Holger Hopp

University of Groningen

Magnitude Estimation

Outline of the talk

- Magnitude Estimation: The background
- ME and linguistic judgements
- ME: The implementation
- ME: A study
- ME: Perspectives and Problems

Grammaticality Judgements

- central in linguistic theory

grammaticality - acceptability - acceptability judgement

Grammaticality Judgements

- idealization of binarity:

  grammatical vs ungrammatical

- gradedness / relative acceptability
  - pragmatic/context-induced
  - frequency
  - prescriptive knowledge
  - preferences
  - ….

(Schuetze 1996)

Grammaticality Judgements

- Chomsky (1975: 131)

“An adequate linguistic theory will have to recognize degrees of grammaticalness.”

Grammaticality Judgements

- The reality so far

Informal ordinal scales

e.g.: 0 - ? - ?? - ?* - * - **

0 - ? - ?? - ?* - * - **

- informal; i.e. no generally established categories (Featherston 2004)
- absolute rating, rather than relative ranking
- no information on relative distances
- limited scale (2-to-6-point scales)
- limited suitability for statistical analysis (in particular, parametric statistics)

An interval scale of linguistic acc. judgements

Magnitude Estimation

(e.g. Bard et al. 1996; Cowart 1997)

Magnitude Estimation in Psychophysics

- Stevens (1956, 1975)

  Proportional associations of numerical judgements with (perceived) physical stimulus, e.g. Line length, brightness, loudness, electric shock, etc.

10

14

Magnitude Estimation in Psychophysics

- Stevens (1956, 1975)

  Proportional associations of numerical judgements with (perceived) physical stimulus, e.g. Line length, brightness, loudness, etc.

Psychophysical Law:

$$\psi = kS^n$$

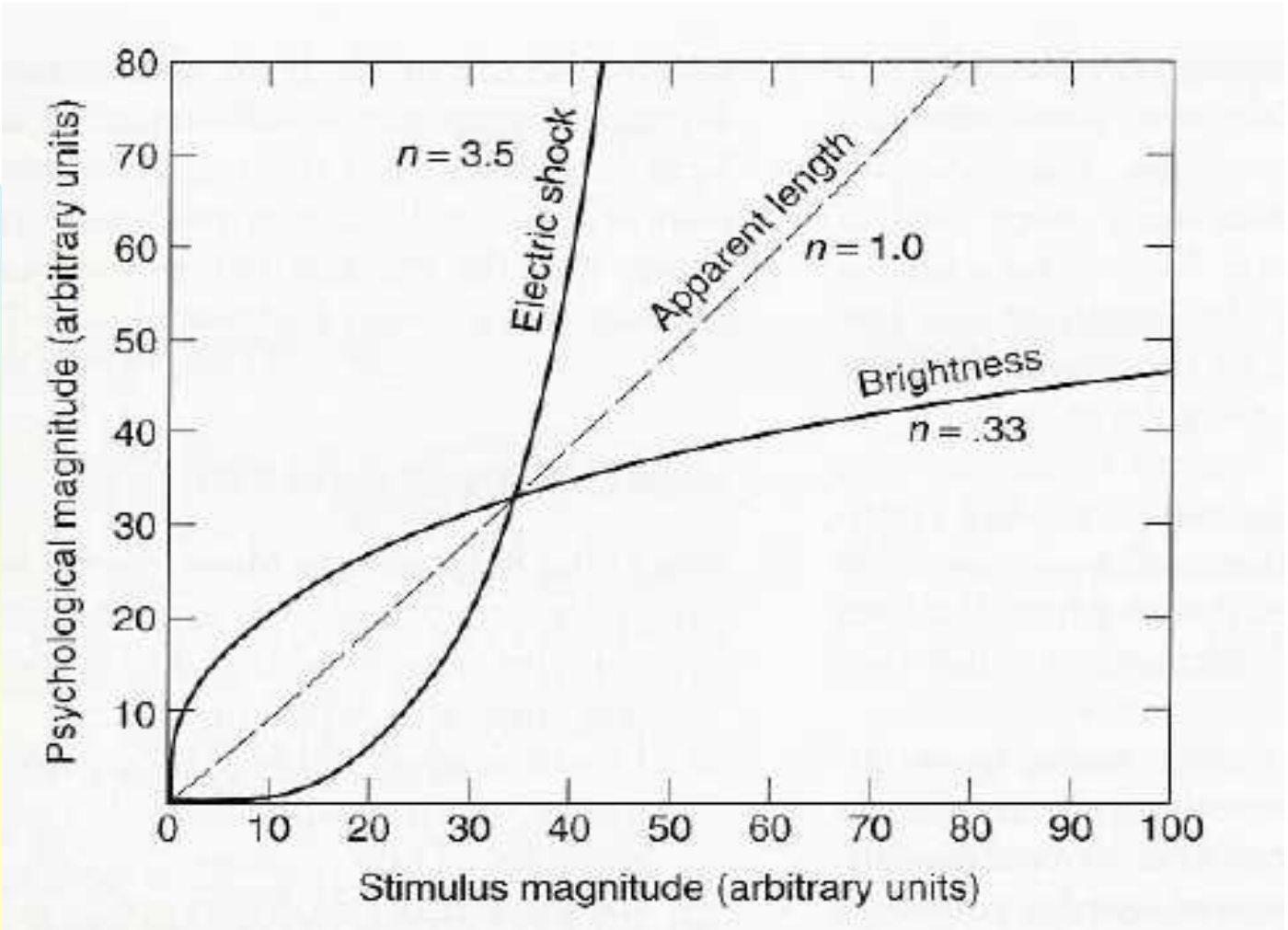
ψ - perceived stimulus magnitude

S - physical magnitude

k - constant

n - characteristic of a given modality (e.g. brightness 0.5; line length 1.0)

Psychophysical ME judgements



Magnitude Estimation in Psychophysics

Psychophysical Law:

$\psi = kS^n$

Power Law:

$\log\psi = \log k + n\log S$

ψ - perceived stimulus magnitude

S - physical magnitude

k - constant

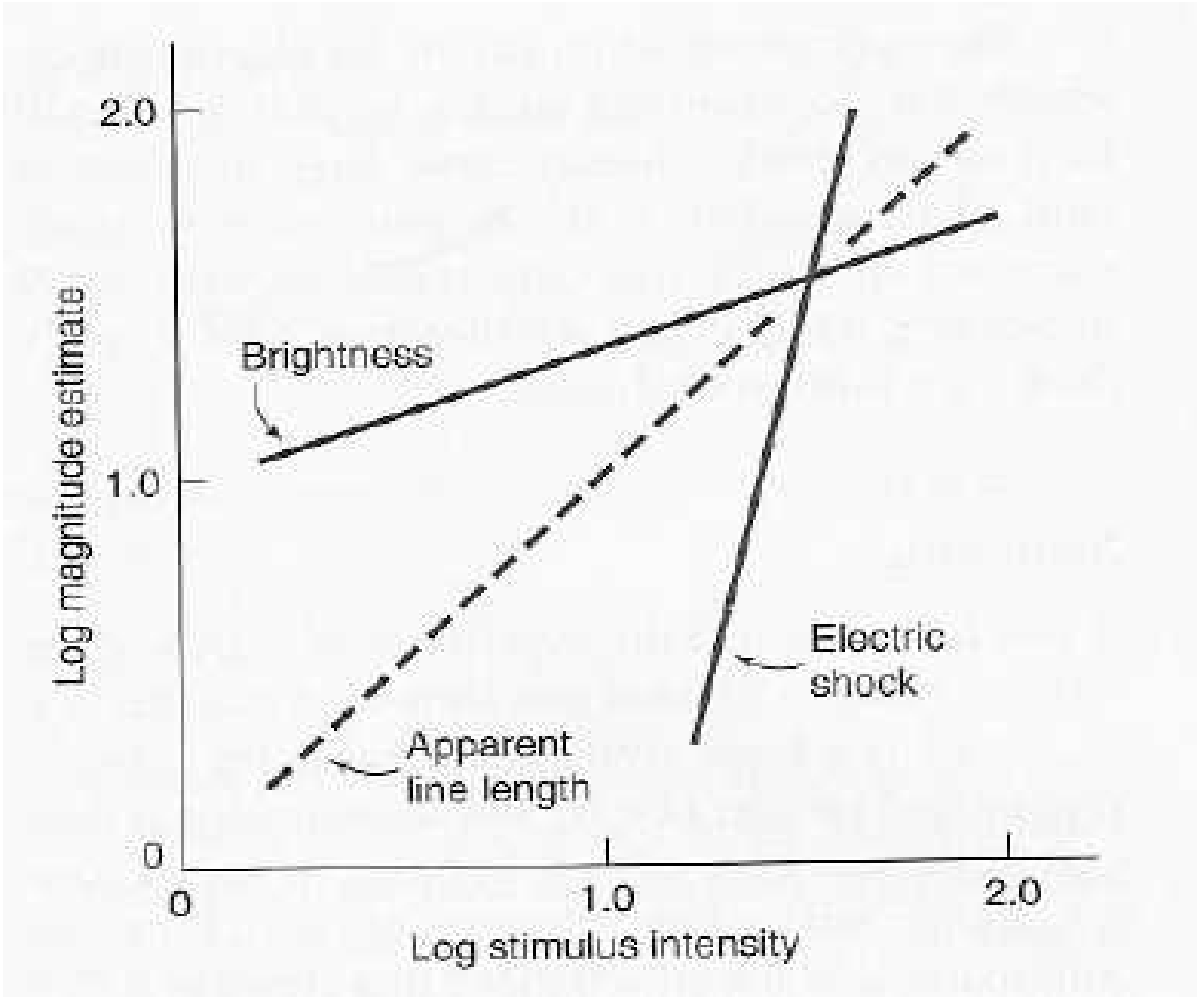
n - characteristic of a given modality (e.g. brightness 0.5; line length 1.0)

The Power Law in Psychophysics

Brightness: n=0.33

Line length: n=1.0

Electric shock: n=3.5



Magnitude Estimation & Linguistic Judgements

- Main difference between psychophysics and linguistics:

No objective measure of comparison for status of linguistic objects

N.B.: Corpus frequency does NOT reflect acceptability patterns

Magnitude Estimation & Linguistic Judgements

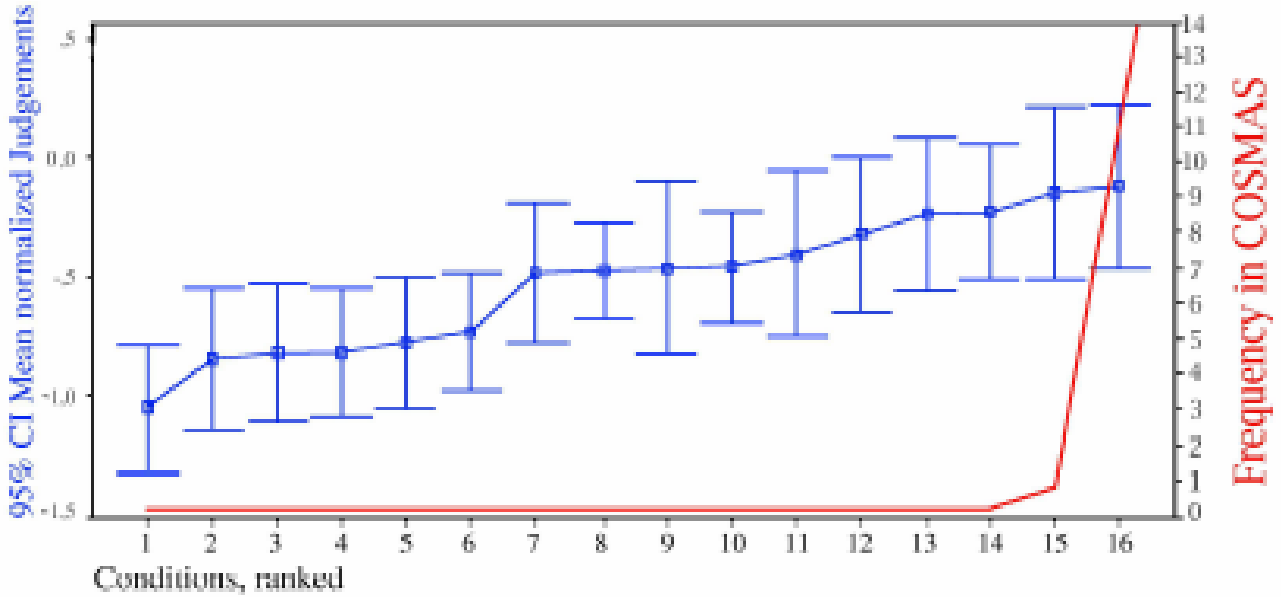
- Featherston (2004):

reflexive pronouns & German ditransitives

COSMAS, IDS (531 million word forms)

ihn ihm	(him.ACC him.DAT)	0
ihn ihm selbst	(him. ACC him. DAT SELF)	0
ihm ihn		0
ihm ihn selbst		0
ihm sich	(him.DAT REFL.ACC)	0
ihm sich selbst		0
ihn sich		1
ihn sich selbst		14

The contrast of frequency and judgement patterns



Magnitude Estimation

- How to validate ME for linguistic stimuli?

Keller (2003):

- objective correlate of acceptability = number of (linguistic) constraint violations (n= 0.36)
- circularity?

Magnitude Estimation

- How to validate ME for linguistic stimuli?

Bard et al. (1996), etc.:

- comparison of ME results with other methods (e.g. 7-point scales, rank ordering, card sorting, cross-modal matching)

Cross-modal matching

- judging (non-) linguistic stimuli by assigning:
  a) line length
  b) numeric value

High correlation between ME data obtained by line lengths and numeric values

Bard et al. (1996)

Magnitude Estimation

Web-based version - DEMO:

http://www.cogsci.ed.ac.uk/~keller/web_exp/demo_magest.instr.html

ME: Experimental variation

- line lengths instead of numbers (increases face validity)
- fixing value of reference item
- leaving vs deleting reference item in view
- discouraging subjects from using 10-point scale
- allow time for calibration of the scale, i.e. discard first few responses

Magnitude Estimation: Summary

- Judgements are anchored by a reference item
- subsequent judgements are relative to previous ones
- judgements are proportional
- the scale has no minimum divisions and is open-ended

Magnitude Estimation: Summary

- Judgements are anchored by a reference item
- subsequent judgements are relative to previous ones
- judgements are proportional
- the scale has no minimum divisions and is open-ended

→ INTERVAL SCALE

An example

Scrambling in German

Scrambling in German

(1) … dass John gestern das Buch las.

that John yesterday the book read

Scrambling in German

(2) … dass John gestern das Buch las.

Scrambling in German

… dass John [das Buch] gestern t las.

Scrambling in German

(2) … dass John [das Buch] gestern t las.

(3) … dass [das Buch] John gestern t las.

Scrambling in context

(4) Wer kauft den/einen Wagen?
Who buys the/a car?

a. ?Maria denkt, dass der VATer den Wagen kauft.
Maria thinks that the father the car buys
b. Maria denkt, dass der VATer einen Wagen kauft.
c. Maria denkt, dass den Wagen der VATer kauft.
d. *Maria denkt, dass einen Wagen der VATer kauft.

Scrambling in context

(5) Was kauft der Vater?
What does the father buy?

a. Maria denkt, dass der Vater den WAgen kauft.
Maria thinks that the father the car buys
b. Maria denkt, dass der Vater einen WAgen kauft.
c. ?*Maria denkt, dass den WAgen der Vater kauft.
d. *Maria denkt, dass einen WAgen der Vater kauft.

Scrambling: A division of labour

Information Structure

(via FOCUS at PF)

Semantics

(Definiteness at/post LF)

SYNTAX

Pilot study: Stimuli

- contexts:

All FOCUS: What happened?
S-FOCUS: Who buys the car?
O-FOCUS: What does the father buy?

- stimuli:

a. Maria denkt, dass der Vater den Wagen kauft.
Maria thinks that the father the car buys
b. Maria denkt, dass den Wagen der Vater kauft.
Maria thinks that the car the father buys

Pilot study: Design

- factorial design (3x2x2)
    - *FOCUS*: All, S-Focus, O-Focus
    - *WORD ORDER*: base vs scrambled
    - *DEFiniteness of object*: +/-def

- additional conditions:
    - *Verb Second* x +/- scrambling
    - *Double nominative* x +/- scrambling

- fillers

Pilot study: Subjects and Method

- 47 native speakers of German

- Latin Square design - 8 lexicalizations
  → 192 test items + fillers

- 8 lists: 24 test items - 10 filler items

- Practice session 1: Judging line lengths (4)
- Practice session 2: Judging sentences (4)
- Experiment: Judging sentences (34 [24])

Results:

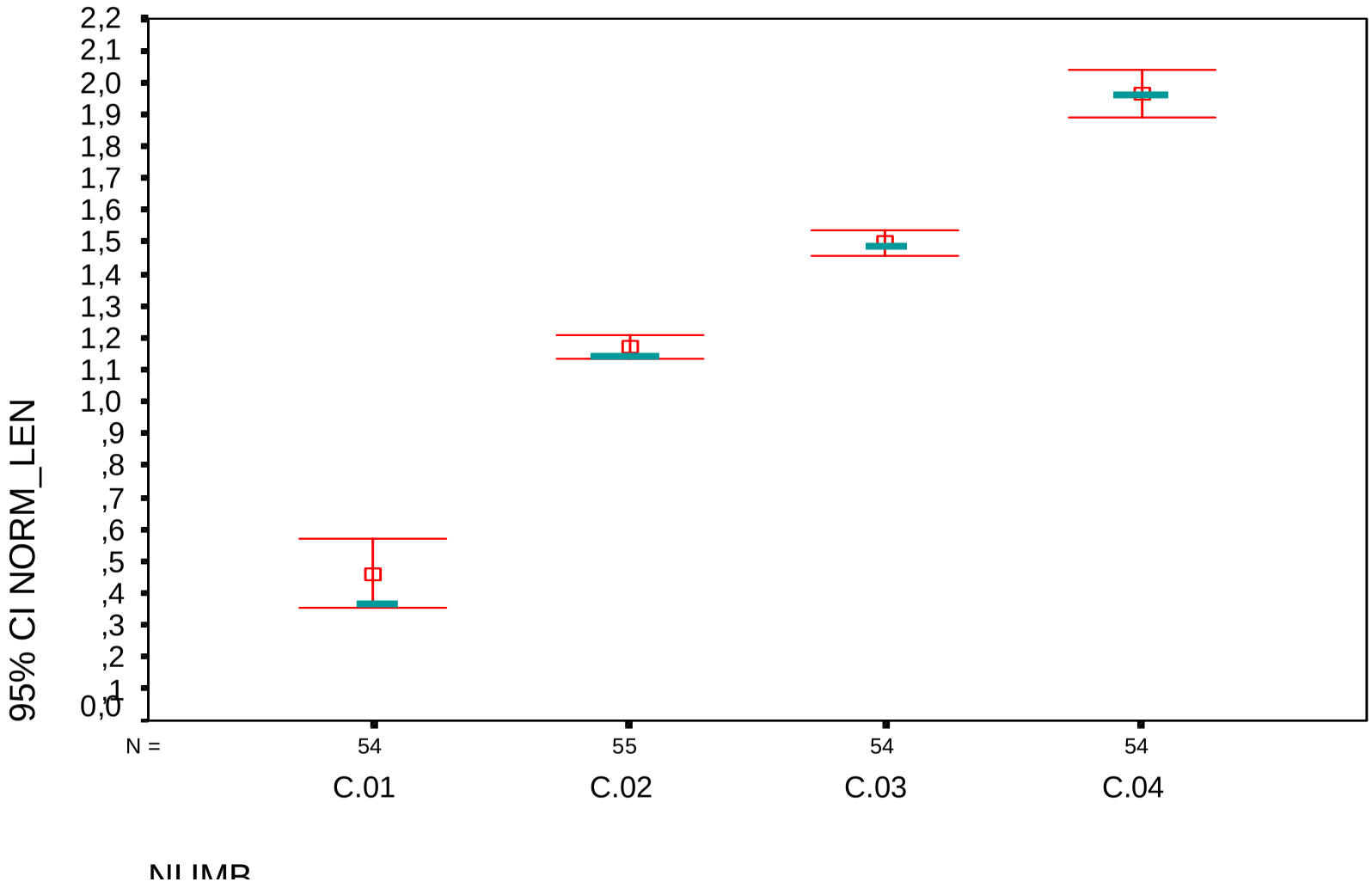
- Normalizing data:

$\psi_N = \psi_S / \psi_R$ (S - Stimulus; R - Reference)

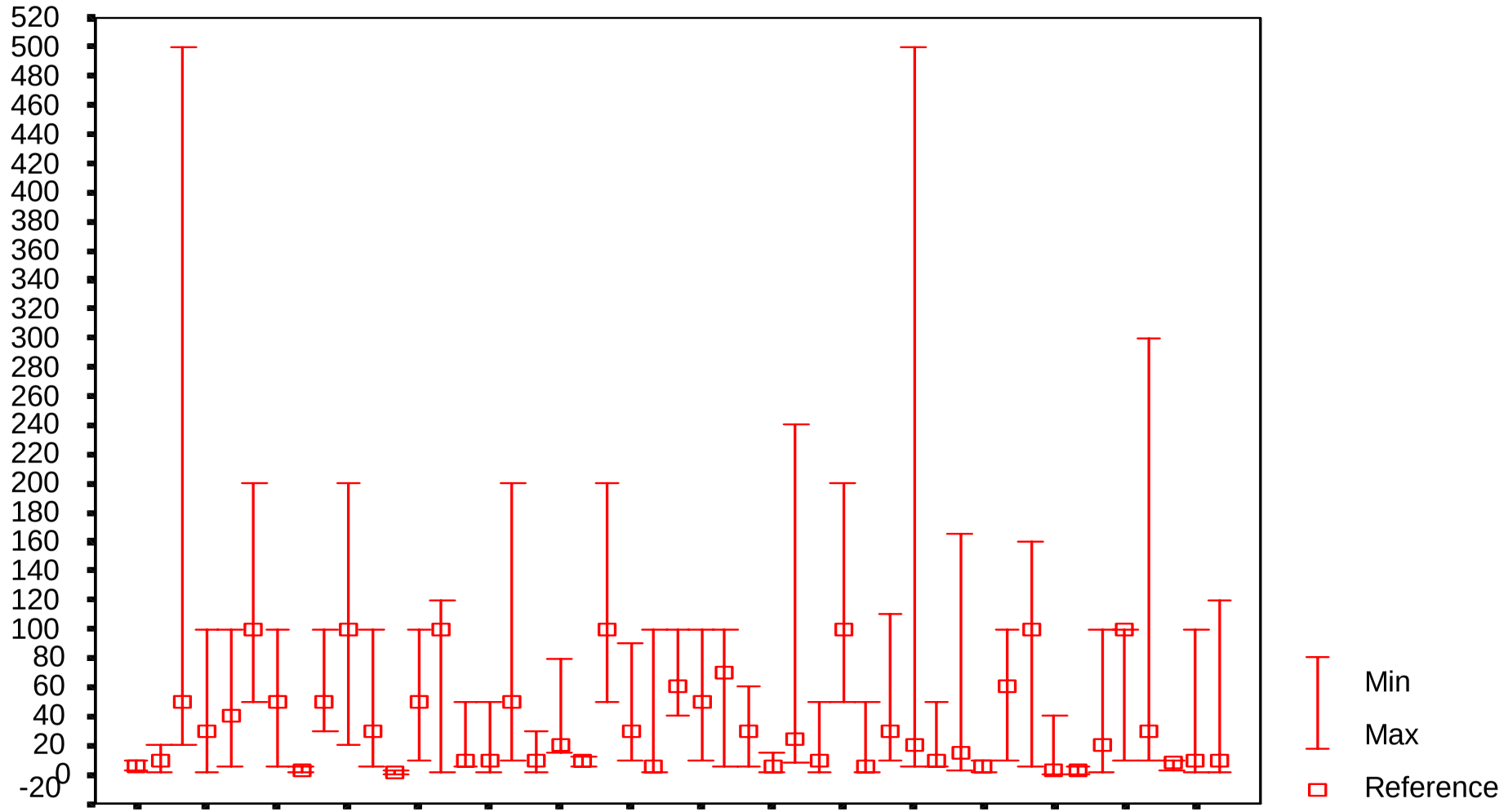
- Geometric means (i.e. log-transformation):

$\log \psi_N$

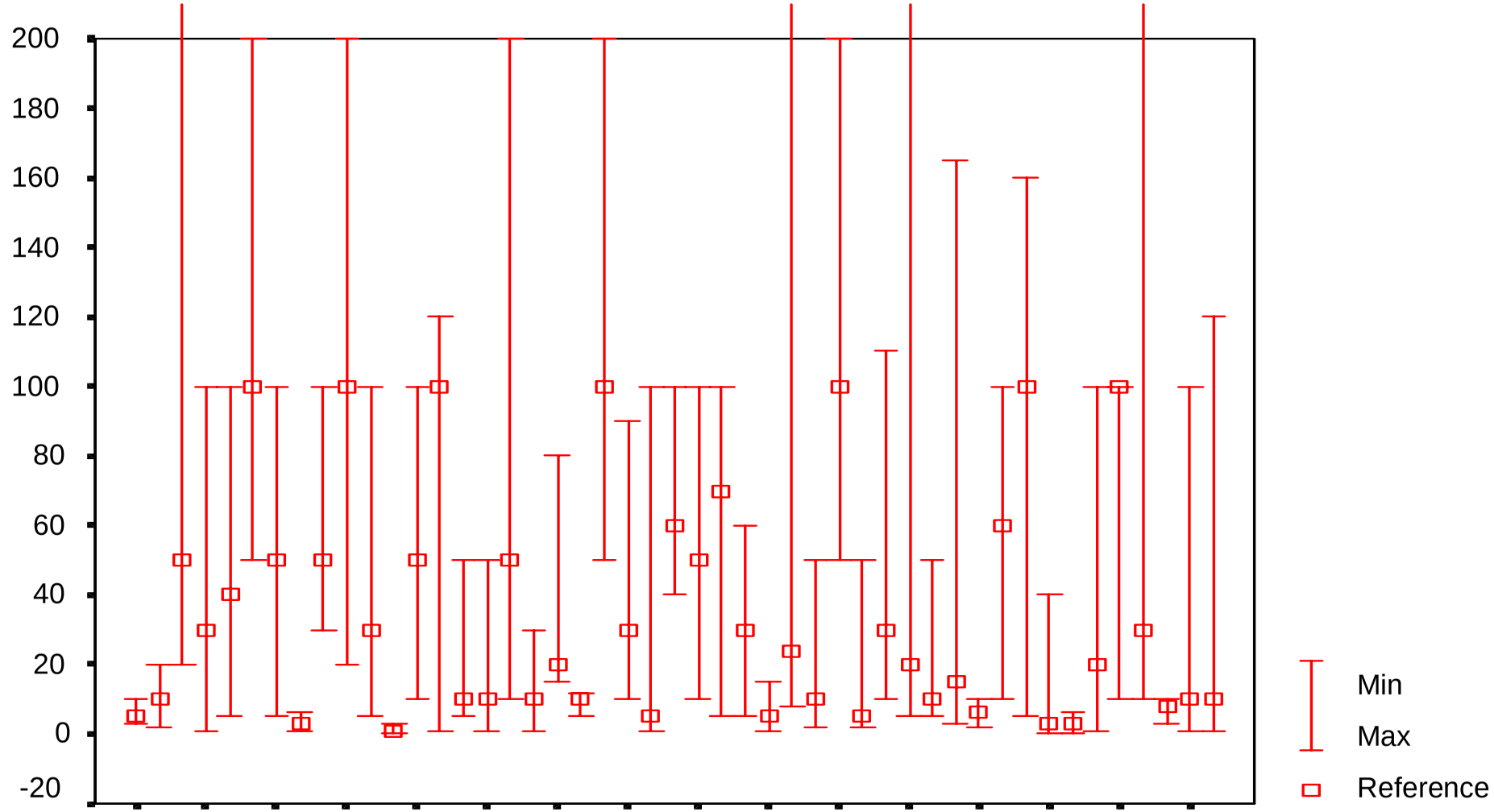
Results: Line judgements



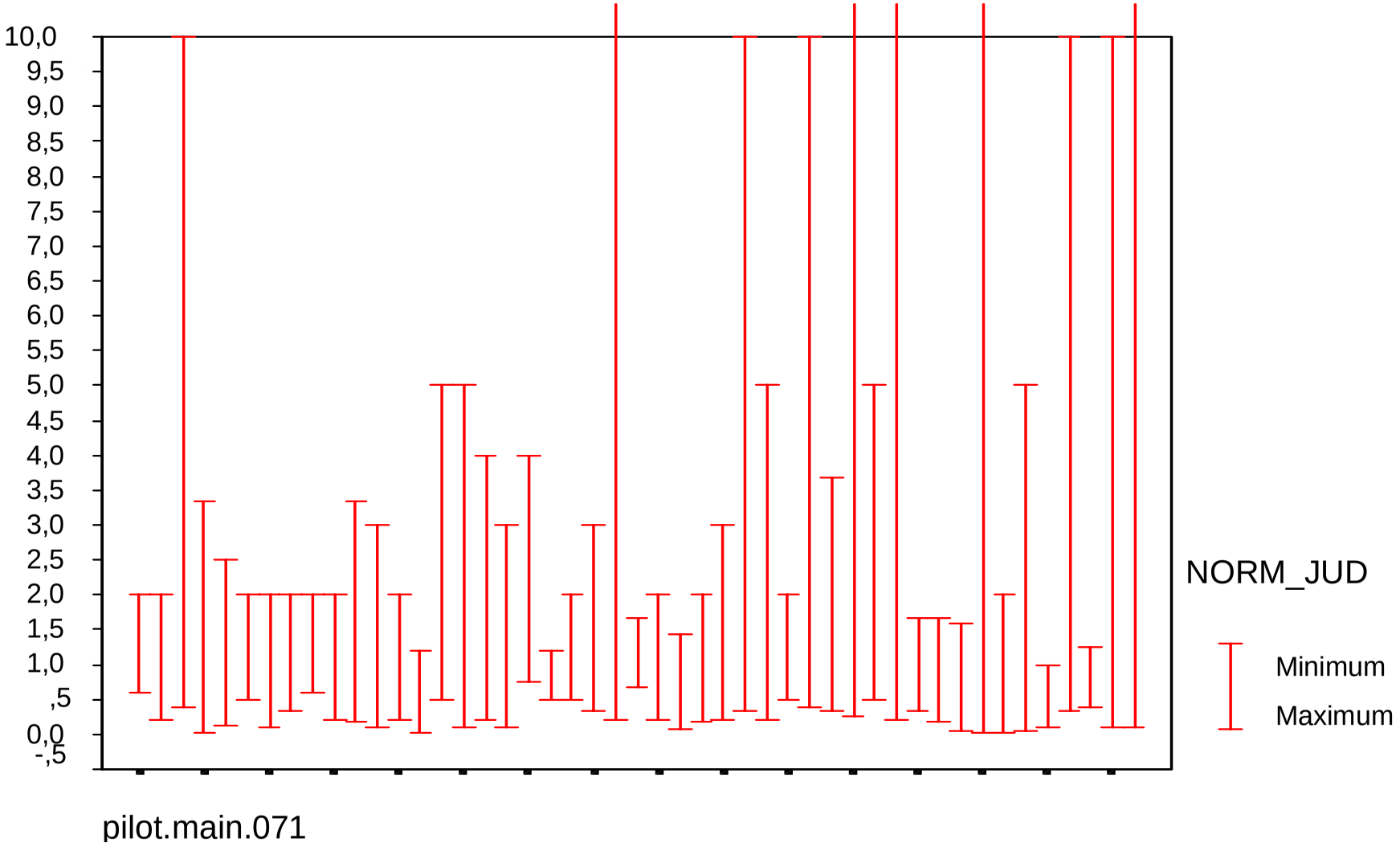
Results: Range of Individual Scales



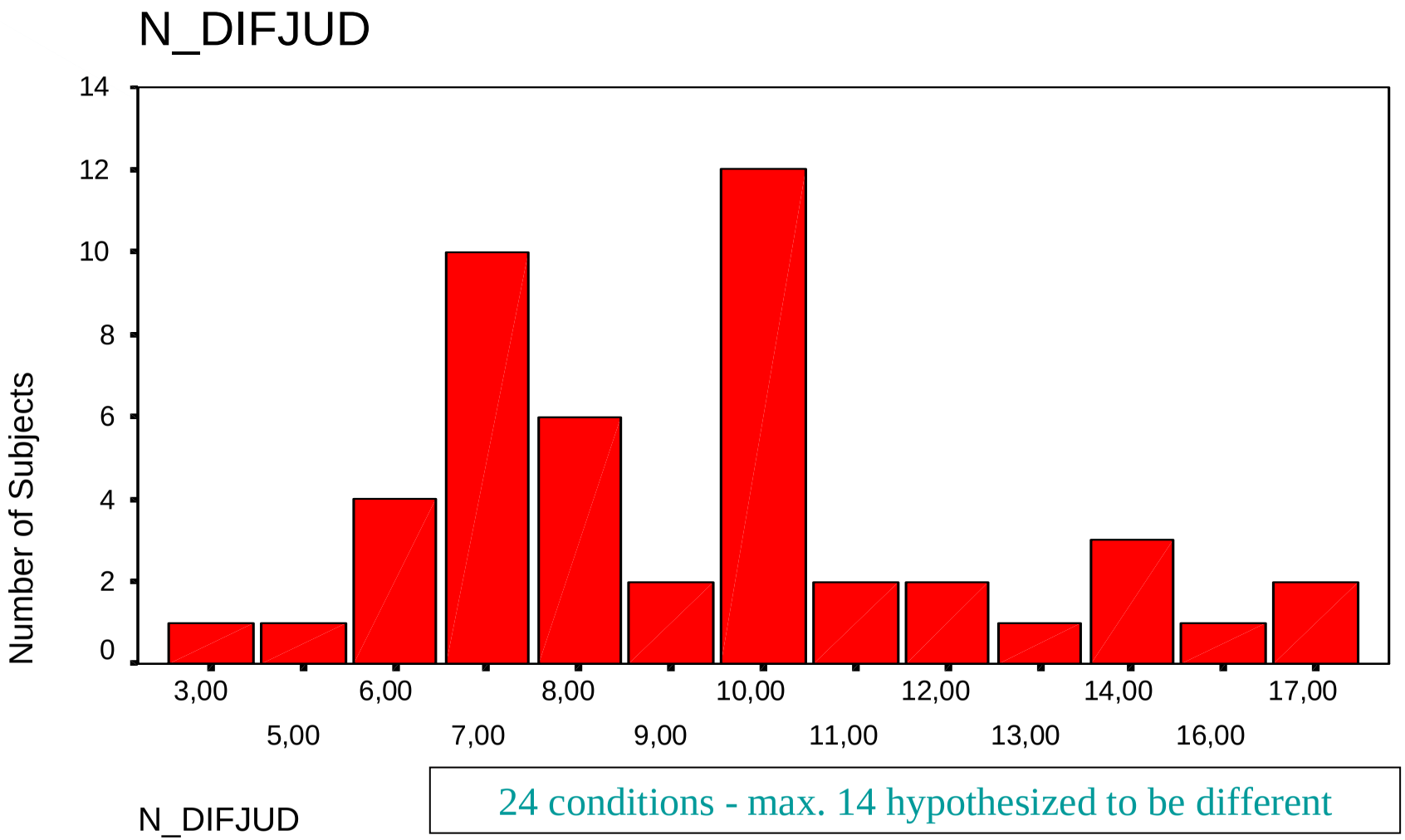
Results: Range of Individual Scales (rescaled)



Range of Individual Scales (Normalized)

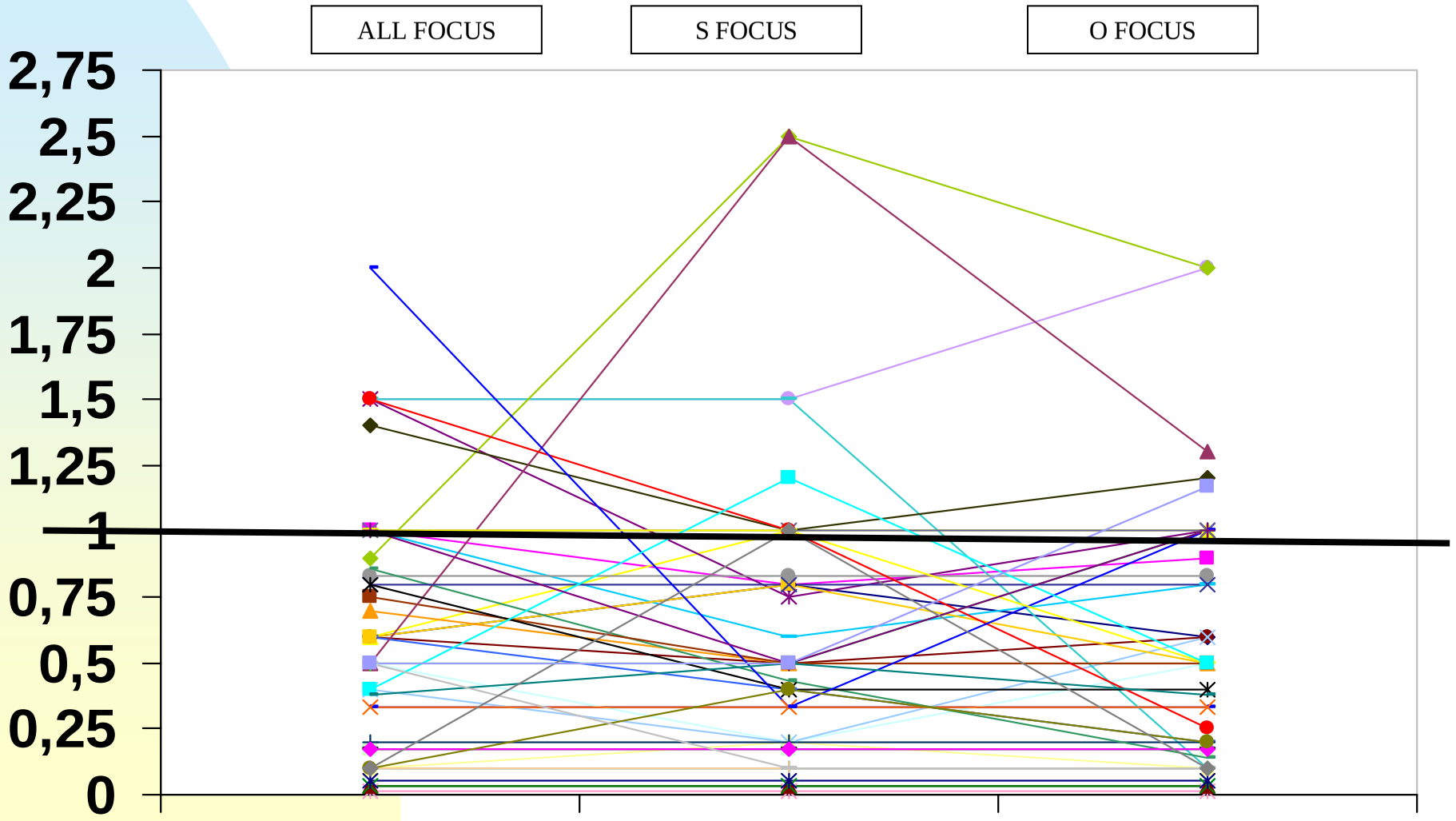


Number of Distinctions by Subjects

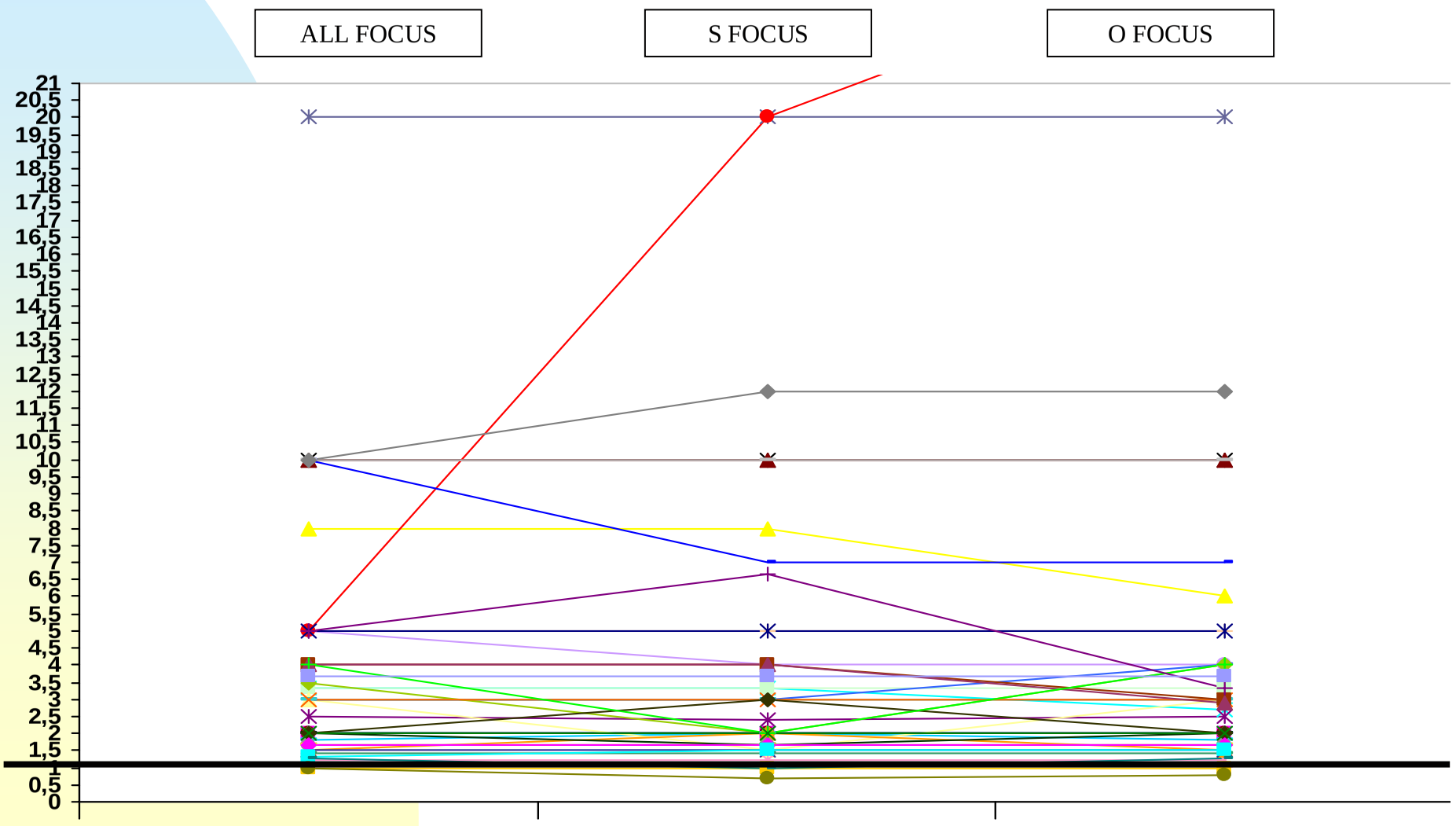


24 conditions - max. 14 hypothesized to be different

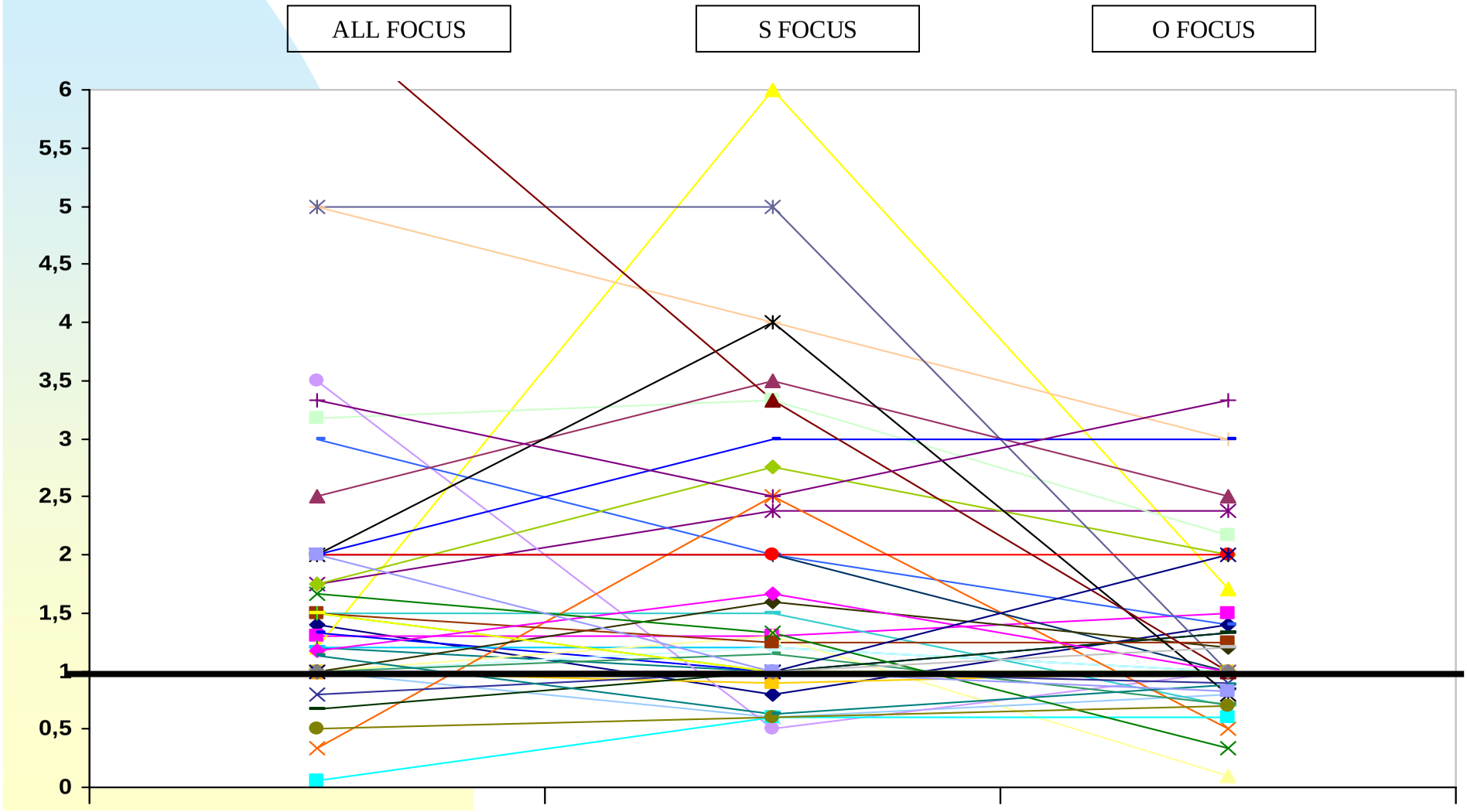
Subject Consistency - Ungrammatical Items



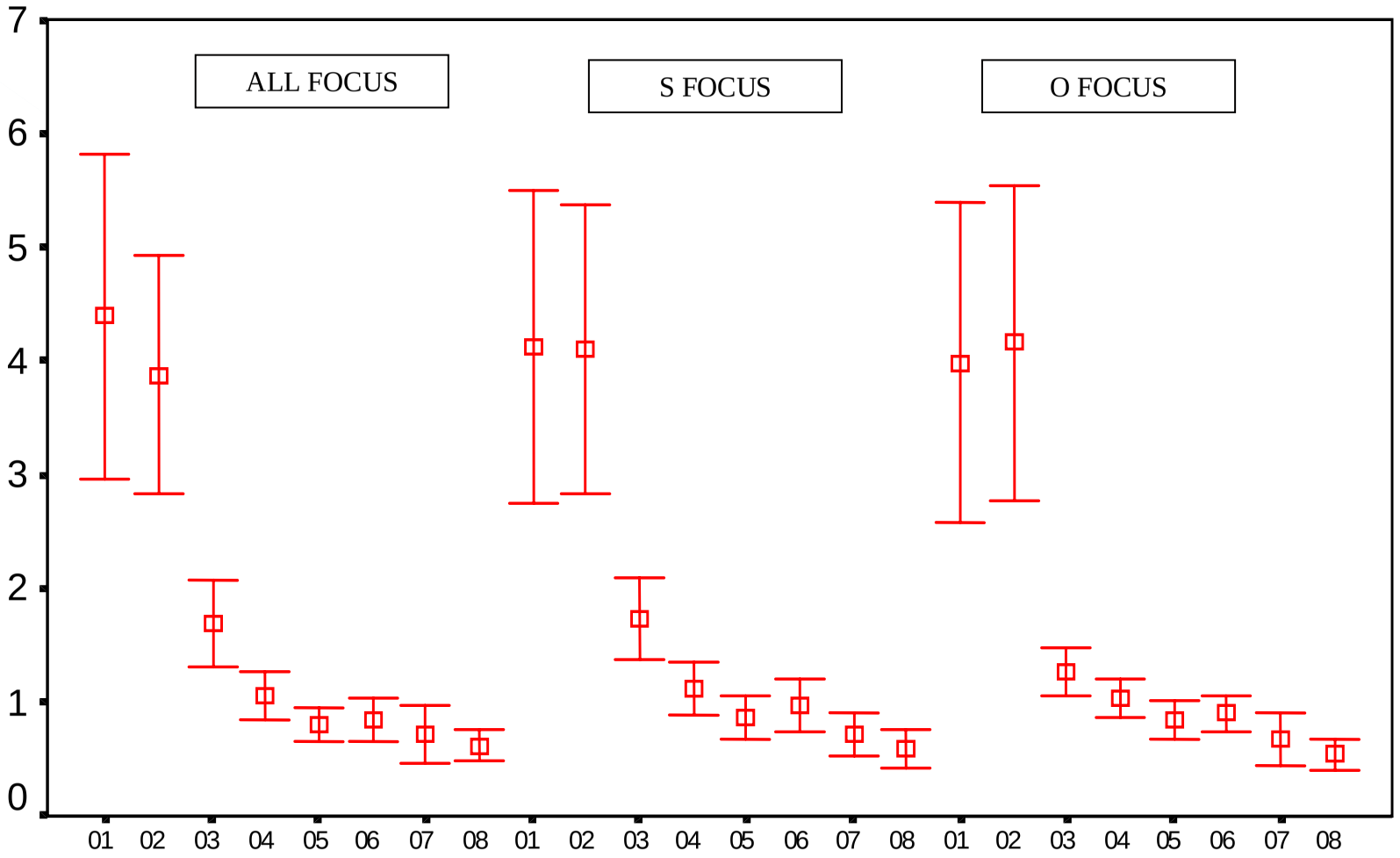
Subject Consistency - Non-scrambled order



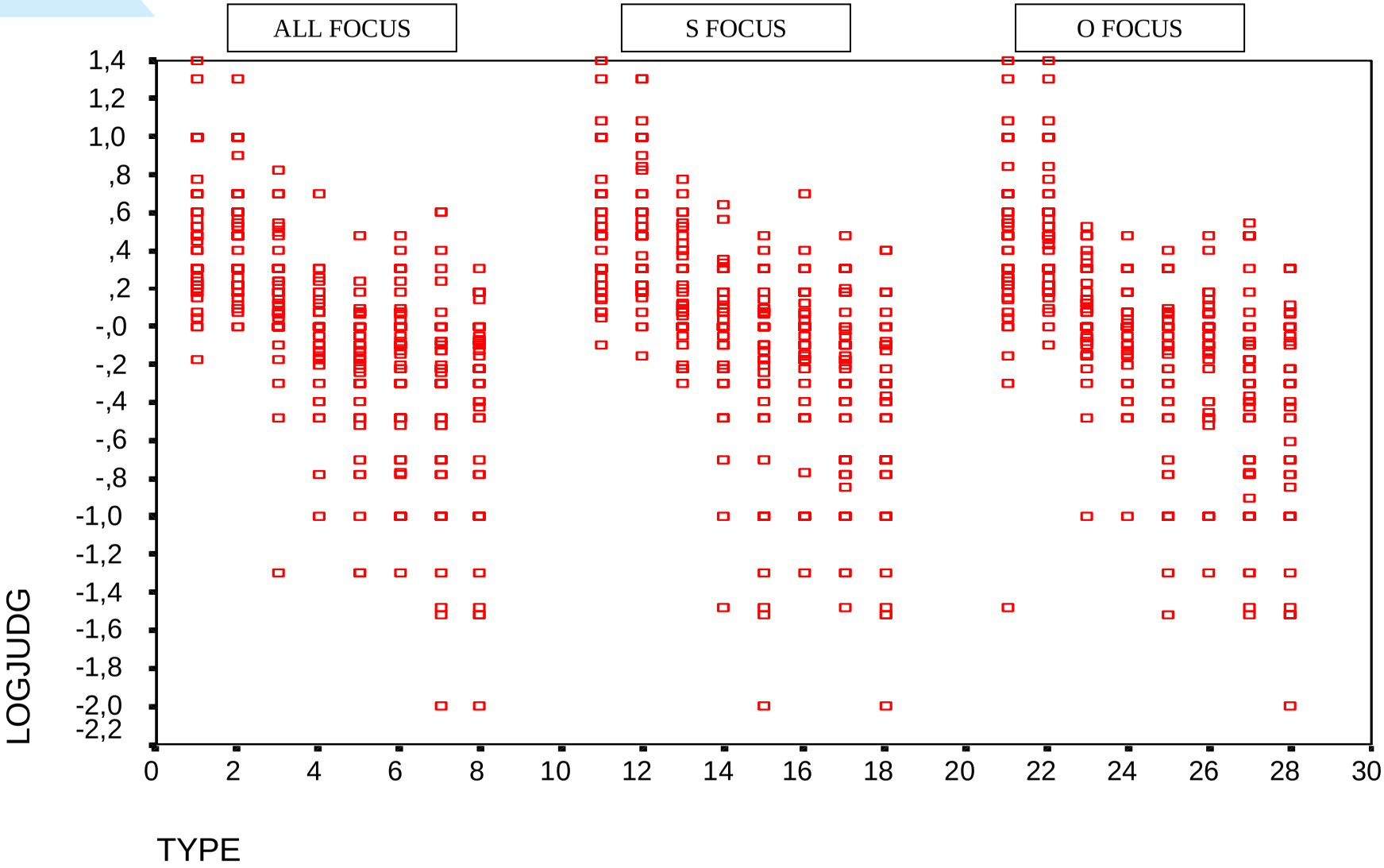
Subject Consistency - Scrambled order



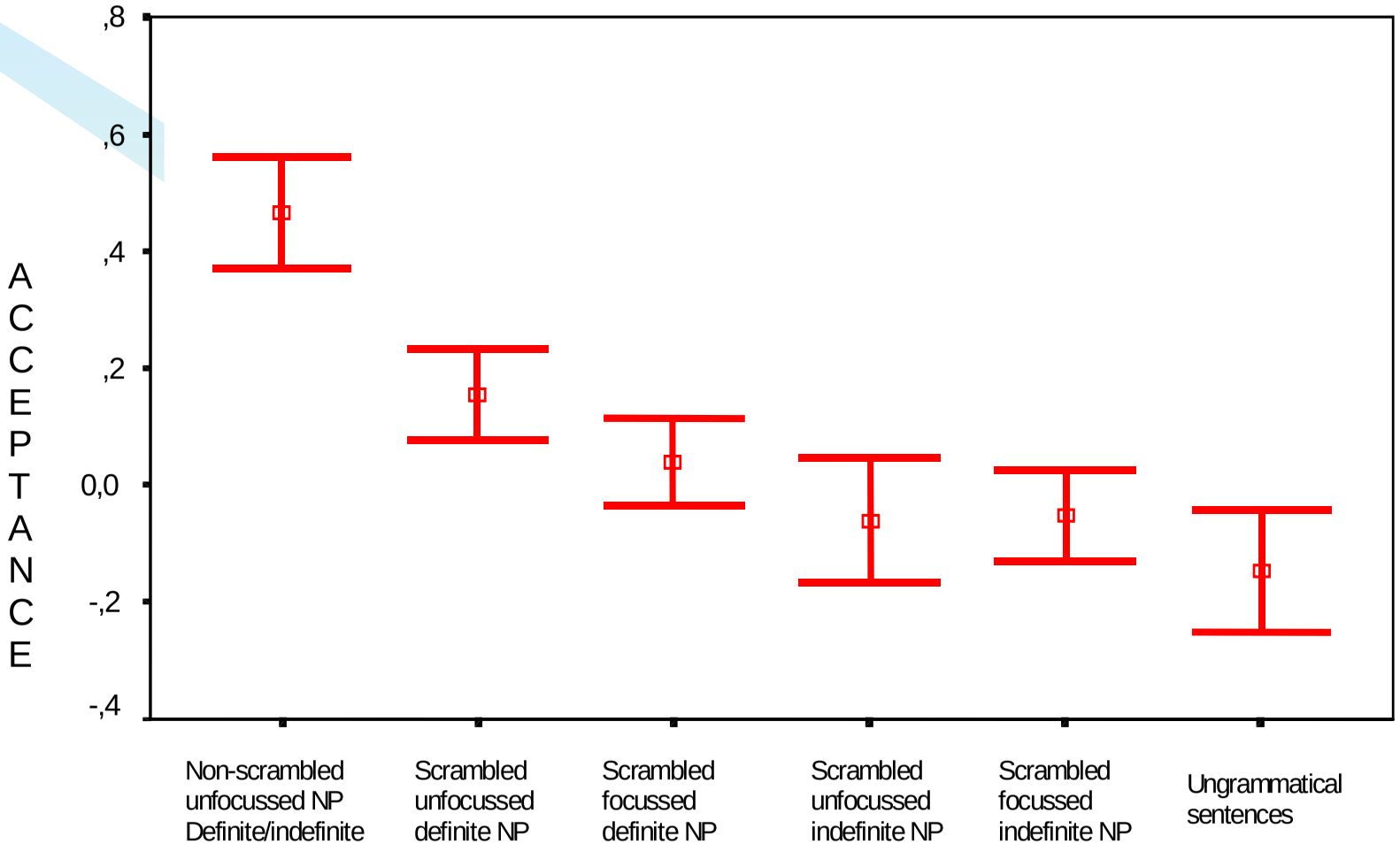
Results: Normalized means



Log-transformed judgements by item



Results: log-transformed normalized means



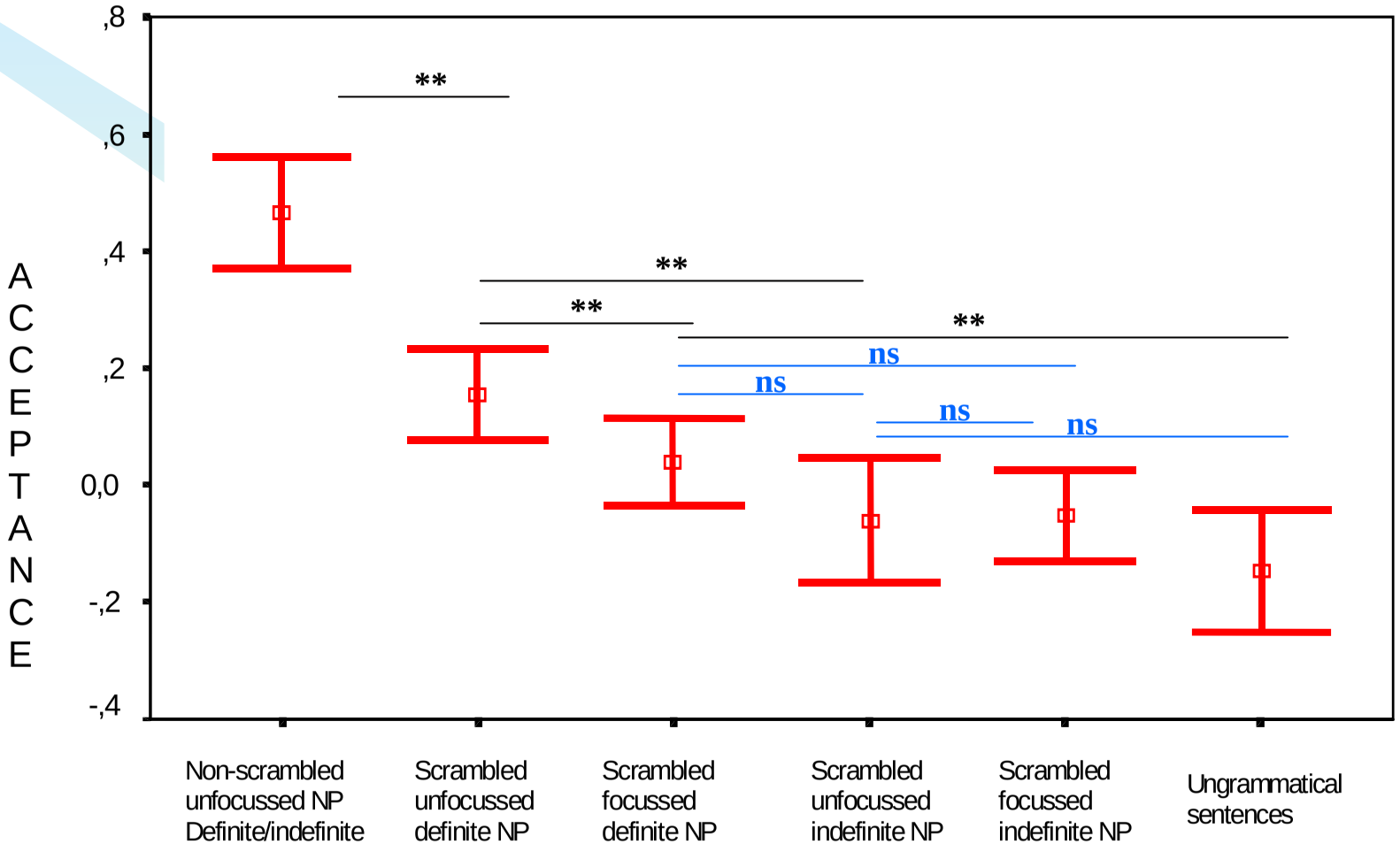
Pilot study

Statistical calculations:

- Repeated Measures ANOVA

- Post-hoc tests (e.g. Tukey)

Results: log-transformed normalized means



Pilot study: Summary

- constant acceptability of base order irrespective of FOCUS/DEF

- cumulativity of some constraints (e.g. WORD ORDER & FOCUS)

- non-cumulativity of other constraints (FOCUS & DEF)

- facilitating effects of FOCUS in appropriate contexts & adverse effect in inappropriate context

- DEF constrains scrambling categorically

→ NO optionality & NO obligatoriness of scrambling in any context

Pilot study: Summary

ranked weighting according the observed interactions between the factors investigated (35).

(35) *Case Violation = Verb Second = Definiteness > Word Order > FOCUS*

Magnitude Estimation: Perspectives

- Classifying constraint interactions:

Soft vs *Hard* Constraints (e.g. Keller 2000)

- Modelling constraint interactions in grammar:

Stochastic OT (e.g. Boersma, Bresnan, etc.)

Linear OT (e.g. Keller)

Weighted Rules (e.g. Uszkoreit)

Magnitude Estimation: Problems

-The robustness of judgements

Replication studies (Bard et al. 1996) and Fanselow et al. (2004) show high reliability of graded judgements

- The effects of the reference sentence

Schlesewsky (2004): Absolute and relative ratings of items change depending on reference sentence. Relative ranking remains constant.

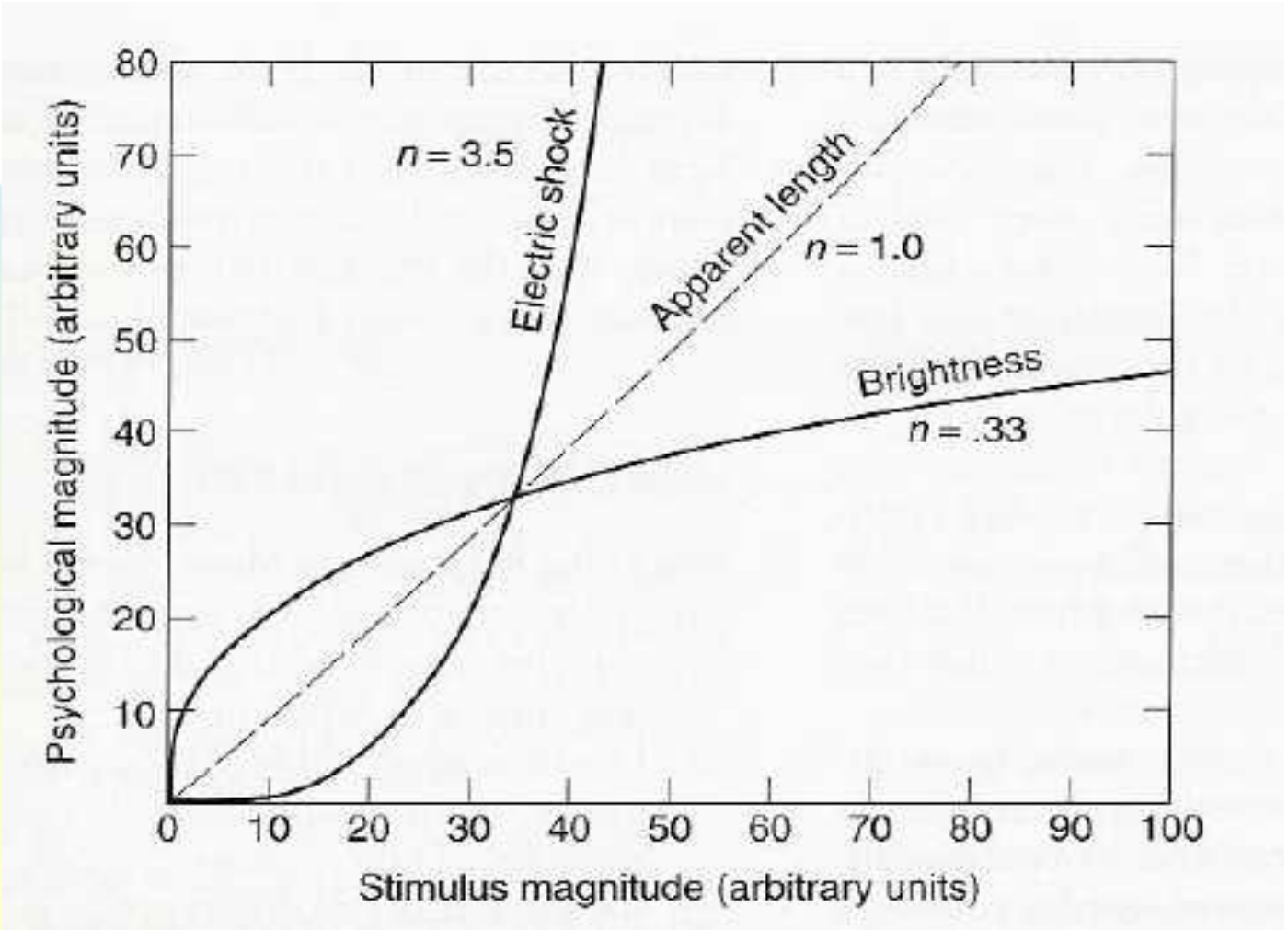
Magnitude Estimation: Problems

Proportionality

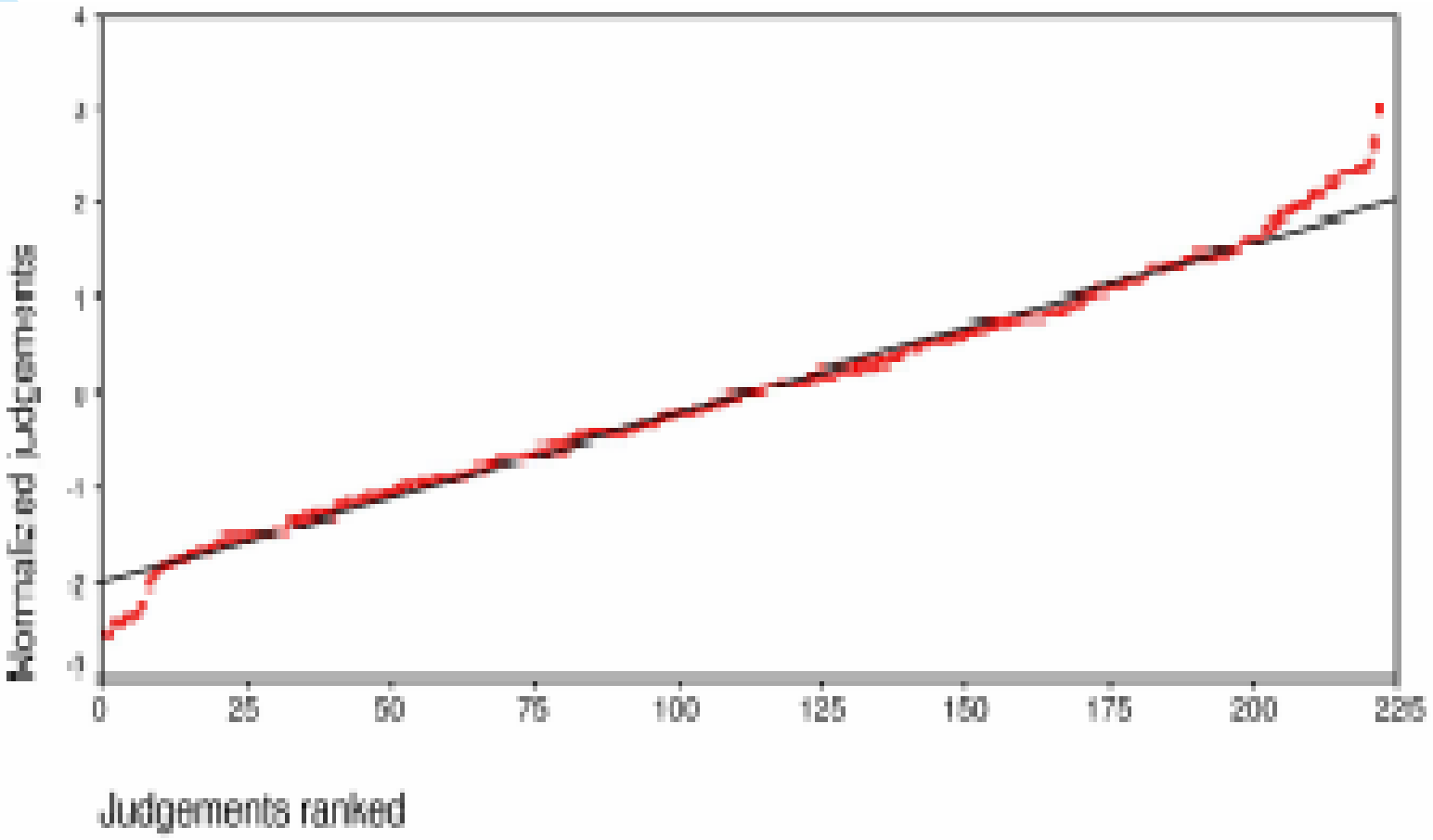
- What does it ‘mean’ for a sentence to be e.g. ‘twice or half as good/acceptable’ as the reference sentence?

- Can subjects judge proportionality, or do they just judge differences?

Psychophysical ME judgements



Linguistic ME judgements



Featherston (2004)

Consequences

?

- Just use normalized judgements, rather than log-transformed judgements
- Use ‘thermometer scale’, i.e. provide two anchoring points with fixed values as reference items (Featherston 2004)
- Treat ME data as (ultimately) reflecting ordinal scale