Stop the morphological cycle, I want to get off: Modeling the development of fusion

Micha Elsner, Martha Booker Johnson,
Stephanie Antetomaso and Andrea D. Sims

THE OHIO STATE UNIVERSITY

An old idea: change in morphological typology

Over time, morphologically agglutinative structures can become fusional…

But this doesn't always happen!

Which language features accelerate these changes?

Which ones slow or stop them?

Agglutination vs. fusion

Indo-European (reconstructed)

	PRS	AOR
1SG	-m-i	-m
2SG	-s-i	-s
3SG	-t-i	-t

Ancient Greek ("give")

	PRS	AOR
1SG	dídō-mi	edō-n
2SG	dídō-s	edō-s
3SG	dídō-si	edō

Agglutination vs. Fusion

Indo-European (reconstructed)

	PRS	AOR
1SG	-m-i	-m
2SG	-s-i	-s
3SG	-t-i	-t

Ancient Greek ("give")

	PRS	AOR
1SG	dídō-mi	edō-n
2SG	dídō-s	edō-s
3SG	dídō-si	edō

But what *is* fusion, really?

Plank (1999) lists many, many properties

Typical of agglutination	**Typical of fusion**
identifiable exponents	fused exponents
no inflection classes	inflection classes
no syncretism	plentiful syncretism
zero and multiple exponence	little zero/multiple exponence
large paradigms	small paradigms
weak phonological cohesion	strong phonological cohesion
many optional elements	few optional elements

But what *is* fusion, really?

Plank (1999) lists many, many properties

Typical of agglutination	**Typical of fusion**
identifiable exponents	fused exponents
no inflection classes	inflection classes
no syncretism	plentiful syncretism
zero and multiple exponence	little zero/multiple exponence
large paradigms	small paradigms
weak phonological cohesion	strong phonological cohesion
many optional elements	few optional elements

A huge list isn't very satisfying

These properties don't always go together…

Or even *usually* go together (Haspelmath 2009, for instance)

It's not clear which of them are **causes** and which are **effects**

So it's hard to frame hypotheses about why they may, or may not, coincide for a particular inflection in a particular language

History is the missing piece

By understanding how these systems arise over time, we can see whether some properties **contribute to** or **prevent** the development of others

(see Nichols 1992, Harris 2008, Murawaki 2018 and others)

One process of change

Indo-European (reconstructed)

	PRS	AOR
1SG	-m-i	-m
2SG	-s-i	-s
3SG	-t-i	-t

One process of change

Indo-European (reconstructed)

	PRS	AOR
1SG	-m-i	-m
2SG	-s-i	-s
3SG	-t-i	-t

change

Proto-Greek (reconstructed)

	PRS	AOR
1SG	-m-i	**-n**
2SG	-s	-s
3SG	-t-i	-t

One process of change

change →

Proto-Greek (reconstructed)

	PRS	AOR
1SG	-m-i	-n
2SG	-s-∅	-s
3SG	-t-i	-t

reanalysis →

Proto-Greek (reconstructed)

	PRS	AOR
1SG	-mi	-n
2SG	-s	-s
3SG	-ti	-t

One process of change

change →

Proto-Greek (reconstructed)

	PRS	AOR
1SG	-m-i	-n
2SG	-s-∅	-s
3SG	-t-i	-t

reanalysis →

Proto-Greek (reconstructed)

	PRS	AOR
1SG	-mi	-n
2SG	-s	-s
3SG	-ti	-t

...followed by dialectal generalization of *-mi*, *-n* to new verbs

The “morphological cycle”

Not the only way morphological fusion can arise…

But *one* important way:

Proposed by Schleicher (1850); newer survey in Igartua (2015)

Morphological change is driven by phonology and linear adjacency

Thus, the **phonological and linear factors are causal!**

Language learners face a choice

Learning outcome 1

	PRS	AOR
1SG	-m-i	**-n**
2SG	**-s-**∅	-s
3SG	-t**-i**	-t

plus **phonological rule**: m# → n
and more...

Learning outcome 2

	PRS	AOR
1SG	-mi	-n
2SG	-s	-s
3SG	-ti	-t

What determines the outcome?

Environments in which the morphemes appear

Frequency of the different combinations

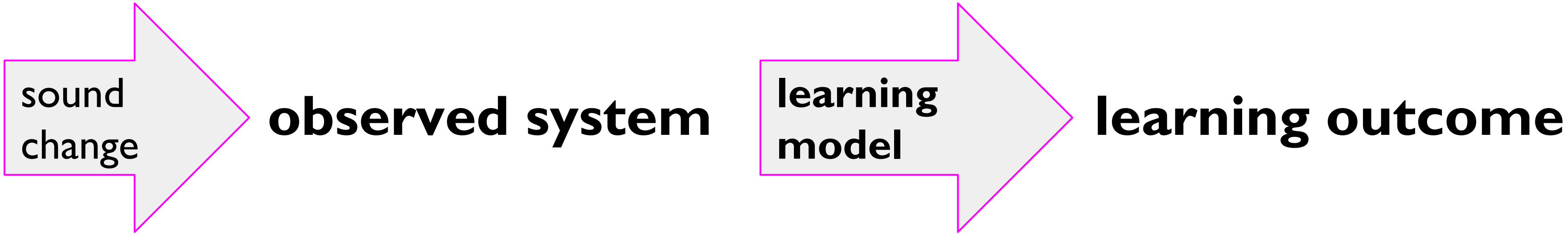
Evidence for the rule elsewhere in the language

Bybee (2002): “Items that are used together, fuse together”

This study

A proof-of-concept simulation using artificial data

Bayesian model of learner predicts when reanalysis might occur



What features of change process / input system predict the outcome?

Two specific claims

Case study 1

Variable realization (morphological slots which may contain ∅) preserves **agglutinative** structures: Plank (1999), Comrie (1989)

Case study 2

Stress-based vowel reduction encourages **fusional** structures: Zingler (2018)

Model framework

High priority for **interpretability**: we want to examine the output and see if the learned system is fusional or not

Makes neural models (Kann et al. 2016) less appealing; use older-fashioned transducer cascade (Cotterell et al. 2015; Dreyer and Eisner 2008) instead

Everything implemented using Carmel toolkit (Graehl et al. 1997, Chiang et al. 2010)

The model

latent variables

Input	lex="give", tense=PRS, prs/num=1SG
	Transducer 1: fusion
Abstract morphs	lex="give", tense=PRS | prs/num=1SG
	Transducer 2: lexicon
Underlying	dídō - mi
	Transducer 3: phonology
Surface	dídōmi

Model biases

Balance between two opposing pressures:

Learn a small inventory of morphemes (pressure for more agglutinative analyses)

Do not learn unnecessary phonological rules (pressure for more fusional analyses)

Case study 1: variable realization

Data loosely based on Kihehe (Bantu, Tanzania; Johnson 2015, Odden and Odden 1999)

twikomala

tu-	i-	komala
SM	PROG	sit

"we are sitting"

Kihehe verbs are marked for subject agreement (SM) and tense/aspect; phonological rules prevent VV on the surface

Ambiguous: may or may not be fused in speakers' minds

Why would variable elements matter?

An agglutinative analysis has (num SMs + num TMA) morphemes;

A fused analysis has (num SMs × num TMA) morphemes

twikomala

tu-	i-	komala
SM	PROG	sit

"we are sitting"

What if the template had more slots?

tu-	??-	i-	komala
SM	??	PROG	sit

More combinations to memorize; fused analysis less appealing

Polysynthesis: many variably-filled slots

Choguita Rarámuri: Uto-Aztecan, Mexico (Caballero 2008)

Table 15: Choguita Rarámuri verbal suffixes

	Category	Suffix	Reference
S1	Inchoative	Inchoative *-ba* (Inch)	§1
S2	Transitives	Pluractional transitive *-ča* (Tr:pl)	§2.1
		Transitive *-na* (Tr)	§2.2
		Transitive *-bu* (Tr)	§2.3
S3	Applicatives	Applicative *-ni* (Appl)	§3.1
		Applicative *-si* (Appl)	§3.2
		Applicative *-wi* (Appl)	§3.3
S4	Causative	Causative *-ti* (Caus)	§4
S5	Applicative	Applicative *-ki* (Appl)	§5
S6	Desiderative	Desideartive *–nare* (Desid)	§6
S7	A. Motion	Associated Motion *–simi* (Mot)	§7
S8	A. Evidential	Auditory Evidential *–čane* (Ev)	§8
S9	Tense, Aspect, Mood, Voice	Past Passive *-ru* (Pst:Pass)	§9.1
		Future Passive *-pa* (Fut:Pass)	§9.2
		Medio-Passive *–riwa* (MPass)	§9.3
		Conditional Passive *–suwa* (Cond:Pass)	§9.4
		Future Singular *-méa, -ma* (Fut:sg)	§9.5
		Future Plural *-po* (Fut:pl)	§9.6
		Motion Imperative *-me* (Mot:Imp)	§9.7
		Conditional *-sa* (Cond)	§9.8
		Irrealis singular *–me* (Irr:sg)	§9.9
		Irrealis plural *-pi* (Irr:pl)	§9.10
S10	Mood	Potential *-ra* (Pot)	§10.1
		Imperative sg. *–ka* (Imp:sg)	§10.2
		Imperative sg. *-sa* (Imp:sg)	§10.3
		Imperative pl. *-si* (Imp:pl)	§10.4
S11	Tense, Aspect, Mood	Reportative *-ra* (Rep)	§11.1
		Past perfective *-li* (Pst)	§11.2
		Past perfective , 1st person *-ki* (Pst:1)	§11.3
		Past imperfective *–e* (Impf)	§11.4
		Progressive *-a* (Prog)	§11.5
		Indirect causative *-nura*	§11.6
S12	Subord.	Temporal *-či* (Temp)	§12.1
		Epistemic *-o* (Ep)	§12.2
		Gerund *-ka* (Sim)	§12.3
		Purposive *-ra* (Pur)	§12.4
		Participial *-ame* (Ptcp)	§12.5

Simulations with artificial data

Each language has:

200 random CV stems

1000 word forms

Model validity check

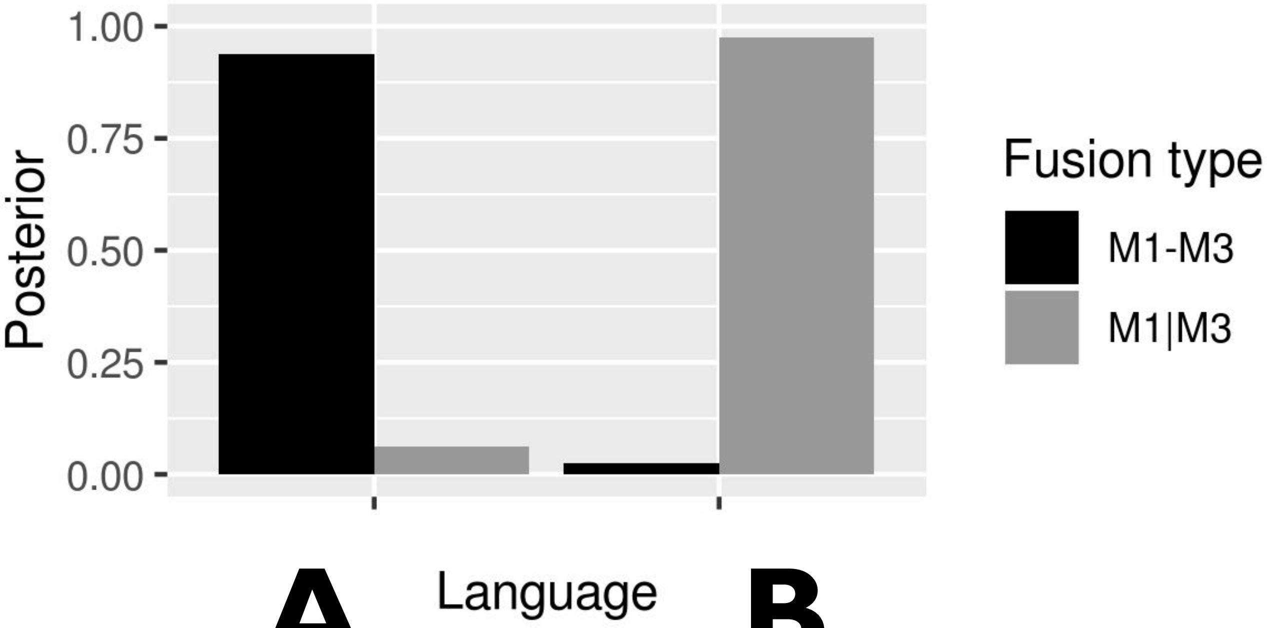
A
agglutinative

M_1	M_2	M_3	lex	under		surface
ta		i	mela	ko-i-mela	⇨	koimela
ko		a	tano	mu-i-mela		muimela
he		de	...			
mu		no				
gu						
si						

B
fusional

M_1 | M_3	lex	under		surface
ya, se, dunu, lanu	mela	ya-mela	⇨	yamela
ha, hi, si, yu…	tano	dunu-mela		dunumela
	...			

As expected...



A

agglutinative

B

fusional

Add some phonology

C

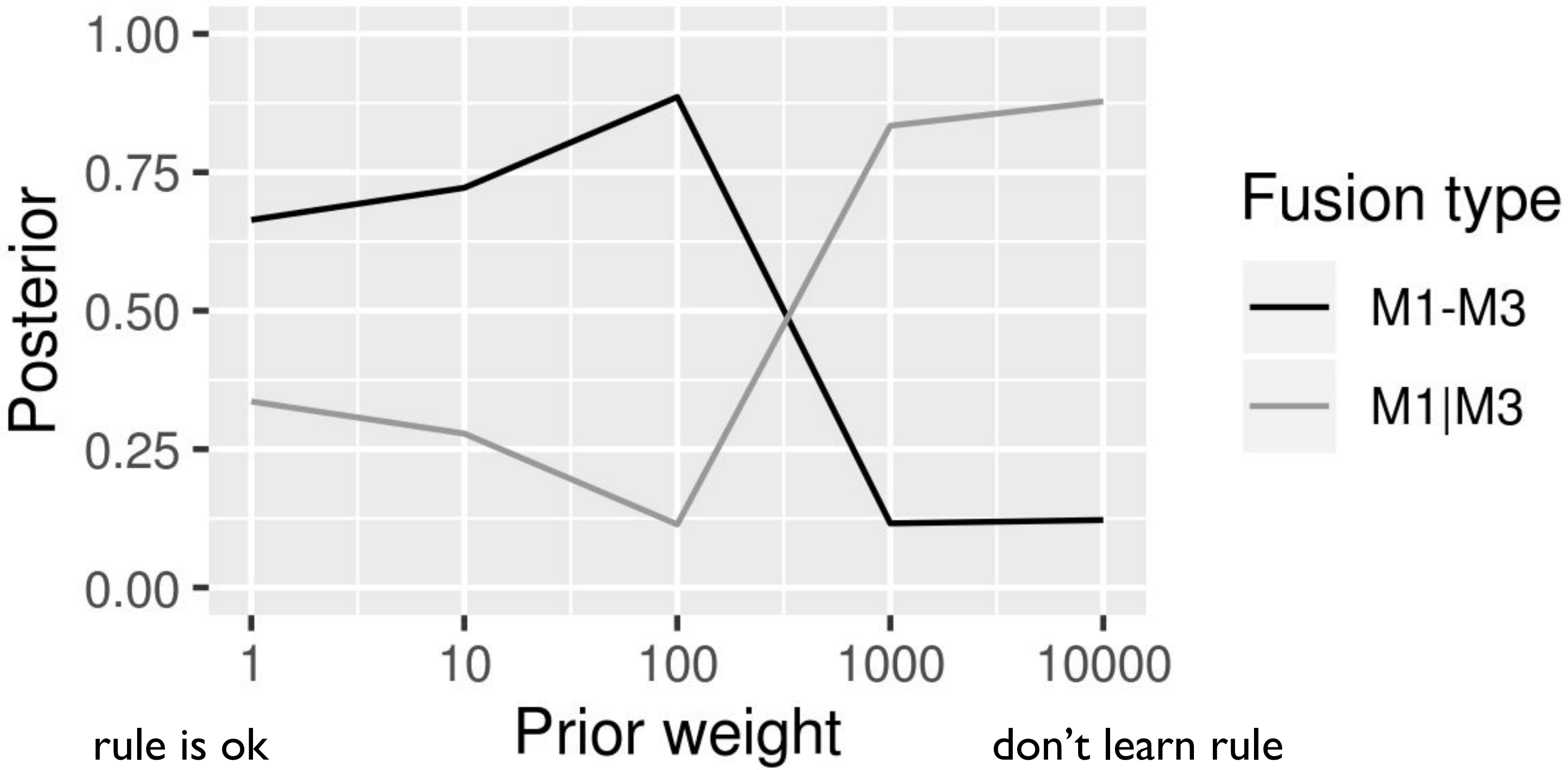
ambiguous

M_1	M_2	M_3	lex	under		surface
ta		i	mela	ko-i-mela	⇨	**ki**mela
ko		a	tano	mu-i-mela		**mwi**mela
he		de	...			
mu		no				
gu						
si						

high V + V → glide + V
low V + V → _ + V

Prior weight sets bias against phonological rule

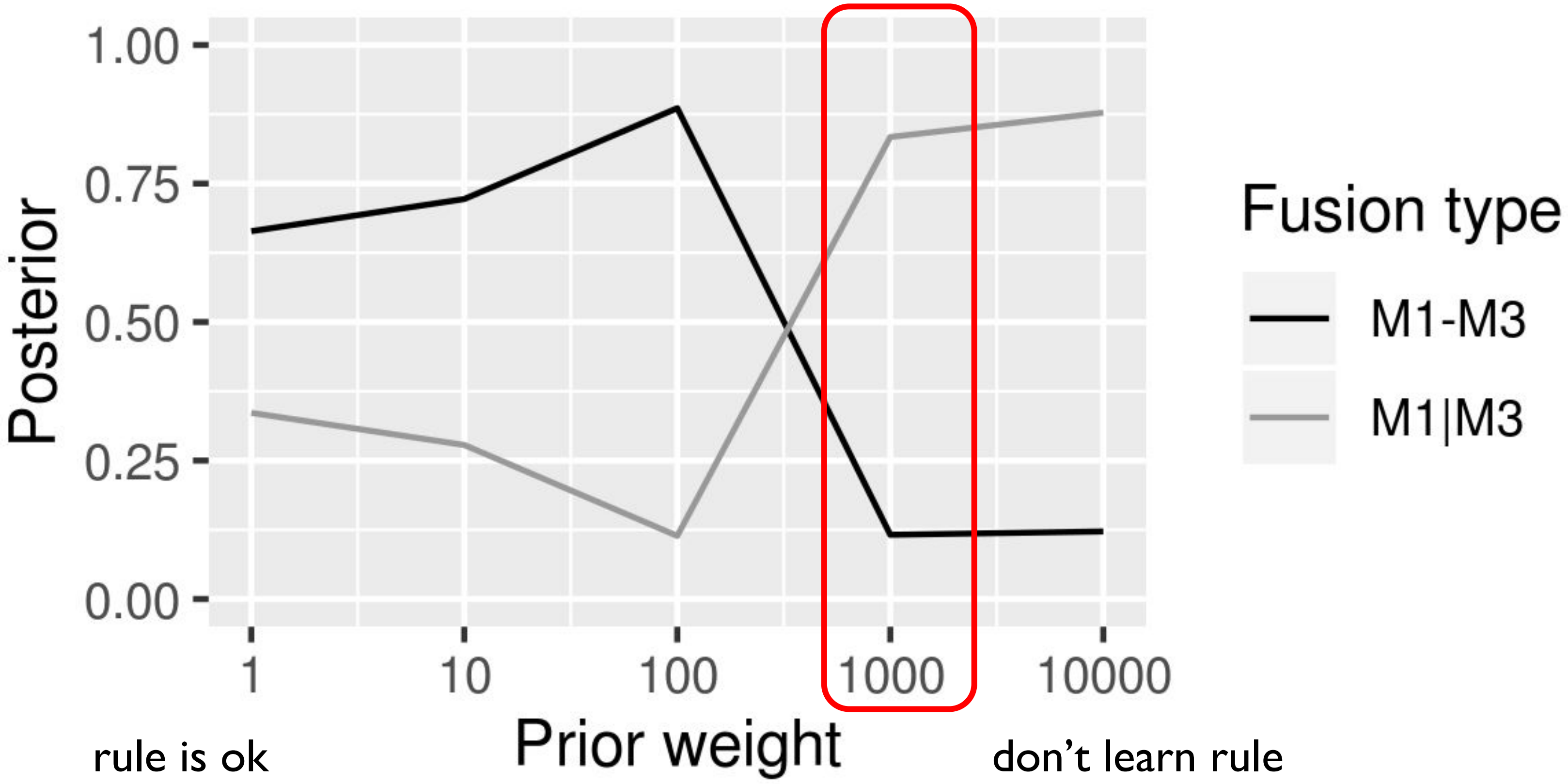
Bias determines outcome



rule is ok

don't learn rule

Fix prior, then vary morphological template



rule is ok

don't learn rule

Add a variable element

D

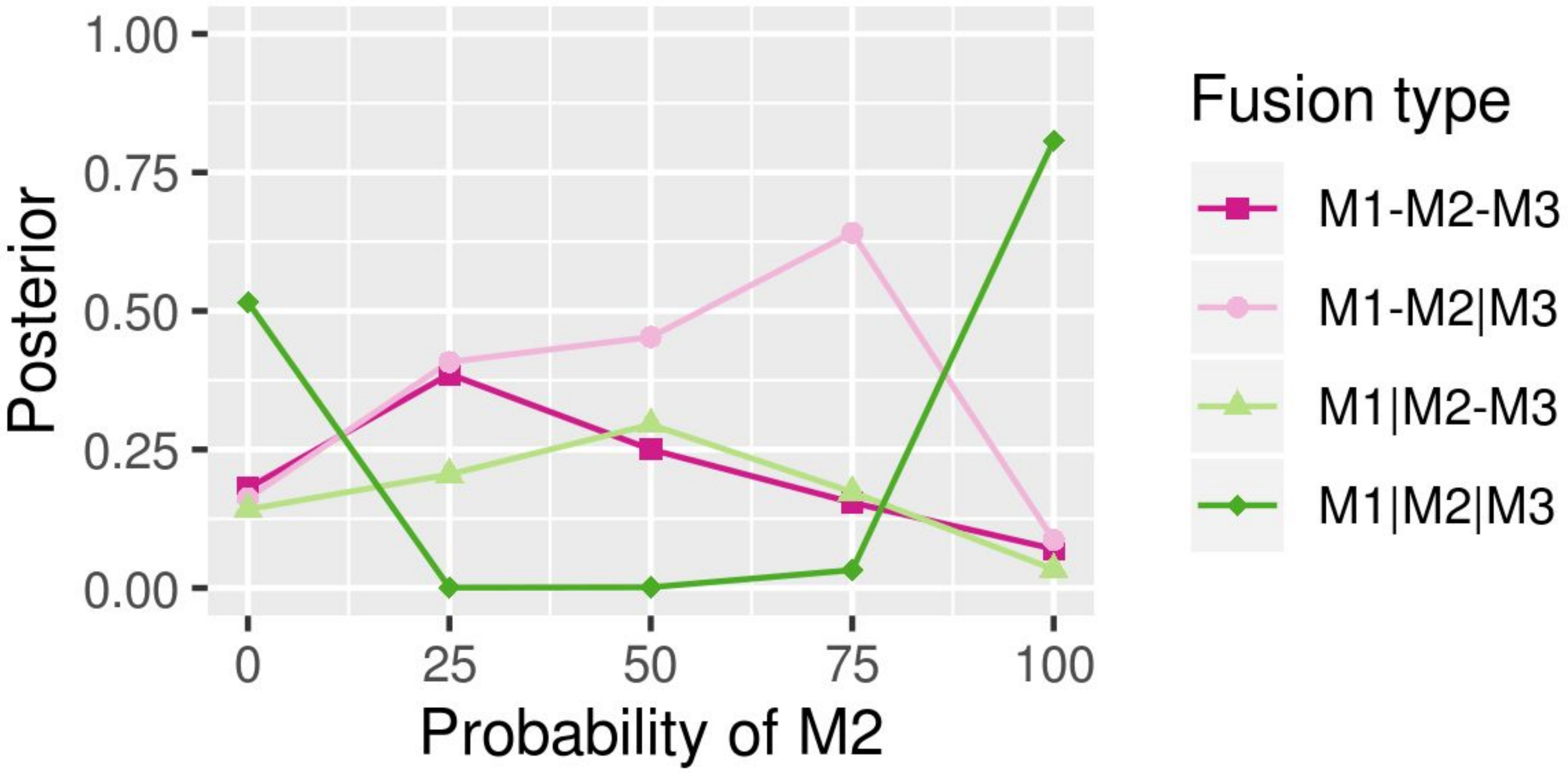
ambiguous

M_1	M_2	M_3	lex	under		surface
ta	**sa**	i	mela	mu-i-mela	⇨	**mw**imela
ko	ø	a	tano	mu-sa-i-mela		**musi**mela
he		de	...			
mu		no				
gu						
si						

M_2 filled with variable probability

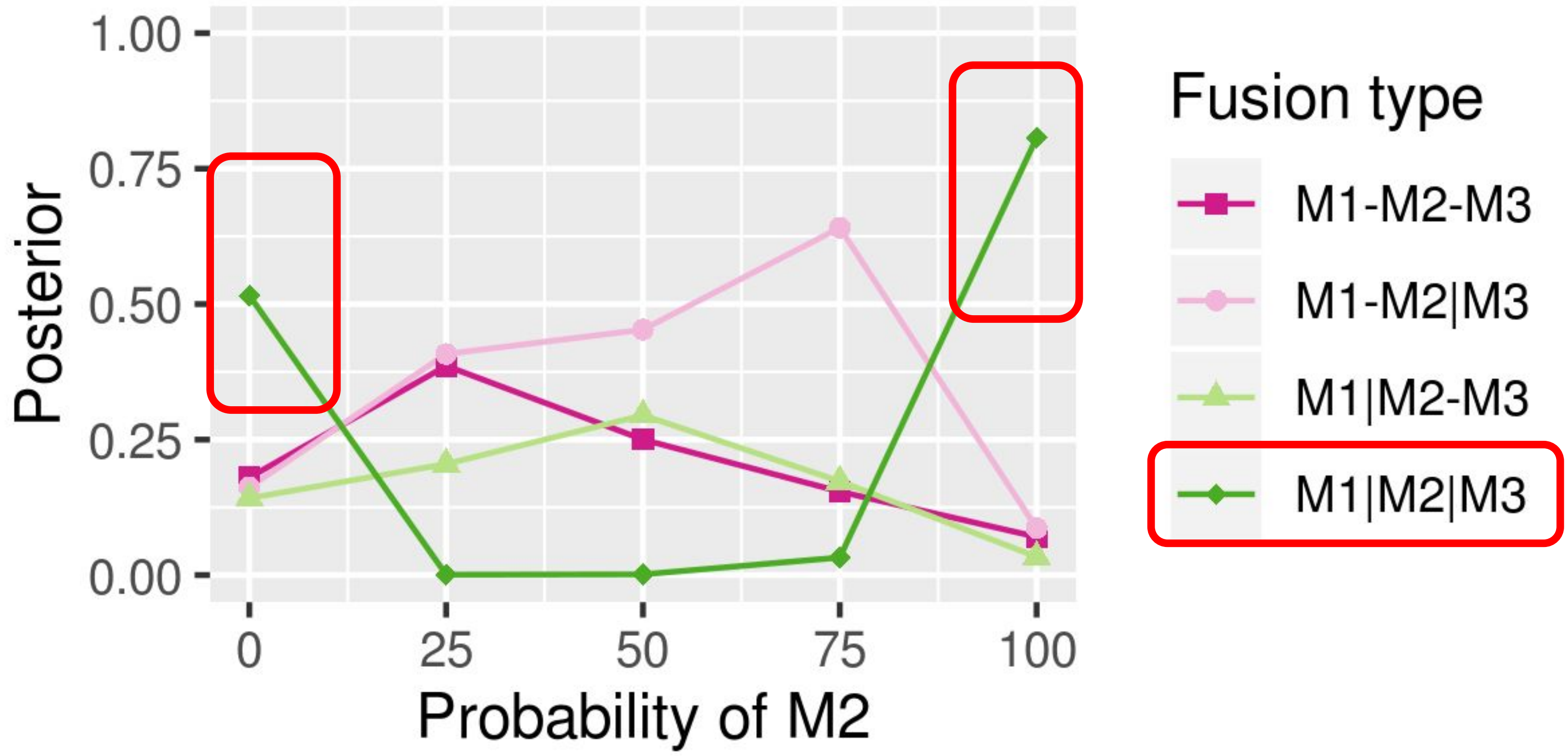
high V + V → glide + V
low V + V → _ + V

Results

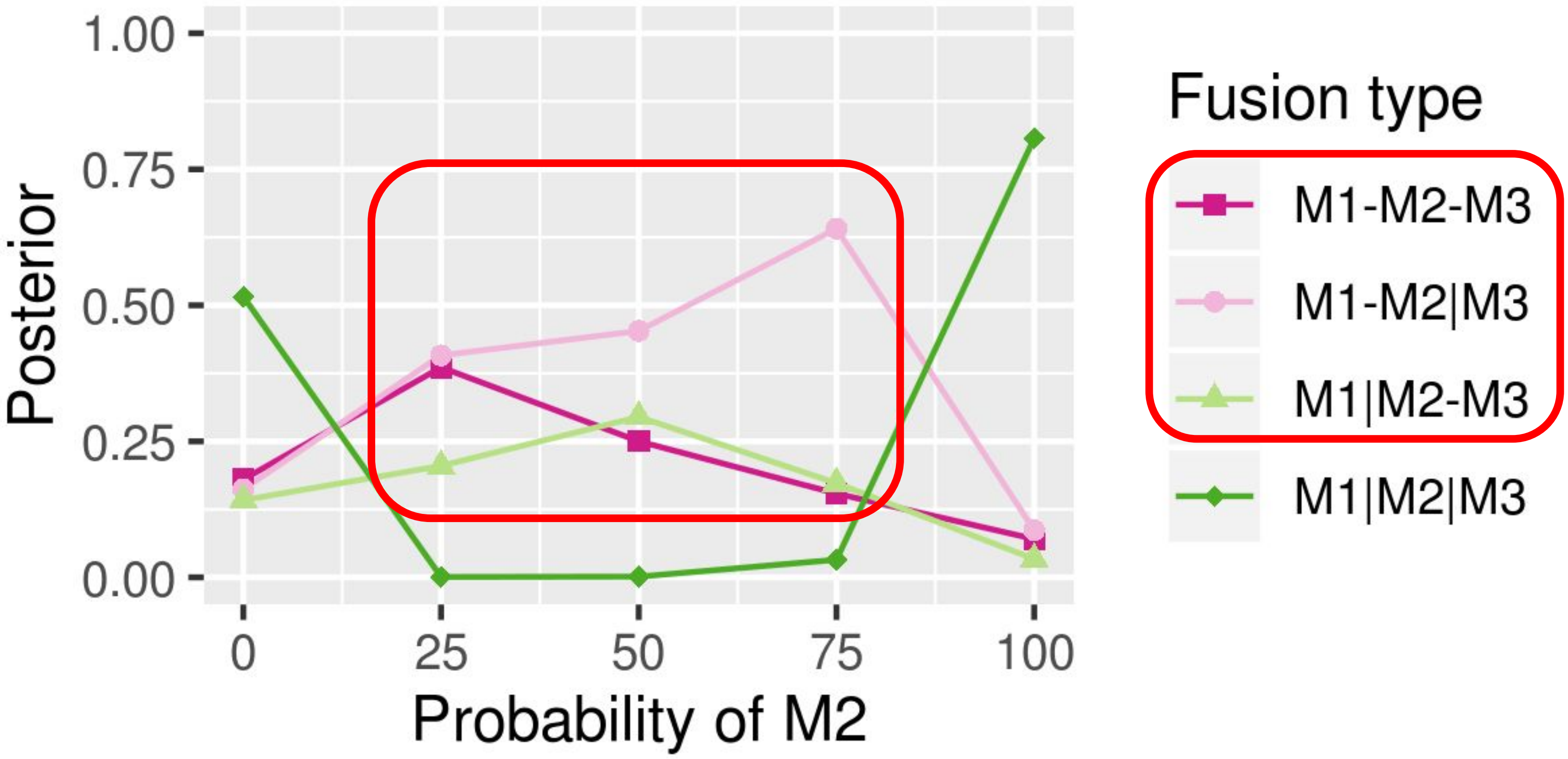


1.00
0.75
0.50
0.25
0.00
Posterior
0
25
50
75
100
Probability of M2
Fusion type
M1-M2-M3
M1-M2|M3
M1|M2-M3
M1|M2|M3

Categorical systems: full fusion



Variable M_2: partial or no fusion



Interim conclusion

Adding variable template slots preserves agglutination

Even when outcome without them would be fusional

When fusion does result, tends to be more local (as in Caballero and Kapatsinski 2019)

Effect critically depends on **variability**, not just extra slots

Case study 2: stress-based reduction

Zingler (2018) argument for maintenance of Turkish agglutination: **vowel harmony** prevents **stress-based reduction**, which would in turn lead to more fusional system

The link between harmony and reduction is unclear… but we'll focus on the second claim, that reduction leads to fusion

Why reduction?

Reducing vowels forces consonants into contact…

- Result: phonological interactions across morpheme boundaries
- Obscures the true underlying forms of morphemes
- Especially when reduction effect targets the same syllables each time

Language E has final stress

	lex	M_1	M_2		underlying
	dite	ta	pi		dite-ko-de
E	...	ko	ka	⇒	
		he	de		
final stress		mu	no		
		gu			
		si			

Language E has final stress

underlying

dite-ko-de

di**te**-ko-**de**

Assign **sS** stress from right

Because there are always two monosyllabic suffixes, this stress system ensures M_1 will be unstressed and M_2 will be stressed

Language E has final stress

underlying

surface

dite-ko-de → di**te**-ko-**de** → dtekde → ddekte

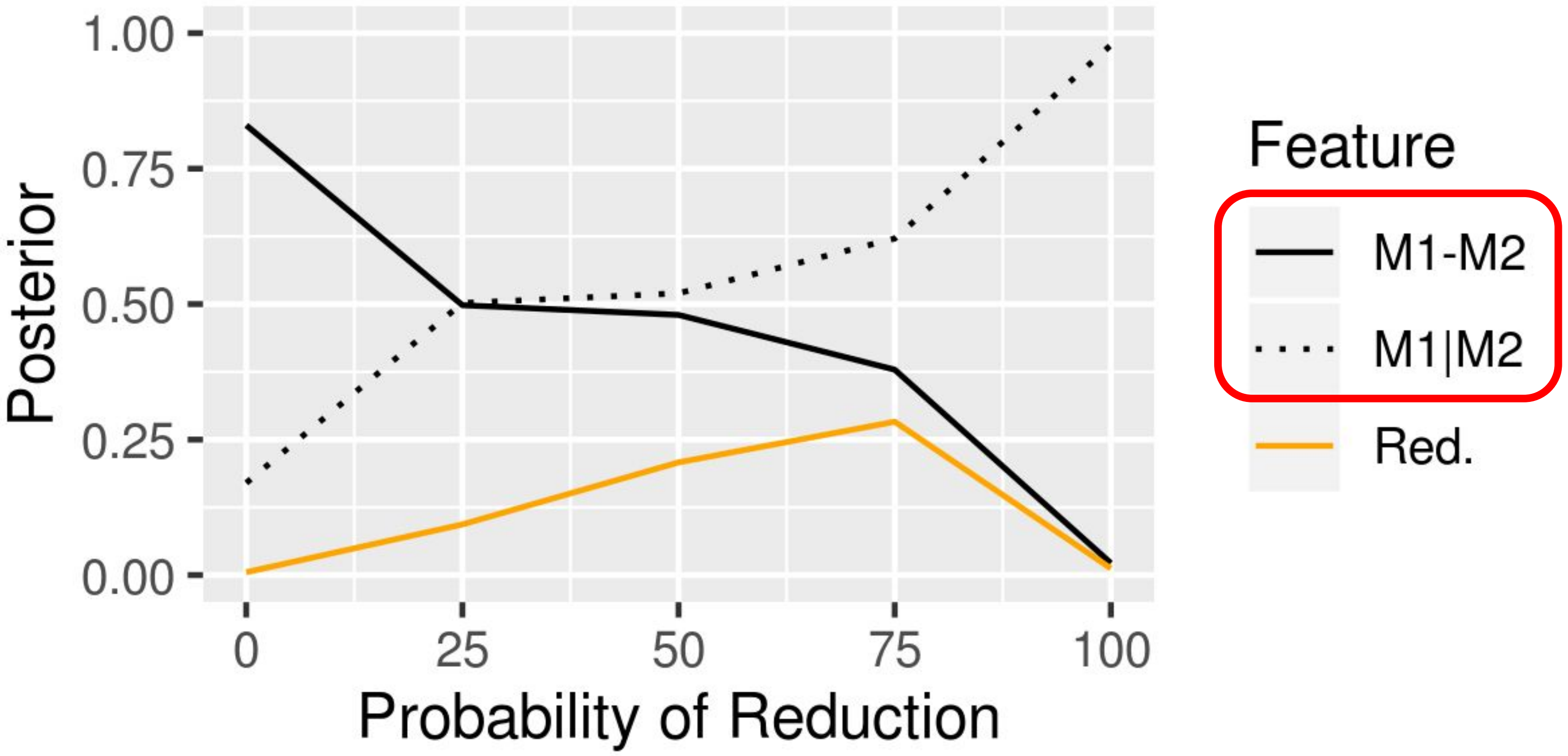
Assign **sS** stress from right

Probabilistically reduce weak vowels

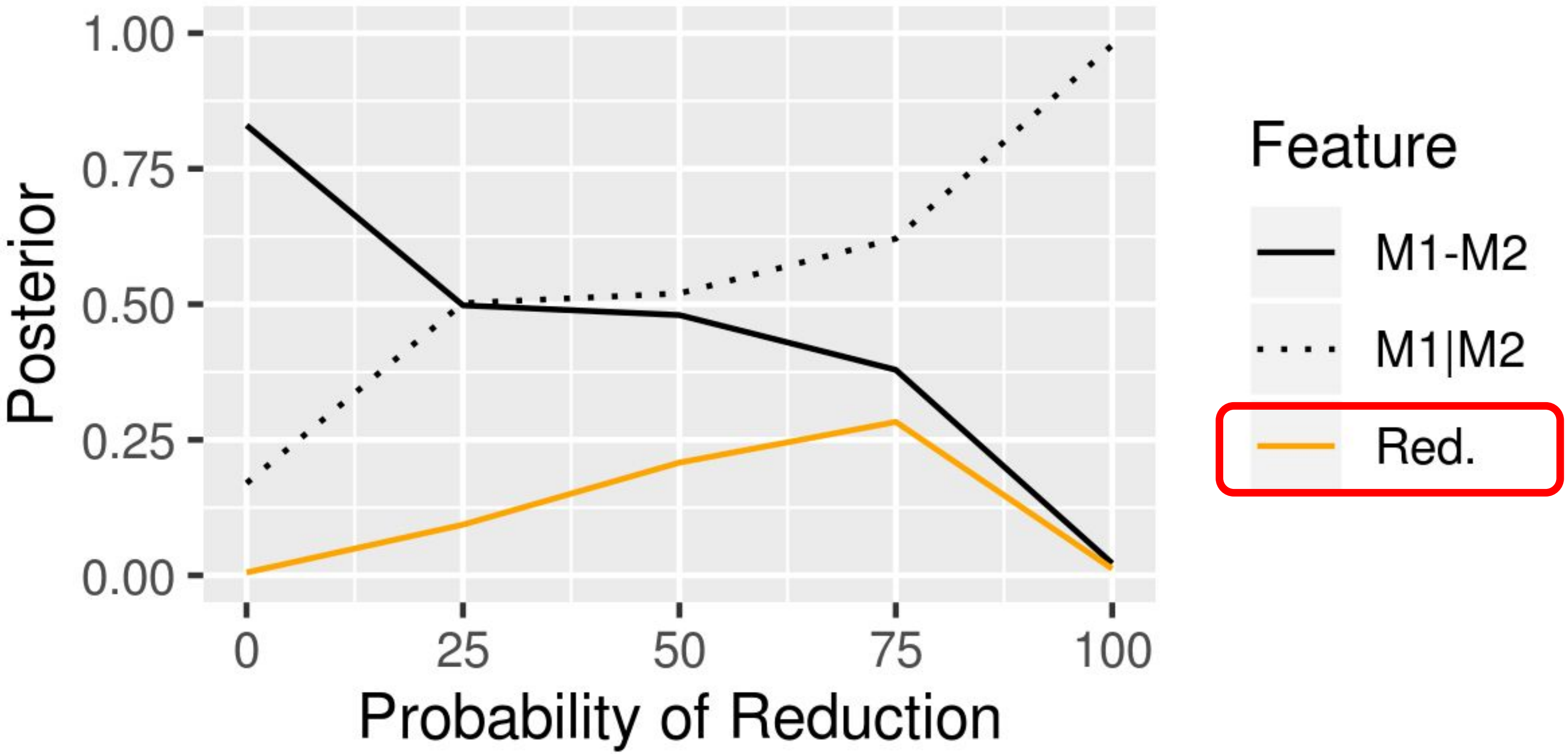
Voicing assimilation

Test variants with different probability of reduction

Reduction encourages fusion



Probability of reduction is underestimated



Reduction is “baked into” the lexicon

Learning outcomes

% reduction	M1=1 (*ta*)	M2=3 (*de*)	M1=1 | M2=3 (*ta-de*)
0	ta	de	-
25	ta (t, te)	de (te)	tade
50	ta (t)	te (de)	tade
75	ta (ti, t)	te (de)	tte
100	-	-	tte

What if stress placement is less predictable?

Occurs in real languages with lexical stress or some kinds of predictable stress systems…

Language with initial stress:

Even and odd-length stems place different stress on suffixes

Language F has initial stress

underlying → surface

dite-ko-de → **di**te-**ko**-de → ditkod → ditkod

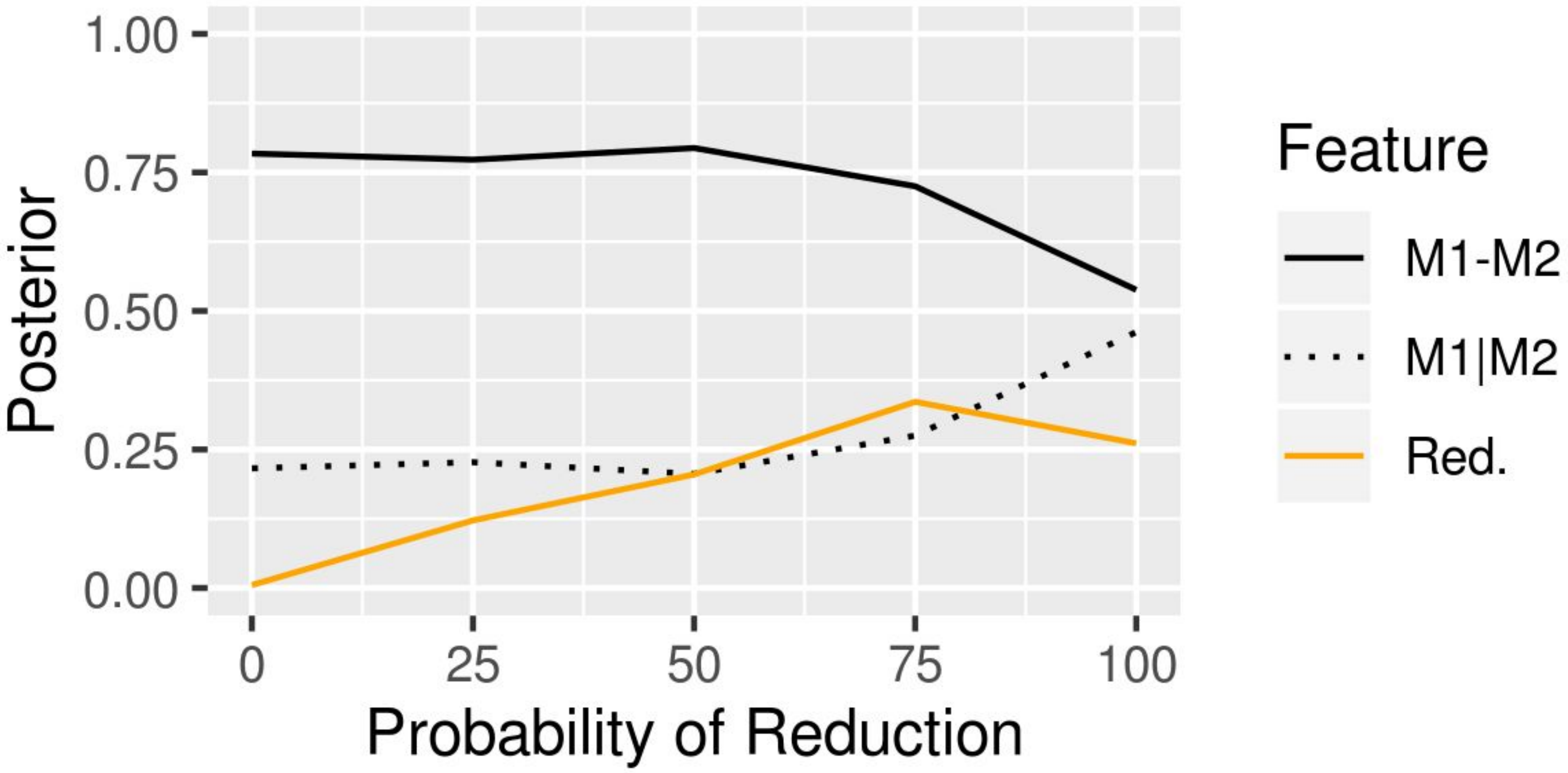
Assign **Ss** stress from left

Probabilistically reduce weak vowels

Voicing assimilation

Test variants with different probability of reduction

Unpredictable stress: less fusion



Plank's list, revisited

Typical of agglutination	Typical of fusion
identifiable exponents	fused exponents
no inflection classes	inflection classes
no syncretism	plentiful syncretism
zero and multiple exponence	little zero/multiple exponence
large paradigms	small paradigms
weak phonological cohesion	strong phonological cohesion
many optional elements	few optional elements

Conclusion

Typical of agglutination

identifiable exponents

no inflection classes

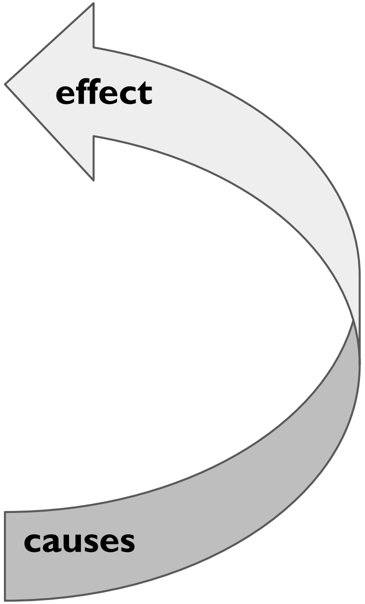
no syncretism

zero and multiple exponence

large paradigms

weak phonological cohesion

many optional elements



Conclusion

But we can also understand why (per Haspelmath and others) “agglutinative” and “fusional” features don’t always cluster…

Many ways for fusion to arise historically

Our model addresses only one mechanism

Future work

Test the model on data from real historical corpora!

Look at other language features (like the rest of Plank's list)

More generally: historical explanations for typological correlations (Harris 2008 and others), combined with models of the learner

Thank you!

And thanks to:

Fall 2018 Seminar on Models of Morphological Learning and Change at Ohio State University

Brian Joseph

Three anonymous reviewers