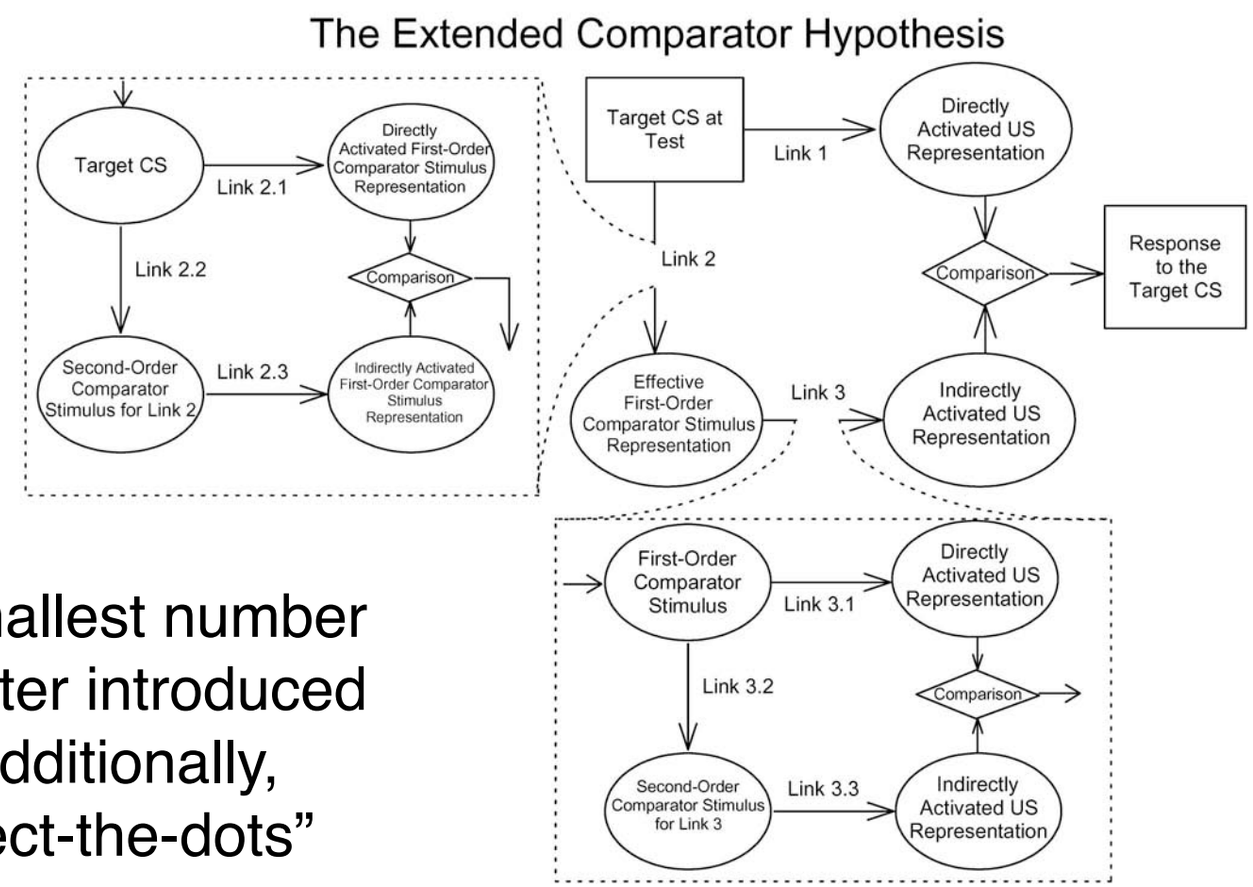
William of Ockham
(1288 - 1348)

“entia non sunt multiplicanda praeter necessitatem”

“entities should not be multiplied beyond necessity.”

“All things being equal, the simplest solution tends to be the best one.”

$$\Delta V = \alpha\beta\left(\lambda - V\right)$$



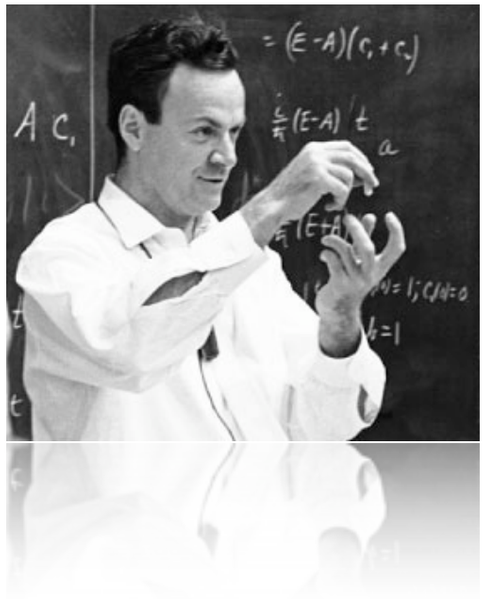
In general, mathematical models with the smallest number of parameters are preferred as each parameter introduced into the model adds some uncertainty to it. Additionally, adding too many parameters leads to “connect-the-dots” curve-fitting which has little predictive power.

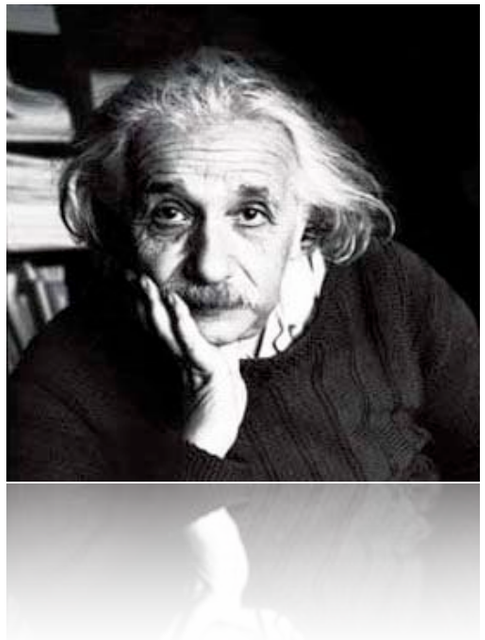
“If a thing can be done adequately by means of one, it is superfluous to do it by means of several; for we observe that nature does not employ two instruments where one suffices.”

- Thomas Aquinas

“The truth always turns out to be simpler than you thought.”

- Richard Feynman





“Things should be as simple as possible, but not simpler.”

- Albert Einstein

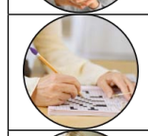
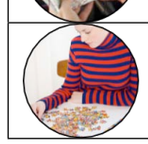
The ‘number of factors’ problem

- Choosing the number of factors requires a personal judgement.
- More factors provide more information, *but* less factors provide a simpler solution.
- How to help decide on the number of factors:
  - Eigenvalues > 1
  - Scree Plot
  - Parallel analysis test
  - Theoretical considerations
- It’s best to compare different solutions before making a final decision.

The ‘number of factors’ problem

Eigenvalues > 1

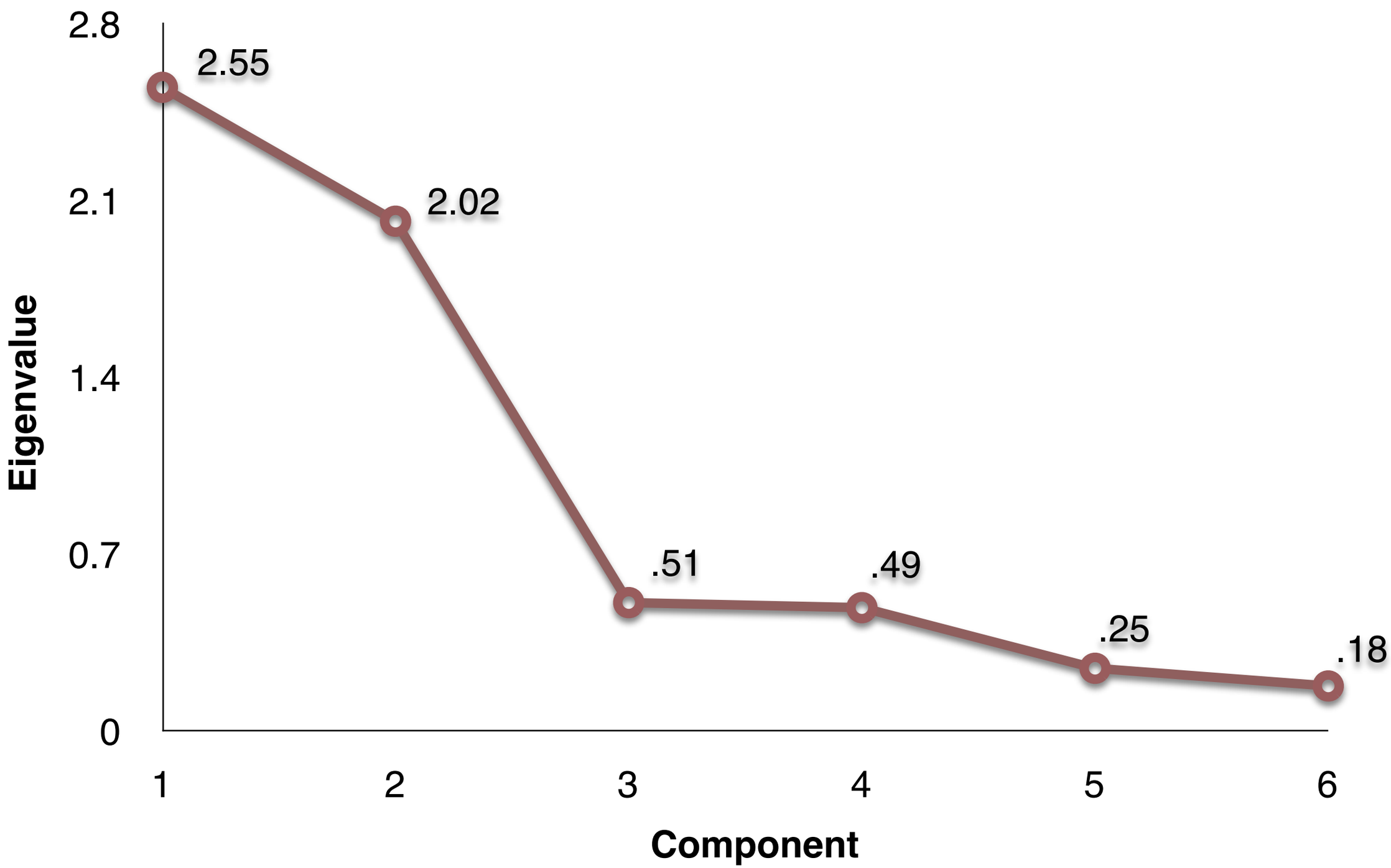
$$\mathbf{L} =$$

	1	2	3	4	5	6
	2.55	0	0	0	0	0
	0	2.02	0	0	0	0
	0	0	0.51	0	0	0
	0	0	0	0.49	0	0
	0	0	0	0	0.25	0
	0	0	0	0	0	0.18

‘eigenvalues greater than one’ rule...

The ‘number of factors’ problem

Scree Plot

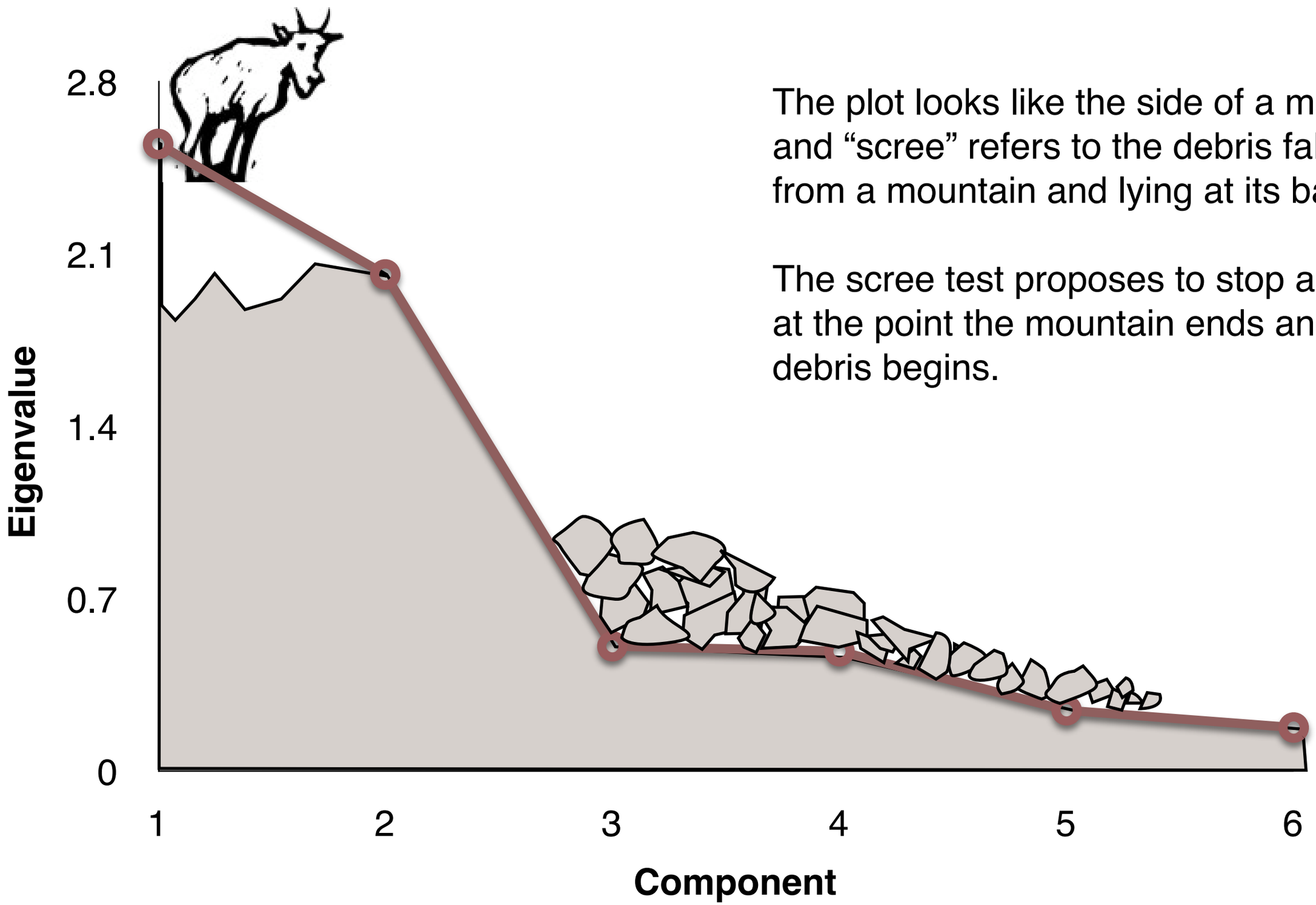


Frank's Slide

On April 29, 1903, at 4:10 a.m., 82 million tonnes (30 million cubic meters) of limestone crashed from the summit of Turtle Mountain and covered approximately three square kilometres of the valley floor. The slide dammed the Crowsnest River and formed a small lake and buried seven houses on the outskirts of the sleeping town of Frank, as well as several rural buildings. Frank was home to approximately 600 people in 1903; of the roughly 100 individuals who lived in the path of the slide, more than 70 were killed.

The ‘number of factors’ problem

Scree Plot



The plot looks like the side of a mountain, and “scree” refers to the debris fallen from a mountain and lying at its base.

The scree test proposes to stop analysis at the point the mountain ends and the debris begins.

The ‘number of factors’ problem

Parallel analysis test

Behavior Research Methods, Instruments, & Computers
2000, 32 (3), 396-402

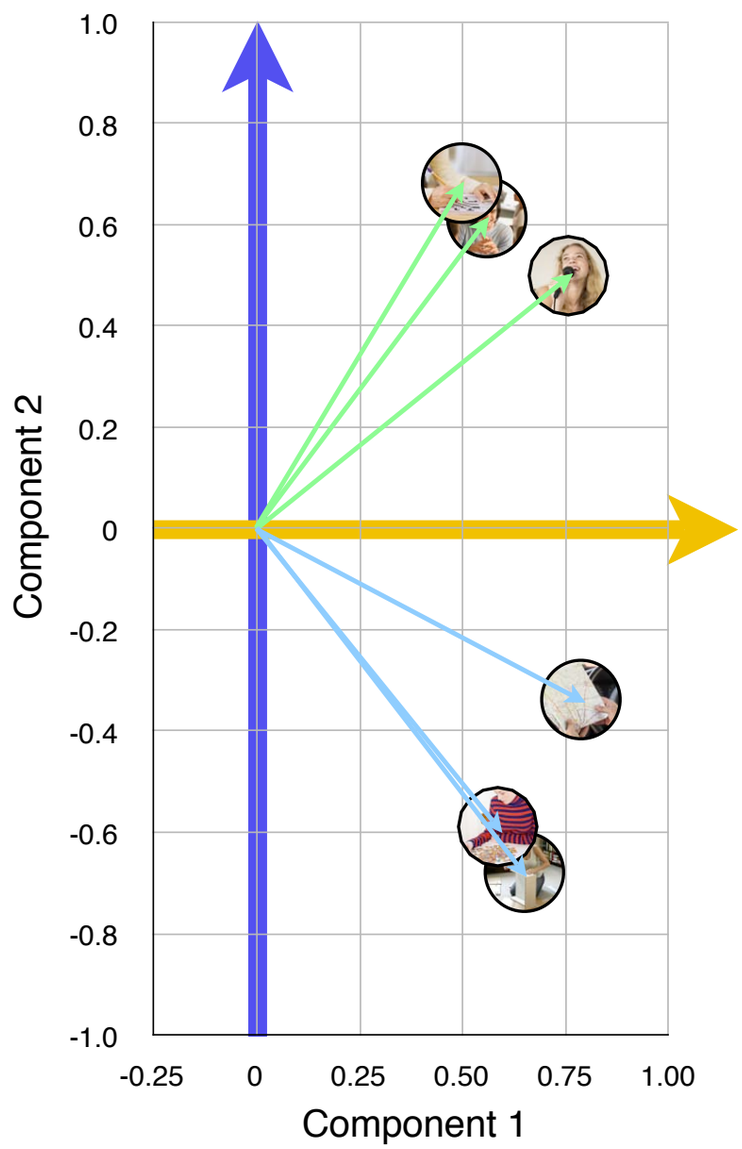
SPSS and SAS programs for determining the number of components using parallel analysis and Velicer’s MAP test

BRIAN P. O’CONNOR
Lakehead University, Thunder Bay, Ontario, Canada

Popular statistical software packages do not have the proper procedures for determining the number of components in factor and principal components analyses. Parallel analysis and Velicer’s minimum average partial (MAP) test are validated procedures, recommended widely by statisticians. However, many researchers continue to use alternative, simpler, but flawed procedures, such as the eigenvalues-greater-than-one rule. Use of the proper procedures might be increased if these procedures could be conducted within familiar software environments. This paper describes brief and efficient programs for using SPSS and SAS to conduct parallel analyses and the MAP test.

http://people.ok.ubc.ca/brioconn/nfactors/nfactors.html

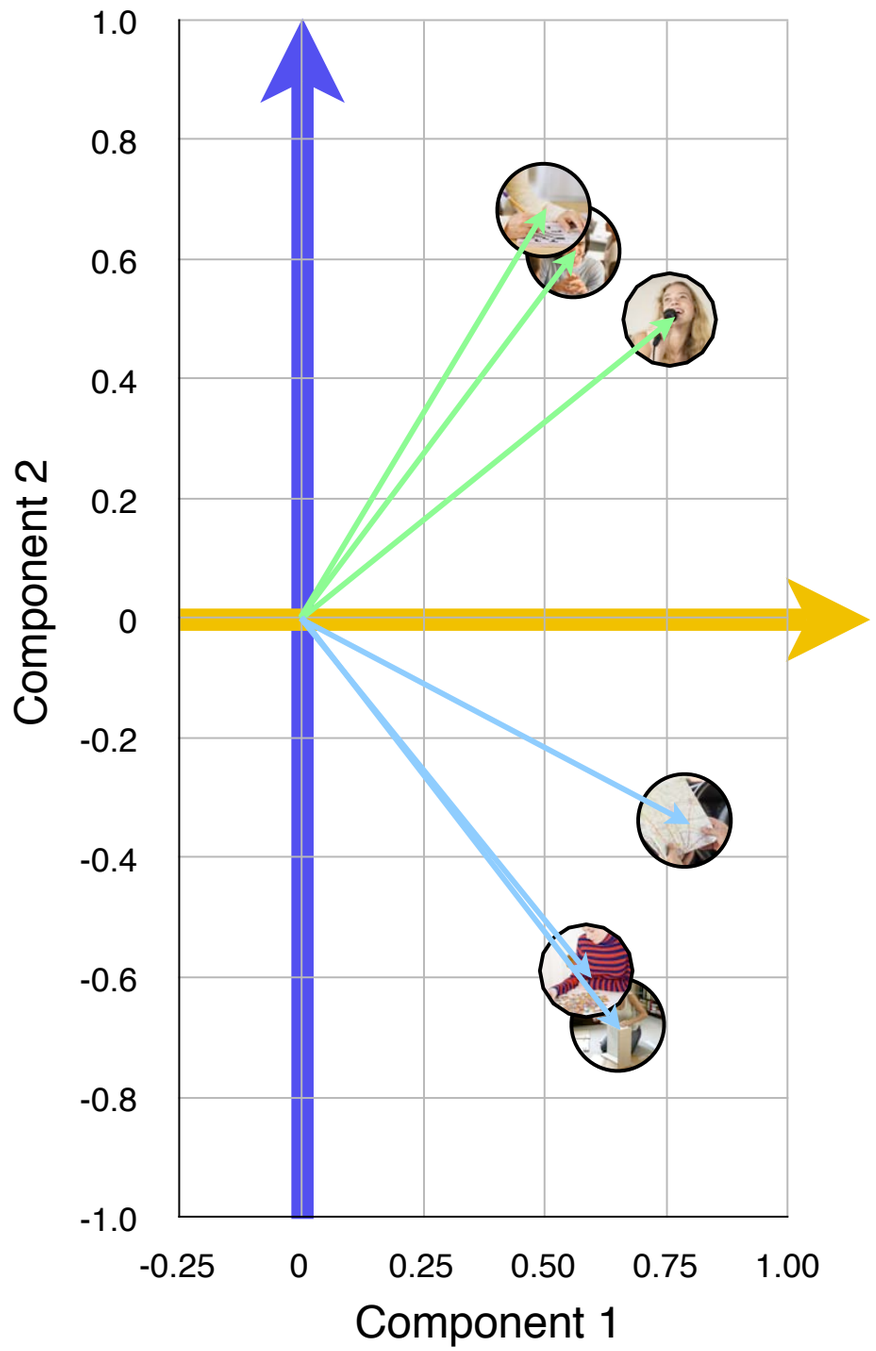
The rotation problem



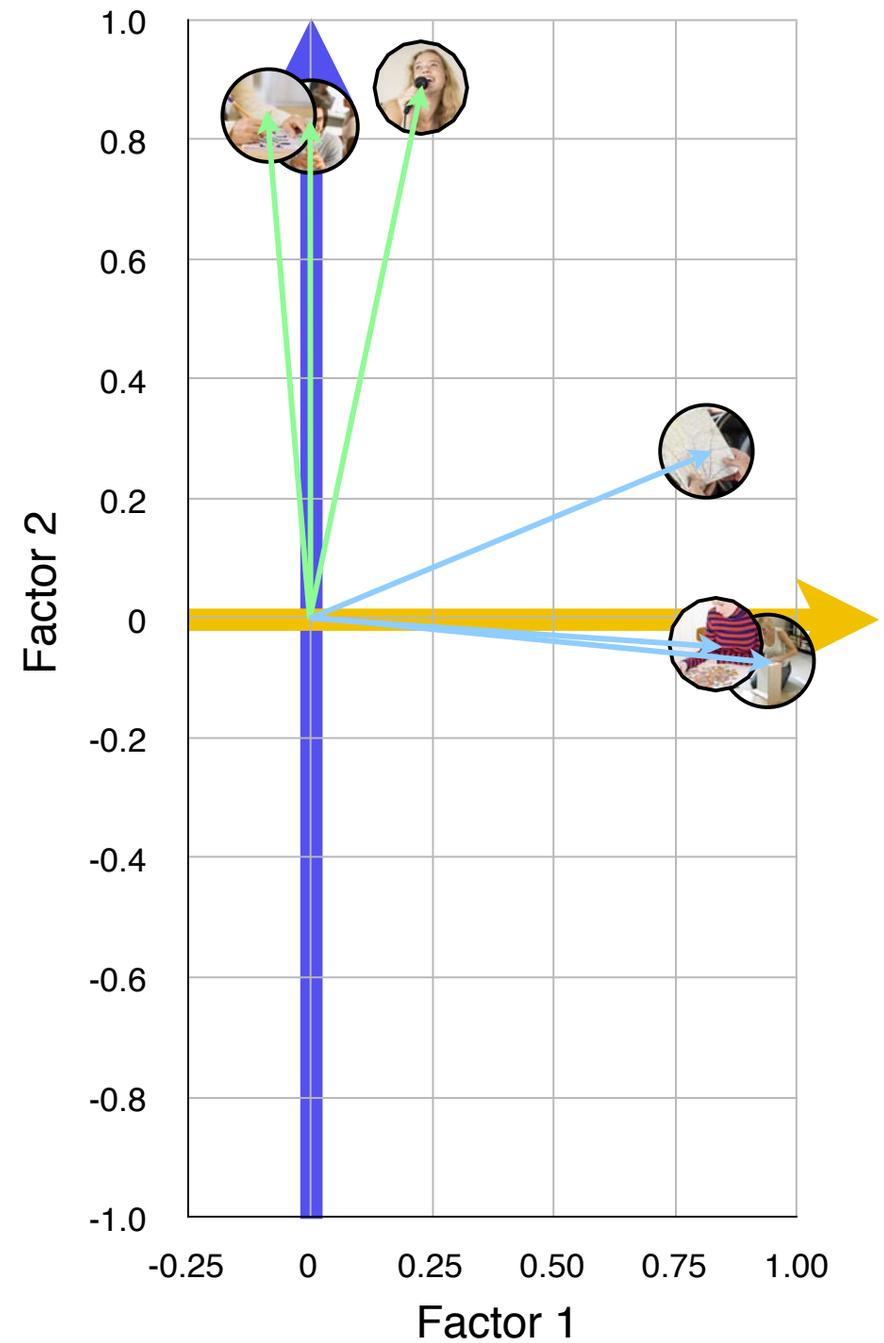
One simple solution to the observation that the components are not aligning with the directions of the two groups of variables is to rotate the axes so that they do align.

- Factors are rotated to make it easier to label
- Two types of rotation are possible:
  1. *Orthogonal*: factors remain uncorrelated.
  2. *Oblique*: factors are allowed to become correlated

The rotation problem

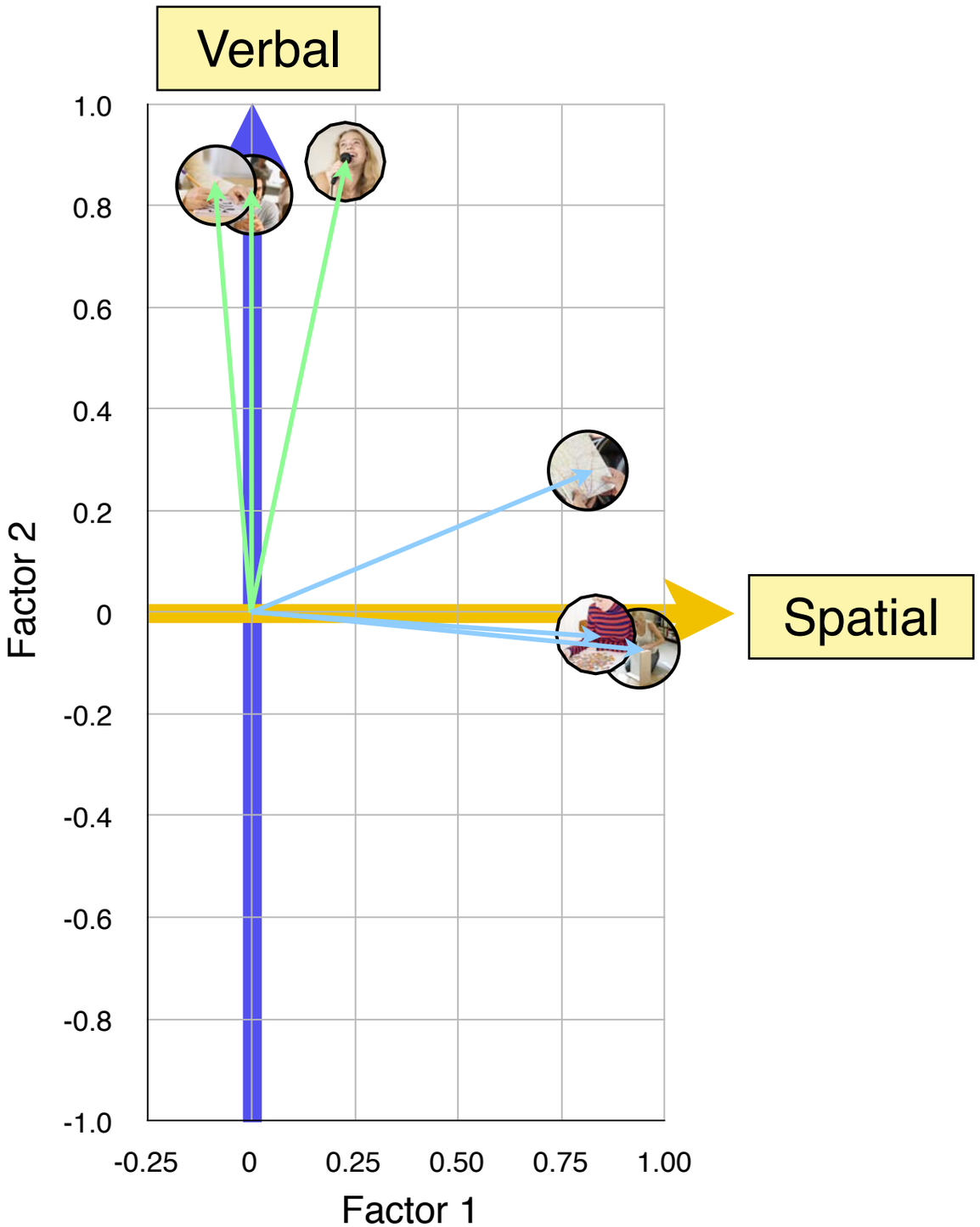


Unrotated Solution



Rotated Solution
(orthogonal)

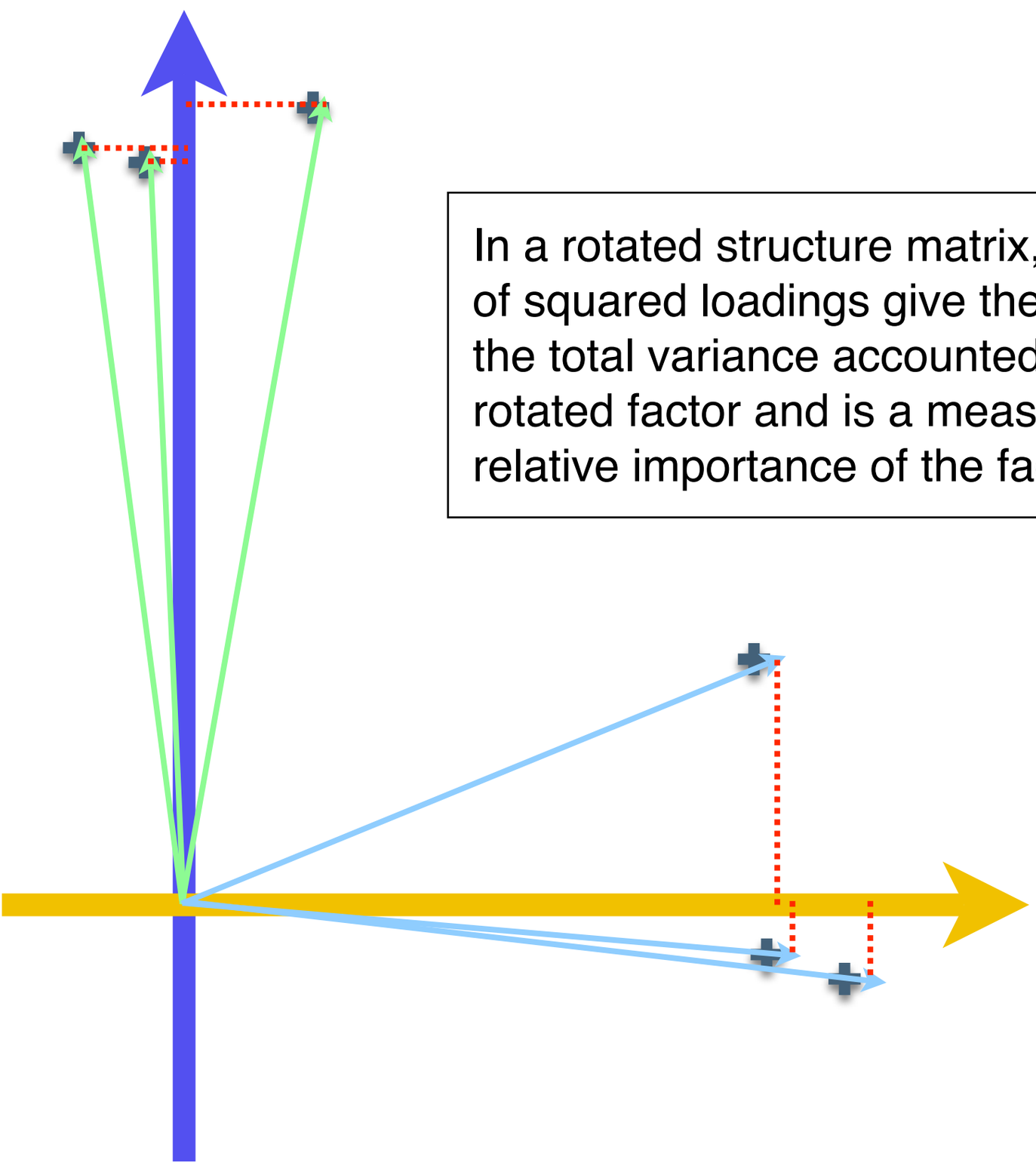
The rotation problem



The meanings of the variables that load highly on a factor are taken to define the meaning of that factor or to label the factor. Obviously, we need to know and are clear about what the variables mean or what the variables measure.

Orthogonal rotation to simple structure:

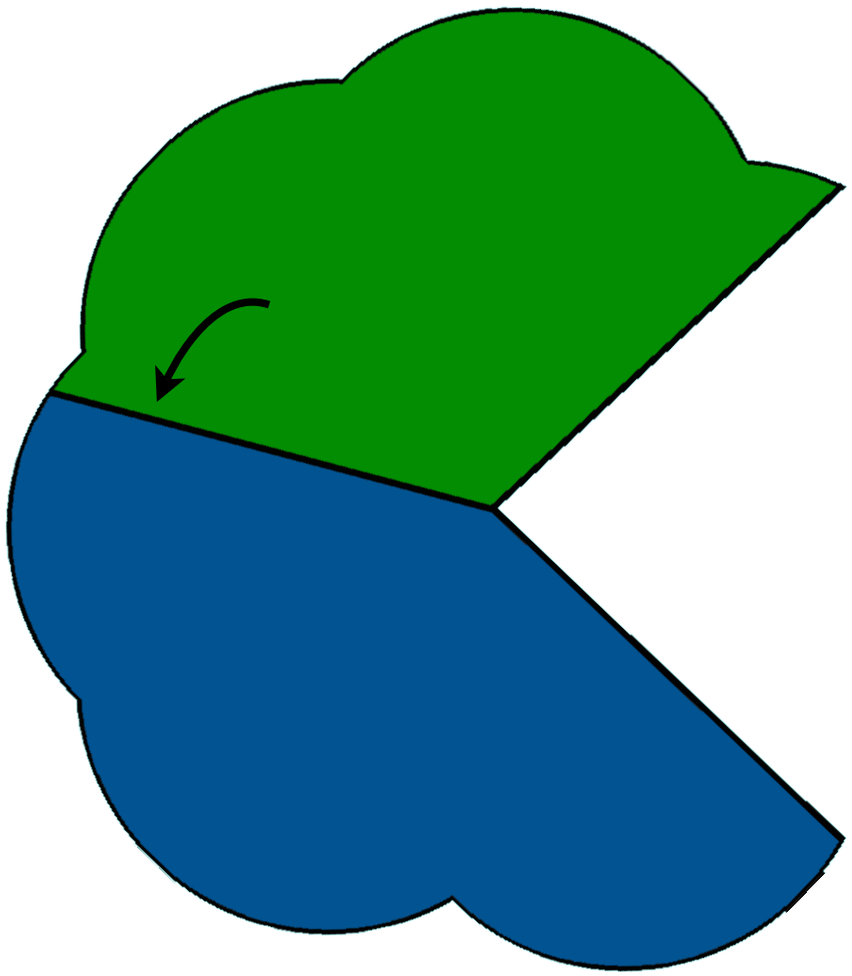
variance accounted for by the factors = SS loadings



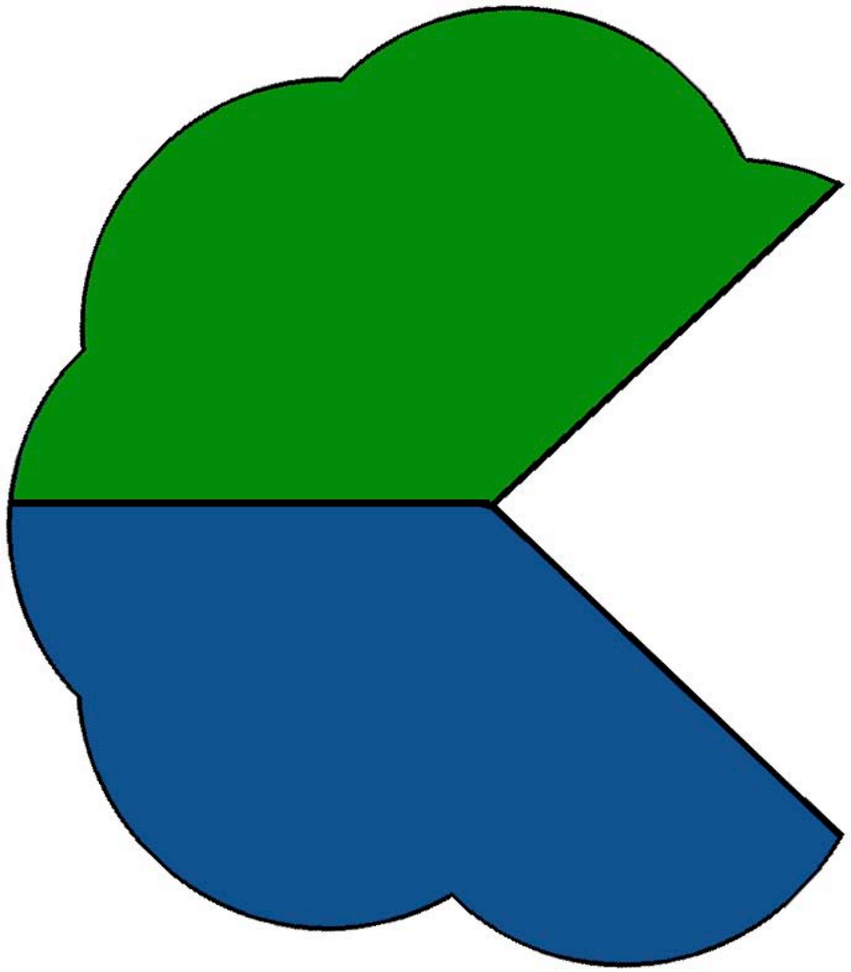
In a rotated structure matrix, the sum of squared loadings give the amount of the total variance accounted for by that rotated factor and is a measure of the relative importance of the factor.

In an unrotated structure matrix, the the eigenvalues are the sum of squared loadings (SSLs) of the variables for each factor.

Orthogonal rotation:
Euler Representation



Unrotated Solution



Rotated Solution
(orthogonal)

Orthogonal rotation:

Matrix Representation

$$\mathbf{A_{rotated}} = \mathbf{A_{unrotated}}\mathbf{\Lambda}$$

$\mathbf{\Lambda}$ = the factor transformation matrix

- represents the angles that the axes are rotated.
- this isn't interpreted.

$\mathbf{A_{rotated}}$

	1	2
	0.23	0.88
	0.00	0.83
	-0.08	0.84
	0.94	-0.07
	0.82	0.28
	0.84	-0.04

=

$\mathbf{A_{unrotated}}$

	1	2
	0.76	0.50
	0.56	0.61
	0.50	0.68
	0.65	-0.68
	0.79	-0.34
	0.59	-0.59

×

$\mathbf{\Lambda}$

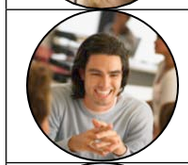
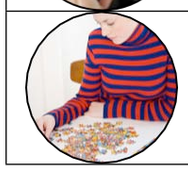
1	2
0.74	0.67
-0.67	0.74

Orthogonal rotation:
Loading Matrices

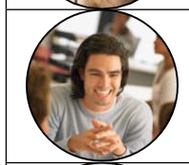
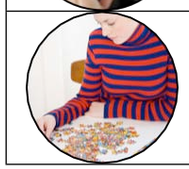
Remember:
These loadings are ‘regression-like’ weights used to estimate the unique contribution of each factor to the variance of the variable.

They represent correlations between the factors and the variables.

$\mathbf{A}_{\mathbf{unrotated}}$

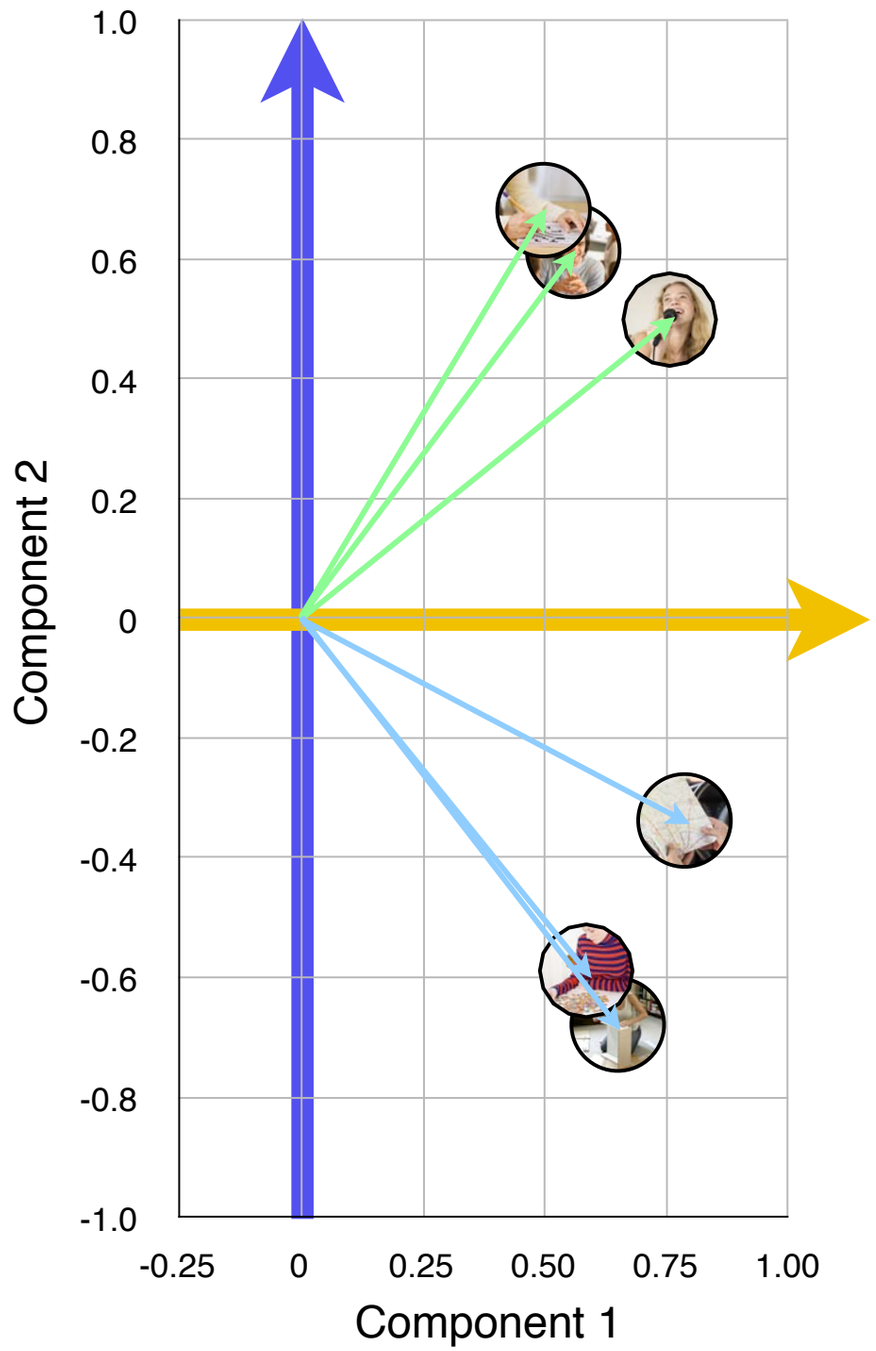
	1	2
	0.76	0.50
	0.56	0.61
	0.50	0.68
	0.65	-0.68
	0.79	-0.34
	0.59	-0.59

$\mathbf{A}_{\mathbf{rotated}}$

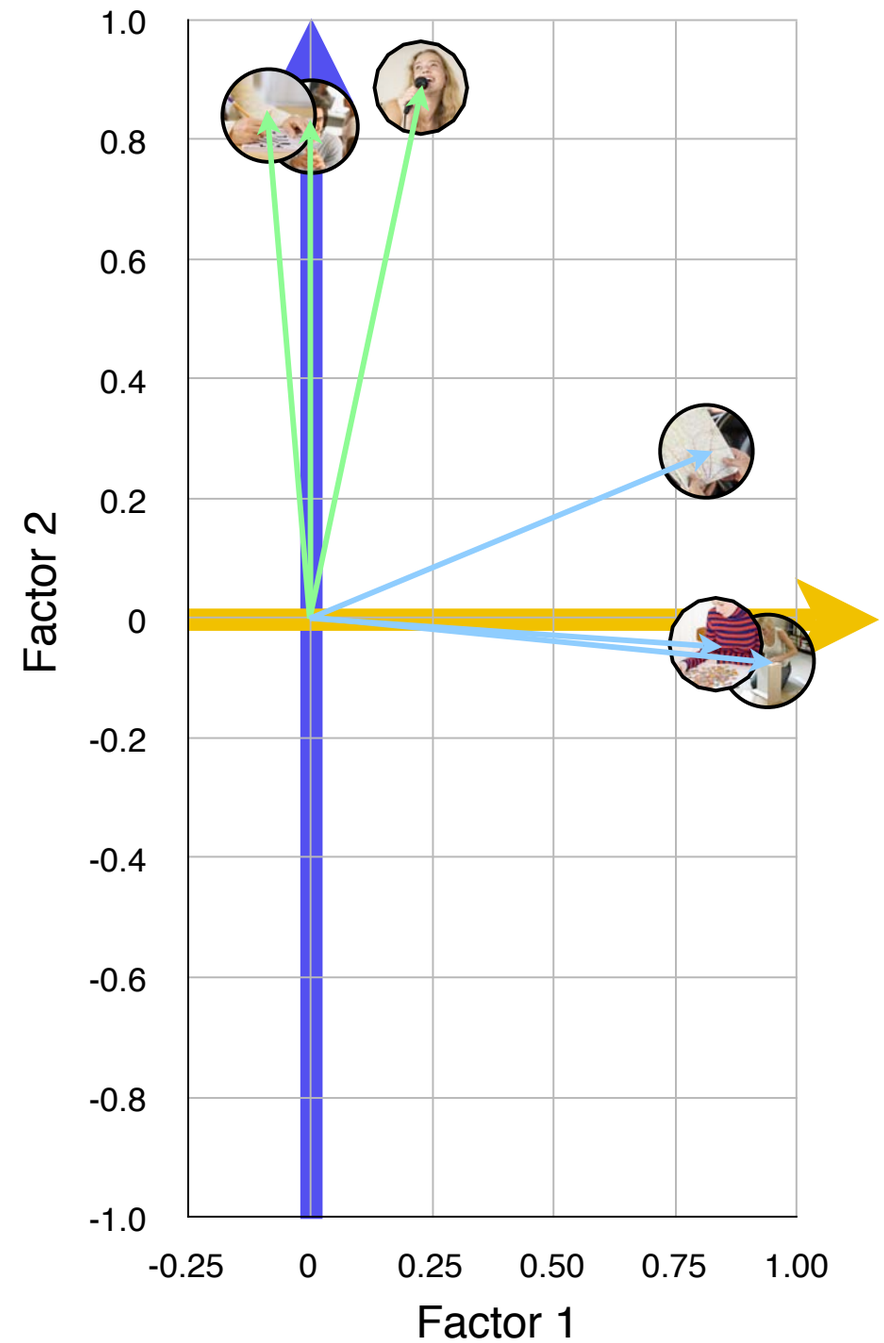
	1	2
	0.23	0.88
	0.00	0.83
	-0.08	0.84
	0.94	-0.07
	0.82	0.28
	0.84	-0.04

High loadings are interpreted

The rotation problem



Unrotated Solution

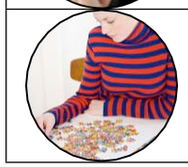


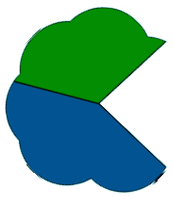
Rotated Solution (orthogonal)

Importance of factors

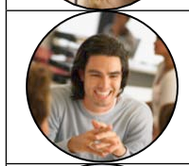
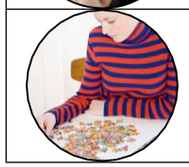
In an orthogonally rotated factor solution, the Sum of Squared Loadings (SSLs) give the amount of variance accounted for by the rotated factor.

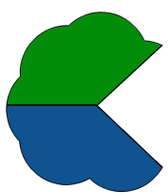
$\mathbf{A}_{\mathbf{unrotated}}$

	1	2
	0.76	0.50
	0.56	0.61
	0.50	0.68
	0.65	-0.68
	0.79	-0.34
	0.59	-0.59
Sum of Square Loadings	2.55	2.02
/6	42.4%	33.7%



$\mathbf{A}_{\mathbf{rotated}}$

	1	2
	0.23	0.88
	0.00	0.83
	-0.08	0.84
	0.94	-0.07
	0.82	0.28
	0.84	-0.04
	2.31	2.26
	38.5%	37.6%



Factor Scores

- We can generate factor scores for each case using the loading matrix.
- This gives us regression-like coefficients for weighing our variable scores to produce factor scores.

$$\mathbf{R}^{-1}\mathbf{A} = \mathbf{B}$$

$\mathbf{R}^{-1}$

2.70	-1.03	-1.18	-0.13	-0.61	-0.05
-1.03	1.76	-0.15	0.24	-0.08	0.02
-1.18	-0.15	1.95	0.63	-0.16	-0.12
-0.13	0.24	0.63	3.47	-1.82	-1.52
-0.61	-0.08	-0.16	-1.82	2.50	0.17
-0.05	0.02	-0.12	-1.52	0.17	1.99

$\times$

$\mathbf{A}_{\mathbf{rotated}}$

	1	2
	0.23	0.88
	0.00	0.83
	-0.08	0.84
	0.94	-0.07
	0.82	0.28
	0.84	-0.04

$=$

$\mathbf{B}$

	1	2
	0.06	0.39
	-0.04	0.37
	-0.08	0.38
	0.42	-0.08
	0.34	0.08
	0.37	-0.06

Factor Scores

- If we multiply each subject's original standardised score by these coefficients, then we get the factor scores.

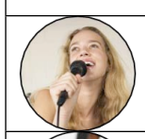
$$\mathbf{ZB} = \mathbf{F}$$

$\mathbf{Z}$

Participant						
1	-0.10	-1.62	-1.78	0.78	1.45	-0.16
2	0.50	-0.19	0.31	-0.07	-0.28	0.37
3	-1.22	-0.32	-1.77	0.93	-0.28	0.37
4	0.77	1.10	-0.62	-0.99	-1.15	-0.30
5	0.34	0.25	1.16	0.87	0.55	0.86
6	0.17	0.24	-0.51	1.31	0.96	1.65
7	1.24	0.76	-1.05	1.35	0.70	1.70
8	0.63	0.37	-0.68	-1.38	-0.17	-1.21
9	0.59	1.30	1.56	-0.90	-0.53	-0.01
10	0.97	1.55	-0.21	0.07	-0.06	-0.18
11	0.74	0.35	-0.32	1.47	1.06	1.28
12	-0.50	0.24	0.02	0.47	0.32	-0.14
13	0.56	0.72	-0.58	-0.49	0.36	-1.63
14	0.63	0.87	0.13	0.02	-0.37	0.12
15	-0.37	-0.60	-0.88	1.58	1.08	1.70
16	0.98	1.62	0.52	-0.72	-1.77	-0.86
17	-0.62	-1.29	-0.06	-0.22	-0.41	-0.60
18	1.98	0.38	1.18	0.96	1.11	0.09
19	0.75	-0.67	1.19	2.56	2.06	2.75
20	-1.47	-0.10	-1.28	-0.02	0.63	-0.60
⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	0.36	0.69	-1.27	0.32	-0.11	-0.14

$\times$

$\mathbf{B}$

	1	2
	0.06	0.39
	-0.04	0.37
	-0.08	0.38
	0.42	-0.08
	0.34	0.08
	0.37	-0.06

$=$

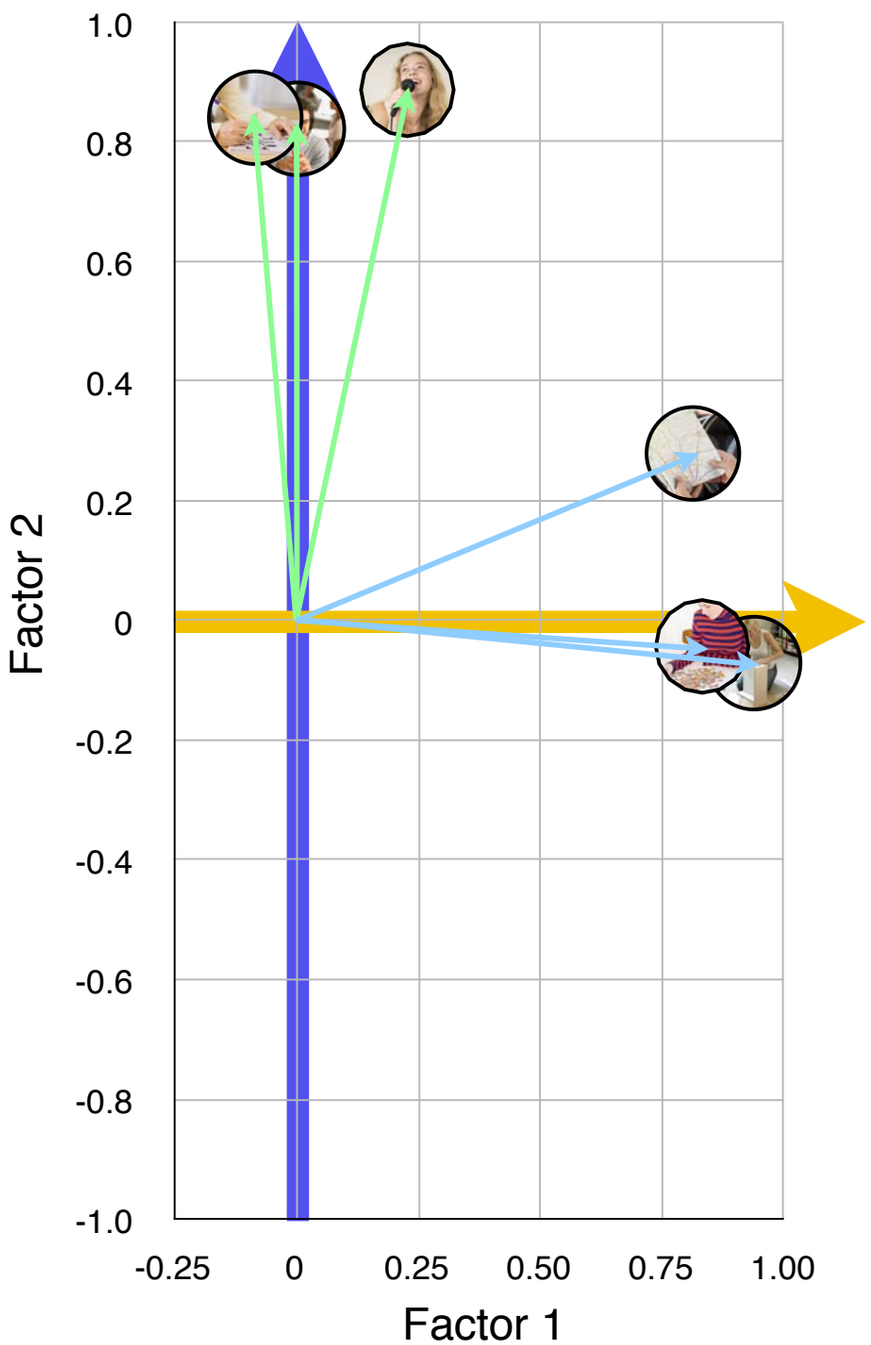
$\mathbf{F}$

Participant	1	2
1	0.96	-1.25
2	0.02	0.20
3	0.51	-1.38
4	-0.87	0.46
5	0.79	0.59
6	1.52	-0.17
7	1.55	0.20
8	-1.01	0.29
9	-0.71	1.33
10	-0.05	0.87
11	1.50	0.18
12	0.21	-0.10
13	-0.63	0.44
14	-0.09	0.58
15	1.73	-0.84
16	-1.28	1.14
17	-0.43	-0.72
18	0.82	1.36
19	2.76	0.29
20	0.01	-1.00
	⋮	⋮
100	0.14	-0.12

Factor Scores

- If we multiply each subject's original standardised score by these coefficients, then we get the factor scores.

$$\mathbf{ZB} = \mathbf{F}$$



This means that the first participant has an estimated standard score of .96 on the first factor (verbal), and -1.25 on the second factor (spatial).

This particular participant seems to be more verbally inclined than they are spatially.

$\mathbf{F}$

Participant	1	2
1	0.96	-1.25
2	0.02	0.20
3	0.51	-1.38
4	-0.87	0.46
5	0.79	0.59
6	1.52	-0.17
7	1.55	0.20
8	-1.01	0.29
9	-0.71	1.33
10	-0.05	0.87
11	1.50	0.18
12	0.21	-0.10
13	-0.63	0.44
14	-0.09	0.58
15	1.73	-0.84
16	-1.28	1.14
17	-0.43	-0.72
18	0.82	1.36
19	2.76	0.29
20	0.01	-1.00
	⋮	⋮
100	0.14	-0.12

Factor Scores

- We can also predict each participant's score on each variable using these factor scores.

$$\mathbf{FA}' = \mathbf{Z}_{\mathbf{predicted}}$$

$\mathbf{F}$

Participant	1	2
1	0.96	-1.25
2	0.02	0.20
3	0.51	-1.38
4	-0.87	0.46
5	0.79	0.59
6	1.52	-0.17
7	1.55	0.20
8	-1.01	0.29
9	-0.71	1.33
10	-0.05	0.87
11	1.50	0.18
12	0.21	-0.10
13	-0.63	0.44
14	-0.09	0.58
15	1.73	-0.84
16	-1.28	1.14
17	-0.43	-0.72
18	0.82	1.36
19	2.76	0.29
20	0.01	-1.00
	⋮	⋮
100	0.14	-0.12

$\times$

$\mathbf{A}'_{\mathbf{rotated}}$

1	0.23	0.00	-0.08	0.94	0.82	0.84
2	0.89	0.83	0.84	-0.07	0.28	-0.04

$=$

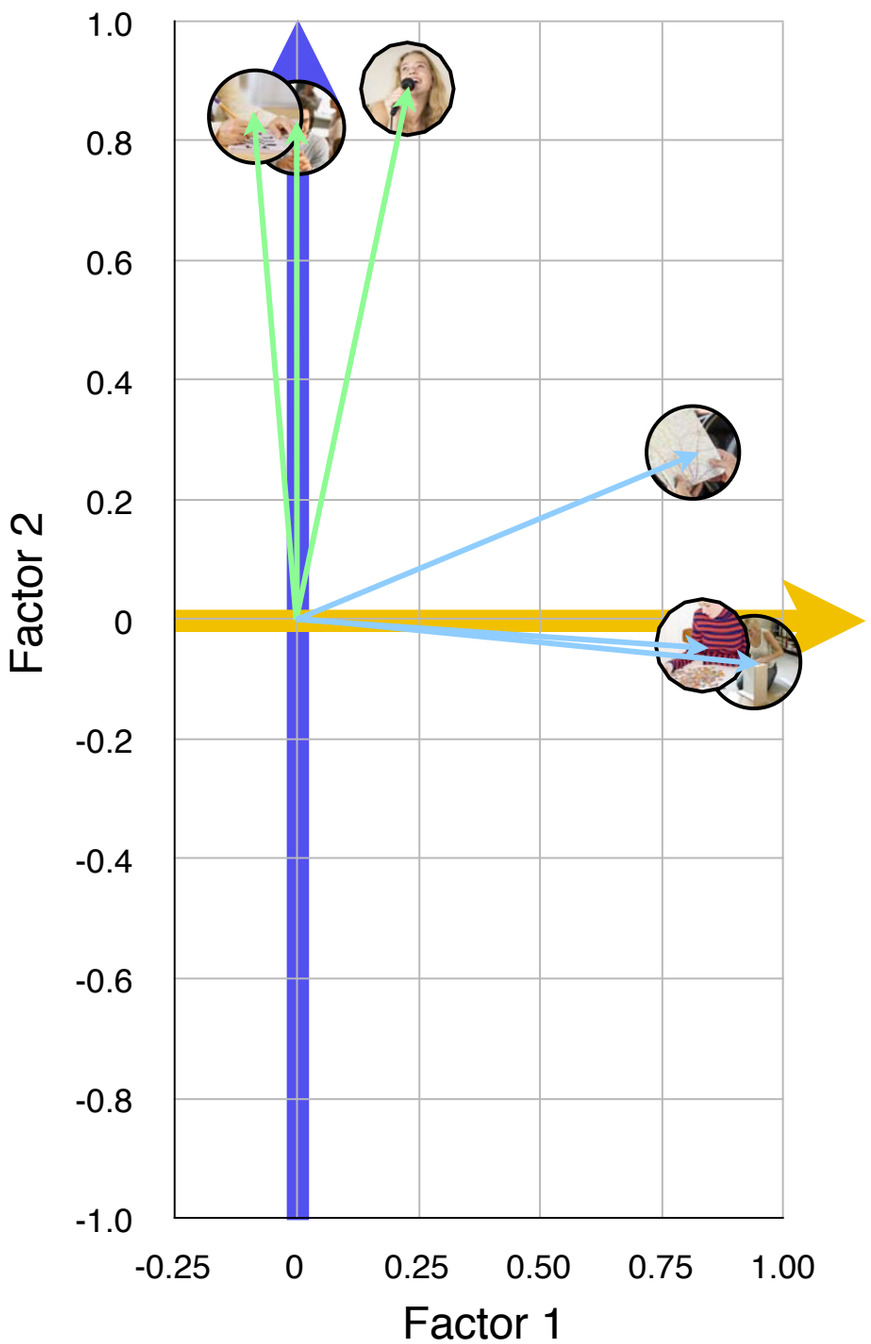
$\mathbf{Z}_{\mathbf{predicted}}$

Participant						
1	-0.88	-1.02	-1.13	0.99	0.44	0.86
2	0.18	0.16	0.16	0.01	0.07	0.01
3	-1.11	-1.14	-1.21	0.58	0.03	0.49
4	0.21	0.38	0.46	-0.85	-0.58	-0.75
5	0.70	0.49	0.43	0.70	0.81	0.63
6	0.20	-0.13	-0.27	1.44	1.20	1.28
7	0.54	0.17	0.04	1.44	1.32	1.29
8	0.03	0.24	0.33	-0.96	-0.74	-0.85
9	1.01	1.09	1.18	-0.76	-0.20	-0.65
10	0.76	0.72	0.74	-0.11	0.20	-0.08
11	0.51	0.16	0.03	1.40	1.28	1.25
12	-0.04	-0.08	-0.10	0.21	0.15	0.18
13	0.24	0.36	0.42	-0.62	-0.39	-0.55
14	0.49	0.47	0.49	-0.12	0.09	-0.10
15	-0.35	-0.69	-0.86	1.69	1.18	1.48
16	0.72	0.93	1.06	-1.28	-0.73	-1.12
17	-0.74	-0.60	-0.57	-0.35	-0.56	-0.33
18	1.39	1.13	1.08	0.67	1.05	0.62
19	0.89	0.25	0.02	2.58	2.34	2.30
20	-0.88	-0.83	-0.84	0.08	-0.27	0.05
⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	-0.07	-0.10	-0.11	0.14	0.08	0.12

Factor Scores

- We can also predict each participant's score on each variable using these factor scores.

$$\mathbf{FA}' = \mathbf{Z}_{\mathbf{predicted}}$$



This means that the first participant is predicted to have a standardised score of -.88 on lyric recollection, -1.02 on holding two conversations at once, etc. for each of the six variables.

$\mathbf{Z}_{\mathbf{predicted}}$

Participant						
1	-0.88	-1.02	-1.13	0.99	0.44	0.86
2	0.18	0.16	0.16	0.01	0.07	0.01
3	-1.11	-1.14	-1.21	0.58	0.03	0.49
4	0.21	0.38	0.46	-0.85	-0.58	-0.75
5	0.70	0.49	0.43	0.70	0.81	0.63
6	0.20	-0.13	-0.27	1.44	1.20	1.28
7	0.54	0.17	0.04	1.44	1.32	1.29
8	0.03	0.24	0.33	-0.96	-0.74	-0.85
9	1.01	1.09	1.18	-0.76	-0.20	-0.65
10	0.76	0.72	0.74	-0.11	0.20	-0.08
11	0.51	0.16	0.03	1.40	1.28	1.25
12	-0.04	-0.08	-0.10	0.21	0.15	0.18
13	0.24	0.36	0.42	-0.62	-0.39	-0.55
14	0.49	0.47	0.49	-0.12	0.09	-0.10
15	-0.35	-0.69	-0.86	1.69	1.18	1.48
16	0.72	0.93	1.06	-1.28	-0.73	-1.12
17	-0.74	-0.60	-0.57	-0.35	-0.56	-0.33
18	1.39	1.13	1.08	0.67	1.05	0.62
19	0.89	0.25	0.02	2.58	2.34	2.30
20	-0.88	-0.83	-0.84	0.08	-0.27	0.05
⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	-0.07	-0.10	-0.11	0.14	0.08	0.12

Note: The closer that these predicted Z scores match the actual Z scores, the better the factor analysis captured the relationship among the variables.

Factor Scores

- We can also predict each participant's score on each variable using these factor scores.

$$\mathbf{FA}' = \mathbf{Z}_{\mathbf{predicted}}$$

We're calculating these using the regression method.

$$Z_j \leftarrow F_1 F_2 \ldots F_m$$

Remember me?

DATA = MODEL + RESIDUAL

...it's what is left after the model has been fitted.

$$\mathbf{Z} \quad = \quad \mathbf{FA}' \quad + \quad \mathbf{E}$$

Communalities

- Consider the previous regression equation:

$$Z_j \leftarrow F_1 F_2 \ldots F_m$$

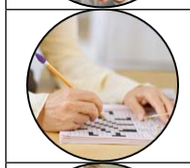
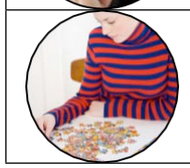
- There's an R^2 value for each variable.
  - How much of the variance of a variable is accounted for by the factors.
  - How much the variable has in common with the factors.
- It is called the *communality* of the variable, h^2.
- For orthogonal solutions,

$$h_j^2 = a_{j1}^2 + \cdots + a_{jm}^2$$

The sum of squared loadings

Communalities

$\mathbf{A}_{\mathbf{rotated}}$

	1	2
	0.23	0.88
	0.00	0.83
	-0.08	0.84
	0.94	-0.07
	0.82	0.28
	0.84	-0.04
Sum of Square Loadings	2.31	2.26
/6	38.5%	37.6%

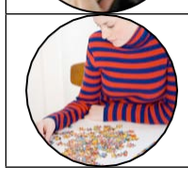
$$\sum a^2 = .23^2 + .88^2 = .836$$

That is, 83.6% of the variance in ‘lyric recall’ is accounted for by Factor 1 plus Factor 2.

This gives us an indication of how much ‘lyric recall’ has in common with the two factors.

Communalities

$\mathbf{A}_{\mathbf{rotated}}$

	1	2	h^2
	0.23	0.88	**0.836**
	0.00	0.83	**0.681**
	-0.08	0.84	**0.714**
	0.94	-0.07	**0.887**
	0.82	0.28	**0.748**
	0.84	-0.04	**0.699**
Sum of Square Loadings	2.31	2.26	**4.564**
/6	38.5%	37.6%	**76.11%**

$$\sum a^2 = .23^2 + .88^2 = .836$$

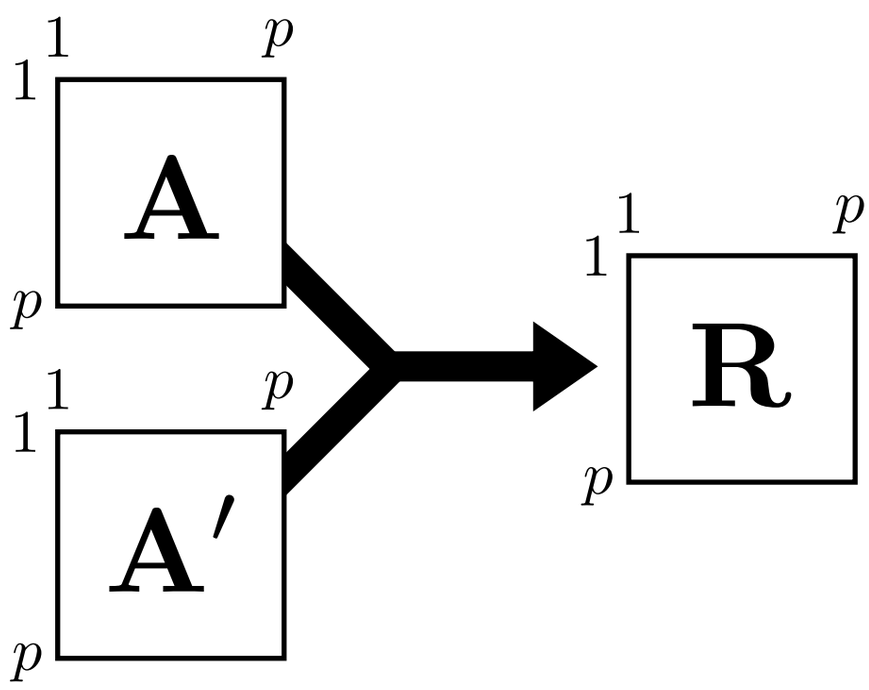
That is, 83.6% of the variance in 'lyric recall' is accounted for by Factor 1 plus Factor 2.

This gives us an indication of how much 'lyric recall' has in common with the two factors.

$$\sum a^2 = .23^2 + .00^2 + -.08^2 + .94^2 + .82^2 + .84^2 = 2.31 = \lambda_1$$

The reproduced correlation matrix

- For a full principal components analysis with p components retained, the original $\mathbf{R}$ matrix could be expressed in terms of the matrix $\mathbf{A}$.

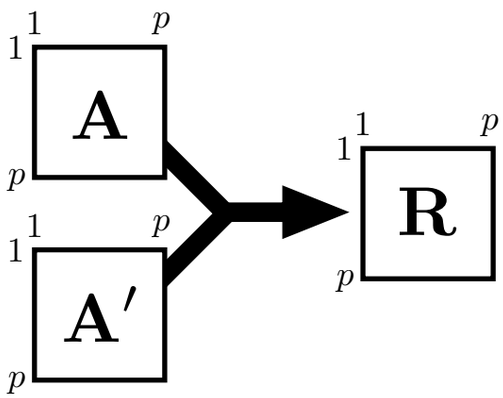


$$\mathbf{A}\mathbf{A}' = \mathbf{R}$$

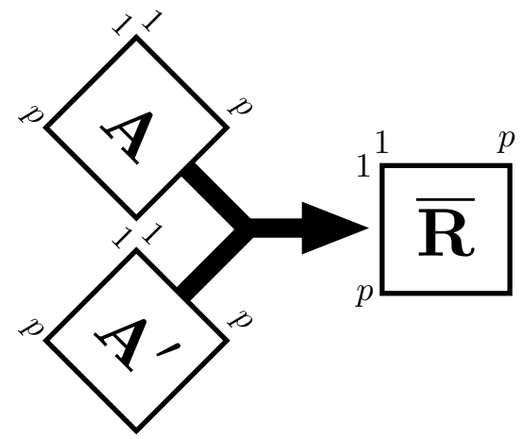
- In factor analysis, when *m* factors are retained, a reproduced correlation matrix can be calculated:

$$\mathbf{A}_{\mathbf{rotated}}\mathbf{A}'_{\mathbf{rotated}} = \overline{\mathbf{R}}$$

The reproduced correlation matrix



$$\mathbf{A}\mathbf{A}' = \mathbf{R}$$



$$\mathbf{A}_{\mathbf{rotated}}\mathbf{A}'_{\mathbf{rotated}} = \overline{\mathbf{R}}$$

Me again...

DATA = MODEL + RESIDUAL

...it's what is left after the model has been fitted.

$$\mathbf{R} = \overline{\mathbf{R}} + \mathbf{E}$$

$$\mathbf{R}_{\mathbf{full}} = \mathbf{R}_{\mathbf{reproduced}} + \mathbf{R}_{\mathbf{residual}}$$

The reproduced correlation matrix

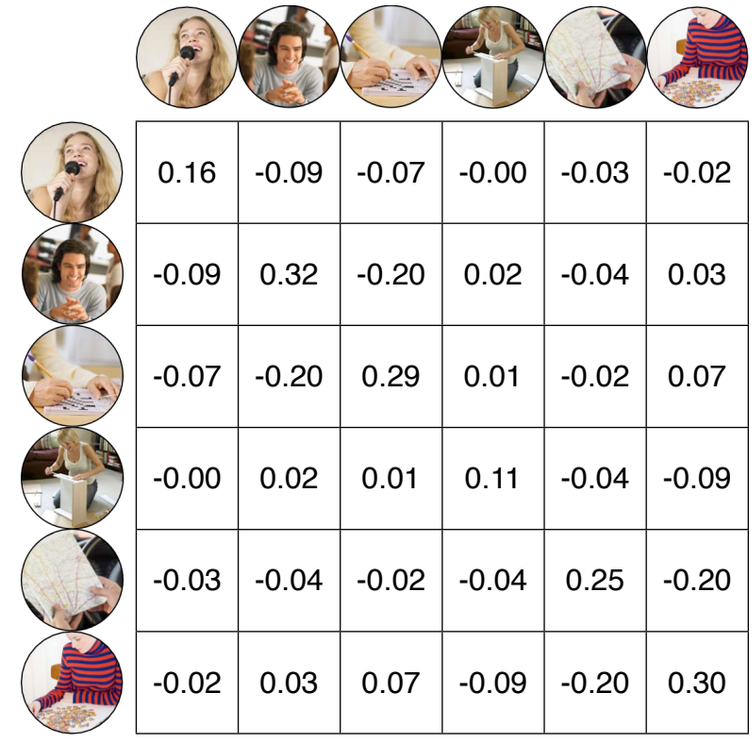
$$\mathbf{R}_{\mathbf{full}} = \mathbf{R}_{\mathbf{reproduced}} + \mathbf{R}_{\mathbf{residual}}$$

1	0.64	0.65	0.15	0.40	0.14
0.64	1	0.49	-0.04	0.19	-0.01
0.65	0.49	1	-0.13	0.15	-0.04
0.15	-0.04	-0.13	1	0.71	0.70
0.40	0.19	0.15	0.71	1	0.47
0.14	-0.01	-0.04	0.70	0.47	1

=

0.84	0.73	0.73	0.15	0.44	0.15
0.73	0.68	0.69	-0.05	0.23	-0.03
0.73	0.69	0.71	-0.14	0.17	-0.11
0.15	-0.05	-0.14	0.89	0.75	0.79
0.44	0.23	0.17	0.75	0.75	0.67
0.15	-0.03	-0.11	0.79	0.67	0.70

+



0.16	-0.09	-0.07	-0.00	-0.03	-0.02
-0.09	0.32	-0.20	0.02	-0.04	0.03
-0.07	-0.20	0.29	0.01	-0.02	0.07
-0.00	0.02	0.01	0.11	-0.04	-0.09
-0.03	-0.04	-0.02	-0.04	0.25	-0.20
-0.02	0.03	0.07	-0.09	-0.20	0.30

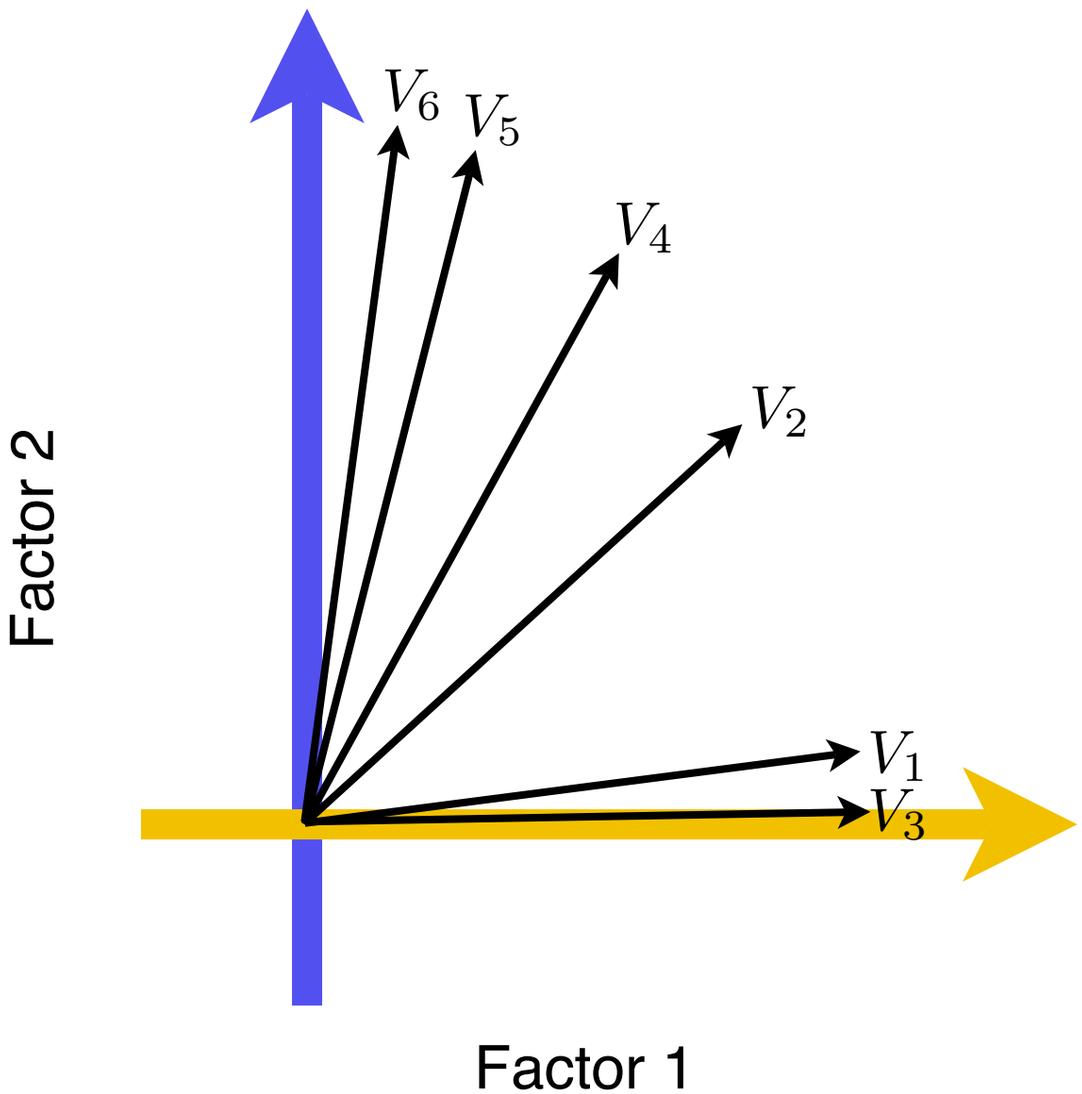
Note: The closer that these reproduced correlations match the original correlations, the better the factor analysis captured the relationship among the variables. This difference comes out in the residual correlations. The goal is to get these as small as possible.

Oblique Rotation

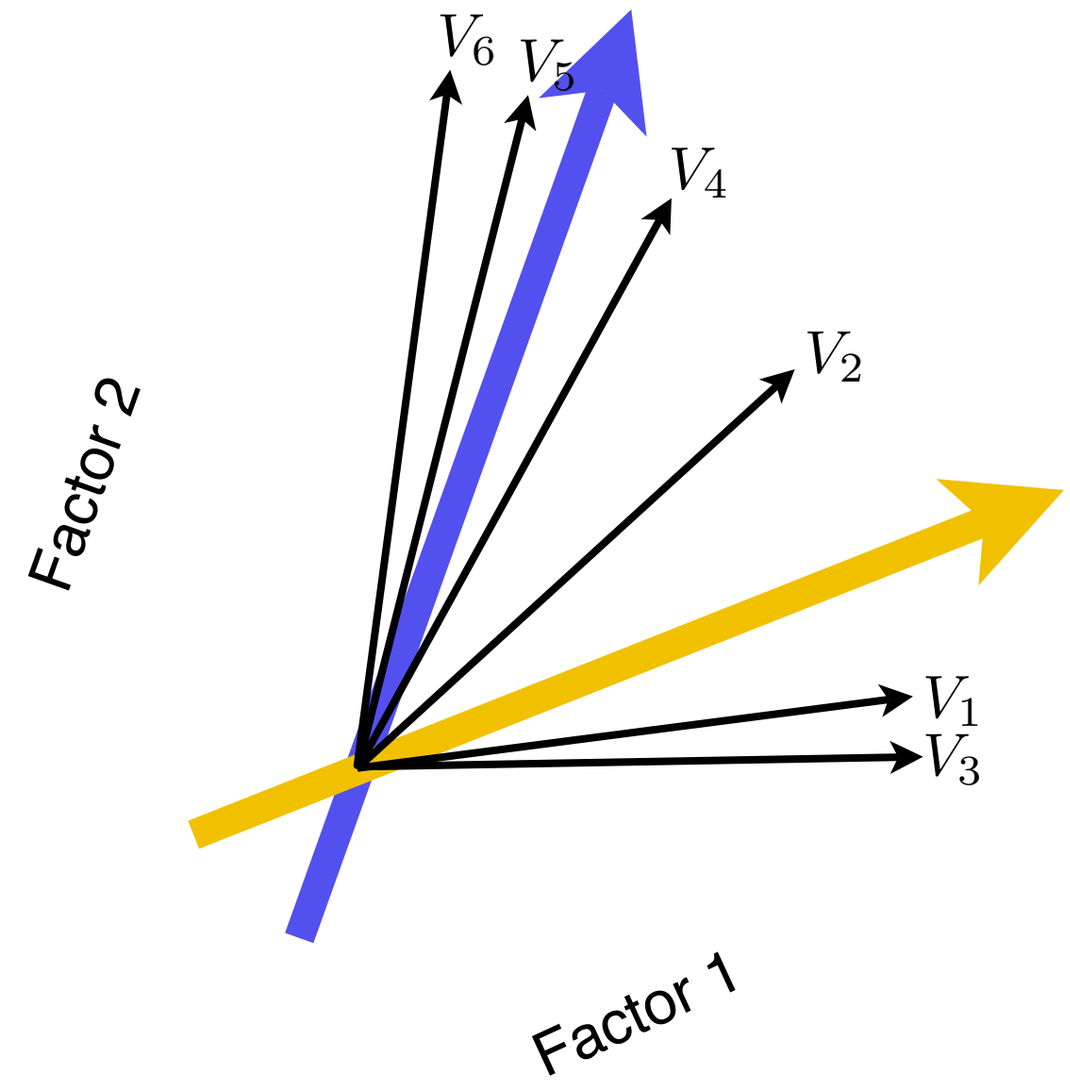
- The aim of a factor analysis is to arrive at a smaller number of interpretable factors. The two key words here are *smaller* and *interpretable*. This led to the two issues:
  - the number of factors problem; and
  - the rotation problem.
- The criteria for interpretability involved simple structure and an orthogonal rotation was one solution. However, an orthogonal solution may not be the most interpretable.

Oblique Rotation

Geometrical Representation



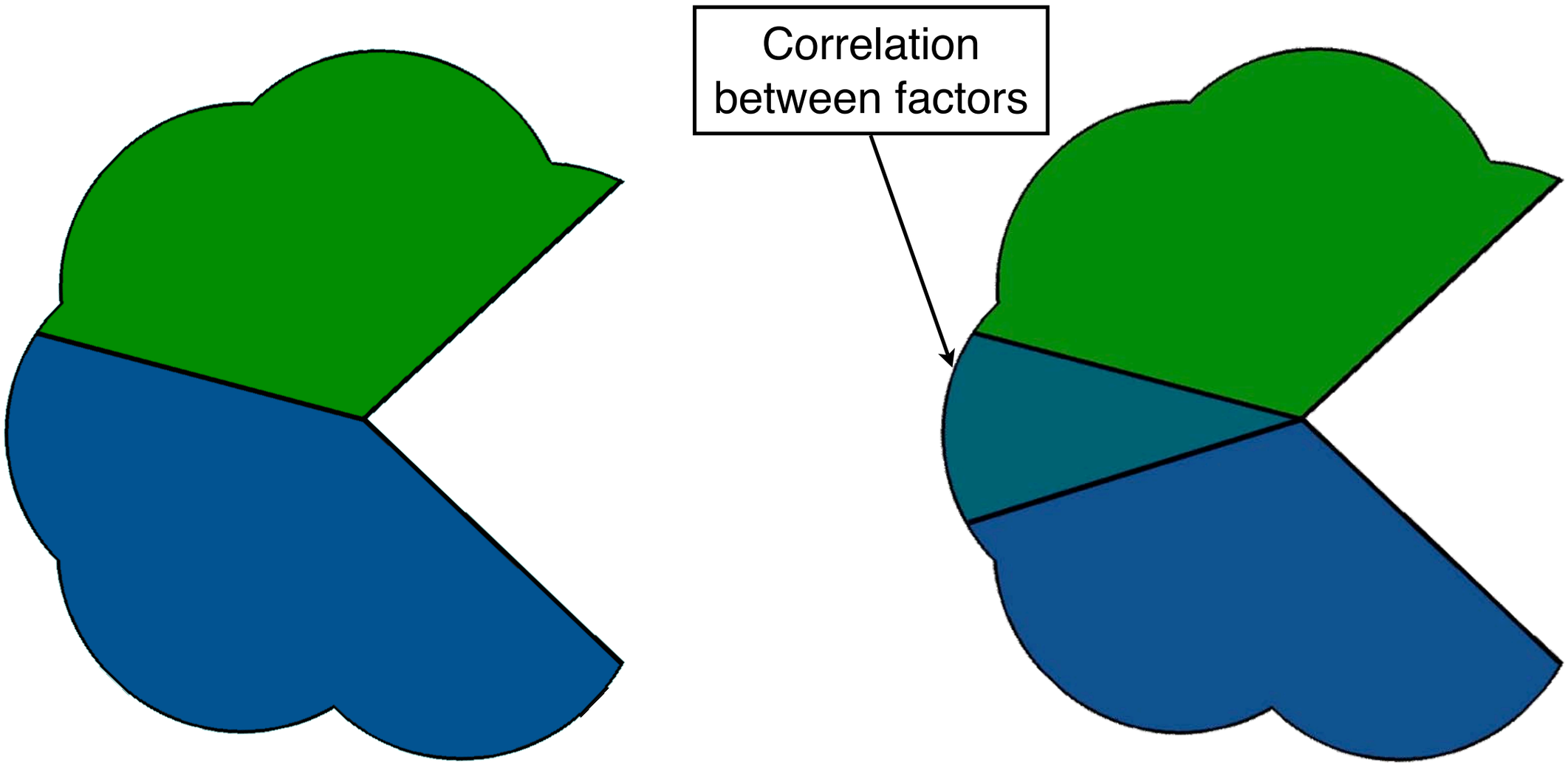
Orthogonal Rotation



Oblique Rotation

Oblique Rotation

Euler Representation



Orthogonal Rotation

Oblique Rotation

Oblique Rotation

Matrix Representation

$$\mathbf{A_{rotated}} = \mathbf{A_{unrotated}}\mathbf{\Lambda}$$

$$\mathbf{R_{reproduced}} = \mathbf{C}\mathbf{A}' = \mathbf{A}\mathbf{\Phi}\mathbf{A}'$$

$\mathbf{\Phi} =$ Factor correlation matrix

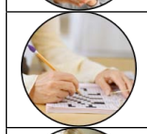
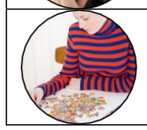
$\mathbf{C} =$ Structure matrix

$\mathbf{A} =$ Pattern matrix

Oblique Rotation

Matrix Representation

$\mathbf{A}_{\mathbf{obliqueRot}}$

	1	2
	0.19	0.88
	-0.03	0.83
	-0.12	0.85
	0.95	-0.11
	0.81	0.24
	0.84	-0.08

Once we've generated the loading matrix which becomes the pattern matrix with an oblique rotation, we can generate the factor scores in the same way as we did previously:

$\mathbf{R}^{-1}$

2.70	-1.03	-1.18	-0.13	-0.61	-0.05
-1.03	1.76	-0.15	0.24	-0.08	0.02
-1.18	-0.15	1.95	0.63	-0.16	-0.12
-0.13	0.24	0.63	3.47	-1.82	-1.52
-0.61	-0.08	-0.16	-1.82	2.50	0.17
-0.05	0.02	-0.12	-1.52	0.17	1.99

$\times$

$\mathbf{A}_{\mathbf{obliqueRot}}$

	1	2
	0.19	0.88
	-0.03	0.83
	-0.12	0.85
	0.95	-0.11
	0.81	0.24
	0.84	-0.08

$=$

$\mathbf{B}$

	1	2
	0.04	0.38
	-0.06	0.37
	-0.10	0.39
	0.42	-0.10
	0.34	0.07
	0.37	-0.08

$$\mathbf{R}^{-1}\mathbf{A} = \mathbf{B}$$

and then the factor scores:

$\mathbf{Z}$

Participant						
1	-0.10	-1.62	-1.78	0.78	1.45	-0.16
2	0.50	-0.19	0.31	-0.07	-0.28	0.37
3	-1.22	-0.32	-1.77	0.93	-0.28	0.37
4	0.77	1.10	-0.62	-0.99	-1.15	-0.30
5	0.34	0.25	1.16	0.87	0.55	0.86
6	0.17	0.24	-0.51	1.31	0.96	1.65
7	1.24	0.76	-1.05	1.35	0.70	1.70
8	0.63	0.37	-0.68	-1.38	-0.17	-1.21
9	0.59	1.30	1.56	-0.90	-0.53	-0.01
10	0.97	1.55	-0.21	0.07	-0.06	-0.18
11	0.74	0.35	-0.32	1.47	1.06	1.28
12	-0.50	0.24	0.02	0.47	0.32	-0.14
13	0.56	0.72	-0.58	-0.49	0.36	-1.63
14	0.63	0.87	0.13	0.02	-0.37	0.12
15	-0.37	-0.60	-0.88	1.58	1.08	1.70
16	0.98	1.62	0.52	-0.72	-1.77	-0.86
17	-0.62	-1.29	-0.06	-0.22	-0.41	-0.60
18	1.98	0.38	1.18	0.96	1.11	0.09
19	0.75	-0.67	1.19	2.56	2.06	2.75
20	-1.47	-0.10	-1.28	-0.02	0.63	-0.60
⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	0.36	0.69	-1.27	0.32	-0.11	-0.14

$\times$

$\mathbf{B}$

	1	2
	0.04	0.38
	-0.06	0.37
	-0.10	0.39
	0.42	-0.10
	0.34	0.07
	0.37	-0.08

$=$

$\mathbf{F}$

Participant	1	2
1	1.02	-1.29
2	0.01	0.20
3	0.58	-1.41
4	-0.90	0.51
5	0.76	0.56
6	1.54	-0.24
7	1.54	0.13
8	-1.02	0.34
9	-0.77	1.36
10	-0.09	0.87
11	1.50	0.12
12	0.22	-0.11
13	-0.65	0.47
14	-0.11	0.58
15	1.78	-0.93
16	-1.33	1.20
17	-0.40	-0.70
18	0.76	1.33
19	2.76	0.17
20	0.06	-1.00
	⋮	⋮
100	0.14	-0.12

$$\mathbf{ZB} = \mathbf{F}$$

Oblique Rotation

Matrix Representation

Now - because we've relaxed the requirement of orthogonal factors by using an oblique rotation, the factors are now correlated with one another (i.e., the angle between the factors is no longer 90˚). And so we can see what the correlation is between our factors:

$$\left(\frac{1}{N-1}\right) \times \mathbf{F}' \times \mathbf{F} = \mathbf{\Phi}$$

$\mathbf{F}'$

Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	⋯	100
1	1.02	0.01	0.58	-0.90	0.76	1.54	1.54	-1.02	-0.77	-0.09	1.50	0.22	-0.65	-0.11	1.78	-1.33	-0.40	0.76	2.76	0.06	⋯	0.14
2	-1.29	0.20	-1.41	0.51	0.56	-0.24	0.13	0.34	1.36	0.87	0.12	-0.11	0.47	0.58	-0.93	1.20	-0.70	1.33	0.17	-1.00	⋯	-0.12

$\mathbf{F}$

Participant	1	2
1	1.02	-1.29
2	0.01	0.20
3	0.58	-1.41
4	-0.90	0.51
5	0.76	0.56
6	1.54	-0.24
7	1.54	0.13
8	-1.02	0.34
9	-0.77	1.36
10	-0.09	0.87
11	1.50	0.12
12	0.22	-0.11
13	-0.65	0.47
14	-0.11	0.58
15	1.78	-0.93
16	-1.33	1.20
17	-0.40	-0.70
18	0.76	1.33
19	2.76	0.17
20	0.06	-1.00
	⋮	⋮
100	0.14	-0.12

$\mathbf{\Phi}$

	1	2
1	1.00	0.09
2	0.09	1.00

Here we see that the correlation between the factors is rather low. The factors are *almost* orthogonal which is substantially less complex.

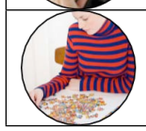
As T&F indicate: Ordinarily one uses orthogonal rotation in a case like this because complexities introduced by oblique rotation are not warranted by such a low correlation among factors.

Oblique Rotation

Matrix Representation

However, if we *do* use an oblique rotation, we also get the structure matrix $\mathbf{C}$, which represents the correlations between the variables and the factors.

$\mathbf{A}_{\mathbf{obliqueRot}}$

	1	2
	0.19	0.88
	-0.03	0.83
	-0.12	0.85
	0.95	-0.11
	0.81	0.24
	0.84	-0.08

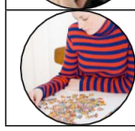
$\times$

$\mathbf{\Phi}$

	1	2
1	1.00	0.09
2	0.09	1.00

$=$

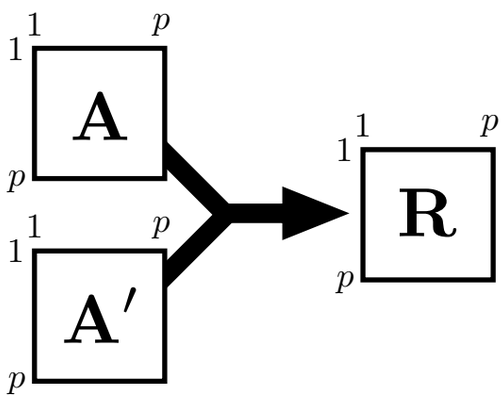
$\mathbf{C}$

	1	2
	0.11	0.87
	-0.11	0.84
	-0.20	0.87
	0.96	-0.20
	0.79	0.17
	0.85	-0.16

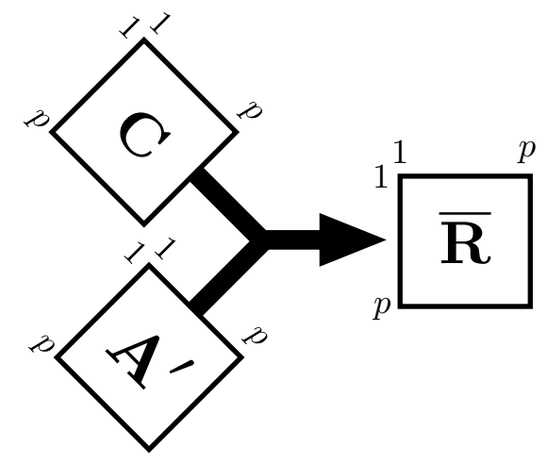
As with the orthogonal rotation, we can generate the “reproduced correlation matrix”...

Oblique Rotation

Matrix Representation



$$\mathbf{A}\mathbf{A}' = \mathbf{R}$$



$$\mathbf{C}\mathbf{A}'_{\mathbf{obliqueRot}} = \overline{\mathbf{R}}$$

Me again...

DATA = MODEL + RESIDUAL

...it's what is left after the model has been fitted.

$$\mathbf{R} = \overline{\mathbf{R}} + \mathbf{E}$$

$$\mathbf{R}_{\mathbf{full}} = \mathbf{R}_{\mathbf{reproduced}} + \mathbf{R}_{\mathbf{residual}}$$

Oblique Rotation

Matrix Representation

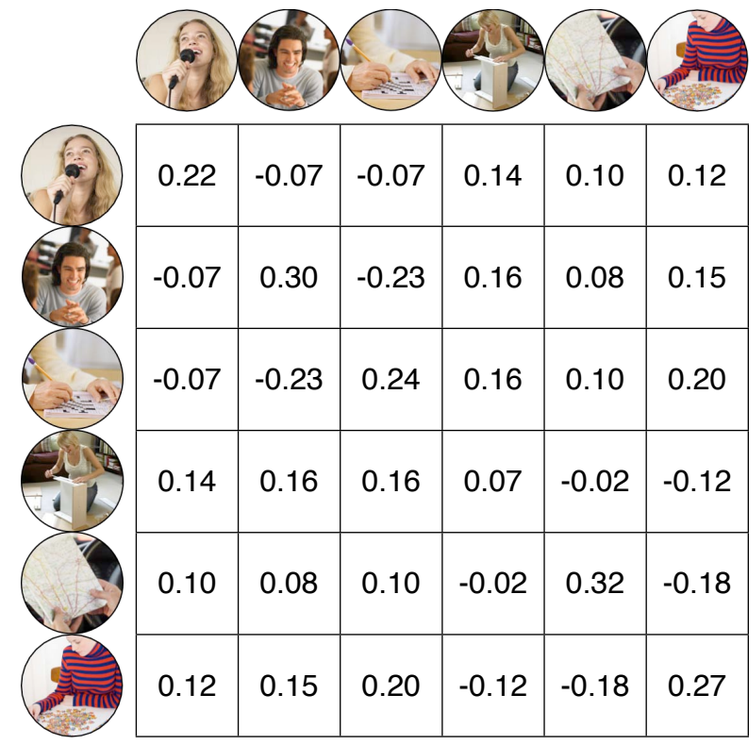
$\mathbf{R_{full}} = \mathbf{R_{reproduced}} + \mathbf{R_{residual}}$

1	0.64	0.65	0.15	0.40	0.14
0.64	1	0.49	-0.04	0.19	-0.01
0.65	0.49	1	-0.13	0.15	-0.04
0.15	-0.04	-0.13	1	0.71	0.70
0.40	0.19	0.15	0.71	1	0.47
0.14	-0.01	-0.04	0.70	0.47	1

$=$

0.78	0.71	0.72	0.01	0.30	0.02
0.71	0.70	0.72	-0.20	0.12	-0.16
0.72	0.72	0.76	-0.29	0.05	-0.24
0.01	-0.20	-0.29	0.93	0.73	0.83
0.30	0.12	0.05	0.73	0.68	0.65
0.02	-0.16	-0.24	0.83	0.65	0.73

$+$



0.22	-0.07	-0.07	0.14	0.10	0.12
-0.07	0.30	-0.23	0.16	0.08	0.15
-0.07	-0.23	0.24	0.16	0.10	0.20
0.14	0.16	0.16	0.07	-0.02	-0.12
0.10	0.08	0.10	-0.02	0.32	-0.18
0.12	0.15	0.20	-0.12	-0.18	0.27

Note: The closer that these reproduced correlations match the original correlations, the better the factor analysis captured the relationship among the variables. This difference comes out in the residual correlations. The goal is to get these as small as possible.

To Recap

Steps in Factor Analysis

a. Selecting and measuring the set of variables, (designing the study).

Often a FA or a PCA is used in an attempt to ‘save’ poorly designed or conceived research. Any variables can be correlated and ‘factored’ and PCA and FA have the ability to create apparent order from real chaos. This has given factor analysis a bad name.

b. Preparing the correlation matrix

The honesty of correlations, and thus the honesty and trustworthiness of the chosen factor solution, depends on the sample size, symmetry of the frequency distributions of the variables (skewness), linearity of the relationships and outliers among the cases. In line with the general policy on data diagnostics, if transformations are used or cases dropped, the analysis should be run with and without the changes to assess the effects of distributional problems and outliers.

To Recap

Steps in Factor Analysis

c. Extracting a set of factors from the correlation matrix, (choice of factor model).

The method of factor extraction depends on the assumptions the researcher makes about the nature of the variance of an item or variable. This leads to the use of a Common Factor Model or a Principal Components Model. We have only discussed the PCA model so far.

d. Determining the number of factors to retain

This is a difficult decision and is influenced not only by the numerical information but also by the nature of the variables and by the substantive theory underlying the study.

To Recap

Steps in Factor Analysis

e. Rotating the factors to increase interpretability (choice of rotation method, orthogonal or oblique)

After extraction, there are an infinite number of rotations available, all accounting for the same amount of variance in the original data but with the factors defined differently.

f. Interpreting the results

There is no criterion beyond interpretability against which to test the solution. A good factor analysis ‘makes sense’. Interpreting the results of PCA and/or FA is often said to be more related to an art than a science.