A corpus-based analysis of ongoing change in the adjective amplifier systems of Hong Kong, Indian, and Philippine English

Dr. Martin Schweinberger

slides available at

www.martinschweinberger.de

m.schweinberger@uq.edu.au

R code upon request

Phenomenon: Adjective Amplification

(1) And you just have to hint well then it's a very good hint (ICE-AUS:S1A-012$A)

(2) They're all really cheap <#> They're all really nice, the t-shirts in there (ICE-AUS:S1A-009$B)

(3) It was so bad (ICE-AUS:S1A-044$B)

Intensification

Related to the semantic category of *degree* (degree adverbs) and ranges from low (downtoning) to high (amplifiers)

(?: 589–590)

- Amplifiers
  - Boosters, e.g. *very*
  - Maximizers, e.g. *completely*
- Downtoners
  - Approximators, e.g. *almost*
  - Compromisers, e.g. *more or less*
  - Diminishers, e.g. *partly*
  - Minimizers, e.g. *hardly*

Motivation

Amplification

- major area of gramm. change (cf. **?**: 441)
- crucial for "social and emotional expression of speakers" (**?**: 258)
- linguistic subsystem which allows precise circumscription of a variable context (**??**: 49)
- ideal case for testing mechanisms underlying language change!

Previous Research

Amplification

- substantial amount of corpus-based research on intensification (e.g ??????)
  → but mostly either focused on individual intensifiers or without regard to the intensified adjectives
- recently amplifier-adjective bigrams have come more into focus (e.g. ???)
- associated with teenage talk and young(ish) (female) speakers
  (?????)

Focus

- Amplifying *really* replaces *very* (lexical replacement)

(see ? for NZE; see ? and ? for North East British English, ? and ? for Toronto English; see ? for South Eastern Ontario English)

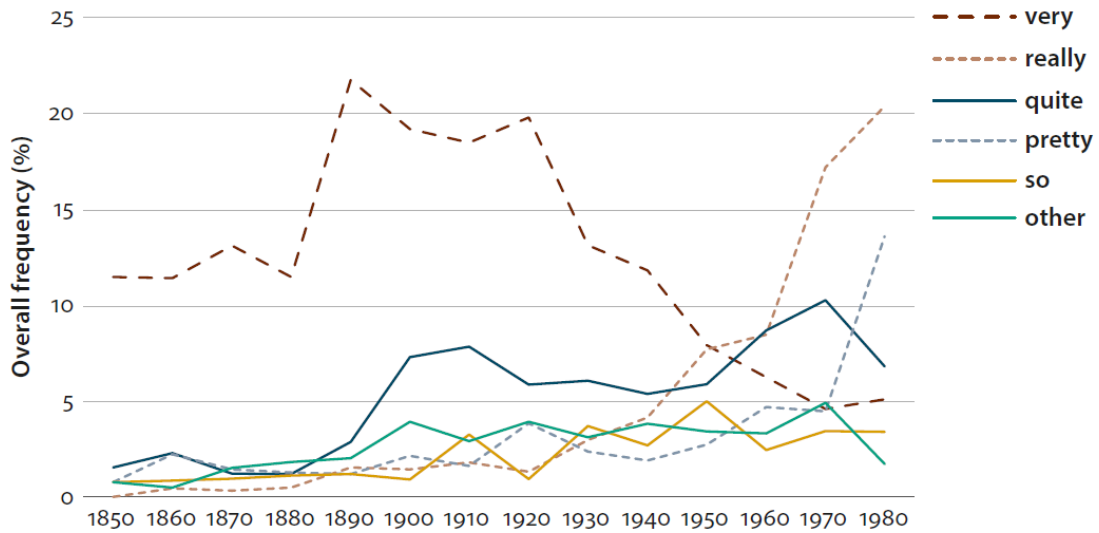


Figure 1: Amplifier variants in NZE across real-time (from ?: 468)

Research Question

Q

Are Asian Englishes in-line with the trend among traditional L1 and other post-colonial varieties of English in that *very* is being replaced by *really*?

Hypotheses

$H_1 1$

HKE, IndE, PhiE align with other varieties:
really is replacing *very*
as the default adjective amplifier

$H_1 2$

Aligned with AmE/AusE → more advanced
Aligned with GBE → more lagging behind

Data and Methodology

Corpus data: International Corpus of English (ICE)

- Hong Kong, Indian, and Philippine ICE components
- Shared design (allows meaningful comparisons between varieties of English)
- One million words (600,000 spoken and 400,000 written) from diverse spoken and written text types (cf. next slide) with each file containing app. 2,000 words.
- Accompanied by metadata and biodata of speaker (extremely interesting resource for variationist analyses)

Corpus data: International Corpus of English (ICE)

Mode	Conversation type	Register	Text type	Number of text files
SPOKEN (300)	Dialogues (180)	Private (100)	Face-to-face conversations	90
			Phonecalls	10
		Public (80)	Classroom Lessons	20
			Broadcast Discussions	20
			Broadcast Interviews	10
			Parliamentary Debates	10
			Legal cross-examinations	10
			Business Transactions	10
	Monologues (120)	Unscripted (70)	Spontaneous commentaries	20
			Unscripted Speeches	30
			Demonstrations	10
			Legal Presentations	10
		Scripted (50)	Broadcast News	20
			Broadcast Talks	20
			Non-broadcast Talks	10

Data Processing

- Spoken private dialogue section of each component
- Part–of-speech tagged (OpenNLP vis R) the
- Retrieved adjectives (PoS–tag JJ)
- Determined whether adjective were preceded by an amplifier (member of a predefined set of amplifiers)
- Sentiment Analysis of adjective types (?)

Data Processing

- Determined if the same amplifier type had occurred within a span of three adjective slots previously ($\rightarrow$ Priming)
- Token freq. of adjective type by age group (?)
- Removed...
  - negated adjectives
  - comparative and superlative forms
  - adjectives that were not amplified by at least two different amplifier types
  - adjectives that were preceded by downtoners
  - strange forms (e.g. *much, many*)

Data Processing

- Semantic classification of adjective (simplified version of **?**, cf. also **????**)
- Manual cross–evaluation of automated classification
- Metadata and speaker information

Variable Coding

Dependent Variable(s)				
Variant	nominal	yes/no occurrence of pre-adjectival *really, so, very*		
Independent Variable(s)				
Age	ordinal	min. young | middle-aged | old	extra linguistic	
Gender	nominal	Female | Male	extra linguistic	
(Education)	nominal	College | NoCollege	extra linguistic	
Priming	nominal	prime | noprime	extra linguistic	
Emotionality	categorical	negative | nonemotional | positive	intra linguistic	
Function	nominal	attributive | predicative	intra linguistic	
SemanticCat.	categorical	semantic category of adj.	intra linguistic	
Gradability	numeric	Gradability score based on BNC	intra linguistic	
Adjective	categorical	bad | funny | good | interesting | nice | other	intra linguistic	
Frequency	numeric	Frequency of adj. by age group	intra linguistic	

Statistical analysis

Statistical analysis: options

Model	Dep. var.	Interactions	Nested data	Output	Overfitting
Tree-based models					
Conditional-Inference Tree (CIT)	✓	✓	X	✓	X
Random Forests (RF)	✓	✓	X	X	✓
Regression models					
Multinomial mixed-effects (GLMMM)	✓	✓	✓	X	✓
Logistic mixed-effects (GLMBLM)	X	✓	✓	✓	✓

→ CIT + run separate GLMBLMs for each variant

Generalized Linear Mixed-Effects Binomial Logistic Regression Model

(??)

What is GLMBLM?

- Standard model for multivariate analyses
- Can handle nested/grouped data structure
- Easy multicollinearity detection

Problems of GLMBLM

- Cannot handle small data sets (well)
- Extremely high β-error rate (?)
  - if sig. effect: ✓
  - if no sig. effect: ???

Results

Results

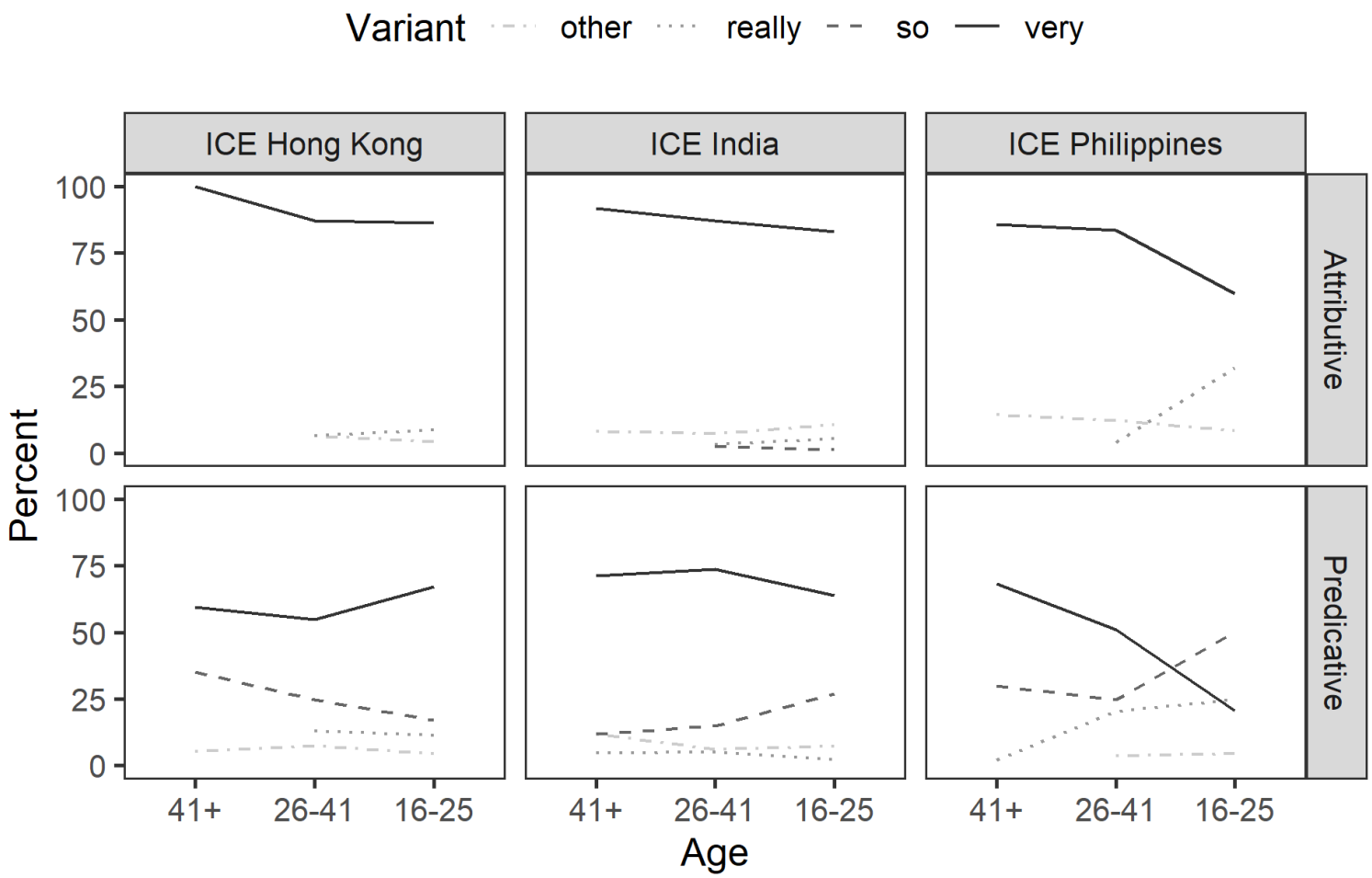


Figure 2: % variants across speaker age by function and corpus.

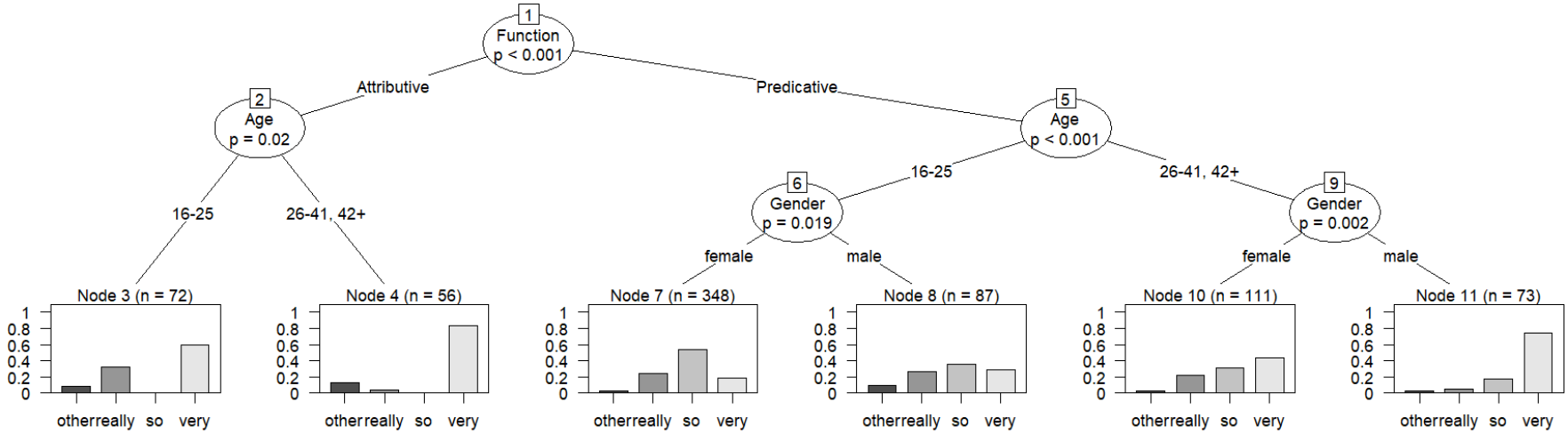


Figure 3: Results of the CIT analysis on ICE-PHI.

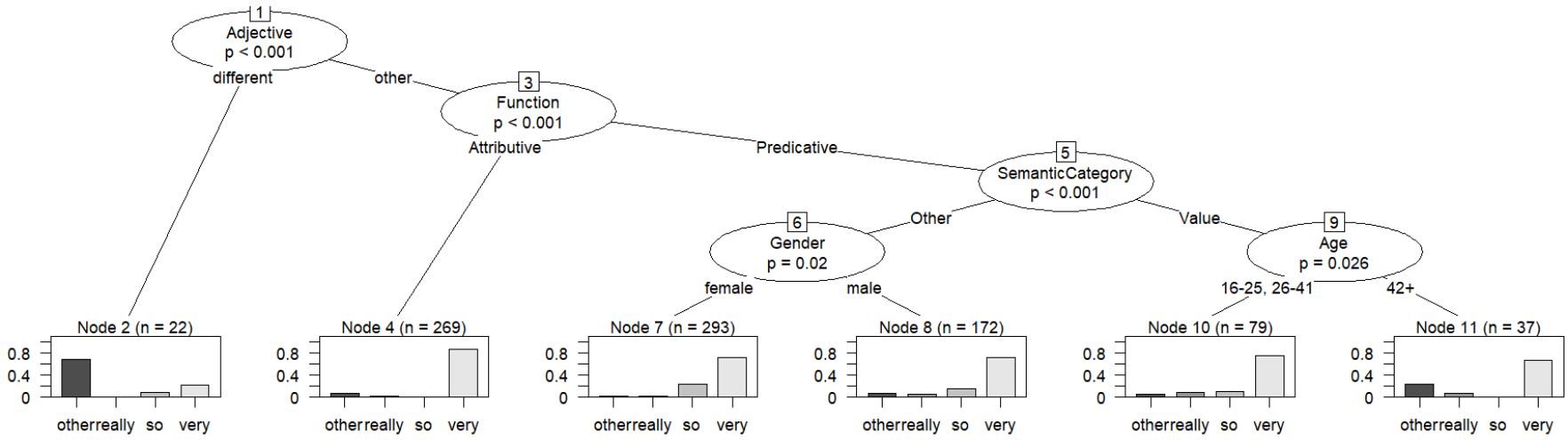


Figure 4: Results of the CIT analysis on ICE-IND.

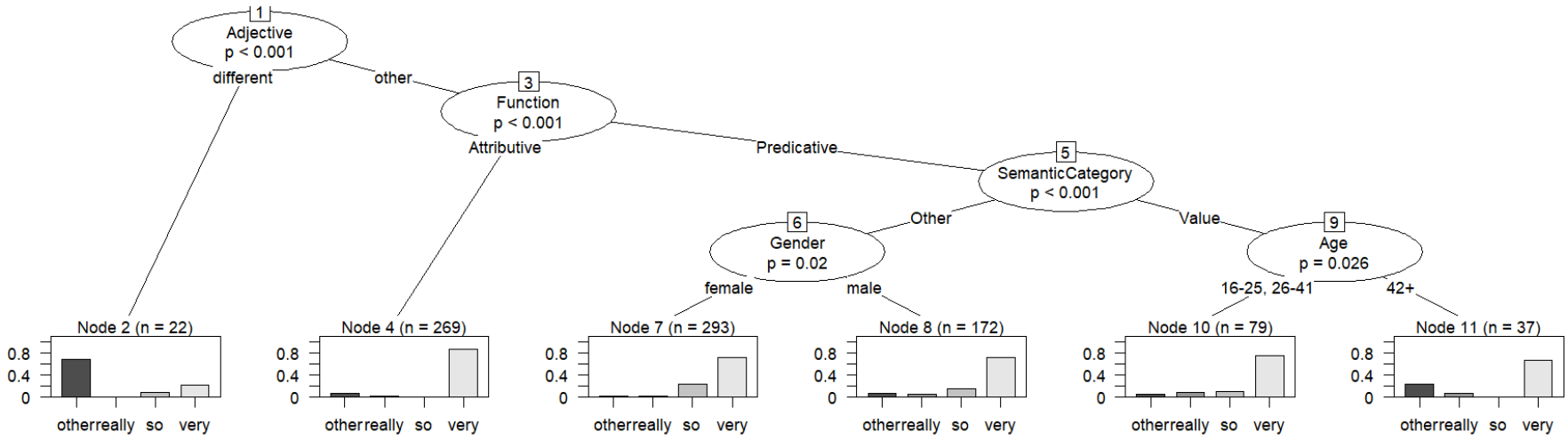


Figure 5: Results of the CIT analysis on ICE-HK.

Results *very* in ICE-PHI

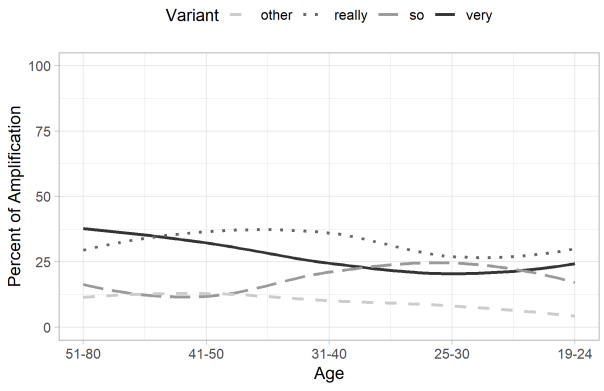


Figure 6: % Variants in CanE.

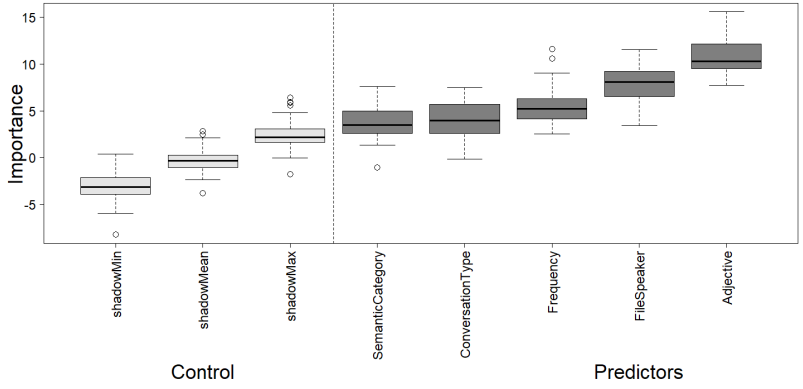


Figure 7: Boruta results for really in CanE.

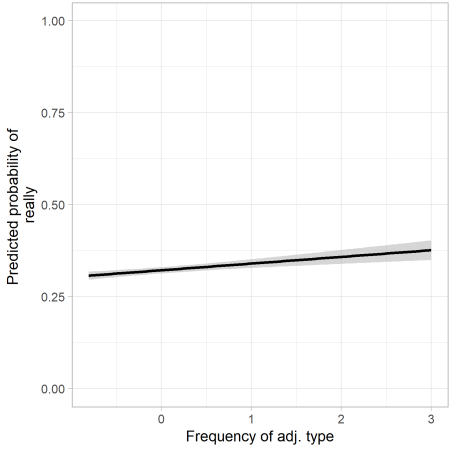


Figure 8: Prob. really in CanE by adj. freq.

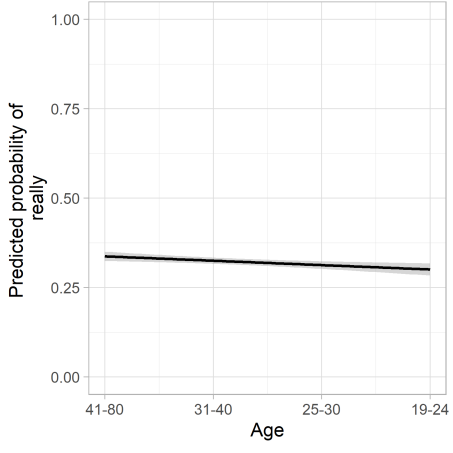


Figure 9: Prob. really in CanE across age.

Discussion & Outlook

Summary

The analysis . . .

- confirms that *really* correlates with adj. freq. (positive correlation between the use of *really* and adjective frequency)
- suggests that lexical replacement is accompanied by (functional) re-organization in addition to diffusion through the speech community (absence of age effects) (see **?**)
- shows that complementing mixed-modeling with Boruta is useful to avoid overlooking significant effects (avoidance of β-errors)

Discussion

- *Really* successfully replaced the dominant form *very* because it collocated with HFAs.
- No signs that *really* of broadening before taking over the system.
- Broadening once dominant (substantiates ? ?)

Argument

1. The co-occurrence with HFAs lead to the innovative variant being used as a more expressive variant to amplify certain HFAs.
2. The frequency of the innovative form increased because it piggybacked on the frequency of the HFA.
3. Increase in use $\rightarrow$ more deeply entrenched.
4. Deeper entrenchment $\rightarrow$ increased ease of retrieval.
5. Higher ease of retrieval $\rightarrow$ advantage over rival variants.
6. Innovative variant broadens because it increasingly co-occurs with more adj. types.

Outlook

Could this be a universal mechanism?

Test if the mechanisms. . .

- can be shown to have worked in analogous changes in English, e.g.
  3^{rd} p. sg. ind. morpheme: <eth> → <(e)s>
- can be shown to have worked in analogous changes in languages other than English

Thank you so, really, very much!

Acknowledgements

I would like to thank...

the ICE teams who have compiled the data for the current study (without them the current study would not have been possible)

my colleagues at UQ

for comments and their feedback on earlier versions of this talk

Aijmer, K. (2011). Are you totally spy? a new intensifier in present-day american english. In S. Hancil (Ed.), *Marqueurs discursifs et subjectivité*, pp. 155–172. Rouen: Universités de Rouen and Havre.

Aijmer, K. (2018). That's well bad. some new intensifiers in spoken in british english. In V. Brezina, R. Love, and K. Aijmer (Eds.), *Corpus Approaches to Contemporary British English*, pp. 60–95. New York and London: Routledge.

Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University press.

Barnfield, K. and I. Buchstaller (2010). Intensifiers on tyneside - longitudinal developments and new trends. *English World-Wide 31*(3), 252–287.

Bauer, L. and W. Bauer (2002). Adjective boosters in the english of young new zealanders. *Journal of English Linguistics 30*, 244–257.

Brinton, L. J. and L. K. Arnovick (2006). *The English Language: A Linguistic History*. Oxford: Oxford University Press.

D'Arcy, A. F. (2015). Stability, stasis and change - the longue duree of intensification. *Diachronica 32*(4), 449–493.

Dixon, R. M. W. (1977). Where have all the adjectives gone? *Studies in Language 1*, 19–80.

Faraway, J. J. (2016). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*, Volume 124. CRC Press.

Fuchs, R. (2016). Register variation in intensifier usage across asian englishes. In H. Pichler (Ed.), *Discourse-Pragmatic Variation and Change: Insights from English*, pp. 185–213. Cambridge: Cambridge University Press.

Fuchs, R. (2017). Do women (still) use more intensifiers than men? *International Journal of Corpus Linguistics 22*(3), 345–374.

Ito, R. and S. Tagliamonte (2003). Well weird, right dodgy, very strange, really cool: Layering and recycling in english intensifiers. *Language in Society 32*, 257–279.

Jockers, M. L. (2017). Syuzhet 1.0.4 now on cran. http://www.matthewjockers.net/2017/12/16/syuzhet-1-0-4/.

Johnson, D. E. (2009). Getting off the goldvarb standard: Introducing rbrul for mixed-effects variable rule analysis. *Language and linguistics compass 3*(1), 359–383.

Kursa, M. B., W. R. Rudnicki, et al. (2010). Feature selection with the boruta package. *J Stat Softw 36*(11), 1–13.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington DC: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.

Lorenz, G. (2002). Really worthwhile or not really significant? a corpus-based approach to the delexicalization and grammaticalization of intensifiers in modern english. In I. Wischer and G. Diewald (Eds.), *New Reflections on Grammaticalization*, pp. 143–161. Amsterdam: John Benjamins.

Macaulay, R. (2006). Pure grammaticalization: The development of a teenage intensifier. *Language Variation and Change 18*, 267–283.

Mair, C. (2004). Corpus linguistics and grammaticalisation theory. statistics, frequencies and beyond. In H. Lindquist and C. Mair (Eds.), *Corpus approaches to grammaticalization in English*, pp. 121–150. Amsterdam and Philadelphia: John Benjamins.

Méndez-Naya, B. (2003). On intensifiers and grammaticalization: The case of swithe. *English Studies 84*, 372–391.

Núñez Pertejo, P. and I. Palacios (2014). That's absolutely crap, totally rubbish. the use of intensifiers absolutely and totally in the spoken language of british adults and teenagers. *Functions of Language 21*(2), 210–237.

Palacios, I. and P. Núñez Pertejo (2012). He's absolutely massive. it's a super day. madonna, she is a wicked singer. youth language and intensification: A corpus-based study. *Text and Talk 32*(6), 773–796.

Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik (1985). *A Comprehensive Grammar of the English Language*. London & New York: Longman.

Schweinberger, M. (2017). Using intensifier-adjective bi-grams to investigate mechanisms of change. Paper presented at ICAME38. Prague, 27/5/2017.

Tagliamonte, S. (2006). "so cool, right?": Canadian english entering the 21st century. *The Canadian Journal of Linguistics/La revue canadienne de linguistique 51*(2), 309–331.

Tagliamonte, S. (2008). So different and pretty cool! recycling intensifiers in toronto, canada. *English Language and Linguistics 12*(2), 361–394.

Tagliamonte, S. and C. Roberts (2005). So weird; so cool; so innovative: The use of intensifiers in the television series friends. *American Speech 80*(3), 280–300.

Tagliamonte, S. A. and D. Denis (2014). Expanding the transmission/diffusion dichotomy: Evidence from canada. *Language 90*(1), 90–136.

Wagner, S. (2017a). Amplifier-adjective 2-grams world-wide: focus on pretty. Paper presneted at ICAME 37. Charles University Prague, 27/5/2017.

Wagner, S. (2017b). Totally new and pretty awesome: Amplifier–adjective bigrams in glowbe. *Lingua 200*, 63–83.

A corpus-based analysis of ongoing change in the adjective amplifier systems of Hong Kong, Indian, and Philippine English

Dr. Martin Schweinberger

slides available at

www.martinschweinberger.de

m.schweinberger@uq.edu.au

R code upon request

Appendix

Variationist Sociolinguistics

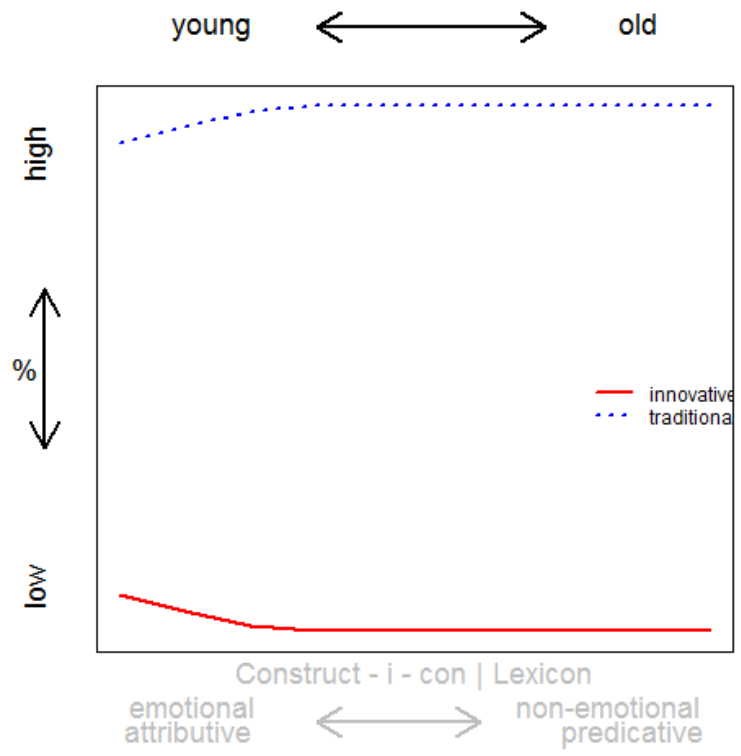
- Language is not homogeneous: variation is ubiquitous
  - Social factors : language use
  - Linguistic variation not random
  - Systematic correlation between certain social factors (age, gender, class, ethnicity, etc.) and language use
- Linguistic differentiation $\leftrightarrow$ social stratification

Diffusion of Innovations

young ⟷ old

upper

working

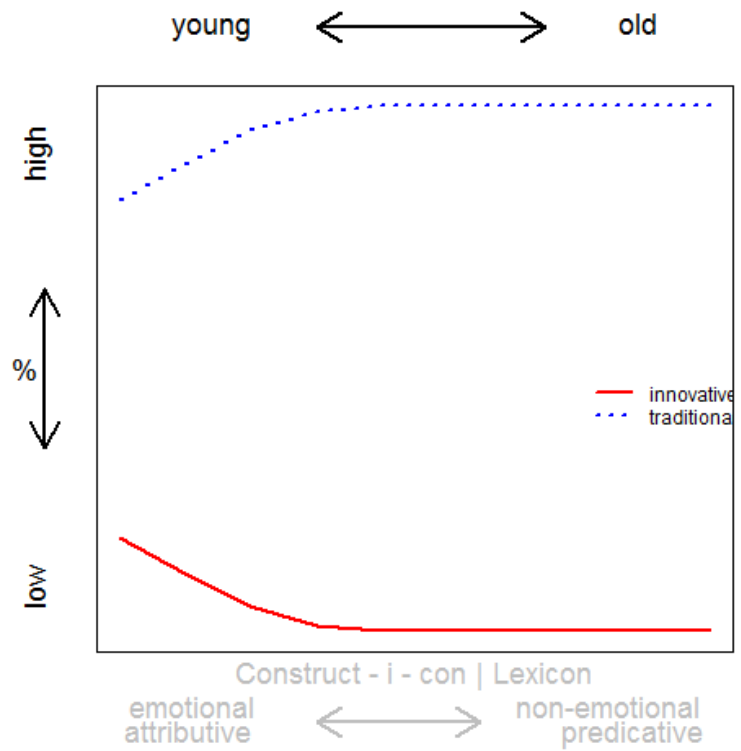


Diffusion of Innovations

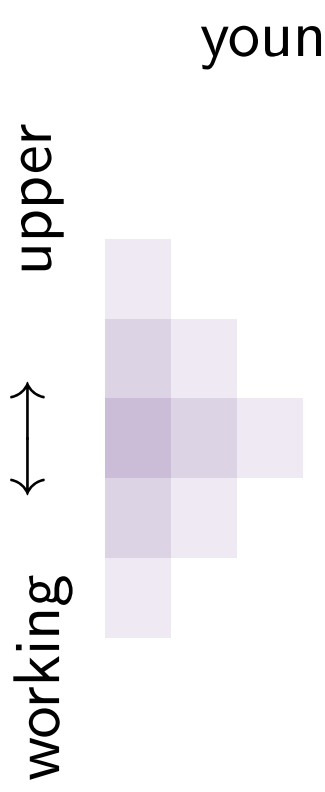
young ⟷ old

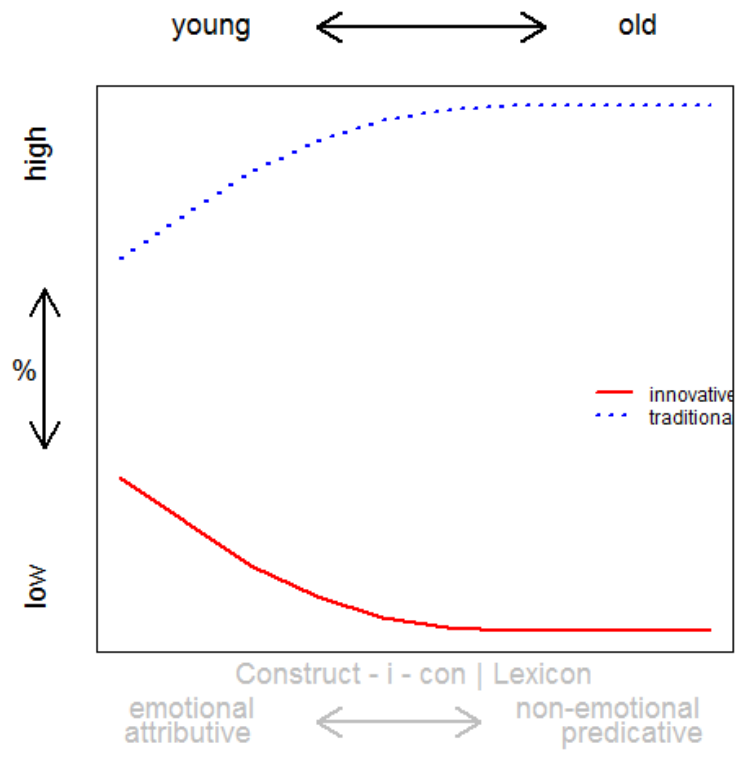
upper

working

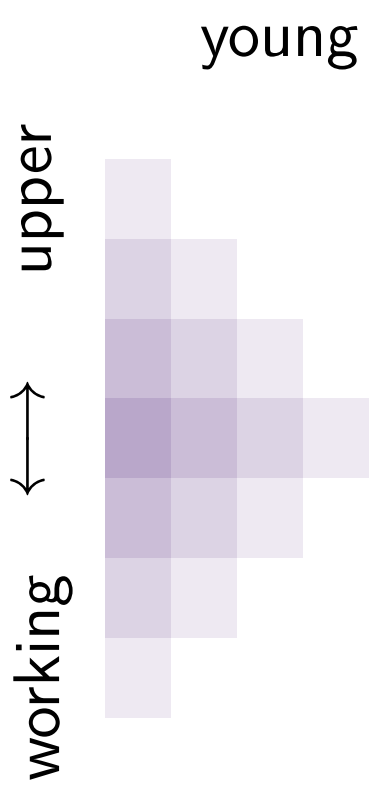


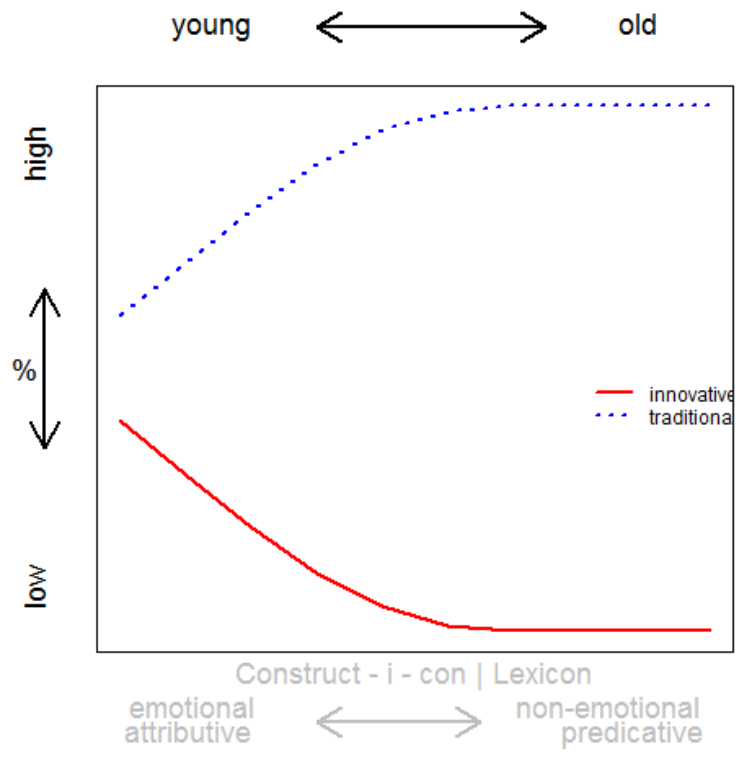
Diffusion of Innovations



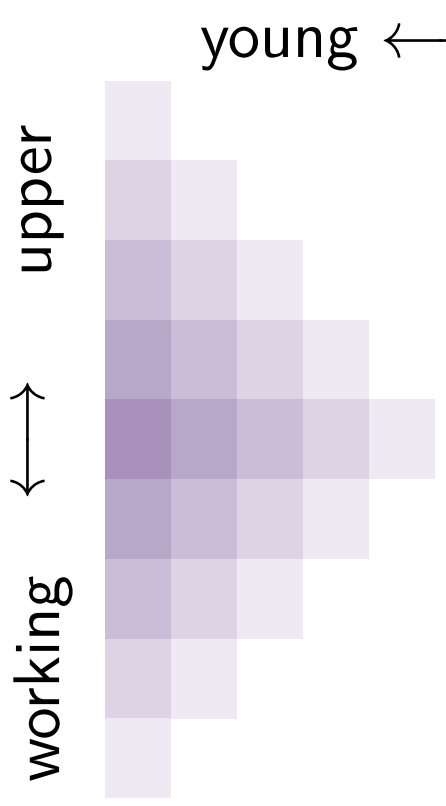


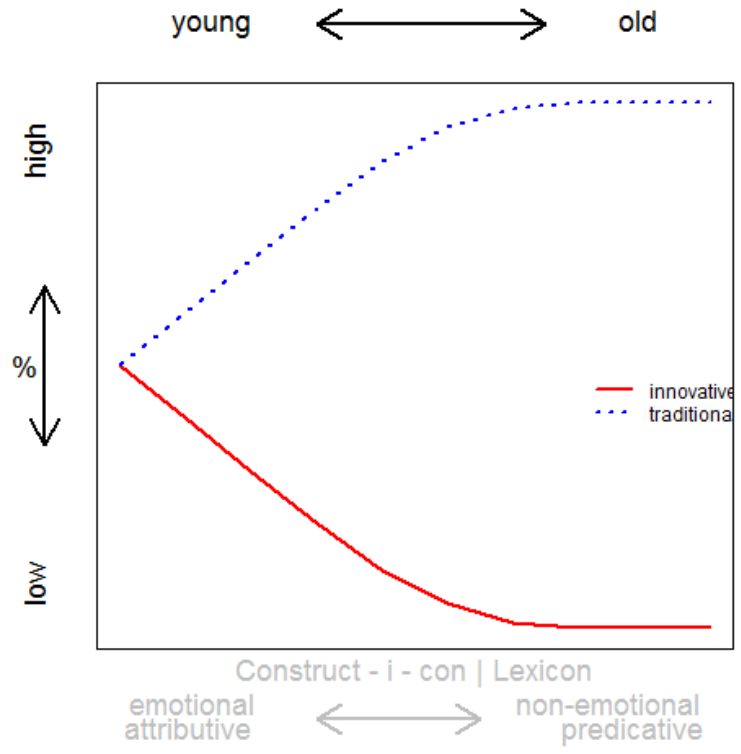
Diffusion of Innovations



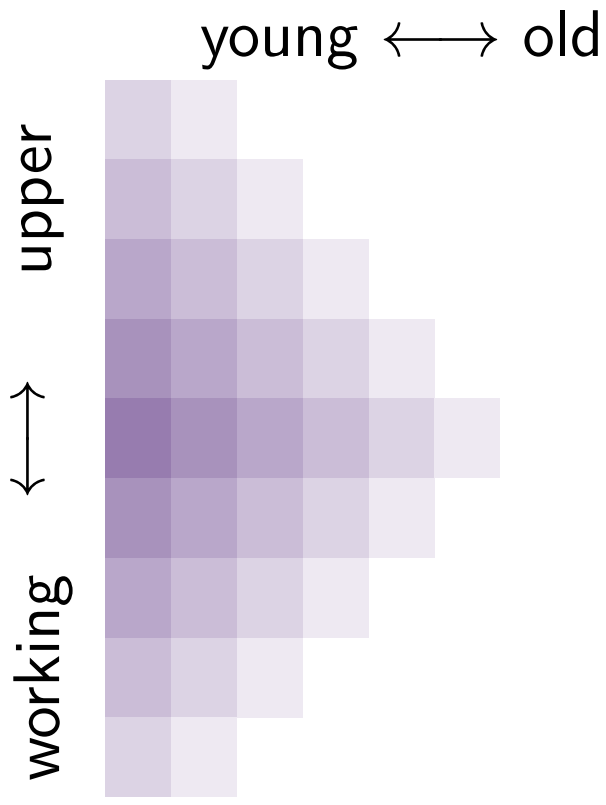


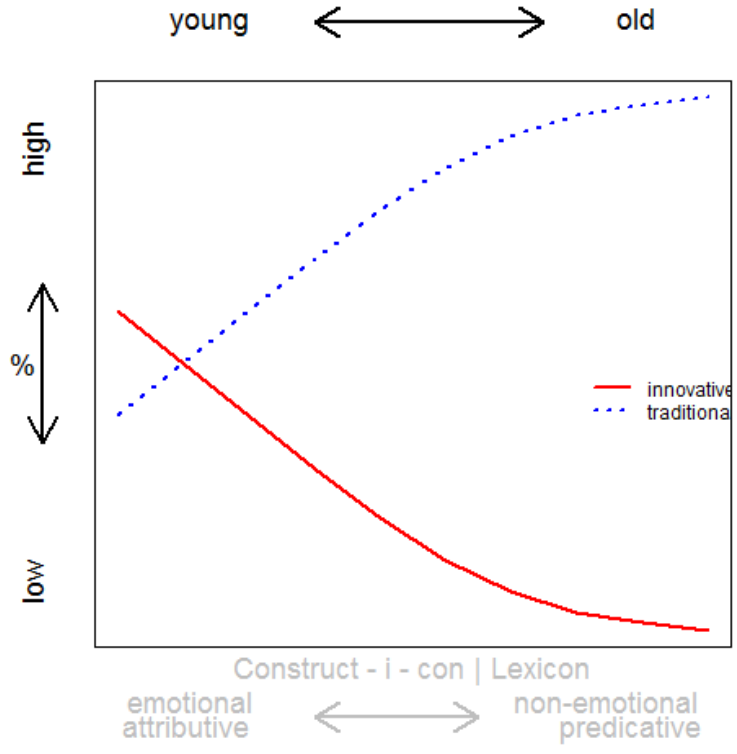
Diffusion of Innovations



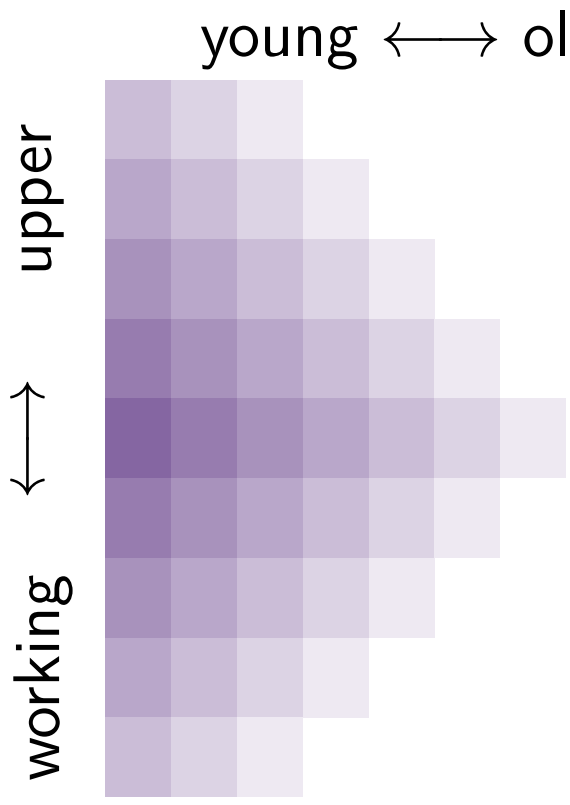


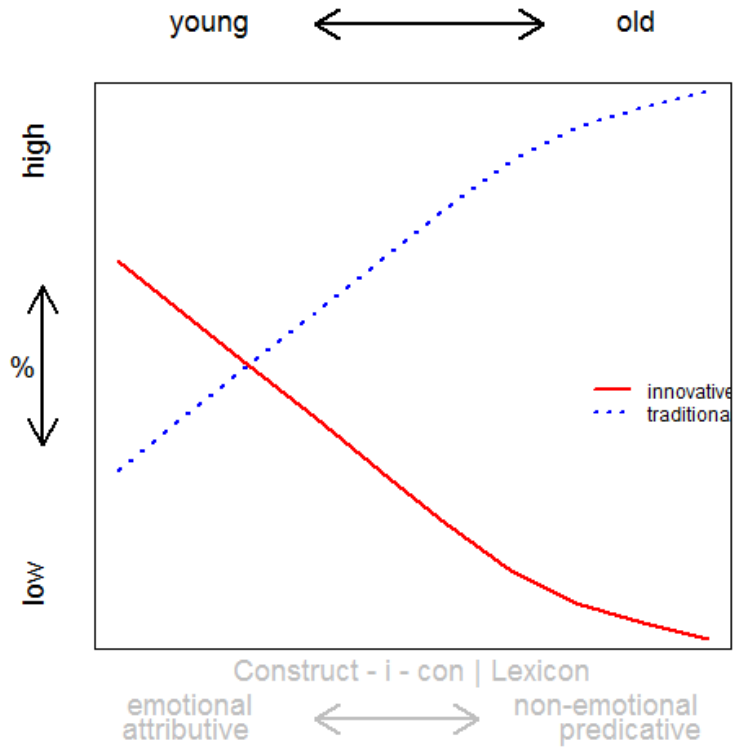
Diffusion of Innovations



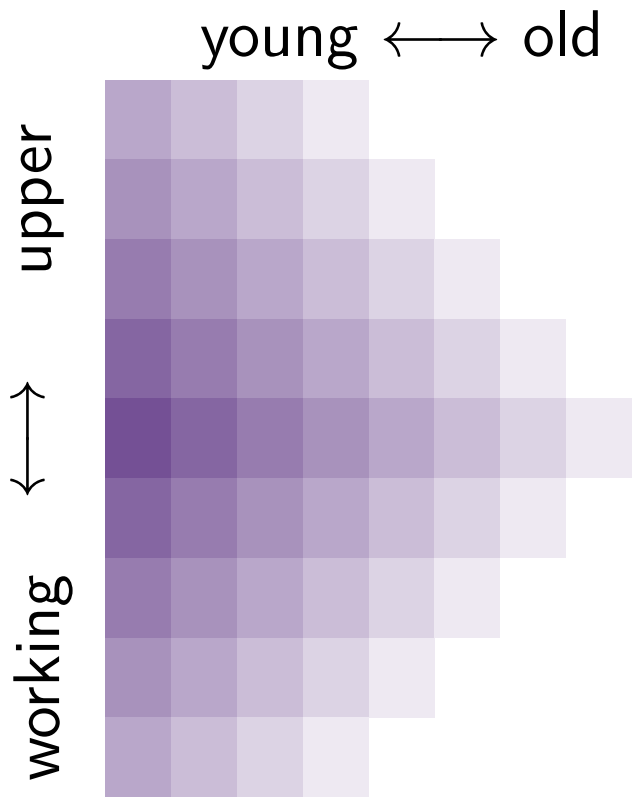


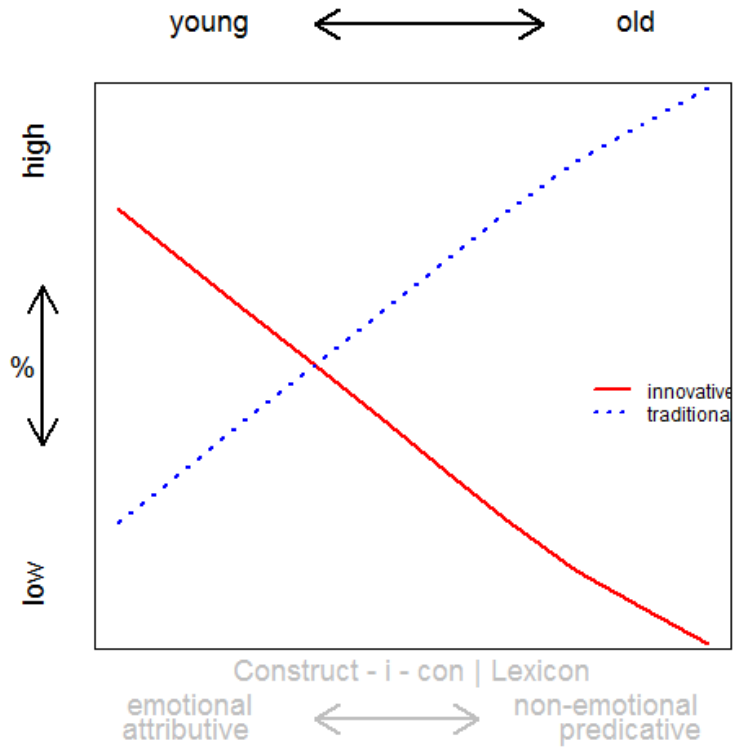
Diffusion of Innovations



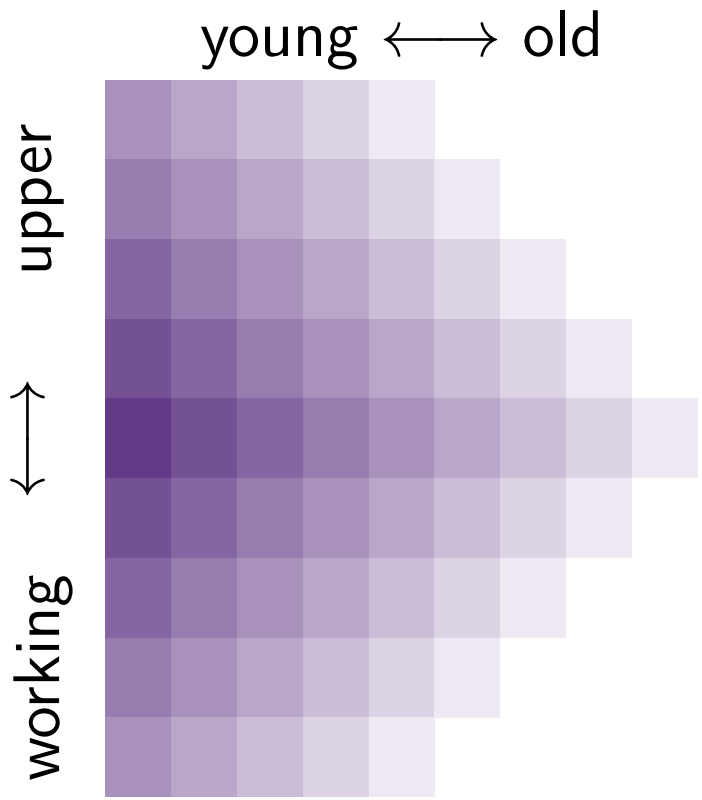


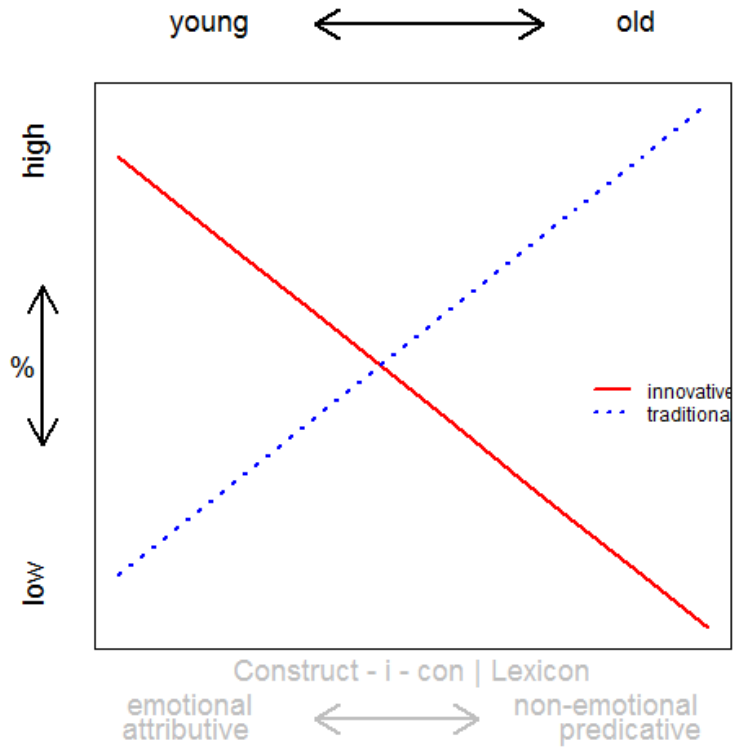
Diffusion of Innovations



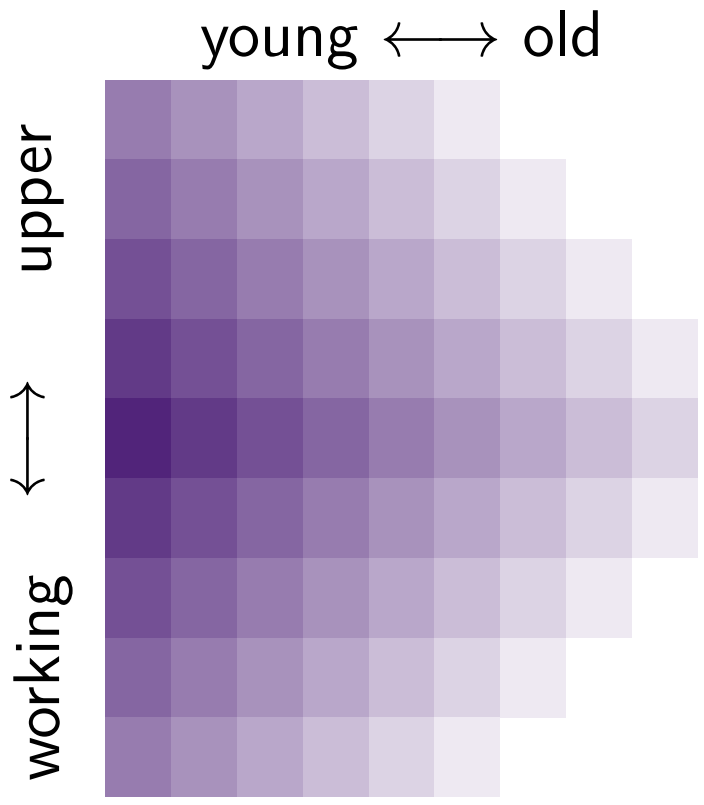


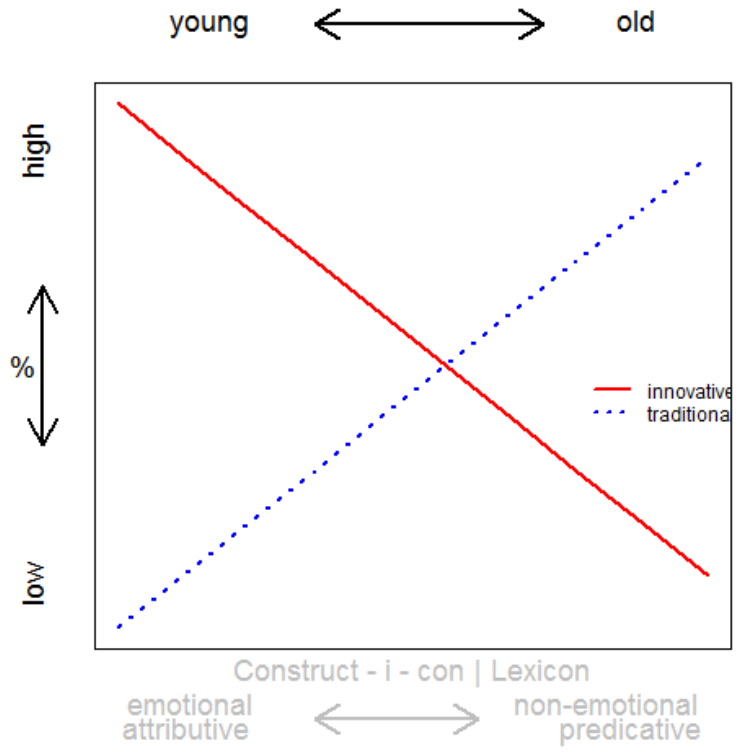
Diffusion of Innovations



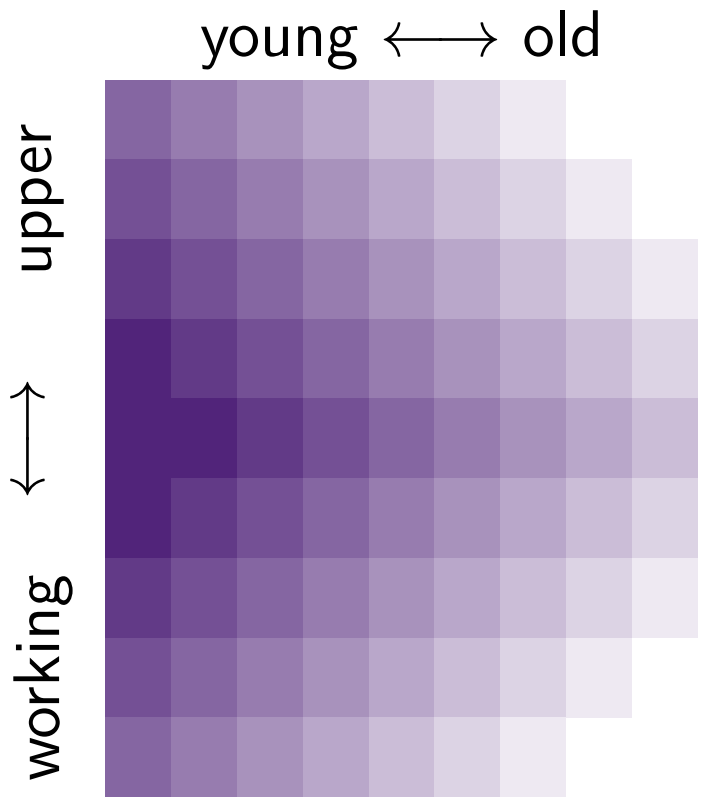


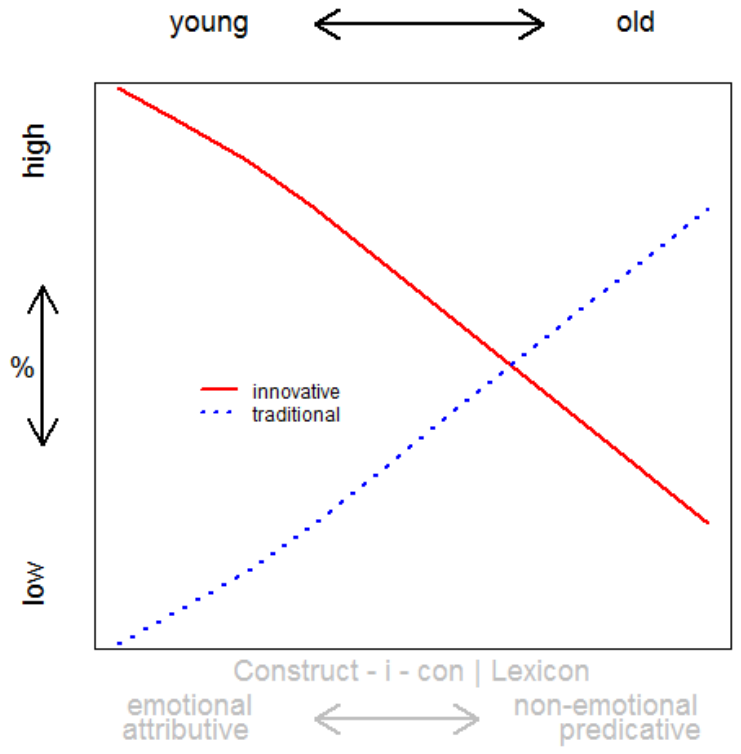
Diffusion of Innovations



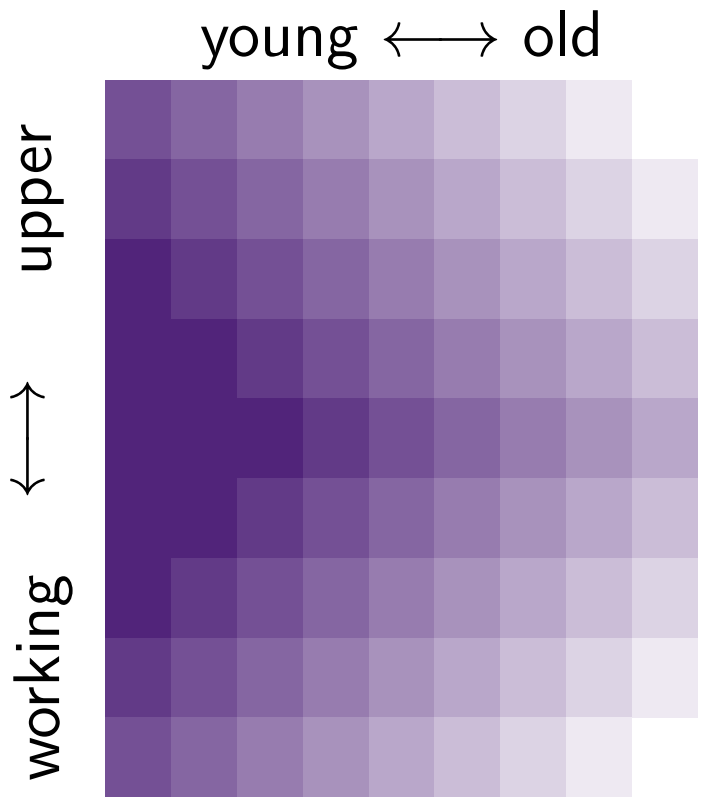


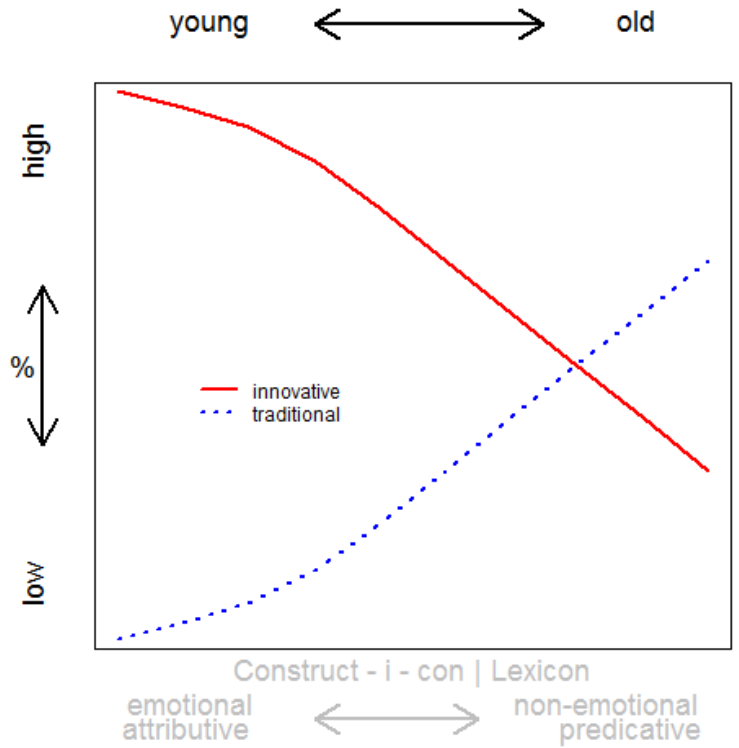
Diffusion of Innovations



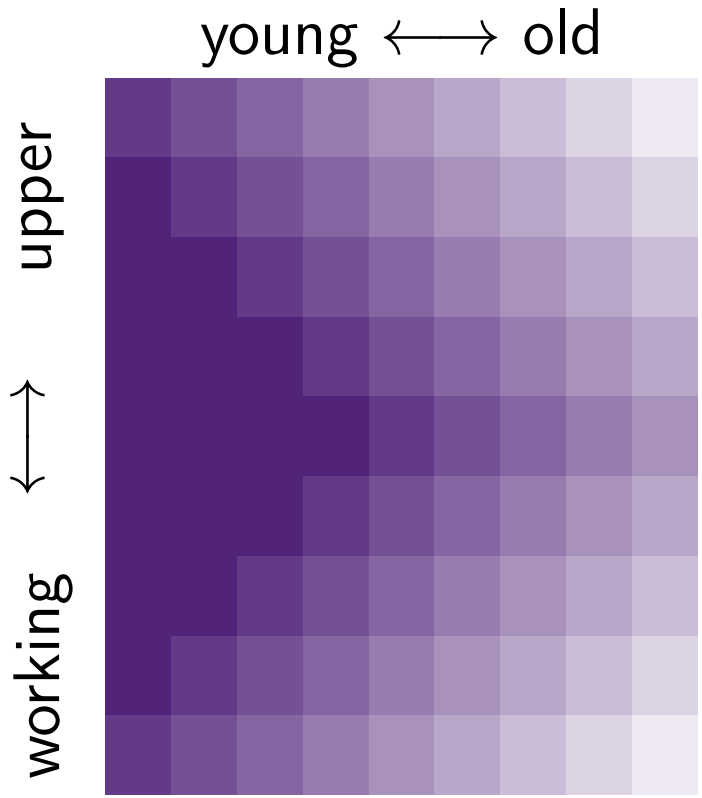


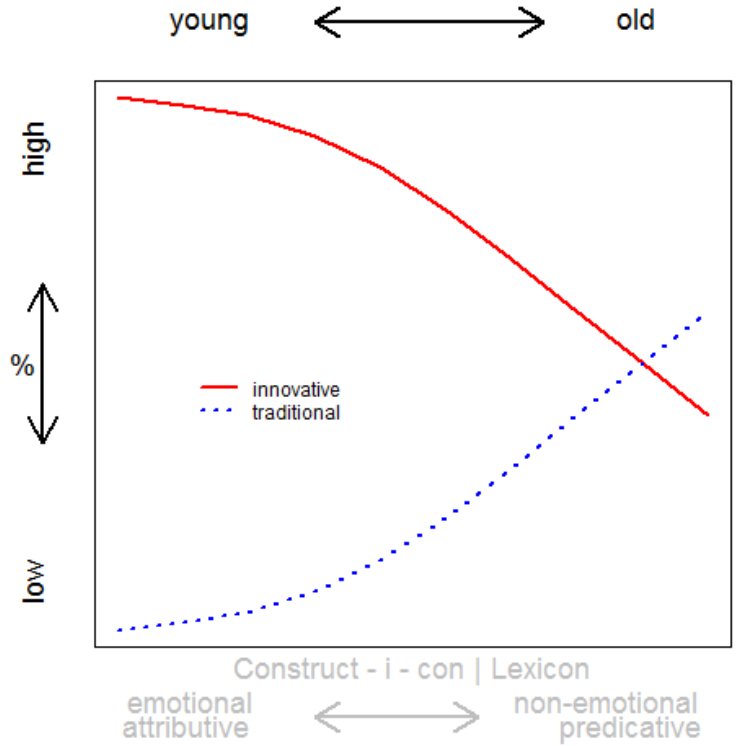
Diffusion of Innovations



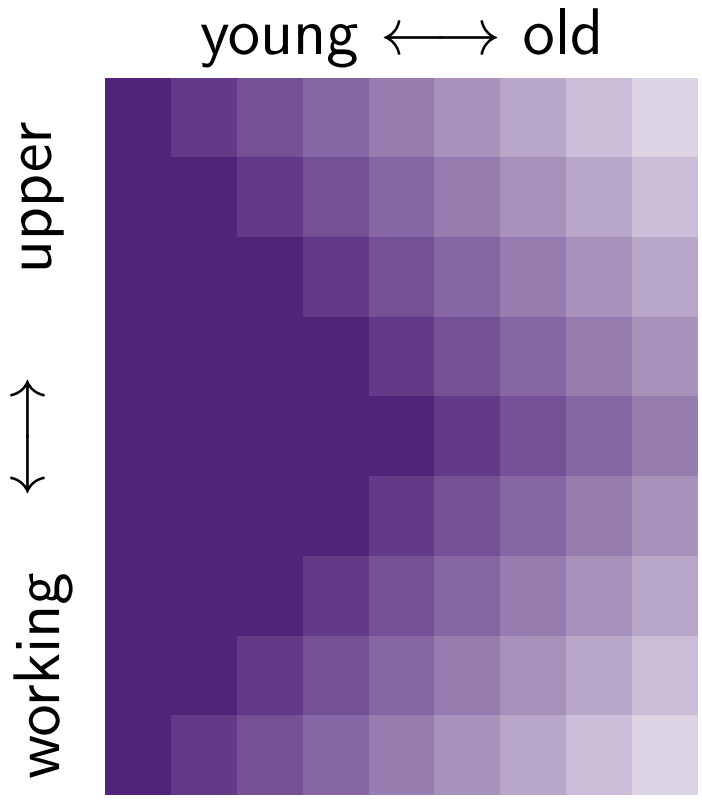


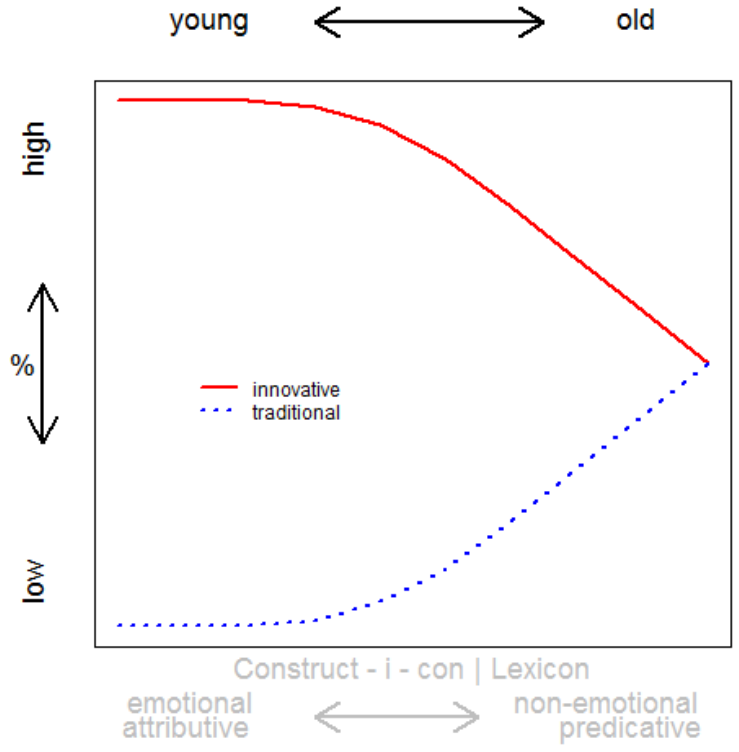
Diffusion of Innovations



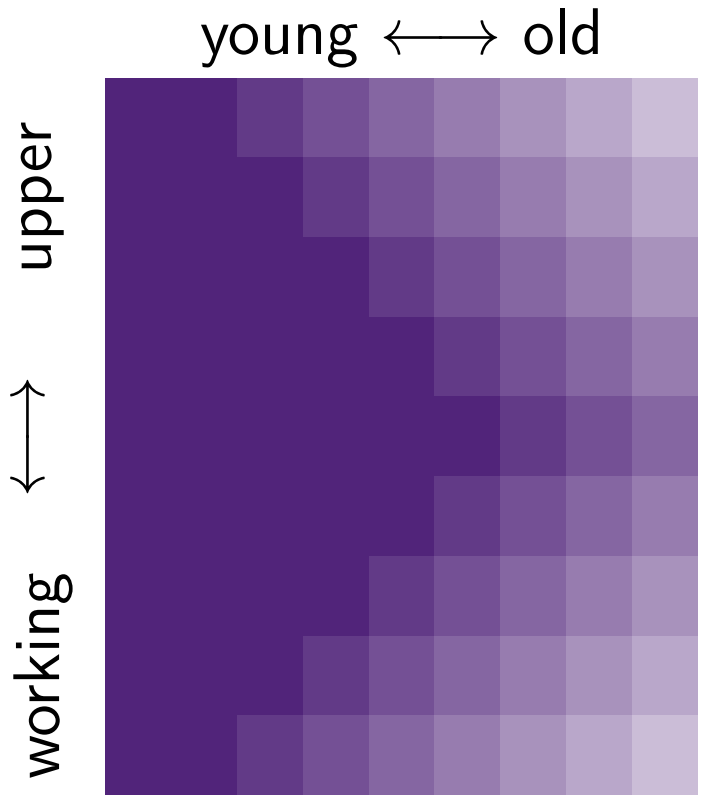


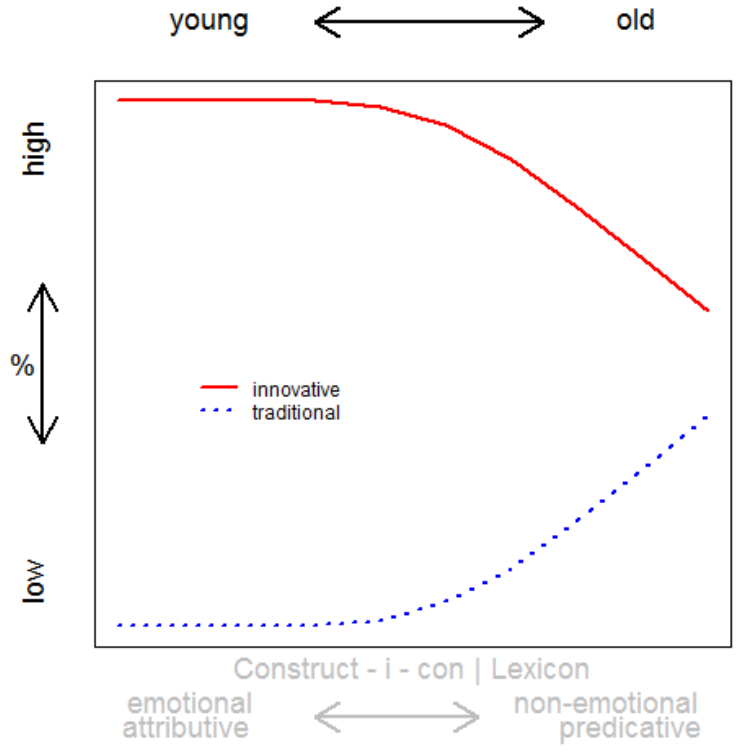
Diffusion of Innovations



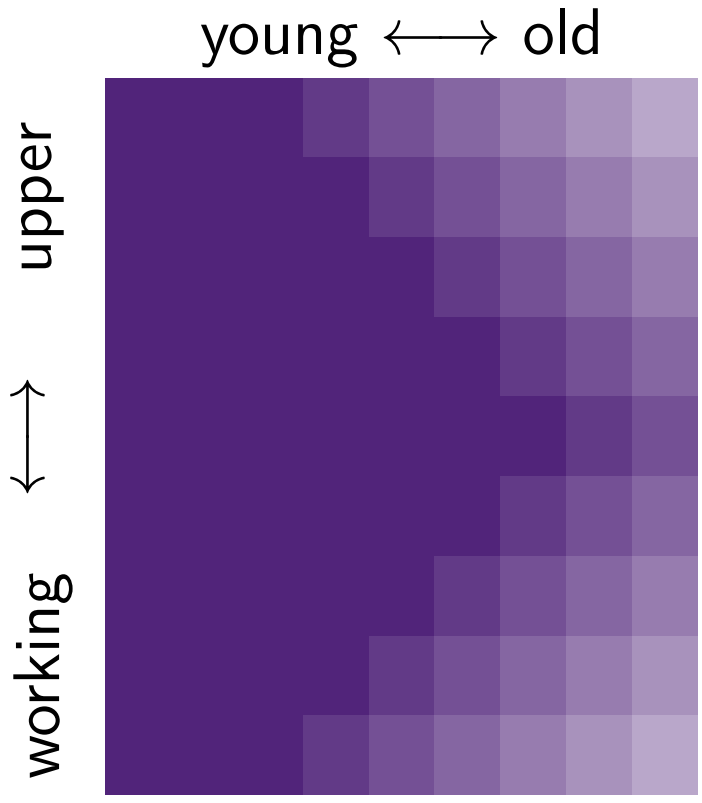


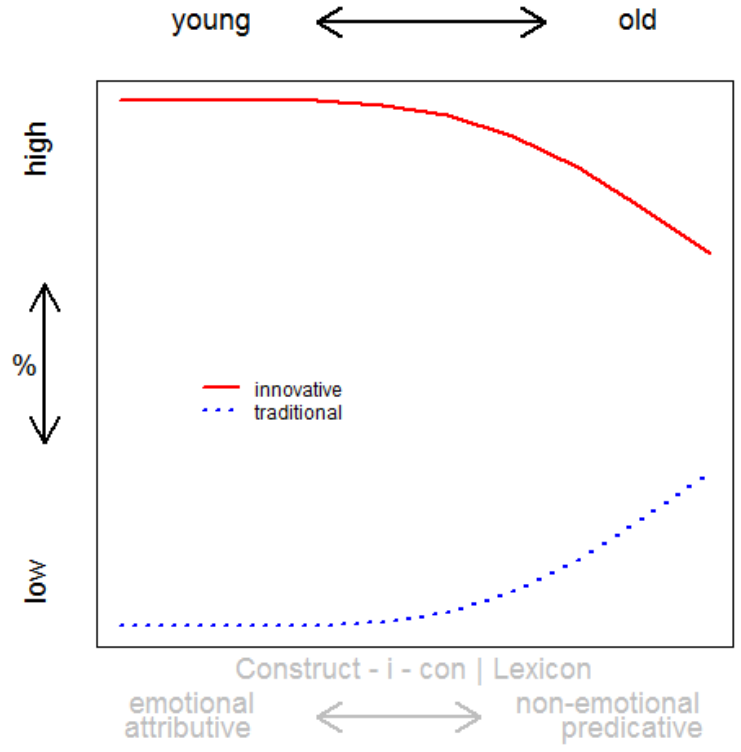
Diffusion of Innovations



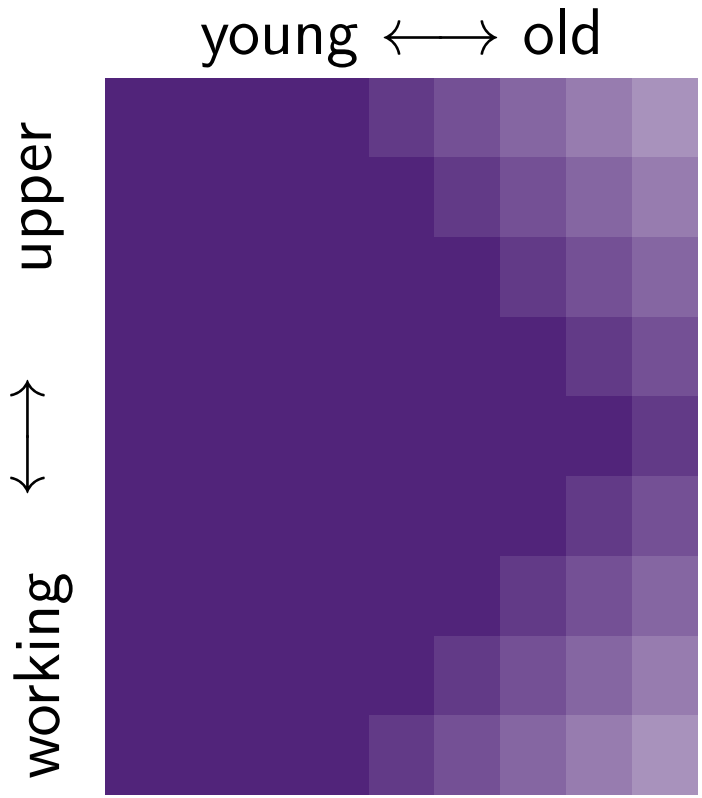


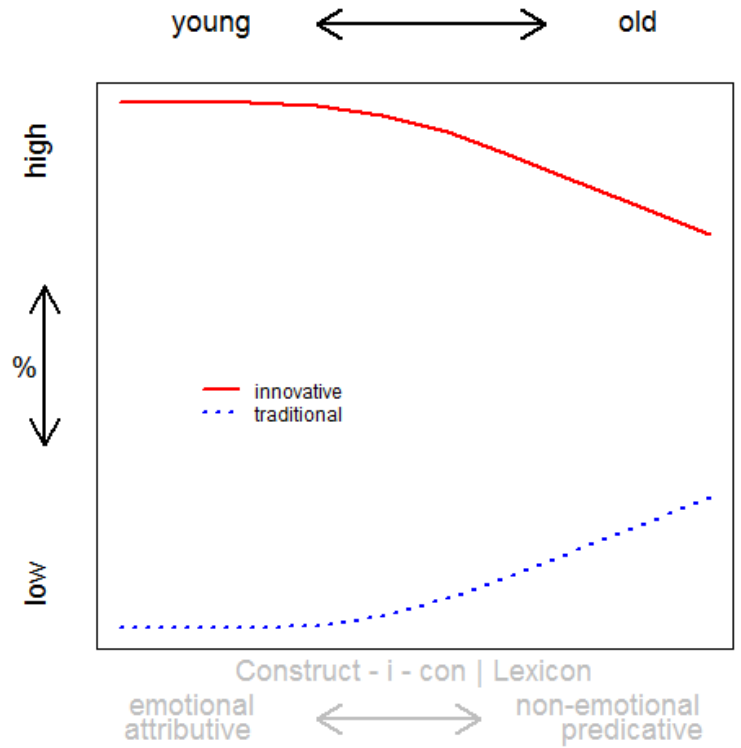
Diffusion of Innovations



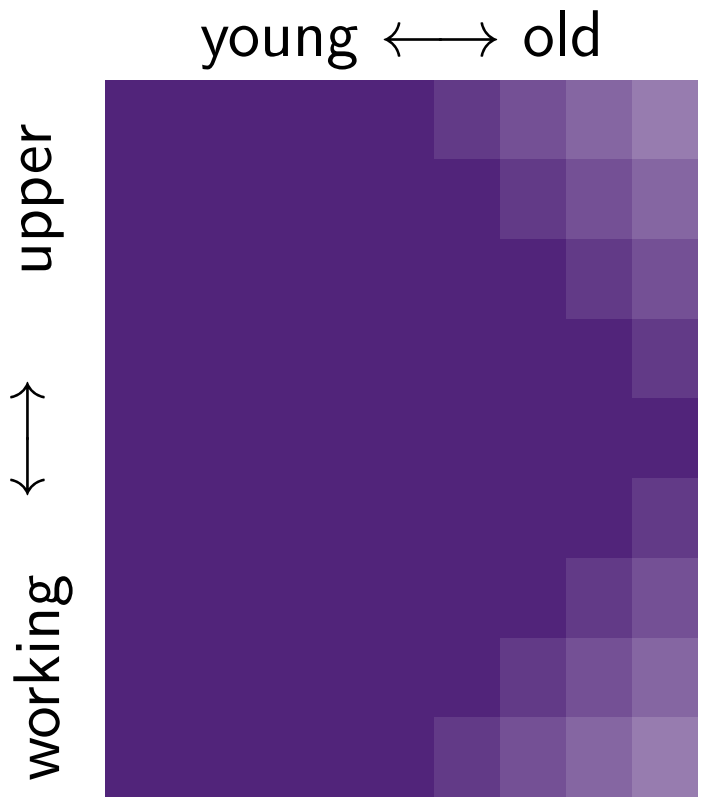


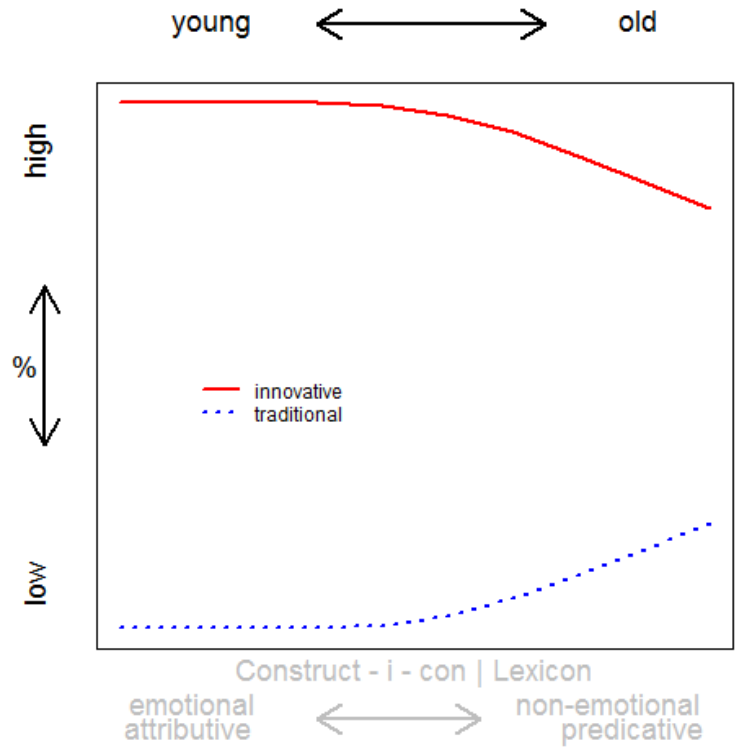
Diffusion of Innovations



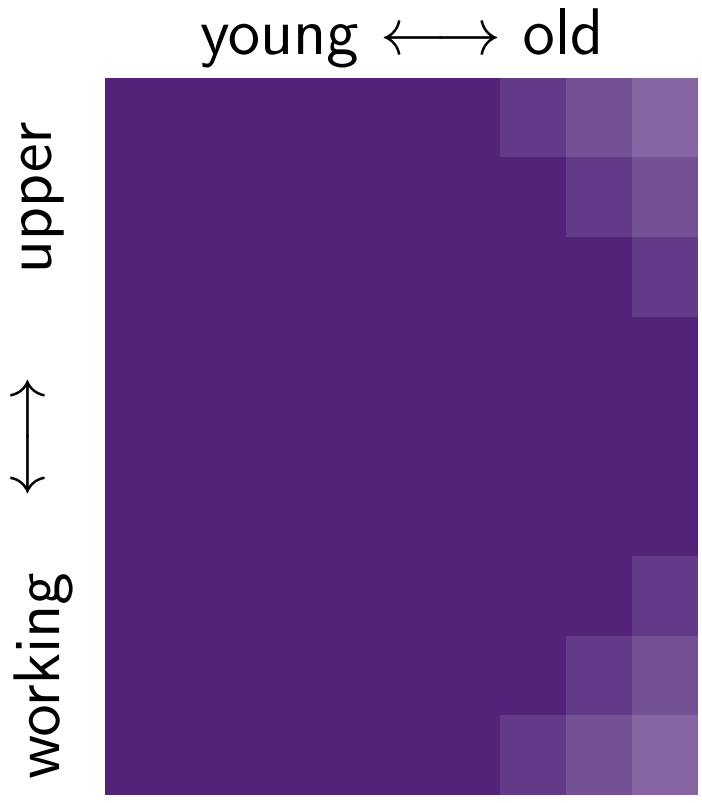


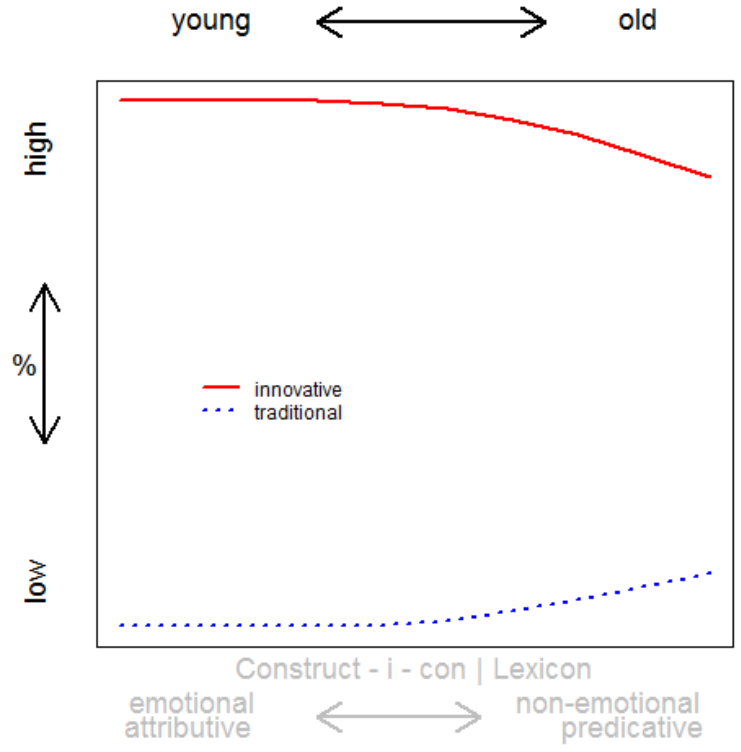
Diffusion of Innovations



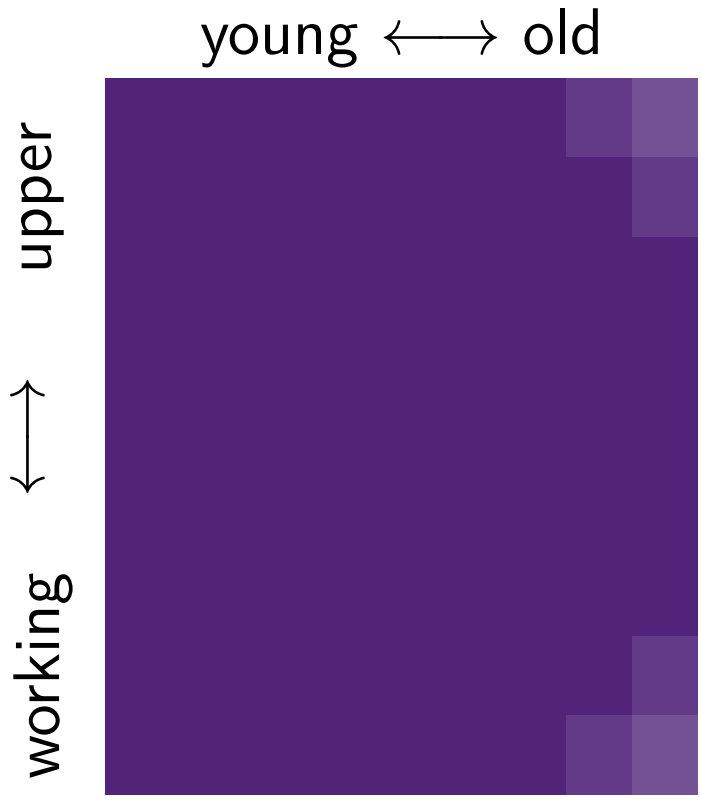


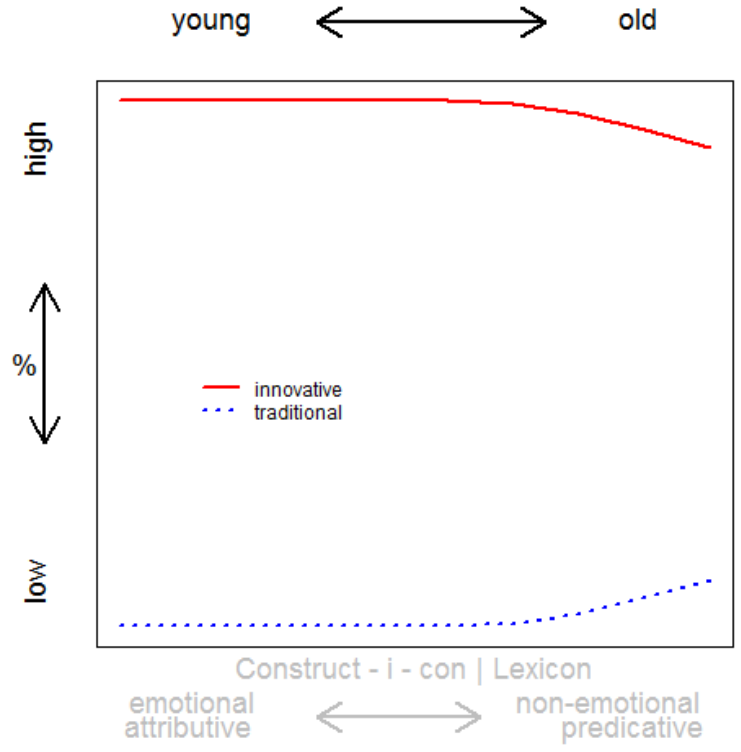
Diffusion of Innovations



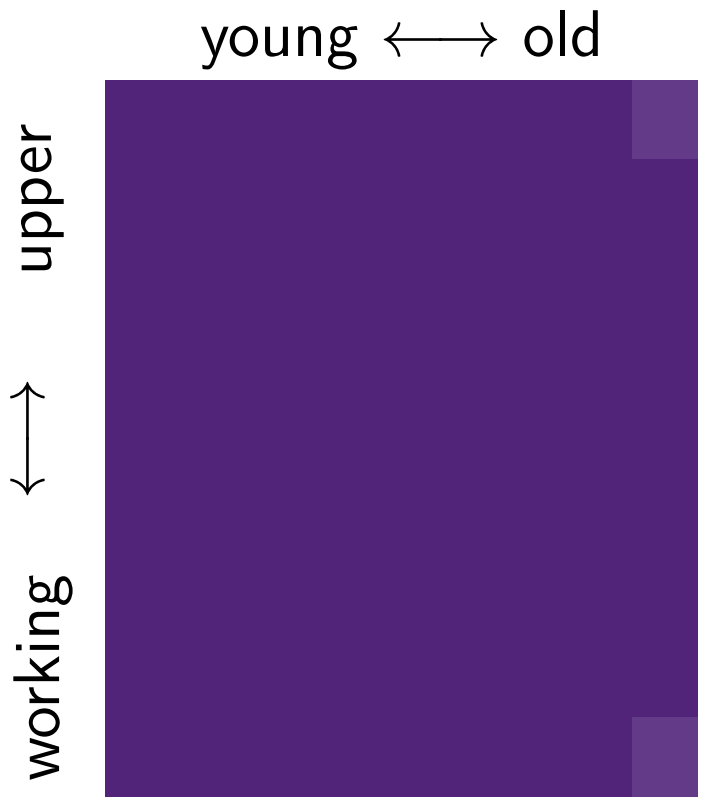


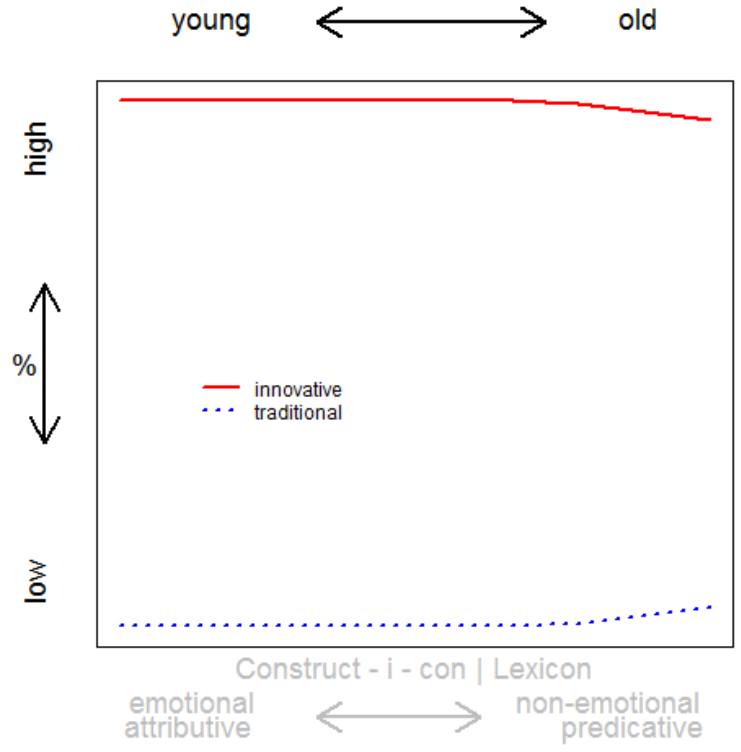
Diffusion of Innovations



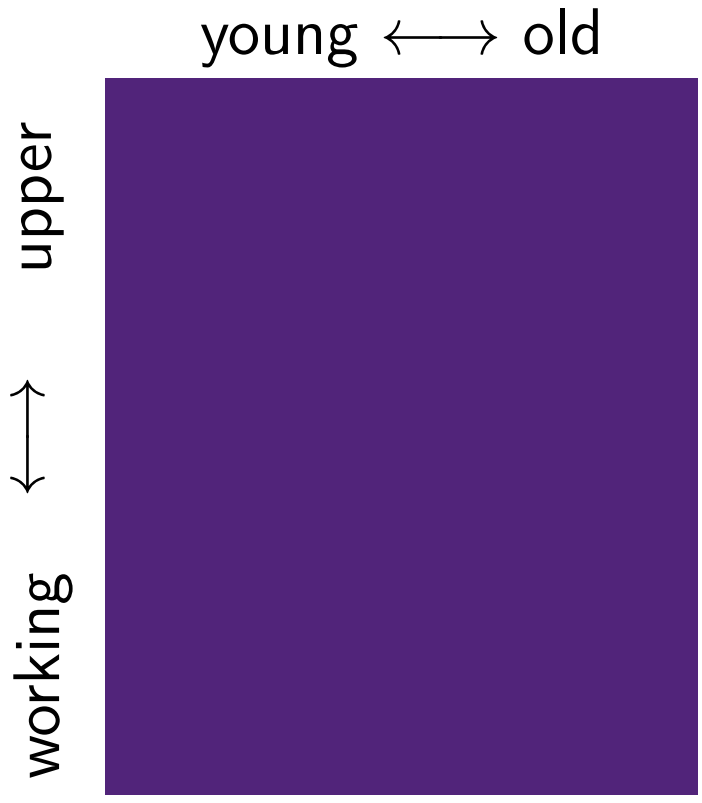


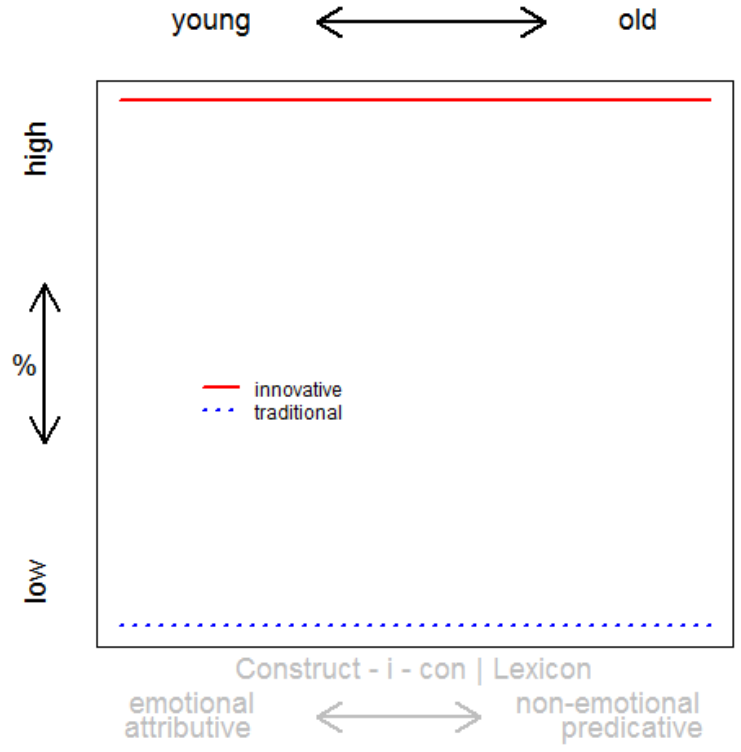
Diffusion of Innovations





Diffusion of Innovations





Mixed-Effects Binomial Logistic Regression

(??)

Figure 10: Difference between models without grouping/nesting and mixed-effects models (with grouping/nesting).