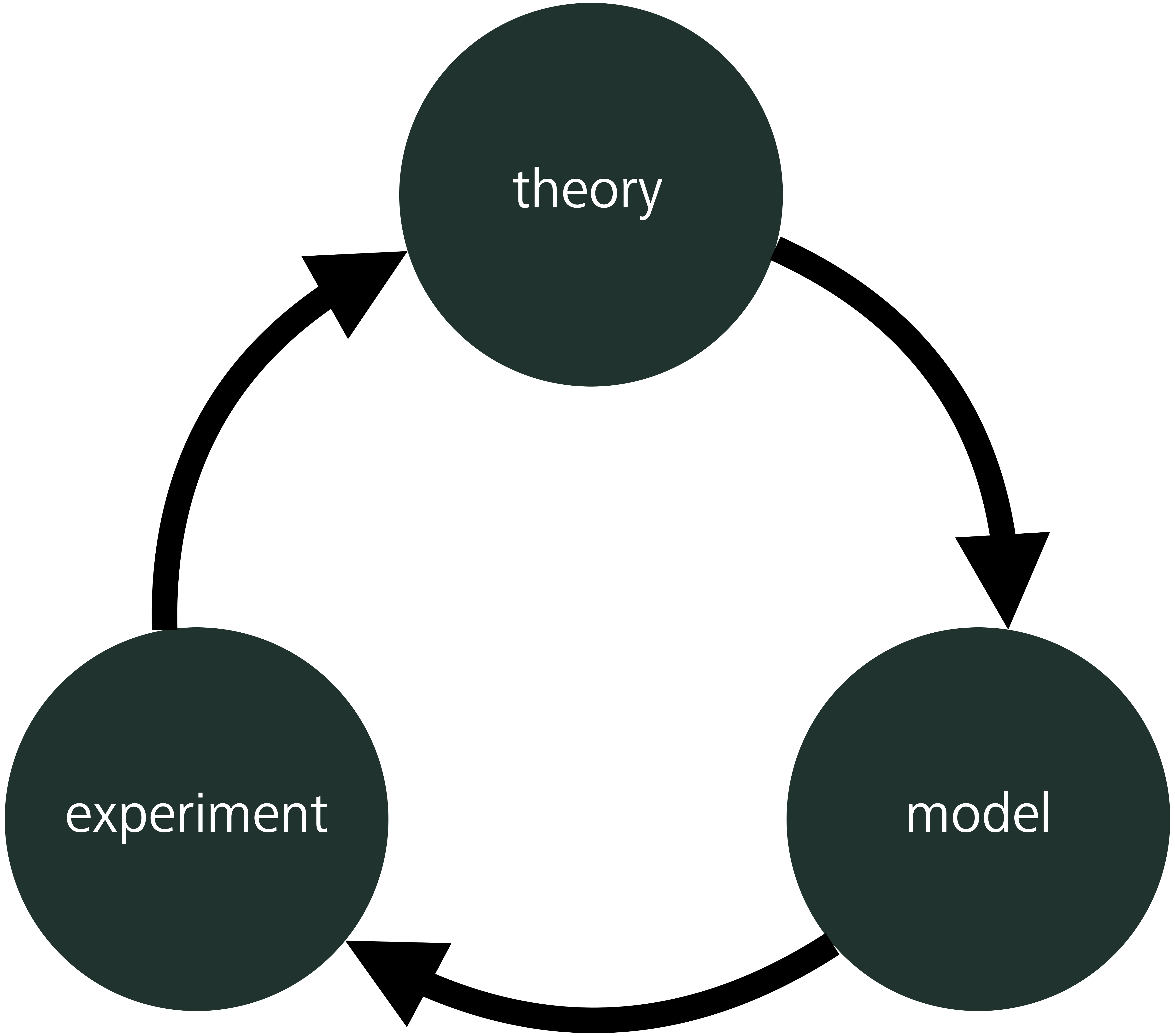
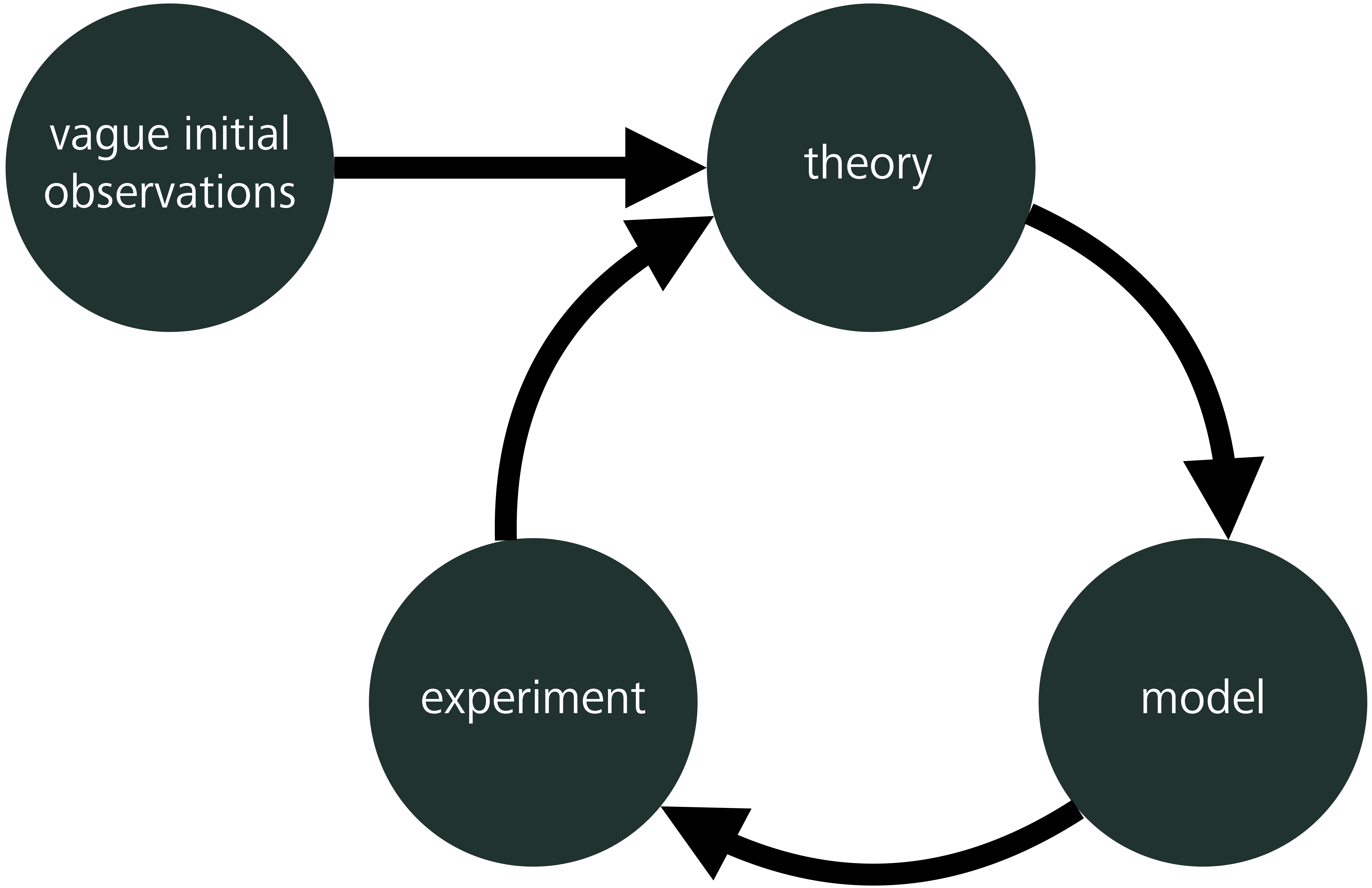
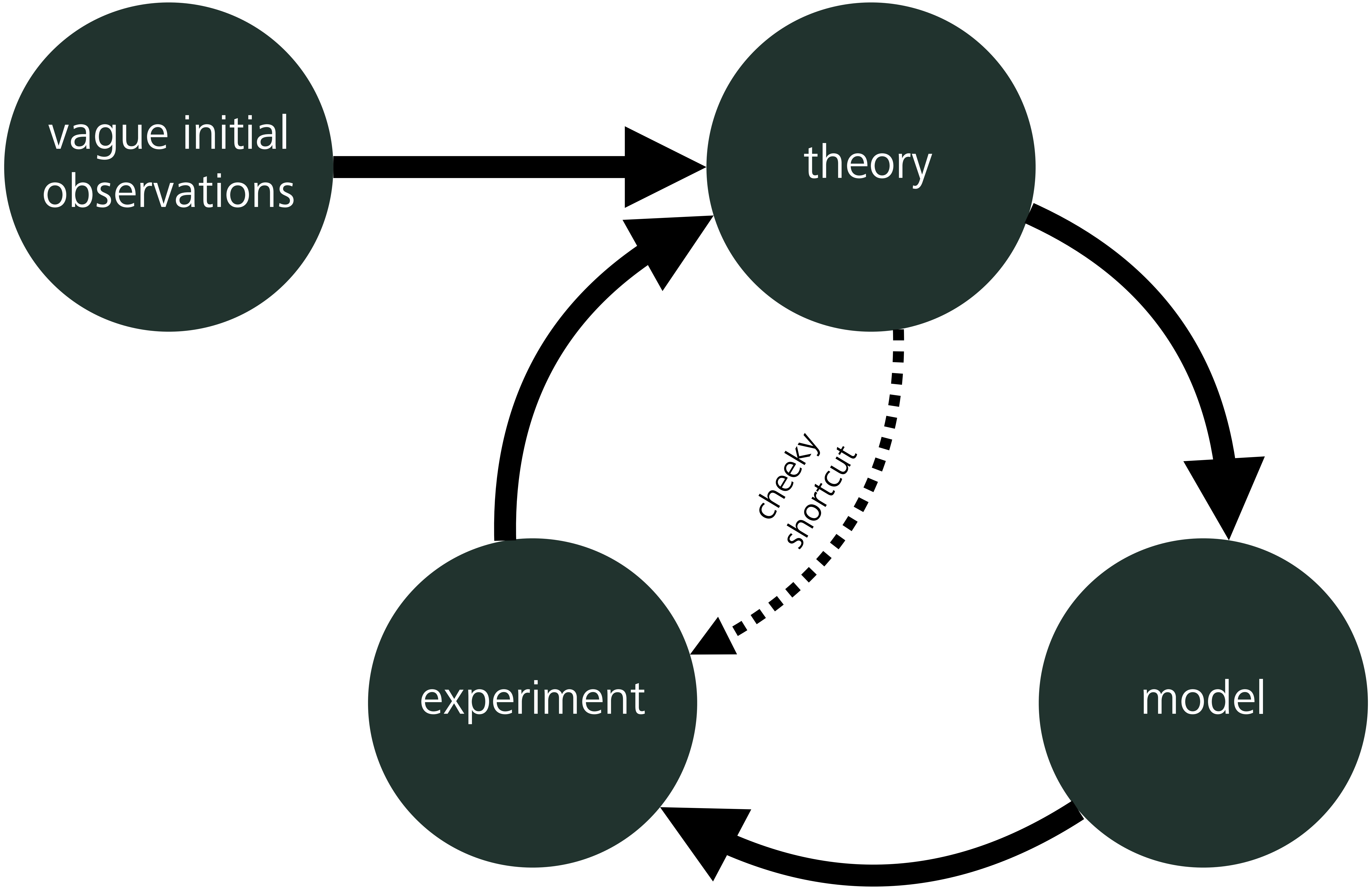
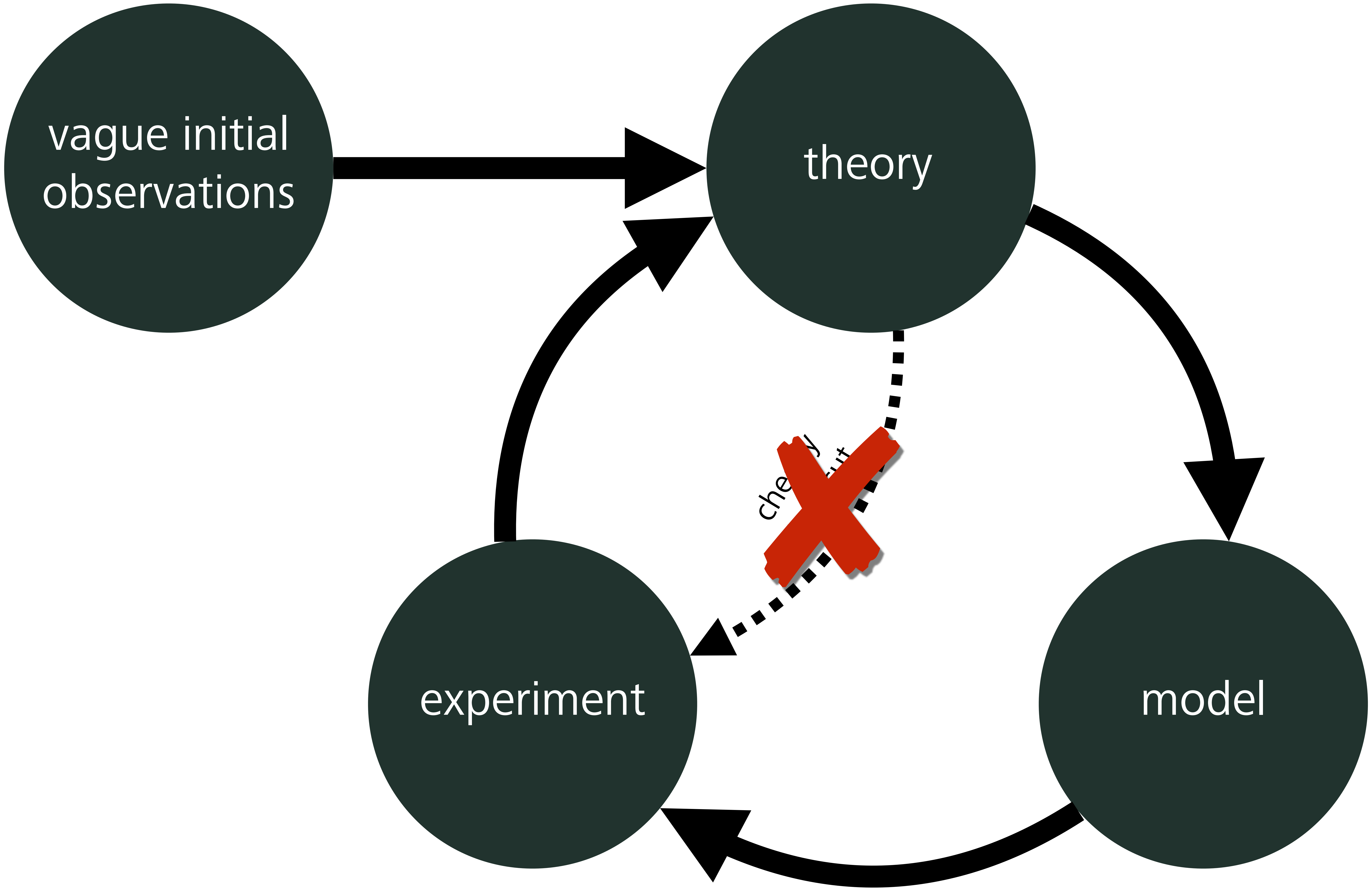
Bayesian models of cognition: From individual biases to emergent cultural phenomena

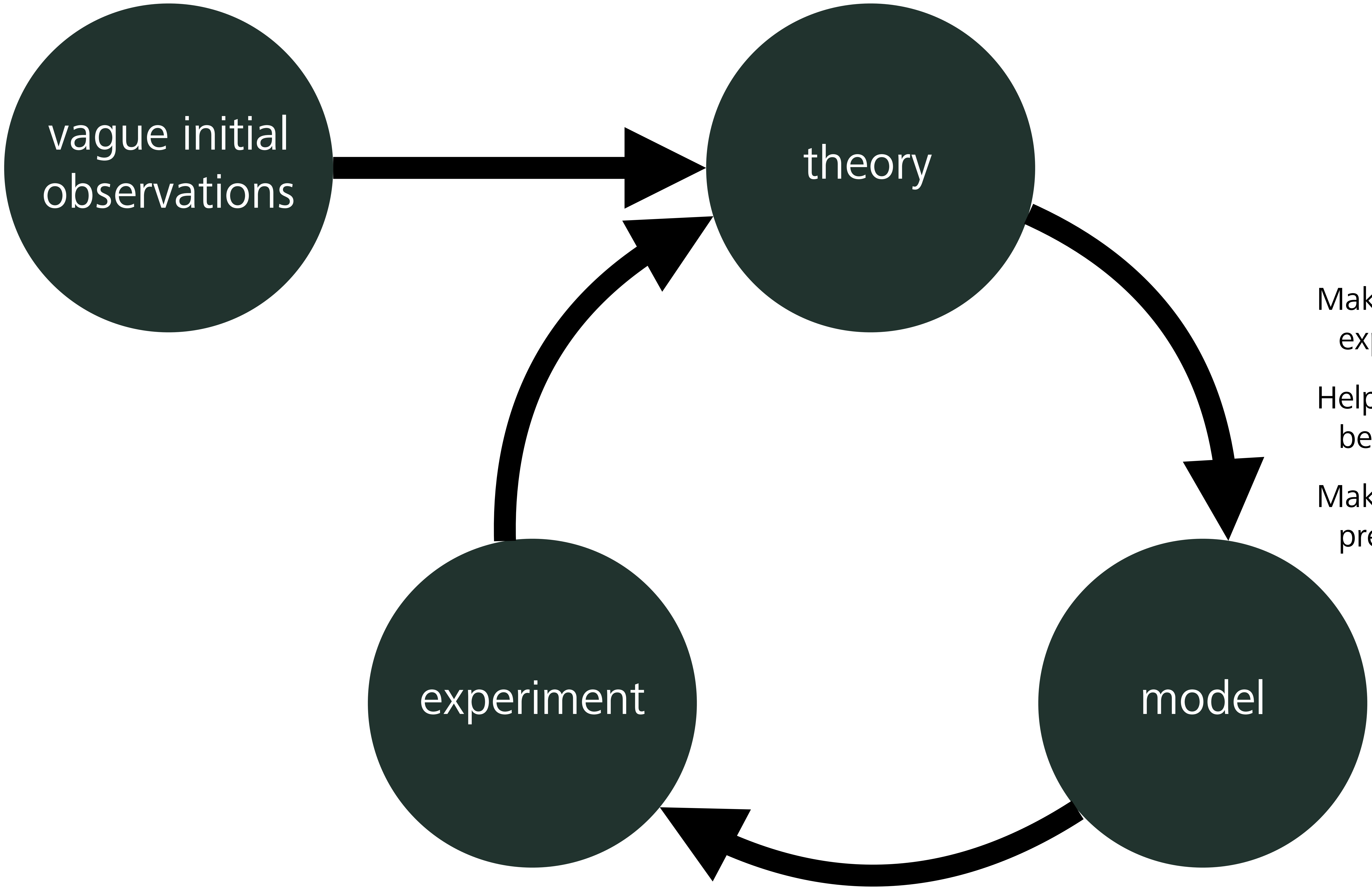
Jon W. Carr
International School for Advanced Studies, Trieste, Italy


theory
model
experiment


vague initial observations
theory
model
experiment

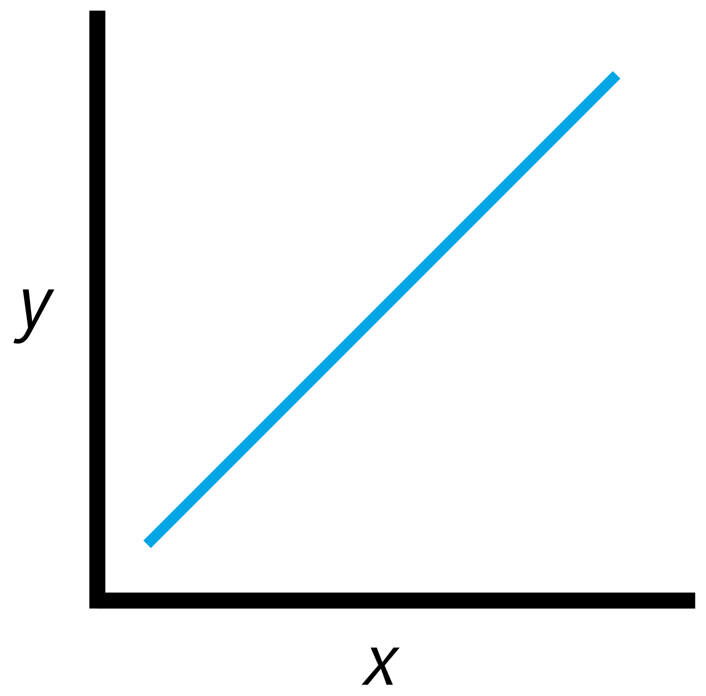

vague initial observations
theory
model
experiment
cheeky shortcut


vague initial observations
theory
model
experiment



- Makes theoretical assumptions explicit
- Helps you refine your theory before moving to experiments
- Makes concrete experimental predictions

linear regression, etc.

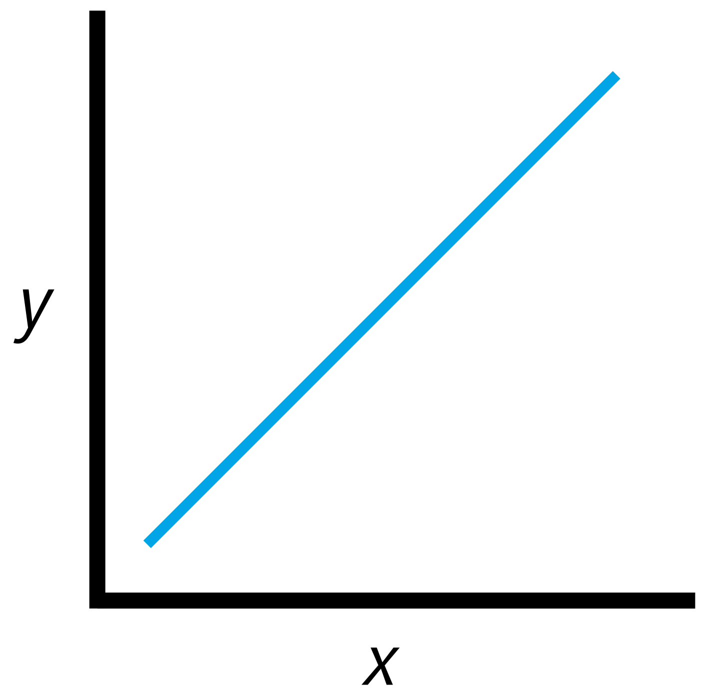


machine learning, etc.

cognitively abstract
simple to understand
not very predictive

cognitively realistic
complex black box
highly predictive

linear regression, etc.

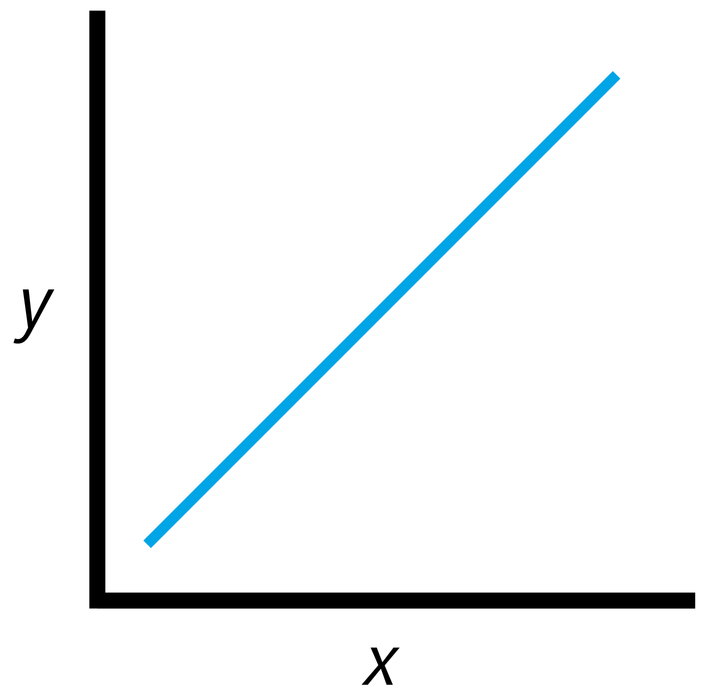


machine learning, etc.

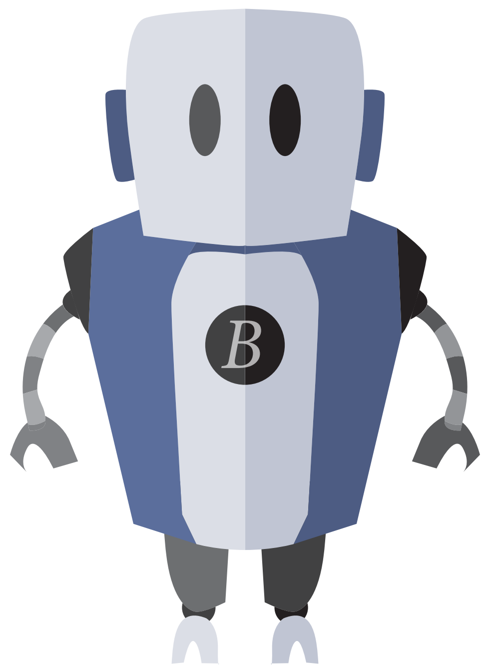
cognitively abstract
simple to understand
not very predictive
uninformative

cognitively realistic
complex black box
highly predictive
uninformative

linear regression, etc.



probabilistic models of cognition

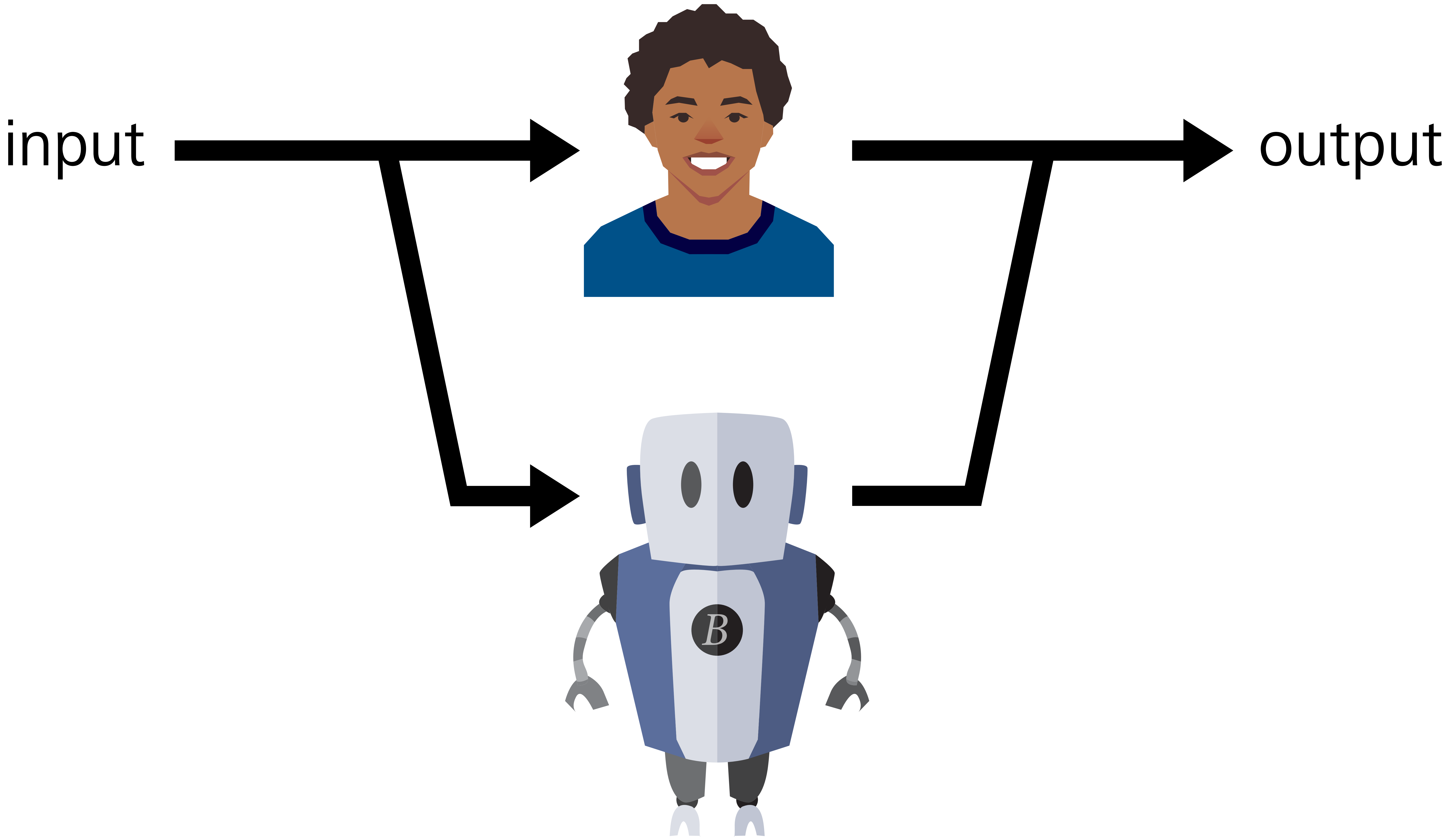


machine learning, etc.

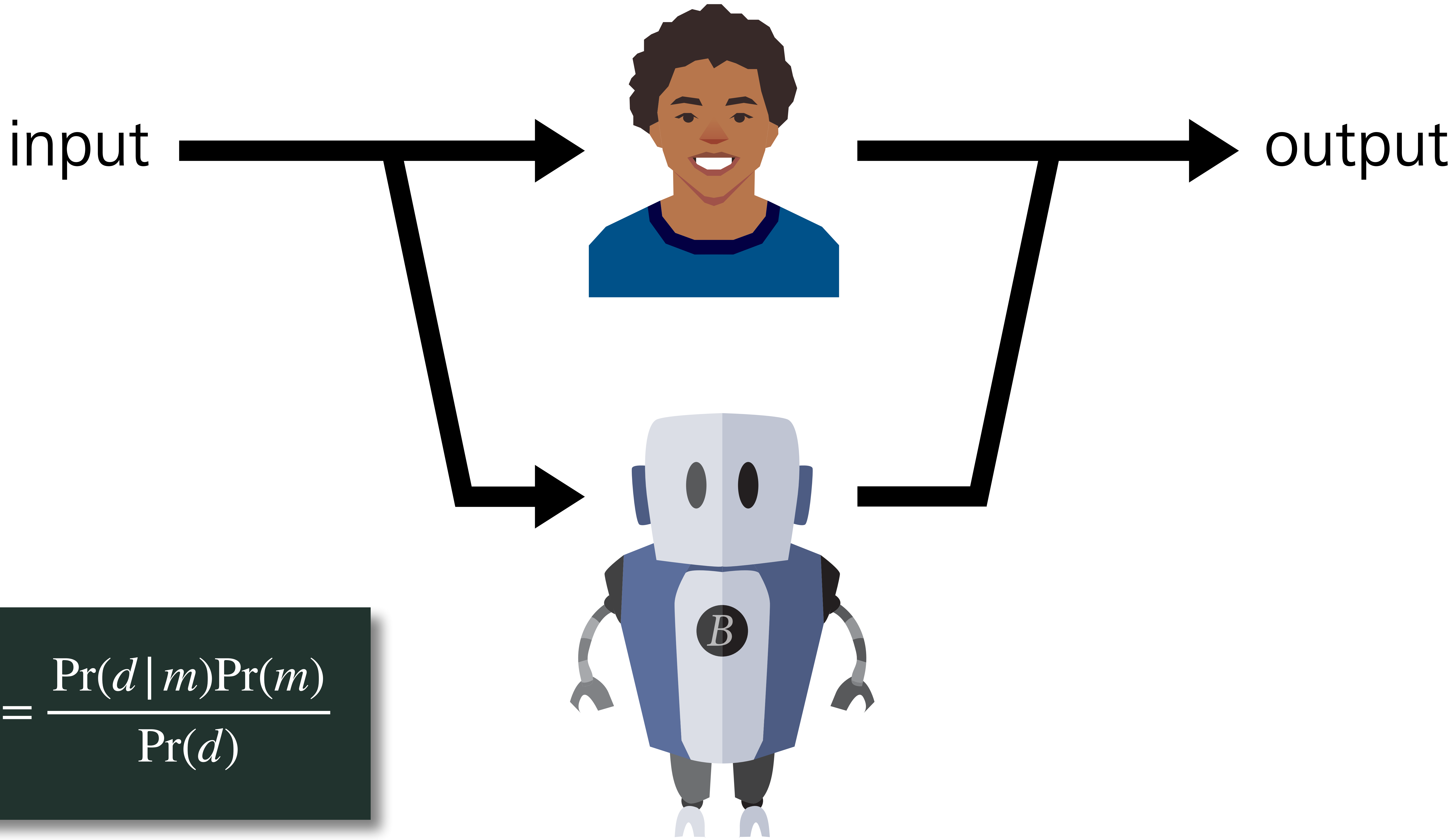
cognitively abstract
simple to understand
not very predictive
uninformative

sweet spot

cognitively realistic
complex black box
highly predictive
uninformative

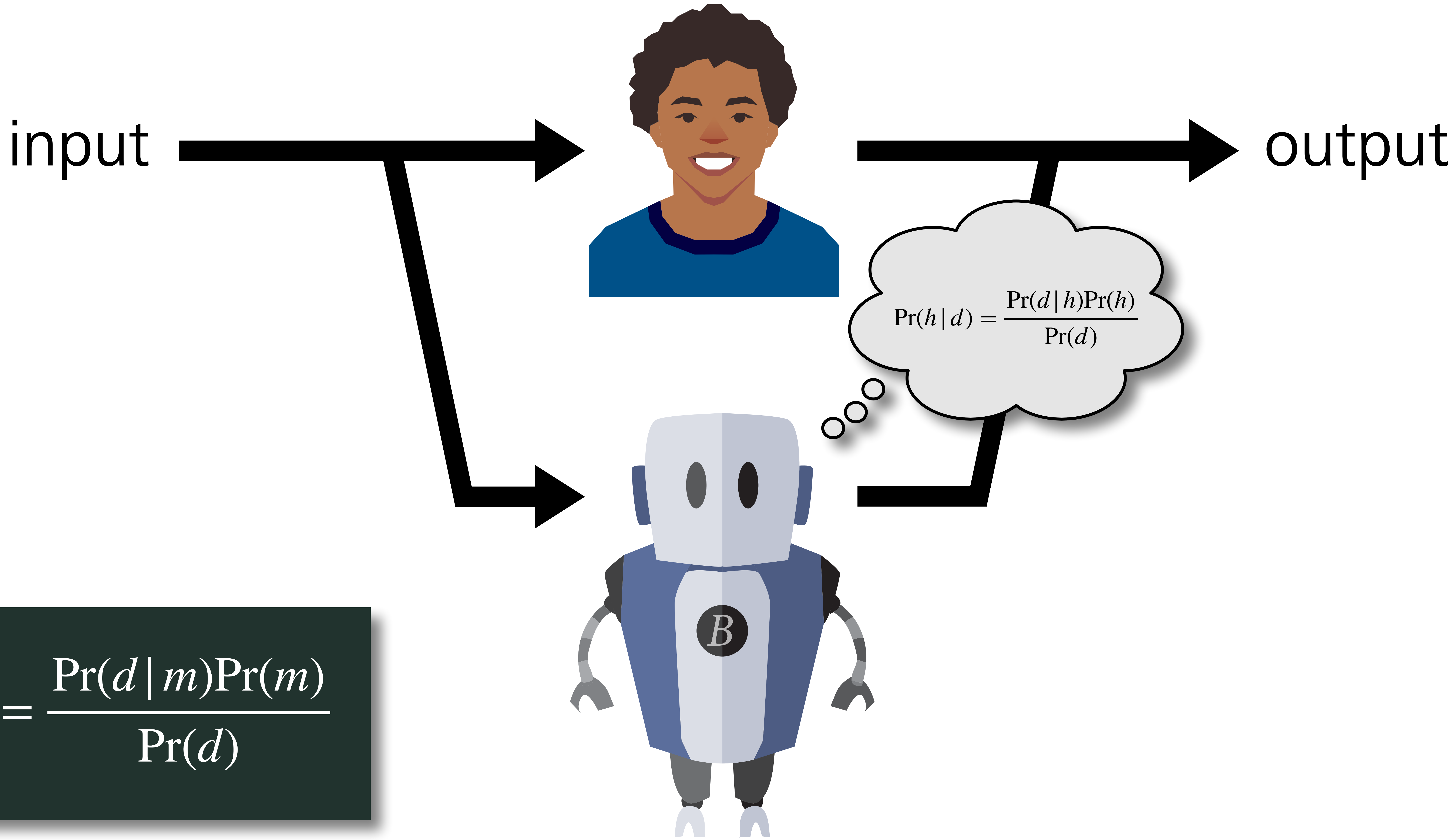


A **mechanistic** proposal about the **probability** of certain outputs being **generated** given certain inputs



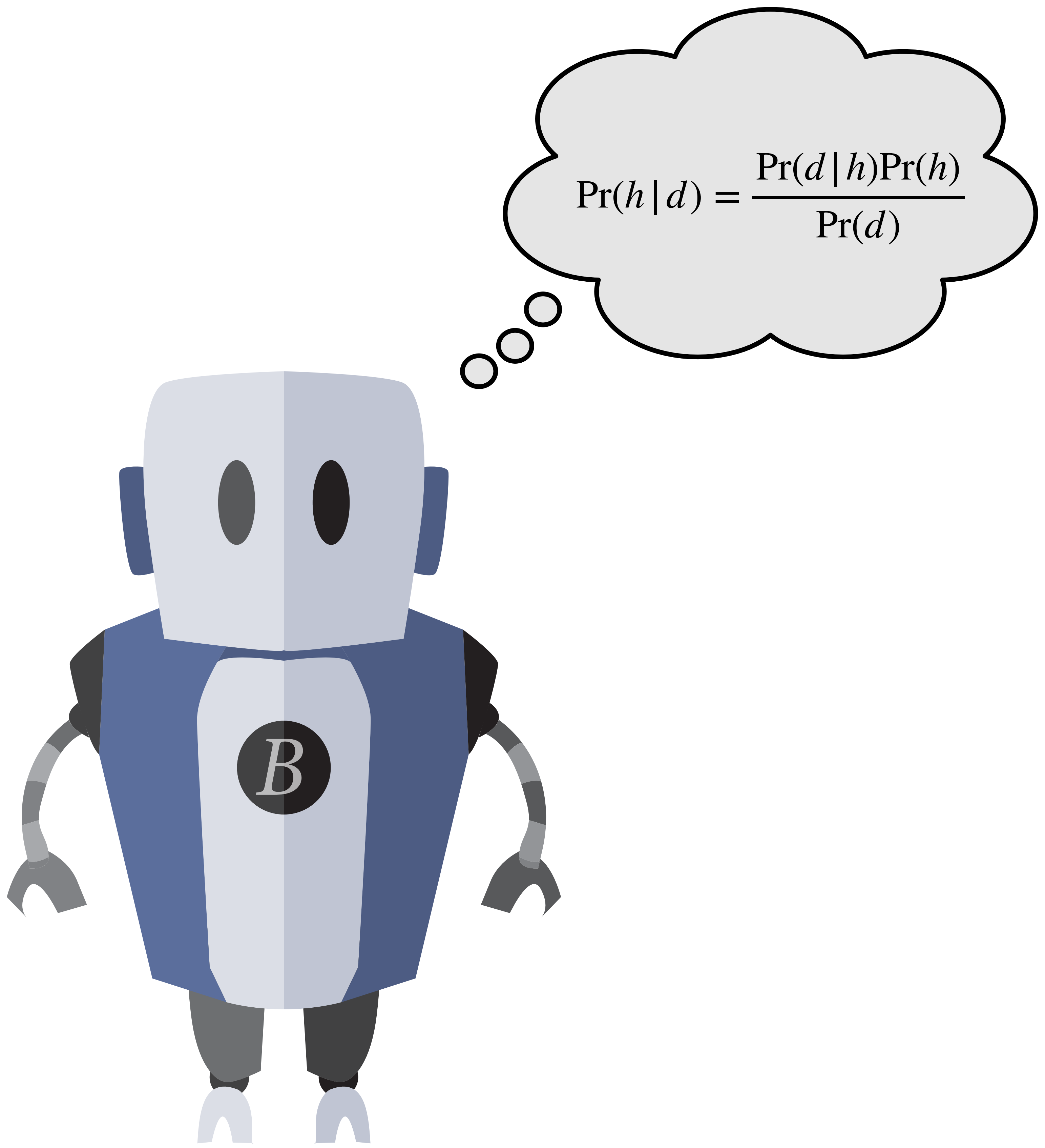
$$\Pr(m \mid d) = \frac{\Pr(d \mid m)\Pr(m)}{\Pr(d)}$$

A **mechanistic** proposal about the **probability** of certain outputs being **generated** given certain inputs

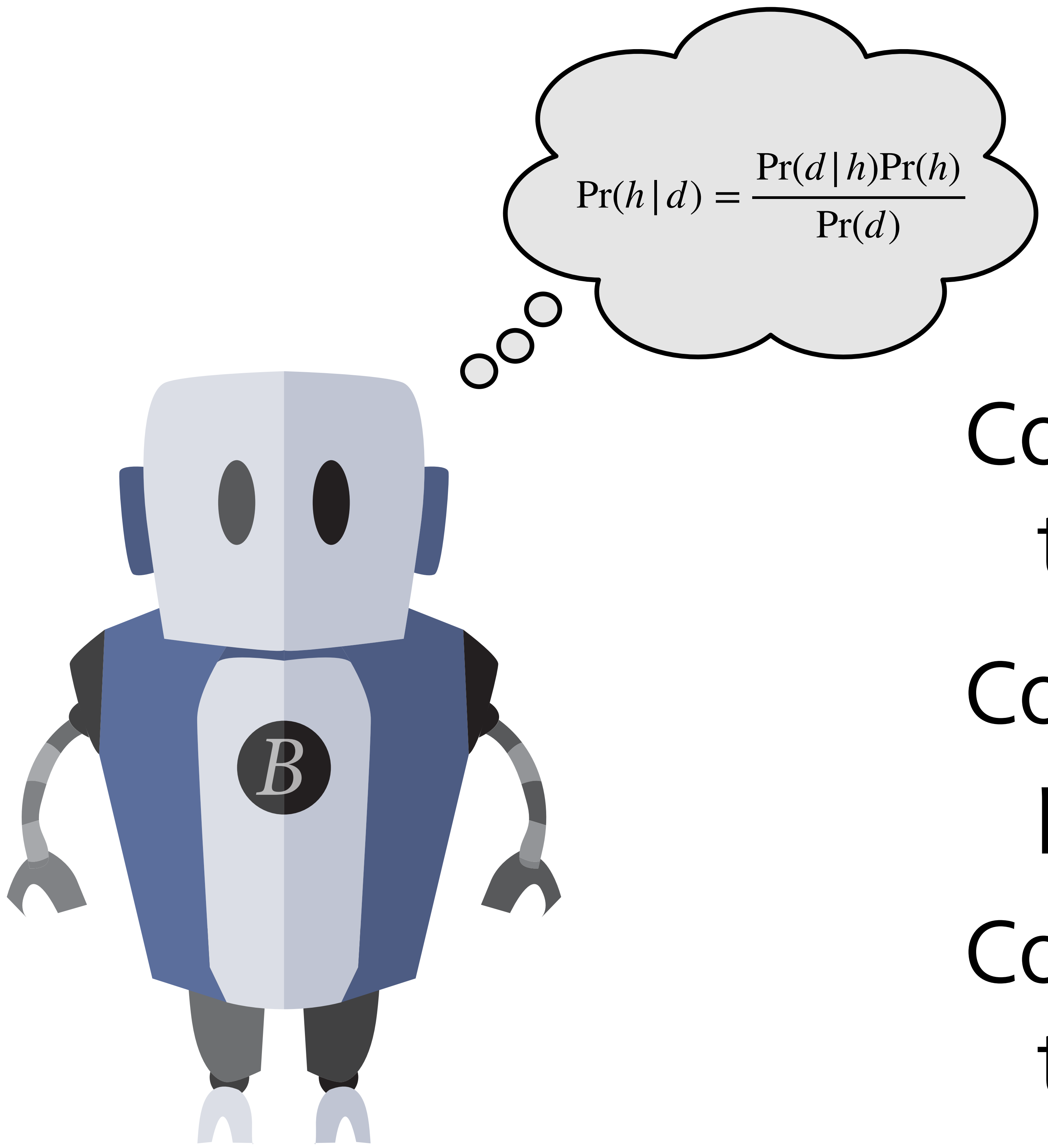


$$\Pr(m \mid d) = \frac{\Pr(d \mid m)\Pr(m)}{\Pr(d)}$$

A **mechanistic** proposal about the **probability** of certain outputs being **generated** given certain inputs



Bayesian models of cognition allow us to explore how **rational agents** combine **observational data** with **prior knowledge** to form **beliefs** about the world



Cognitive processes are represented in the **likelihood** function

Cognitive biases are represented by the **prior**

Cognitive variables are made explicit by the model **parameters**

Bayesian models of cognition allow us to explore how **rational agents** combine **observational data** with **prior knowledge** to form **beliefs** about the world

	Project 1: Visual word identification	Project 2: Evolution of semantic categories
hypothesis space	*words*	*languages*
likelihood	*interesting*	*boring*
prior	*boring*	*interesting*
parameters	*individual-level*	*group-level*
data	*sensory input*	*meaning-signal pairs*

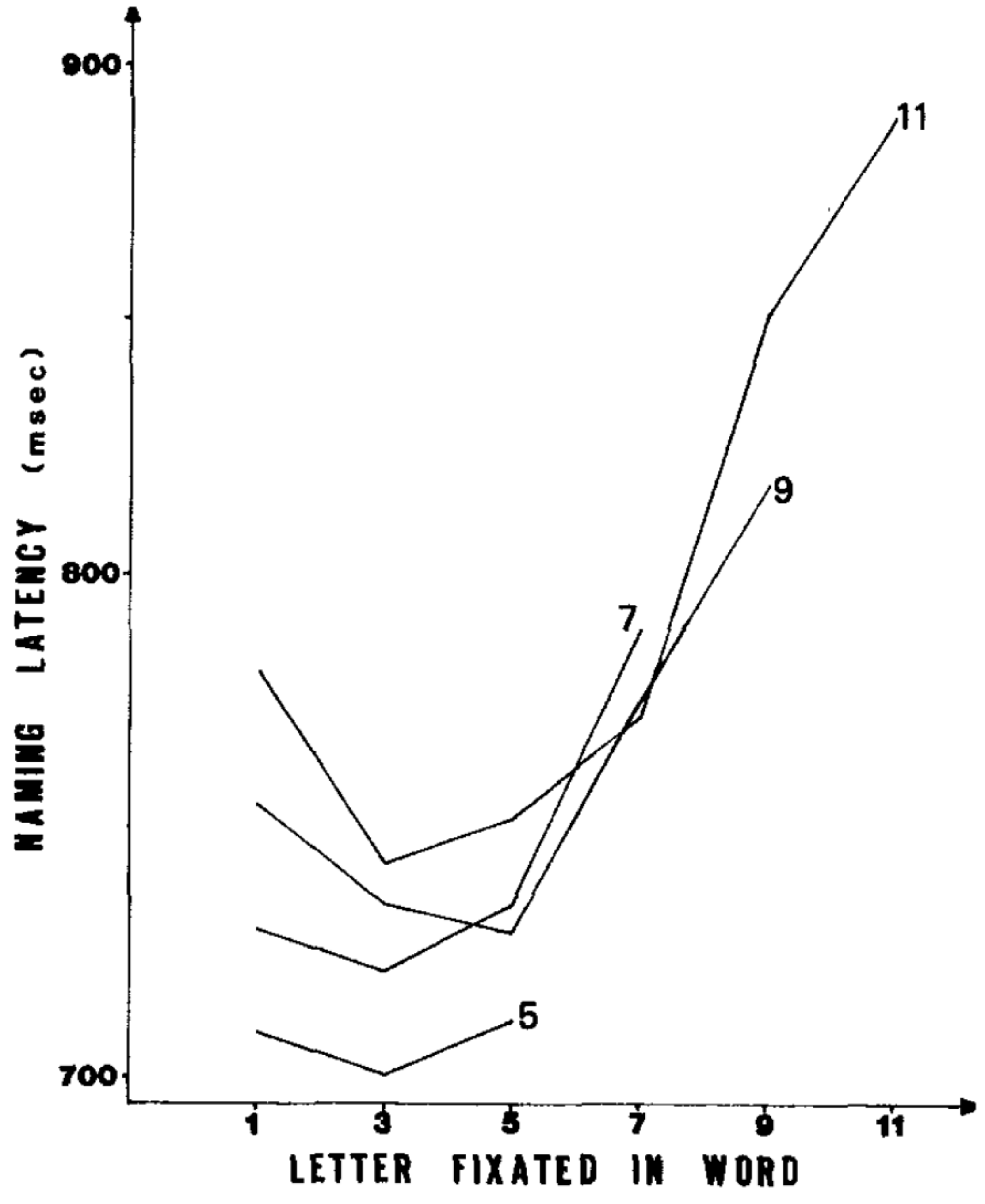
Orthographic informativity and the optimal viewing position


Jon Carr

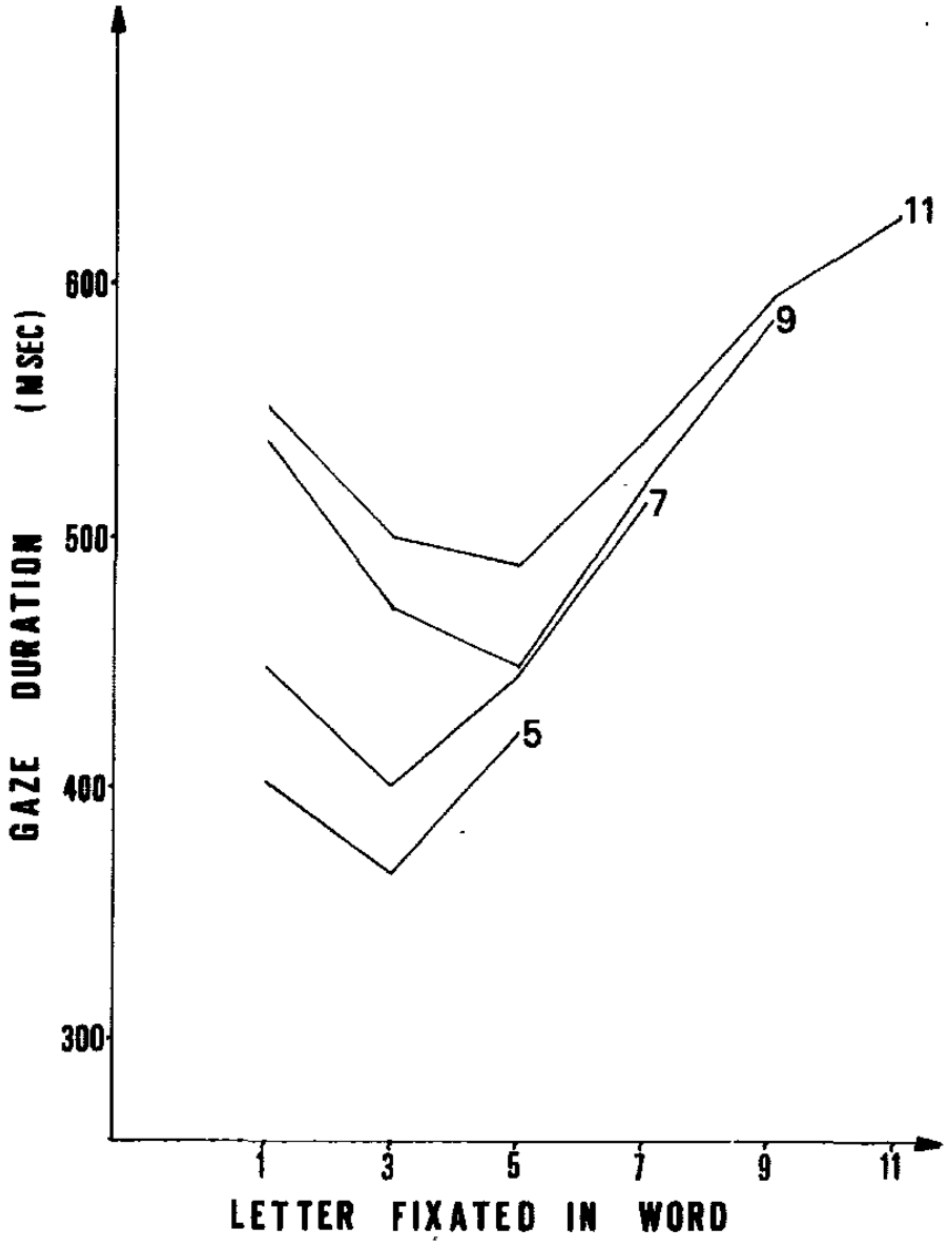

Davide Crepaldi

Optimal viewing position: Readers are better at identifying words when fixating them left-of-center

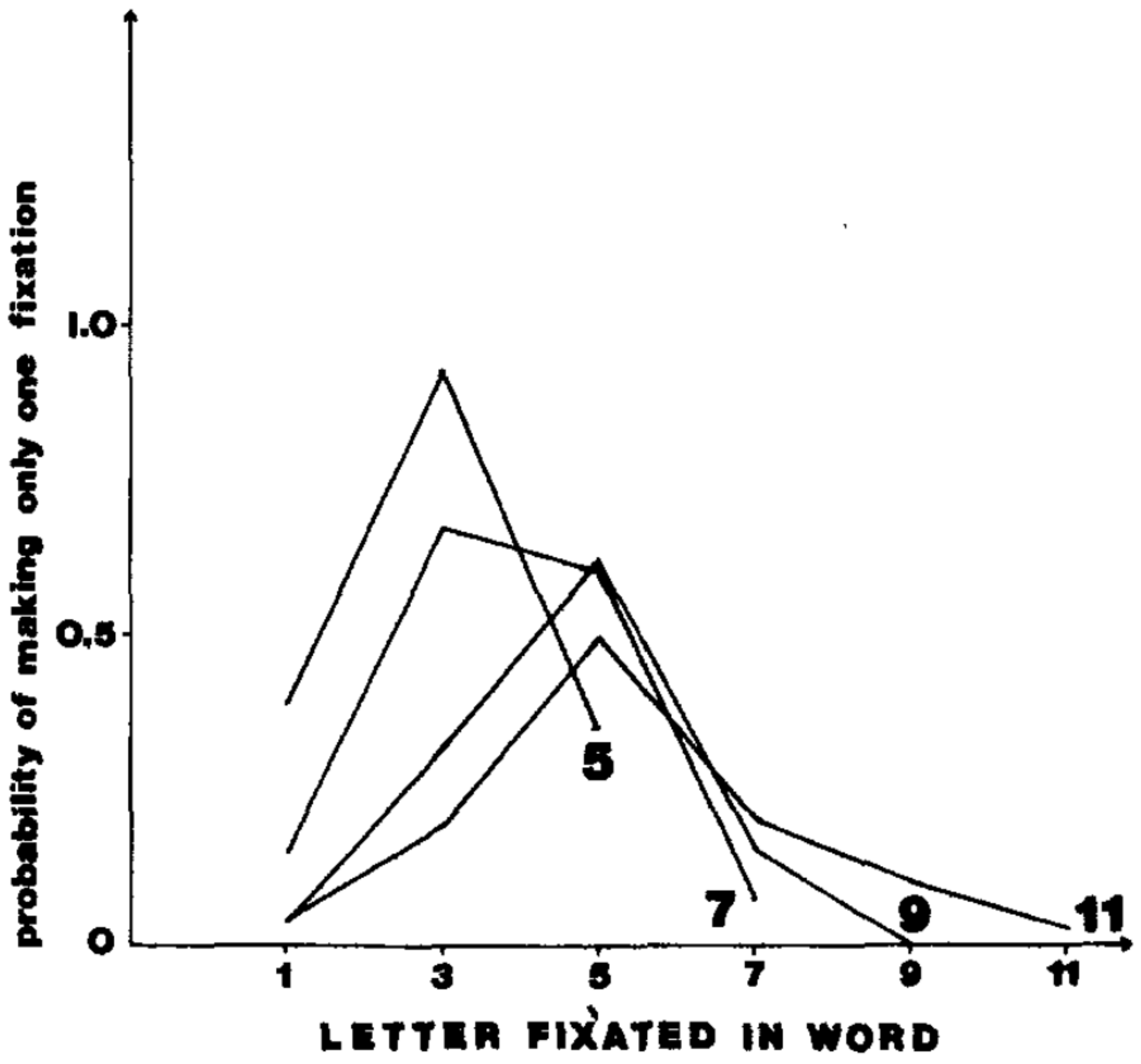
Naming latency



Gaze duration



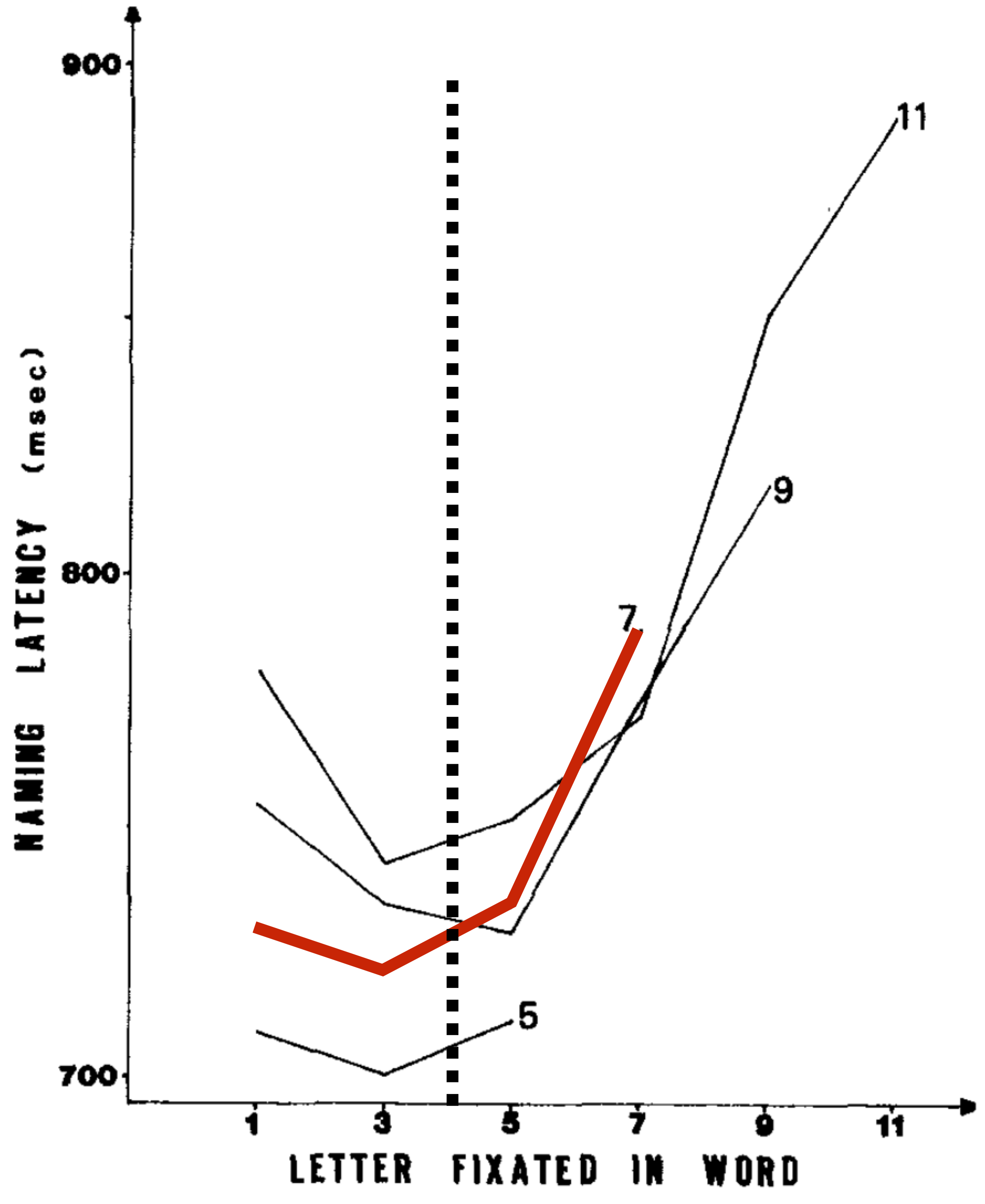
Probability of one fixation



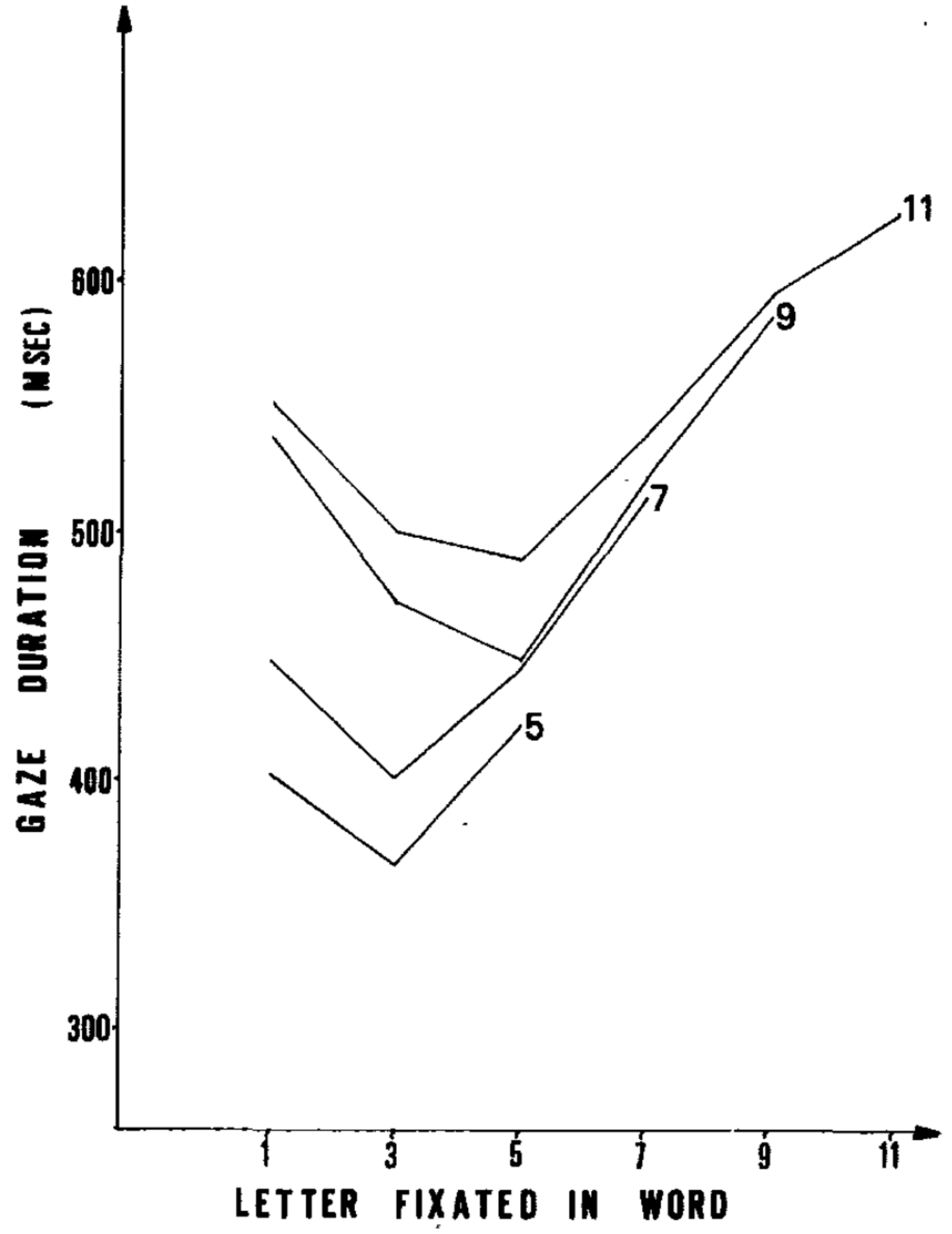
O'Regan et al. (1984)

Optimal viewing position: Readers are better at identifying words when fixating them left-of-center

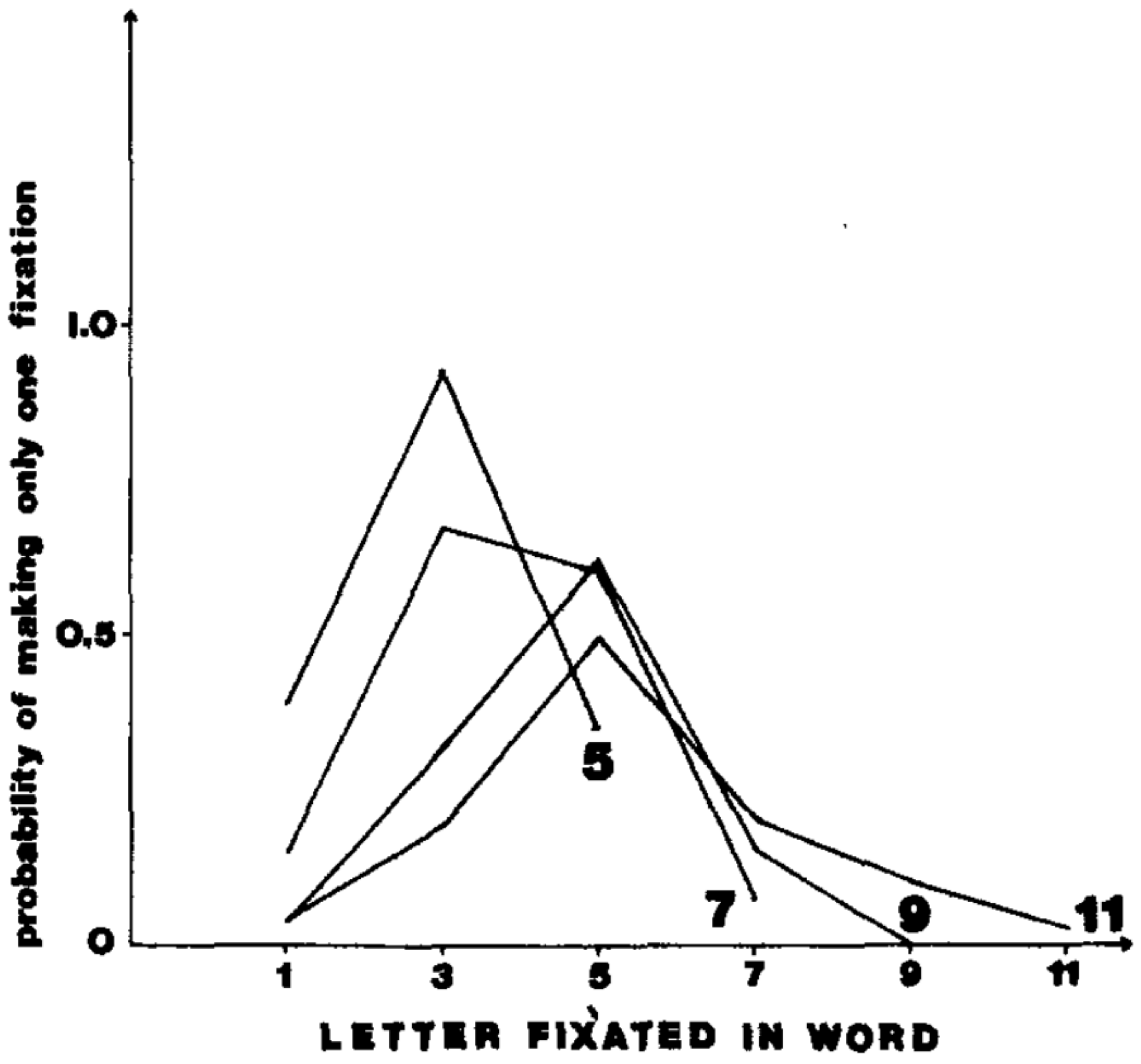
Naming latency



Gaze duration

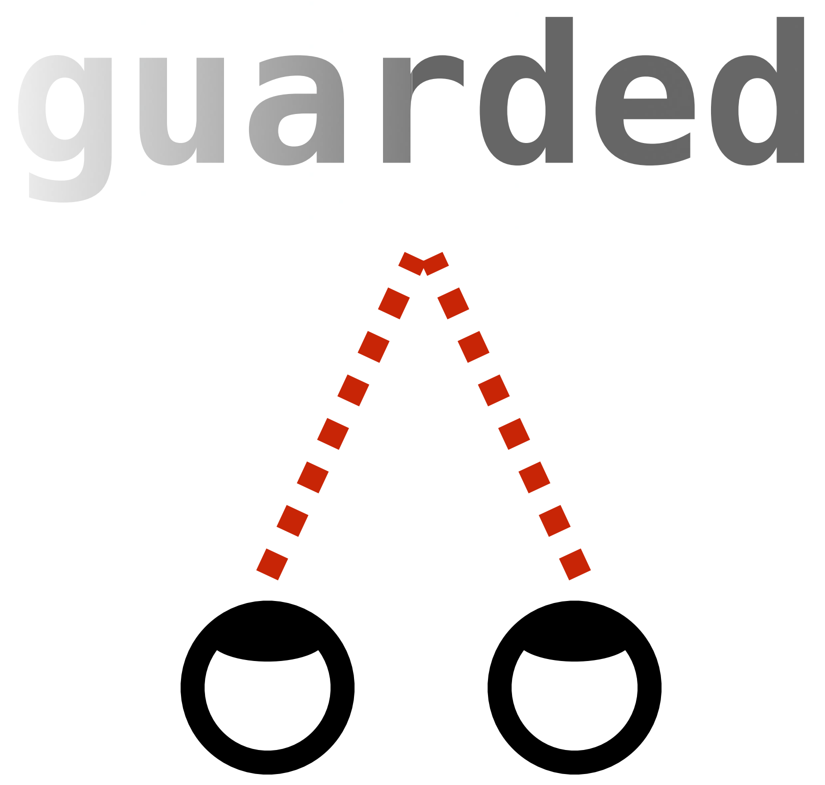


Probability of one fixation

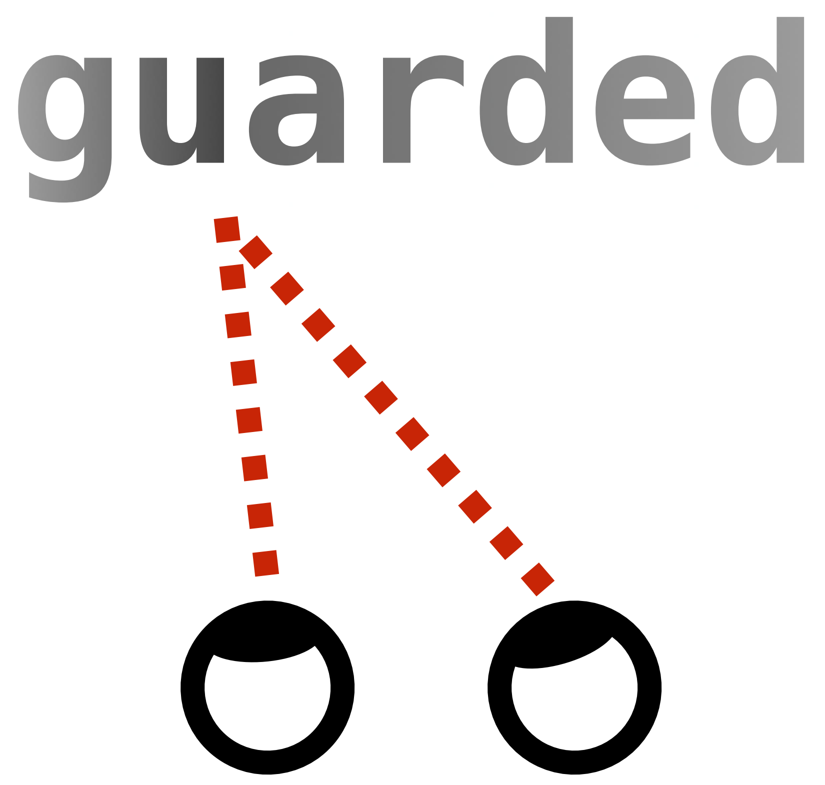


O'Regan et al. (1984)

Theory 1: Bias in the human perceptual span

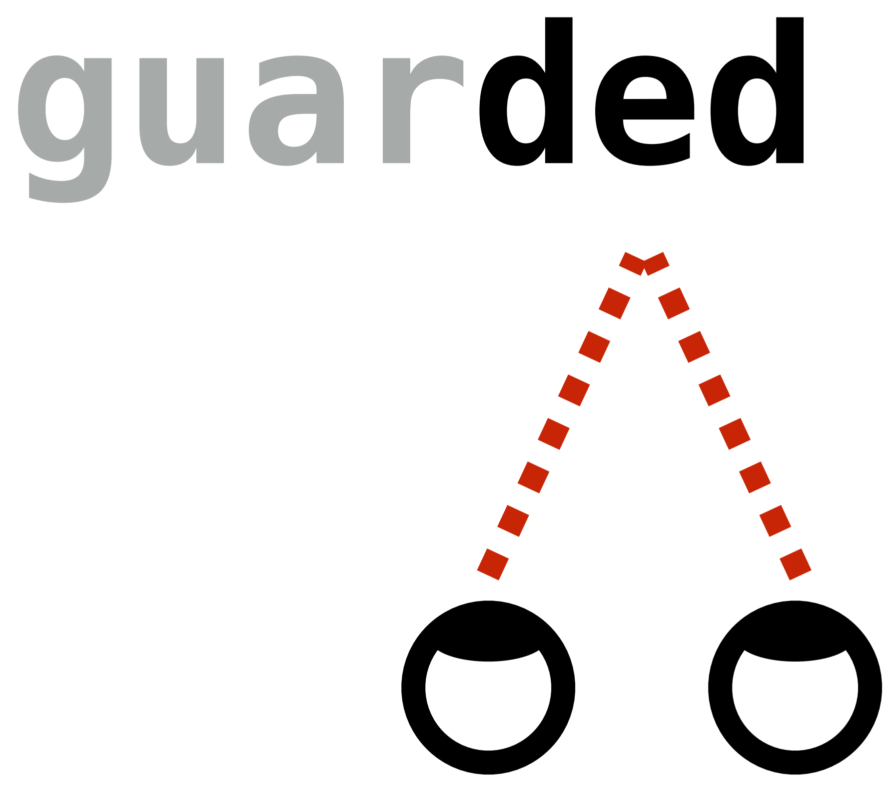


Theory 1: Bias in the human perceptual span



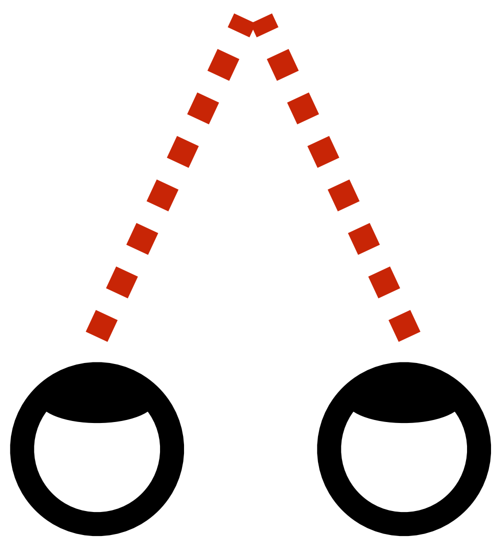
Fixating left-of-center counteracts the asymmetry of the human perceptual span

Theory 2: Bias in the structure of the lexicon



Theory 2: Bias in the structure of the lexicon

guarded

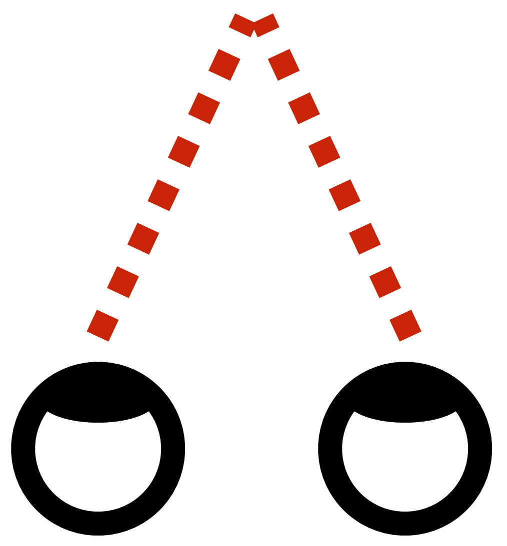


????ded

Candidate words: *alluded*, *amended*, *avoided*, *awarded*, *bearded*, *blended*, *blinded*, *boarded*, *bounded*, *braided*, *branded*, *crowded*, *decided*, ..., *wielded*, *wounded*, *yielded*

Theory 2: Bias in the structure of the lexicon

guarded



gua????

Candidate words: *guarded*, *guanine*

Fixating left-of-center reduces the reader's uncertainty about word identity

Human perceptual biases

e.g. asymmetry of the perceptual span

Structure of the lexicon

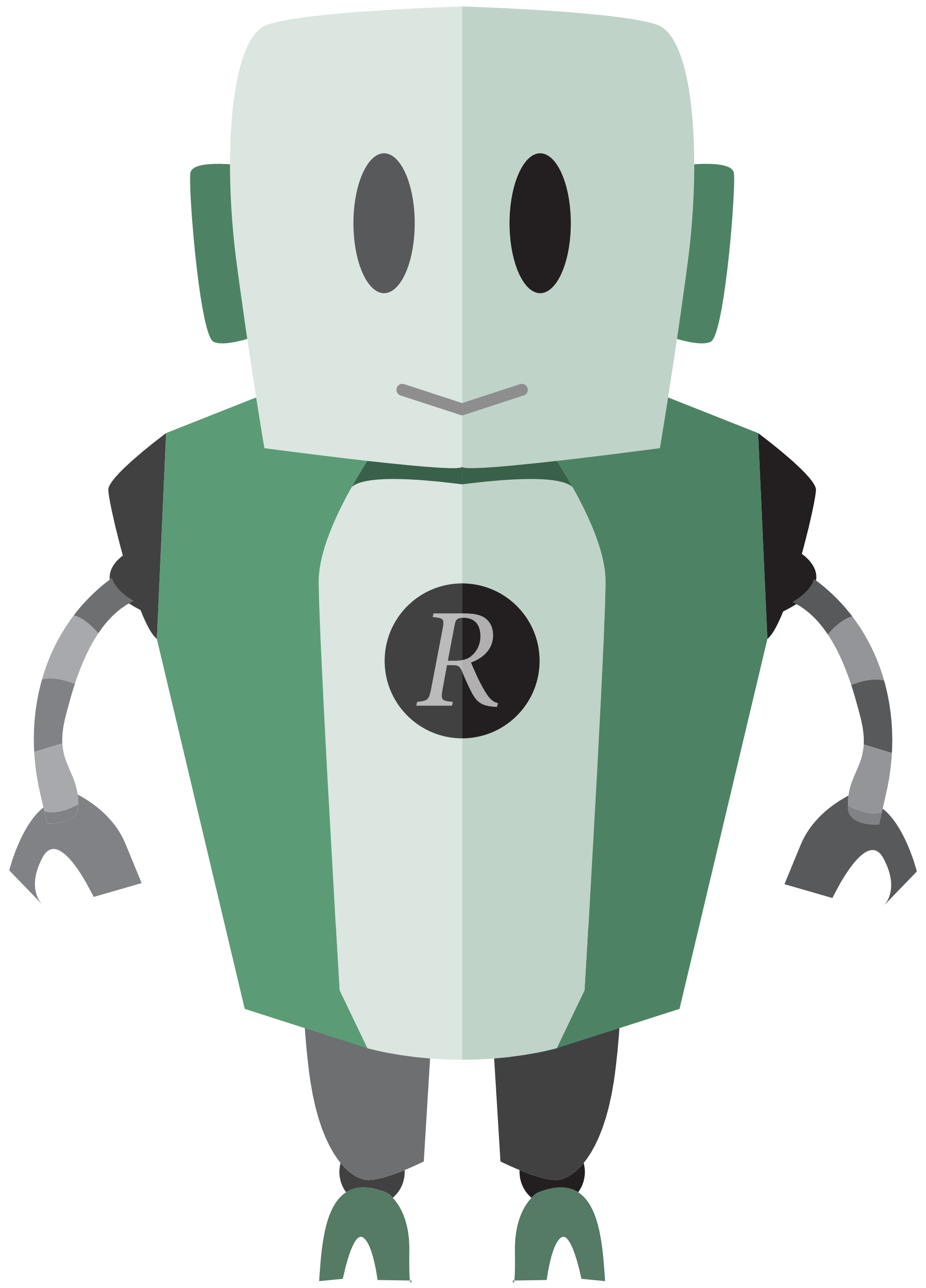
e.g. suffixing vs. prefixing

Optimal viewing position in visual word identification

left-of-center vs. right-of-center

Pattern of fixations in reading

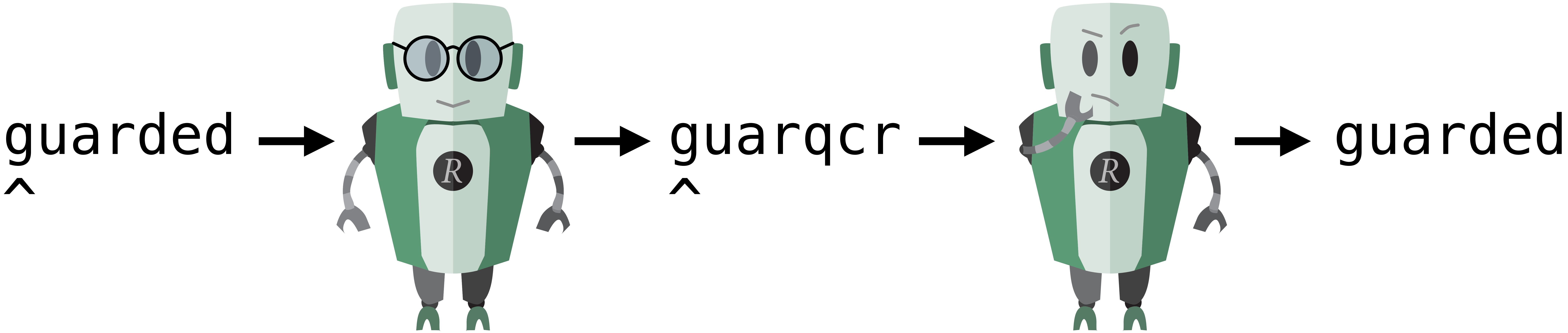
e.g. initial landing position, gaze duration, probability of abandonment

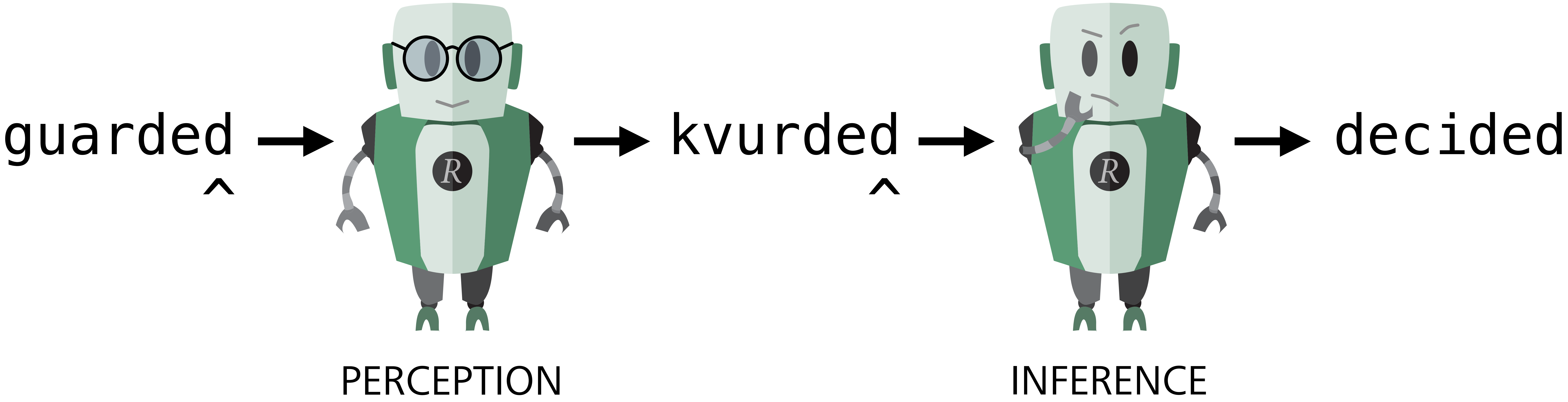


$$S = \{\texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{z}\}$$

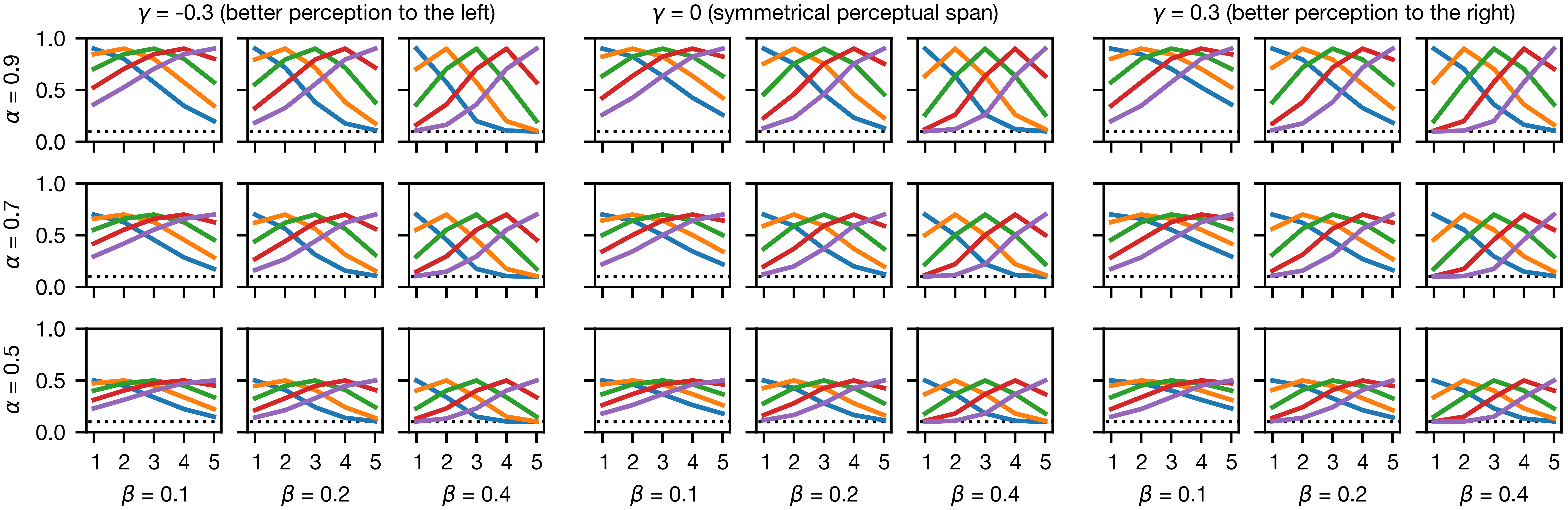
$$\xrightarrow{m = 7}$$

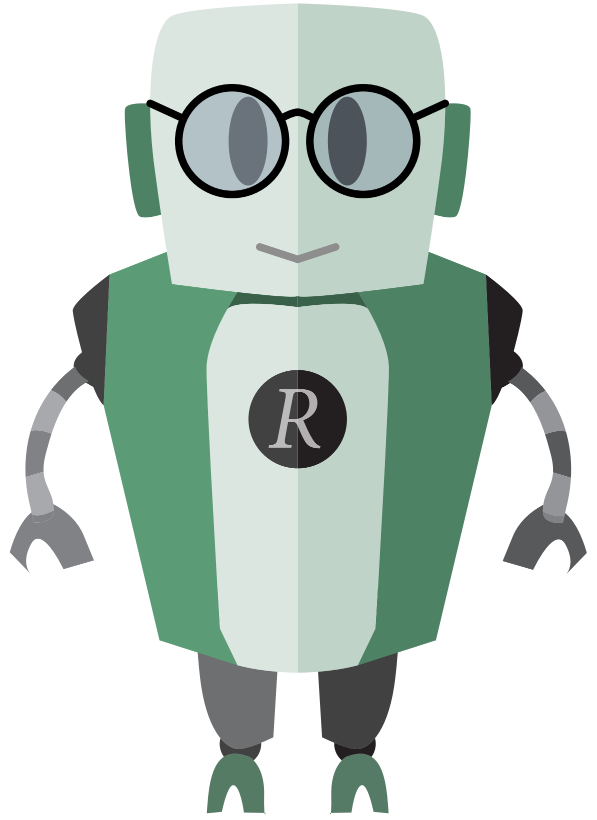
$$W = \{\texttt{abandon}, \ldots, \texttt{zombies}\}$$


guarded
^
guarqcr
^
guarded

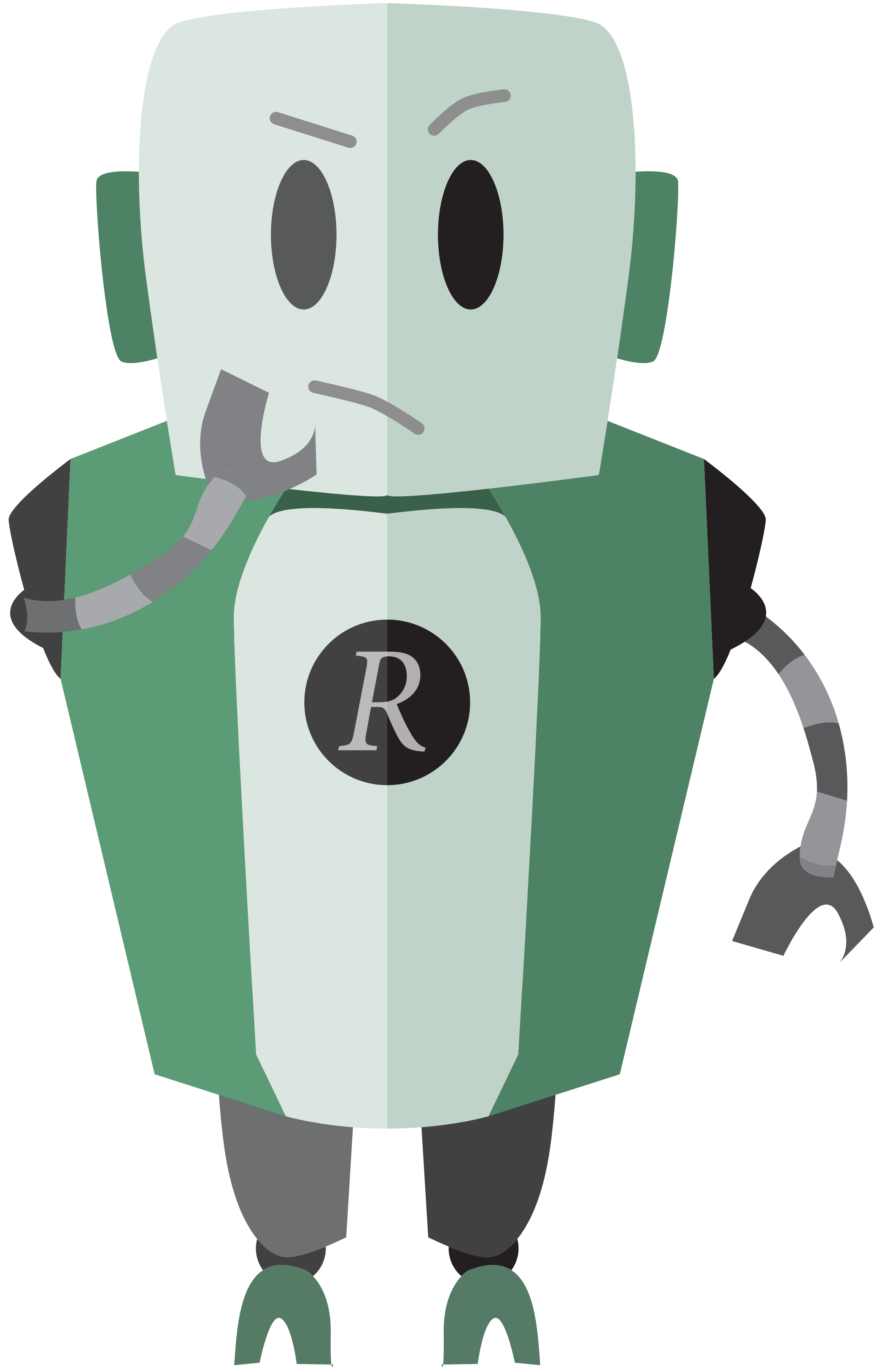


guarded
^
kvurded
^
decided
PERCEPTION
INFERENCE



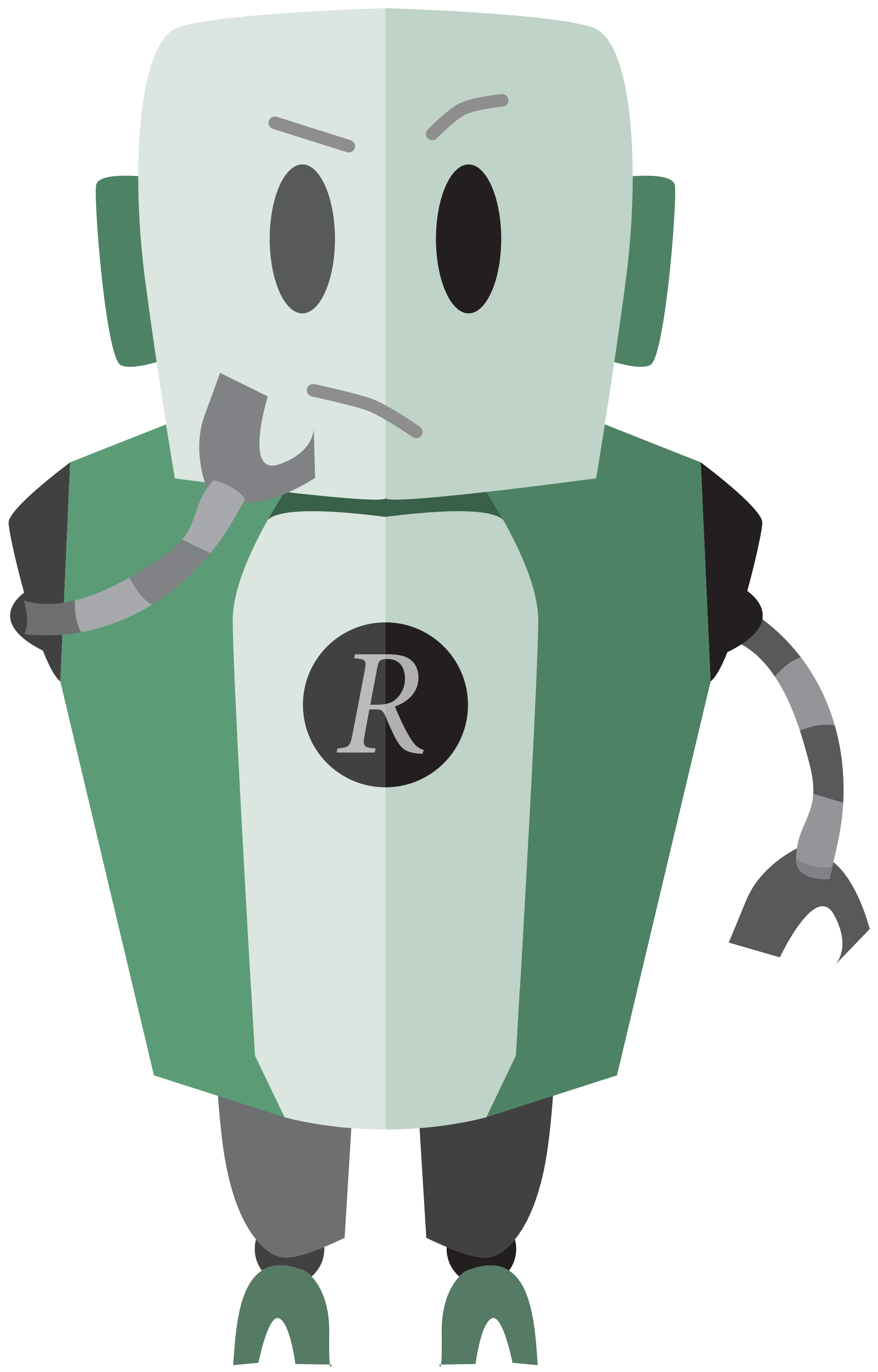


$$\Phi(i\,|\,j) = \begin{cases} \left(\alpha - \frac{1}{|S|}\right)\exp[\beta(\gamma - 1)(i - j)] + \frac{1}{|S|} & \text{if } i > j \\ \left(\alpha - \frac{1}{|S|}\right)\exp[\beta(\gamma + 1)(i - j)] + \frac{1}{|S|} & \text{otherwise} \end{cases}$$



posterior · *likelihood* · *prior*

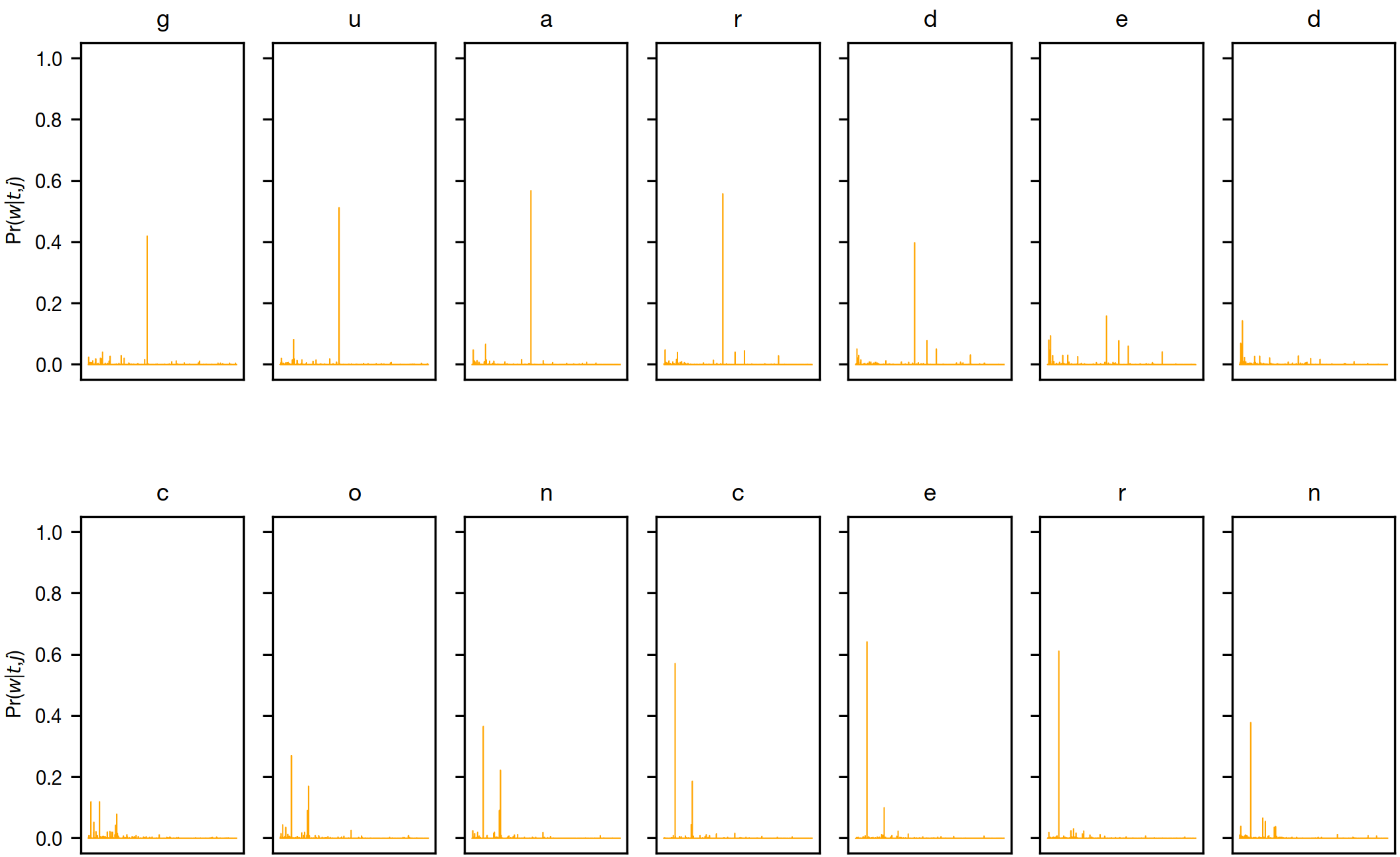
$$\underbrace{\Pr(w \mid p, j)}_{\text{posterior}} \propto \underbrace{\Pr(p \mid w, j)}_{\text{likelihood}}\underbrace{\Pr(w)}_{\text{prior}}$$

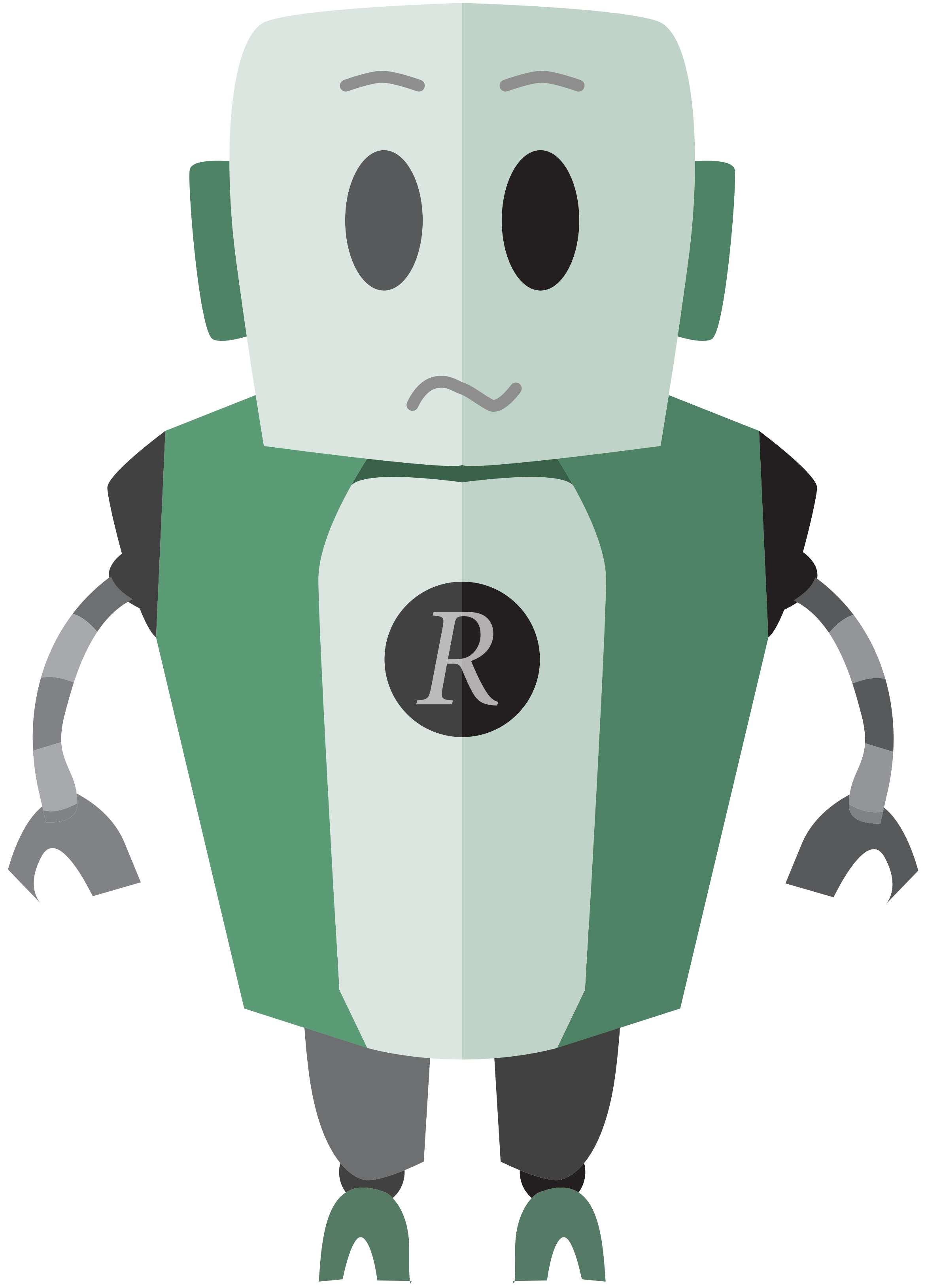


$$\overset{\textit{posterior}}{\Pr(w \mid p, j)} \propto \overset{\textit{likelihood}}{\Pr(p \mid w, j)}\overset{\textit{prior}}{\Pr(w)}$$

$$\Pr(p \mid w, j) = \prod_{i=1}^{m} \begin{cases} \Phi(i \mid j) & \text{if } p(i) = w(i) \\ \dfrac{1 - \Phi(i \mid j)}{|S| - 1} & \text{otherwise} \end{cases}$$

$$\Pr(w) = \frac{\mathrm{F}(w)}{\sum_{w' \in W} \mathrm{F}(w')}$$


g
u
a
r
d
e
d
c
o
n
c
e
r
n
Pr(w|t,j)
Pr(w|t,j)
1.0
0.8
0.6
0.4
0.2
0.0

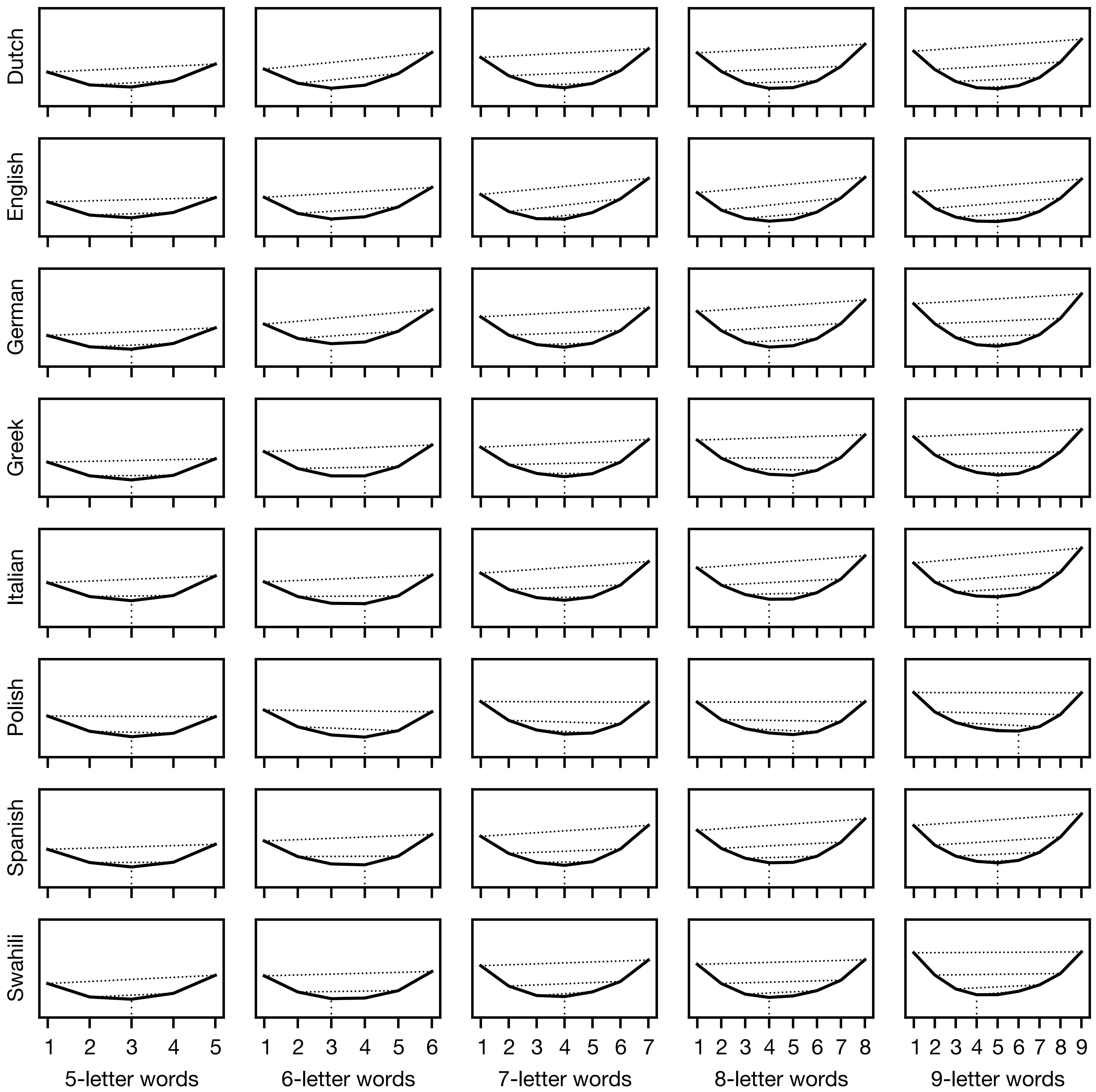


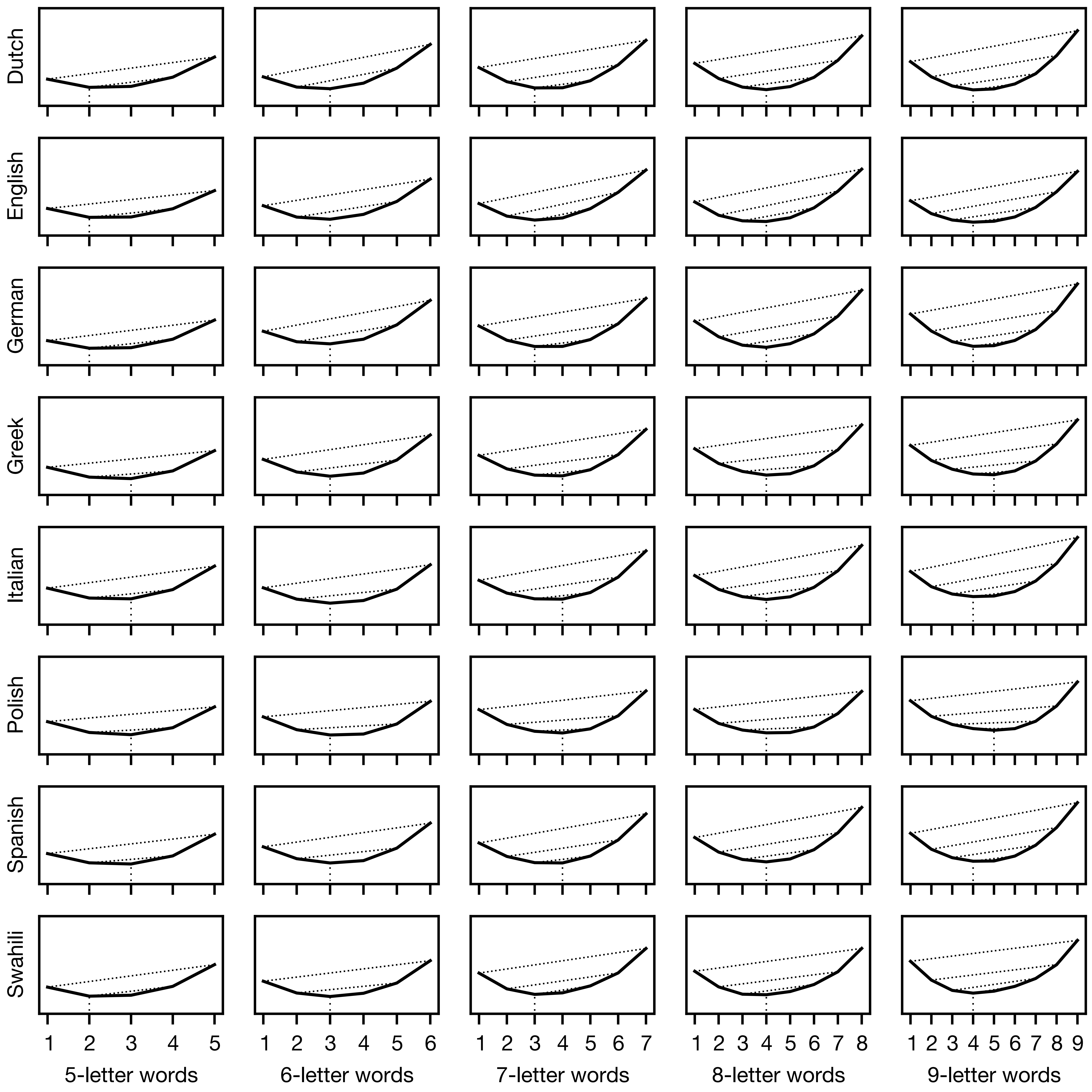
Uncertainty Expected entropy of the reader's posterior, when fixating in a given location

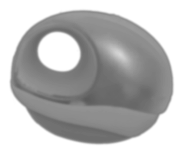
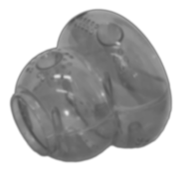
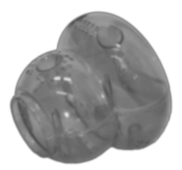
$$\mathrm{U}(j) = \sum_{t \in W} \sum_{p \in P} \Pr(t)\Pr(p \mid t, j)\mathrm{H}(W \mid p, j)$$

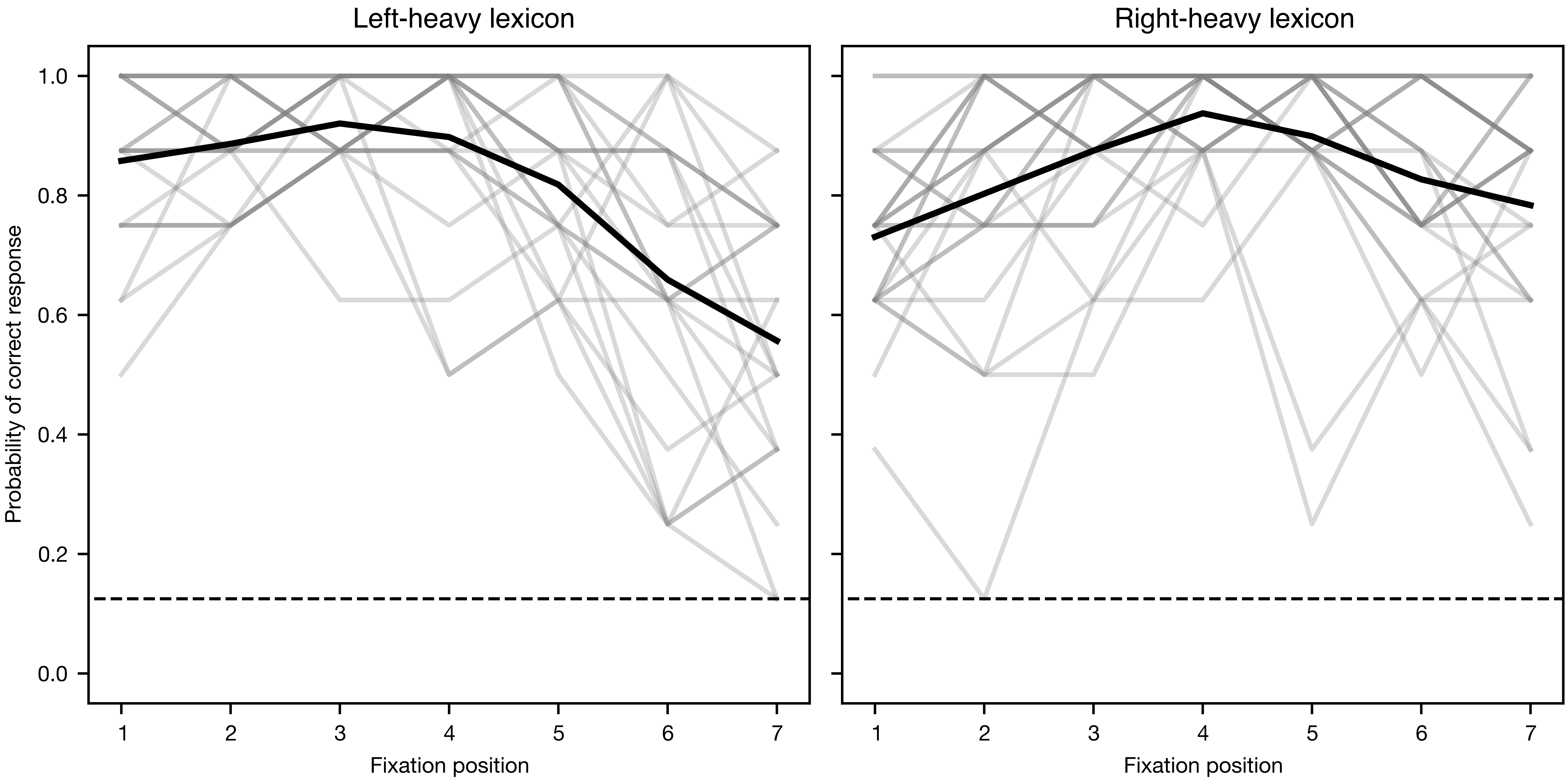
$$\mathrm{H}(W \mid p, j) = -\sum_{w \in W} \Pr(w \mid p, j)\log[\Pr(w \mid p, j)]$$

$$\text{optimal viewing position} = \arg\min_{j=1}^{m} \mathrm{U}(j)$$

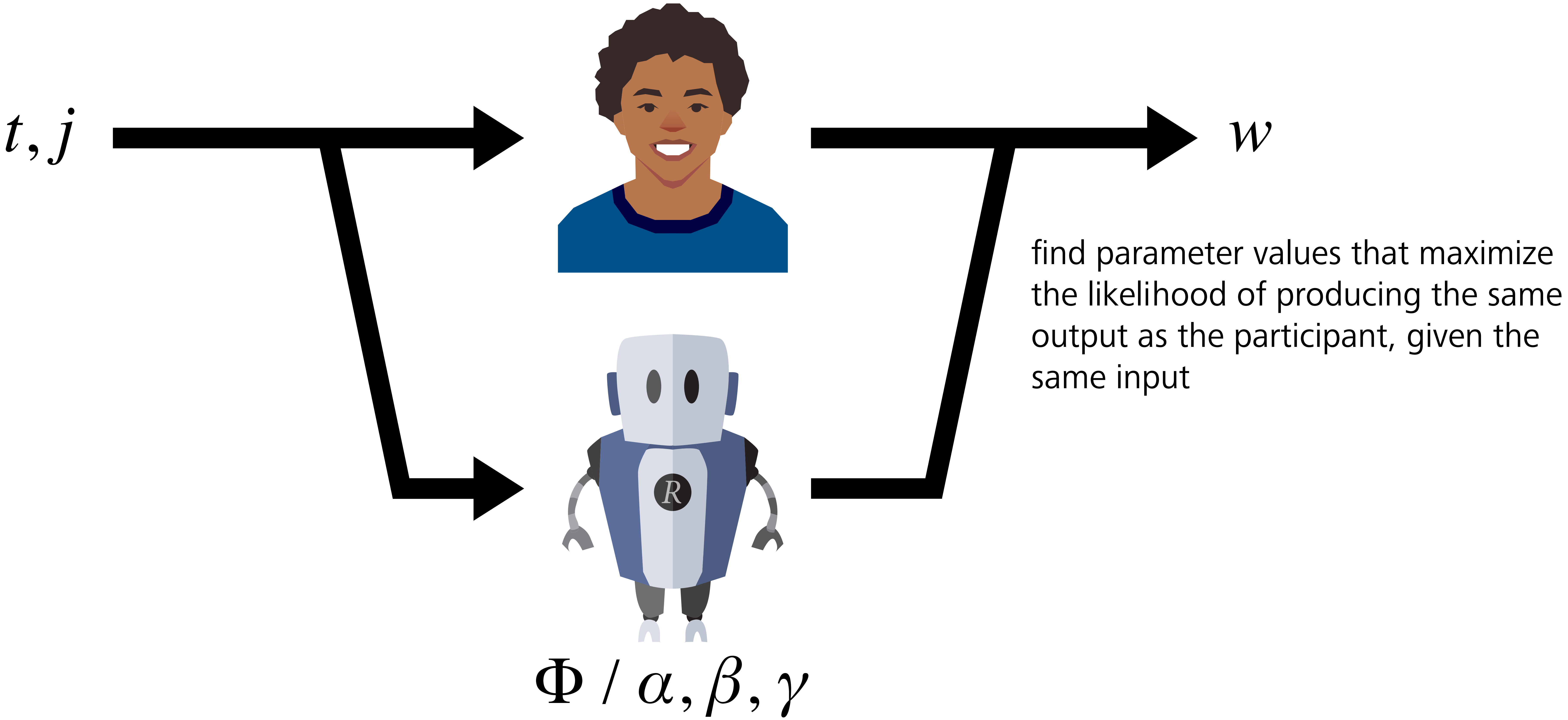

Dutch
English
German
Greek
Italian
Polish
Spanish
Swahili
1 2 3 4 5
5-letter words
1 2 3 4 5 6
6-letter words
1 2 3 4 5 6 7
7-letter words
1 2 3 4 5 6 7 8
8-letter words
1 2 3 4 5 6 7 8 9
9-letter words

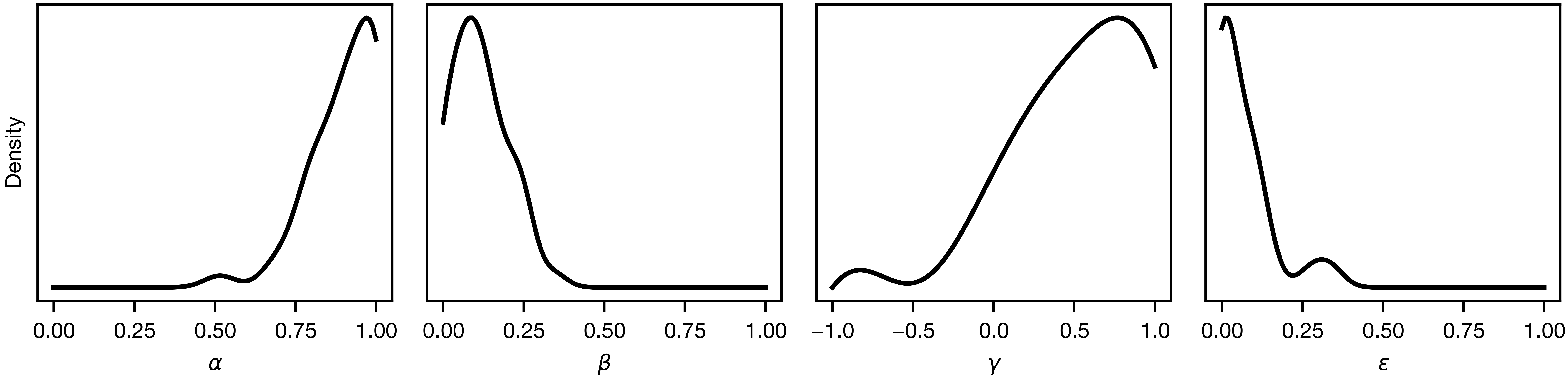

Dutch
English
German
Greek
Italian
Polish
Spanish
Swahili
1 2 3 4 5
5-letter words
1 2 3 4 5 6
6-letter words
1 2 3 4 5 6 7
7-letter words
1 2 3 4 5 6 7 8
8-letter words
1 2 3 4 5 6 7 8 9
9-letter words

	left-heavy lexicon	right-heavy lexicon	
	SNIDABS	SBADINS	
	SPODABS	SBADOPS	
	STUGABS	SBAGUTS	
	SVYGABS	SBAGYVS	
	STIKEBS	SBEKITS	
	SNOKEBS	SBEKONS	
	SVUMEBS	SBEMUVS	
	SPYMEBS	SBEMYPS	

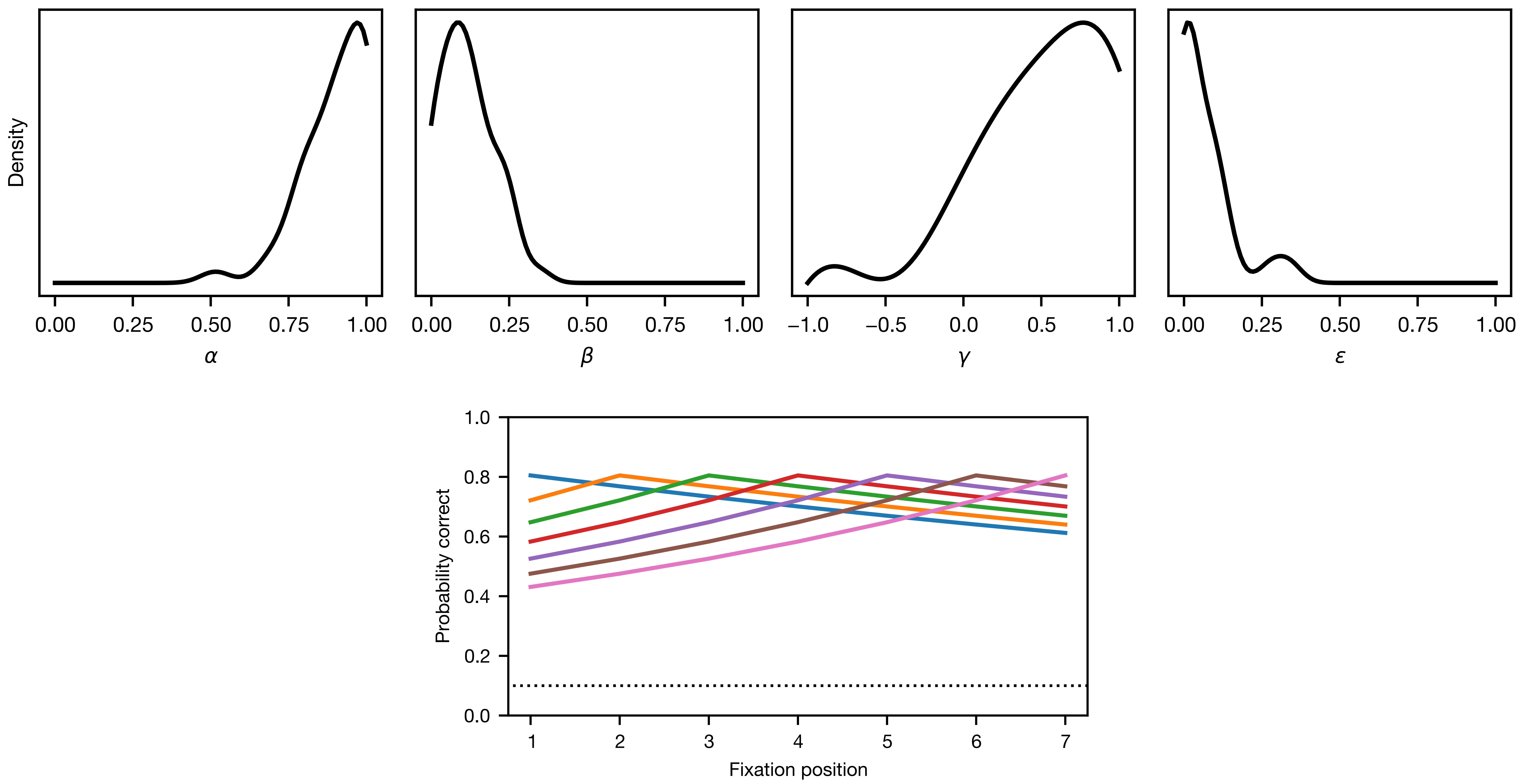


Left-heavy lexicon
Right-heavy lexicon
Probability of correct response
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
Fixation position
Fixation position

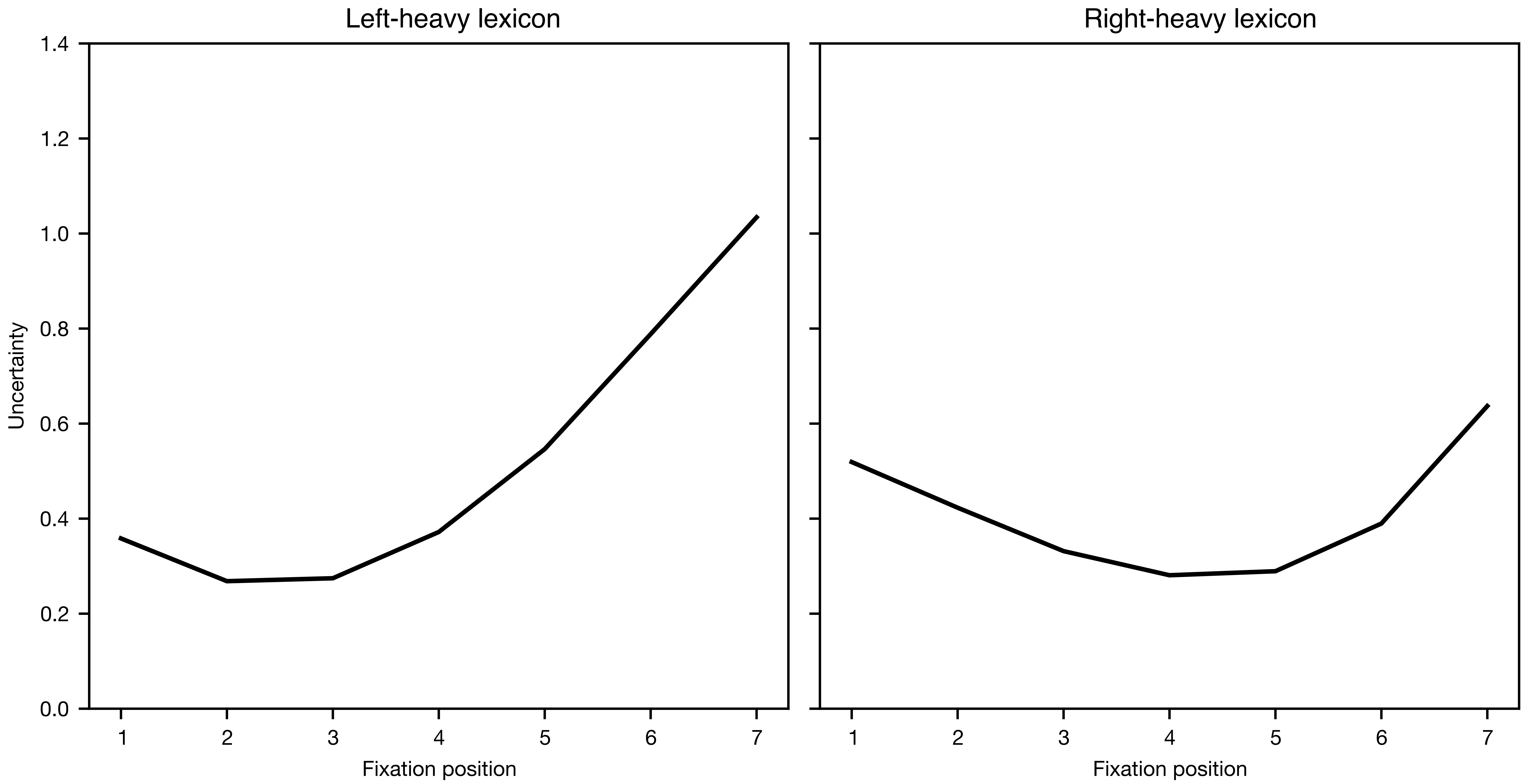

t, j
w
find parameter values that maximize the likelihood of producing the same output as the participant, given the same input
R
Φ / α, β, γ



Density
0.00
0.25
0.50
0.75
1.00
α
0.00
0.25
0.50
0.75
1.00
β
−1.0
−0.5
0.0
0.5
1.0
γ
0.00
0.25
0.50
0.75
1.00
ε



Density
0.00
0.25
0.50
0.75
1.00
α
β
−1.0
−0.5
0.0
0.5
1.0
γ
ε
Probability correct
0.2
0.4
0.6
0.8
1 2 3 4 5 6 7
Fixation position


Left-heavy lexicon
Right-heavy lexicon
Uncertainty
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1
2
3
4
5
6
7
Fixation position
Fixation position

Simplicity and informativeness in semantic category systems


Jon Carr


Kenny Smith

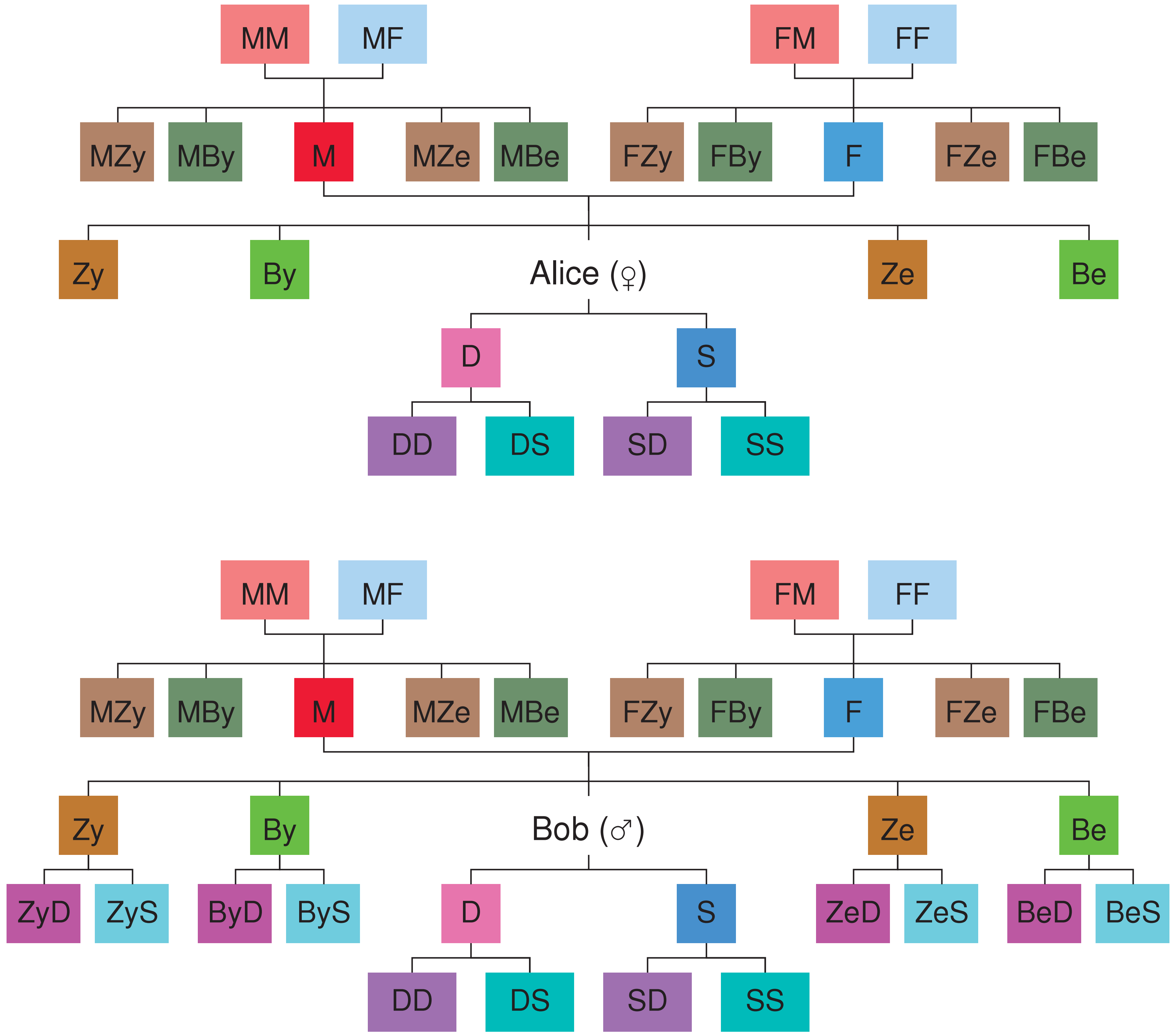

Jenny Culbertson


Simon Kirby

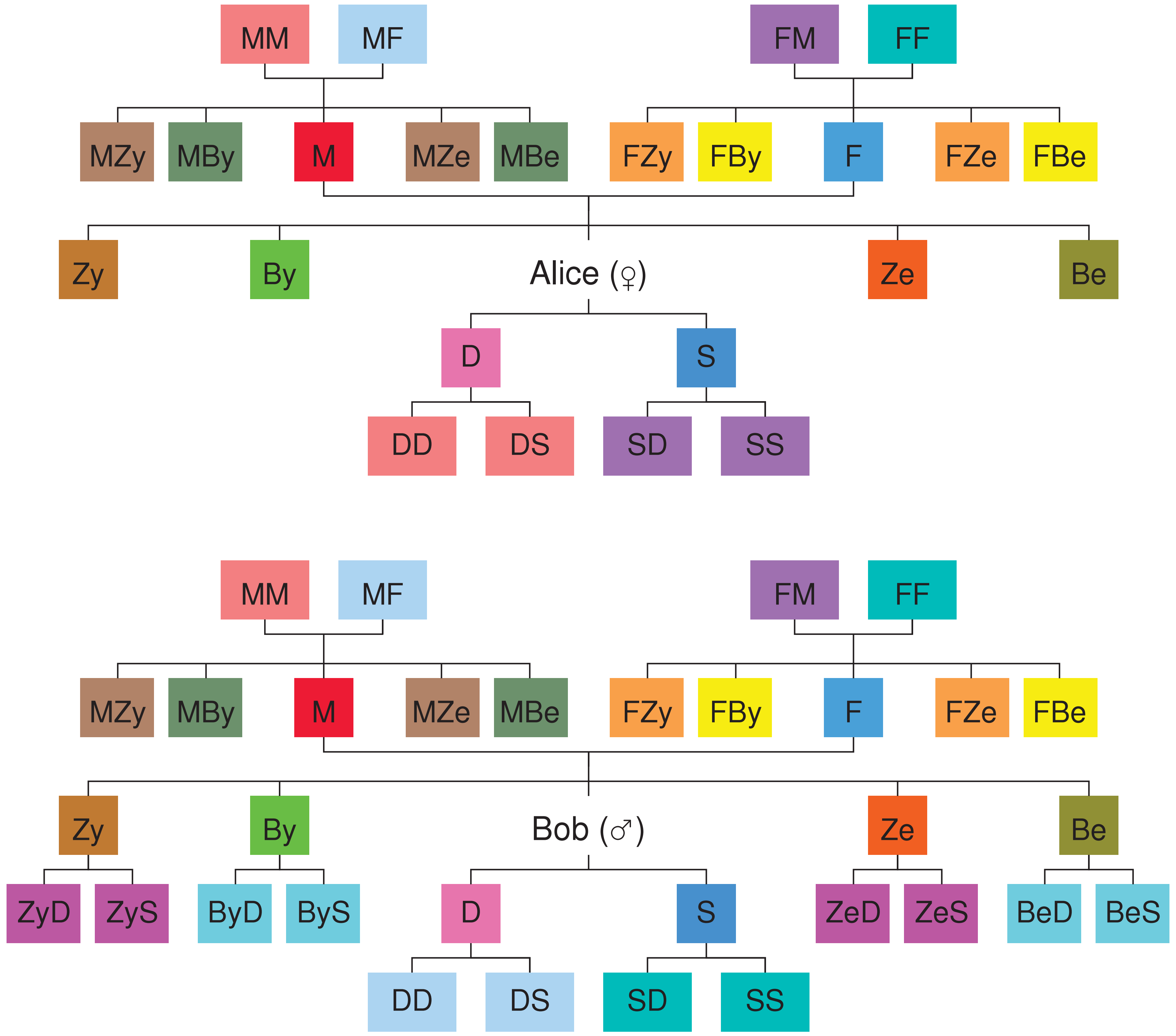
Kinship terms are simple and informative

Kemp & Regier (2012)

English



Northern Paiute

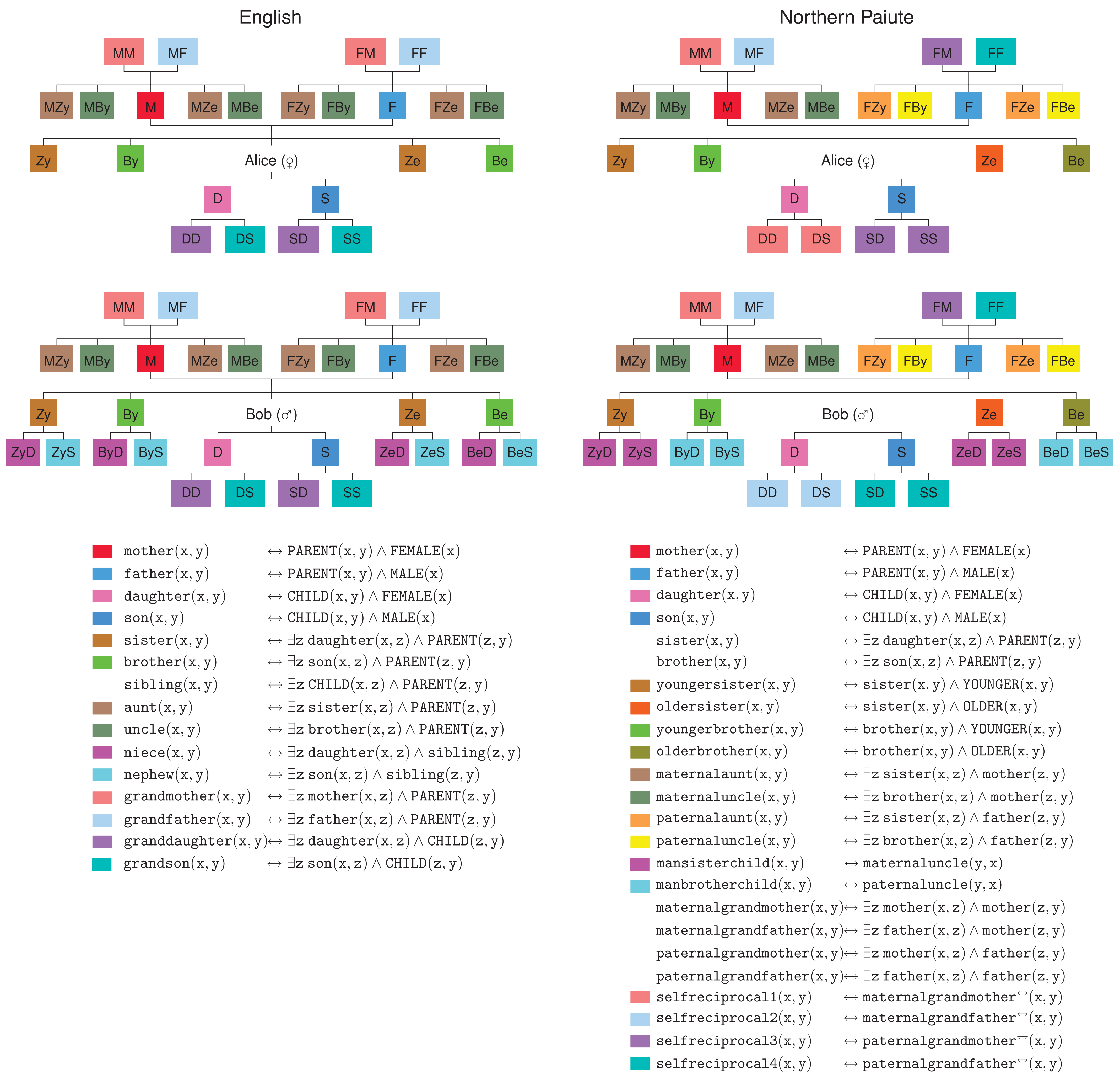


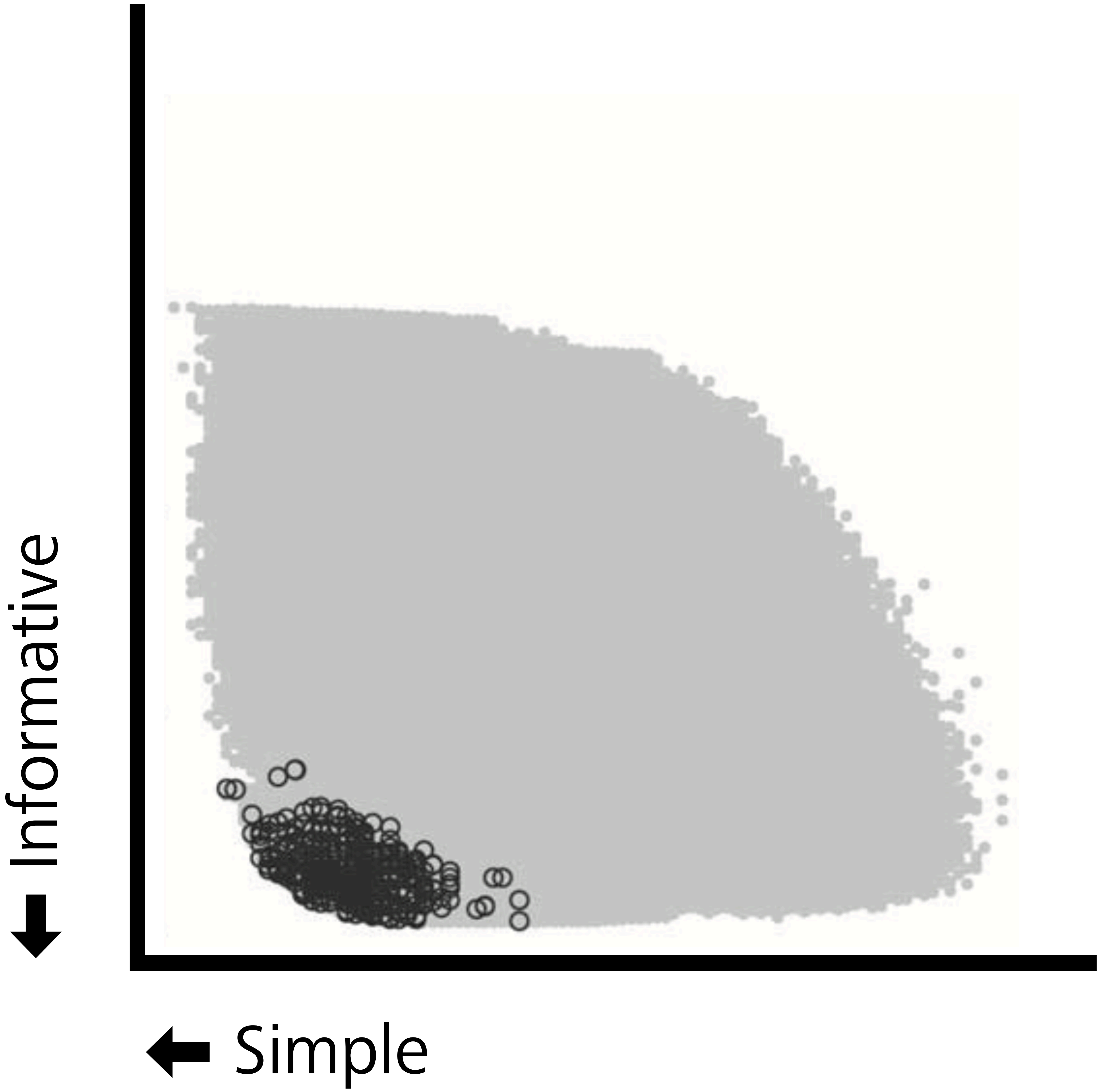
$\mathtt{mother}(x,y)$	$\leftrightarrow \mathtt{PARENT}(x,y) \land \mathtt{FEMALE}(x)$
$\mathtt{father}(x,y)$	$\leftrightarrow \mathtt{PARENT}(x,y) \land \mathtt{MALE}(x)$
$\mathtt{daughter}(x,y)$	$\leftrightarrow \mathtt{CHILD}(x,y) \land \mathtt{FEMALE}(x)$
$\mathtt{son}(x,y)$	$\leftrightarrow \mathtt{CHILD}(x,y) \land \mathtt{MALE}(x)$
$\mathtt{sister}(x,y)$	$\leftrightarrow \exists z\, \mathtt{daughter}(x,z) \land \mathtt{PARENT}(z,y)$

$\mathtt{mother}(x,y)$	$\leftrightarrow \mathtt{PARENT}(x,y) \land \mathtt{FEMALE}(x)$
$\mathtt{father}(x,y)$	$\leftrightarrow \mathtt{PARENT}(x,y) \land \mathtt{MALE}(x)$
$\mathtt{daughter}(x,y)$	$\leftrightarrow \mathtt{CHILD}(x,y) \land \mathtt{FEMALE}(x)$
$\mathtt{son}(x,y)$	$\leftrightarrow \mathtt{CHILD}(x,y) \land \mathtt{MALE}(x)$
$\mathtt{sister}(x,y)$	$\leftrightarrow \exists z\, \mathtt{daughter}(x,z) \land \mathtt{PARENT}(z,y)$

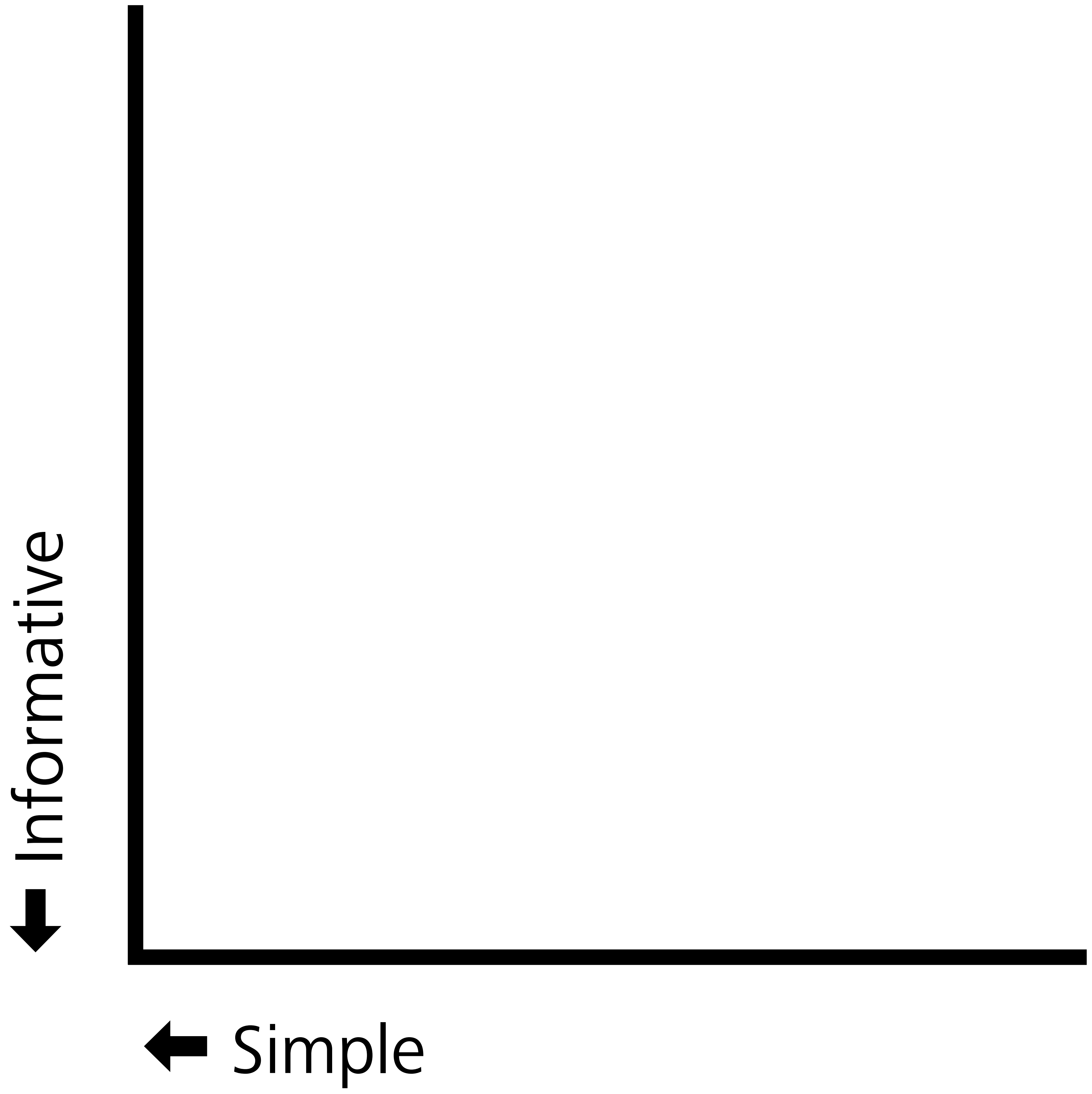
Kinship terms are simple and informative

Kemp & Regier (2012)

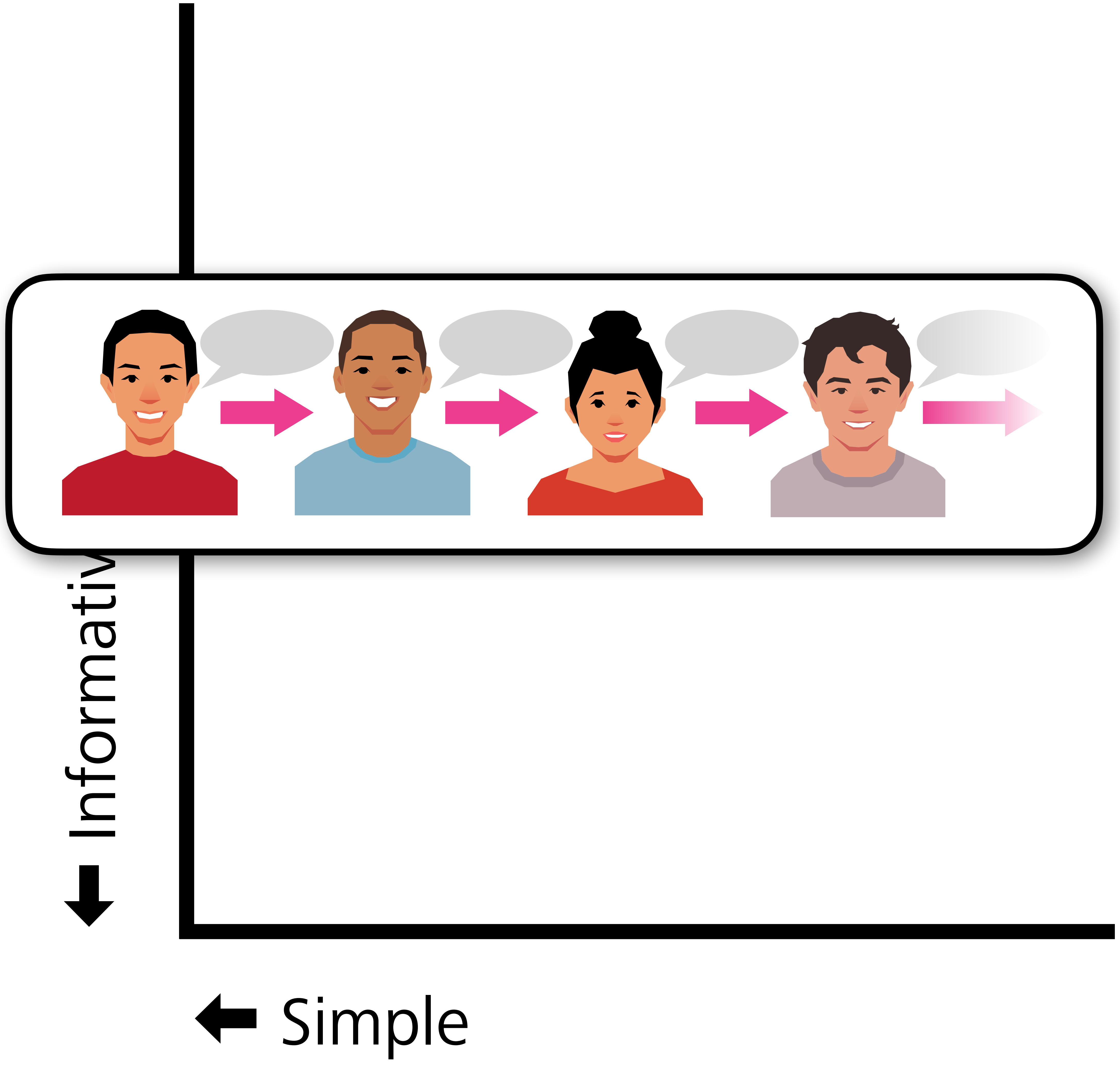




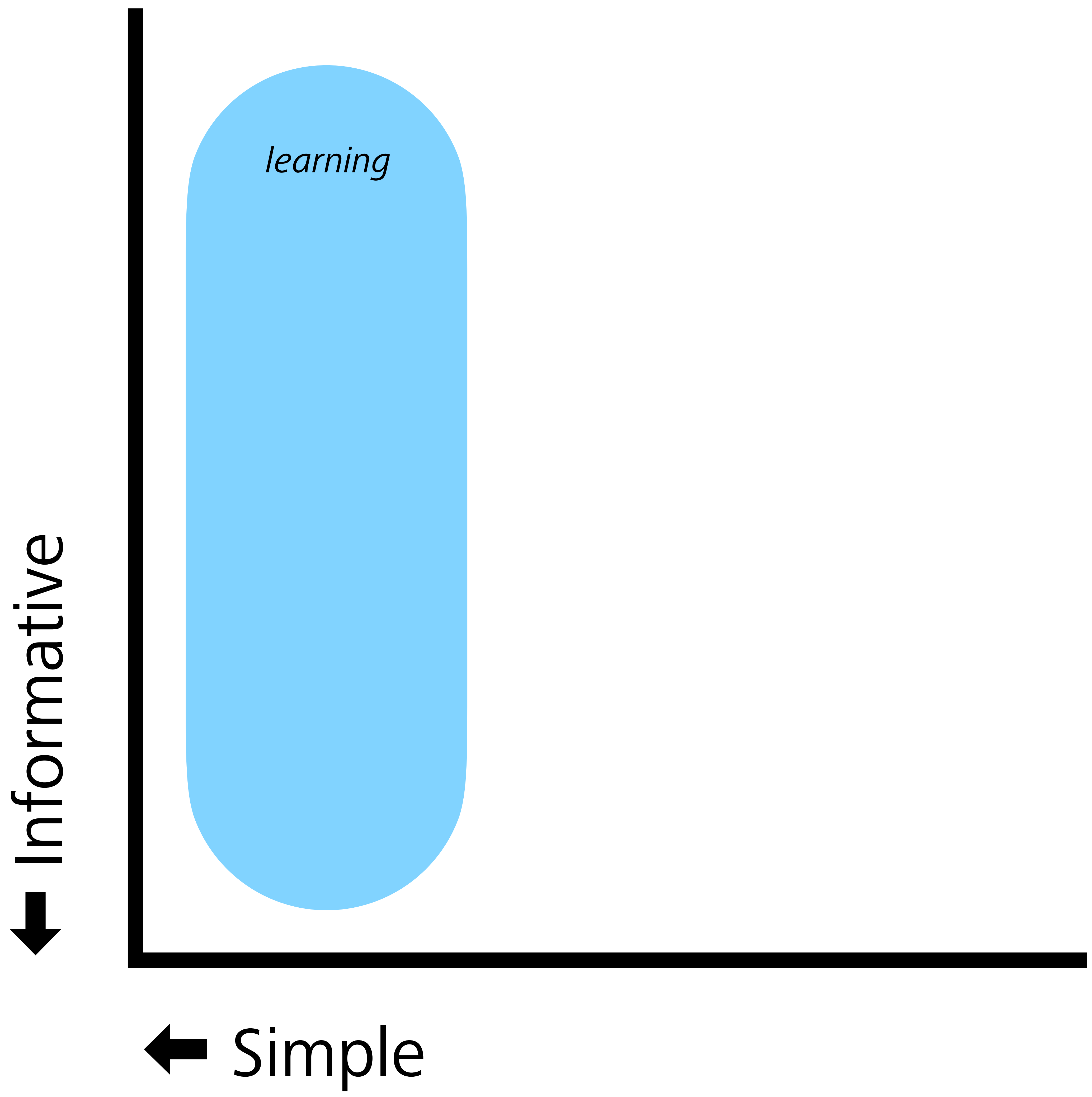
Learning & interaction: Pressures for simplicity and informativeness



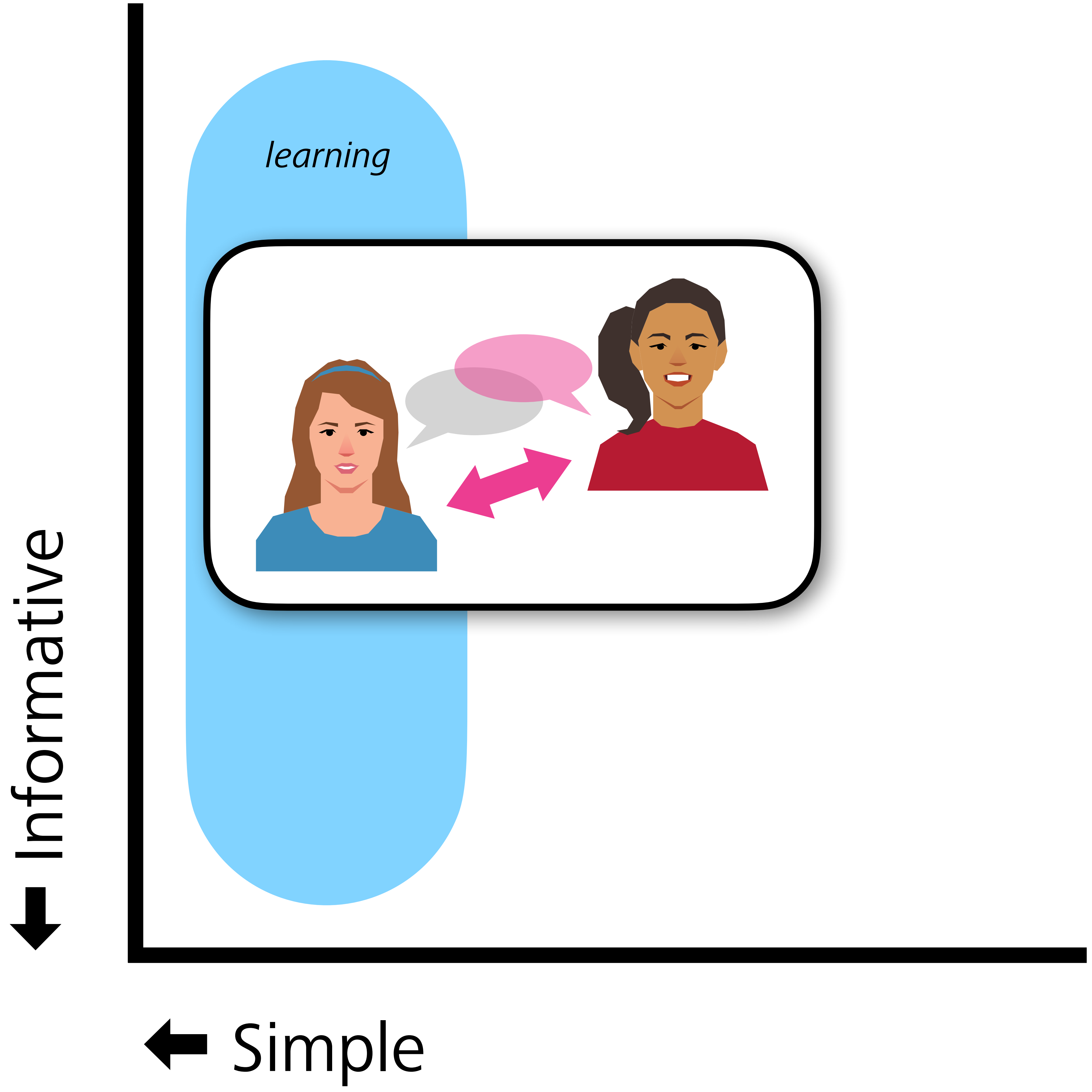
Learning & interaction: Pressures for simplicity and informativeness



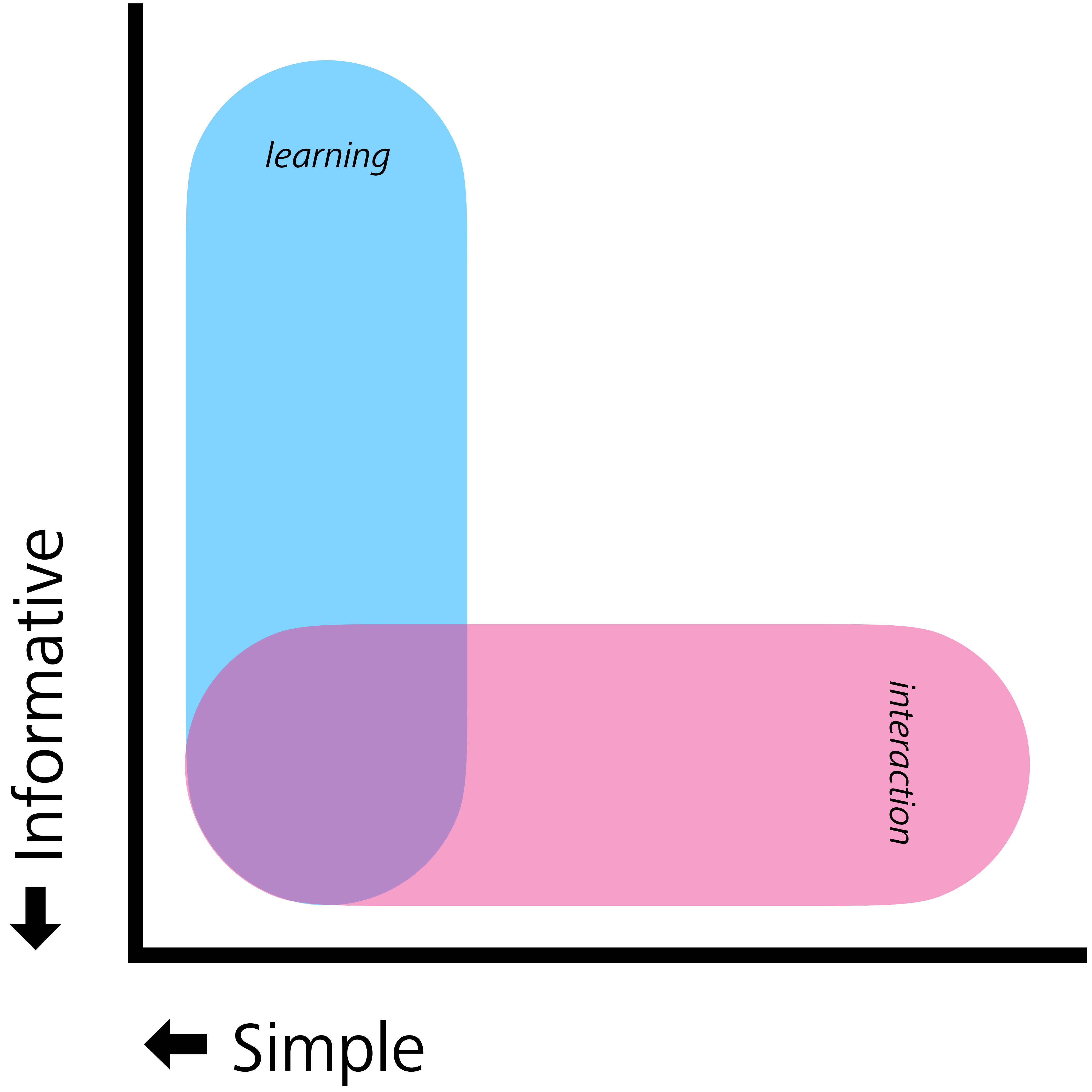
Learning & interaction: Pressures for simplicity and informativeness



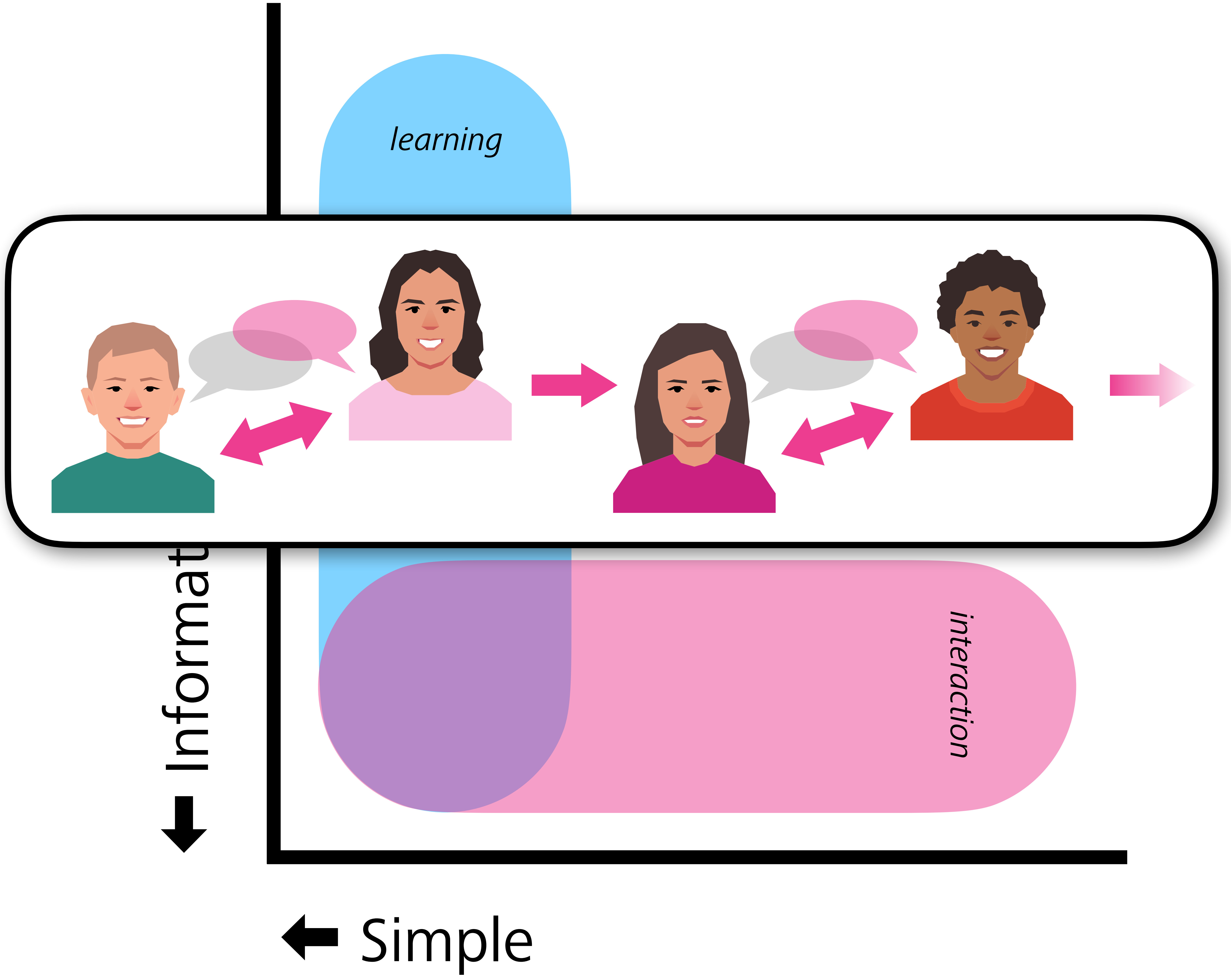
Learning & interaction: Pressures for simplicity and informativeness



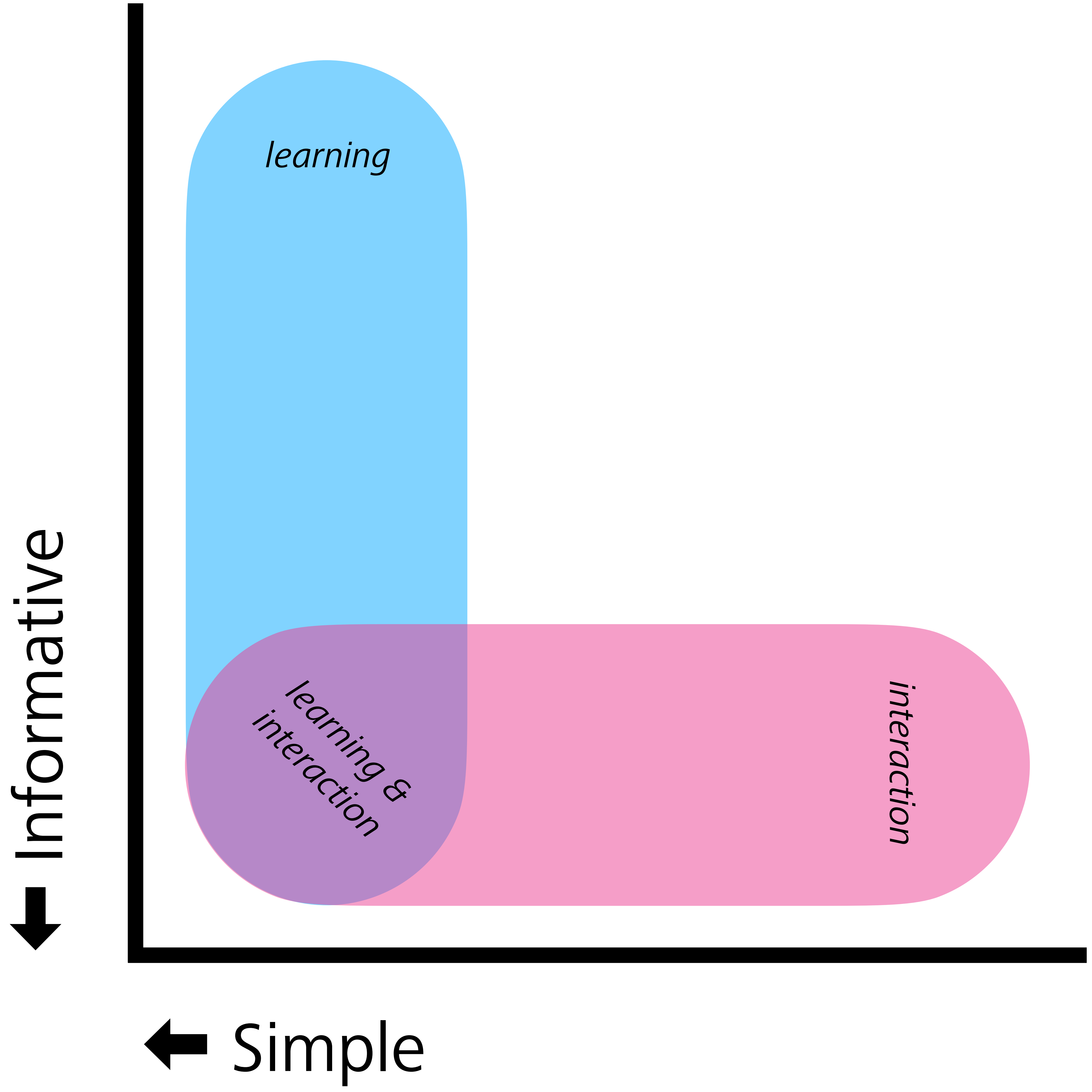
Learning & interaction: Pressures for simplicity and informativeness



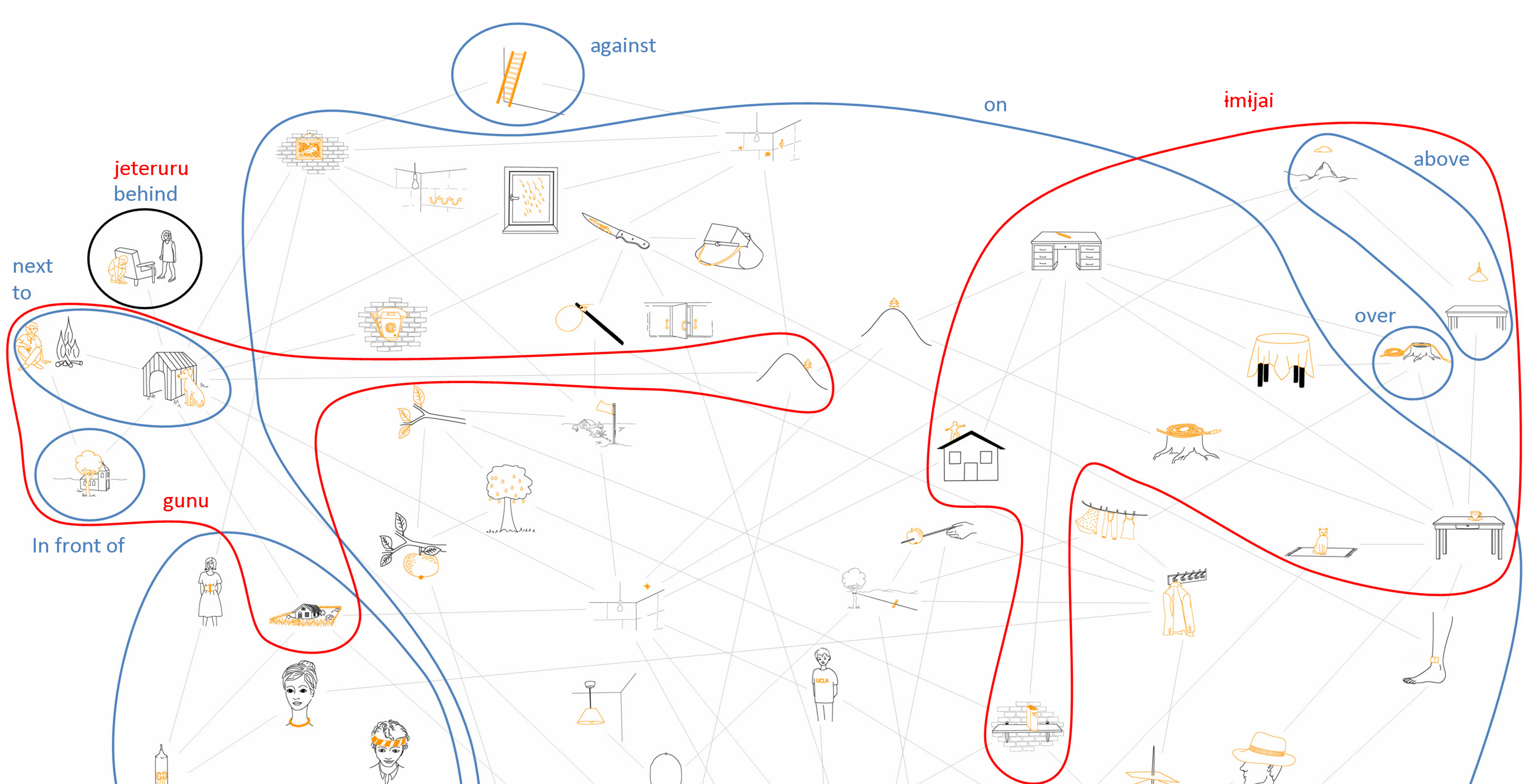
Learning & interaction: Pressures for simplicity and informativeness



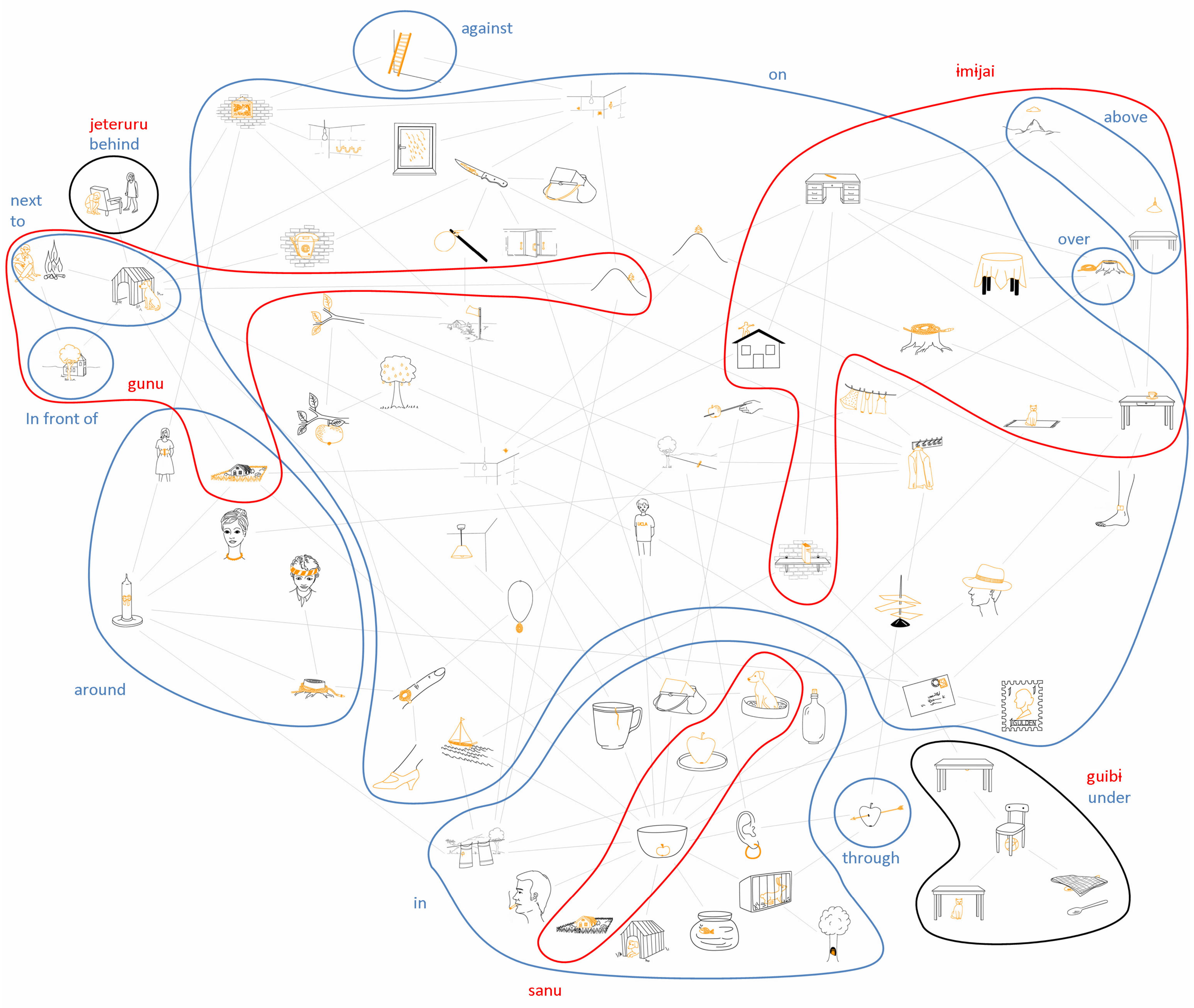
Learning & interaction: Pressures for simplicity and informativeness

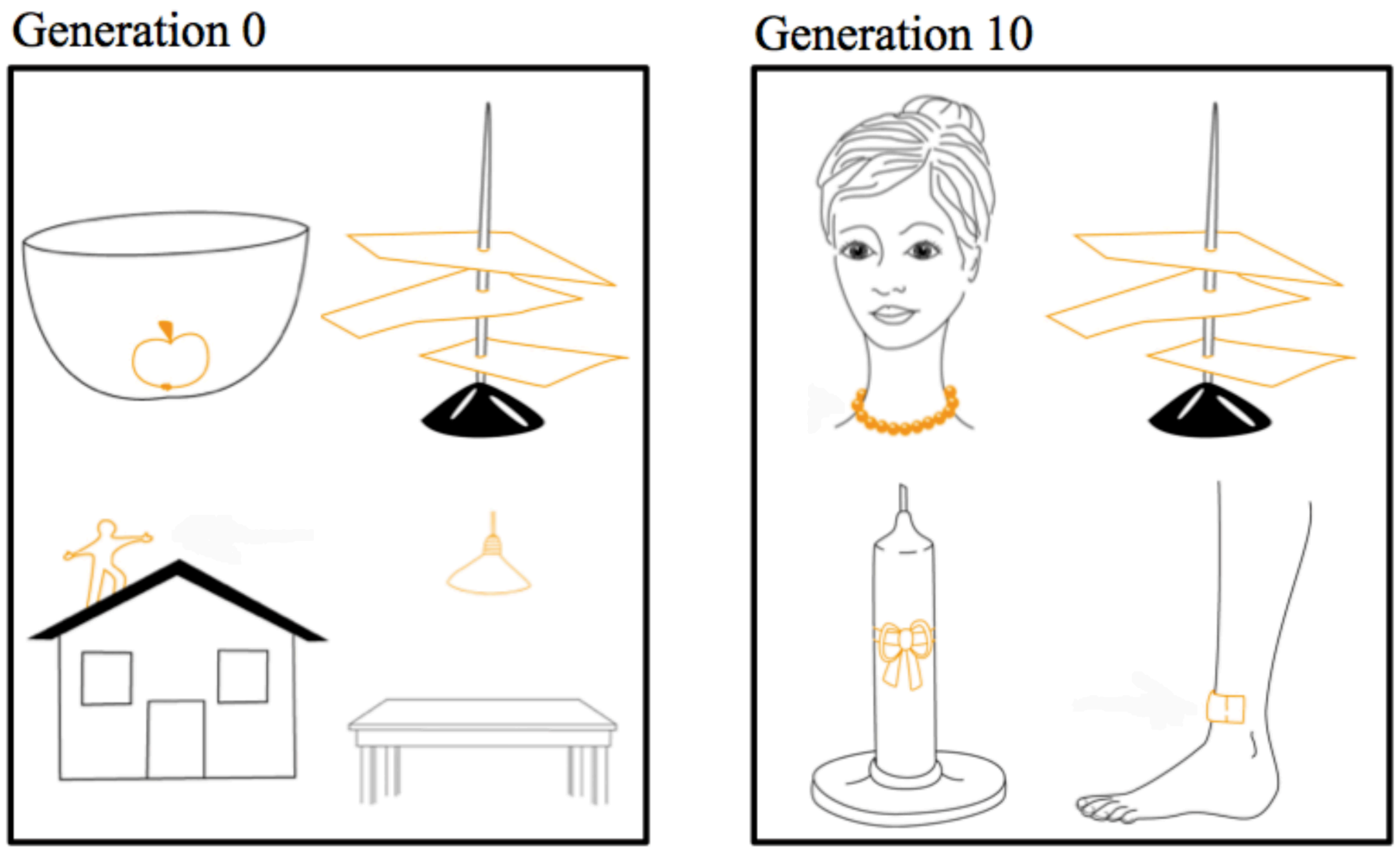


Can iterated learning give rise to informative categories?



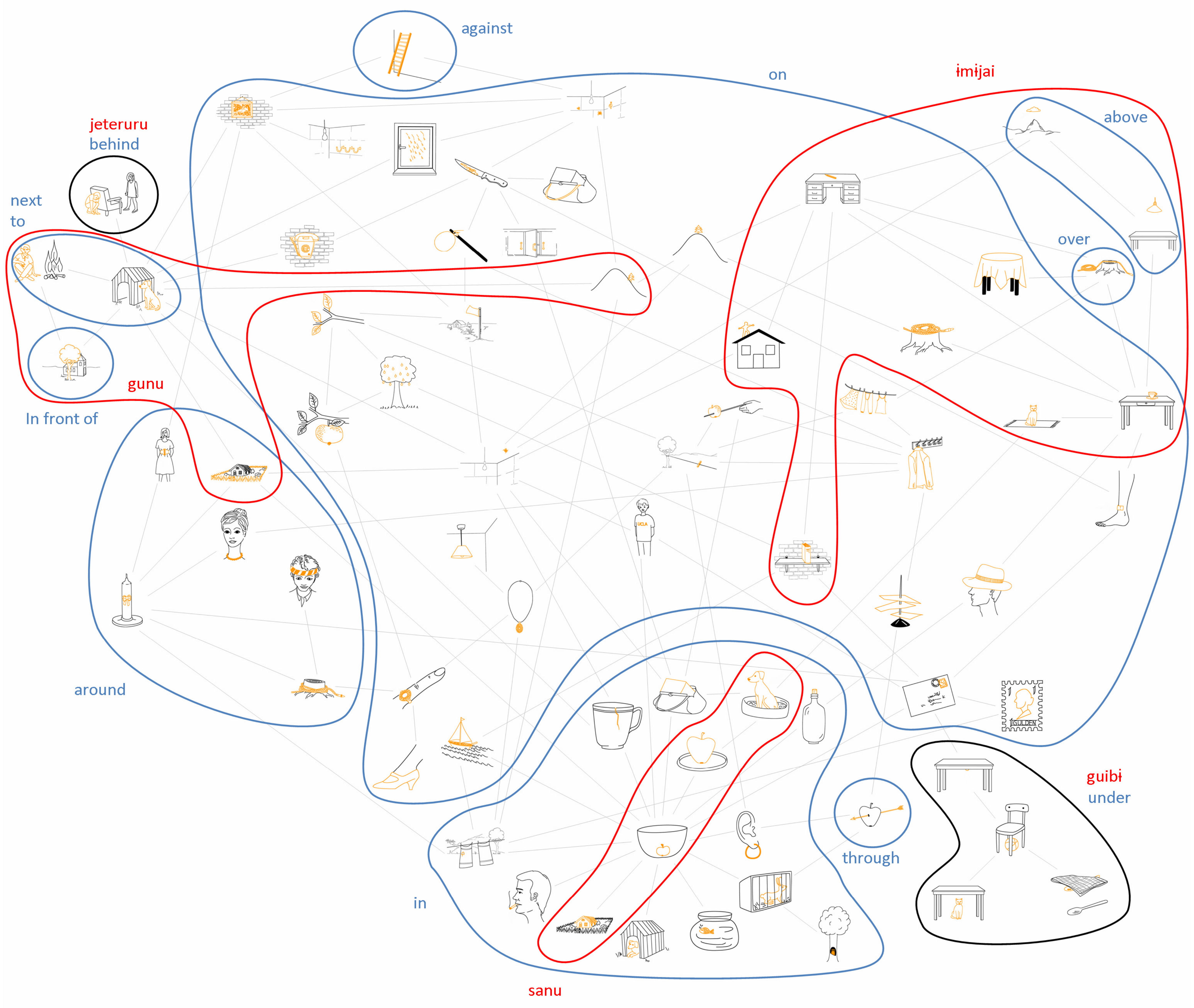
Can iterated learning give rise to informative categories?

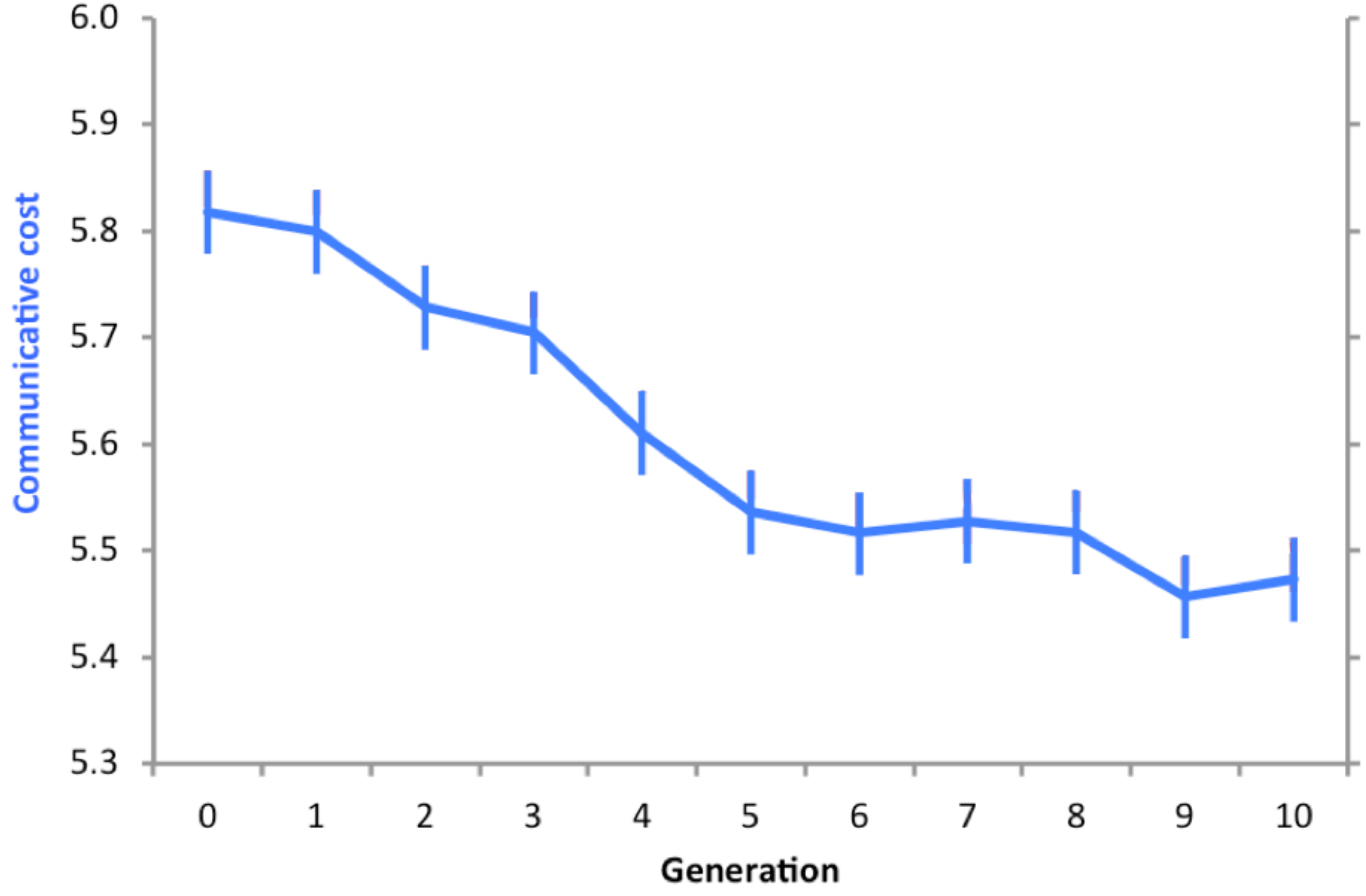




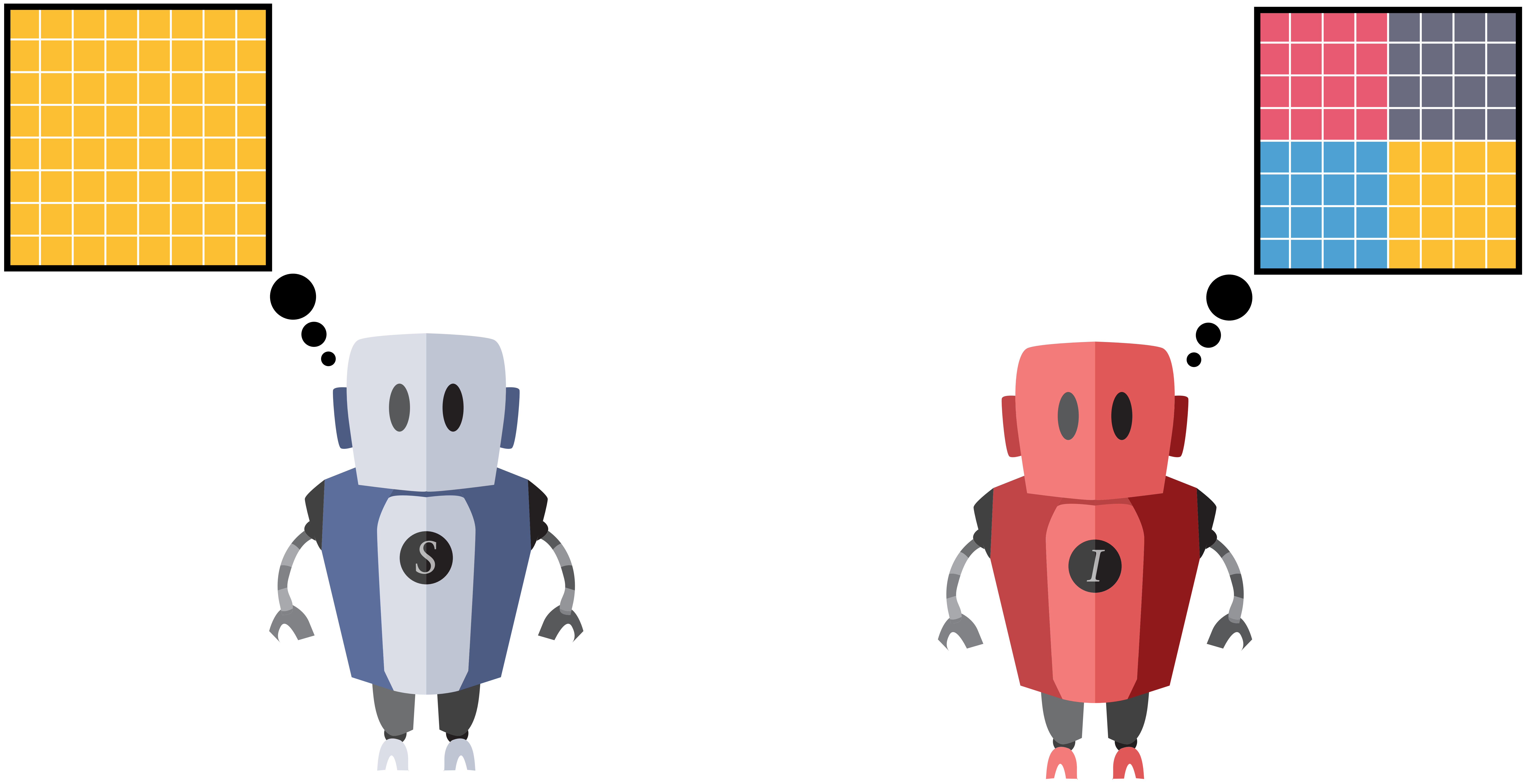
Carstensen, Xu, Smith, Regier (2015)

Can iterated learning give rise to informative categories?



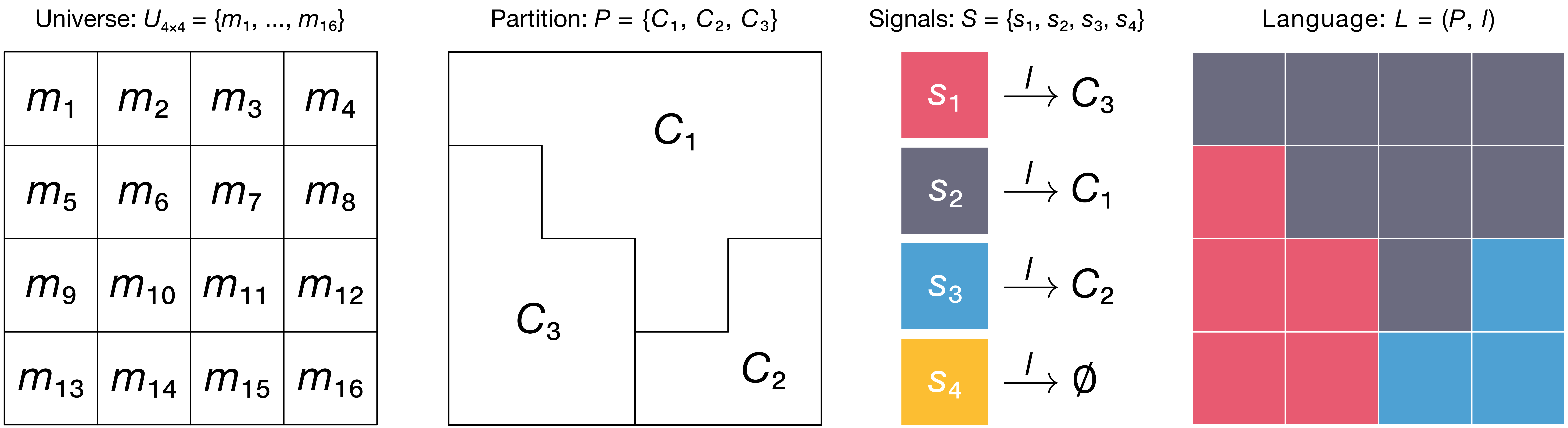


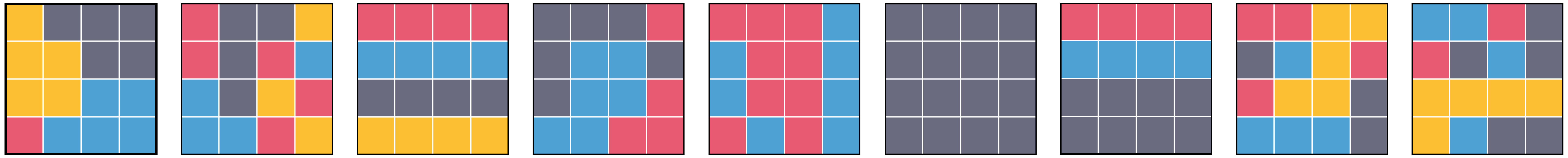
Carstensen, Xu, Smith, Regier (2015)

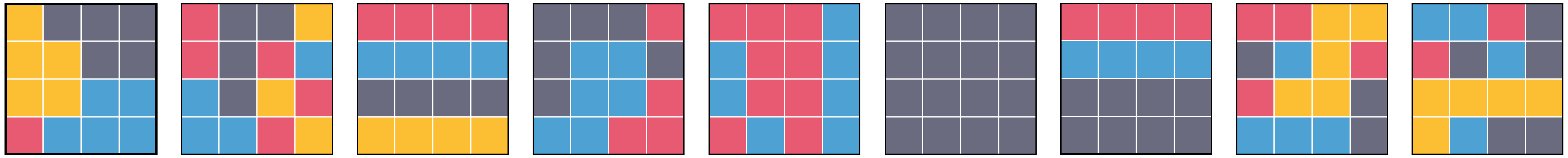


Simplicity bias

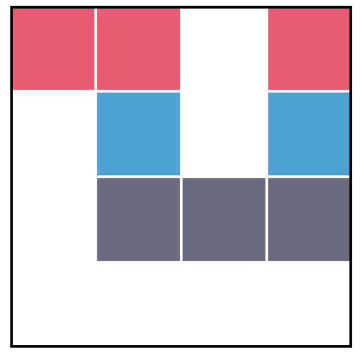
Informativeness bias

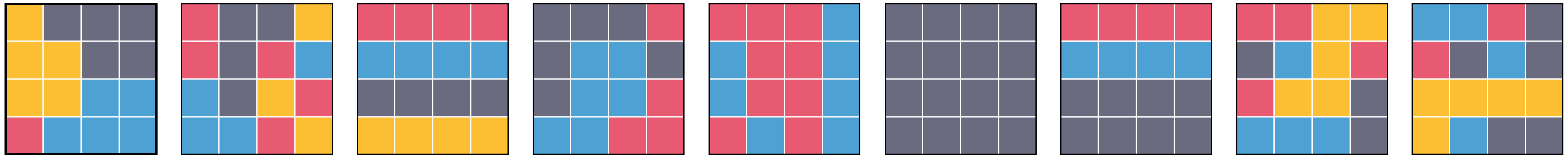

Universise: $U_{4\times4} = \{m_1, ..., m_{16}\}$
m_1 m_2 m_3 m_4
m_5 m_6 m_7 m_8
m_9 m_{10} m_{11} m_{12}
m_{13} m_{14} m_{15} m_{16}
Partition: $P = \{C_1, C_2, C_3\}$
C_1
C_3
C_2
Signals: $S = \{s_1, s_2, s_3, s_4\}$
$s_1 \xrightarrow{l} C_3$
$s_2 \xrightarrow{l} C_1$
$s_3 \xrightarrow{l} C_2$
$s_4 \xrightarrow{l} \emptyset$
Language: $L = (P, l)$

$\mathcal{L} = \{$  $\cdots\}$

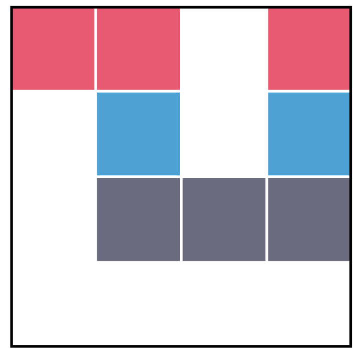
$\mathcal{L} = \{$  $\cdots\}$

$$D = [\langle m_1, s_1\rangle, \langle m_2, s_2\rangle, \langle m_3, s_3\rangle, ..., \langle m_n, s_n\rangle]$$

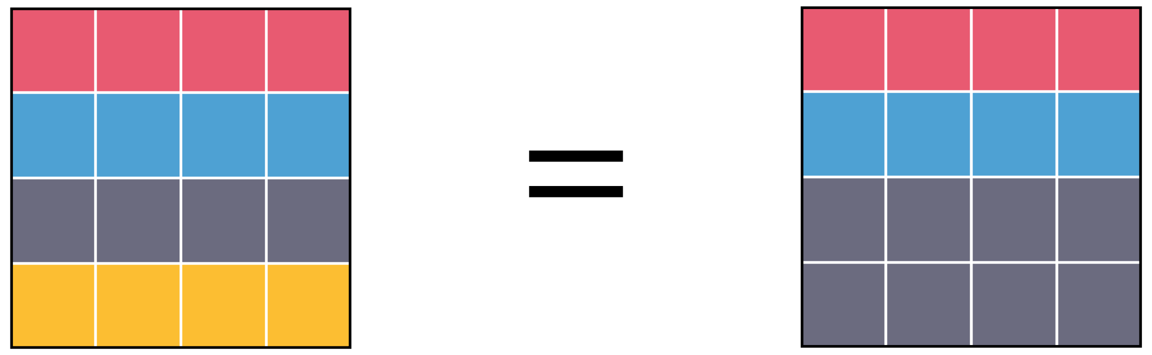


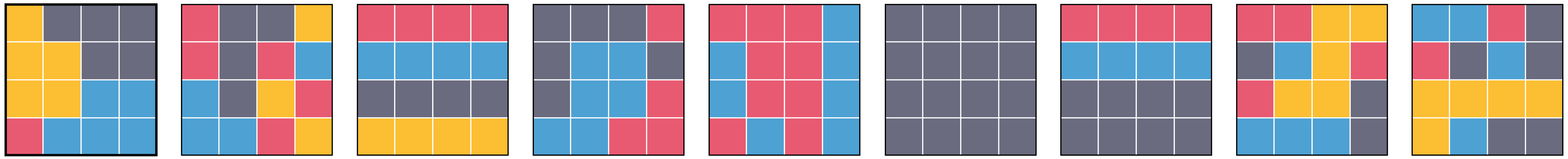
$\mathcal{L} = \{$  $\cdots\}$

$$D = [\langle m_1, s_1\rangle, \langle m_2, s_2\rangle, \langle m_3, s_3\rangle, \ldots, \langle m_n, s_n\rangle]$$

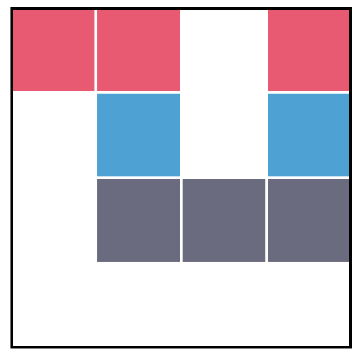


$$\text{likelihood}(D|L) = \prod_{\langle m,s\rangle} P(s|L,m)$$

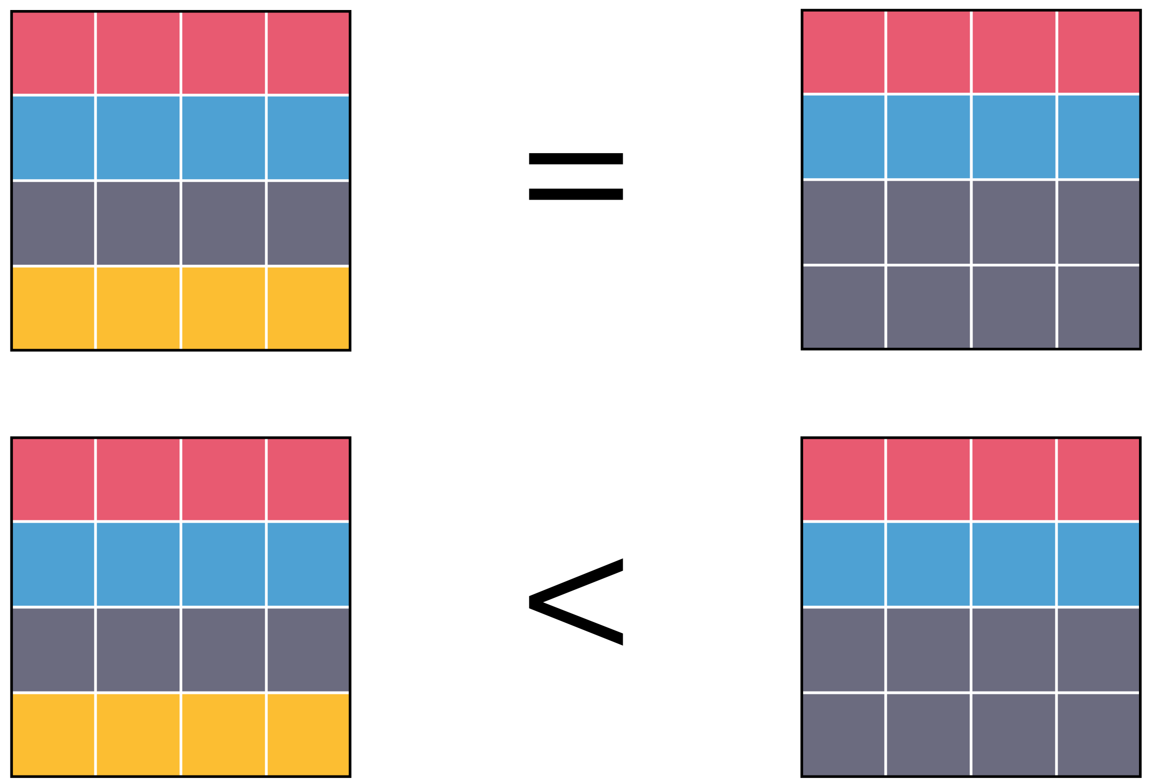


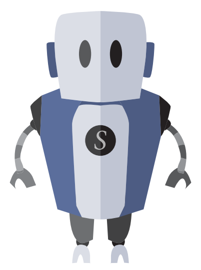
$$\mathcal{L} = \{$$  $$\cdots\}$$

$$D = [\langle m_1, s_1\rangle, \langle m_2, s_2\rangle, \langle m_3, s_3\rangle, \ldots, \langle m_n, s_n\rangle]$$

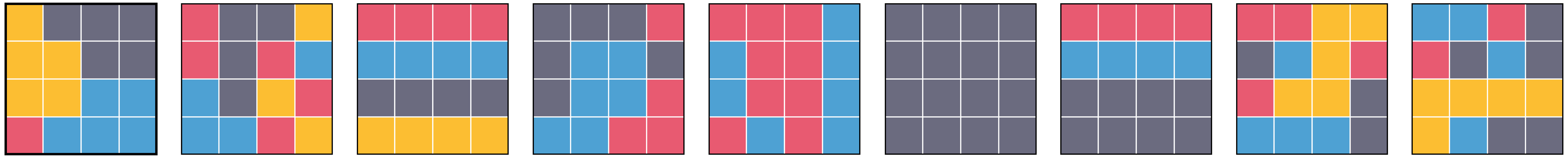


$$\text{likelihood}(D|L) = \prod_{\langle m,s\rangle} P(s|L, m)$$

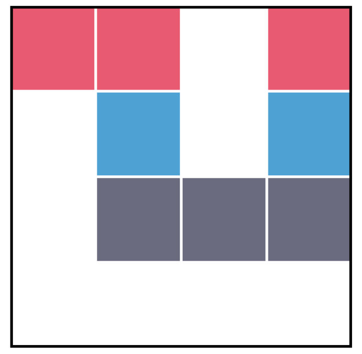




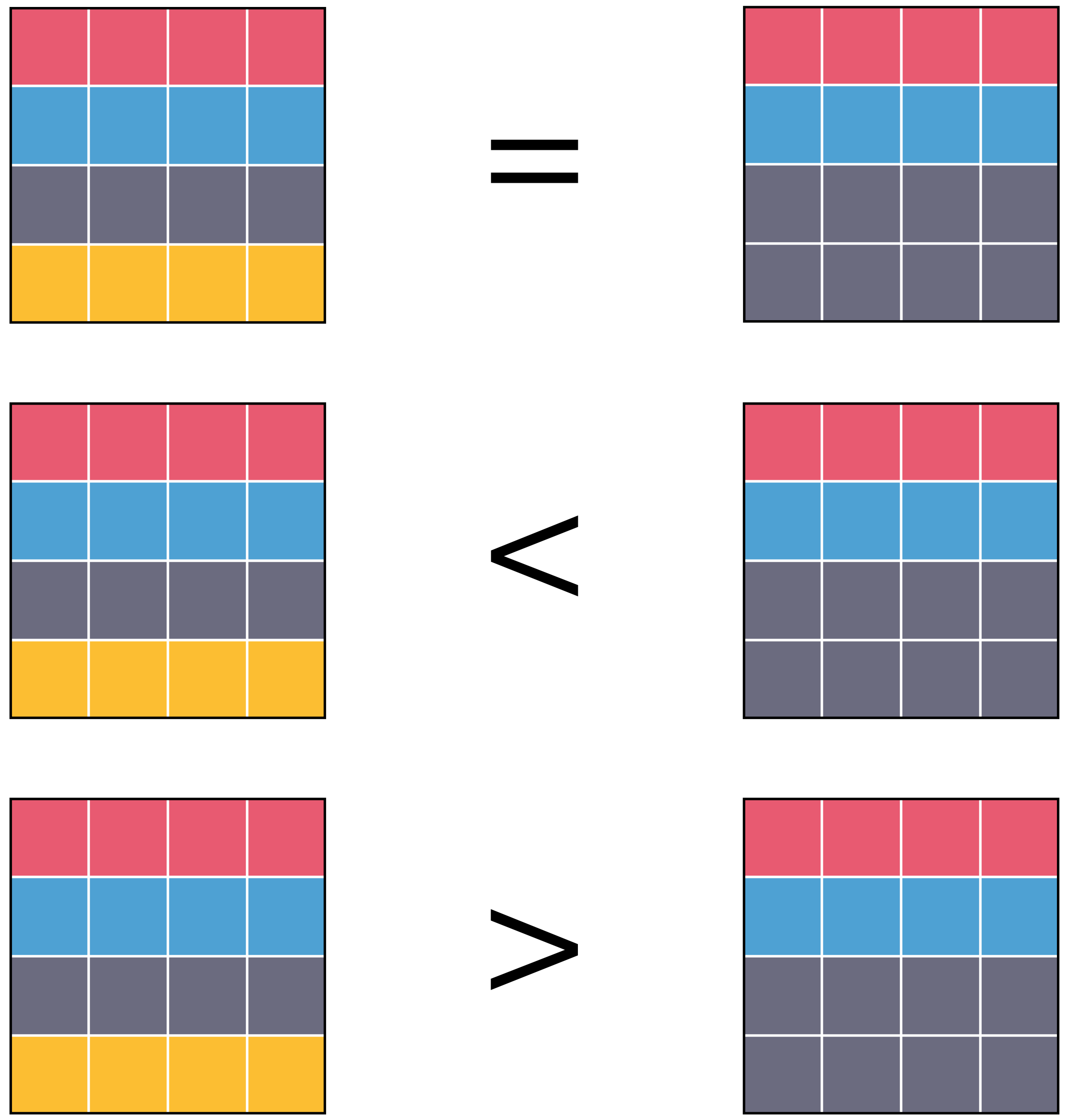
$$\text{prior}(L) \propto 2^{-\text{complexity}(L)}$$

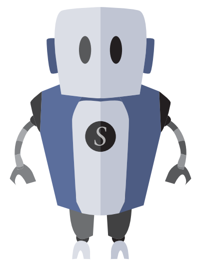
$\mathcal{L} = \{$  $\cdots\}$

$$D = [\langle m_1, s_1\rangle, \langle m_2, s_2\rangle, \langle m_3, s_3\rangle, \ldots, \langle m_n, s_n\rangle]$$

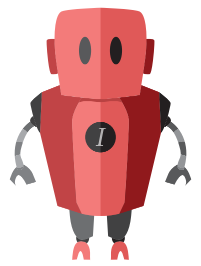


$$\text{likelihood}(D|L) = \prod_{\langle m,s\rangle} P(s|L,m)$$



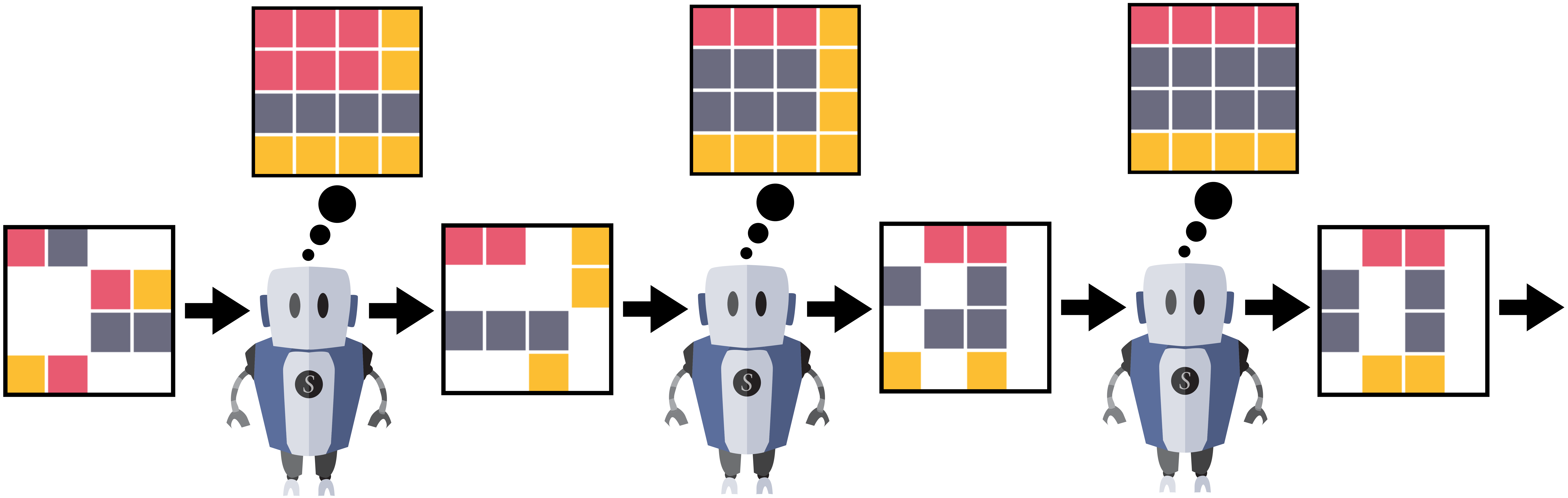


$$\text{prior}(L) \propto 2^{-\text{complexity}(L)}$$

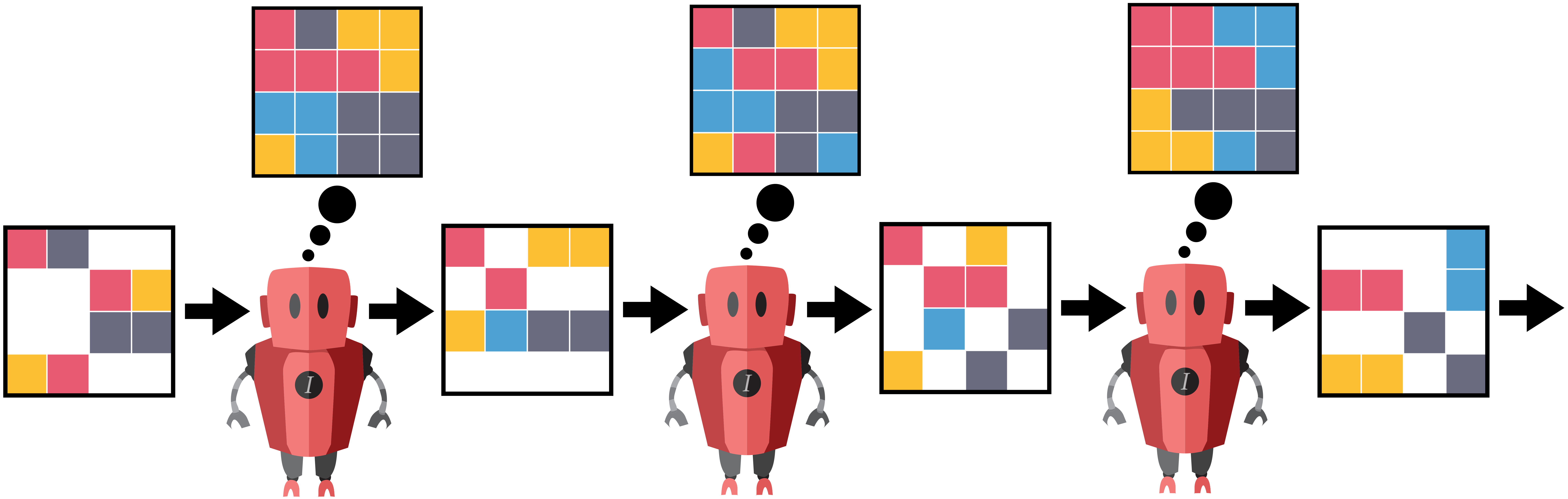


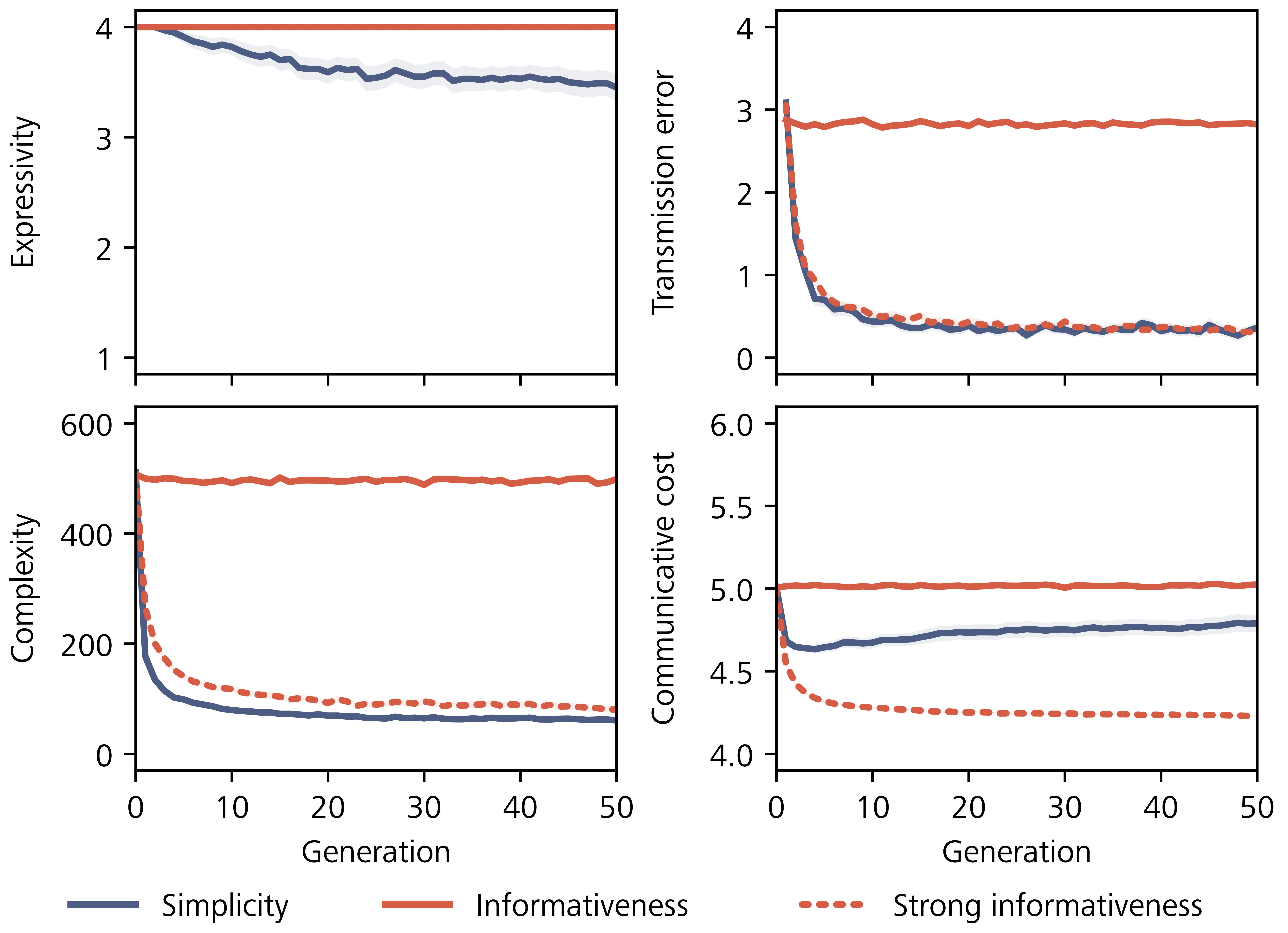
$$\text{prior}(L) \propto 2^{-\text{cost}(L)}$$

Bayesian iterated learning under a **simplicity prior**

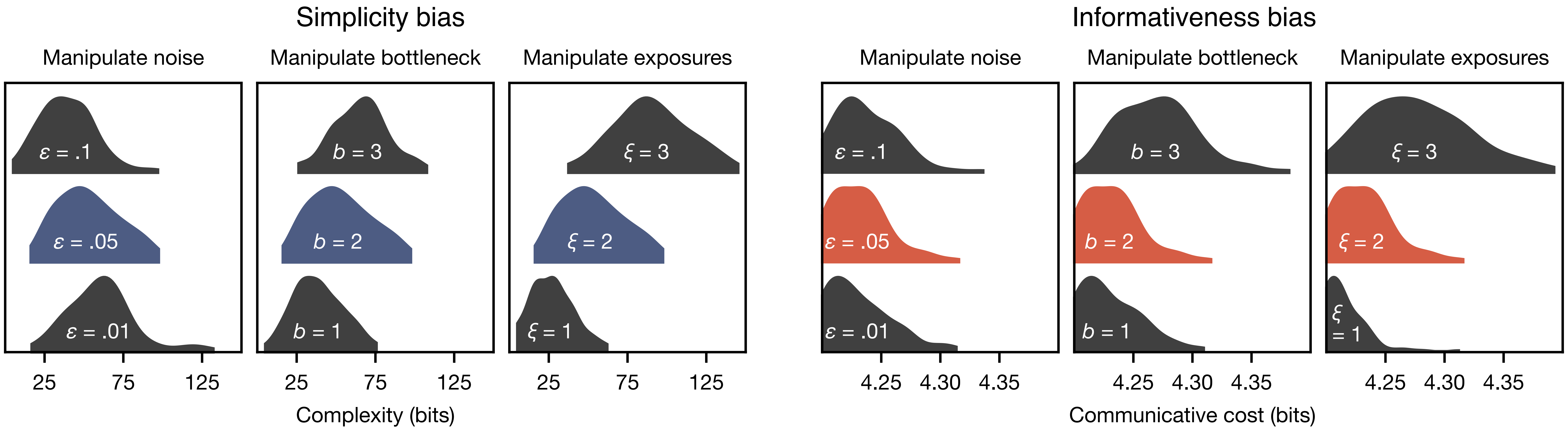


Bayesian iterated learning under an **informativeness prior**

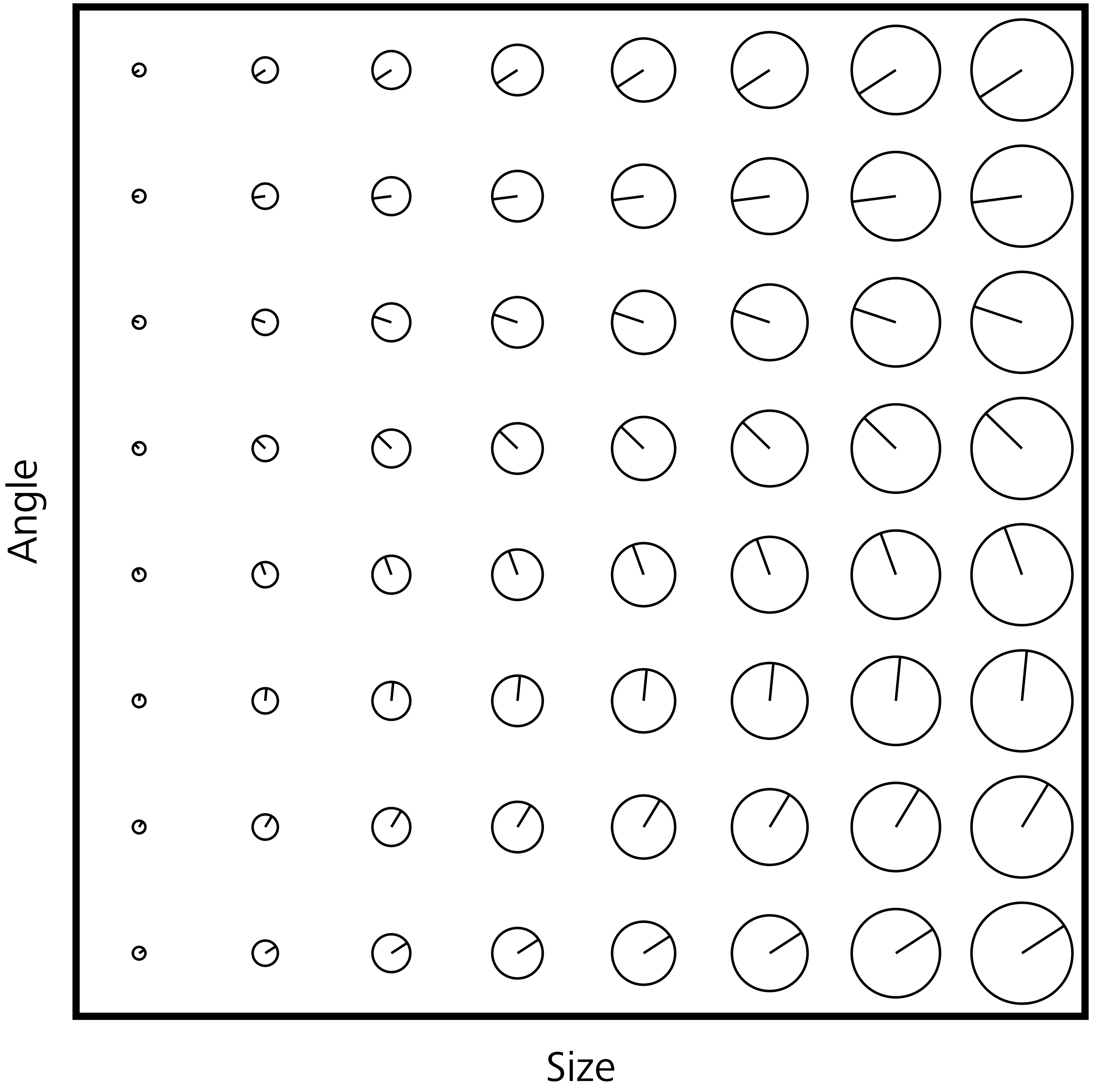


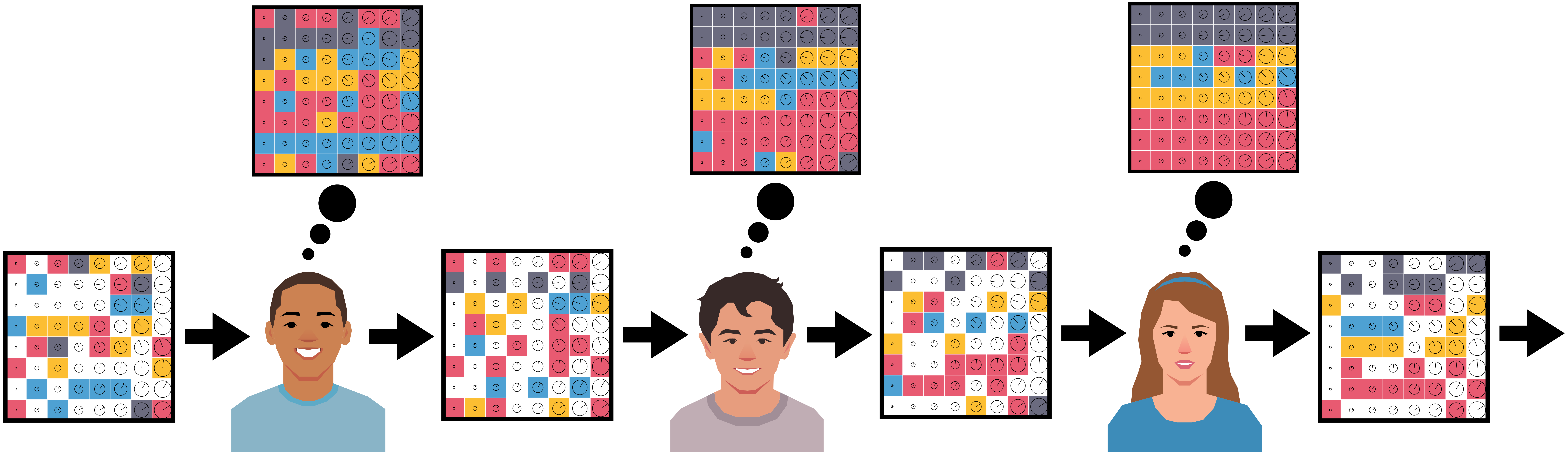


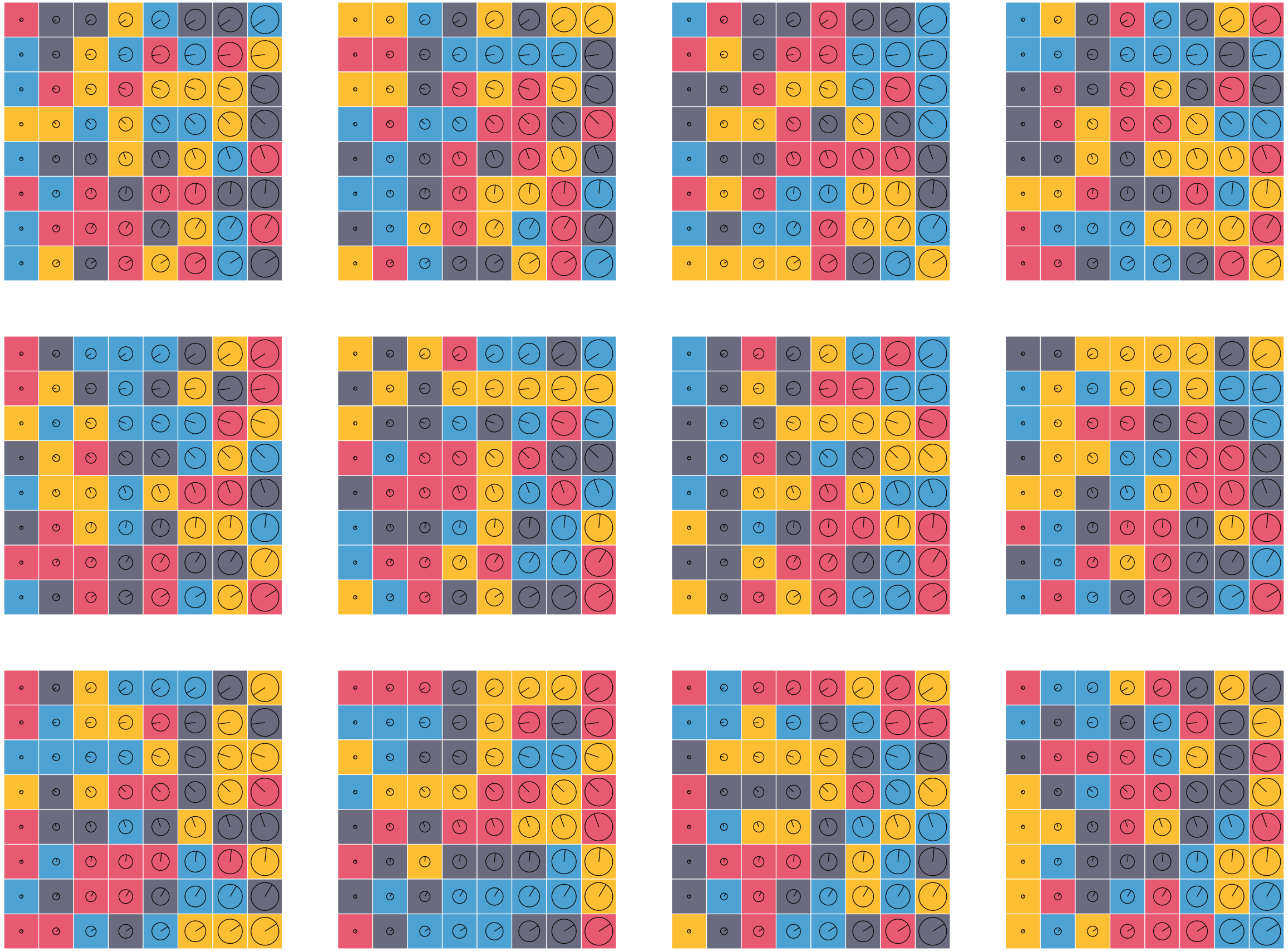
Expressivity
Transmission error
Complexity
Communicative cost
Generation
Generation
Simplicity
Informativeness
Strong informativeness

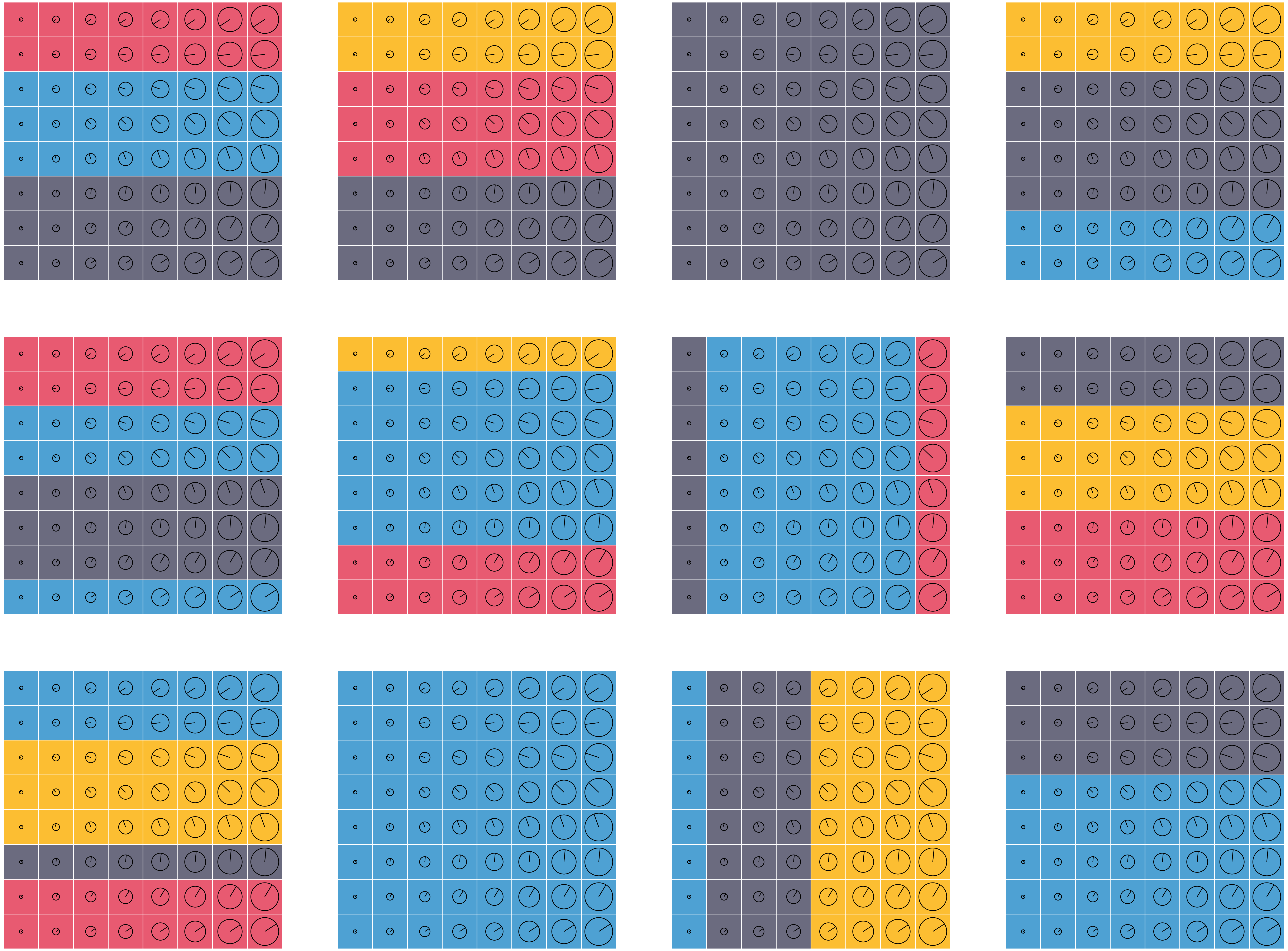


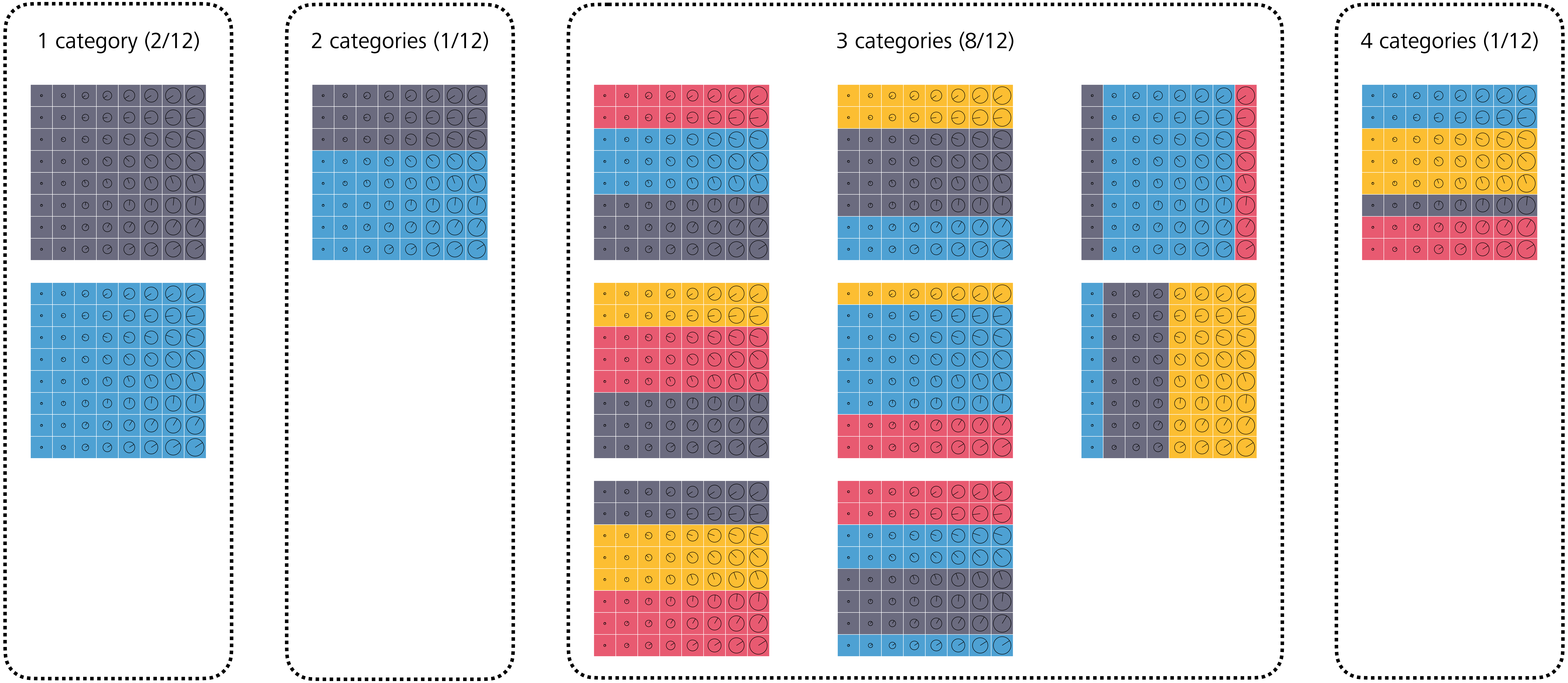
Simplicity bias
Manipulate noise
Manipulate bottleneck
Manipulate exposures
ε = .1
ε = .05
ε = .01
b = 3
b = 2
b = 1
ξ = 3
ξ = 2
ξ = 1
25
75
125
Complexity (bits)
Informativeness bias
Manipulate noise
Manipulate bottleneck
Manipulate exposures
ε = .1
ε = .05
ε = .01
b = 3
b = 2
b = 1
ξ = 3
ξ = 2
ξ = 1
4.25
4.30
4.35
Communicative cost (bits)

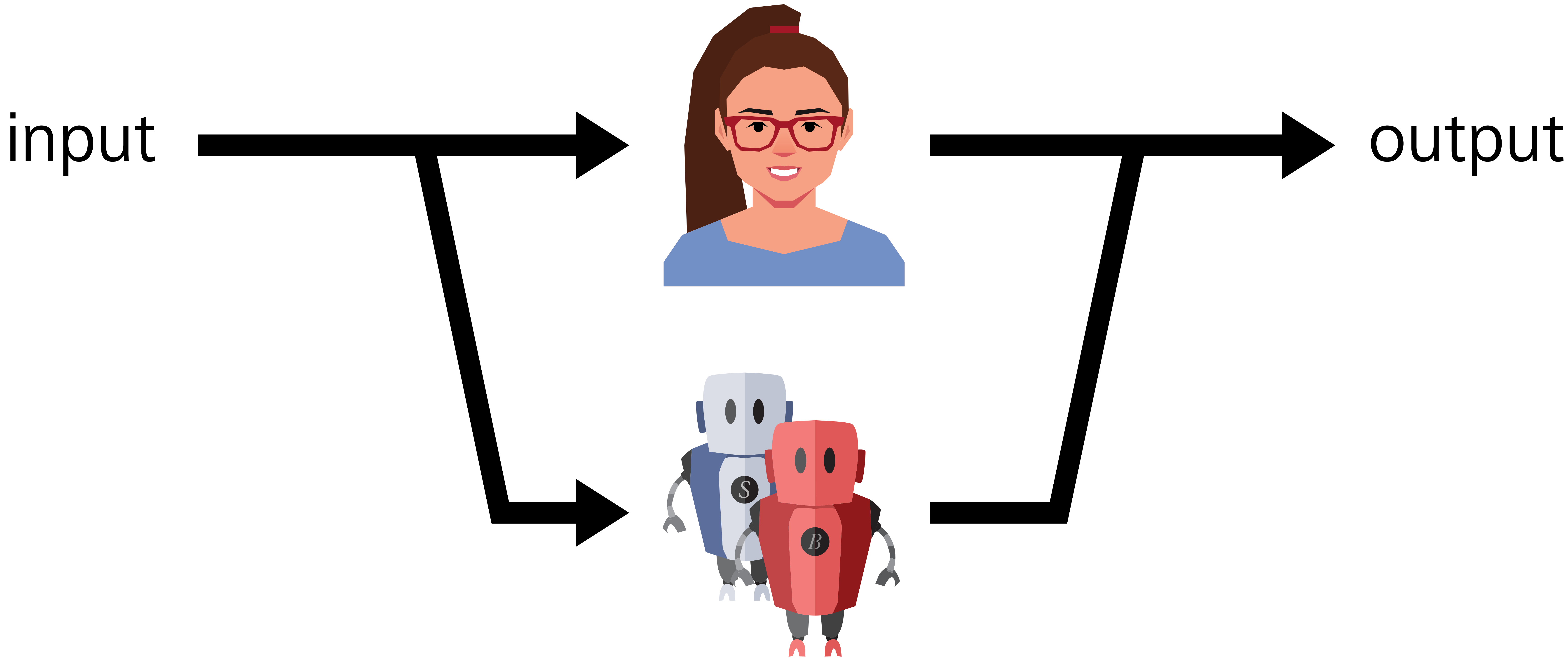

Angle
Size

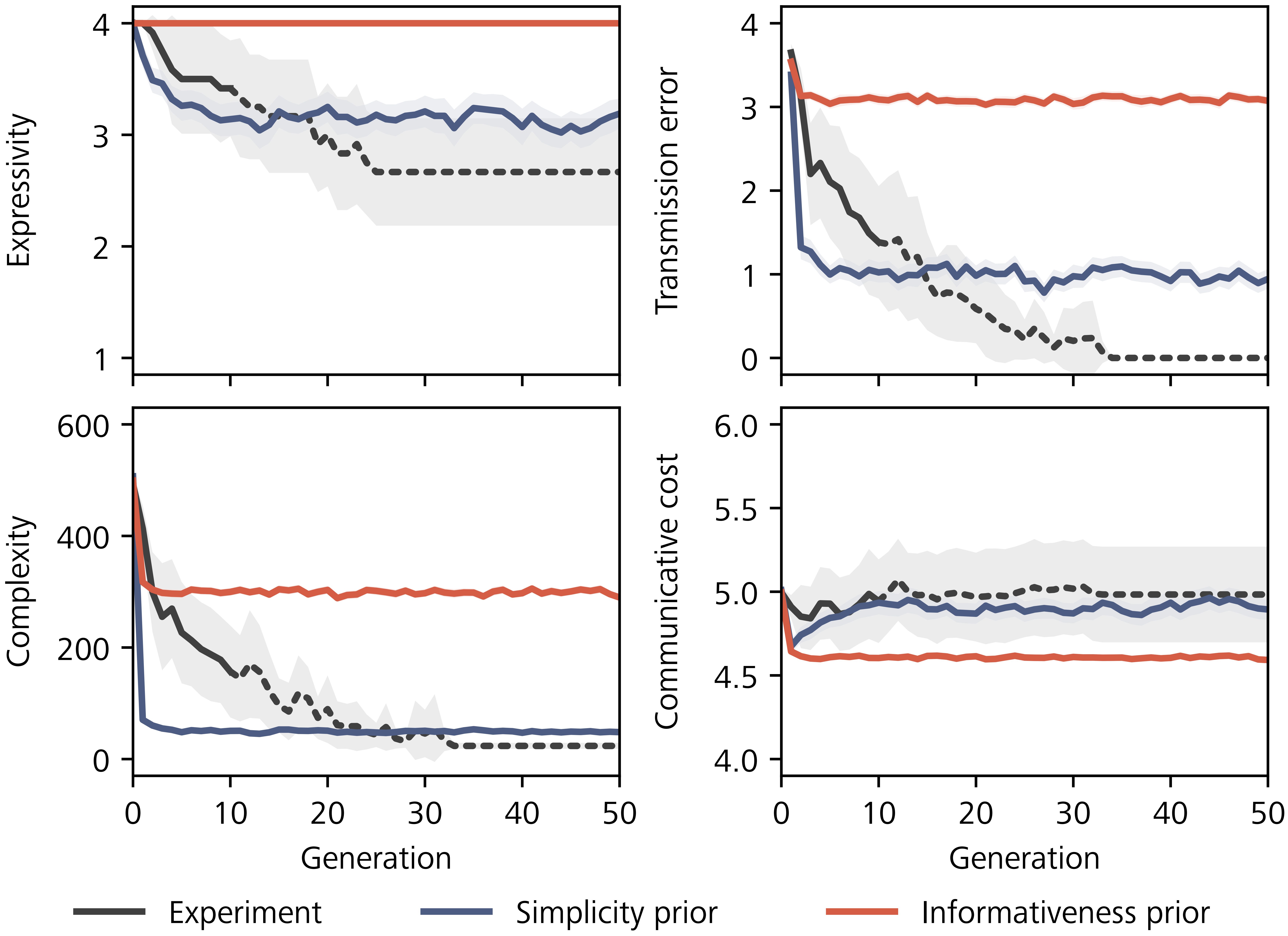




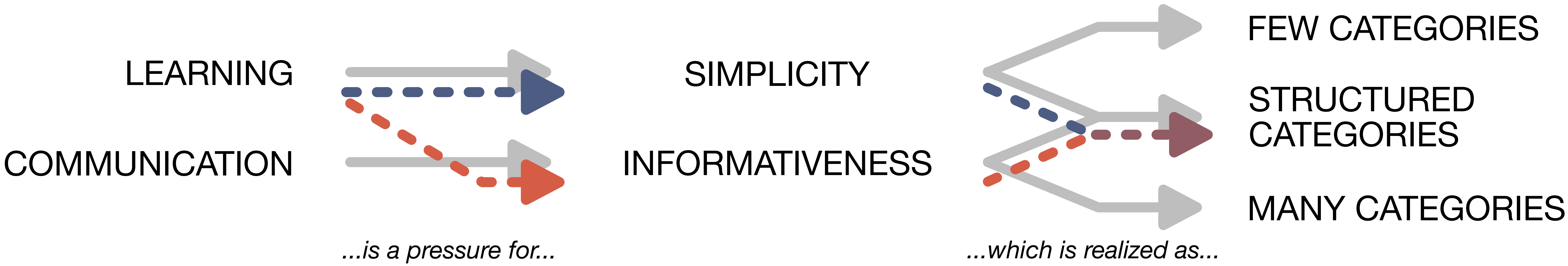


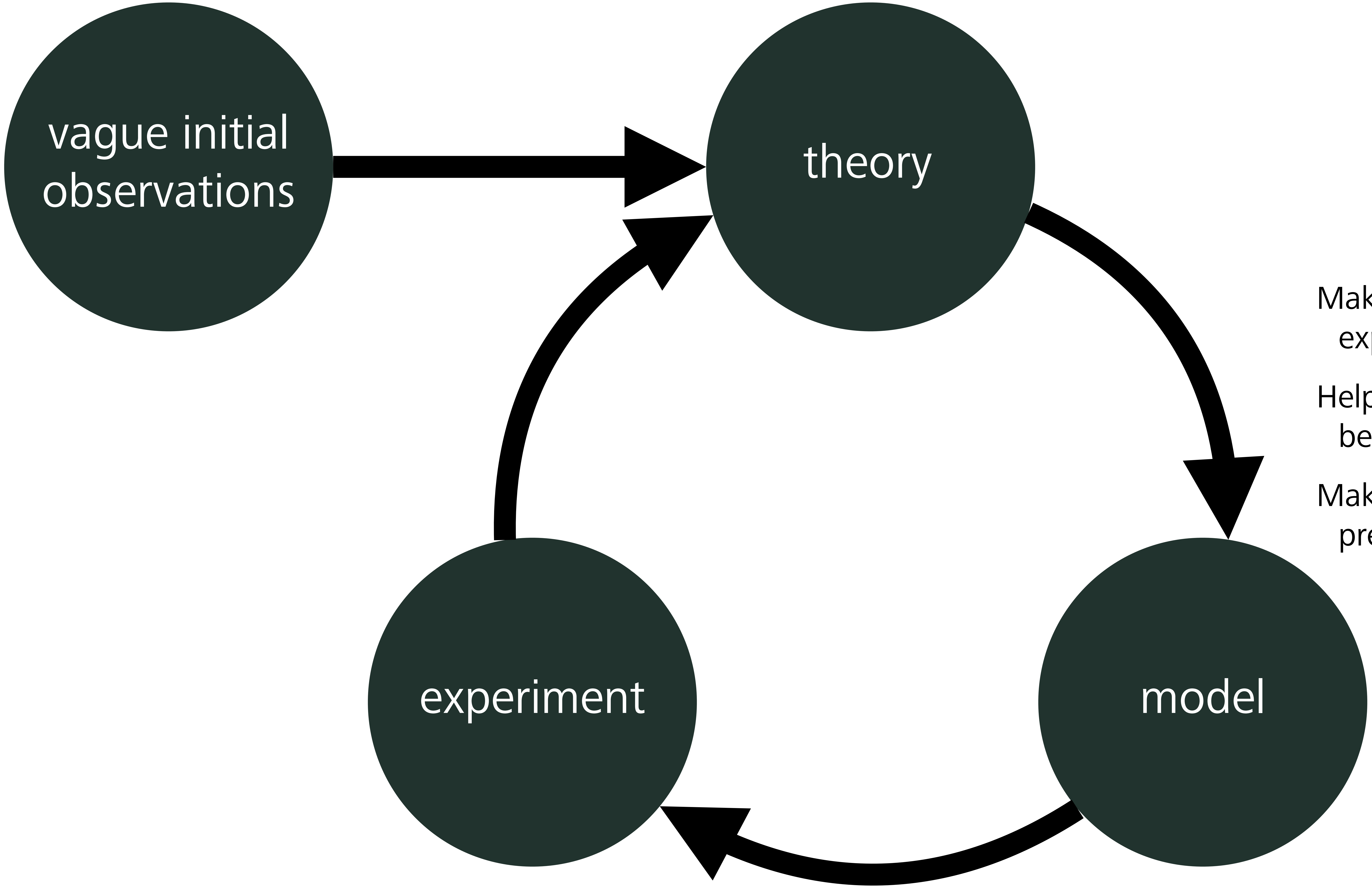

1 category (2/12)
2 categories (1/12)
3 categories (8/12)
4 categories (1/12)


input
output



Expressivity
Transmission error
Complexity
Communicative cost
Generation
Generation
Experiment
Simplicity prior
Informativeness prior


LEARNING
COMMUNICATION
...is a pressure for...
SIMPLICITY
INFORMATIVENESS
...which is realized as...
FEW CATEGORIES
STRUCTURED CATEGORIES
MANY CATEGORIES



Makes theoretical assumptions explicit

Helps you refine your theory before moving to experiments

Makes concrete experimental predictions