

Judgment's Studies

Computational Social Intelligence - Lecture 09

Prof. Alessandro Vinciarelli
School of Computing Science &
Institute of Neuroscience and Psychology

<http://www.dcs.gla.ac.uk/vincia>
Alessandro.Vinciarelli@glasgow.ac.uk



University
of Glasgow

EPSRC

Engineering and Physical Sciences
Research Council



This lecture is based on the following text (available on Moodle):

- R.R.Rosenthal, "Conducting Judgment Studies: Some Methodological Issues", in "The New Handbook of Methods in Nonverbal Behavior Research", J.A.Harrigan, R.Rosenthal and K.R.Scherer (eds.), pp 199-211 (included), pp 213-214 (Cronbach's Alpha), 2008.

Outline

- Introduction
- Reliability
- Conclusions

Outline

- Introduction
- Reliability
- Conclusions

The Big-Five Inventory 10

ID	Item	SD	D	NA	A	SA
1	I am reserved					
2	I am generally trusting					
3	I am lazy					
4	I am relaxed, I handle stress well					
5	I have few artistic interests					
6	I am outgoing, sociable					
7	I tend to find faults with others					
8	I do a thorough job					
9	I get nervous easily					
10	I have an active imagination					

Rammstedt and John, "Measuring Personality in One Minute or Less: A 10-item short version of the BFI", Journal of Research in Personality, 41(1): 203-212, 2007

The Big-Five Inventory 10

ID	Item	SD	D	NA	A	SA
1	This person is reserved					
2	... is generally trusting					
3	... is person lazy					
4	... is relaxed, handles stress well					
5	... has few artistic interests					
6	... is outgoing, sociable					
7	... tends to find faults with others					
8	... does a thorough job					
9	... gets nervous easily					
10	... has an active imagination					

Rammstedt and John, "Measuring Personality in One Minute or Less: A 10-item short version of the BFI", Journal of Research in Personality, 41(1): 203-212, 2007

Judgment Studies

“The term ‘judgment studies’ refers most generally to those studies in which behaviors, persons, objects or concepts are evaluated by one or more judges, raters, coders, or categorizers, referred to collectively as ‘judges’.”

R.R.Rosenthal, “Conducting Judgment Studies: Some Methodological Issues”,
in “The New Handbook of Methods in Nonverbal Behavior Research”,
J.A.Harrigan, R.Rosenthal and K.R.Scherer (eds.), 2008.

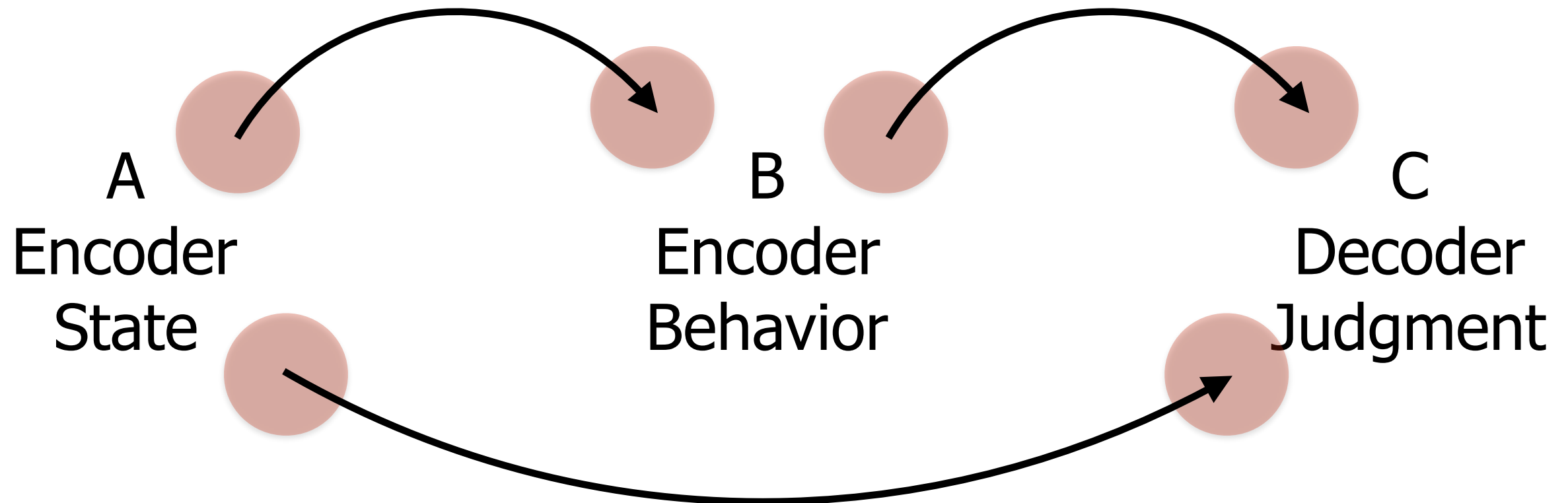
Types of Judgment Studies

Dimensions	Examples
Type of Variable	Dependent vs Independent
Measurement Units	Physical vs Psychological
Reliability	Lower vs Higher
Social Meaning	Lower vs Higher

R.R.Rosenthal, "Conducting Judgment Studies: Some Methodological Issues",
in "The New Handbook of Methods in Nonverbal Behavior Research",
J.A.Harrigan, R.Rosenthal and K.R.Scherer (eds.), 2008.

How does the
encoder manifest her/
his state through her/
his behaviour?

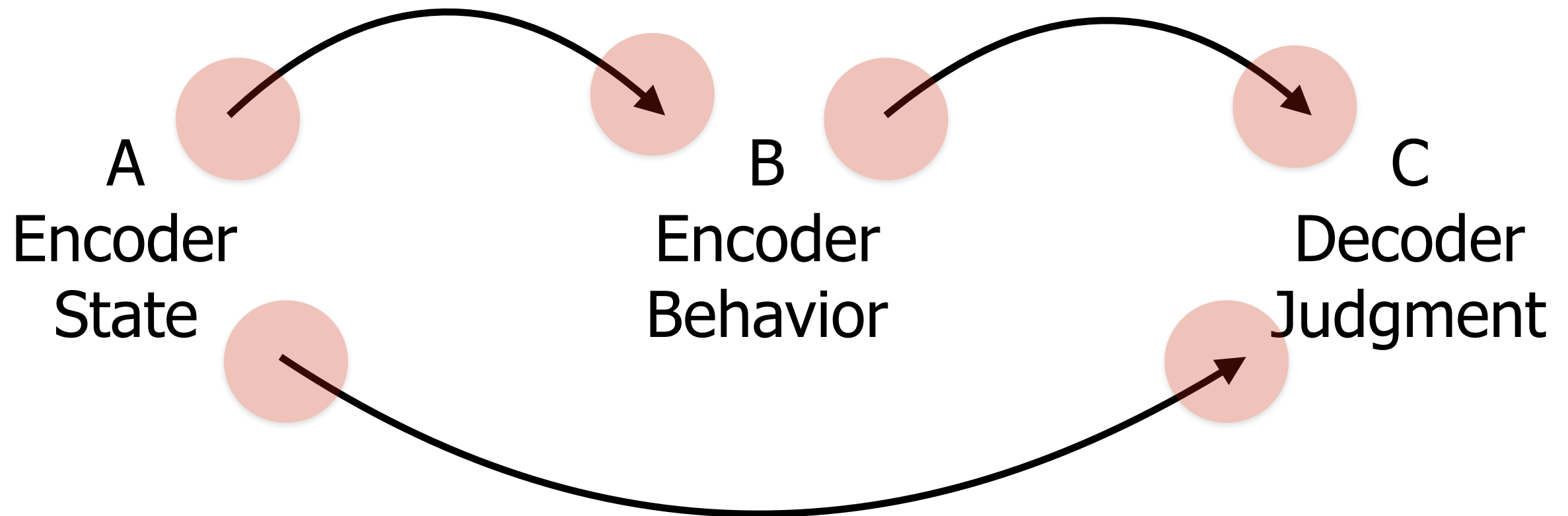
How do decoders
interpret the
behaviour of the
encoder



What is the state the
decoder attributes to
the encoder?

The State is the
independent variable,
the behaviour the
dependent one

The behaviour is the
independent variable,
the judgment the
dependent one



The State is the
independent variable,
the judgment is the
dependent one

Main Issues in Judgment Studies

- Reliability: How reliable are the judgments?
How many judges are necessary to obtain reliable judgments?
- Selection: How to select the judges?
- Composition: How to combine different judgments to form composite variables?

Outline

- Introduction
- **Reliability**
- Conclusions

Reliability

- The consensus among multiple judges suggests that there is consistency between observations and judgments;
- The reliability can be thought of as the measure of the consensus among multiple judges;
- In principle, the higher the consensus, the higher the reliability.

Percentage of times
two judges agree

Number of times two
judges agree

The diagram shows the formula $R = \left(\frac{A}{A + D} \right) 100$. Three arrows point from descriptive text to parts of the formula: one from 'Percentage of times two judges agree' to R , one from 'Number of times two judges agree' to the numerator A , and one from 'Number of times two judges disagree' to the denominator $A + D$. The variables A and D are each enclosed in a light red circle.

$$R = \left(\frac{A}{A + D} \right) 100$$

Number of times two
judges disagree

Decision of Judge A



	Frown	No-Frown
Frown	98	1
No-Frown	1	0

Decision of Judge B

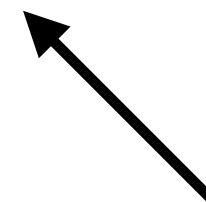


Decision of Judge C



	Frown	No-Frown
Frown	49	1
No-Frown	1	49

Decision of Judge D



$$R = 98\%$$

Decision of Judge A
(1 for Frown and -1
for No-Frown)

Decision of Judge A
(1 for Frown and -1
for No-Frown)

$$\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

$$\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}$$

Average decision of
Judge A

Average decision of
Judge B

Decision of Judge C
(1 for Frown and -1
for No-Frown)

Decision of Judge D
(1 for Frown and -1
for No-Frown)

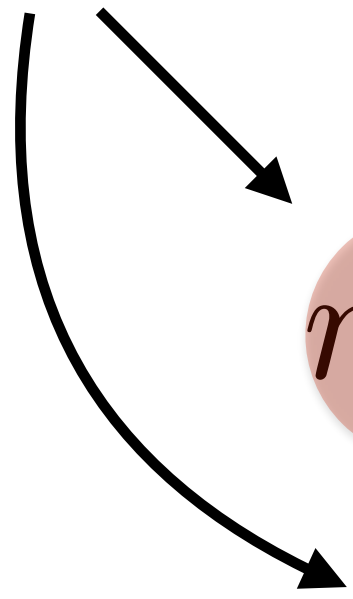
$$\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

$$\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}$$

Average decision of
Judge C

Average decision of
Judge D

The level of agreement is the same but the correlation is different



$r_{AB} = -0.01$

$r_{CD} = 0.96$

The correlation takes into account the variance in the judgment

Limits

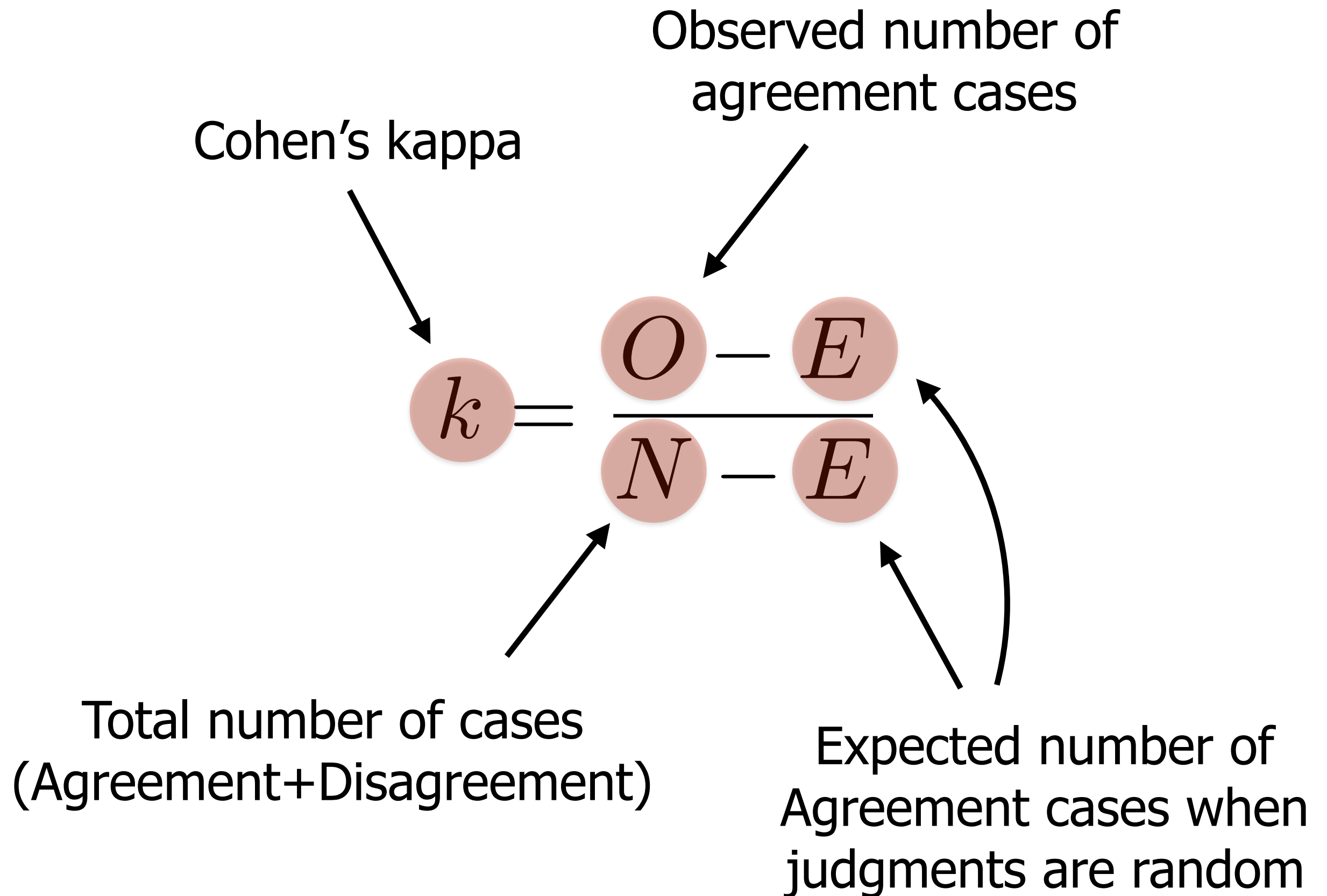
- The percentage of Agreement can be high simply because there is no variance in the judgments;
- If there is no variance, it is not possible to say what happens when there are different judgments;
- The same value of R can correspond to different values of correlation.

The decision of
Judge A

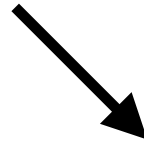
	Schizophrenic	Neurotic	Normal	Brain Damaged
Schizophrenic	13	0	0	12
Neurotic	0	12	13	0
Normal	0	13	12	0
Brain Damaged	12	0	0	13

The decision of
Judge B

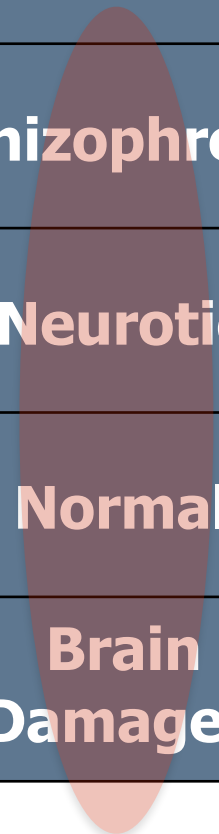
Agreement



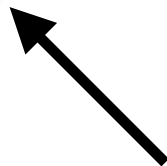
The decision of
Judge A



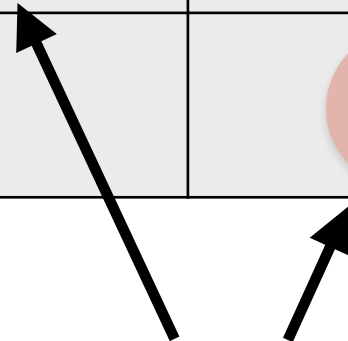
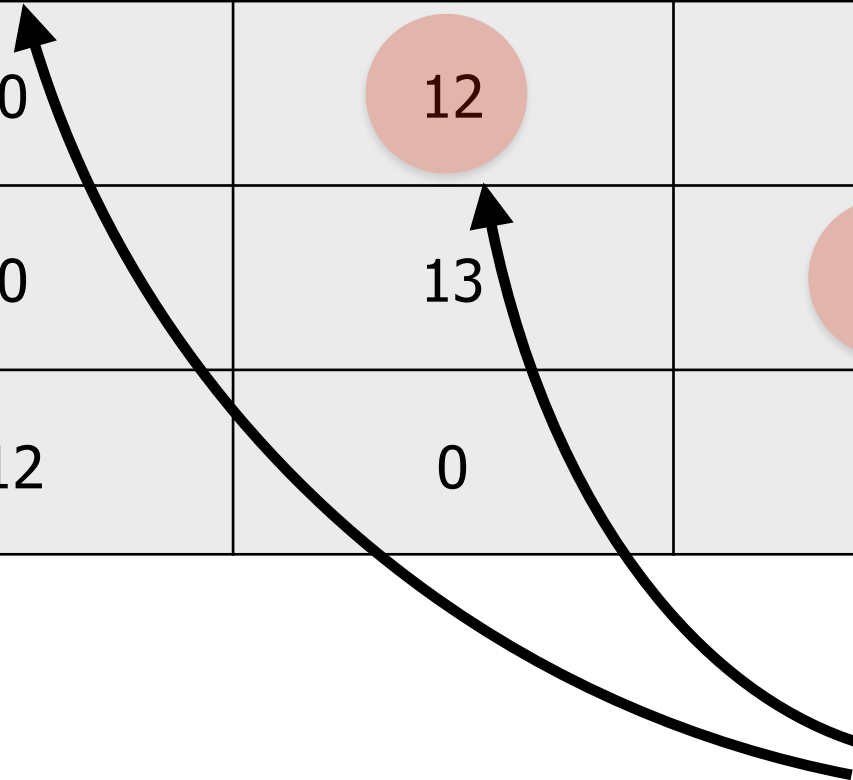
	Schizophrenic	Neurotic	Normal	Brain Damaged
Schizophrenic	13	0	0	12
Neurotic	0	12	13	0
Normal	0	13	12	0
Brain Damaged	12	0	0	13



The decision of
Judge B



Agreement



Expected number of
agreement cases in
cell "ii"

Marginal of Column
and Row "i"

The diagram illustrates the formula for the expected number of agreement cases in cell "ii", E_{ii} . The formula is presented in two equivalent forms: $E_{ii} = \frac{R_i}{N} \frac{C_i}{N} N$ and $E_{ii} = \frac{R_i C_i}{N}$. Each term in the formula is enclosed in a red oval. Arrows from the text labels point to these terms: "Expected number of agreement cases in cell 'ii'" points to E_{ii} ; "Marginal of Column and Row 'i'" points to R_i and C_i ; "Probability of falling in Column 'i' and Row 'i'" points to the two N terms in the first form; and "Total number of cases" points to the N term in the second form.

$$E_{ii} = \frac{R_i}{N} \frac{C_i}{N} N = \frac{R_i C_i}{N}$$

Probability of falling
in Column "i" and
Row "i"

Total number of cases

$$k = \frac{O - E}{N - E} = \frac{50 - 25}{100 - 25} = 0.\bar{3}$$

The decision of
Judge A

	Schizophrenic	Neurotic	Normal	Brain Damaged
Schizophrenic	13	0	0	12
Neurotic	0	12	13	0
Normal	0	13	12	0
Brain Damaged	12	0	0	13

The decision of
Judge B

$k=0.04$

Limits

- The value of k compares the observed agreement and the agreement expected when the judgments are random;
- When there are more than two categories (2x2 tables), it is not clear whether the kappa value applies equally to all of them.

The average
correlation between
judges

The correlation
between judges "i"
and "j"

$$r = \frac{2 \sum_{i=1}^N \sum_{j=i+1}^N r_{ij}}{N(N-1)}$$

The number of judges

The Effective (or
Spearman Brown)
Reliability

The average
correlation between
judges

The diagram illustrates the Spearman-Brown Reliability formula, $R_{SB} = \frac{Nr}{1 + (N - 1)r}$. The formula is presented with its components highlighted by red circles and annotated with arrows. The left side of the equation, R_{SB} , is enclosed in a red circle and labeled 'The Effective (or Spearman Brown) Reliability'. The right side of the equation is a fraction. The numerator, Nr , is enclosed in a red circle and labeled 'The average correlation between judges'. The denominator, $1 + (N - 1)r$, is also enclosed in a red circle. Within the denominator, the term $(N - 1)r$ is further highlighted by a red circle and labeled 'The number of judges'. Arrows point from the descriptive text to their respective parts in the formula.

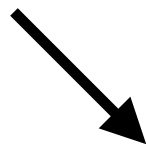
$$R_{SB} = \frac{Nr}{1 + (N - 1)r}$$

The number of judges

Limits

- The effective reliability provides an indication of a how much associated are the judgments of two random judges;
- It is an average value that does not say whether all judges are equally correlated with one another.

The Three Judges



Encoders	A	B	C	Total
1	5	6	7	18
2	3	6	4	13
3	3	4	6	13
4	2	2	3	7
5	1	4	4	9

$$S_{tot}^2$$

$$S_B^2$$



The variance of the scores for one judge

The variance of the total for each encoder

The Cronbach's alpha

The number of judges

Sum over all judges

The variance of the total for each encoder

The variance of the scores for one judge

$$\alpha = \left(\frac{N}{N-1} \right) \frac{S_{tot}^2 - \sum_j S_j^2}{S_{tot}^2}$$

Limits

- It avoids the calculation of multiple correlations when the number of judges is high;
- It tends to give the same values as the other reliability measures considered in this lecture (it is affected by the same limitations).

Outline

- Introduction
- Reliability
- **Conclusions**

Conclusions

- Judgment studies allow one to answer questions on how inner states are expressed and perceived;
- Reliability measures are expected to quantify the extent to which multiple judgments agree with one another;
- It is not sufficient that multiple judges agree, they must have high mutual correlation.

Thank You!