Understanding Typological Distributions

III

Balthasar Bickel

University of Leipzig

Wrapping up and going on

- Distributional skewings can be tested by GLMs in which the response is the odds for families to show specific skewings, i.e.

$$log(\frac{\pi(\text{skewed towards q})}{\pi(\text{skewed towards } \neg \text{q} \mid \text{diverse})}) = \alpha + \beta_i p_i ... + \beta_k p_k$$

(Abbreviate this as $q \sim p_i...p_k$.)

- Statistical models ≠ causal models!
- Instead, statistical models need theoretical interpretation and motivation, i.e. typological theories
- A typological theory is testable iff we can derive from it a set of statistical models, with well-defined variables *p* and *q*.
- Testable typological theories explain what's where why.

Two main classes of typological theories (as I see it)

1. ‘Match’ (naturalist, functionalist) theories: some distributions are more likely than others because they are more ‘natural’, i.e. better tuned to the way our brain processes language and to our cognitive abilities
   - typical variables in naturalistic models: structural and discourse properties, perhaps also types of social structures (or cognitive models of them)
2. ‘Spread’ (replicationalist) theories: distributions reflect the spread of structures in time and space, i.e. descent and language contact
   - typical variables in resulting models: structural and discourse properties, socio-geographic areas; family skewing independent of predictors

- The most interesting research designs combine variables from both theory classes in one model!

Match theories

- Key idea: some distributions match better ...
  - the nature of the brain
  - the nature of communication
  - the nature of society
- A classical example: according to Hawkins, the nature of incremental processing is better matched by OV structures with A≠O coding than by OV structures with A=P in morphology. We can *derive from this* the testable model:

$$log(\frac{\pi(\text{nonneutral})}{\pi(\text{neutral} \mid \text{diverse})}) = \alpha + \beta_1 \text{FINAL} + \beta p_j ... + \beta_k p_k$$

- for which we found robust statistical support.

Match theories and relativism

- *A common misunderstanding:* Match theories must posit universal ‘natures’, i.e. ‘communication’ and ‘society’ must be in the singular.

- *But:* a distribution can universally match universal cognition just as well as it can variably match variable cognition. In either case, what is truly universal is the ‘match’ between a given type of cognition and a given linguistic distribution.

- In fact, some of the best universal ‘matches’ come from research designs in which both cognition types and linguistic structures vary.

- The key point is that they co-vary: $q \sim p$!

A (by now) classical example: spatial relativity

Table 1. Verbal and non-verbal codification of spatial configurations

language	preferred system(s) of verbal codification	system(s) of non-verbal codification	
		predicted	found
Arrernte	A	A	most A
Hai//om	A, (I)	A	most A
Tzeltal	A	A	A
Longgu	A	A	A, also R
Dutch	R	R	R
Japanese	R	R	R, also A
Kilivila	I, R	R	A, also R
Belhara	A	A	most A
Tamil (rural)	A	A	A
Tamil (city)	R	R	R
Kgalagadi	R (A, I)	R	R, also A
Mopan	I	ad hoc	R, A
Totonac	I	ad hoc	R, A

A = absolute reference system
R = relative reference system
I = intrinsic reference system

Other examples from recent relativity research

- Motion (Slobin 1996, 2000, Gennari et al. 2002)
- Objects (Lucy 1992, Imai 2000)
- Color (Roberson et al. 2000 etc.)
- Referential Density (Bickel 2003, Stoll & Bickel 2006)

Referential Density

Belhare (Kiranti, Sino-Tibetan)
pʌila . . . aː . . . ambibu phig-he kinahuŋgo
first PTCL mango [ABS] [3s.A-]pick.from.above-PT[3O] SEQ
otutuiʔ = na jhola-e ukt-he
quite.big = ART[S] bag-LOC [3s.A-]take.down-PT[3O]
inetnahuŋgo dhaki-e leŋs-e
then closely.weaved.basket-LOC [3s.A-]put-PT[3O]
il-lam il-lam sas-sa-ba leŋs-e ʌni . . .
DIST:DEM-MED DIST:DEM-MED pull-CONV-LOC [3s.A-]put-PT[3O] and.then
riksa, eː saikil-lamma, saikil-lamma ta-he
rikshaw PTCL bicycle-MED bicycle-MED [3s.S-]come-PT
kinahuŋgo . . . ⟨B99.4.1–5⟩
SEQ

‘First, . . . uh . . . [someone] picked mangos and took [them] down in a big bag. Then [s/he] put [them] into a basket. [Someone] moved over [an animal] by pulling from over there, and then [someone] came on a rikshaw, uh . . . on a bike, on a bike and then . . .’

Maithili (Indo-Aryan, Indo-European; Nepal)
ek-ṭā ām-ke gāch rah-ai. ā . . . a . . . a . . .
one-CL mango-GEN tree[NOM] be-3NH.NOM[PR] PTCL
ām me ek e-goṭā chaurā ām tor-ait
mango in one one-CL boy[NOM] mango[NOM] pluck-IP
rah-ai
AUX-3NH.NOM[-3NH.NONNOM.PR]
ā . . . u ām toir-ke ṭokari me rakh-ne
PTCL 3NH.NOM mango[NOM] pluck-CONV basket in keep-INF
jāi che-l-ai. omaharse e-goṭā chaurā
AUX AUX-PT-3NH.NOM[-3NH.NONNOM] and.then one-CL boy[NOM]
e-l-ai,
come-PT-3NH.NOM
laḍkā sāikal par caḍh-ne, ā . . . u ek-ṭā am-ke
boy.H[NOM] bike on ride-INF PTCL 3NH.NOM one-CL mango-GEN
ṭokari corā-ke cail ge-l-ai . . . ⟨M3.6.1–6⟩
basket[NOM] steal-CONV move.IP AUX-PT-3NH.NOM

‘There is a mango tree and . . . uh . . . uh . . . in the mangos, one, a boy is picking mangos. And when picking mangos, he put them into a basket. Then a boy came, a young man riding on a bike, and he stole one basket of mangos, and took off . . .’

$$\text{RD} = \frac{N \text{ (overt argument NPs)}}{N \text{ (available argument positions)}}$$

Referential Density

- Experiment with 10 speakers from (sofar) 7 languages
- Plausible predictors:
  - Text length
  - Sociology of communication: close-knit vs. loose
  - Some structural property of grammar: case-based agreement requires NP information, and this primes activation of NP structures in production

Case-based agreement in Maithili (IE)

a. (tũ) bimār ch-æ?
   2nhNOM sick be-2nhNOM
   'Are you sick?'

b. (torā) khuśi ch-au?
   2nhDAT happy be-2nhNONNOM
   'Are you happy?'

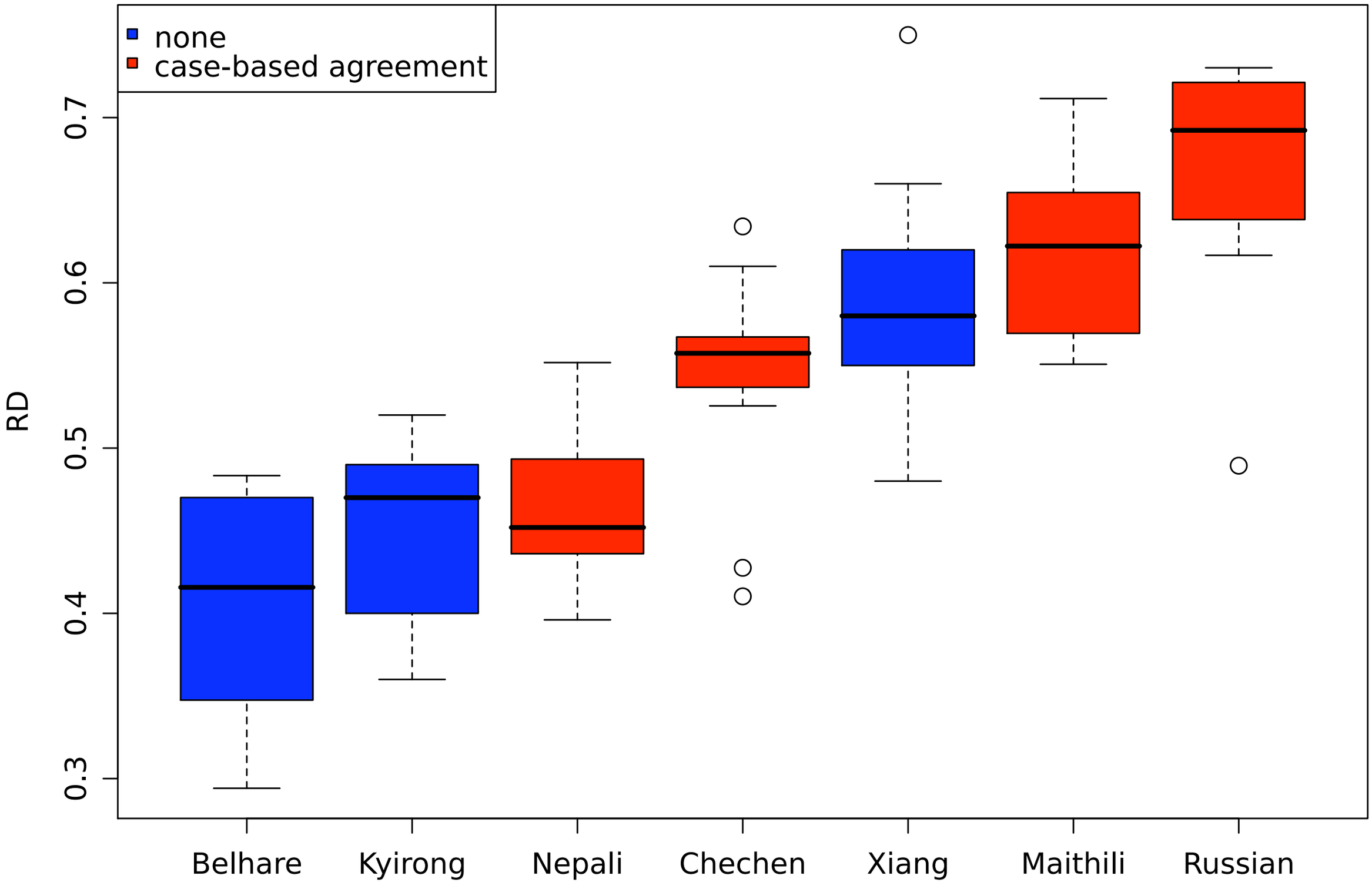
Non-case-based agreement in Belhare (ST)

a. han khar-e-ga i?
   2s[-NOM] go-PST-2s Q
   'Did you go?'

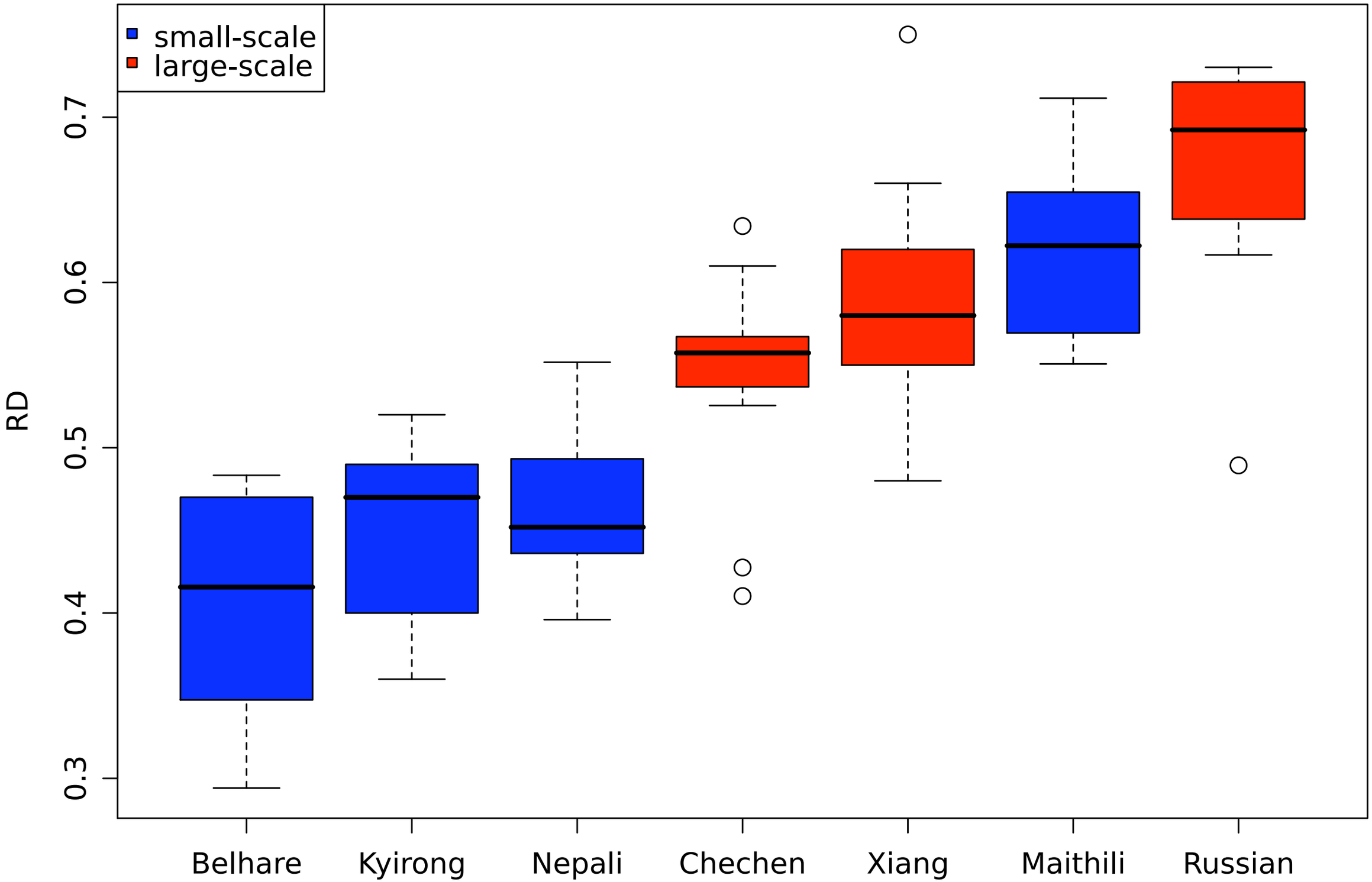
b. han-na un lur-he-ga i?
   2s-ERG 3s[-NOM] [3sA-]tell-PST-2s Q
   'Did you tell him/her?'

c. ciya (hanna-ha) n-niũa tis-e-ga i?
   tee[-NOM] 2s-GEN 2sPOSS-mind easy-PST-2s Q
   'Did you like the tea?'

Referential Density



Referential Density



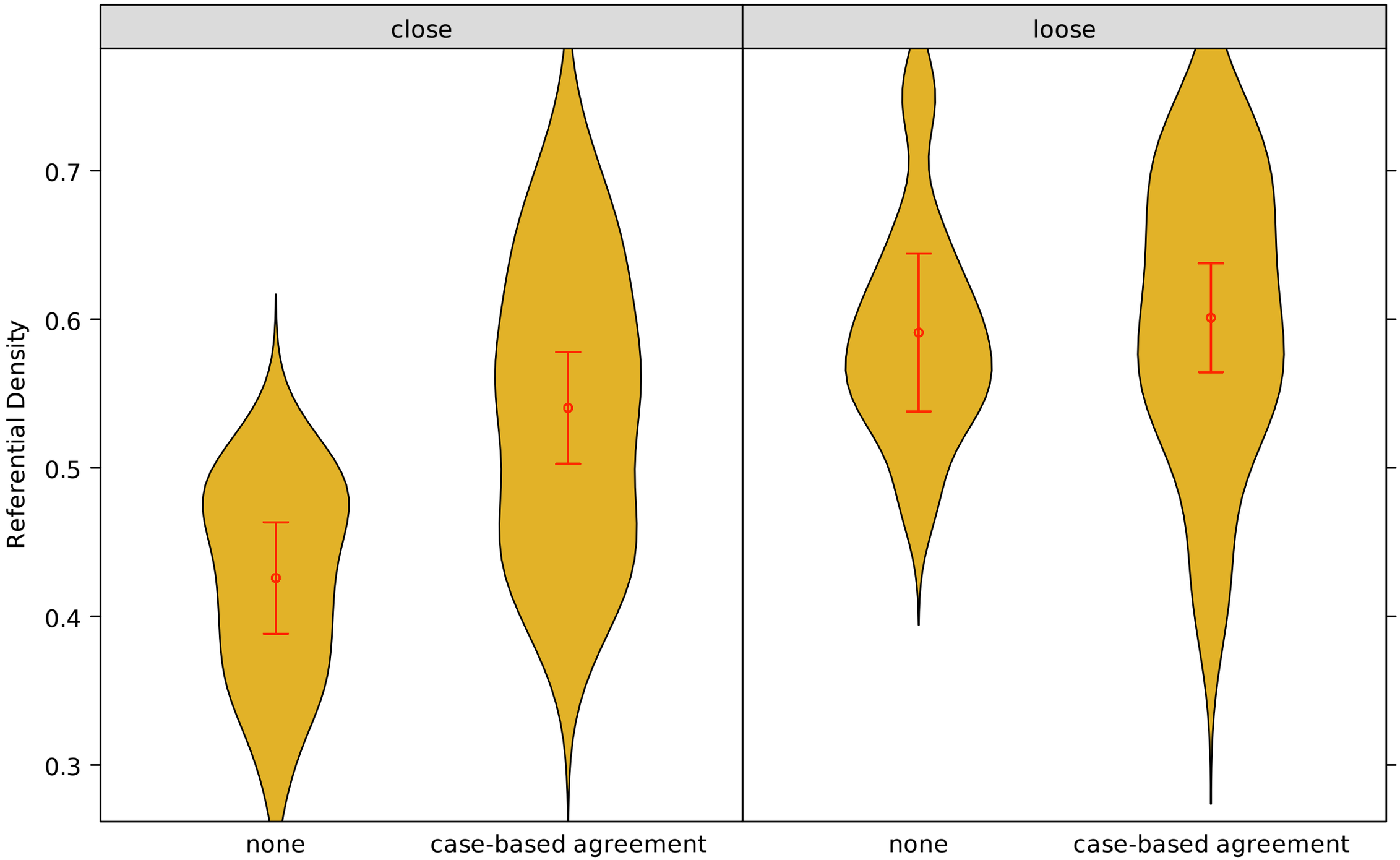
Referential Density: Results

- Model: $\mu(RD)=\alpha+\beta_1 SOC+\beta_2 SYN+\beta_3 LENGTH+\beta_3 SOC*SYN+\beta_4 SOC*LENGTH+\beta_5 SYN*LENGTH+\beta_6 SOC*SYN*LENGTH$
- β4...6 are not significantly different from zero;
- But β_3 is significant: $F(1)=14.45$, $p<.001$
- Better model:

$\mu(RD)=\alpha+\beta_1 SOC+\beta_2 SYN+\beta_3 LENGTH+\beta_3 SOC*SYN$

Referential Density: Results

$\hat{\mu}$(RD) = .43 + 1.6SOC + .12SYN -.10SOC*SYN



Referential Density: Factorial analysis

- μ(RD)=$\alpha+\beta_1$SYN+β_2LENGTH, SOC=close
  - β_2 ns.; β_1: $F(1)=19.5$, $p<.001$

- μ(RD)=$\alpha+\beta_1$SYN+β_2LENGTH, SOC=loose
  - $\beta_{1..2}$ ns., but also no reversal of the SYN effect!

Referential Density: Conclusions

1. Syntactic type has an effect, but it is blurred (not contradicted!) by the social type effect under loose/large-society condition.

2. Social structure also has a — presumably universal — effect!

- This suggests that aspect of social structure and cultural types may be competing predictors of typological distributions.

- Possible other cases:
  - kinship and marriage models and kinship terminology
  - generational groupings and Lardil-style ‘kintax’ (see Evans 2003 in *Ann. Rev. Anth.* for recent discussion)
  - social stratification and honorific grammar

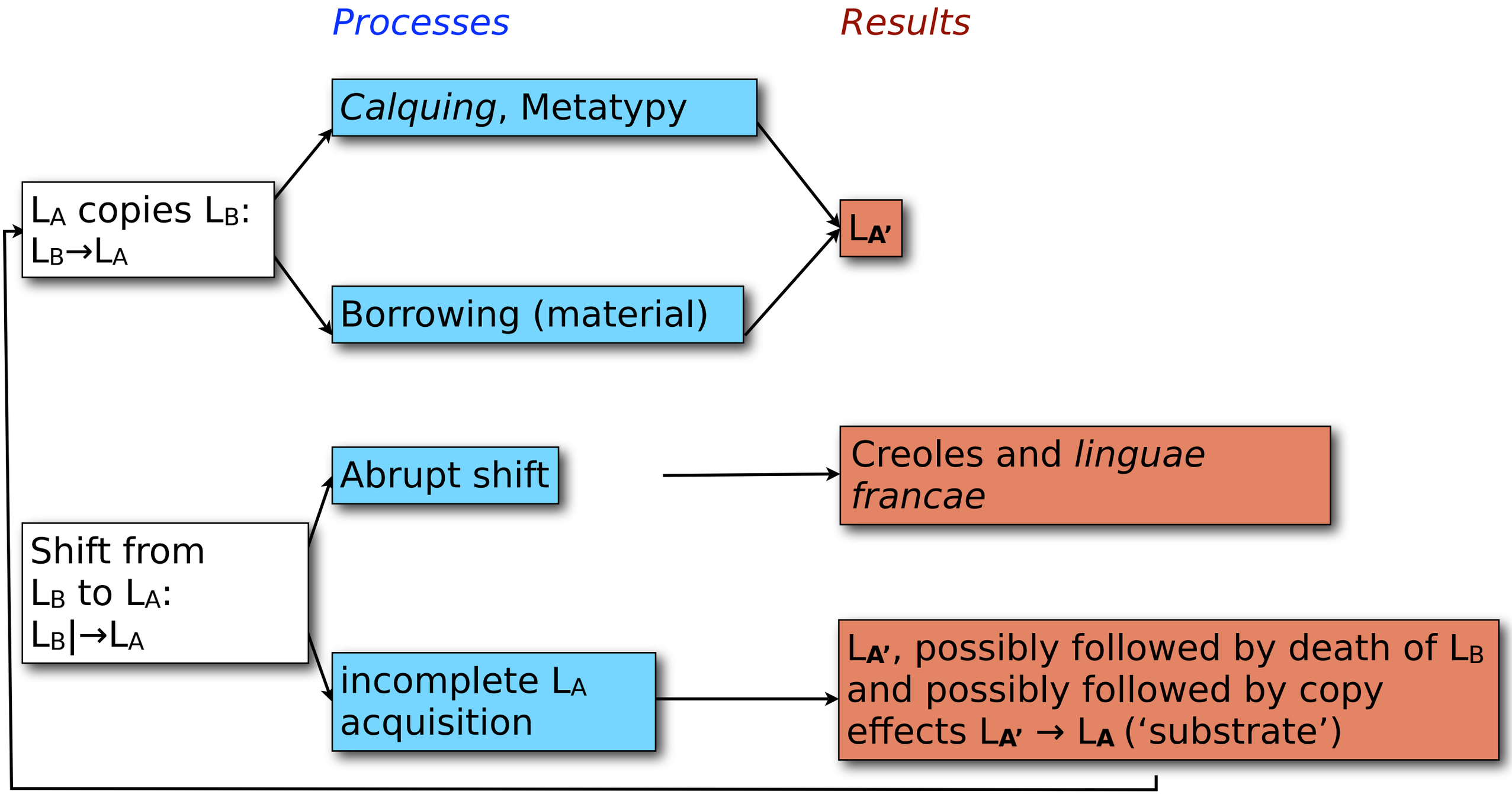
Match theories and relativism: what's behind it?

- *Assumption*: conceptual/pragmatic structure ('thinking') ≠ semantic structure ('speaking') (e.g. Levinson 1996)
- *Observation*: processing and acquisition is easier if the semantic structure matches the conceptual structure: "Thinking for Speaking" (Slobin 1996)
- *Hypothesis*: given certain types of frequently used structures, expect certain conceptual structures to dominate: q ~ p!

A note on sampling

- Small can be better than big!

Spread theories: processes underlying diffusion



Inspired by the debates between Thomason 2001 and Ross 2007

Spread theories

- From all we know, contact is frequent and universal, and it is likely that it has been frequent and universal throughout the history of our species.
- Problem: contrary to a widespread belief, we cannot infer past contact from linguistic distributions!

Problem 1: circularity

- Suppose we find a set of frequency peaks in a set of variables in some geographical region, an ‘isopleth’ (van der Auwera 1998, Haspelmath 2001, on Europe)
- A competing explanation is that the variables are universally correlated.
- When testing universal correlations, we must control for areas....

Problem 2: What counts as evidence?

- Suppose we find evidence in 5 variables (as is the case in the Balkan area)
- In a survey of 200+ variables (the size of modern databases), we can expect to make 5 hits by chance!

Problem 3: testability

- If an area is defined by linguistic data, how can it be tested?
- We can always discern clusters on a map...!

Response: Predictive Areality Theory (Bickel & Nichols 2006)

- Each variable has its own history of and potential for change and spread.
- Each distribution requires its own explanation.
- Variables may or may not have overlapping distributions.
- If variables have an overlapping distribution, shared history is one plausible explanation = a Predictive Areality Theory.
- Any Predictive Areality Theory must be grounded in what we know from population history through archaeology, genetics, ecology, geography, economics, demography, etc.

Predictive Areality Theory (Bickel & Nichols 2006)

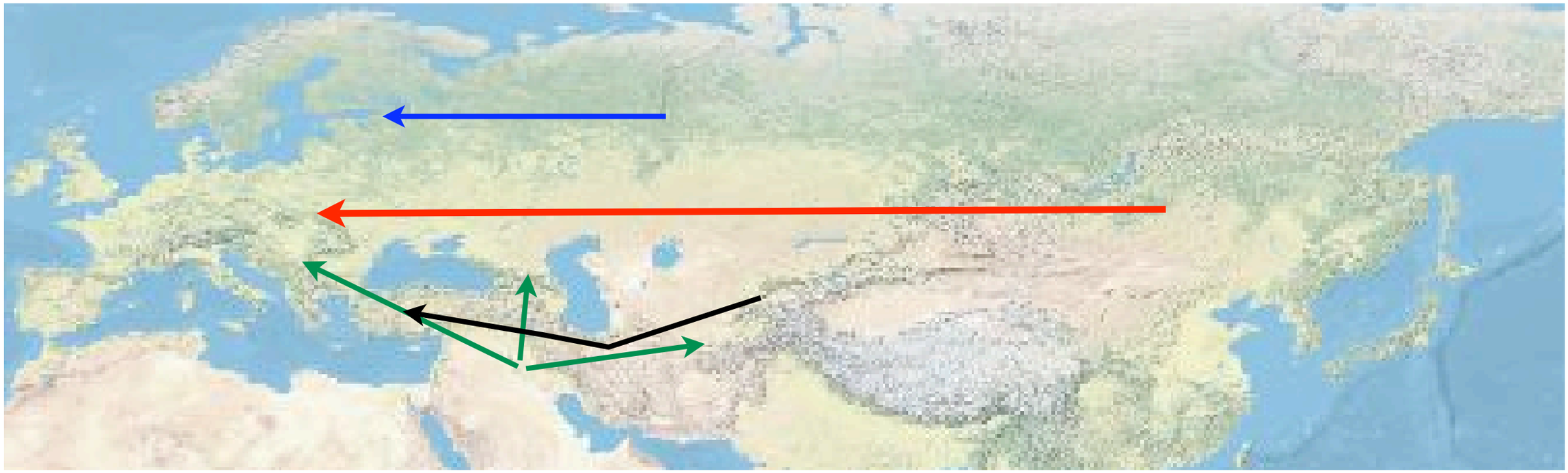
- Under this approach, areality is not a typological fact, but a predictor variable predicting typological distributions. (This is the exact opposite of what areality was under classical approaches!)
- $g(E(q)) = \alpha + \beta_1 AREA + \beta_j p_j ... + \beta_k p_k$, where
  - $p_j ... p_k$ control for structural pressure and where
    - inheritance is controlled for either by the Skewed Family Method or the G-Sampling Method
- The such models (many q, high β_1's) are statistically supported, the more robust is the theory.

A case study: Eurasia

- Extralinguistic bases for the theory:
  - historical records, archeology and known effects of cultural spreads
  - human populations genetics

Eurasia: known and reconstructed history

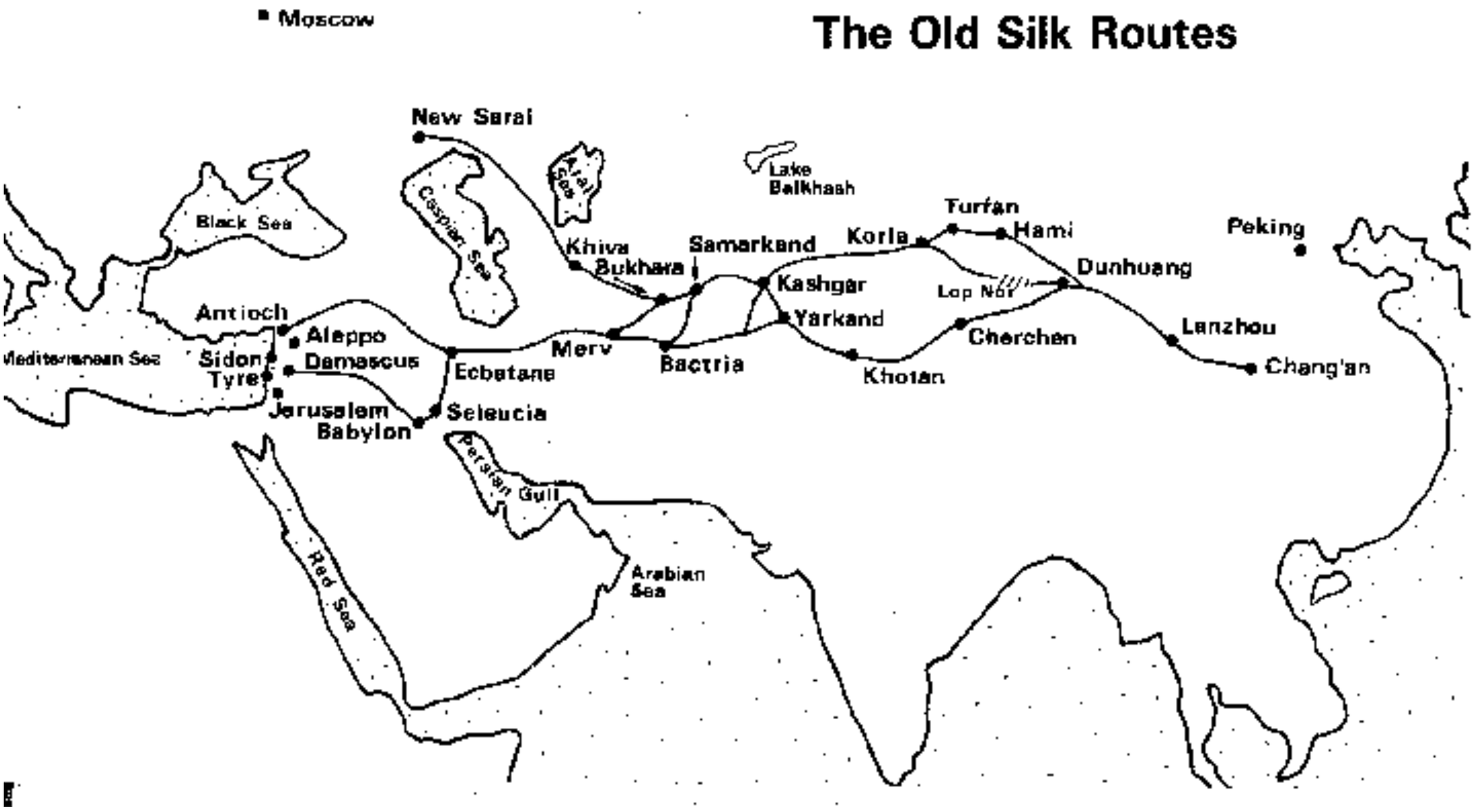
- Repeated conquests and migrations, mostly male-dominated



— PIE > Iranian > Turkic > Mongolian (Golden Horde)
— Uralic
— Anatolian > Armenian > Iranian > Turkic > Mongolian
— Loanwords and technology (agriculture!)

Eurasia: known and reconstructed history

- The Silk Road(s)

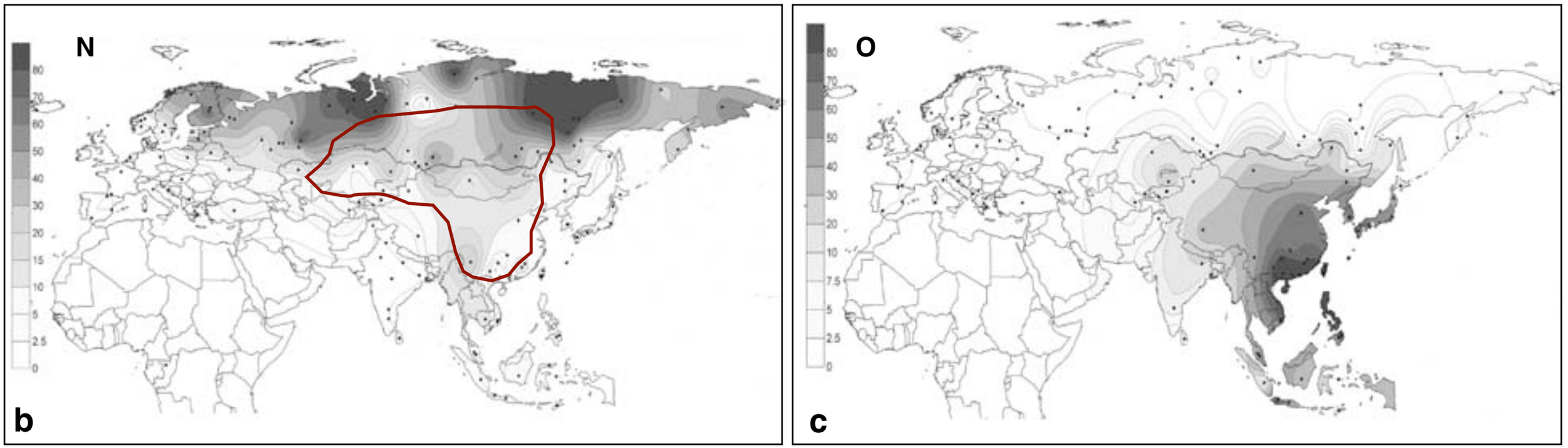


Eurasia: known and reconstructed history

- Religious traditions in South, East and Southeast Asia
- Religious traditions in Mesopotamia and Western Eurasia
- Religious traditions in Central Asia and the Himalayan Plateau (and adjacent regions)
- Culinary traditions in South and Southeast Asia

Eurasia: genetics

N vs. O Y-chromosomal haplogroup split about 34ky ago suggests a “counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe” (Rootsi et al. 2007):



— Approximate spread zone of the most widely spread mtDNA haplogroups

Eurasia: genetics

- Known cases of language shift:

	Genealogy	mtDNA	Y
Finnish	Uralic	European	Partly Sibirian
Hungarian	Uralic	European	
Azerbaijani	Turkic	Caucasian	

- Probably many more such cases since language loyalty is known to have been very low in Northern Eurasia (Nichols 1998)

Eurasia

- All this evidence together suggests strong effects of EURASIA onto the distribution of many typological variables.
- Preliminary results from the *World Atlas of Language Structures:* 30 out of 100 test positive for EURASIA effects, after
  - controling for known universal correlations and
  - adjusting significance levels for accidental hits due to multiple testing
- and we keep finding more, e.g. in the case ~ word order model!

Overall conclusions

- Understanding typological distributions requires
  - developing explicit theories about explanatory factors:
    - ‘match’ factors: principles leading to universally uniform structural pressure in diachrony
    - ‘spread’ factors: effects of language contact and faithful inheritance within families
  - deriving statistical models from this and
  - testing these models against fine-grained variables in sufficiently rich databases (where ‘rich’ is not necessarily the same as big!)

Some common misconceptions

- "Typology is about synchronic classification of languages."

  ✶*No, it is about measuring similarities of linguistic structures across and within languages and explaining them in terms of diachronic factors*

- "Typology is functionalist."

  ✶*Only true for 'match' theories; not for 'spread' theories, and both are important!*

- "Typology doesn't care about formal (mathematical) models."

  ✶*No, at least Distributional Typology relies on the mathematical modeling of similarities and of distributions.*

- "Typology doesn't care about within-language diversity."

  ✶*No, if we use the right set-up of variables and databases!*