Convexity and expressivity in the simplicity–informativeness tradeoff

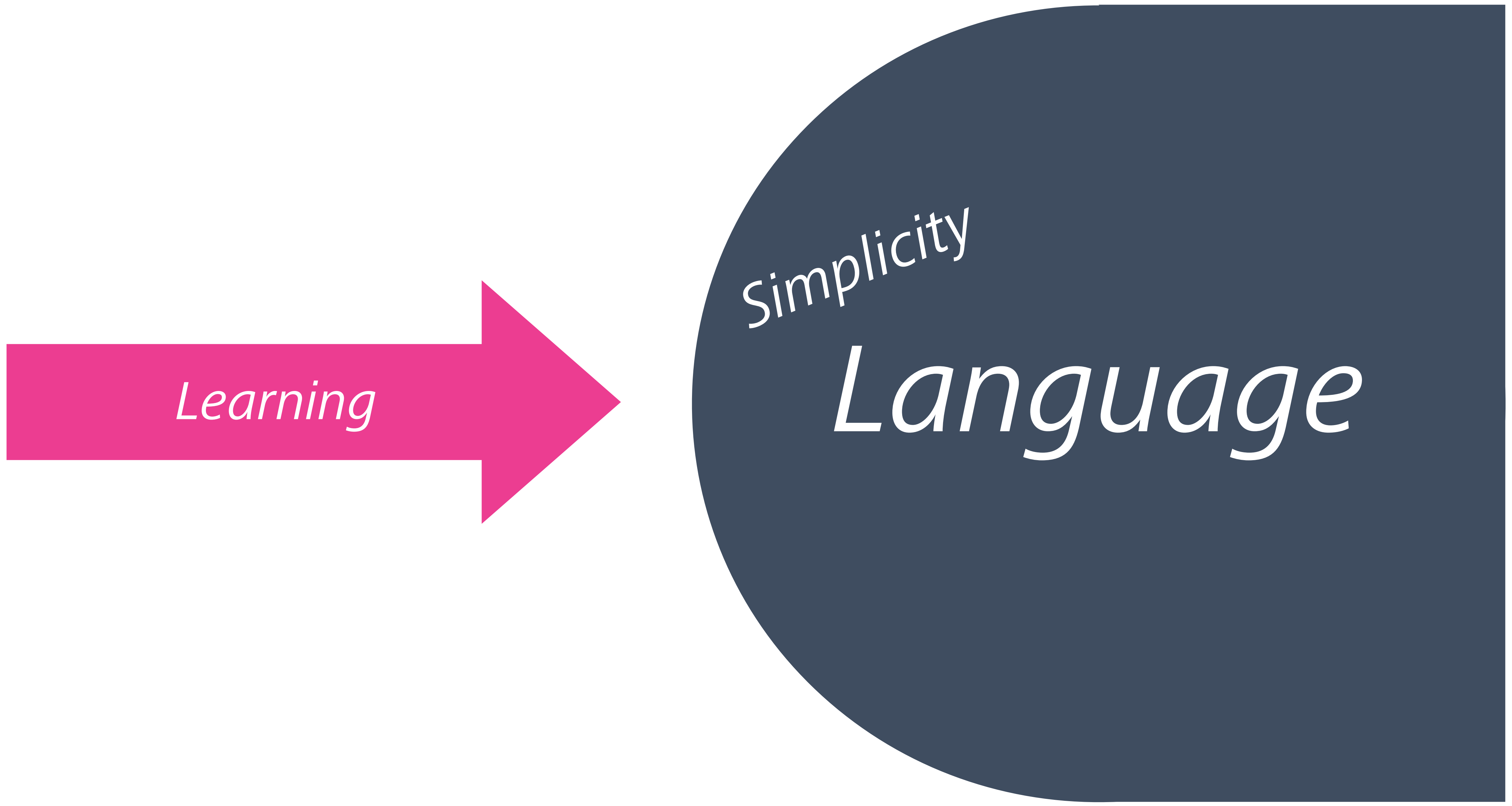
Jon W. Carr, Kenny Smith, Jennifer Culbertson, Simon Kirby

Centre for Language Evolution
School of Philosophy, Psychology and Language Sciences
University of Edinburgh

Pressures shaping language

Language

Pressures shaping language

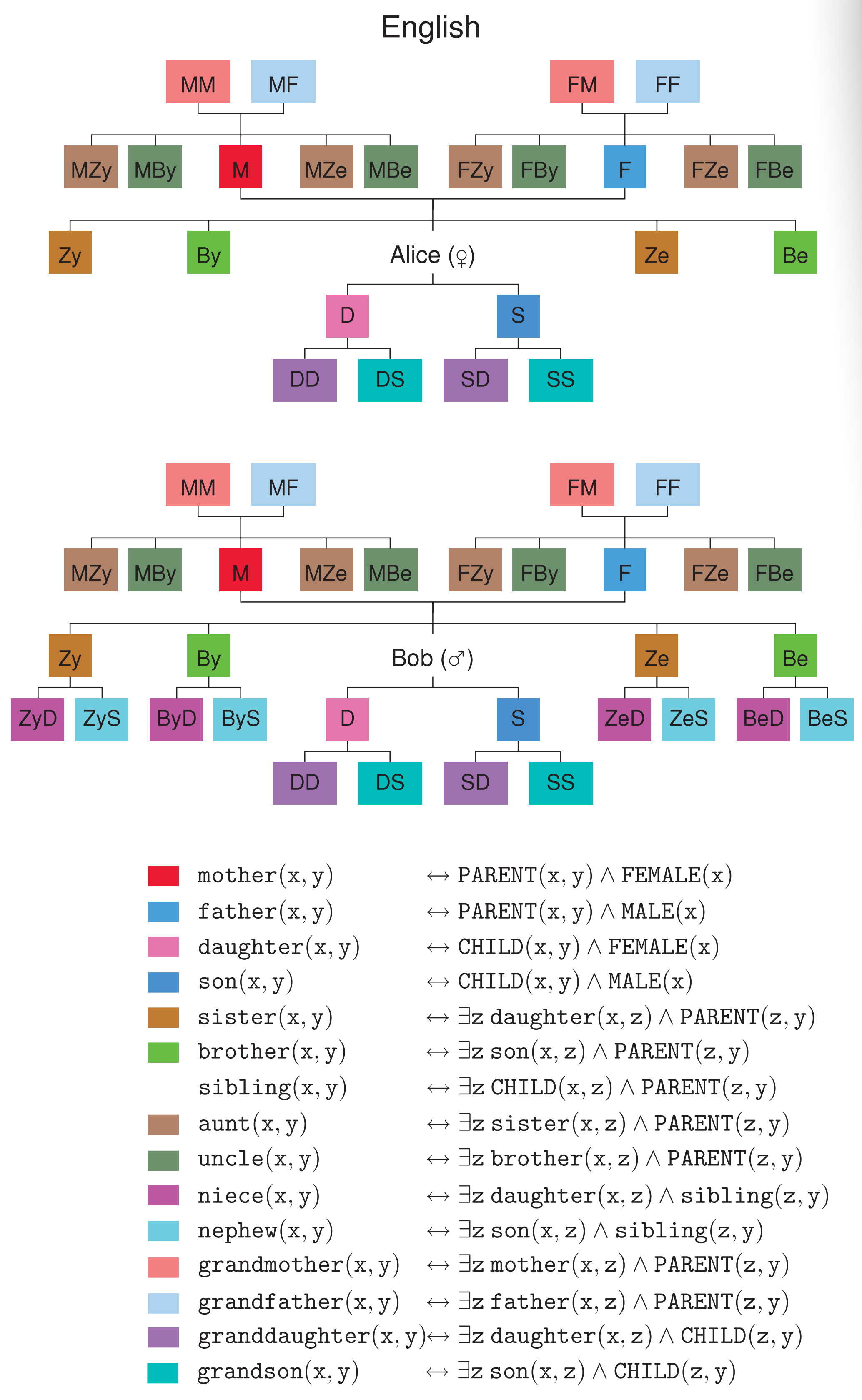


Pressures shaping language

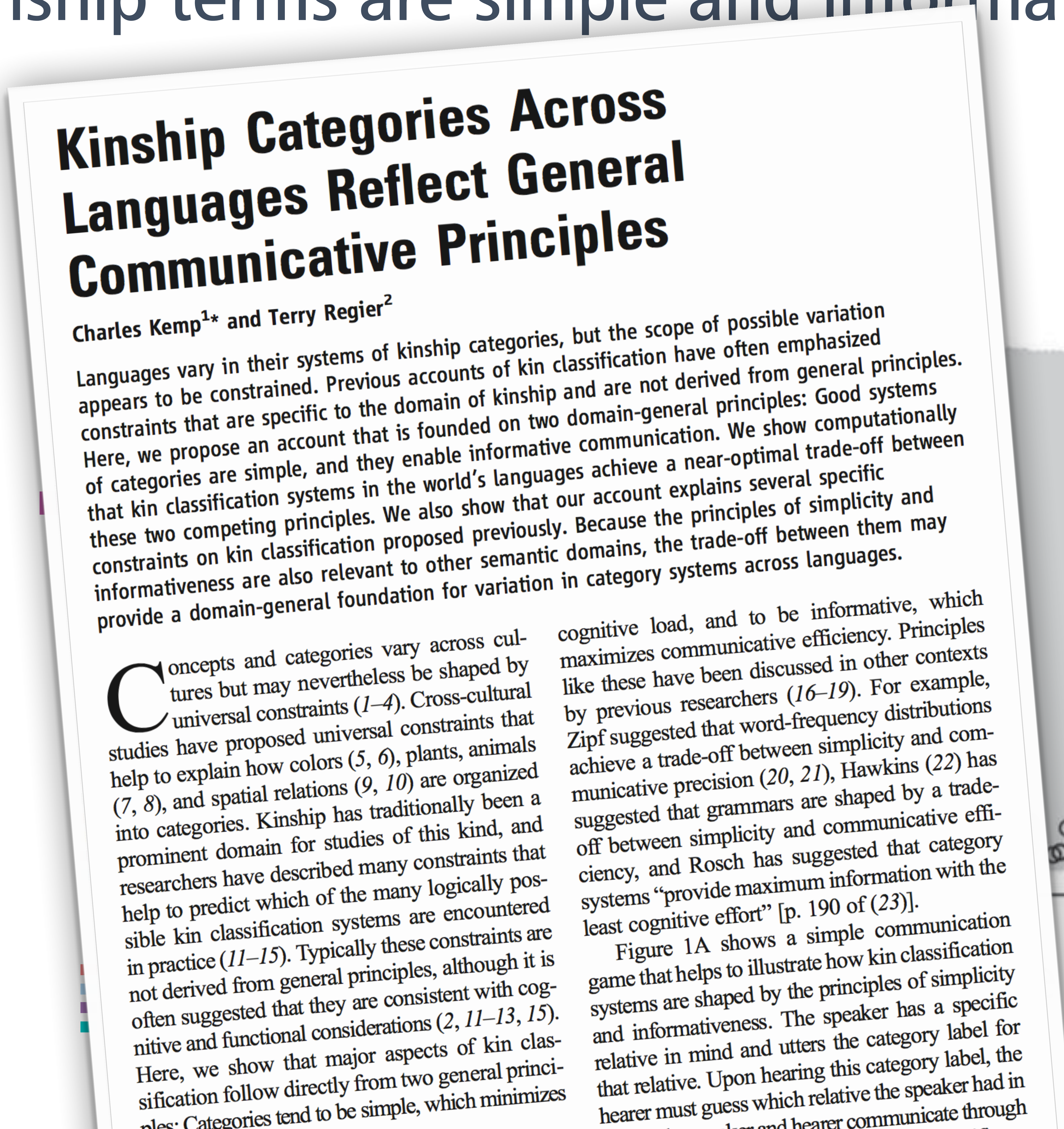
Pressures shaping language

The simplicity–informativeness tradeoff

Kinship terms are simple and informative



Kemp & Regier (2012)



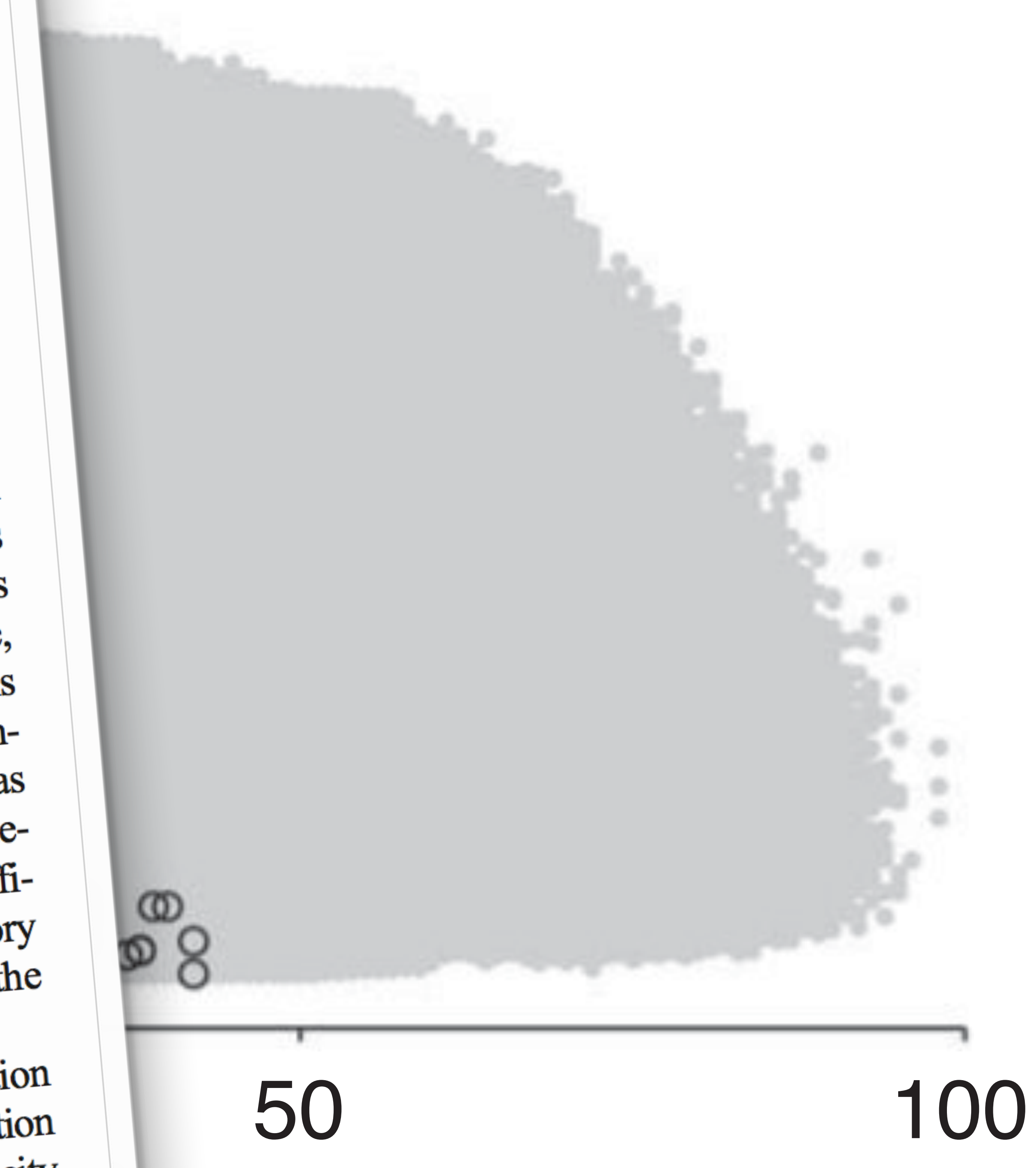
Kinship Categories Across Languages Reflect General Communicative Principles

Charles Kemp[1]* and Terry Regier[2]

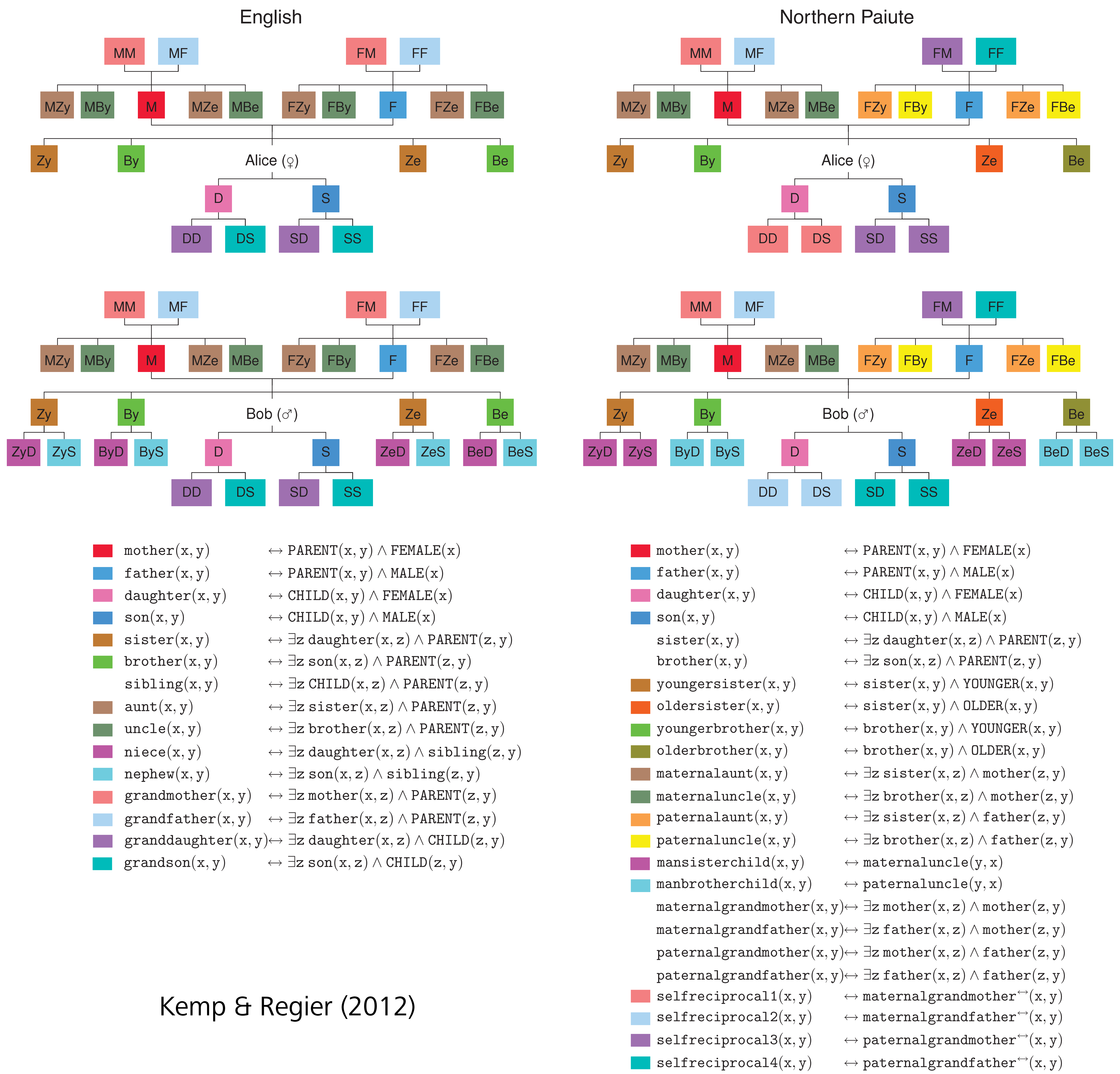
Languages vary in their systems of kinship categories, but the scope of possible variation appears to be constrained. Previous accounts of kin classification have often emphasized constraints that are specific to the domain of kinship and are not derived from general principles. Here, we propose an account that is founded on two domain-general principles: Good systems of categories are simple, and they enable informative communication. We show computationally that kin classification systems in the world's languages achieve a near-optimal trade-off between these two competing principles. We also show that our account explains several specific constraints on kin classification proposed previously. Because the principles of simplicity and informativeness are also relevant to other semantic domains, the trade-off between them may provide a domain-general foundation for variation in category systems across languages.

Concepts and categories vary across cultures but may nevertheless be shaped by universal constraints (1–4). Cross-cultural studies have proposed universal constraints that help to explain how colors (5, 6), plants, animals (7, 8), and spatial relations (9, 10) are organized into categories. Kinship has traditionally been a prominent domain for studies of this kind, and researchers have described many constraints that help to predict which of the many logically possible kin classification systems are encountered in practice (11–15). Typically these constraints are not derived from general principles, although it is often suggested that they are consistent with cognitive and functional considerations (2, 11–13, 15). Here, we show that major aspects of kin classification follow directly from two general principles: Categories tend to be simple, which minimizes cognitive load, and to be informative, which maximizes communicative efficiency. Principles like these have been discussed in other contexts by previous researchers (16–19). For example, Zipf suggested that word-frequency distributions achieve a trade-off between simplicity and communicative precision (20, 21), Hawkins (22) has suggested that grammars are shaped by a trade-off between simplicity and communicative efficiency, and Rosch has suggested that category systems "provide maximum information with the least cognitive effort" [p. 190 of (23)].

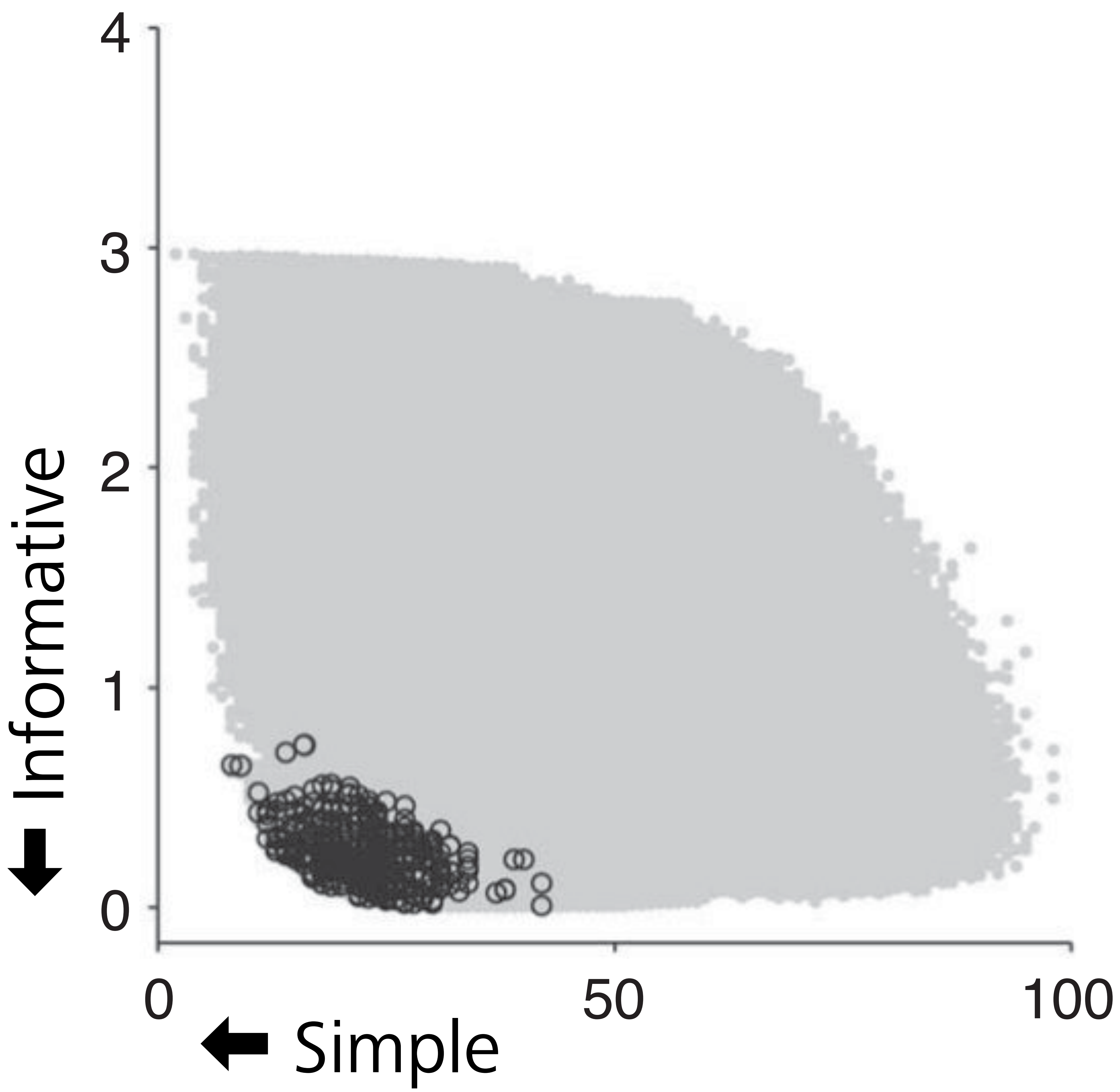
Figure 1A shows a simple communication game that helps to illustrate how kin classification systems are shaped by the principles of simplicity and informativeness. The speaker has a specific relative in mind and utters the category label for that relative. Upon hearing this category label, the hearer must guess which relative the speaker had in mind.



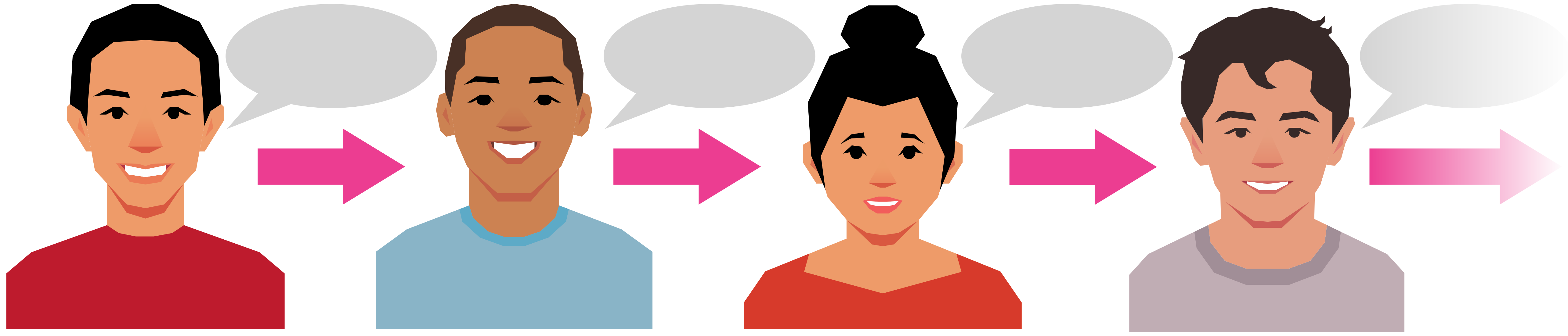
Kinship terms are simple and informative



Kemp & Regier (2012)

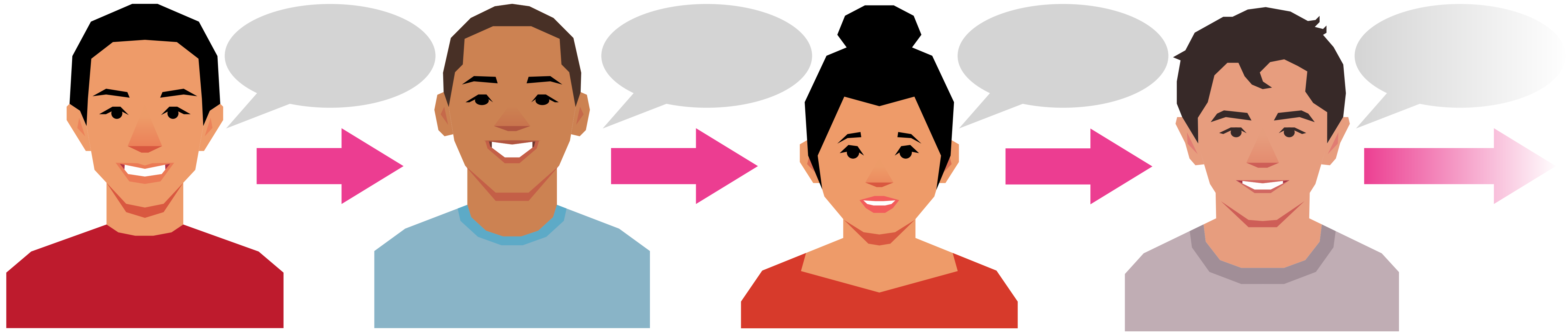


Learning vs. Communication

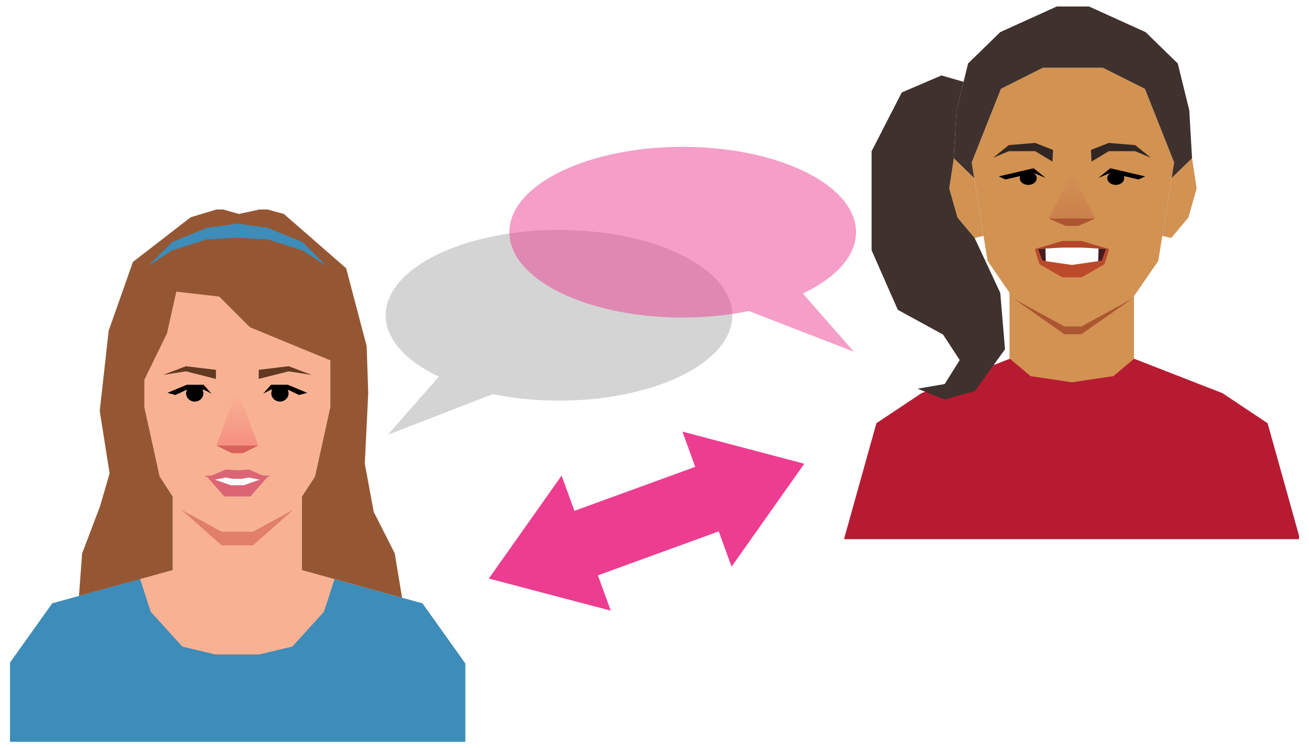


Iterated learning

Learning vs. Communication

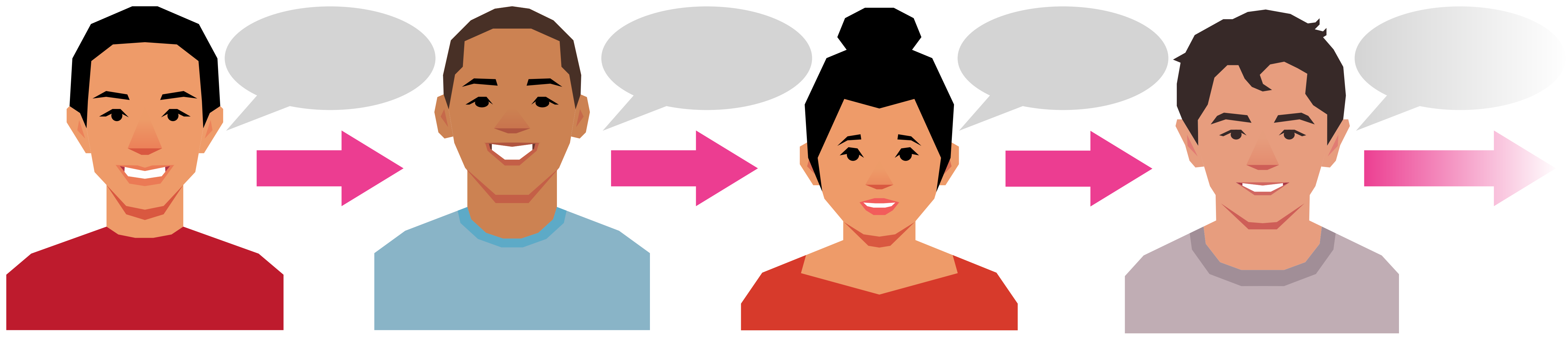


Iterated learning

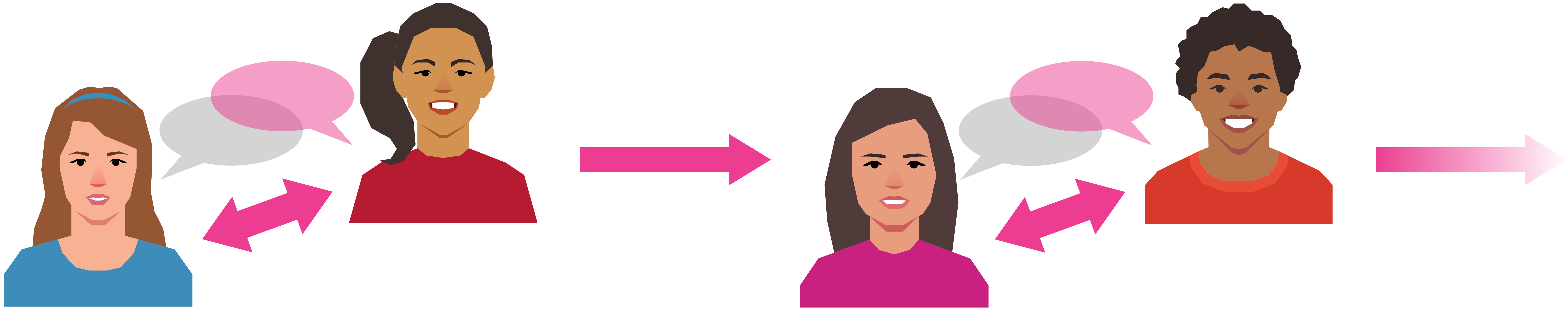


Communication

Learning vs. Communication

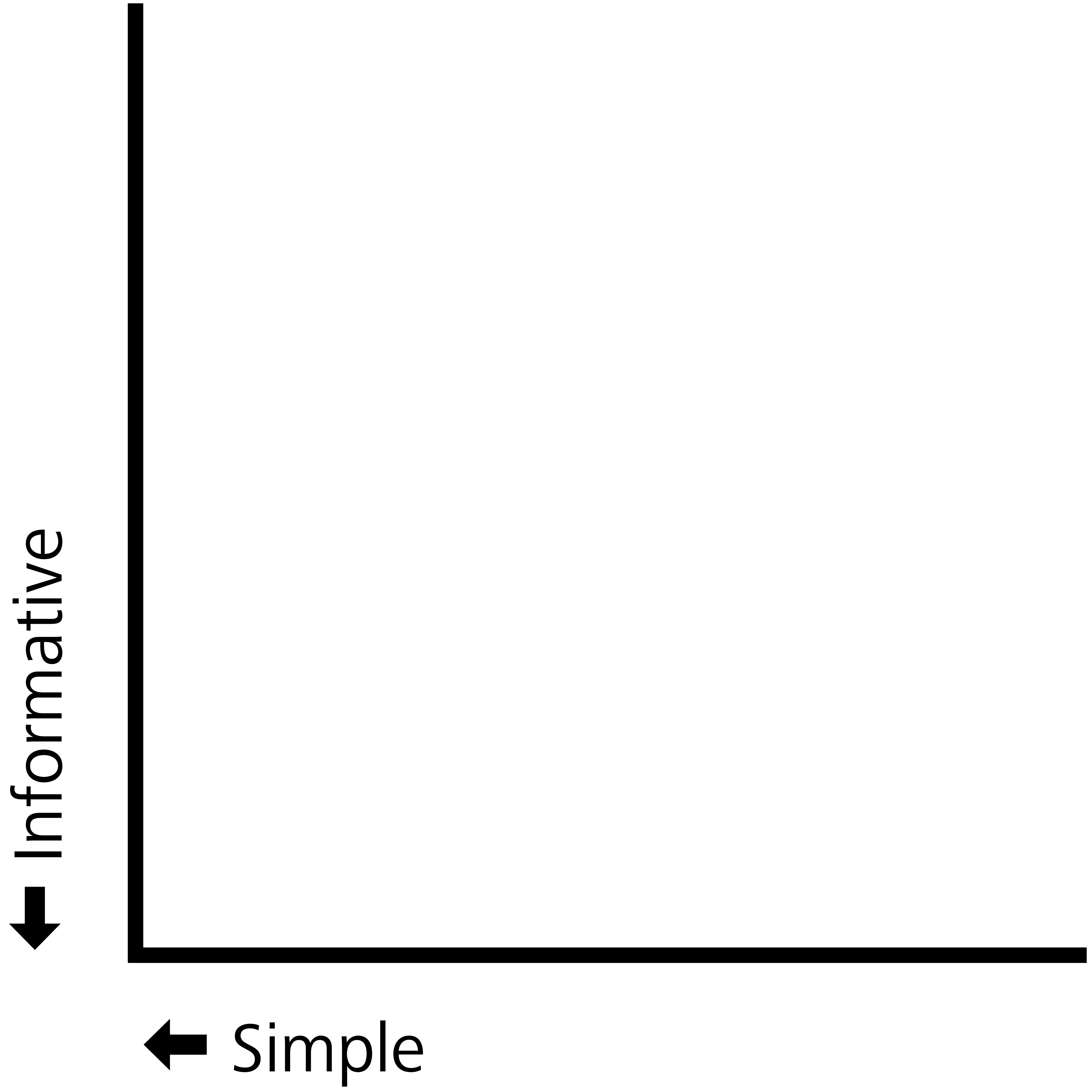


Iterated learning

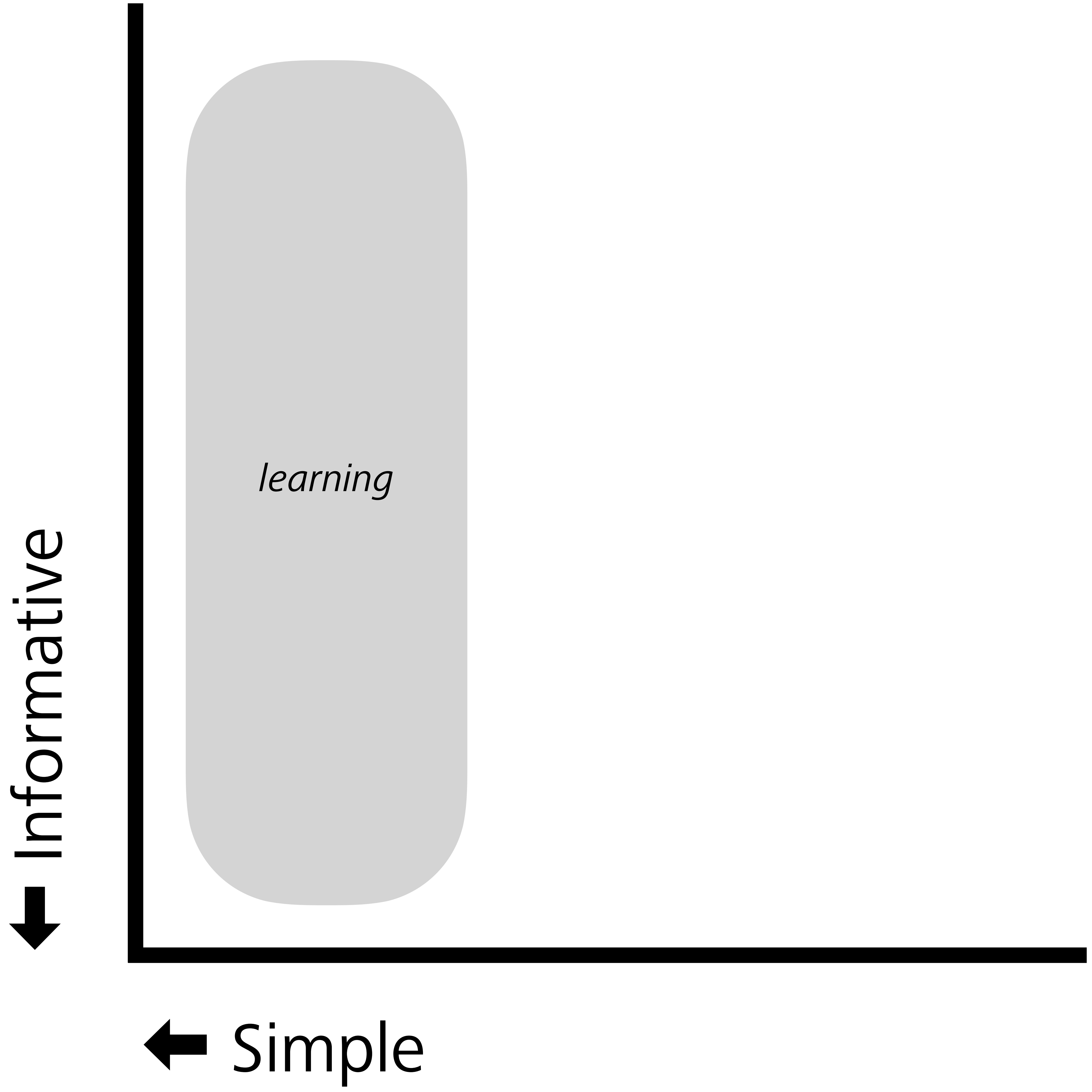


Iterated learning & Communication

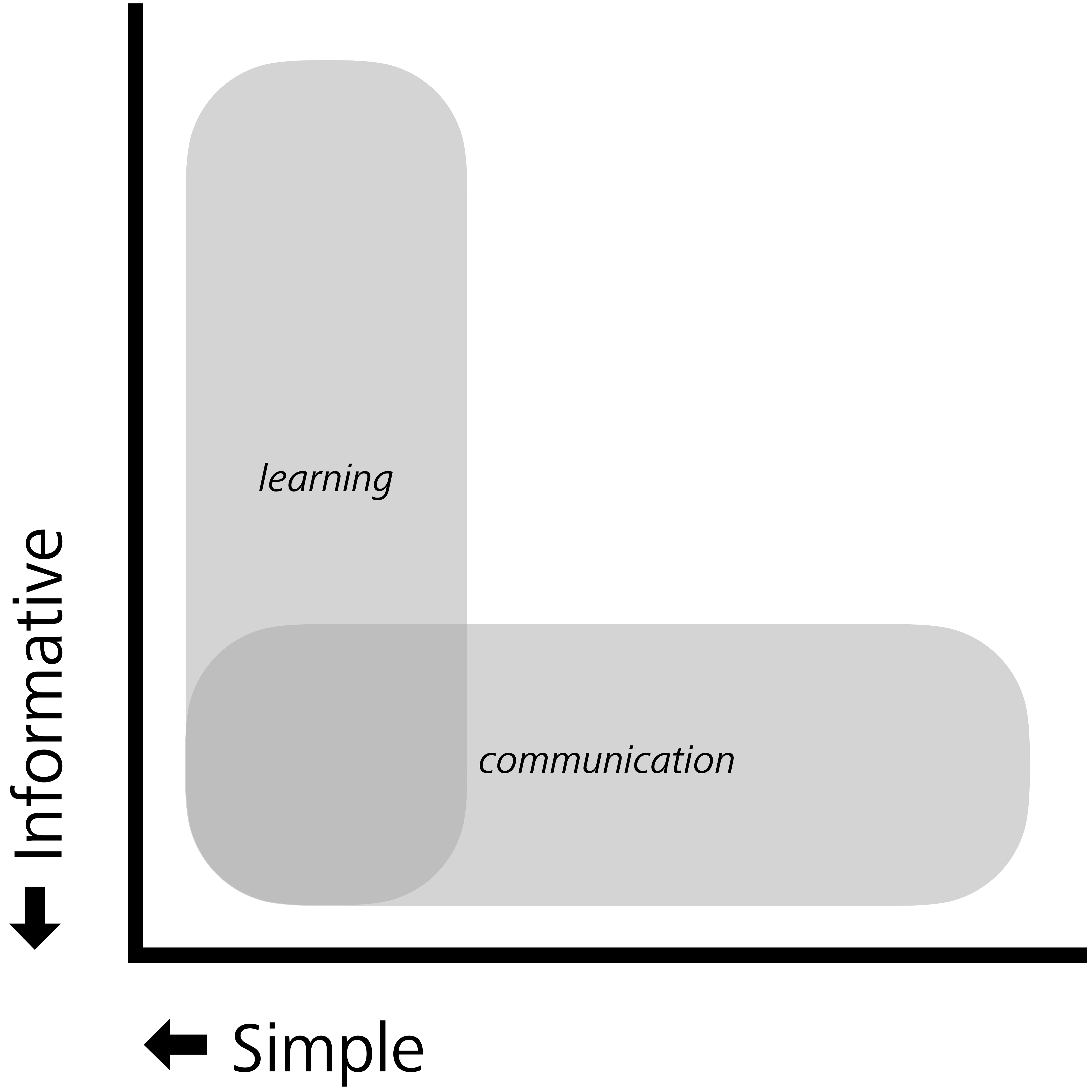
Learning and communication pressures



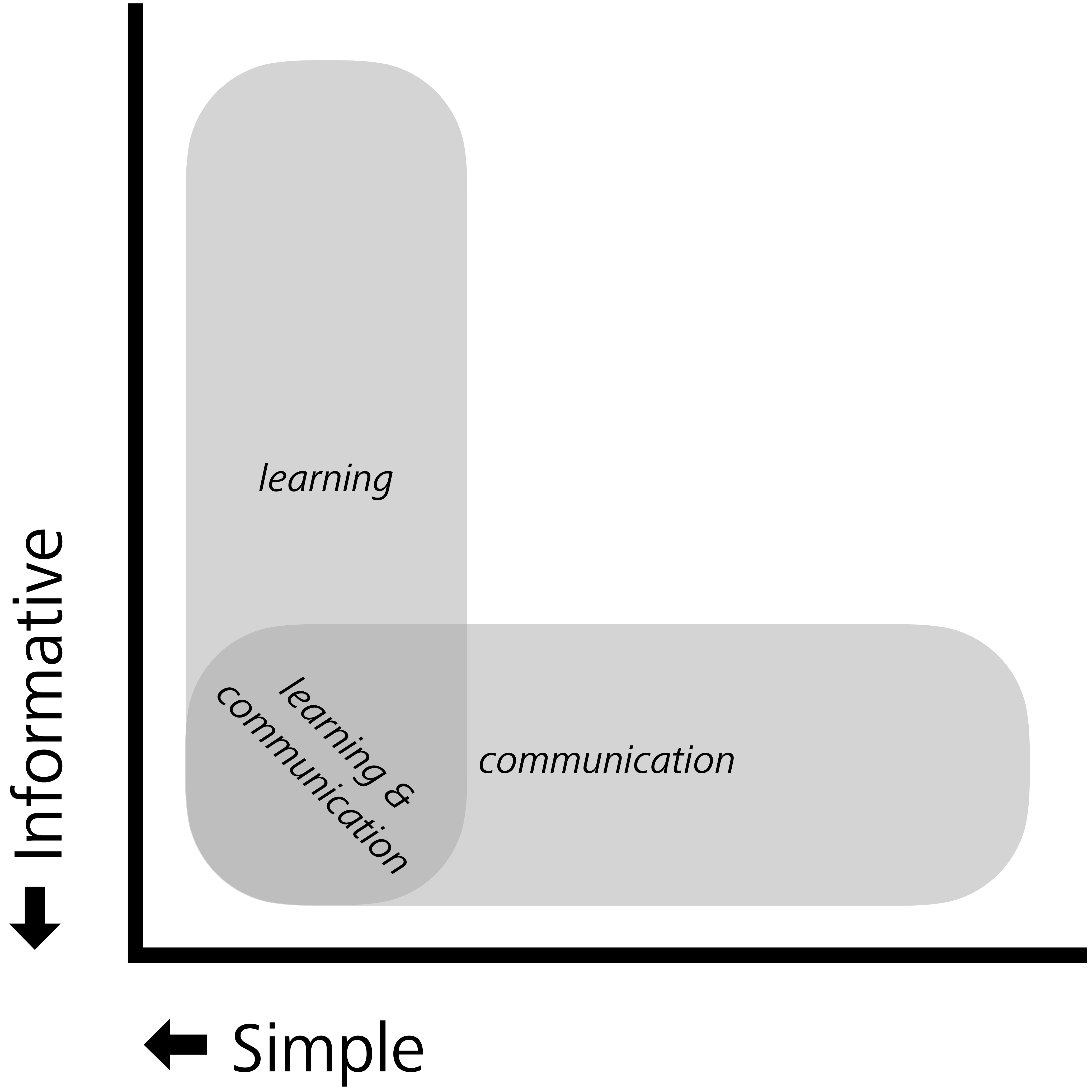
Learning and communication pressures



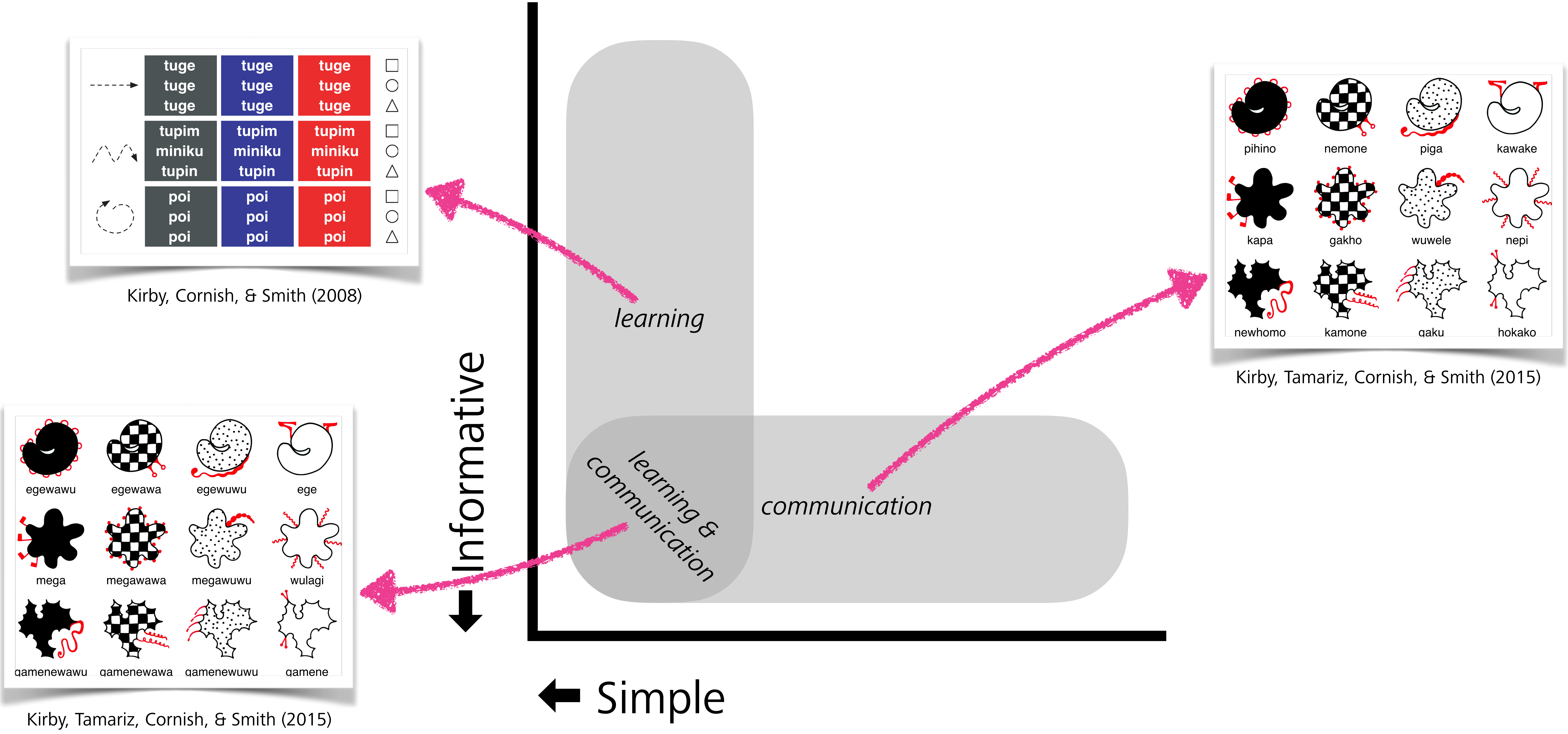
Learning and communication pressures



Learning and communication pressures



Learning and communication pressures



Simplicity

The Minimum Description Length principle

$$\mathrm{DL}(H|D) = \mathrm{DL}(D|H) + \mathrm{DL}(H)$$

The Minimum Description Length principle

$$\mathrm{DL}(H|D) = \mathrm{DL}(D|H) + \mathrm{DL}(H)$$

$$\mathrm{posterior}(H|D) = \mathrm{likelihood}(D|H) \times \mathrm{prior}(H)$$

The Minimum Description Length principle

$$\mathrm{DL}(H|D) = \mathrm{DL}(D|H) + \mathrm{DL}(H)$$

$$\mathrm{posterior}(H|D) = \mathrm{likelihood}(D|H) \times 2^{-\mathrm{DL}(H)}$$

The Minimum Description Length principle

$$\mathrm{DL}(H|D) = \mathrm{DL}(D|H) + \mathrm{DL}(H)$$

$$\mathrm{posterior}(H|D) = \mathrm{likelihood}(D|H) \times 2^{-\mathrm{DL}(H)}$$

Any regularities in data can be used to compress that data

The more regularities there are, the more the data can be compressed

The Minimum Description Length principle

$$\mathrm{DL}(H|D) = \mathrm{DL}(D|H) + \mathrm{DL}(H)$$

Any regulari

The more re ... ompressed

For example…

```
010010111110010000110001000101101100001111010001
```

```
print('010010111110010000110001000101101100001111010001')
```

```
010101010101010101010101010101010101010101010101
```

```
print('01'*24)
```

The Minimum Description Length principle

$$\mathrm{DL}(H|D) = \mathrm{DL}(D|H) + \mathrm{DL}(H)$$

$$\mathrm{posterior}(H|D) = \mathrm{likelihood}(D|H) \times 2^{-\mathrm{DL}(H)}$$

Any regularities in data can be used to compress that data

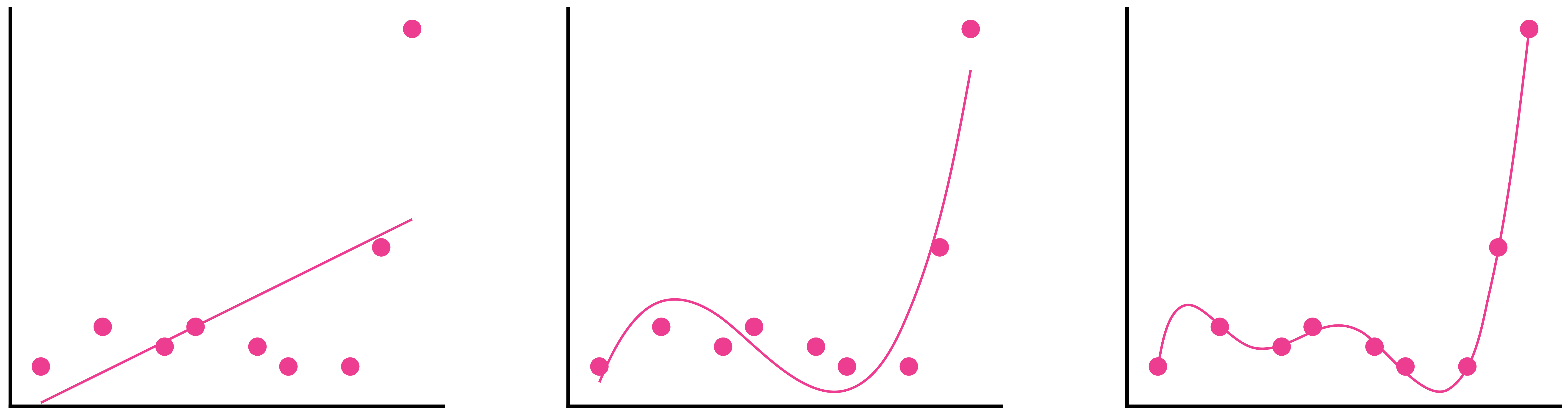
The more regularities there are, the more the data can be compressed

We equate **learning** with **finding regularity**: The more the data can be compressed, the more we have learned from that data

In other words, the more regularity we can identify, the more we have understood (learned) about the process generating the data

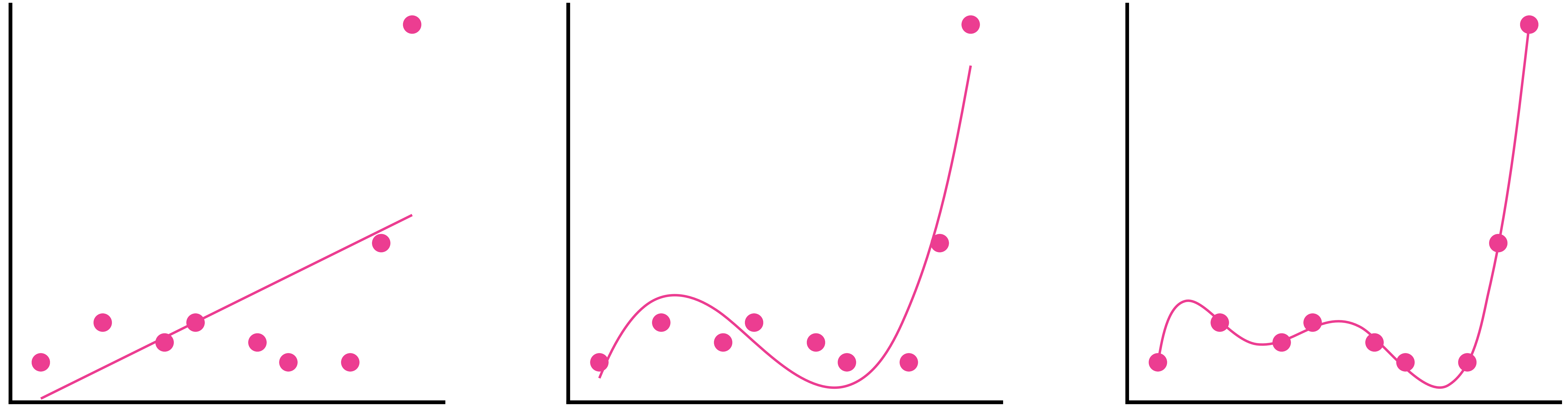
The Minimum Description Length principle

$$\mathrm{DL}(H|D) = \mathrm{DL}(D|H) + \mathrm{DL}(H)$$



The Minimum Description Length principle

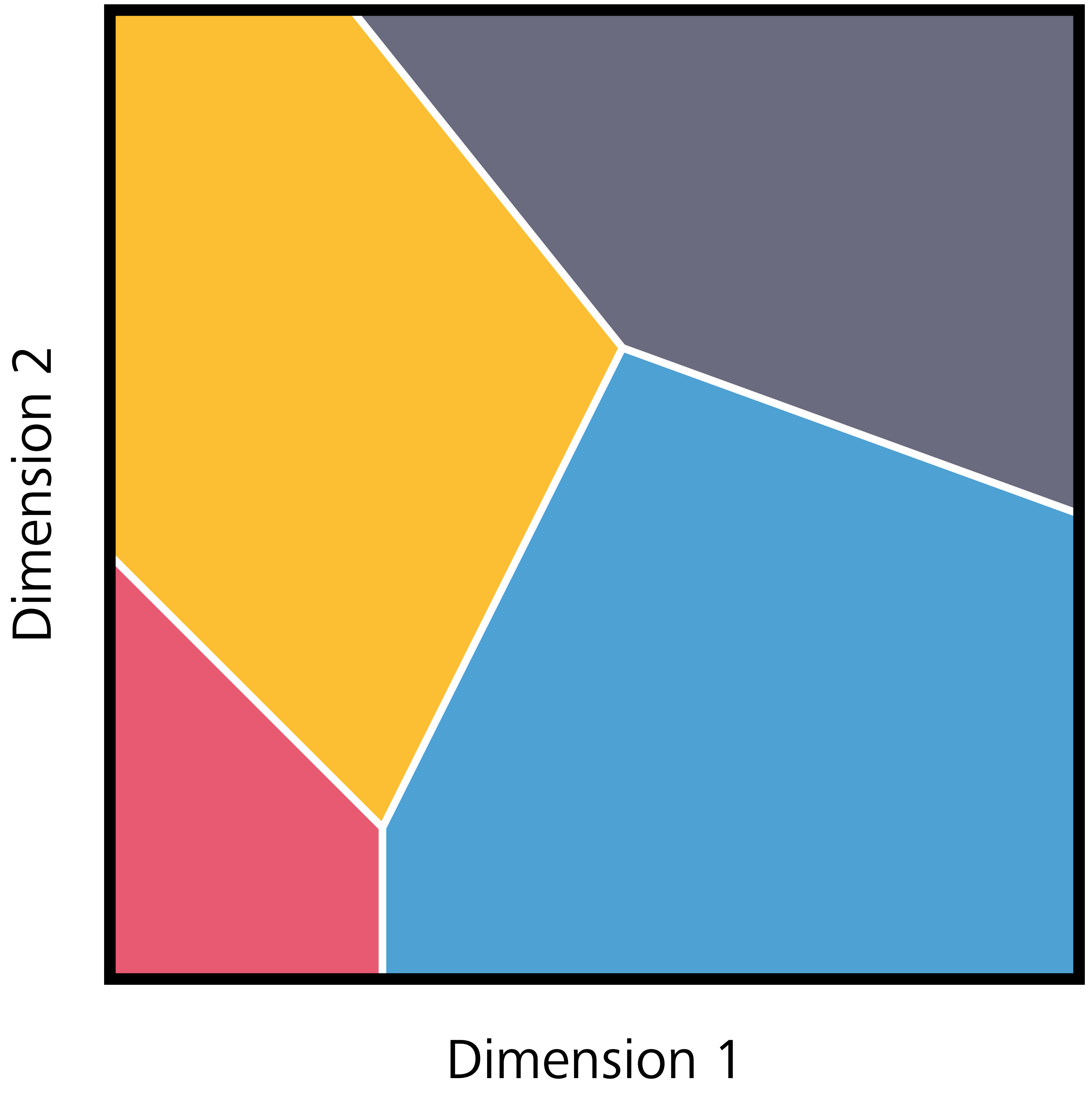
$$\mathrm{DL}(H|D) = \mathrm{DL}(D|H) + \mathrm{DL}(H)$$



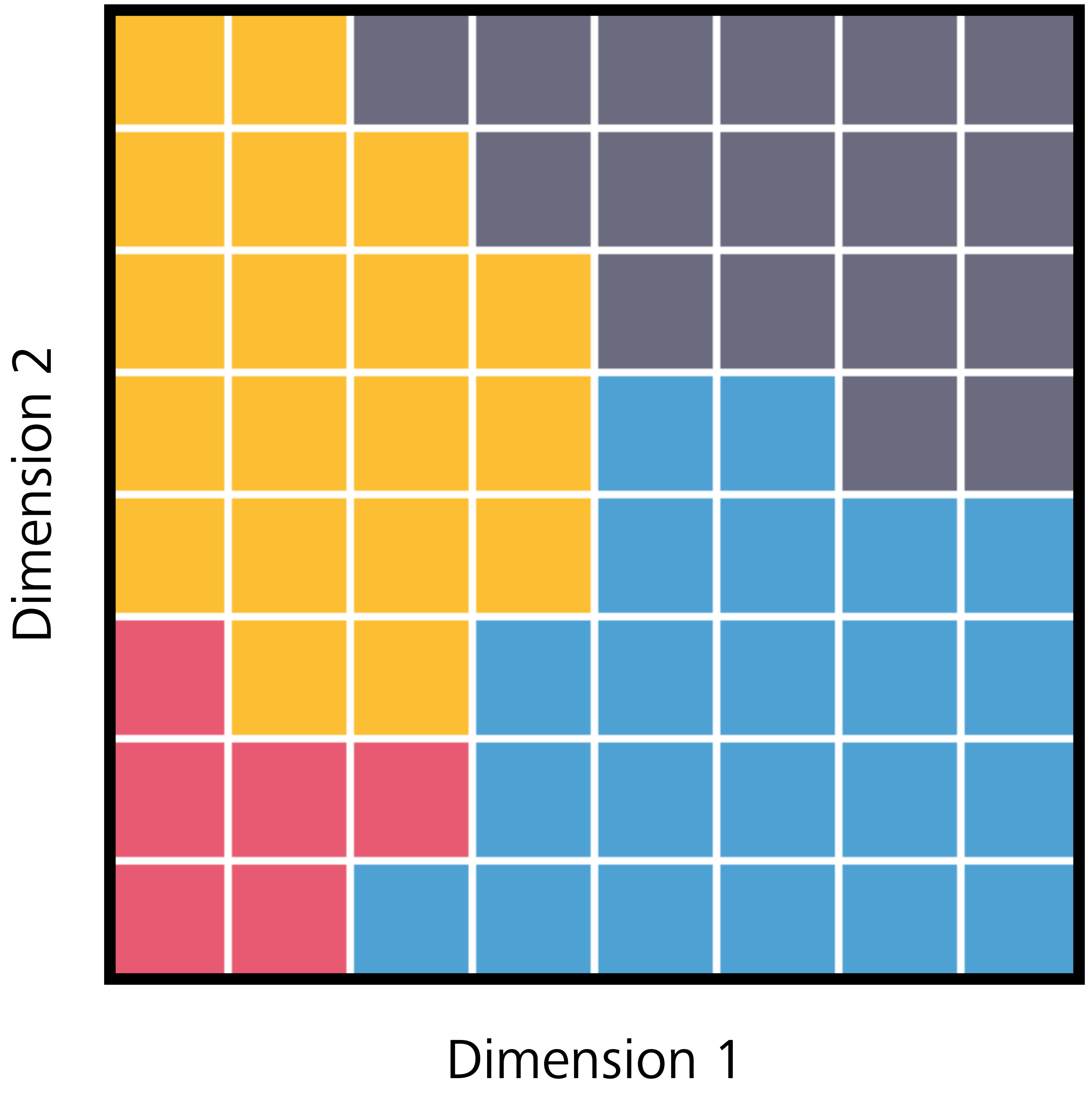
$\mathrm{DL}(D\|H)$	7 bits	2 bits	0 bits
$\mathrm{DL}(H)$	1 bit	3 bits	6 bits
$\mathrm{DL}(H\|D)$	8 bits ✗	5 bits ✓	6 bits ✗

Bayesian model

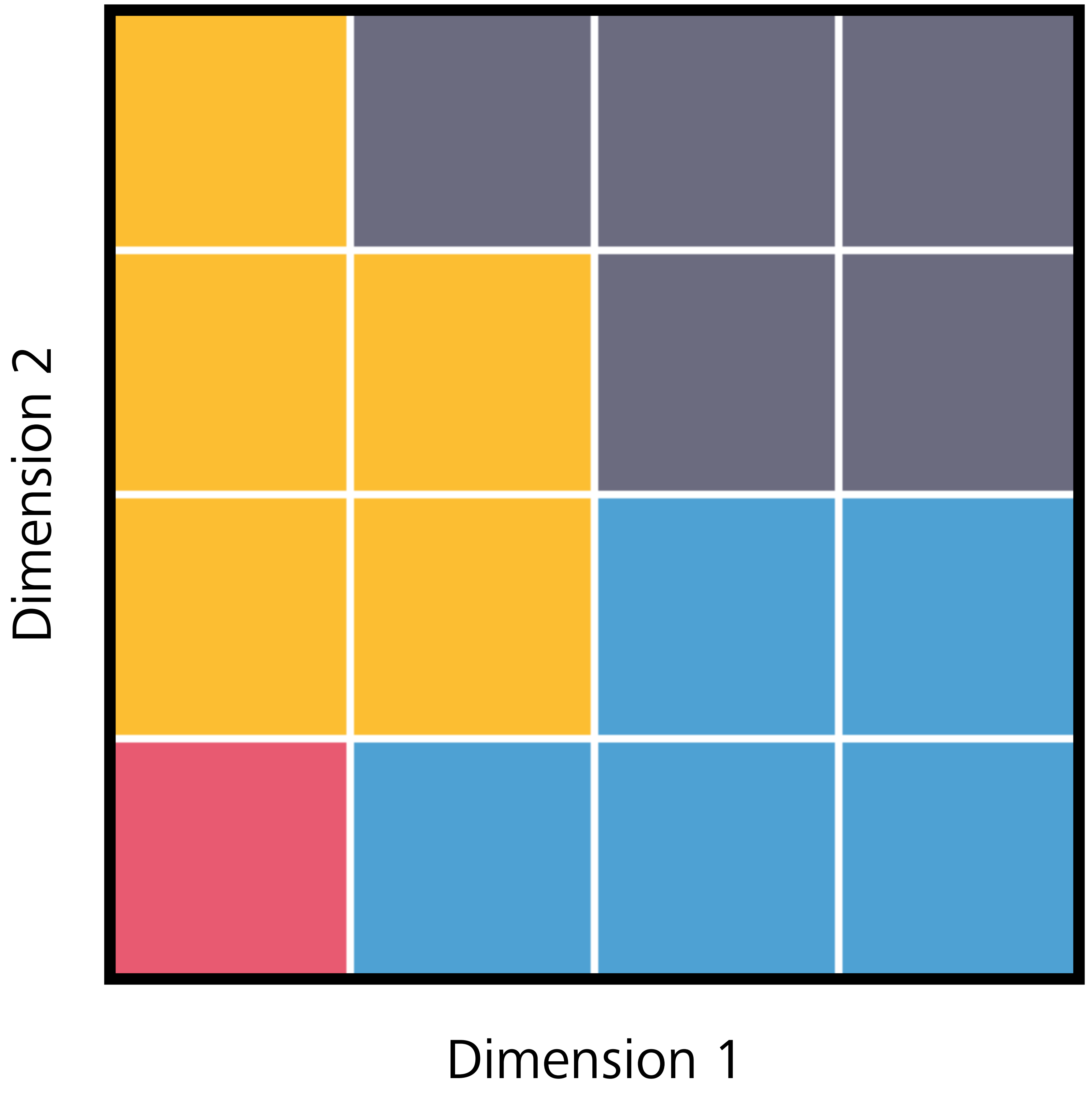
Conceptual spaces



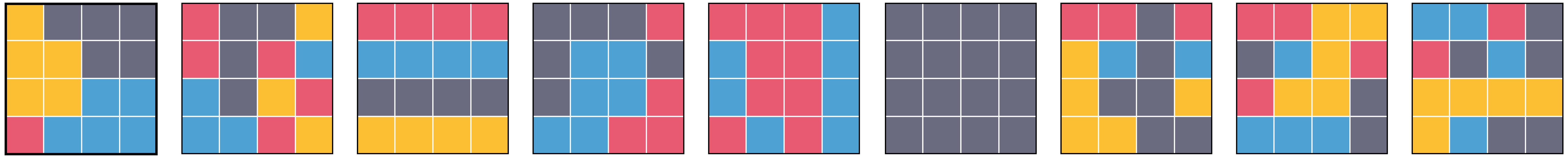
Conceptual spaces



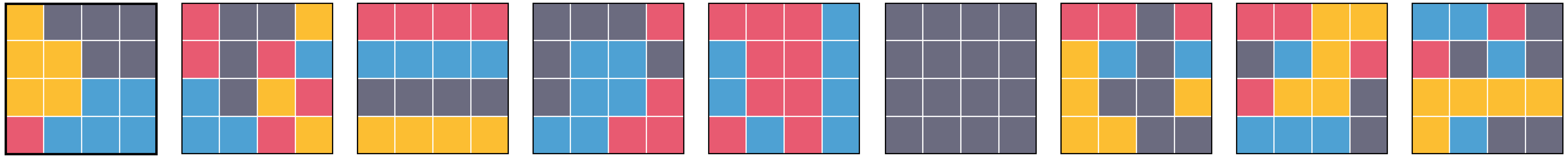
Conceptual spaces



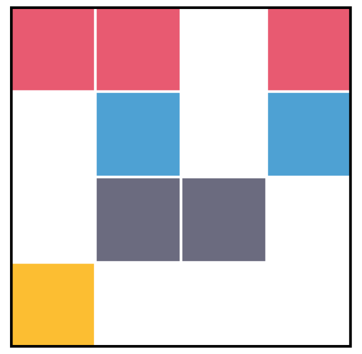
Bayesian inference

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

Bayesian inference

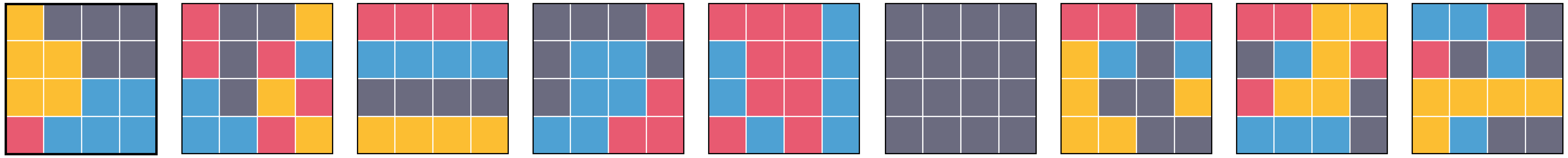
$\mathcal{L} = \{$  $\cdots\}$

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

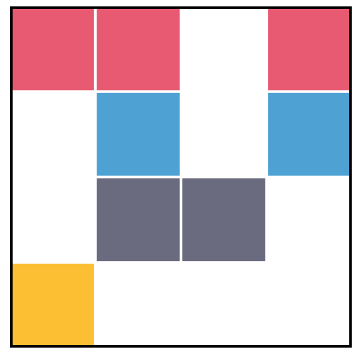


Bayesian inference

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

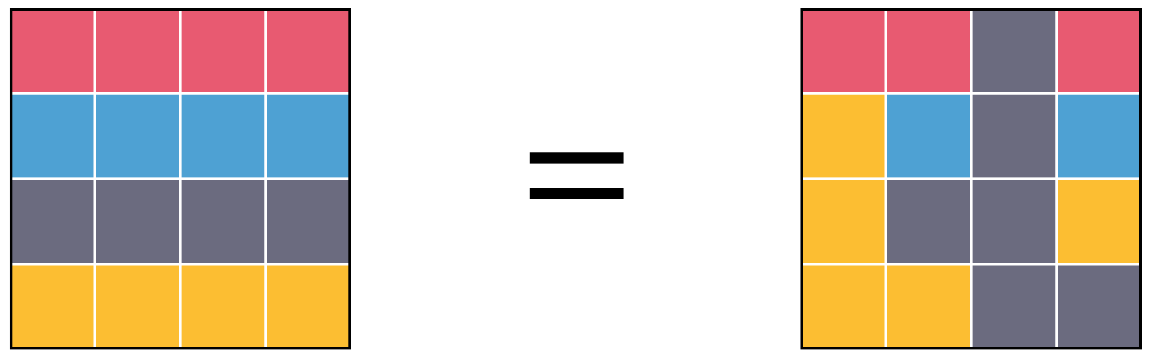


$$\cdots\}$$

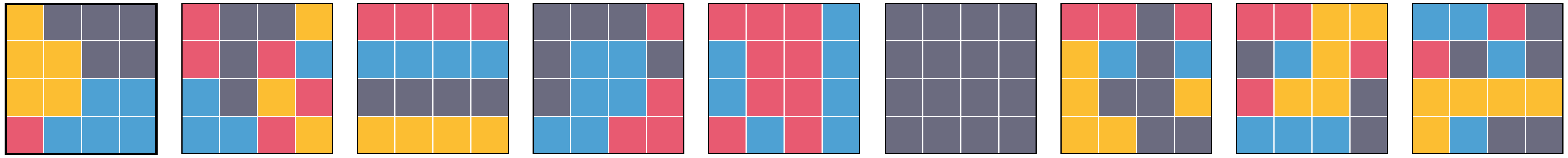


$$D = [\langle m_1, s_1 \rangle, \langle m_2, s_2 \rangle, \langle m_3, s_3 \rangle, \ldots, \langle m_n, s_n \rangle]$$

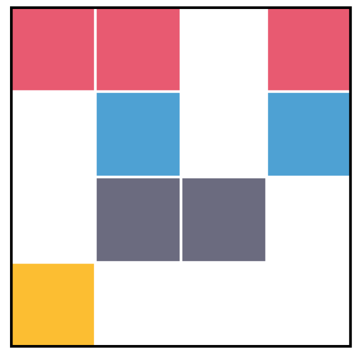
$$\text{likelihood}(D|L) \propto \prod_{\langle m,s \rangle \in D} \frac{1}{|M|} P(s|L, m)$$



Bayesian inference

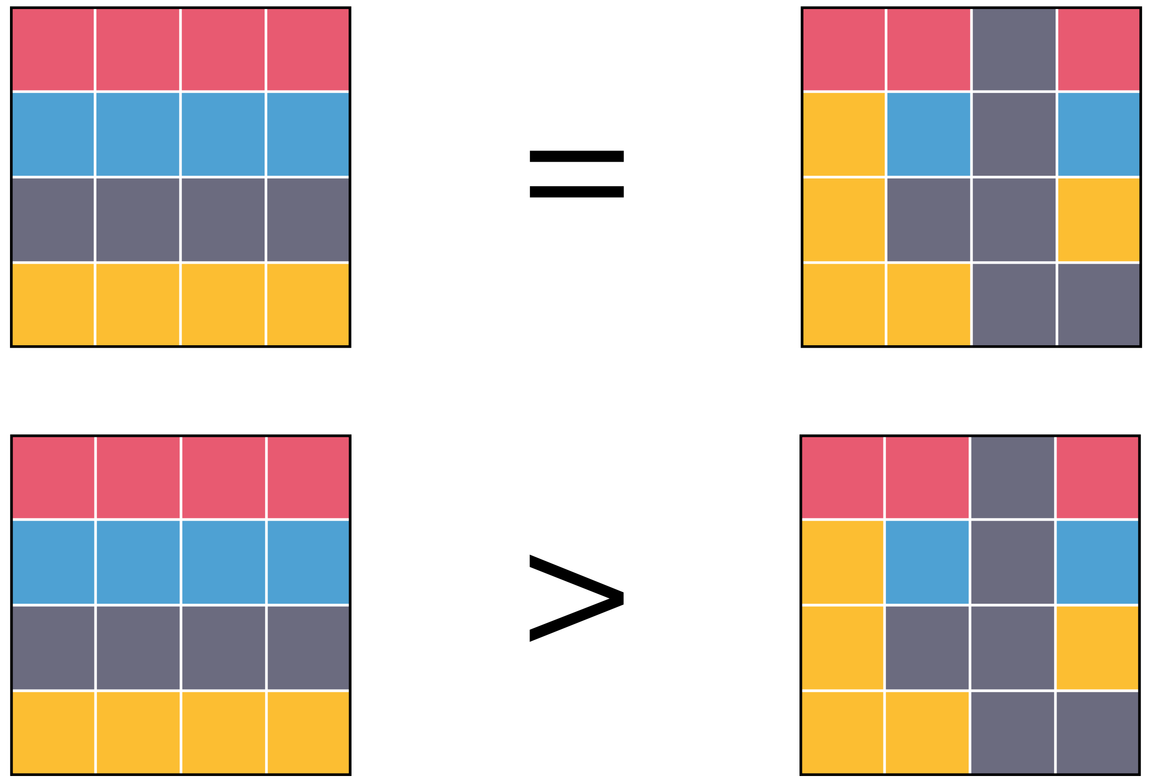
$\mathcal{L} = \{$  $\cdots\}$

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

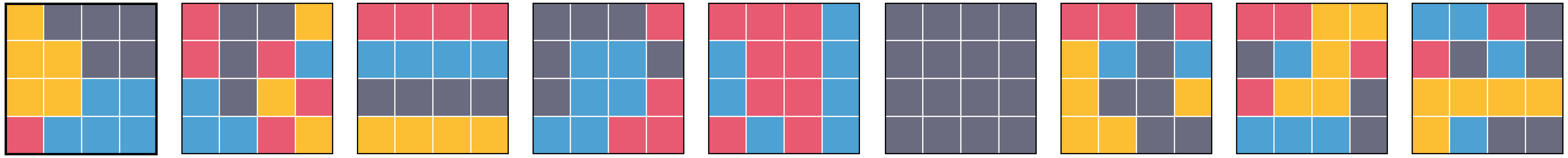


$$\text{likelihood}(D|L) \propto \prod_{\langle m,s \rangle \in D} \frac{1}{|M|} P(s|L, m)$$

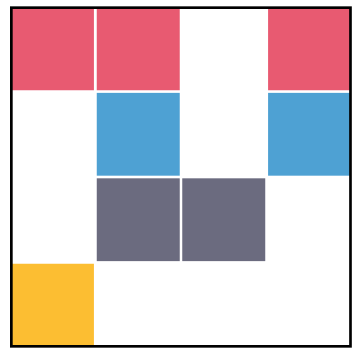
$$\text{prior}(L) \propto 2^{-\text{DL}(L)}$$



Bayesian inference

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

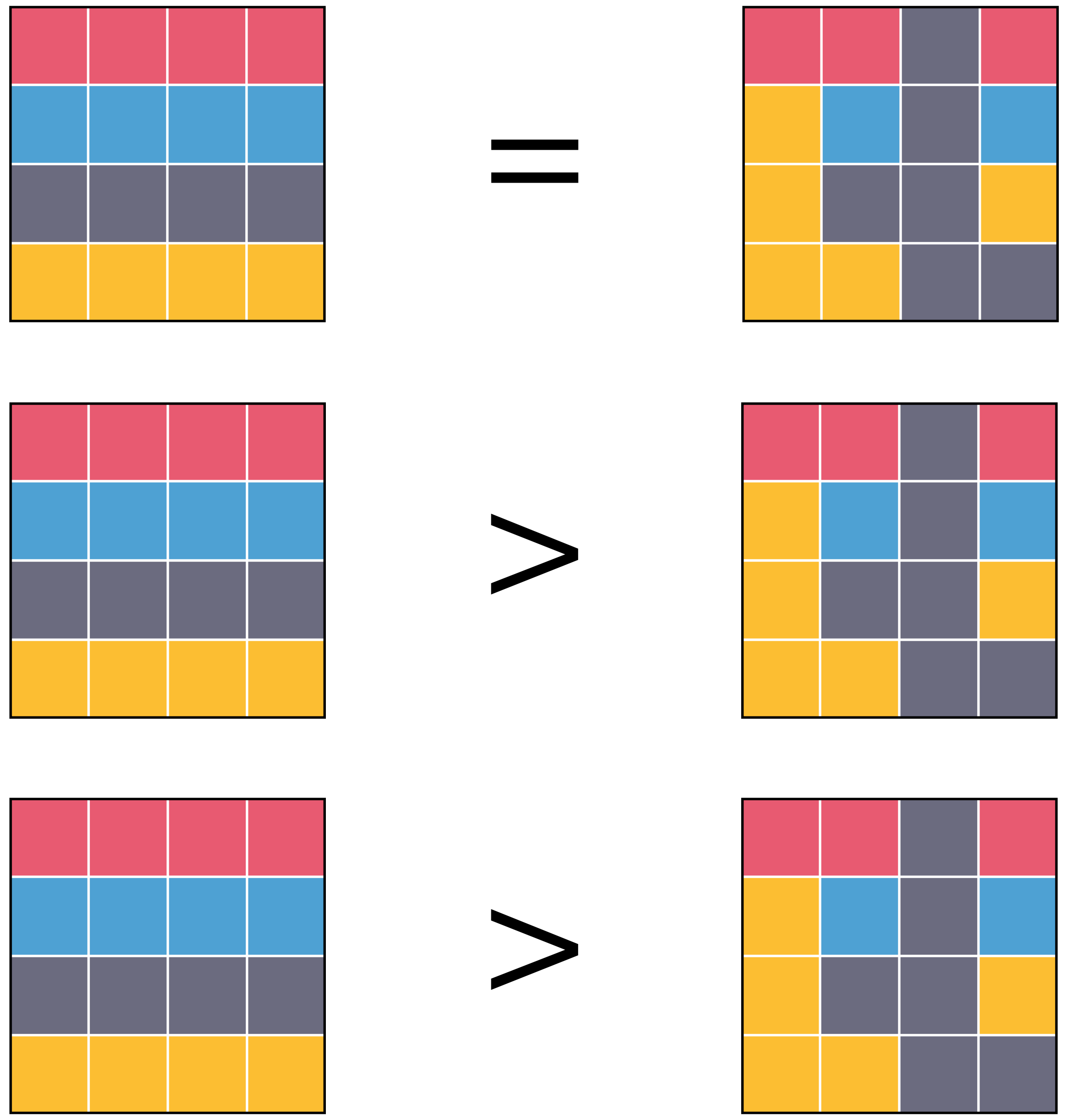
$$D = [\langle m_1, s_1\rangle, \langle m_2, s_2\rangle, \langle m_3, s_3\rangle, ..., \langle m_n, s_n\rangle]$$



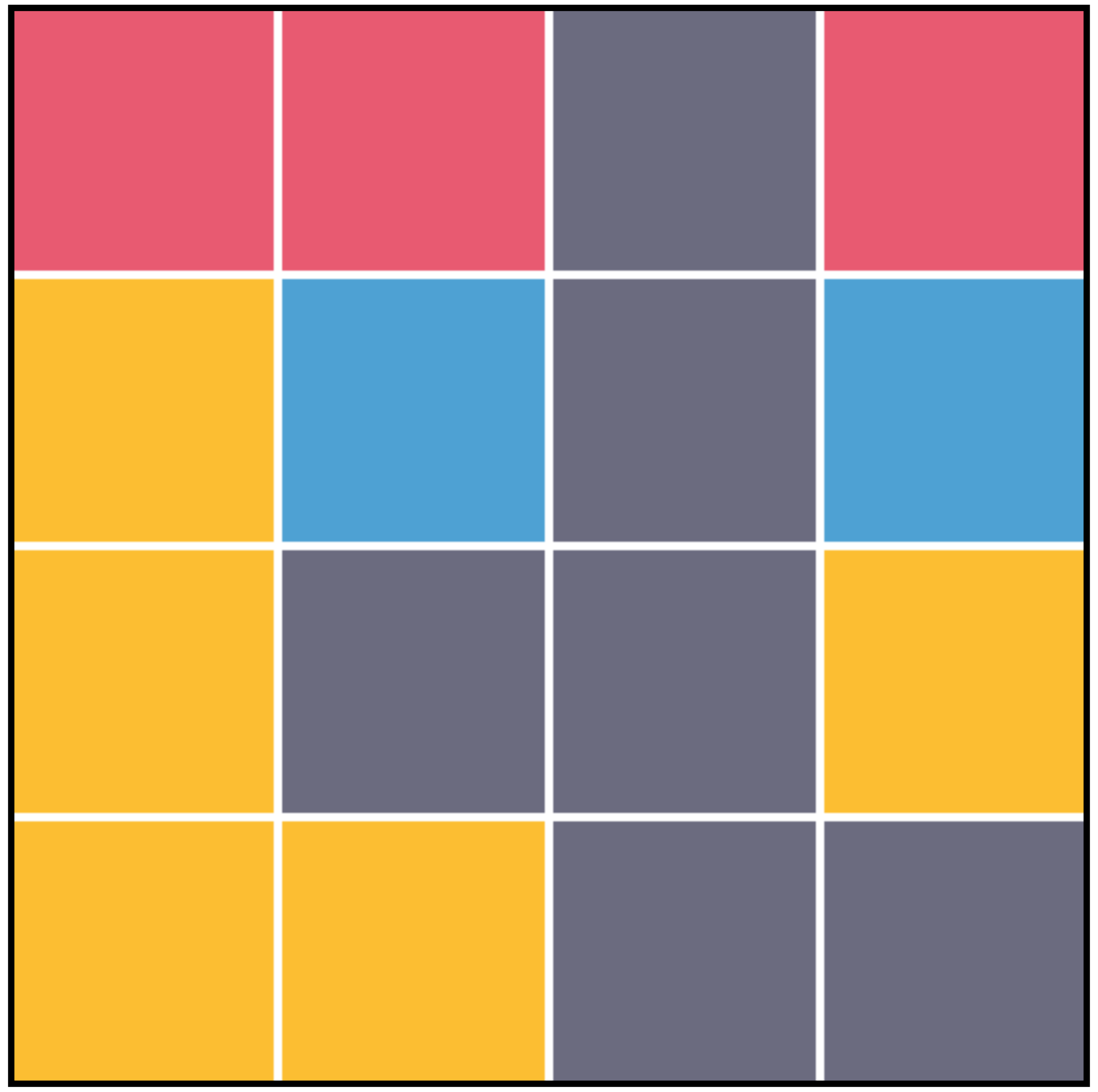
$$\text{likelihood}(D|L) \propto \prod_{\langle m,s\rangle \in D} \frac{1}{|M|} P(s|L,m)$$

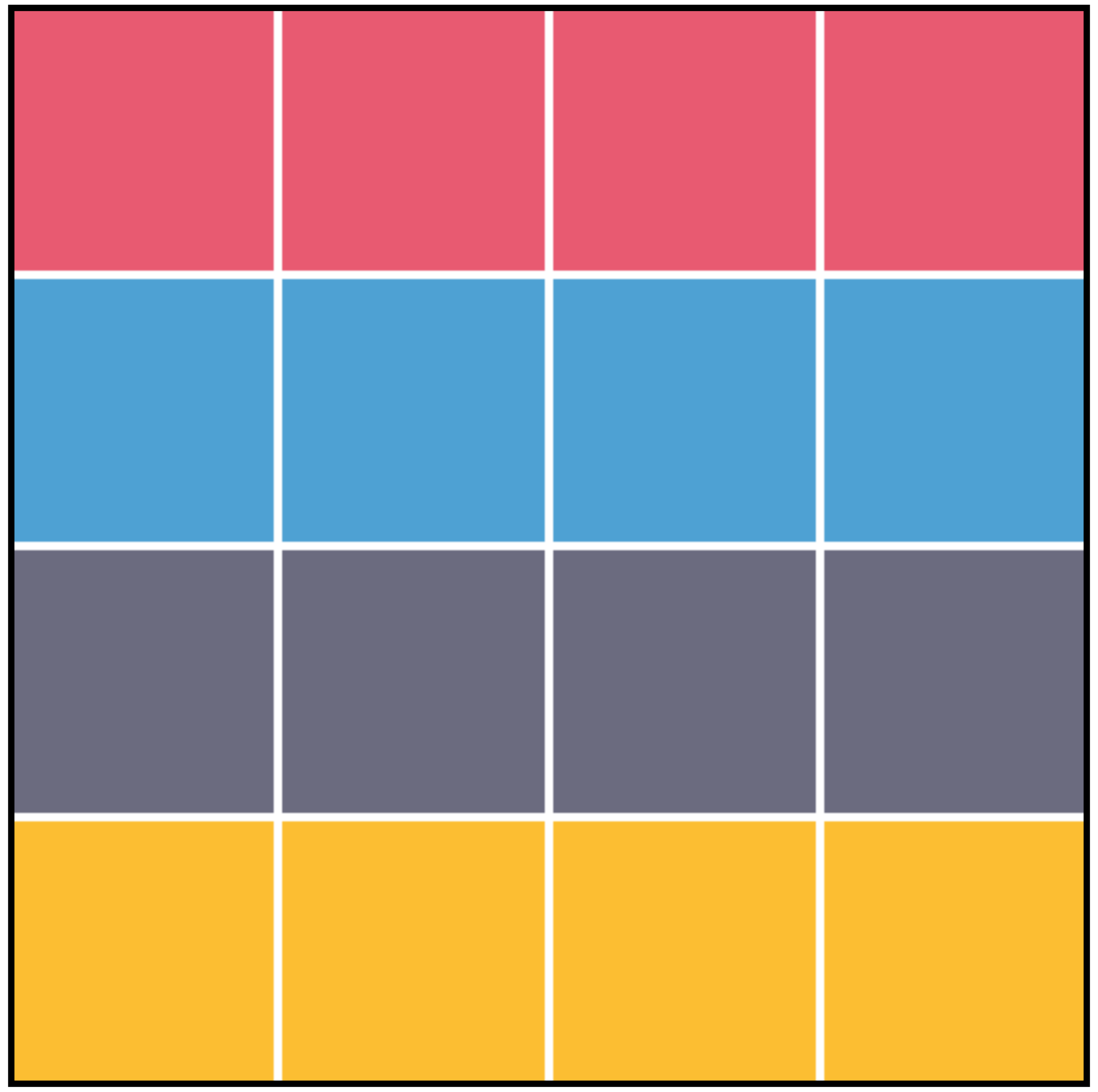
$$\text{prior}(L) \propto 2^{-\text{DL}(L)}$$

$$\text{posterior}(L|D) = \text{likelihood}(D|L) \times \text{prior}(L)$$



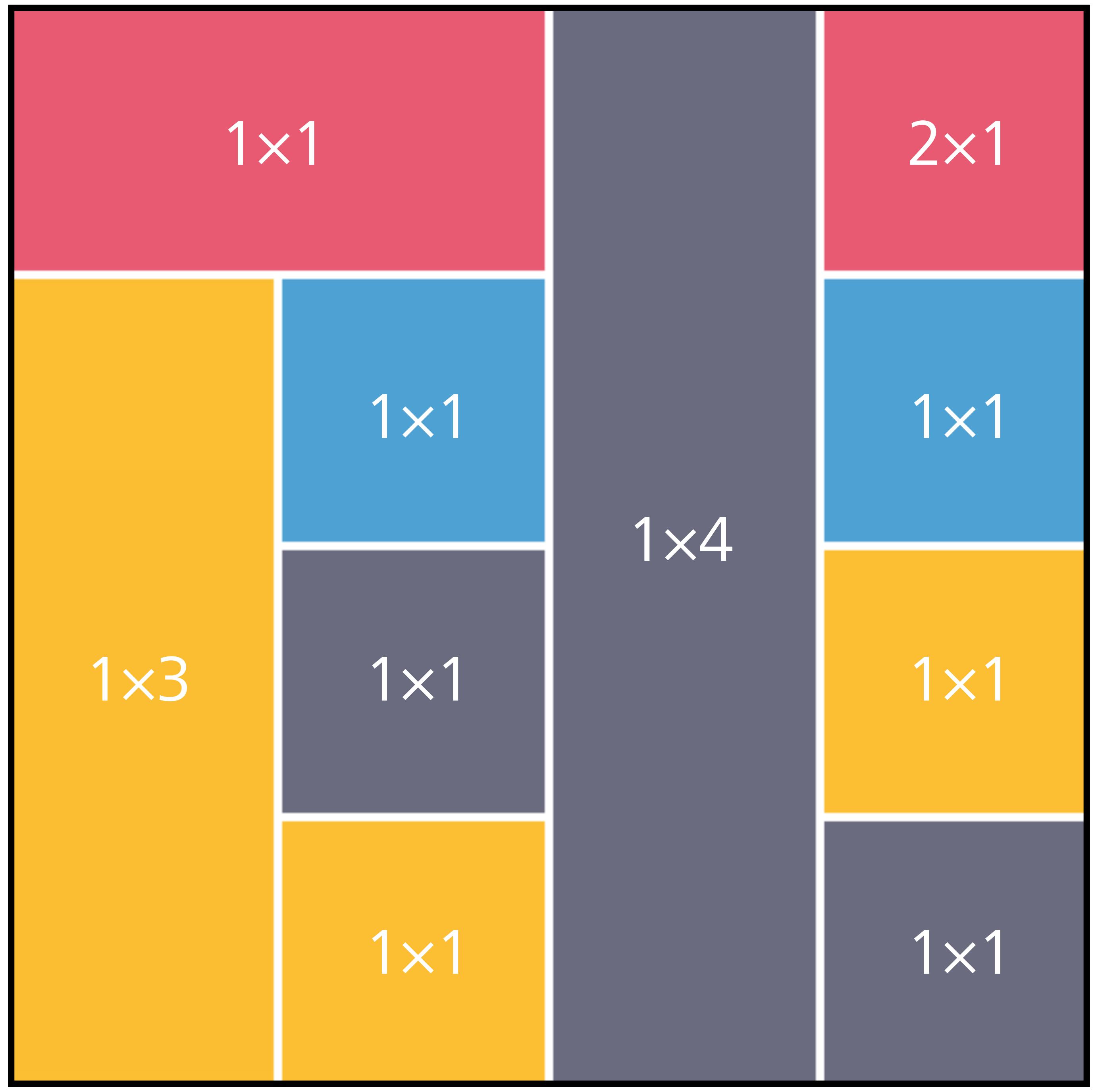
Computing DL(L): The rectangle code



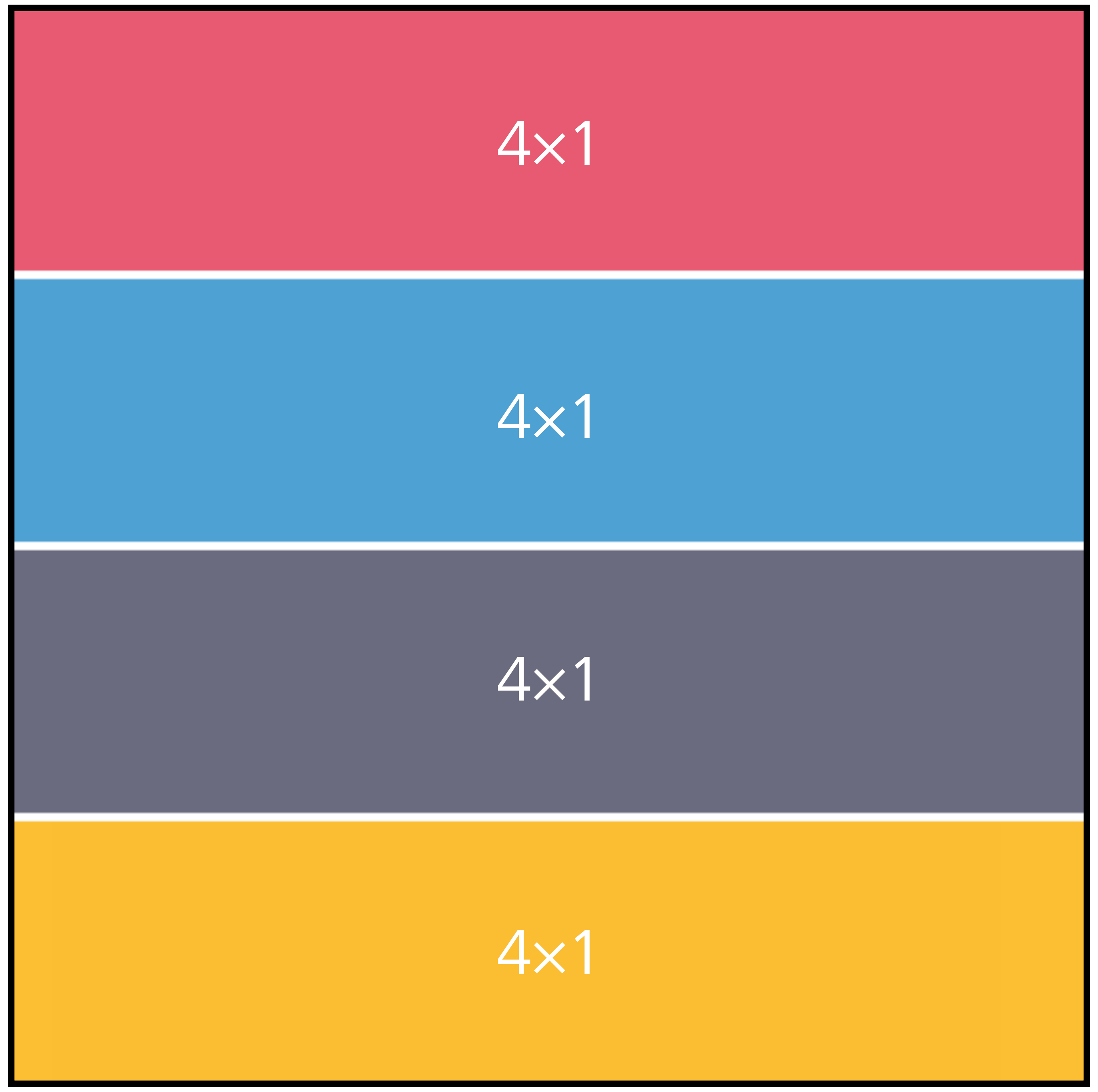


Fast & Feldman (2002)

Computing DL(*L*): The rectangle code



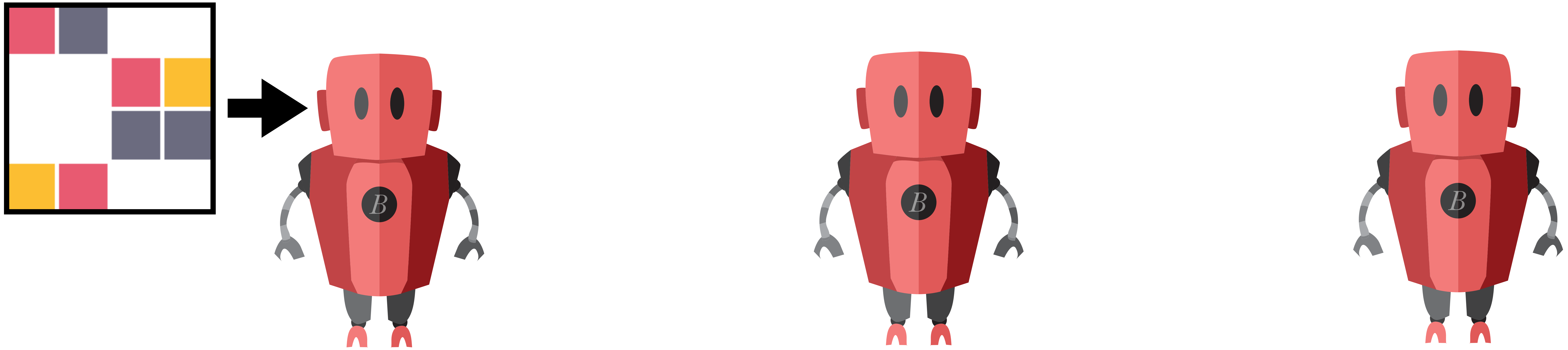
76.58 bits



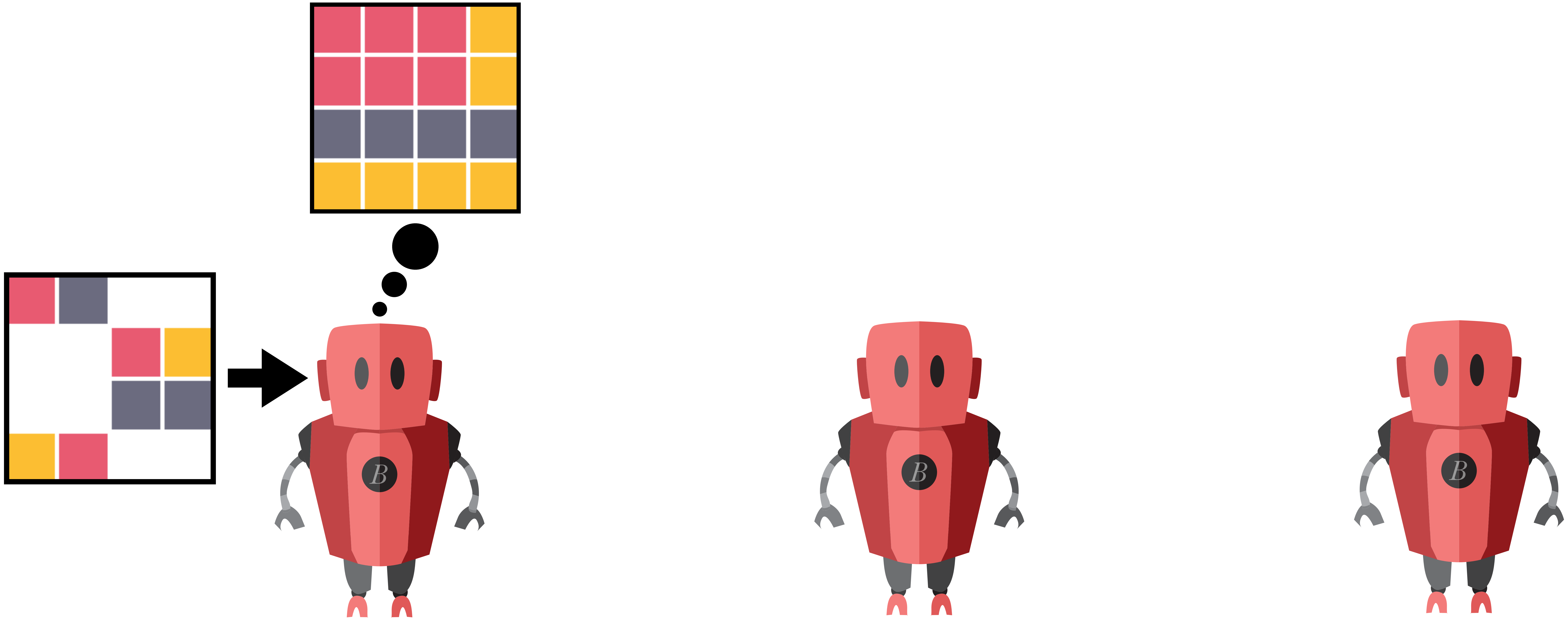
24 bits

Fast & Feldman (2002)

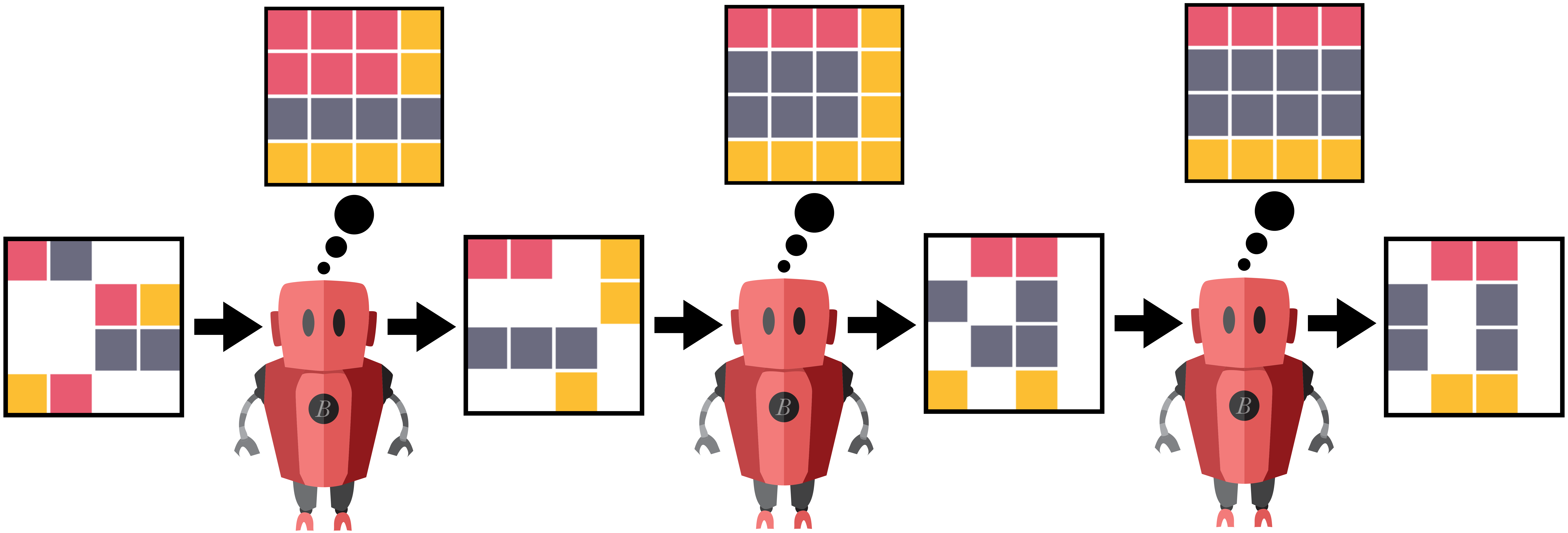
Bayesian iterated learning

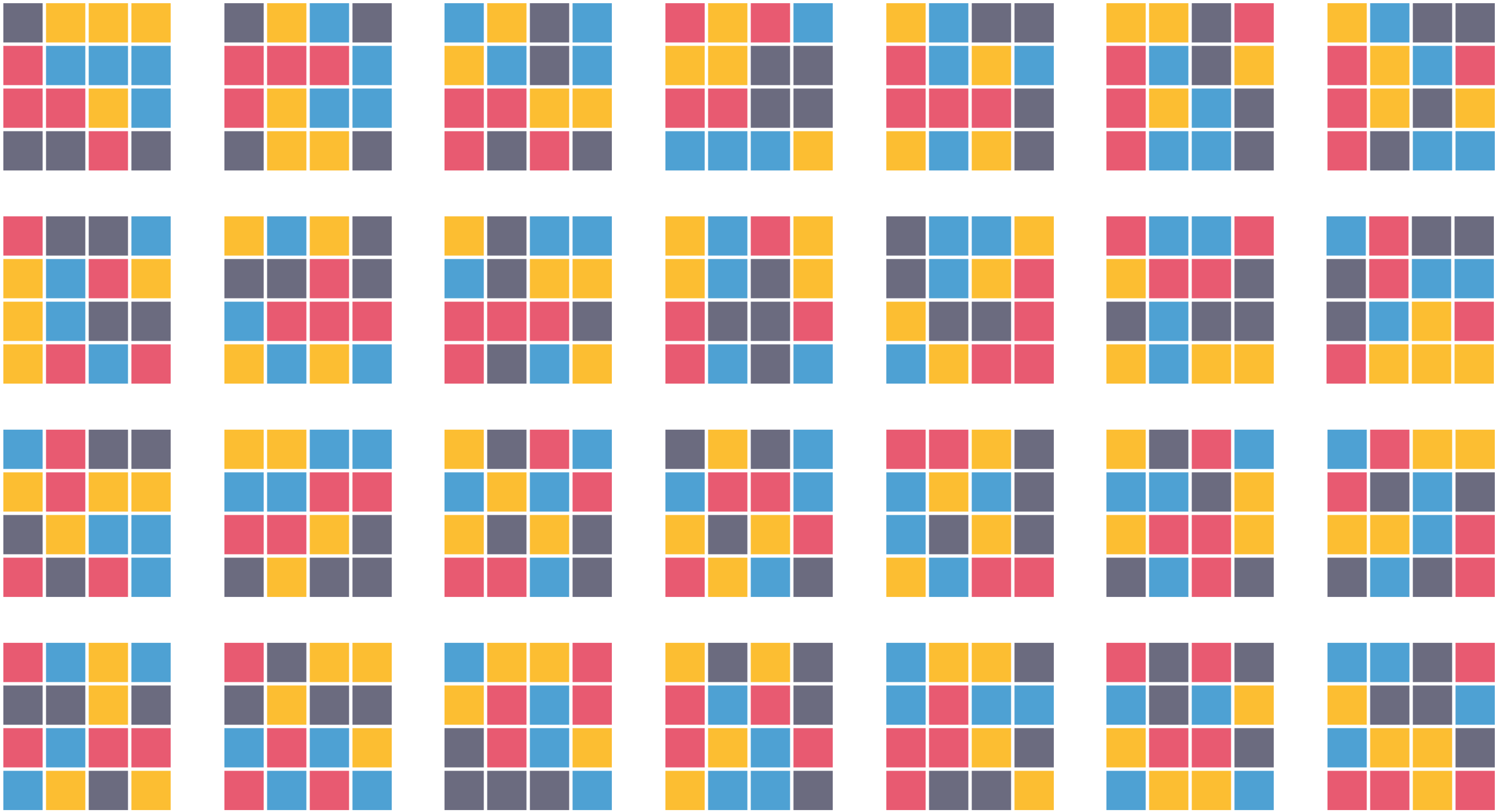


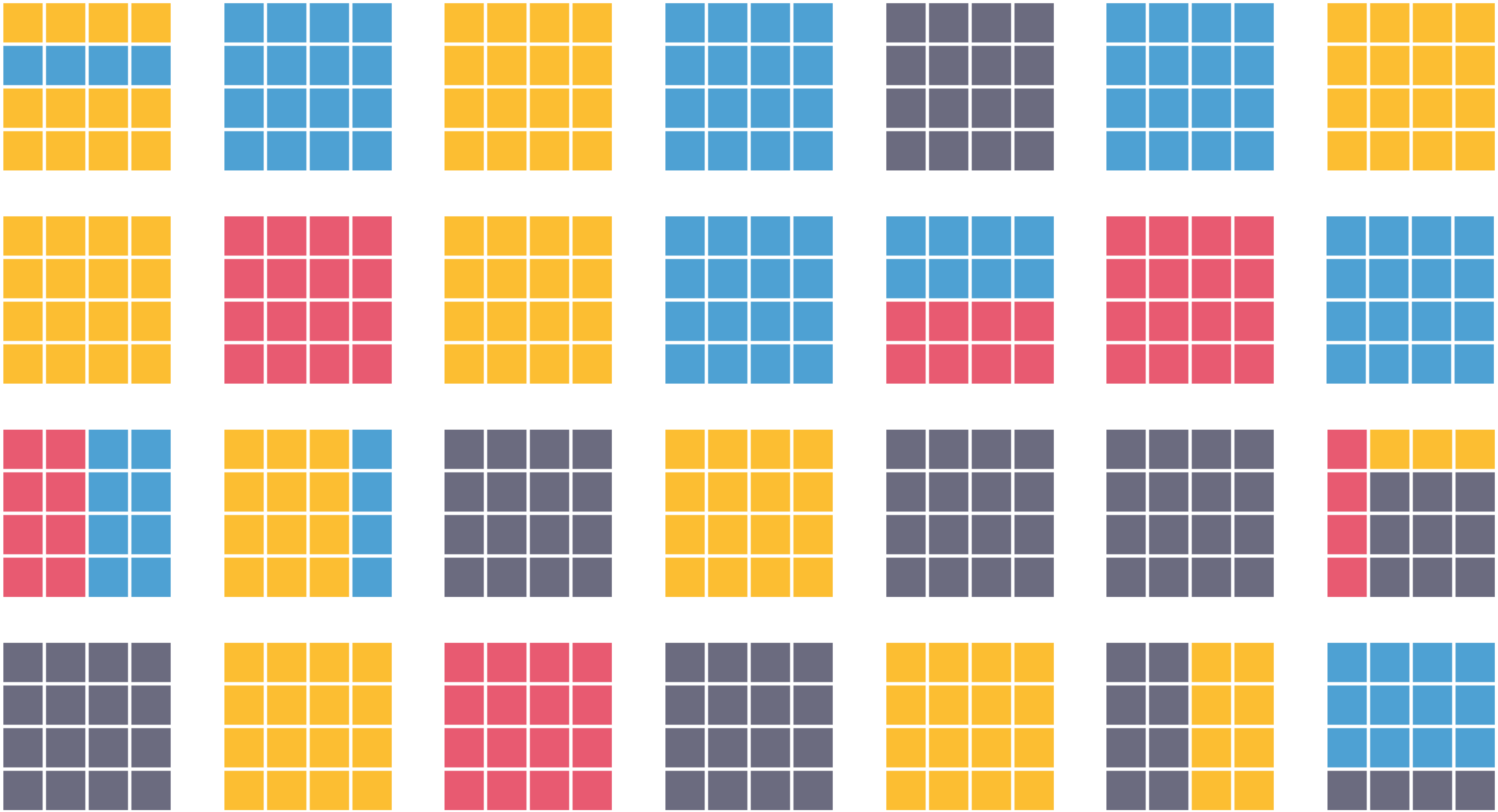
Bayesian iterated learning



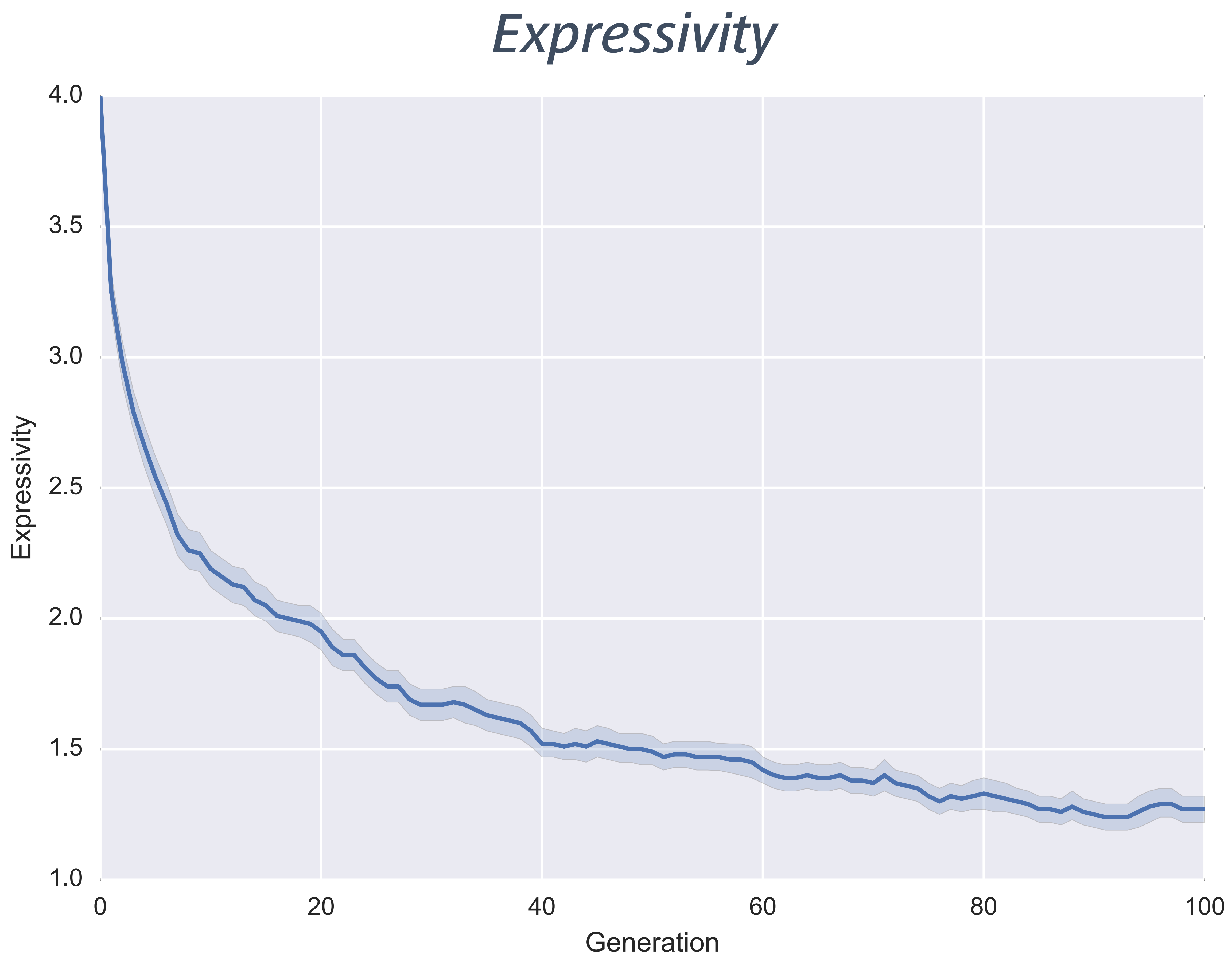
Bayesian iterated learning

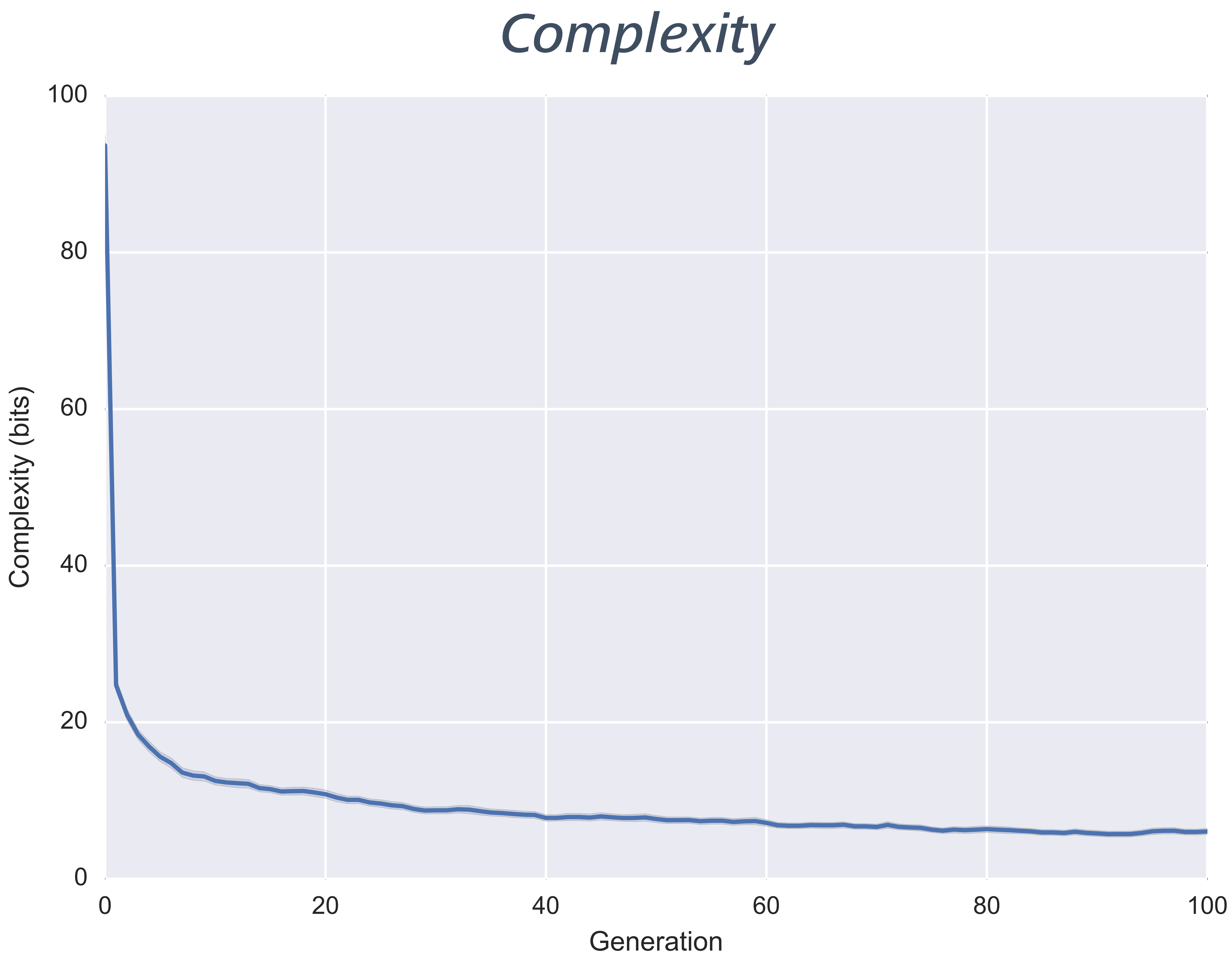






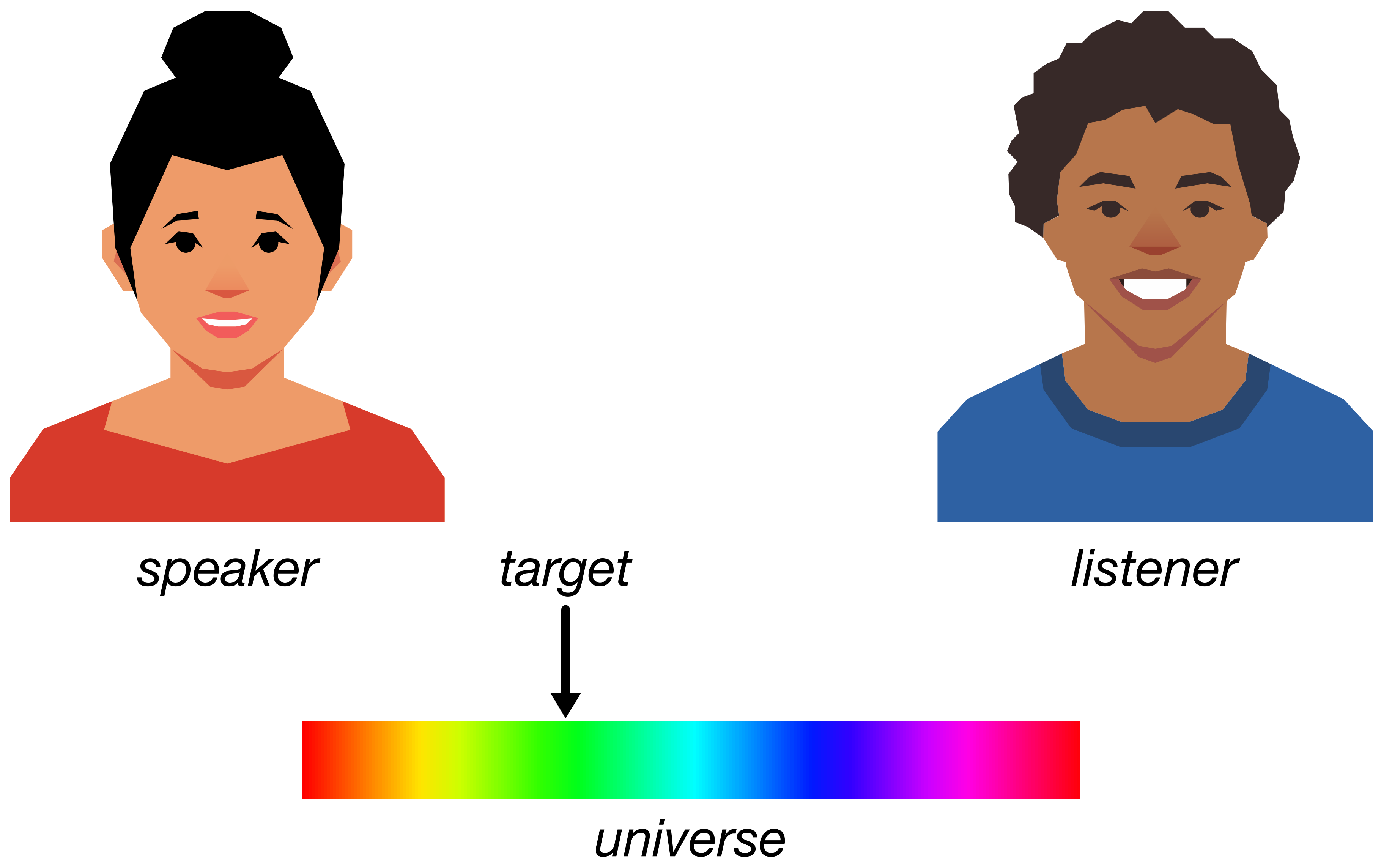
Iterated learning converges to the prior



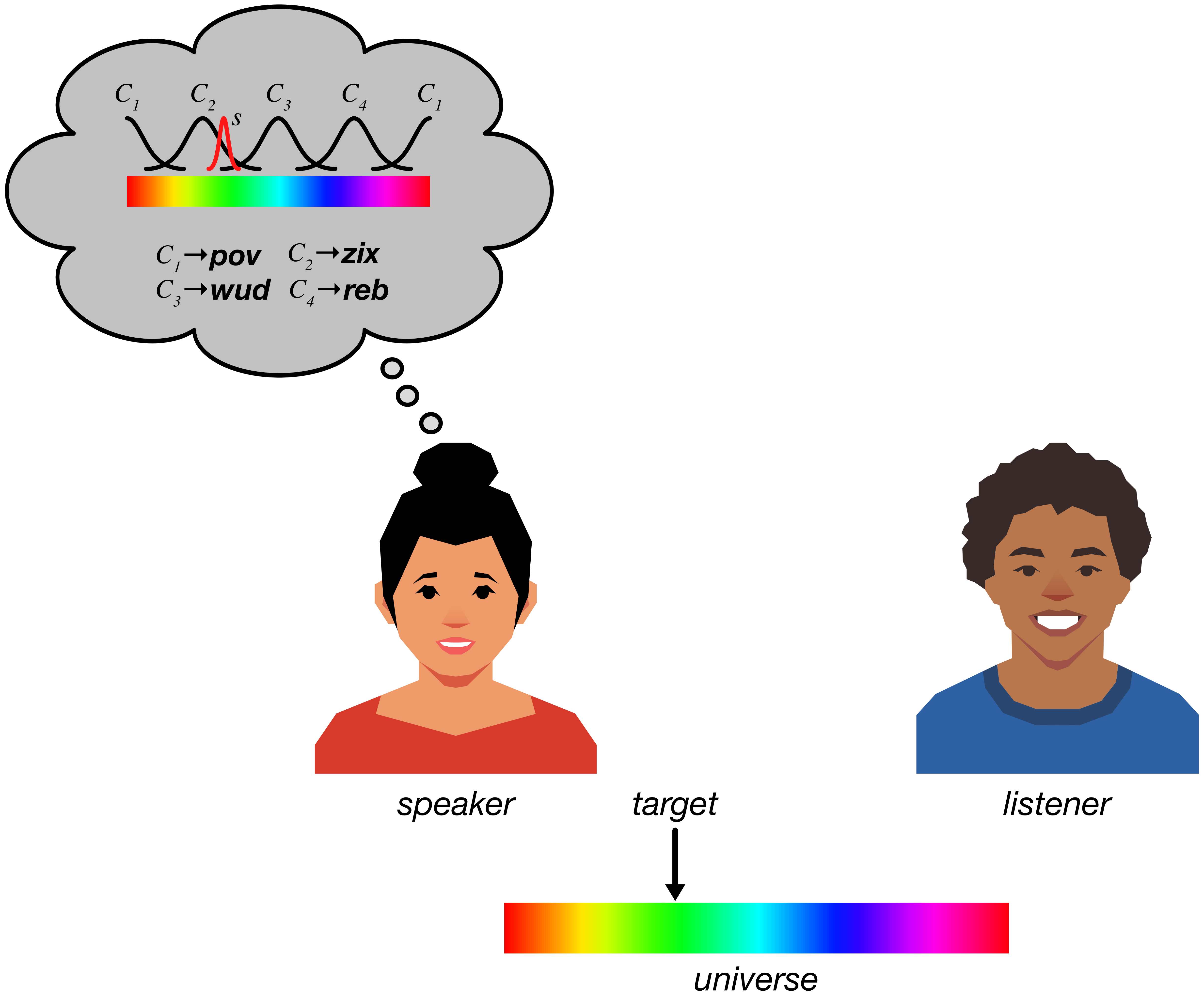


Informativeness

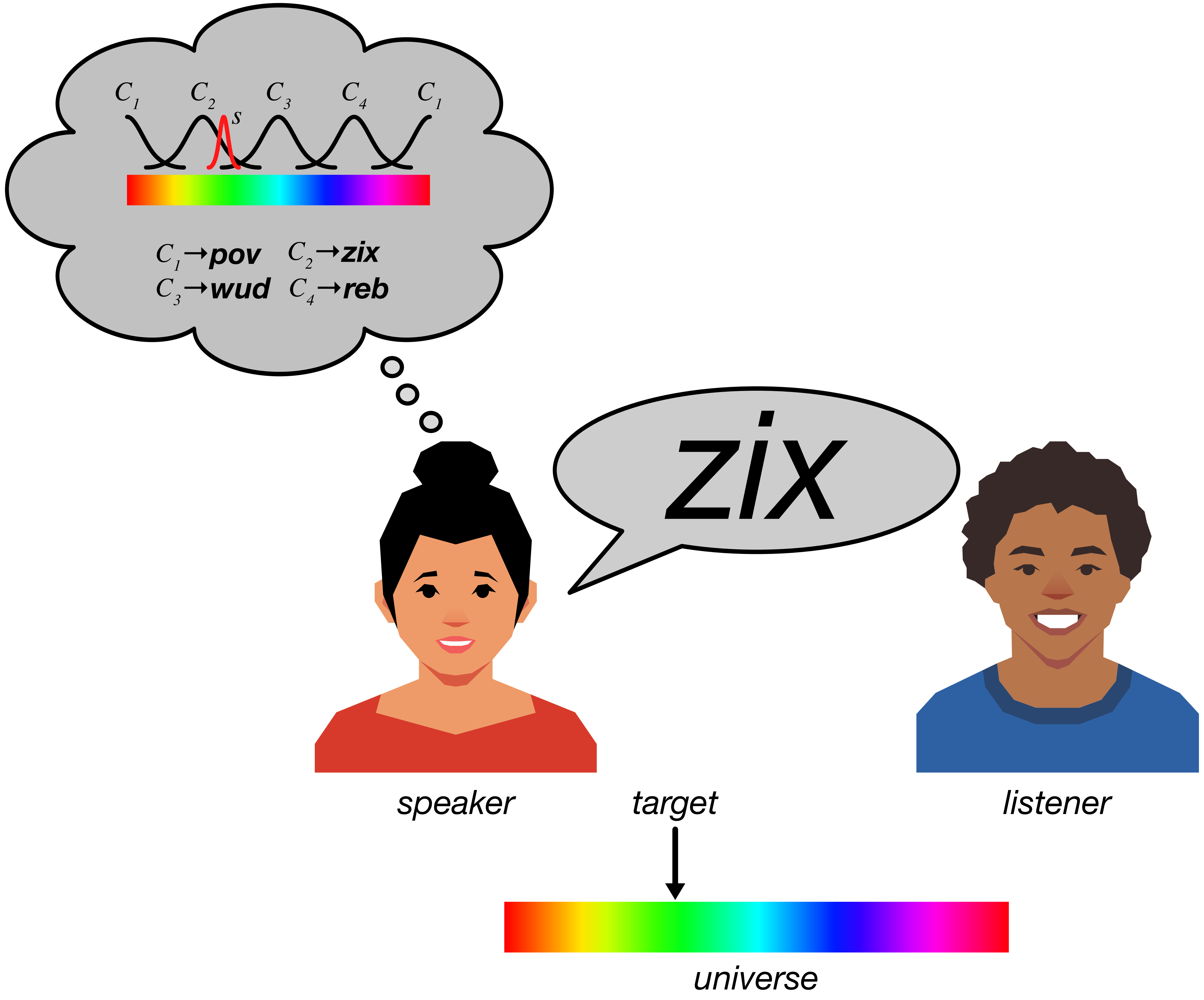
Regier et al.’s informativeness model



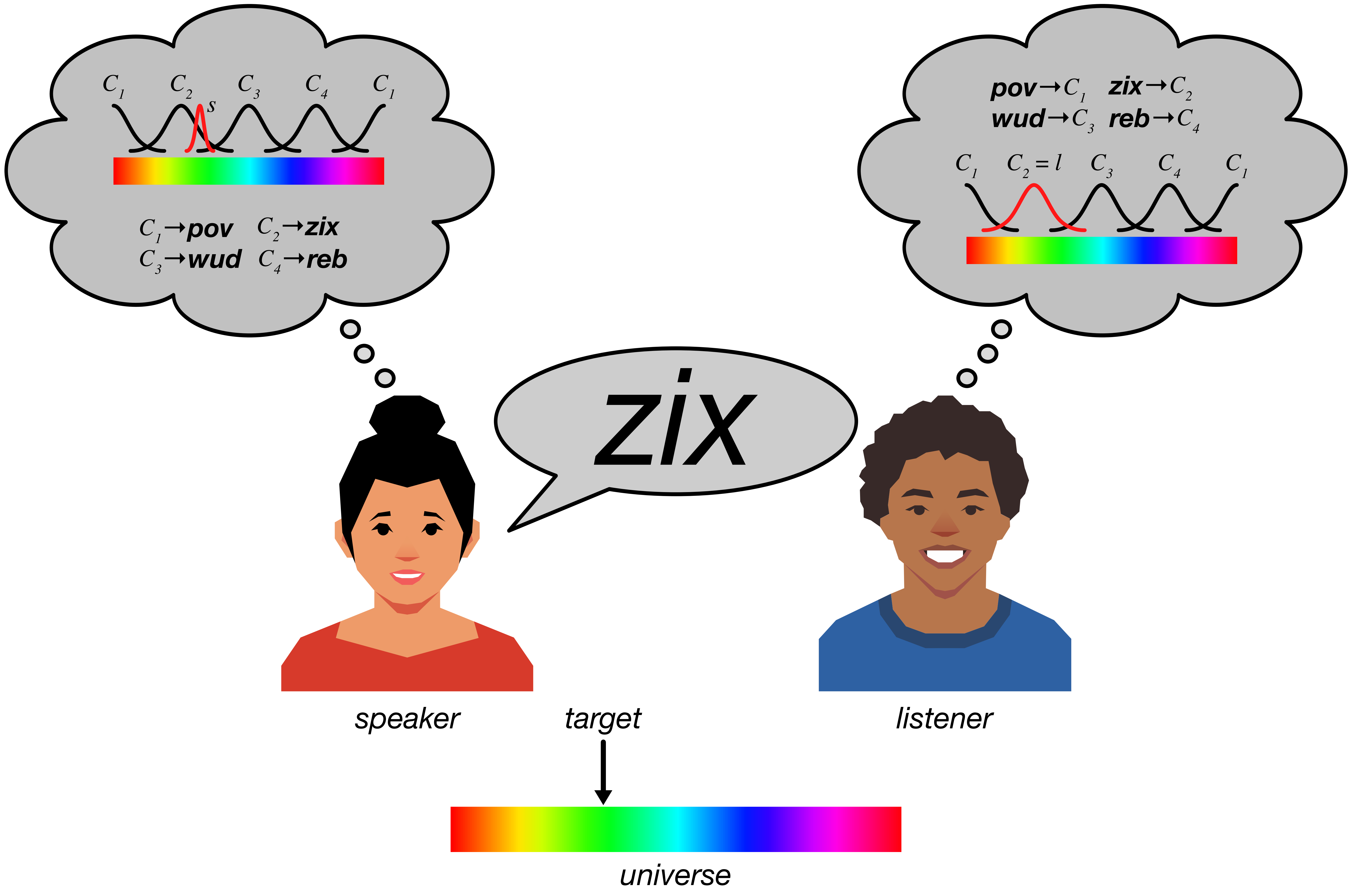
Regier et al.'s informativeness model



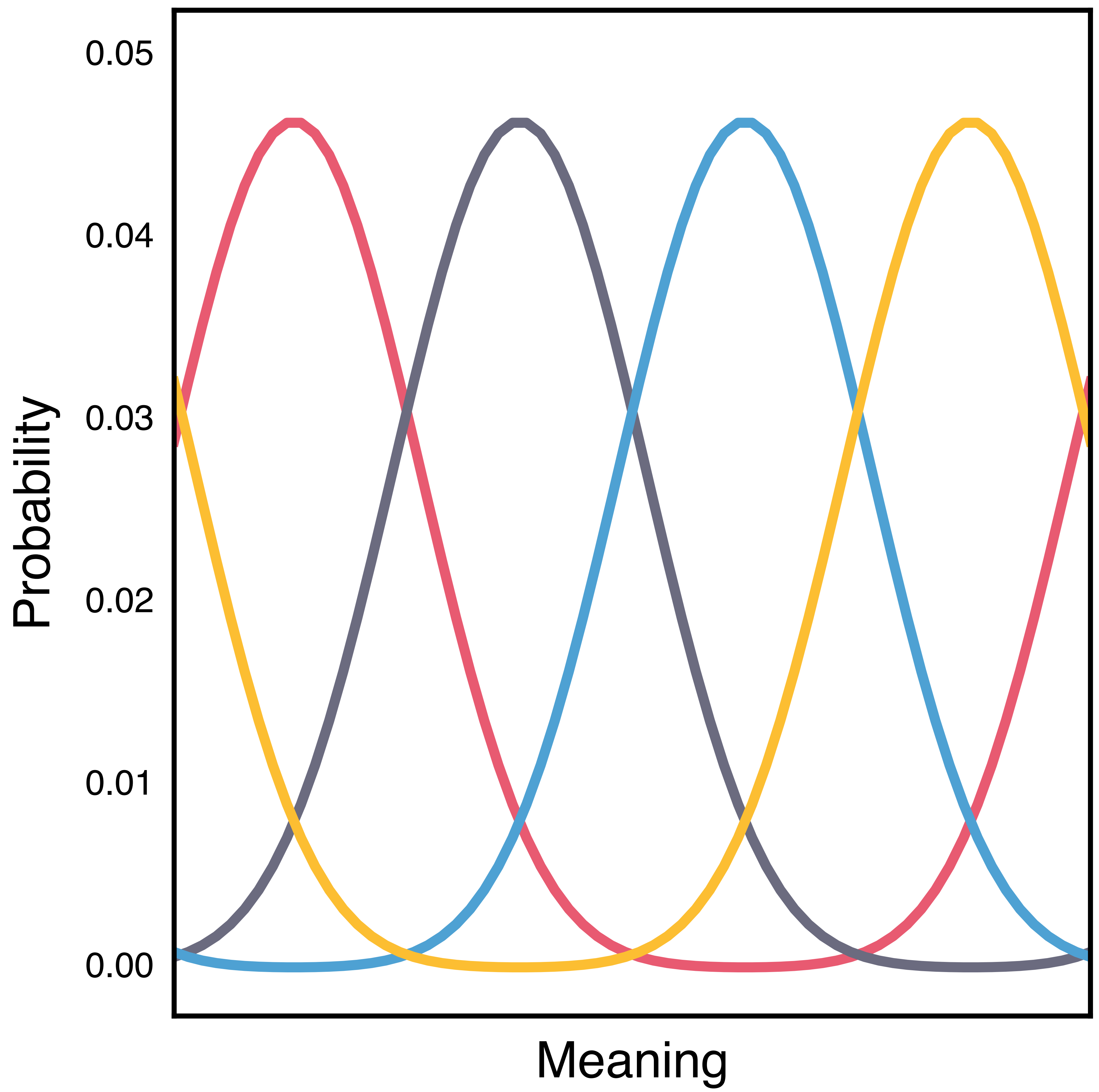
Regier et al.'s informativeness model



Regier et al.'s informativeness model



Communicative cost



$$C_j(i) \propto \sum_{c \in C_j} e^{-\gamma d(i,c)^2}$$

$$K(\mathcal{L}) := \sum_{i \in U} P(i) \cdot -\log C(i)$$

Expressivity A system of many categories is more informative than a system of few categories

Compactness A system of compact categories is more informative than a system of noncompact categories

Can iterated learning give rise to informative languages?

Language evolution in the lab tends toward informative communication

Alexandra Carstensen[1] (abc@berkeley.edu)
Jing Xu[4] (jing.xu@jhmi.edu)
Cameron T. Smith[2] (vmpfc1@berkeley.edu)
Terry Regier[2,3] (terry.regier@berkeley.edu)

Department of Psychology,[1] Department of Linguistics,[2] Cognitive Science Program[3]
University of California, Berkeley, CA 94720 USA

Department of Neurology, Johns Hopkins University, Baltimore, MD 21287 USA[4]

Abstract

Why do languages parcel human experience into categories in the ways they do? Languages vary widely in their category systems but not arbitrarily, and one possibility is that this constrained variation reflects universal communicative needs. Consistent with this idea, it has been shown that attested category systems tend to support highly informative communication. However it is not yet known what process *produces* these informative systems. Here we show that human simulation of cultural transmission in the lab produces systems of semantic categories that converge toward greater informativeness, in the domains of color and spatial relations. These findings suggest that larger-scale cultural transmission over historical time could have produced the diverse yet informative category systems found in the world's languages.

Keywords: Informative communication, language evolution, iterated learning, cultural transmission, spatial cognition, color naming, semantic universals.

The origins of semantic diversity

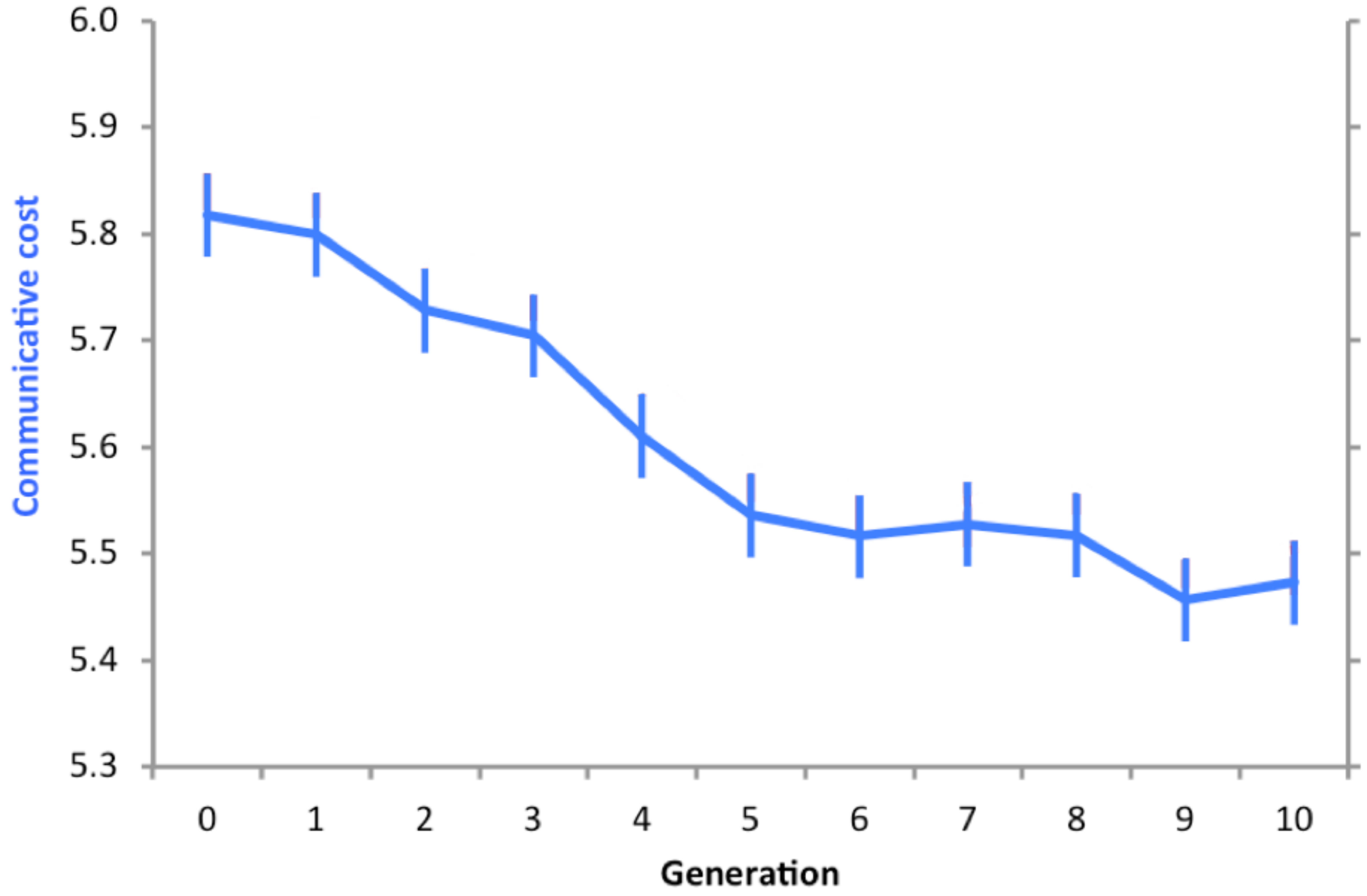
Languages vary widely in their fundamental units of meaning—the concepts and categories they encode in single

Regier's (2012) kinship study, Levinson (2012) pointed out that although that research explains cross-language semantic variation in communicative terms, it does not tell us "where our categories come from" (p. 989); that is, it does not establish what *process* gives rise to the diverse attested systems of informative categories. Levinson suggested that a possible answer to that question may lie in a line of experimental work that explores human simulation of cultural transmission in the laboratory, and "shows how categories get honed through iterated learning across simulated generations" (p. 989). We agree that prior work explaining cross-language semantic variation in terms of informative communication has not yet addressed this central question, and we address it here.

Iterated learning and category systems

The general idea behind iterated learning studies is that of a chain or sequence of learners. The first person in the chain produces some behavior; the next person in the chain observes that behavior, learns from it, and then produces behavior of her own; that learned behavior is then observed by the next person in the chain, who learns from it, and so on. This experimental paradigm is meant to capture in

Can iterated learning give rise to informative languages?



Carstensen, Xu, Smith, Regier (2015)

Experiments

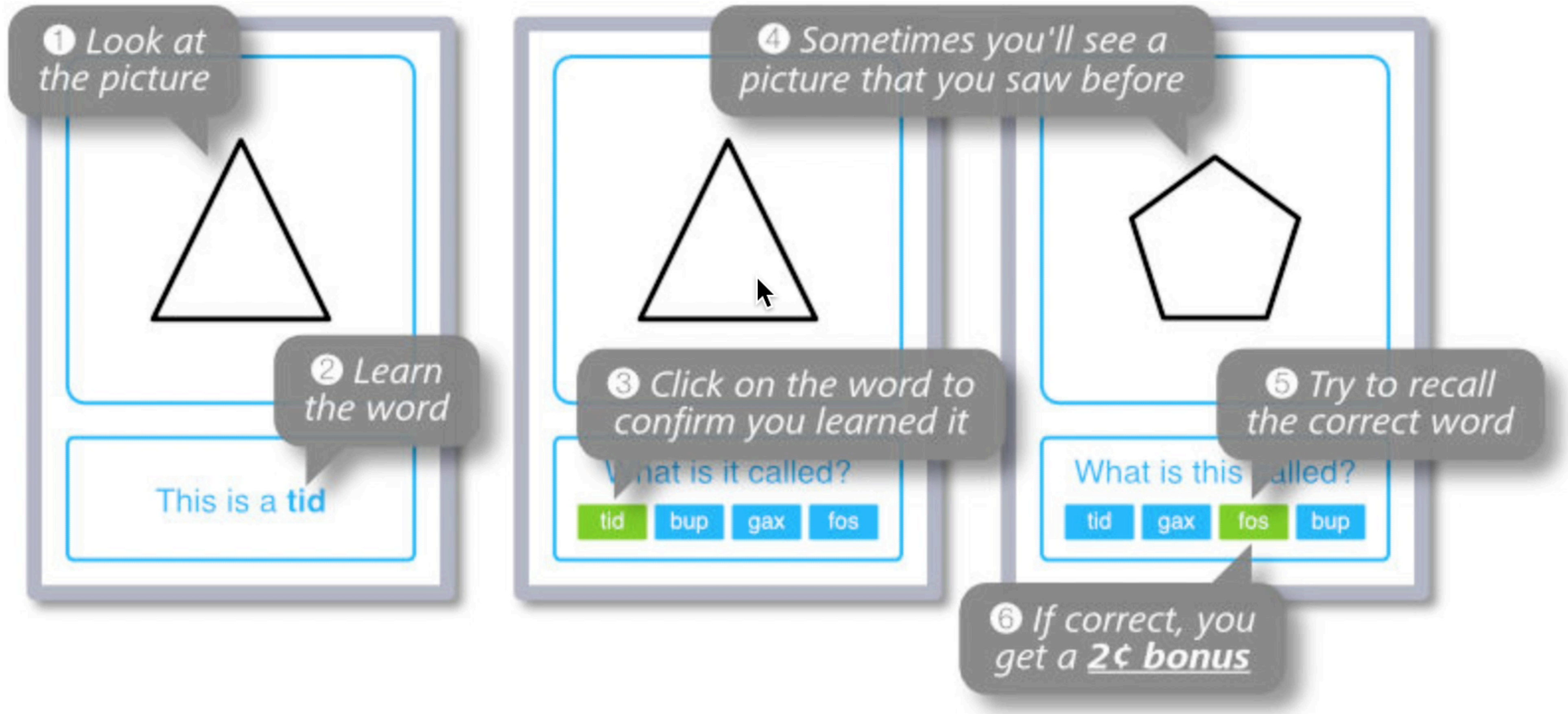
Training phase

localhost/~jon/shepard/

Stage 1: Training

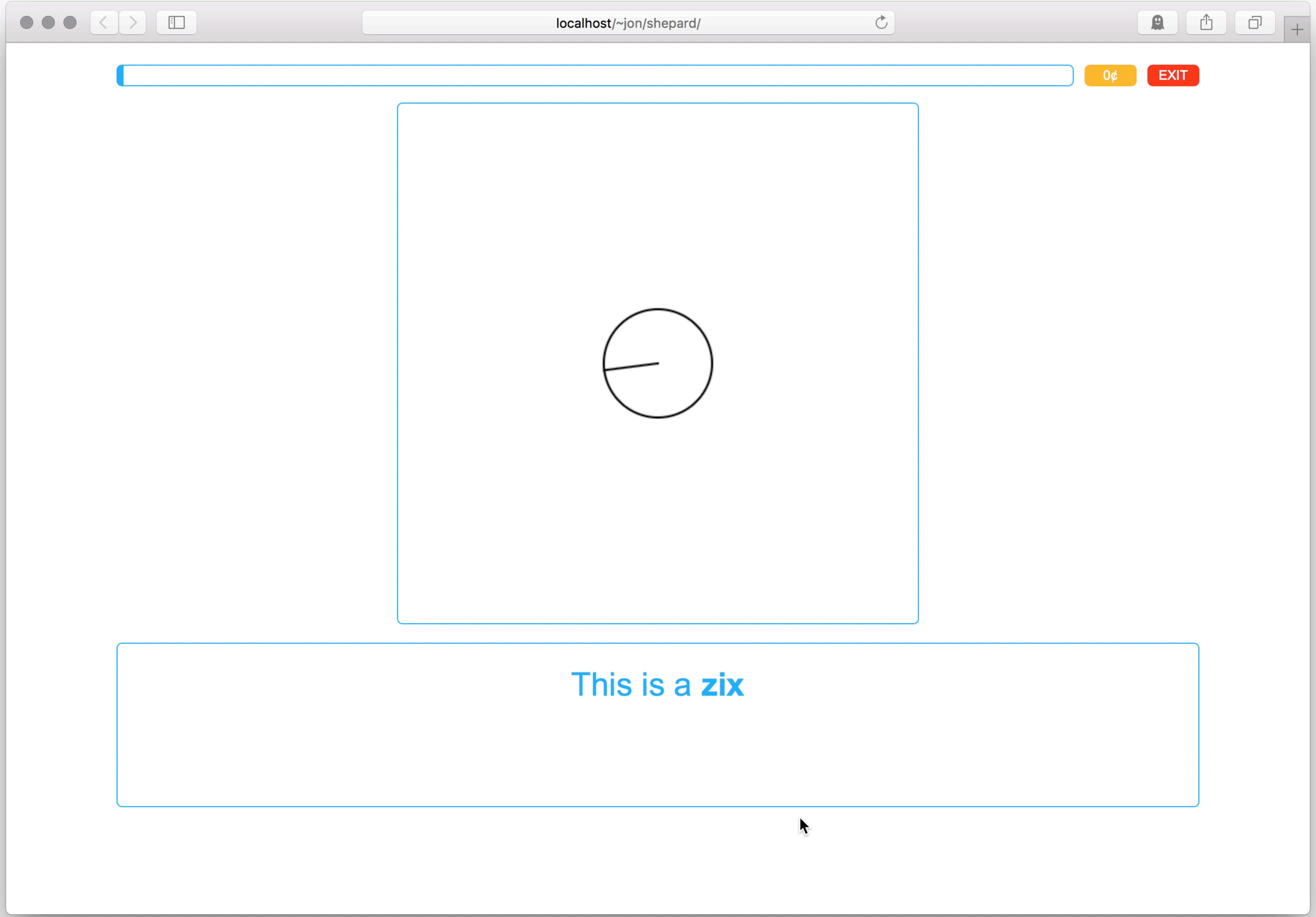
15 minutes

You are going to learn a simple language. **We will train you on 4 words** in the language and **we will test how well you are learning the words**. Try to learn the language as well as you can and **aim to be accurate in your answers**. You will receive a **2¢ bonus payment** for every correct test answer. If you decide to stop the task, please click the EXIT button so that someone else can take part.



START

Training phase



Test phase

localhost/~jon/shepard/

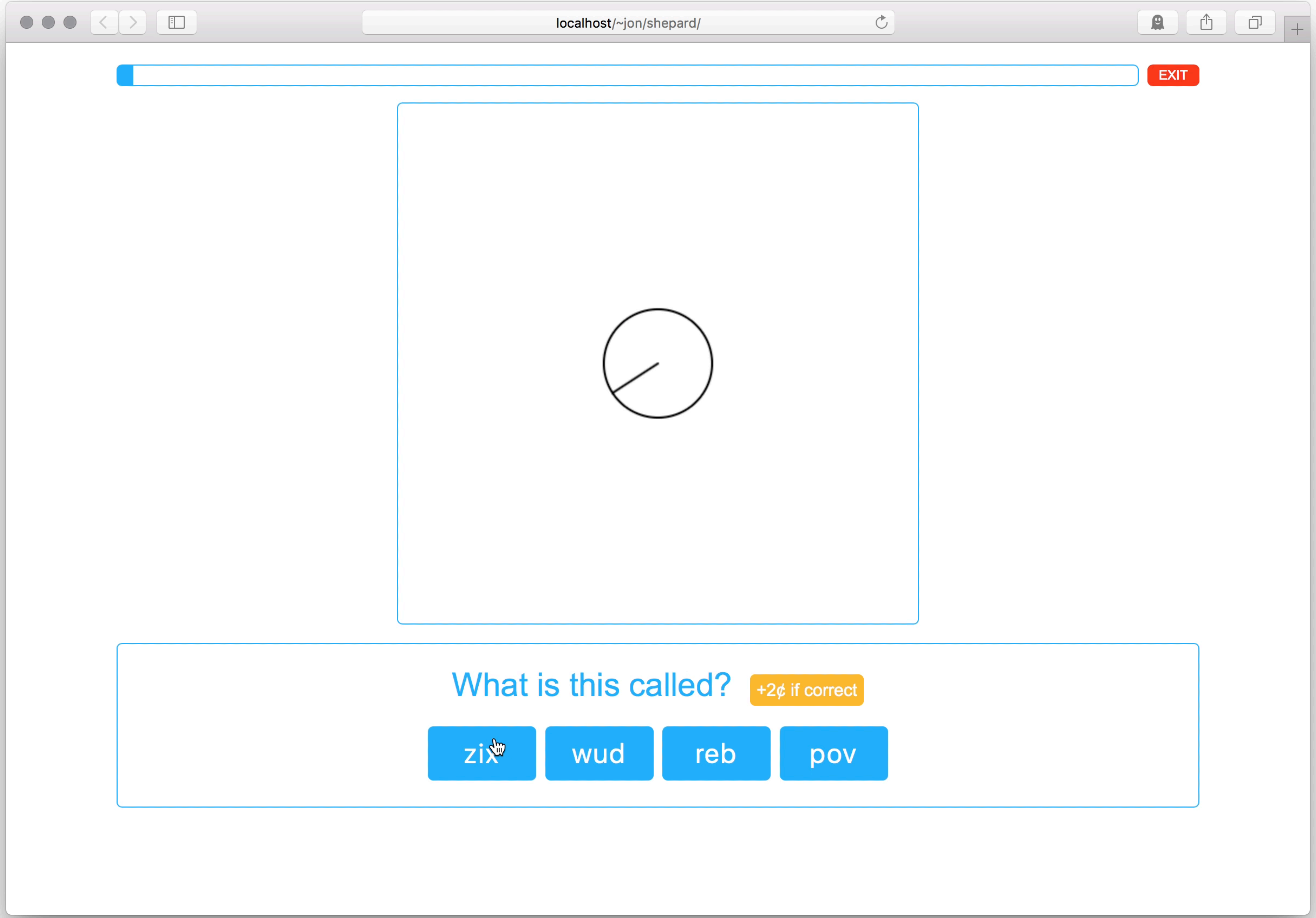
Stage 2: Test

5 minutes

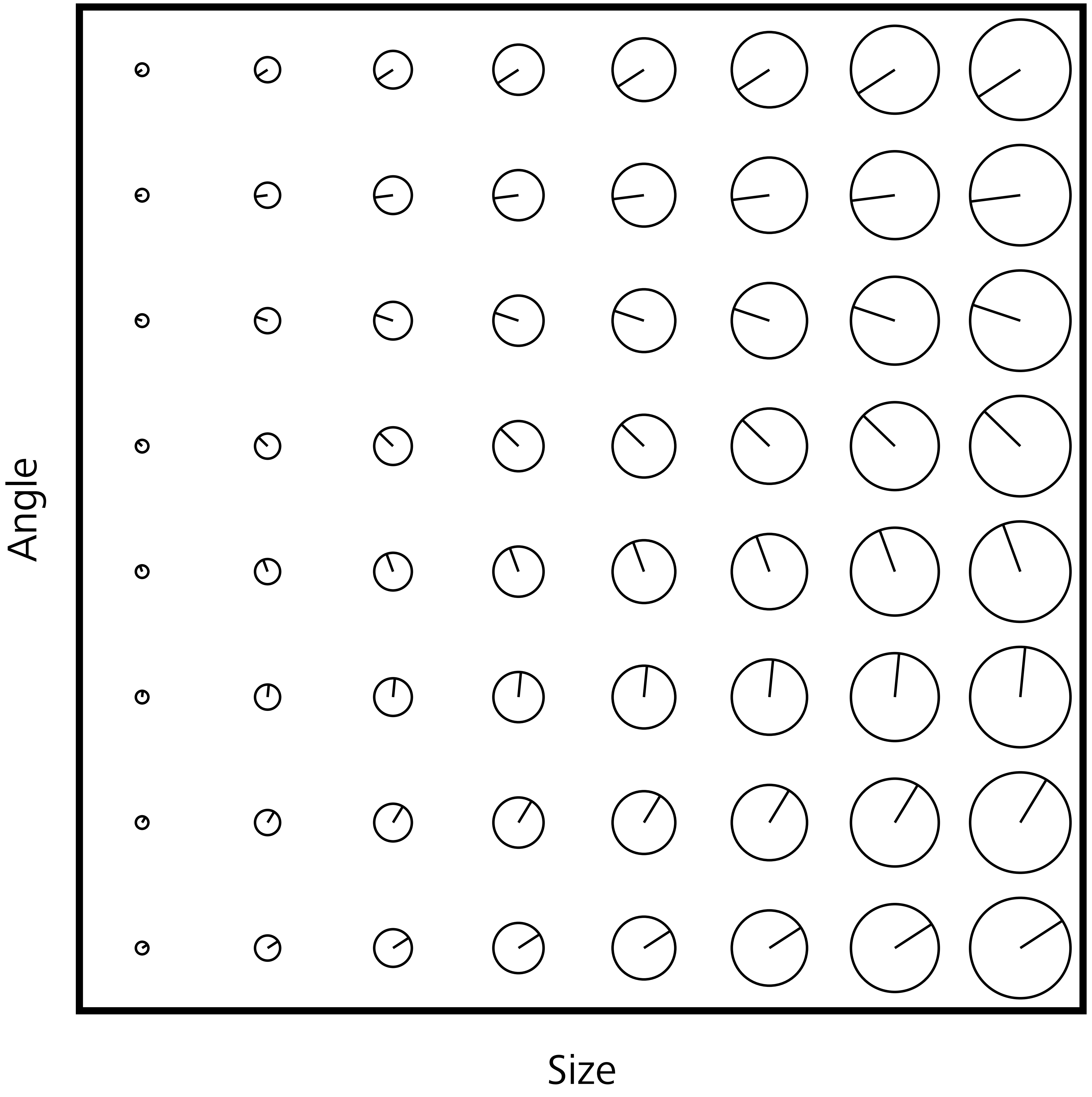
You have now completed the training stage! Next **we will test you on the language** that you just learned. For each picture, try to click on the correct word. You will get a **2¢ bonus payment** for every correct answer. It is therefore possible to earn **up to $1.28** in this stage of the task. However, this time **we will <u>not</u> tell you if you are correct or incorrect**. You will find out at the end how many you got correct.

START

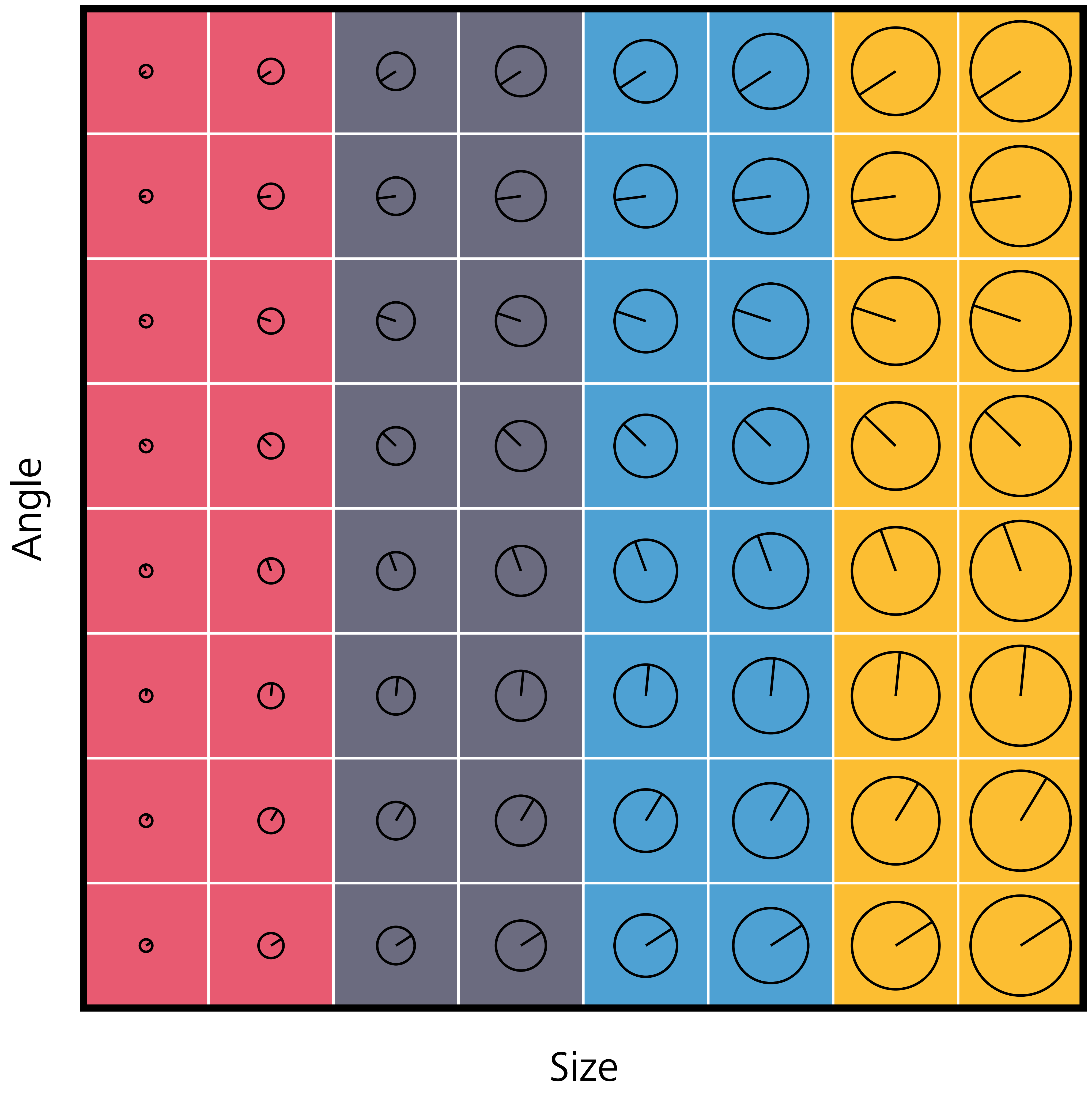
Test phase


localhost/~jon/shepard/
EXIT
What is this called?
+2¢ if correct
zix
wud
reb
pov

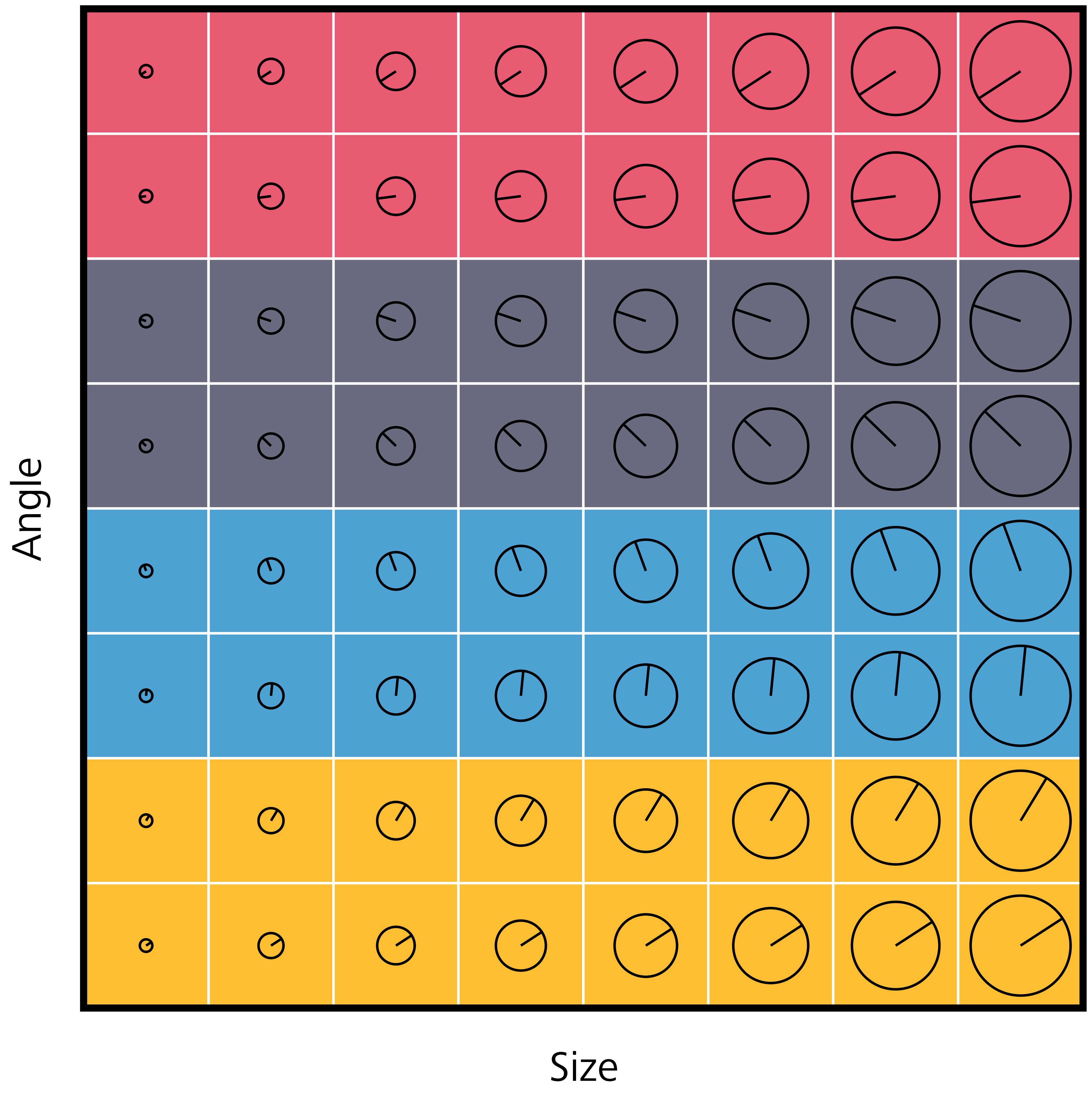
Stimuli



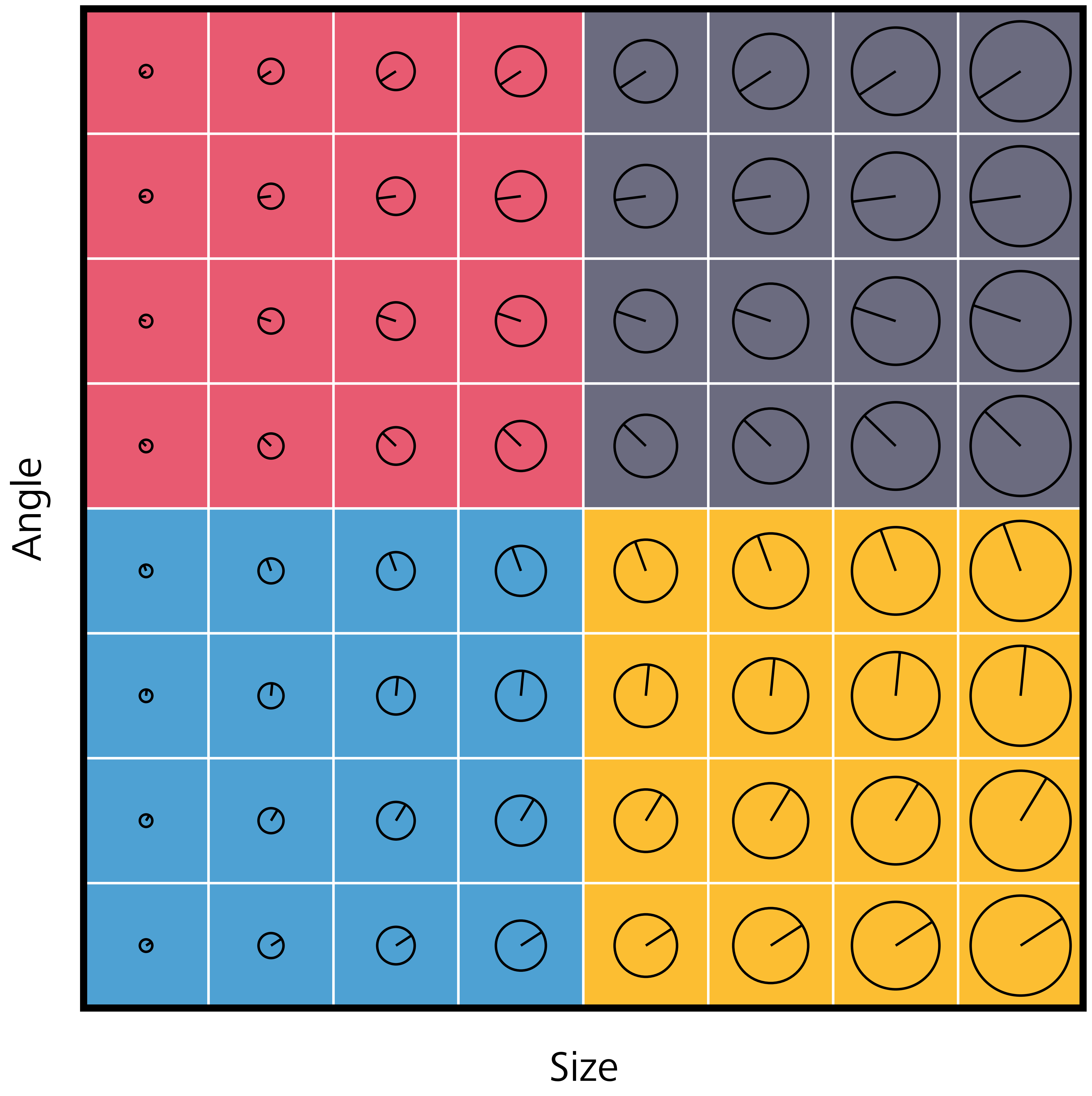
Stimuli



Stimuli

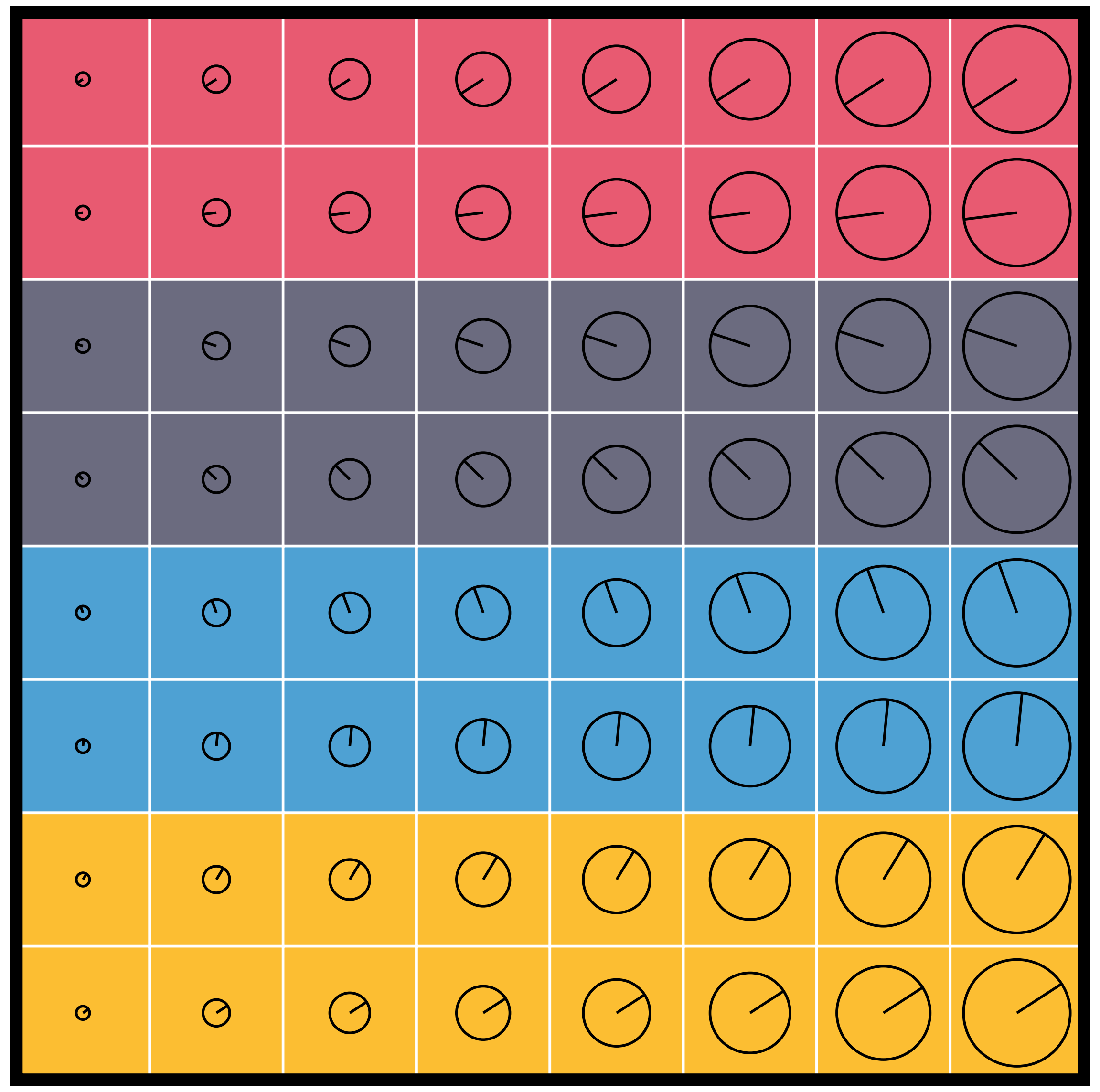


Stimuli

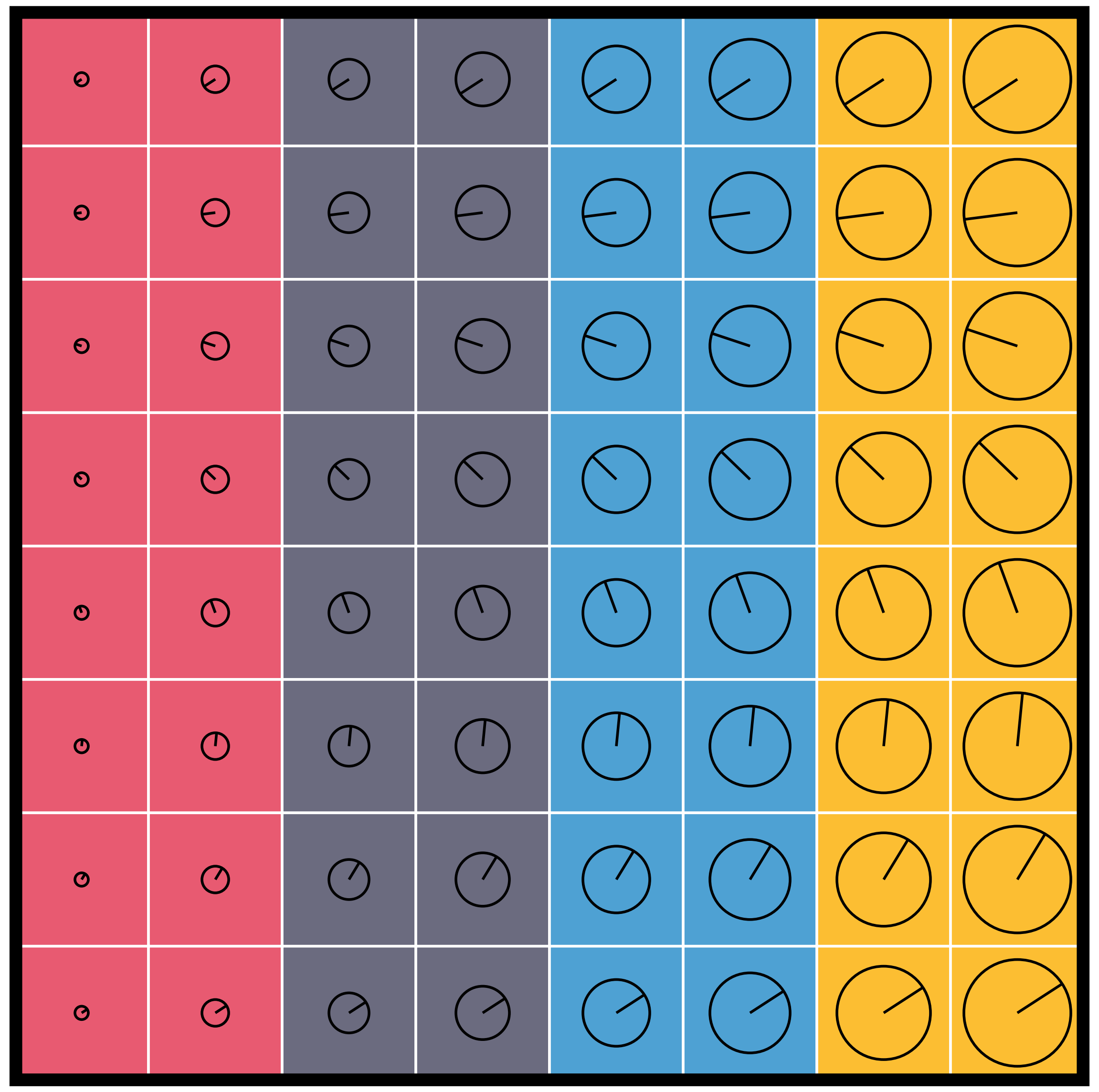


Which is easiest to learn?

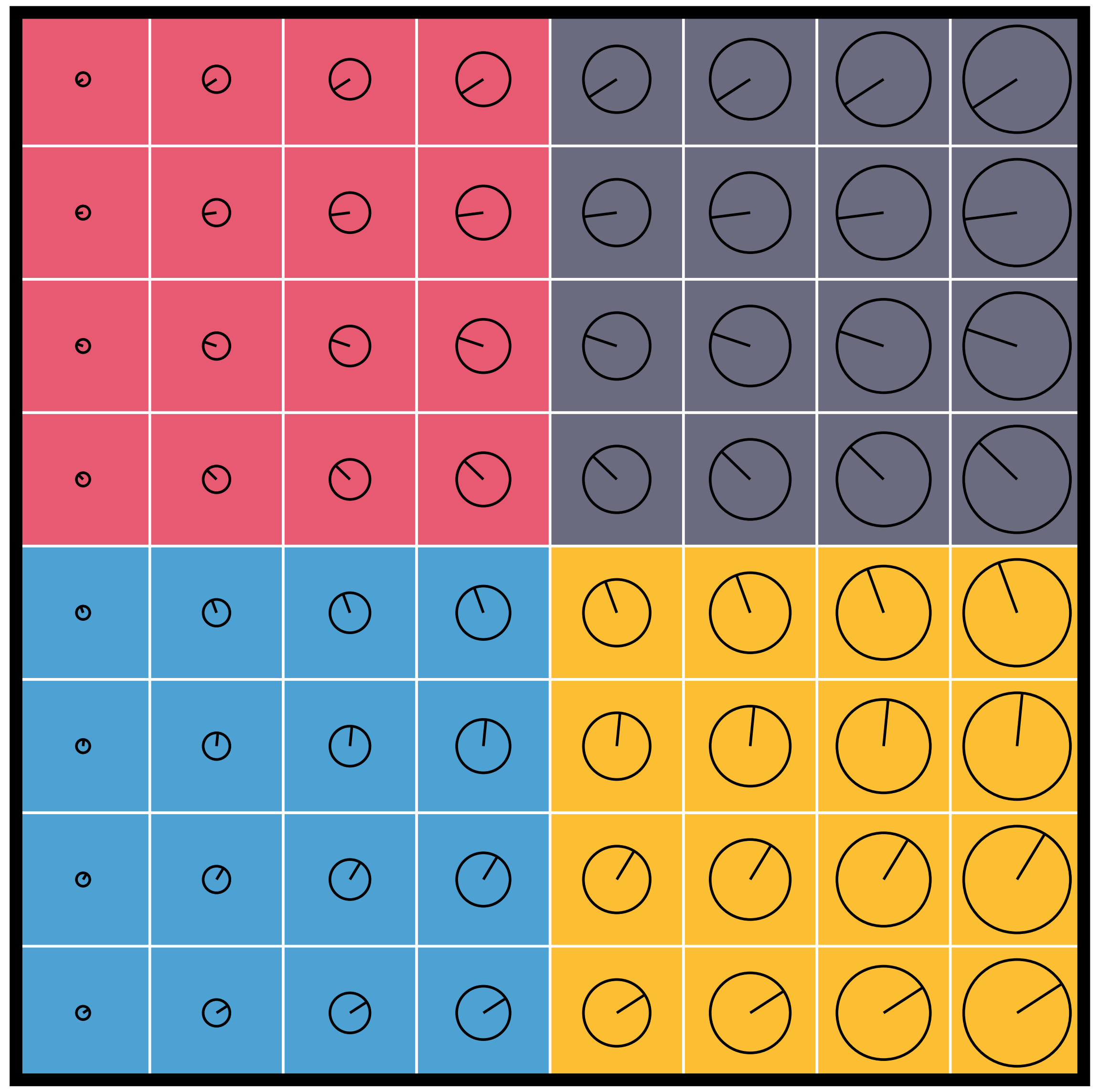
Angle-only



Size-only

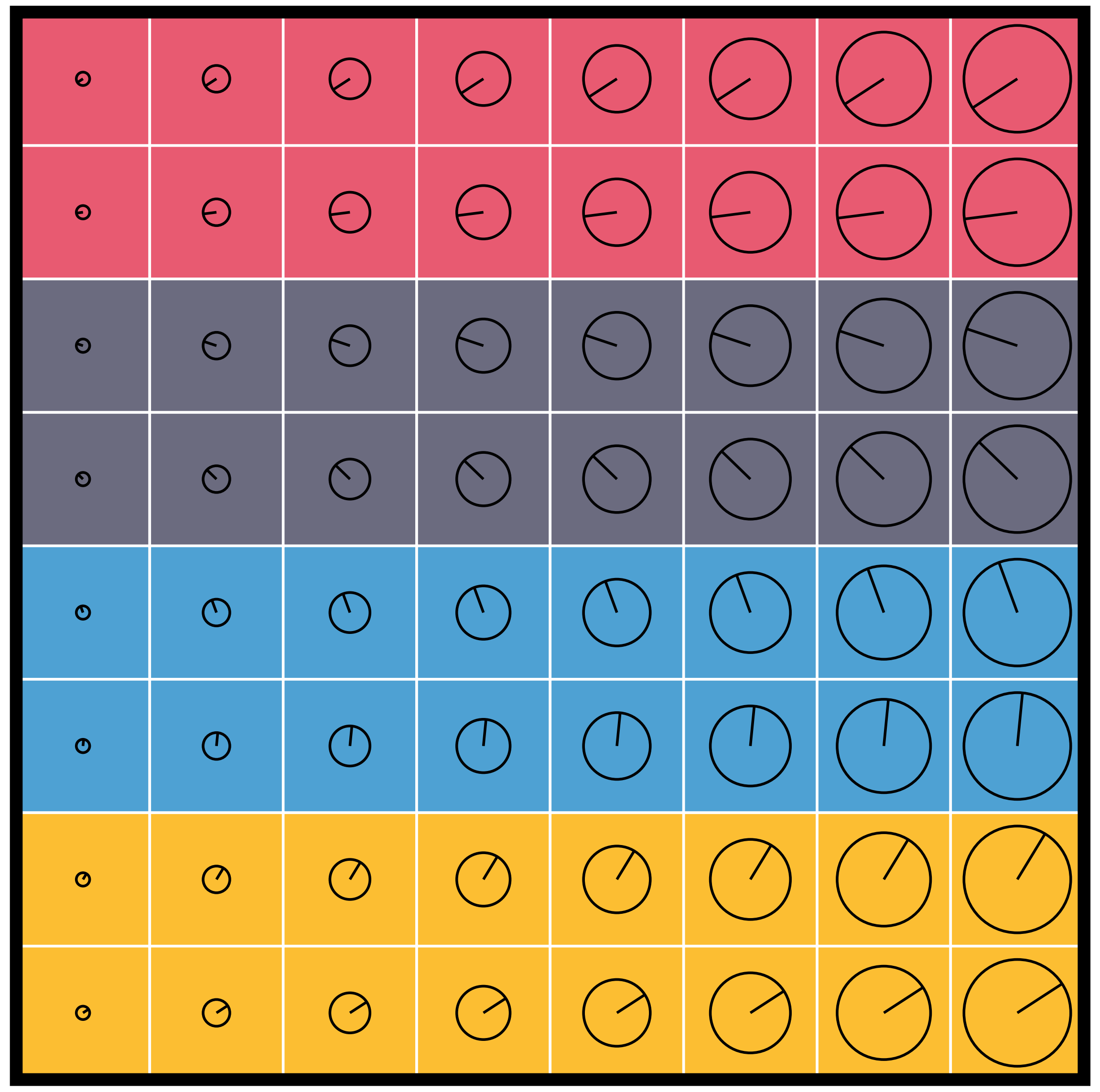


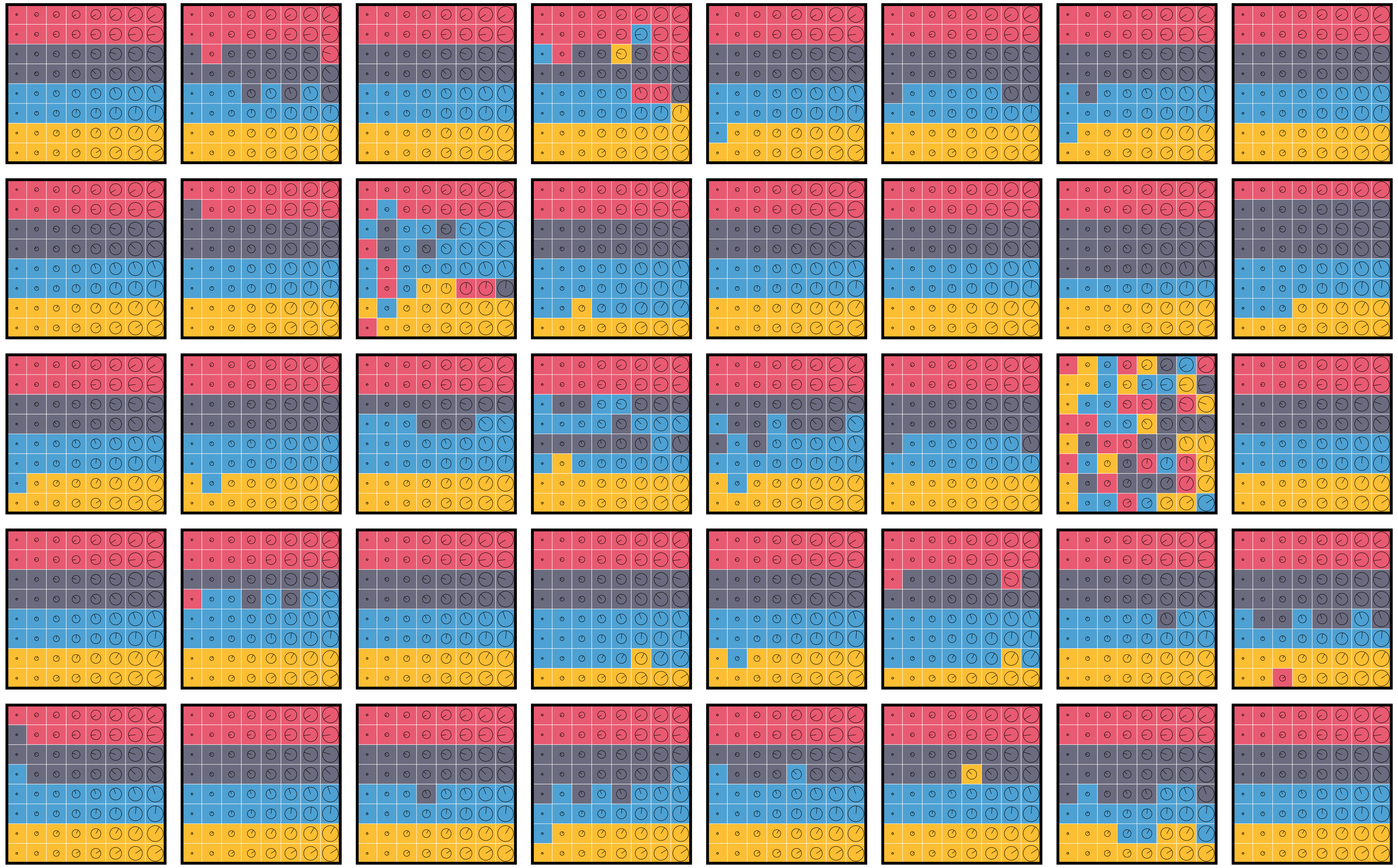
Angle & Size



Results

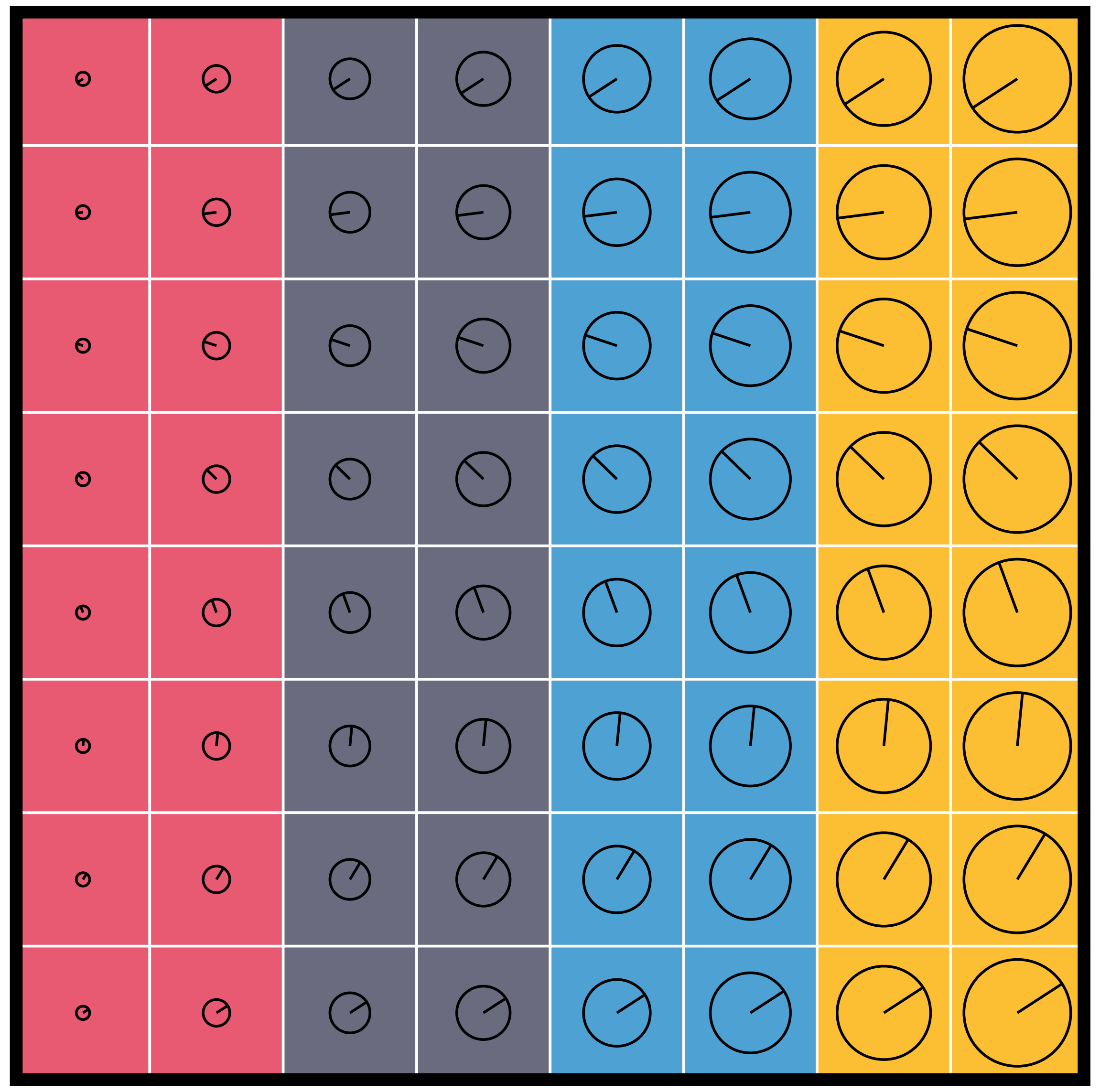
Angle-only



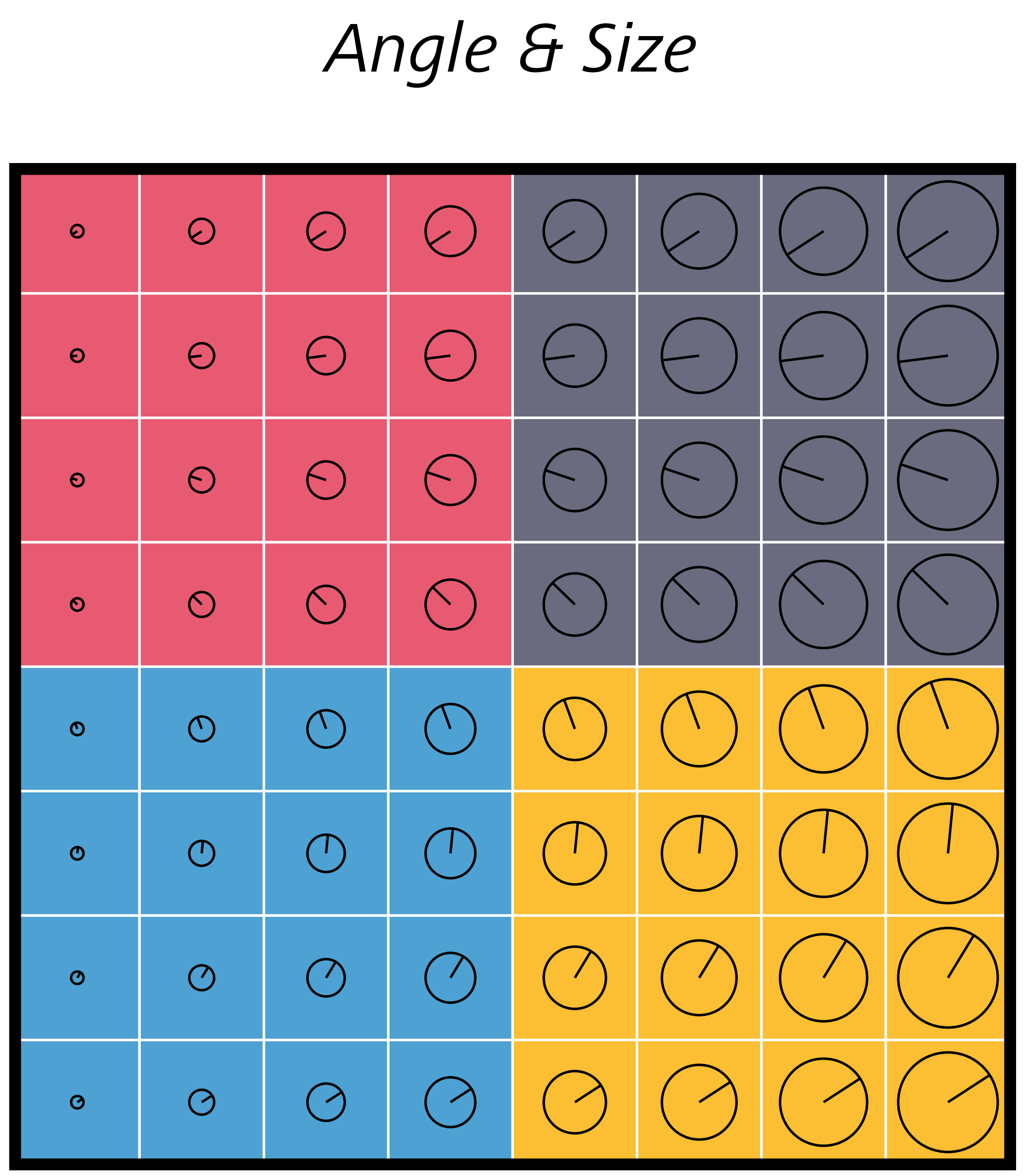


Results

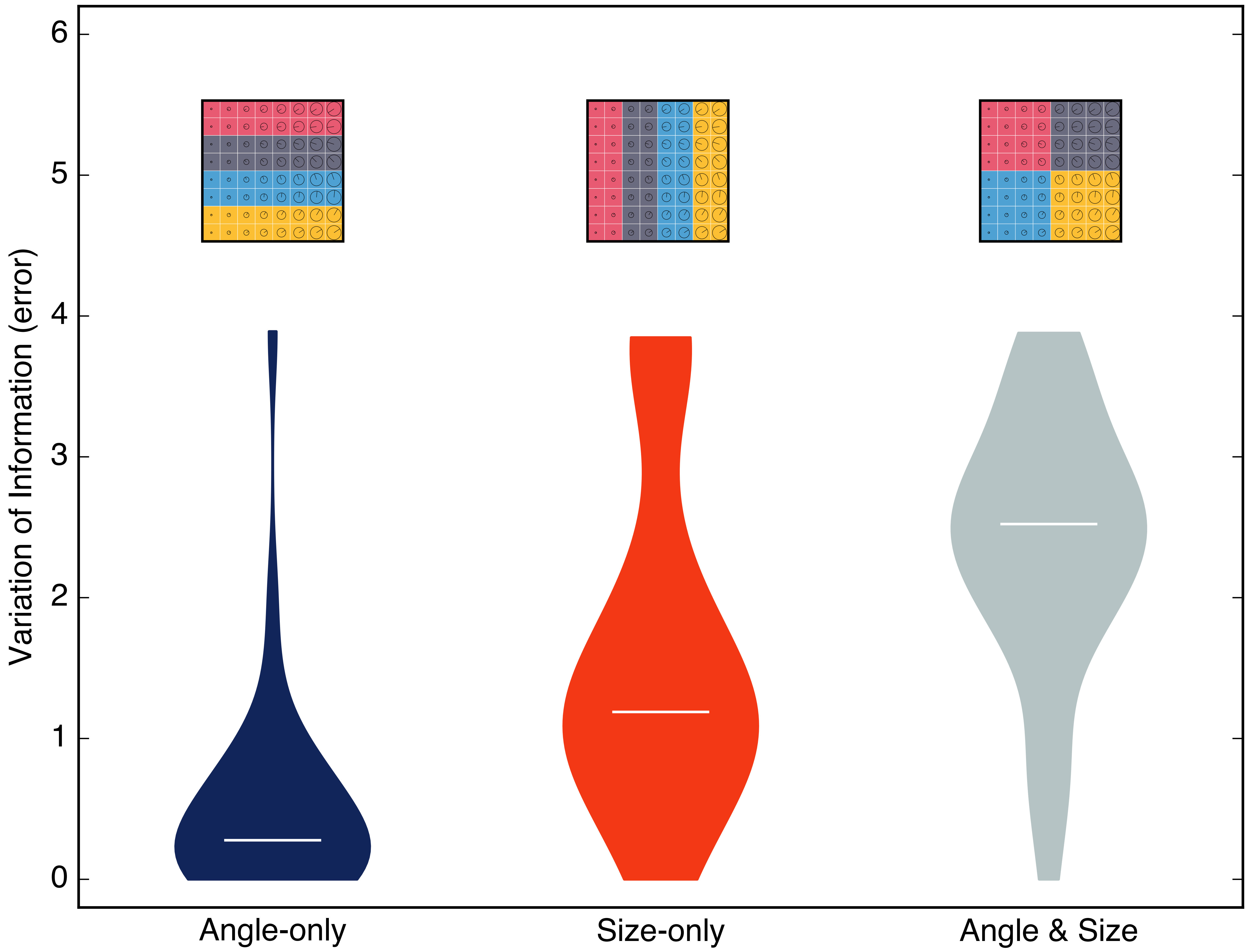
Size-only



Results

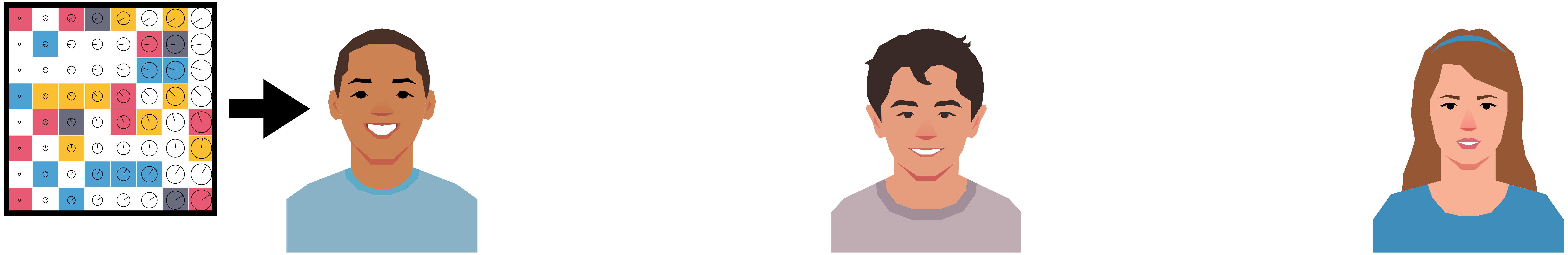


Result: Learnability advantage for the less informative systems

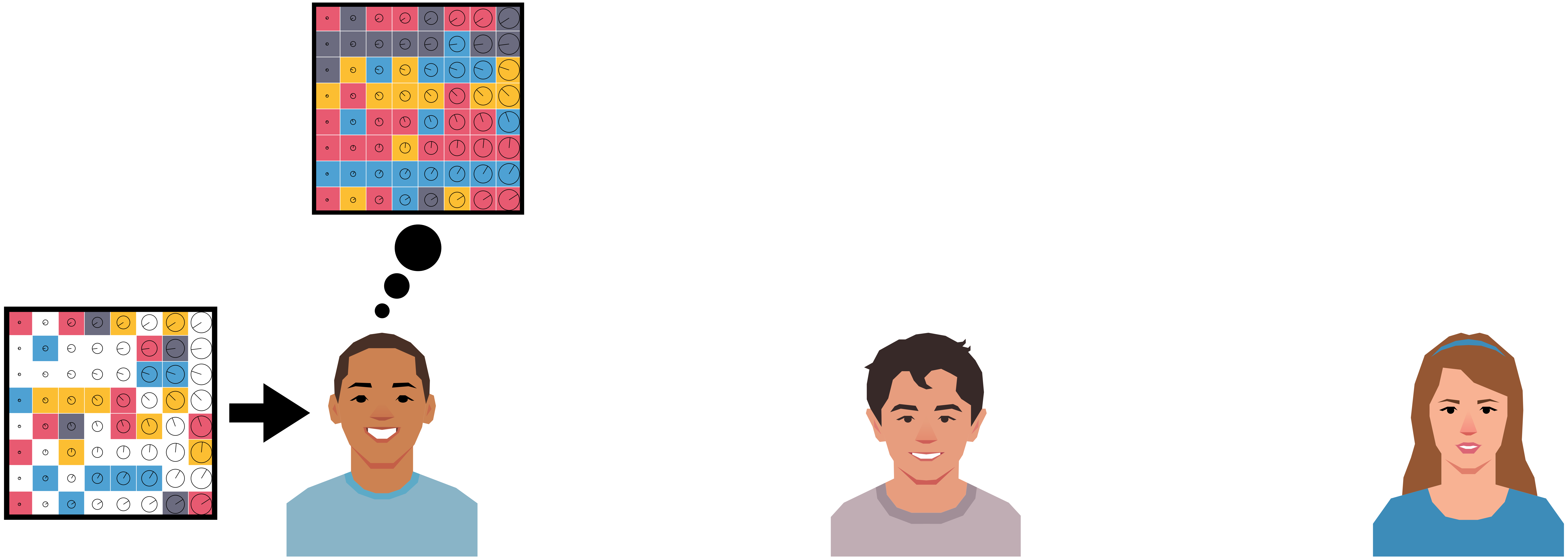


Experiment 2

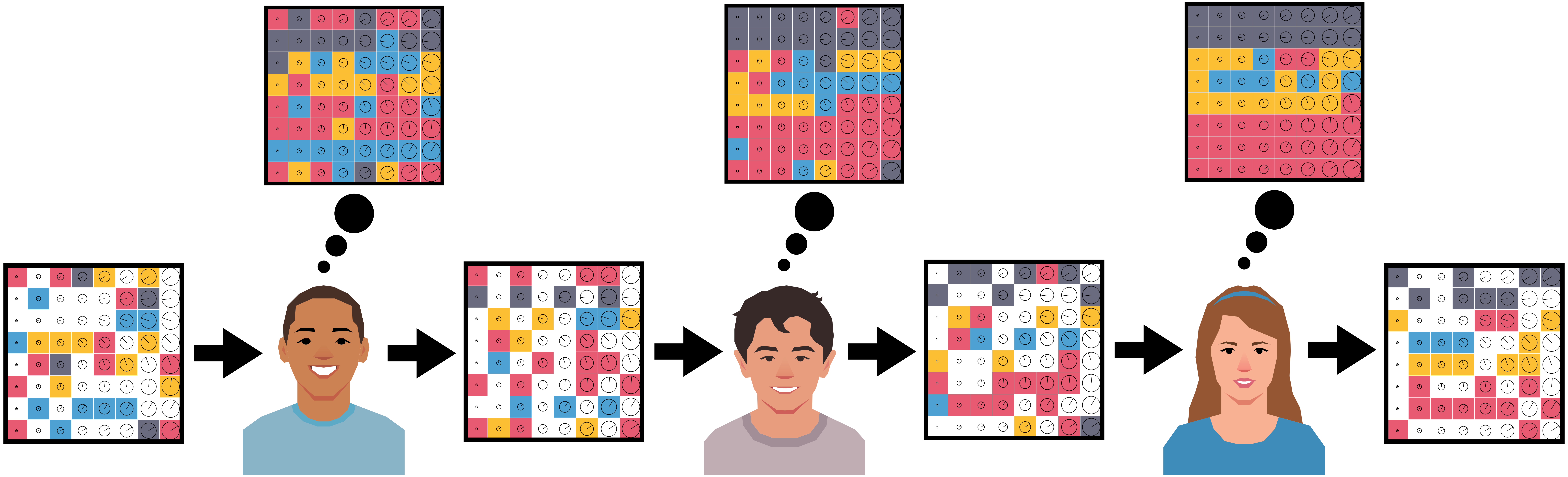
Iterated learning with humans

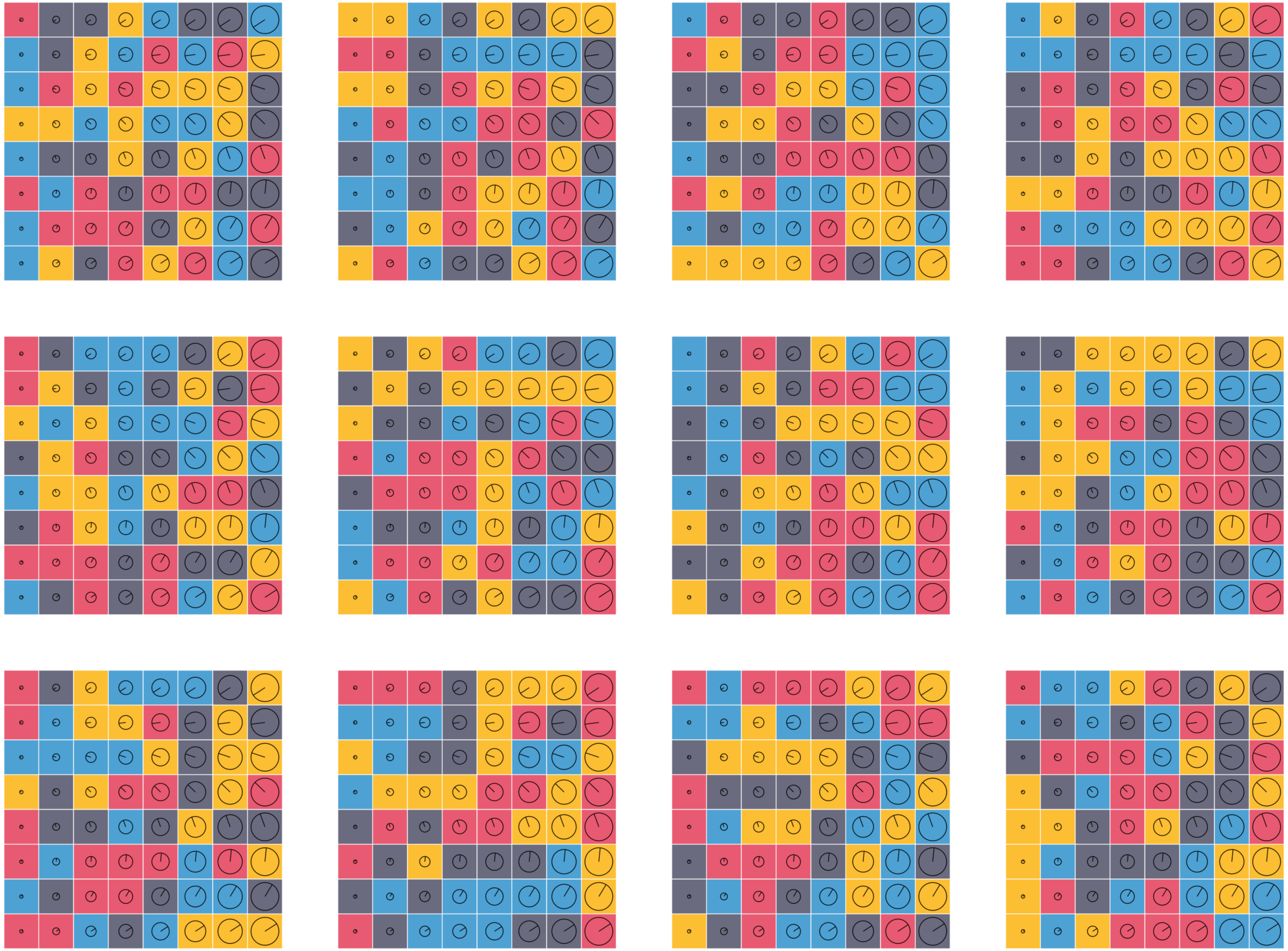


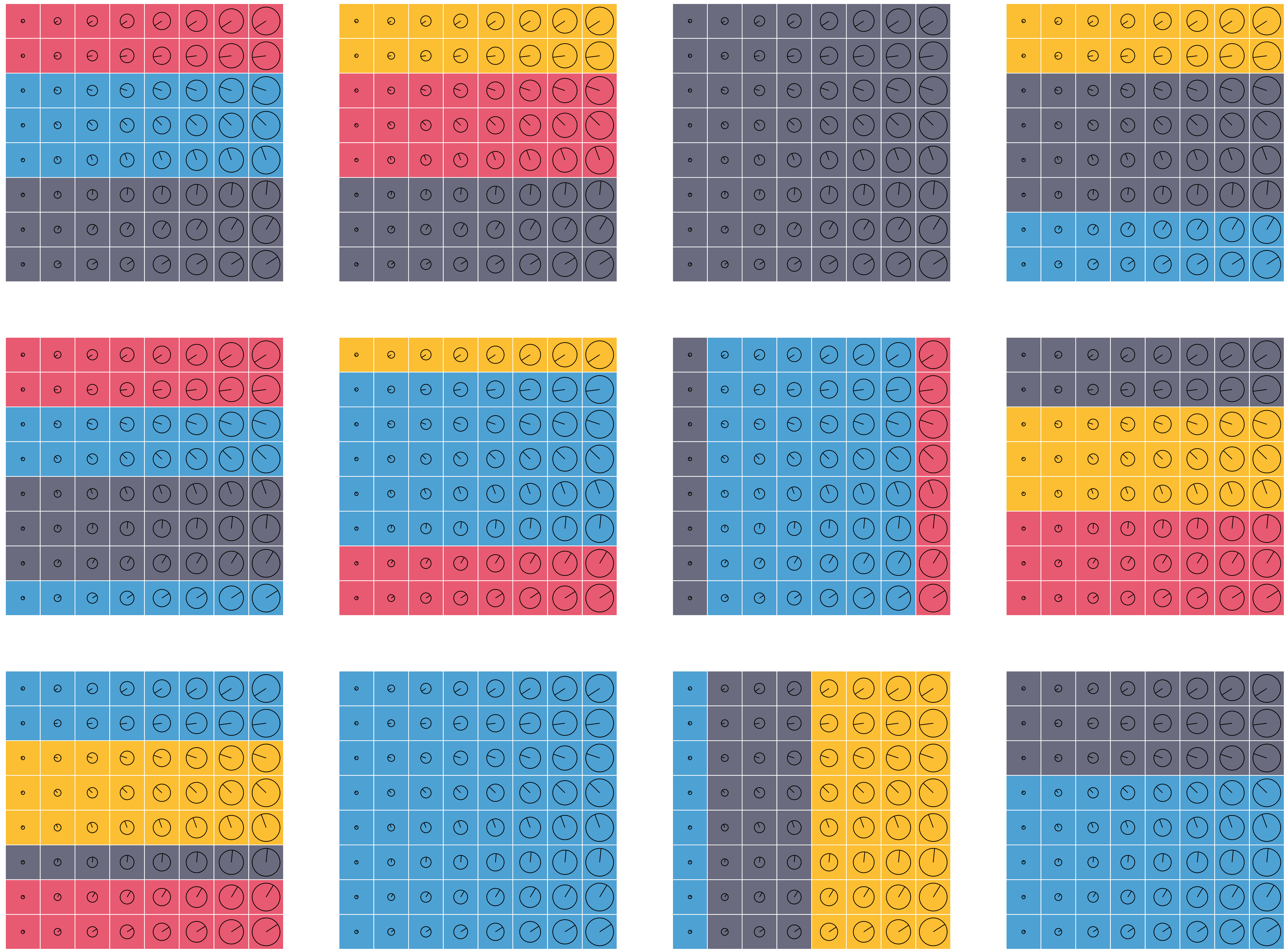
Iterated learning with humans



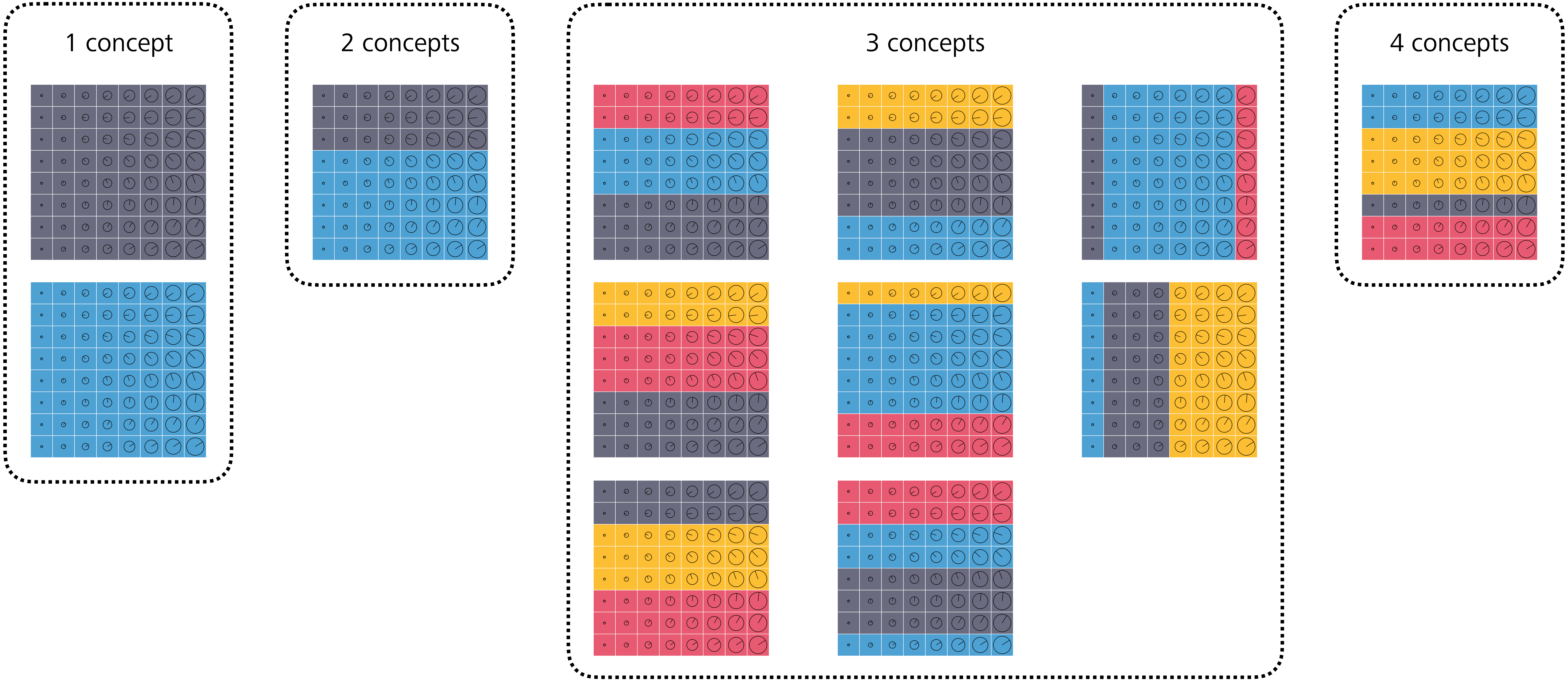
Iterated learning with humans



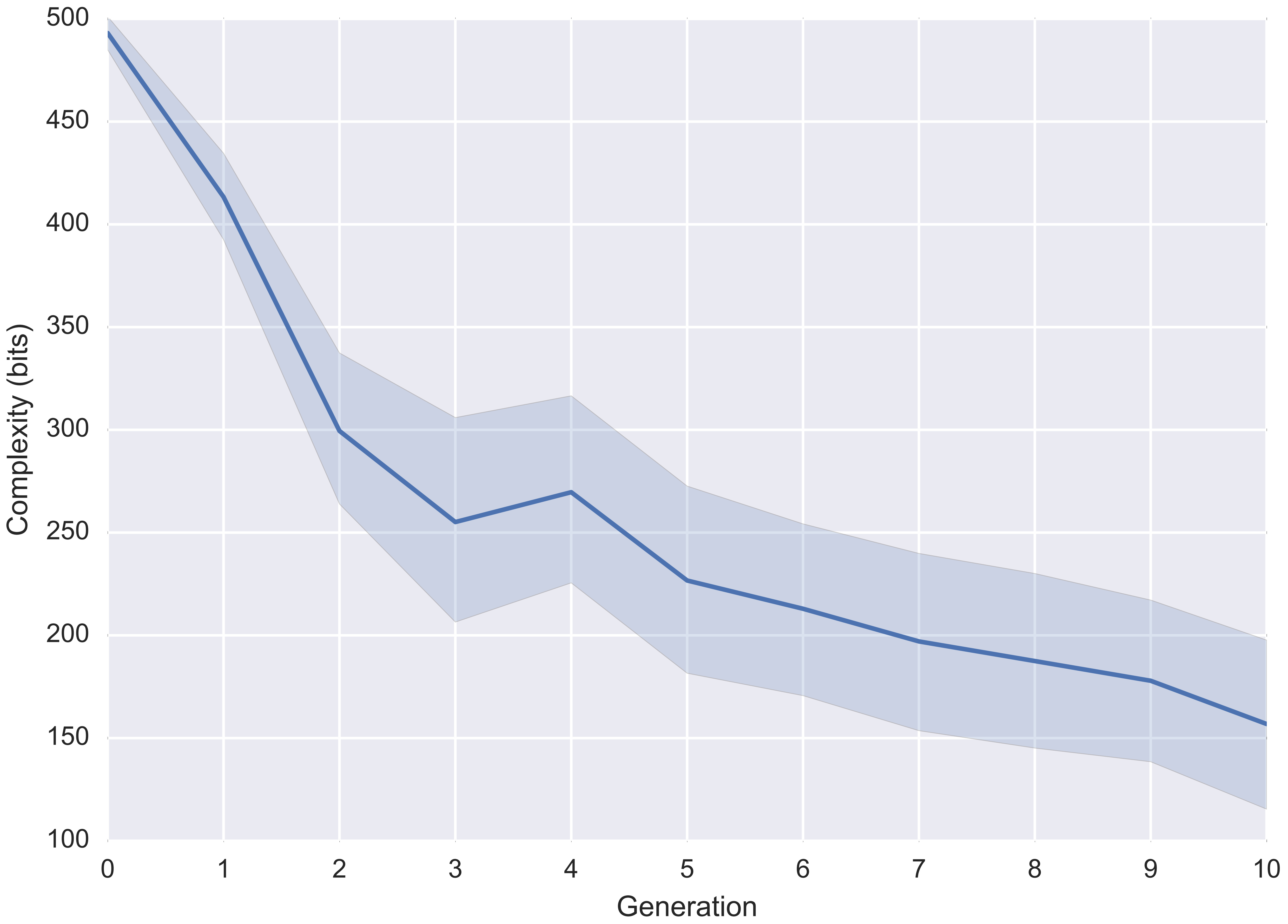




Results

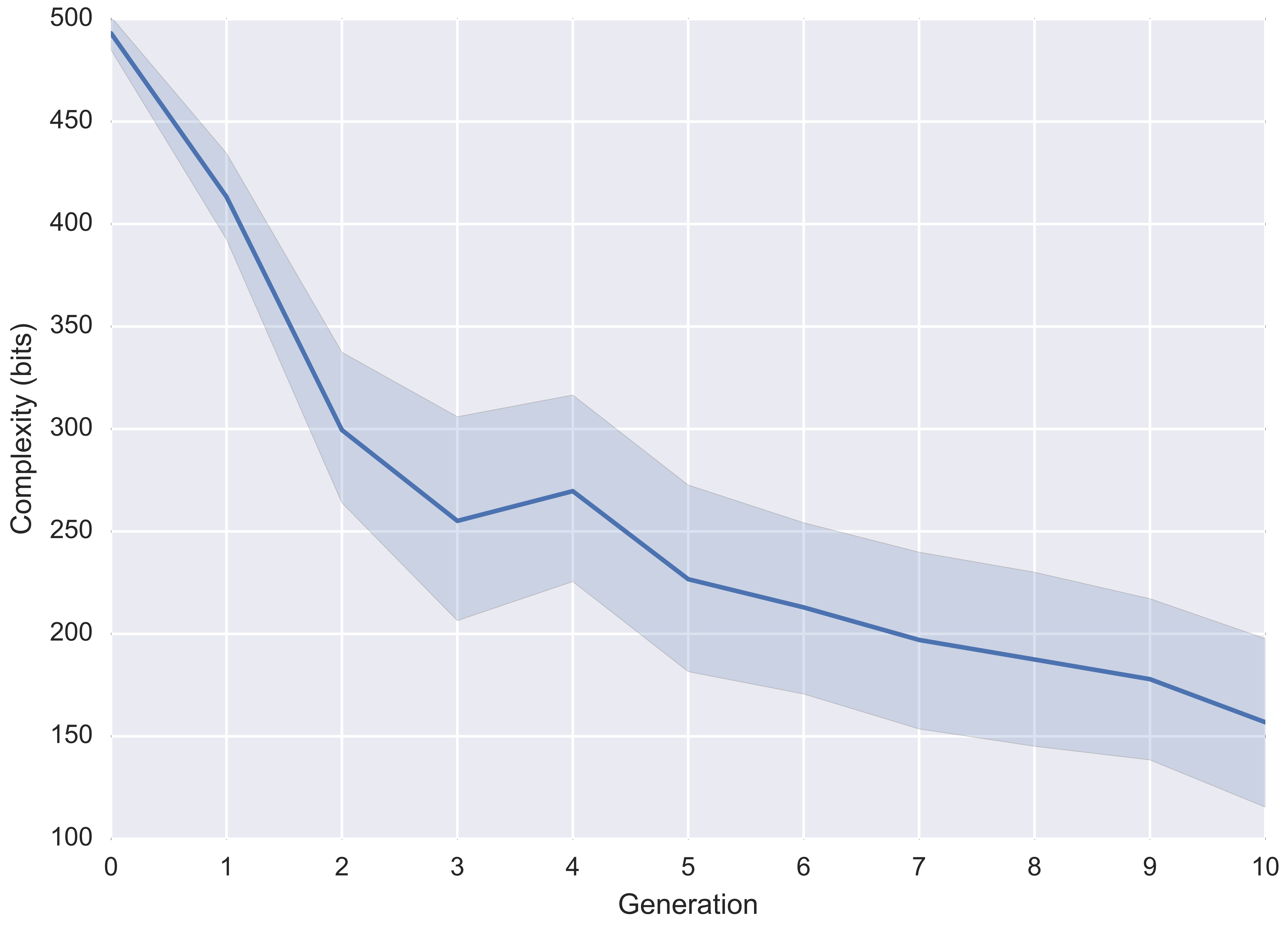


Simplicity

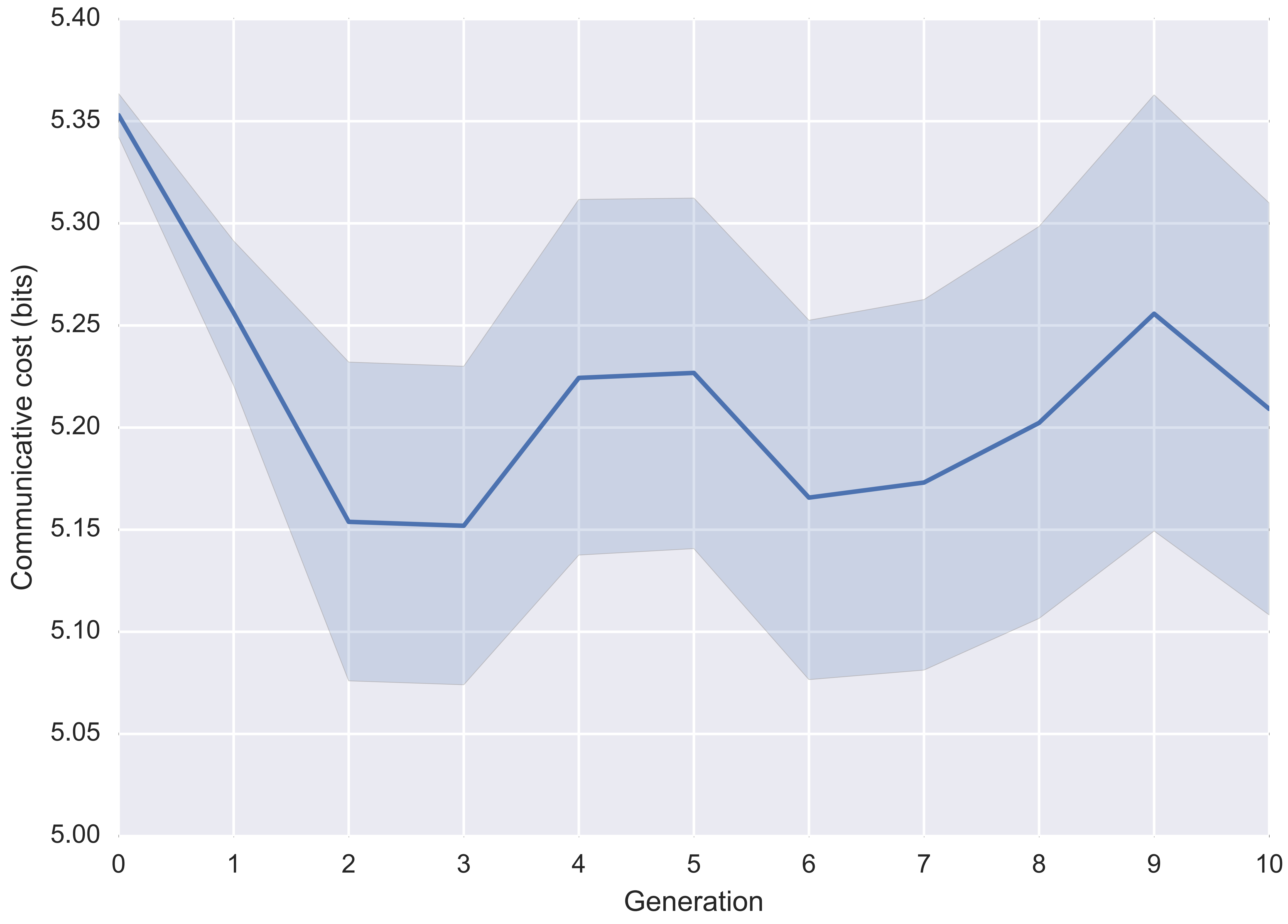


500
450
400
350
300
250
200
150
100
Complexity (bits)
0
1
2
3
4
5
6
7
8
9
10
Generation

Simplicity

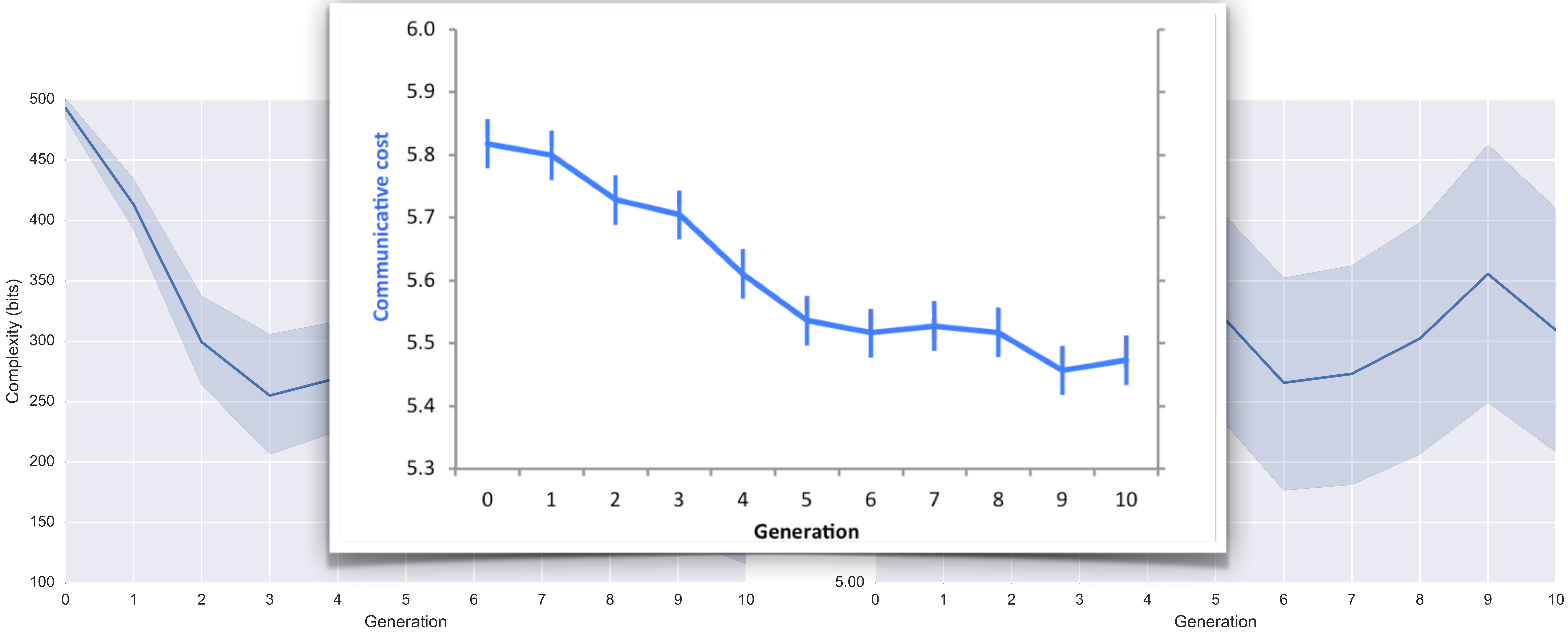


Informativeness

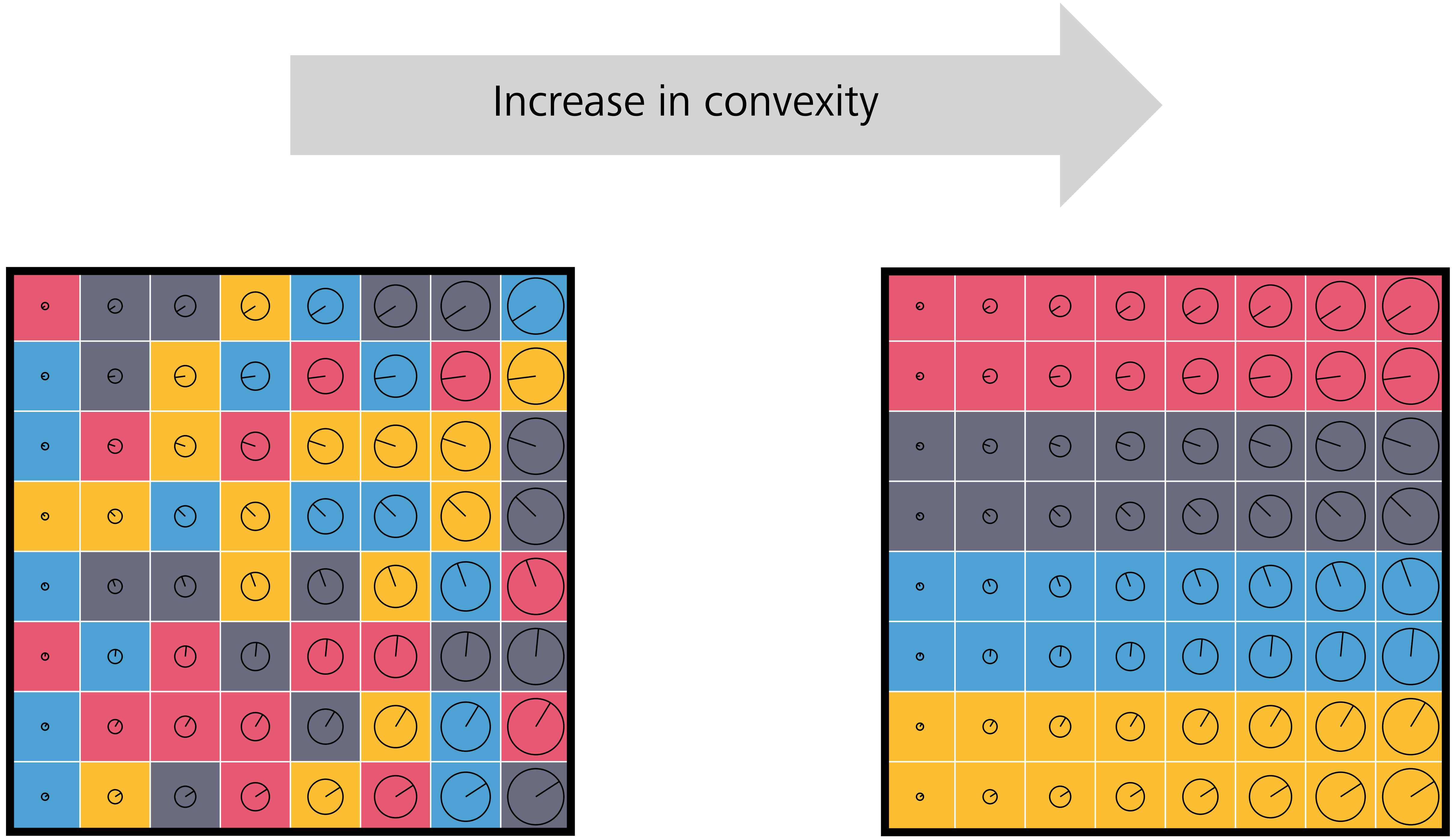


Simplicity

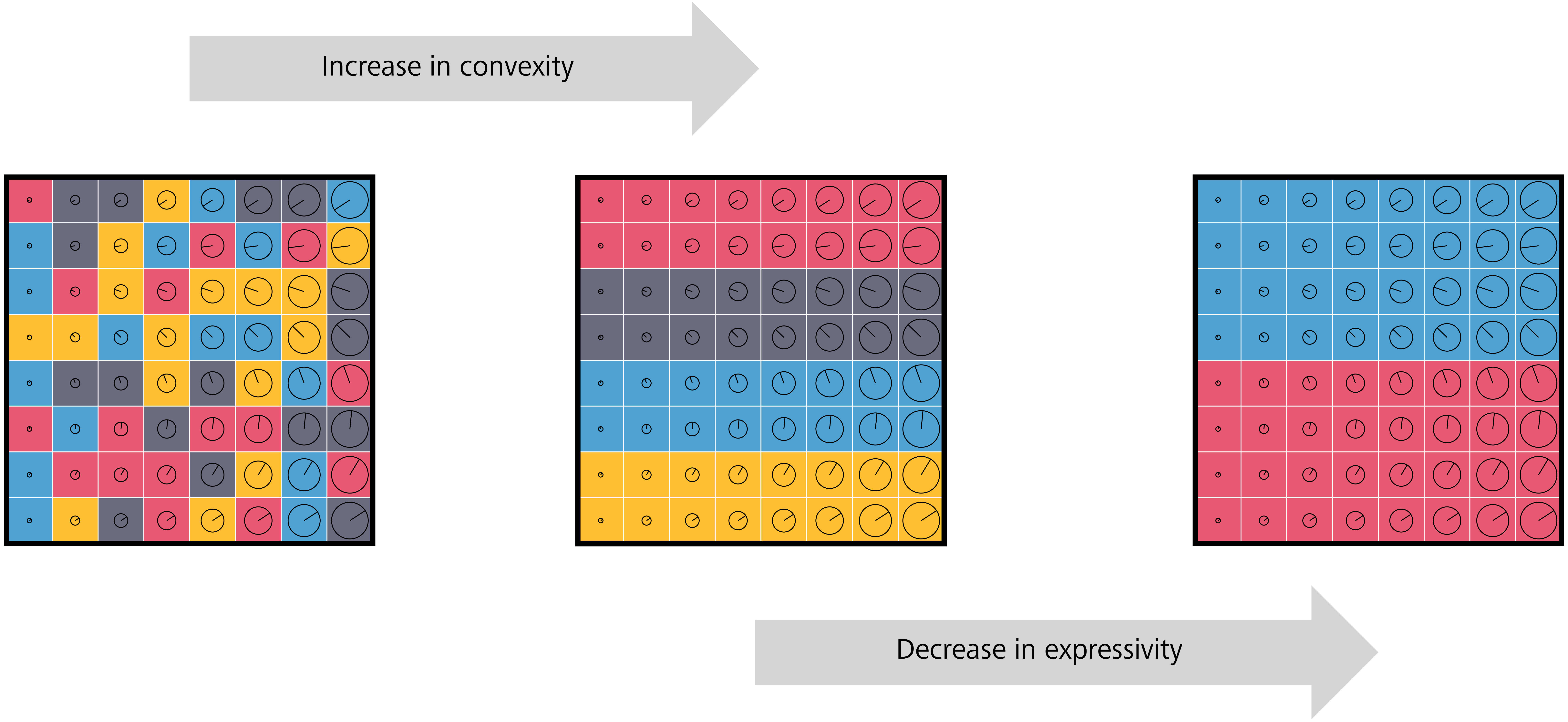
Informativeness



Two ways of achieving simplicity



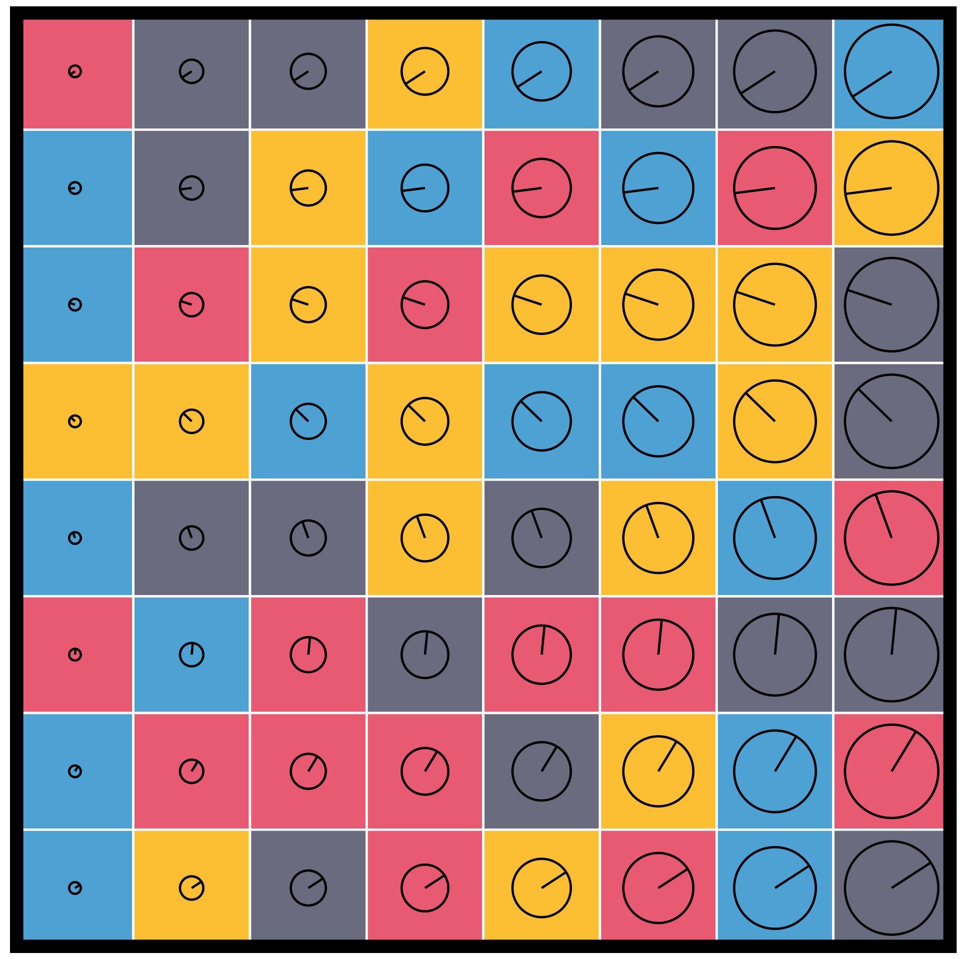
Two ways of achieving simplicity

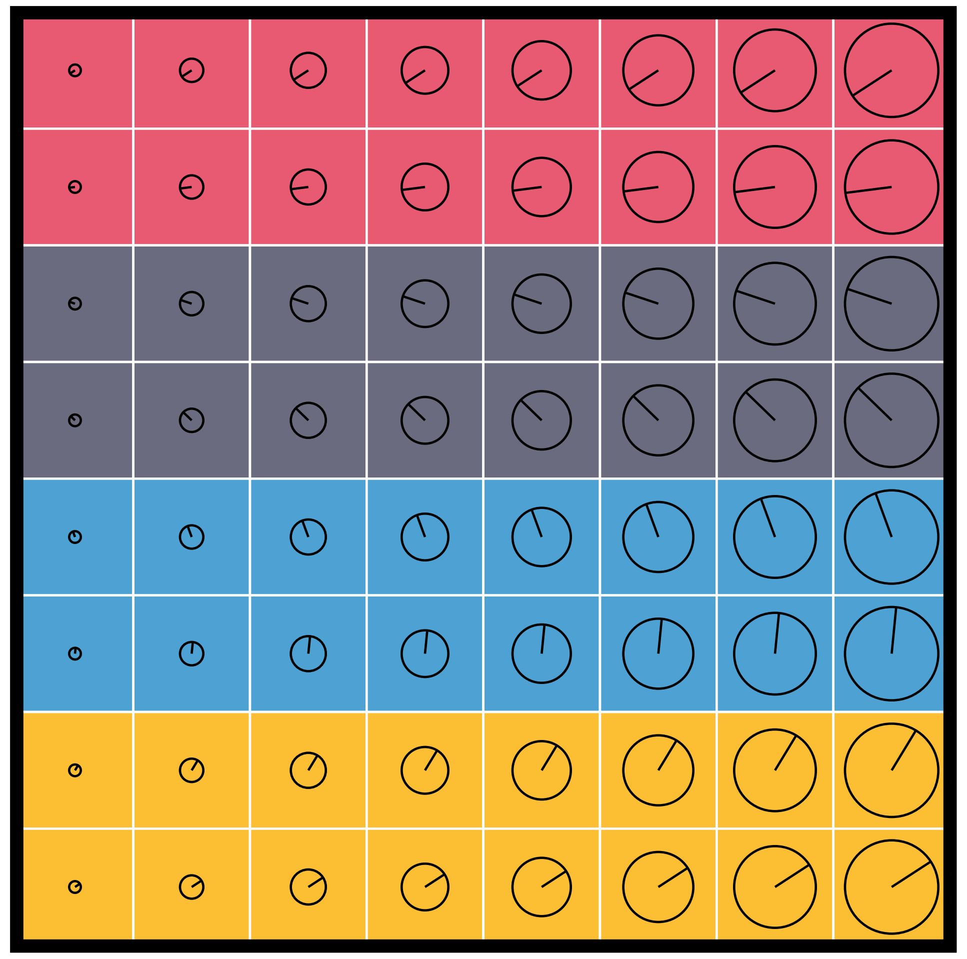


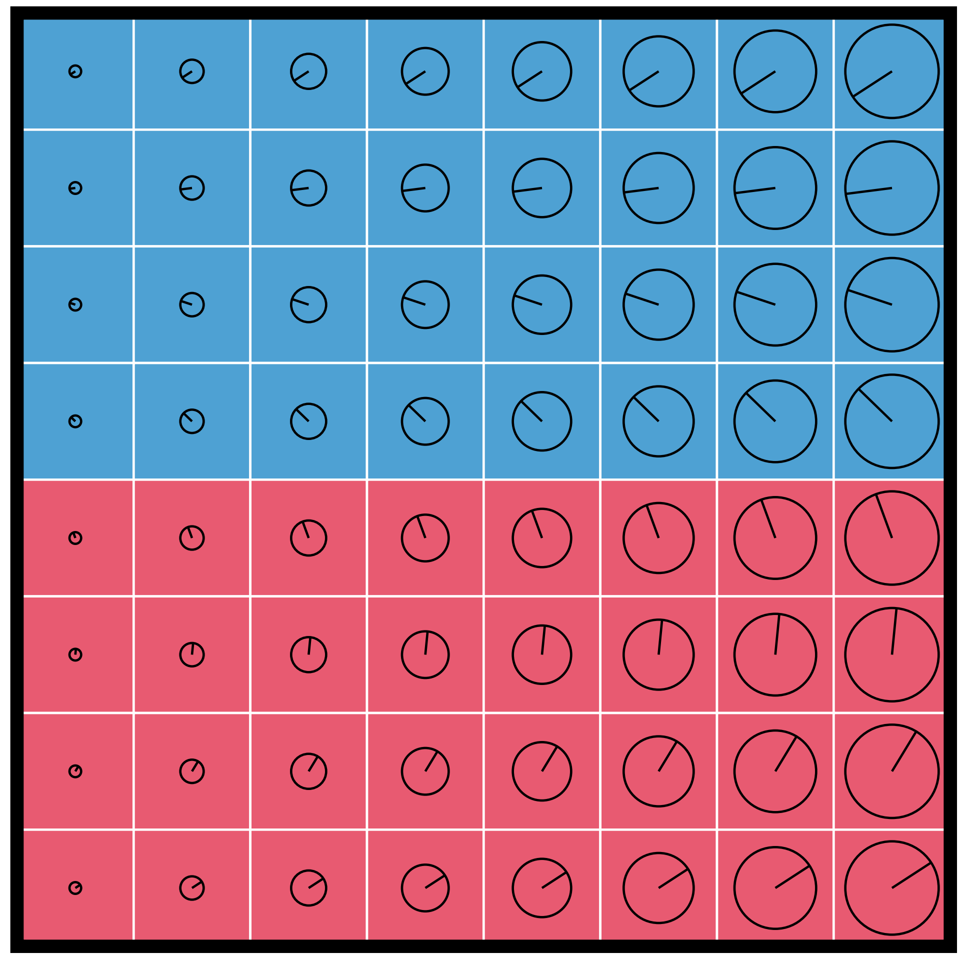
Two ways of achieving simplicity

Increase in convexity

increases informativeness







Decrease in expressivity

decreases informativeness

Conclusions

Languages are shaped in the simplicity–informativeness tradeoff by pressures from learning and communication

Learning contains a simplicity bias to prevent overfitting noise, and to aid reasoning about unseen meanings

Iterated learning converges to the prior bias, favouring languages that are as simple as possible:

> **Loss of expressivity:** Loss of words/concepts to aid learning
>
> **Convex categories:** Reorganization of the space to aid learning

In the process, some informativeness may come along for the ride, potentially obscuring the causal mechanism in experimental work

Vielen Dank!