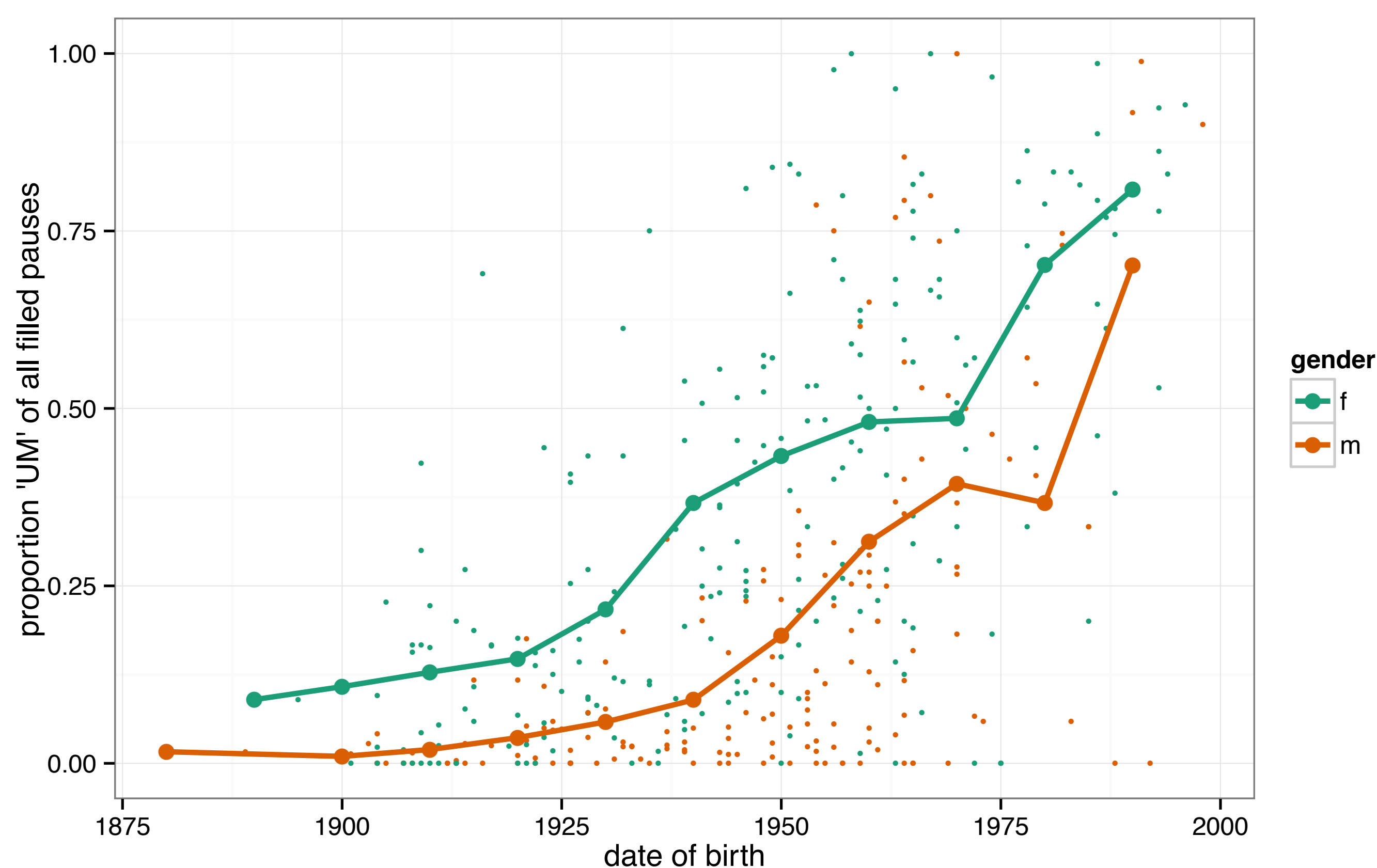




## Introduction

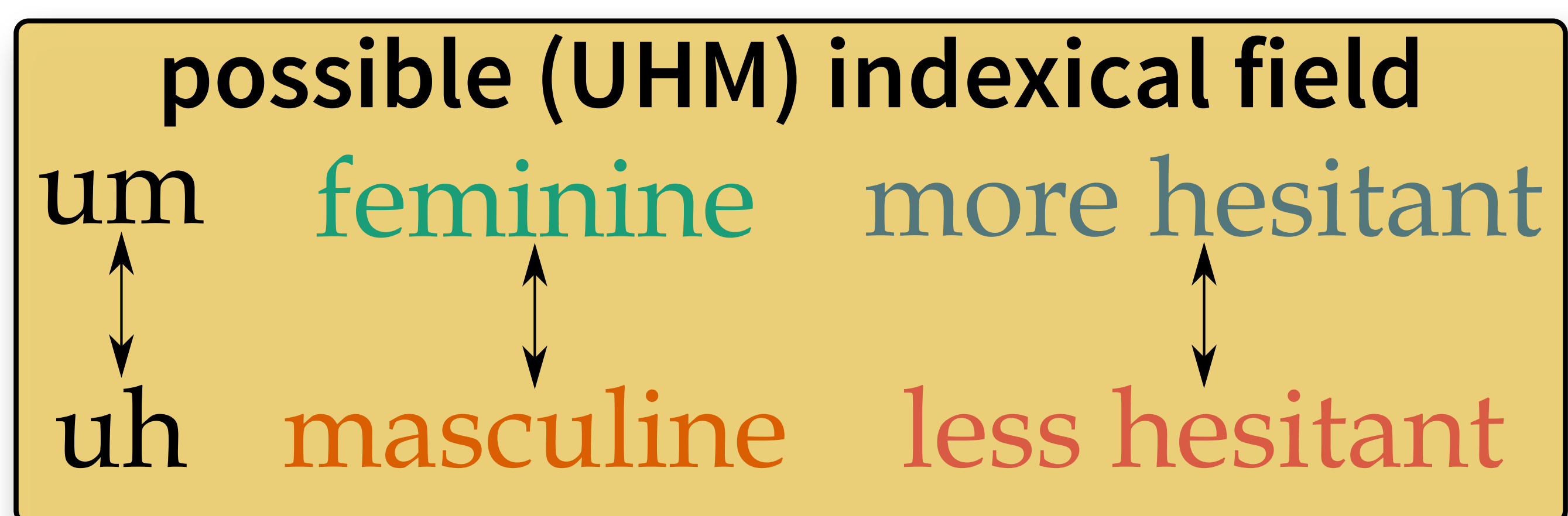
A cautionary tale in imputing a signaling strategy from usage data

## A Change in Progress

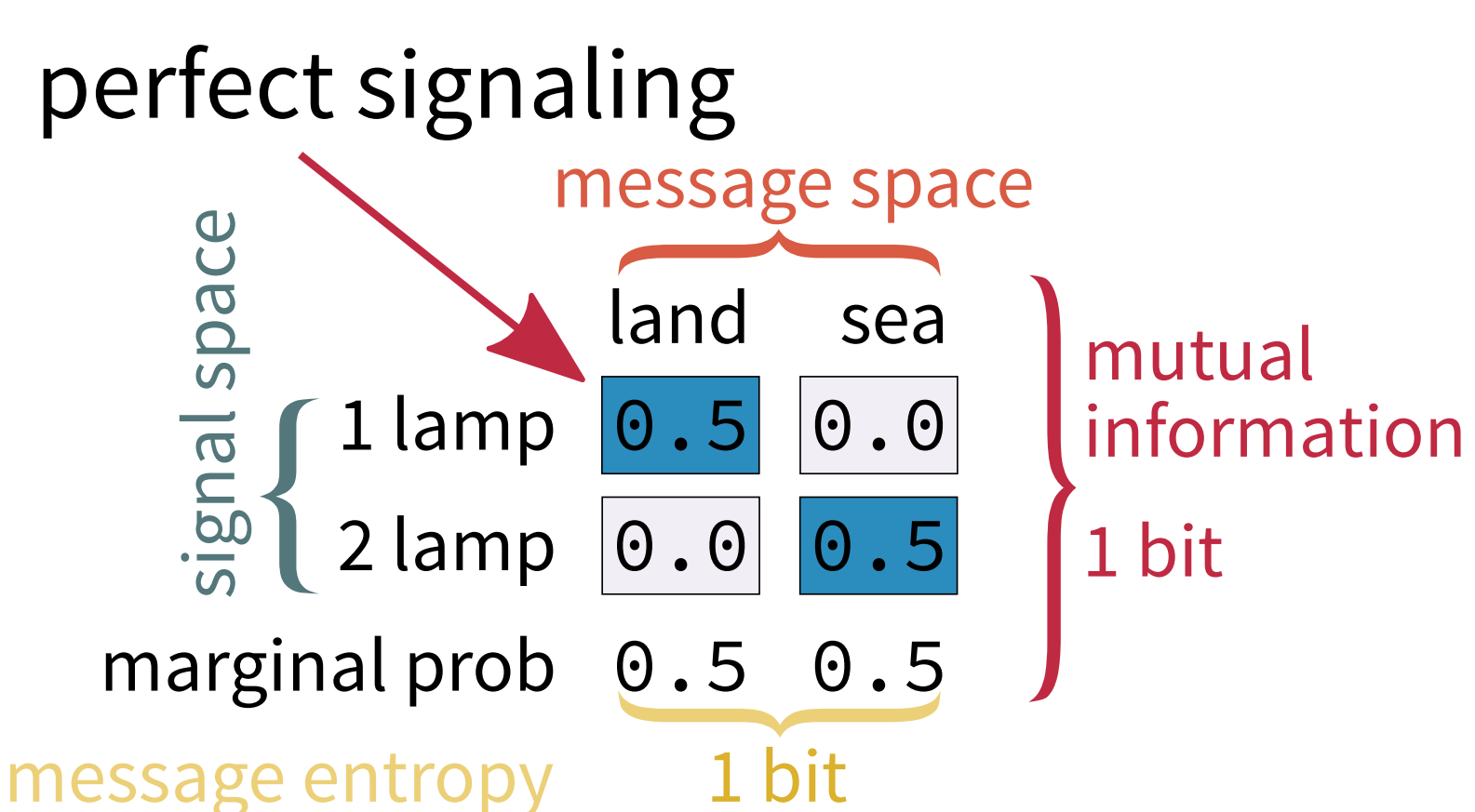


- female led
- greater 'UM' with greater processing difficulty (Clark & Fox Tree, 2002)

## (UHM) Indexicality?

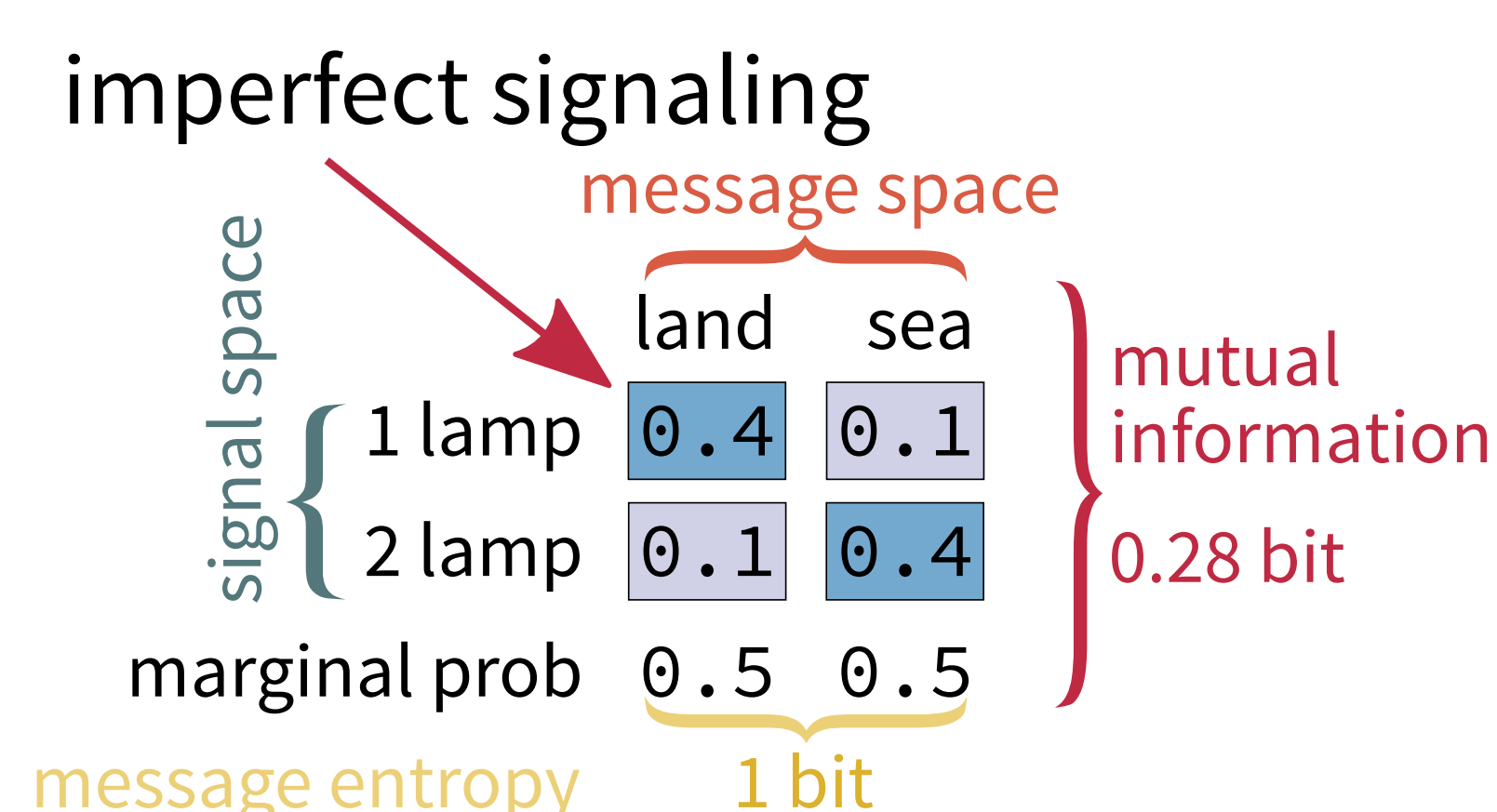


## Information Theory



With perfect signalling the MI = the amount of information to be communicated (the message entropy).

As the signal gets noisier, MI < the amount of information to be communicated.



## Application to Indexicality

- estimate MI between (UHM) and gender.
- compare to other gender signals
- properties of given names

## Data

**filled pauses**  
From the Philadelphia Neighborhood Corpus  
22,967 filled pauses  
395 speakers  
DOB 1889 - 1998

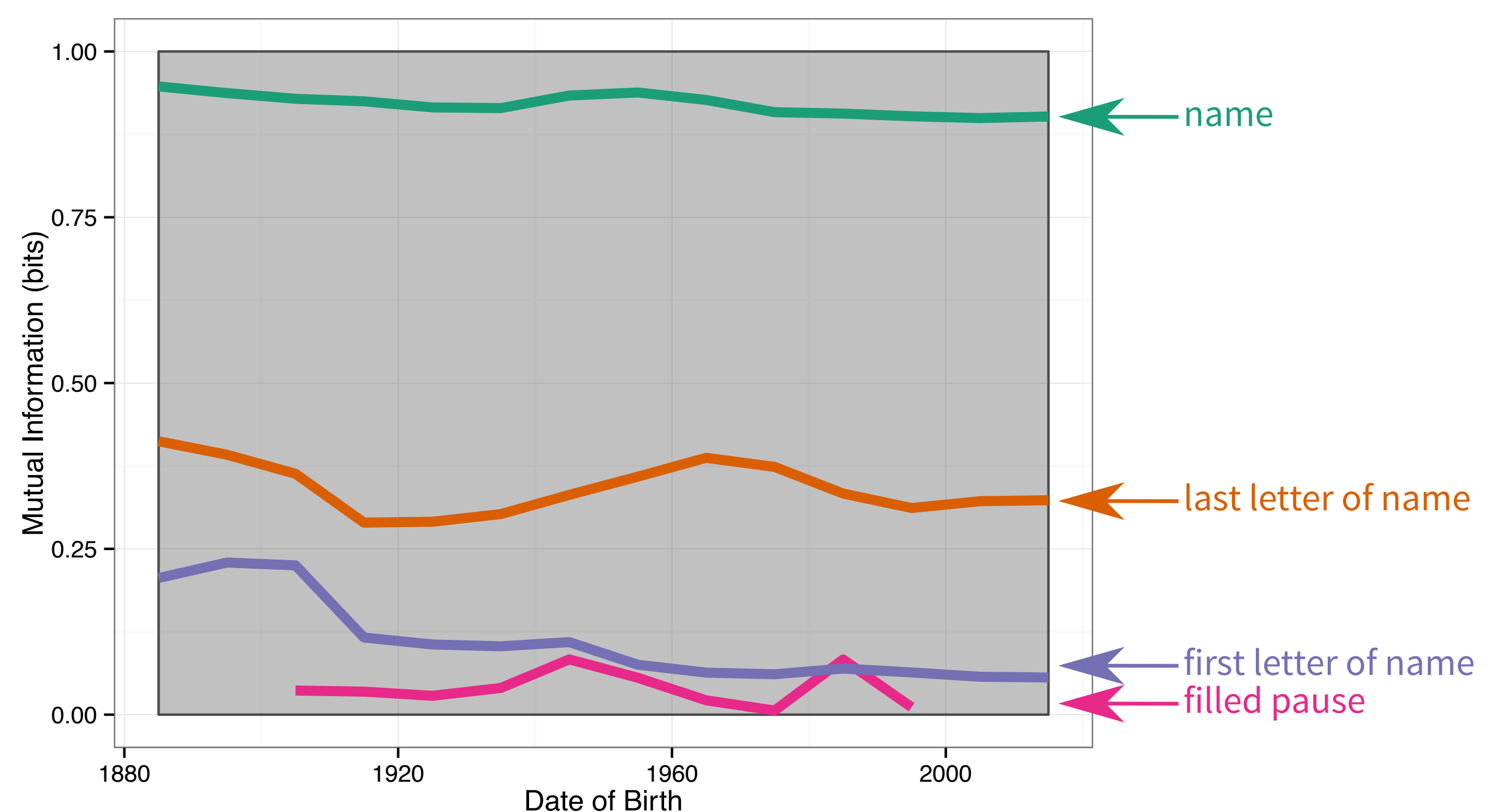
**given names**  
babynames R package  
92,600 unique baby names from the social security administration  
most common names from 1880-2013

## Sample Results

Speakers born in 1950s

signal space	message space	female	male	mutual information
		UM	0.22	
	UH	0.28	0.41	
marginal prob		0.50	0.50	
message entropy		1 bit		

## Results



## Conclusions

Utilizing Information Theory, we can compare linguistic variation we think speakers are using to signal to other signals.

Despite the relatively robust gender effect on 'UM' usage, it is a fairly weak signal for gender.

informative = (filled pause | gender)  
not informative = (gender | filled pause)