Informativeness: A review of work by Regier and colleagues *(and a response)*

Jon W. Carr

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



What shapes language?

CENTRE for LANGUAGE EVOLUTION

compressibility

CENTRE for LANGUAGE EVOLUTION

expressivity

simplicity

Language & Cognition Lab

informativeness

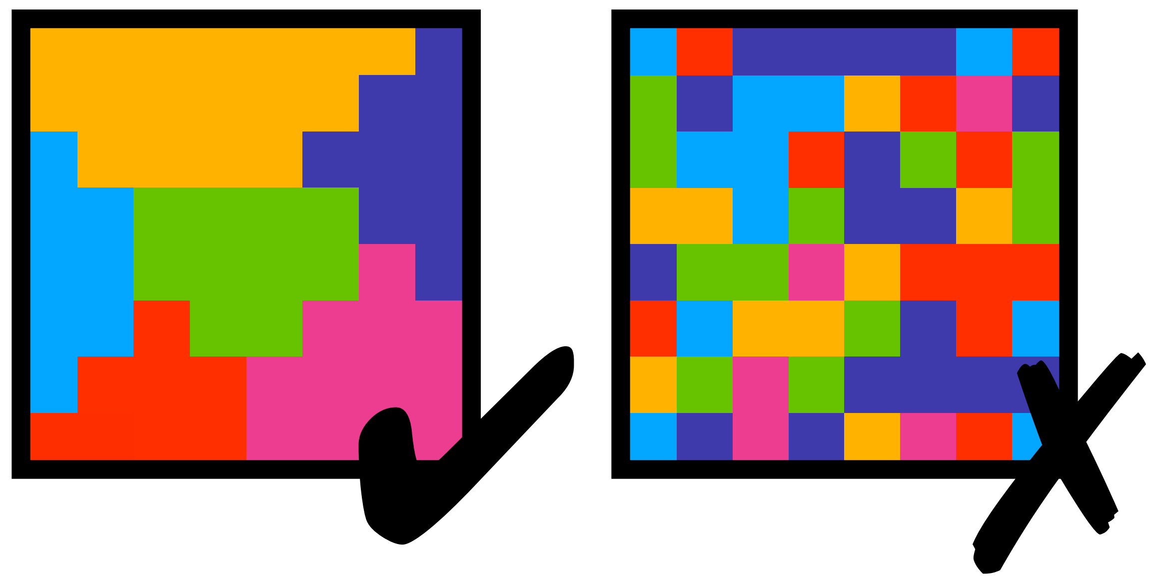
Language & Cognition Lab

How do learning and communication shape the structure of semantic categories?

a pressure for simplicity

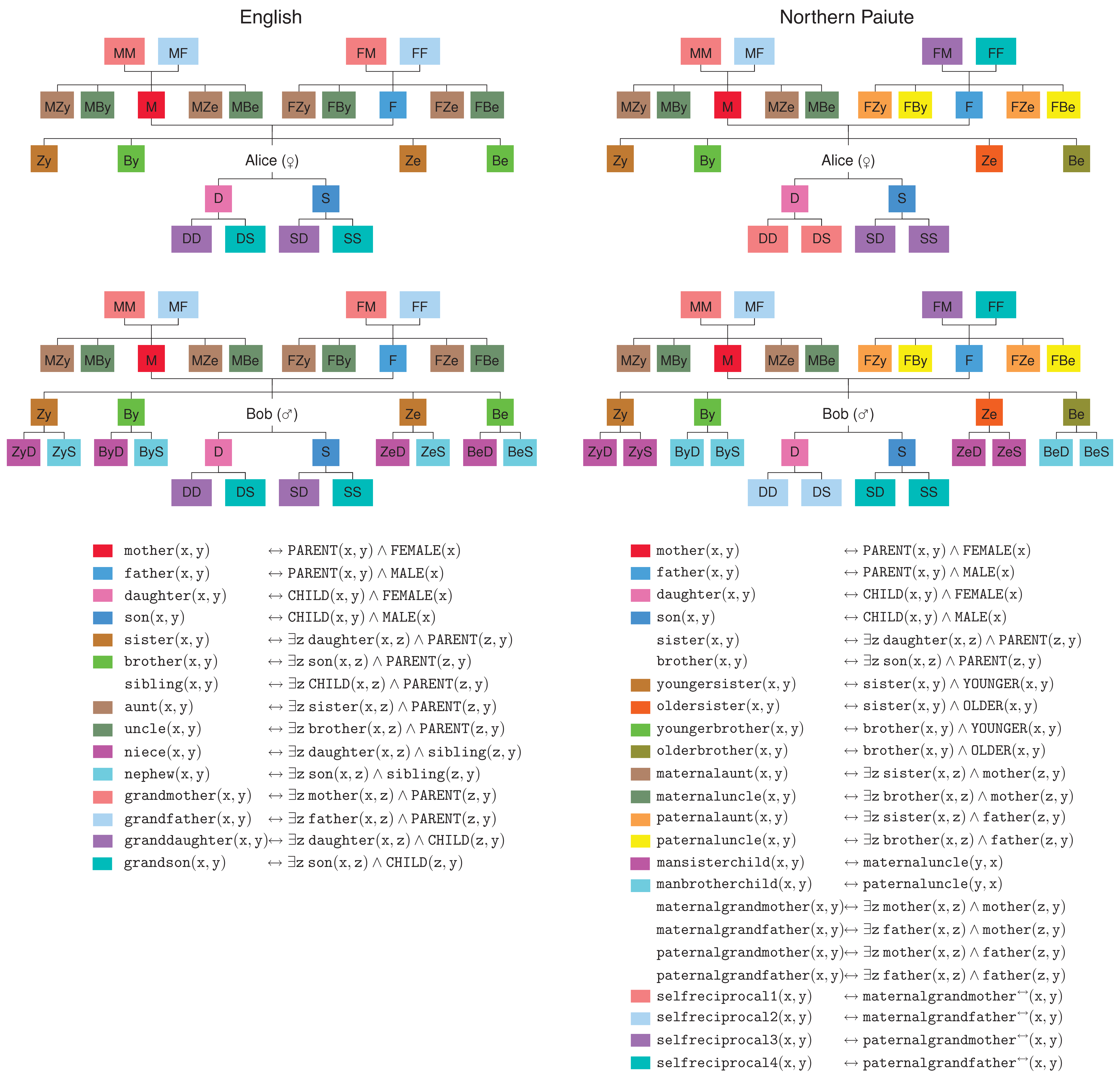
How do learning and communication shape the structure of semantic categories?

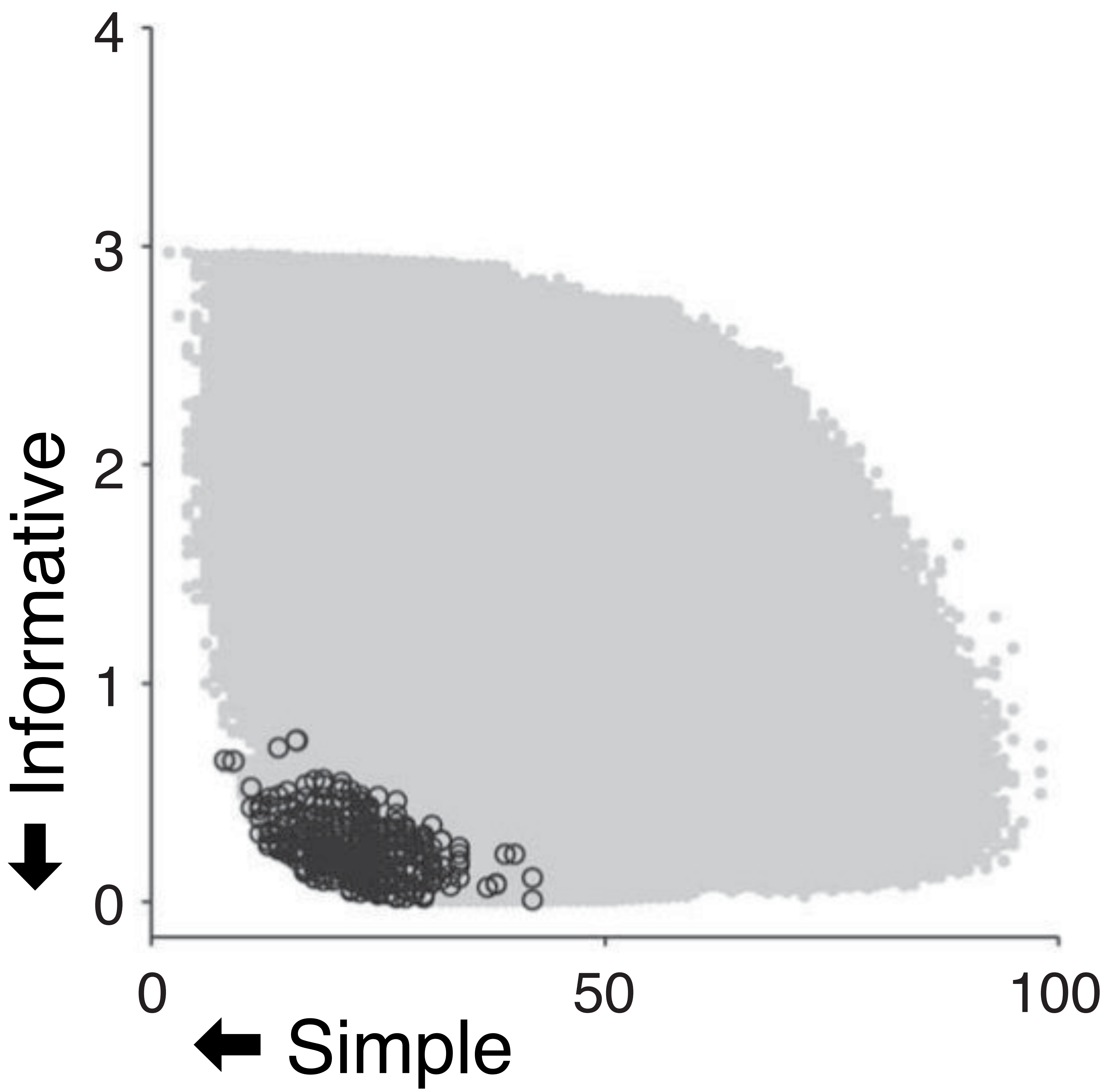
a pressure for informativeness



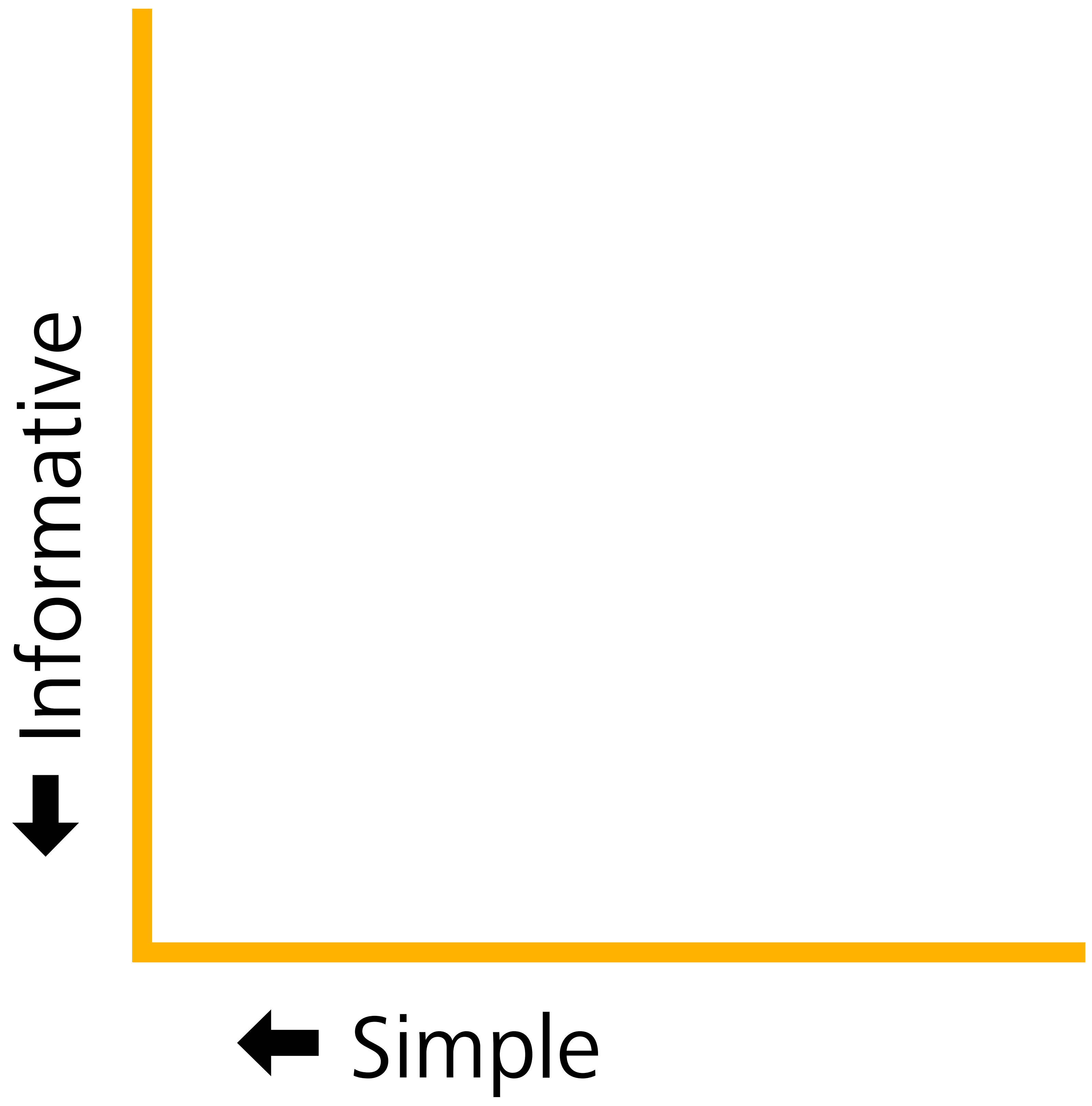
Kinship terms are simple and informative

Kemp & Regier (2012)

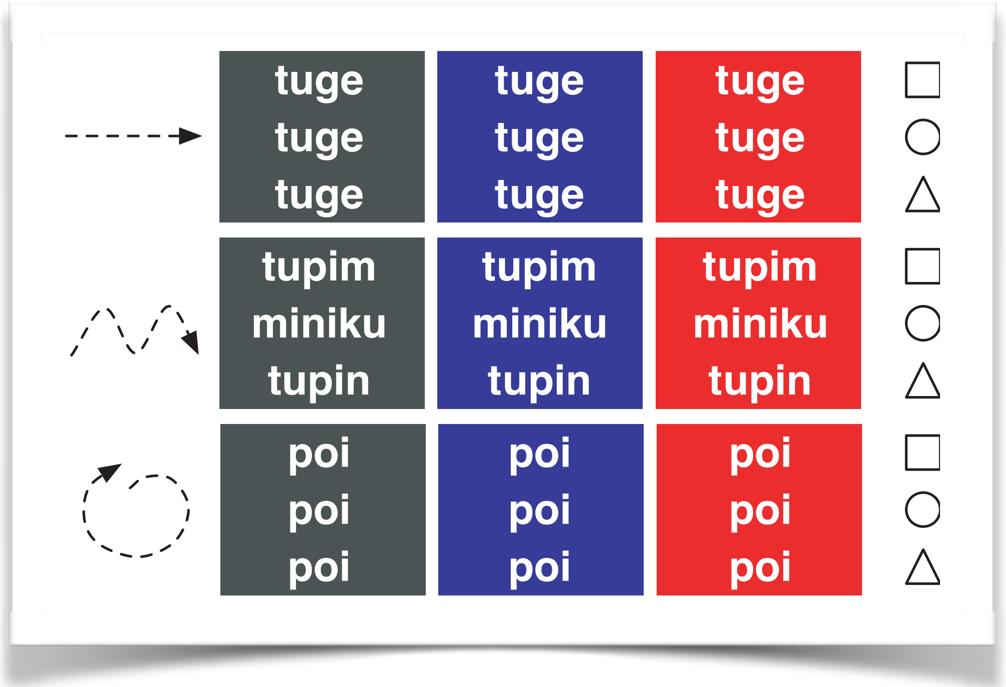




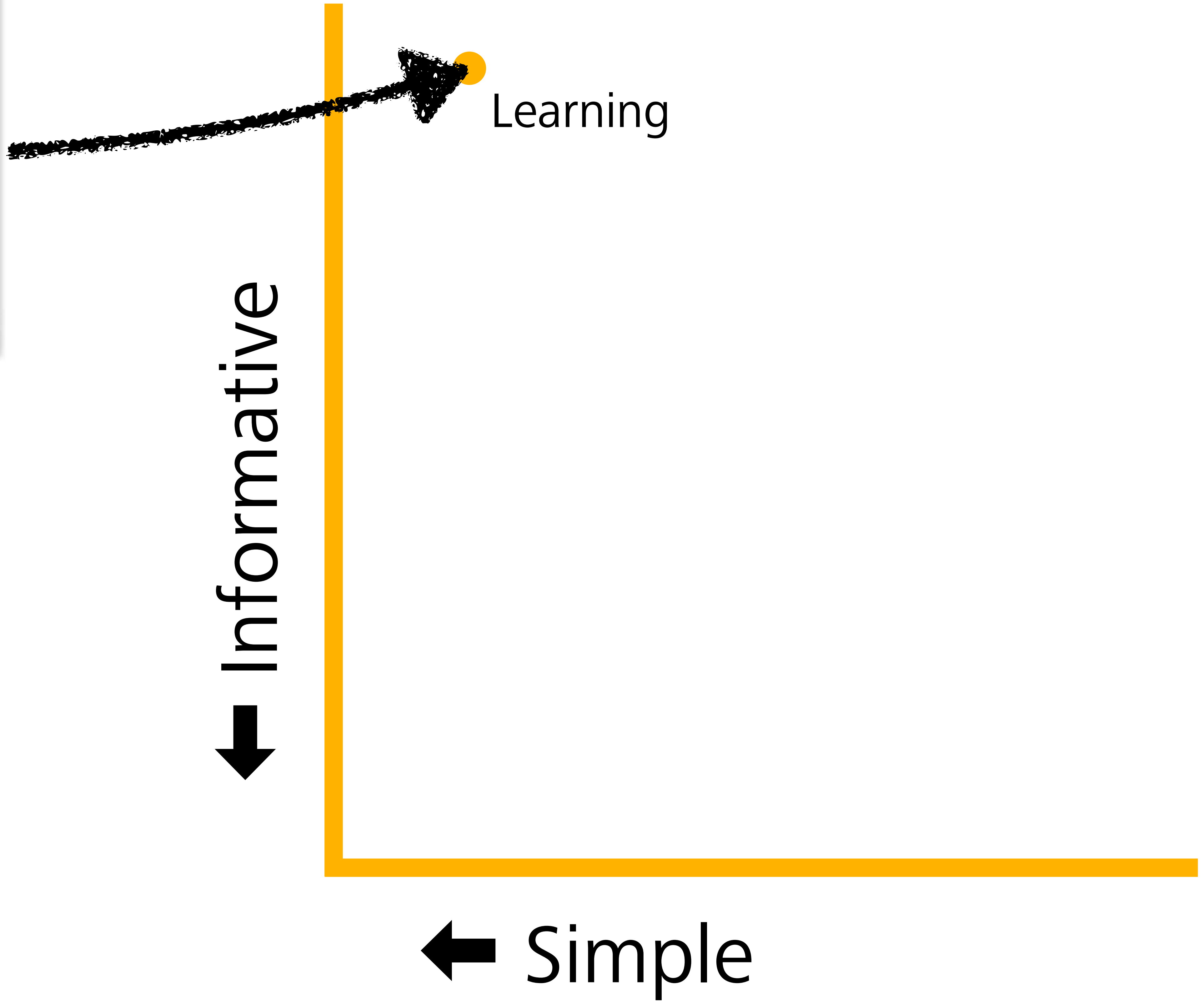
Learning and communication in the CLE framework



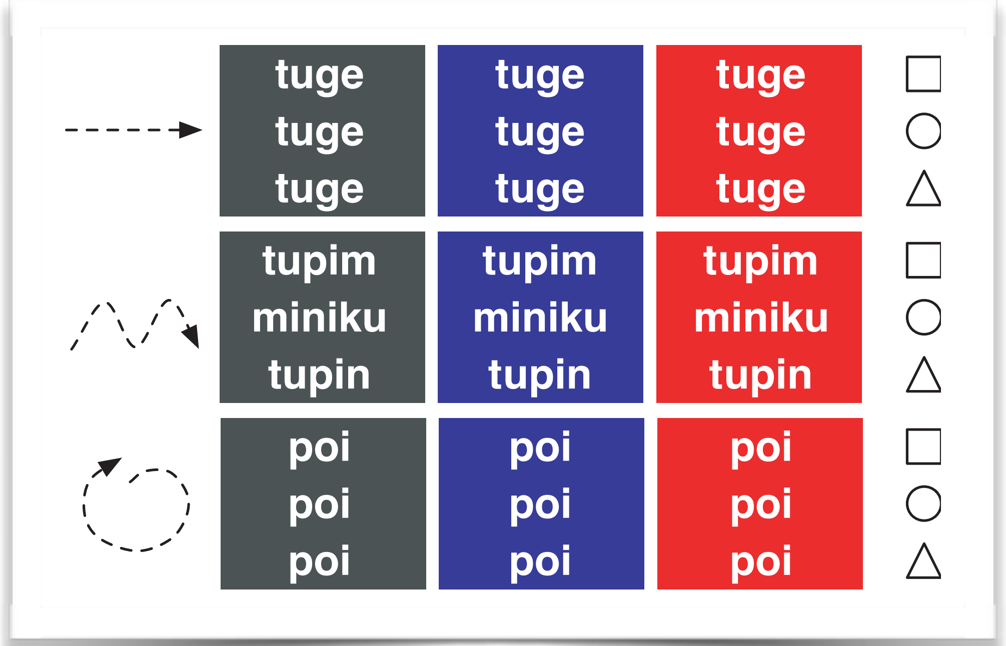
Learning and communication in the CLE framework



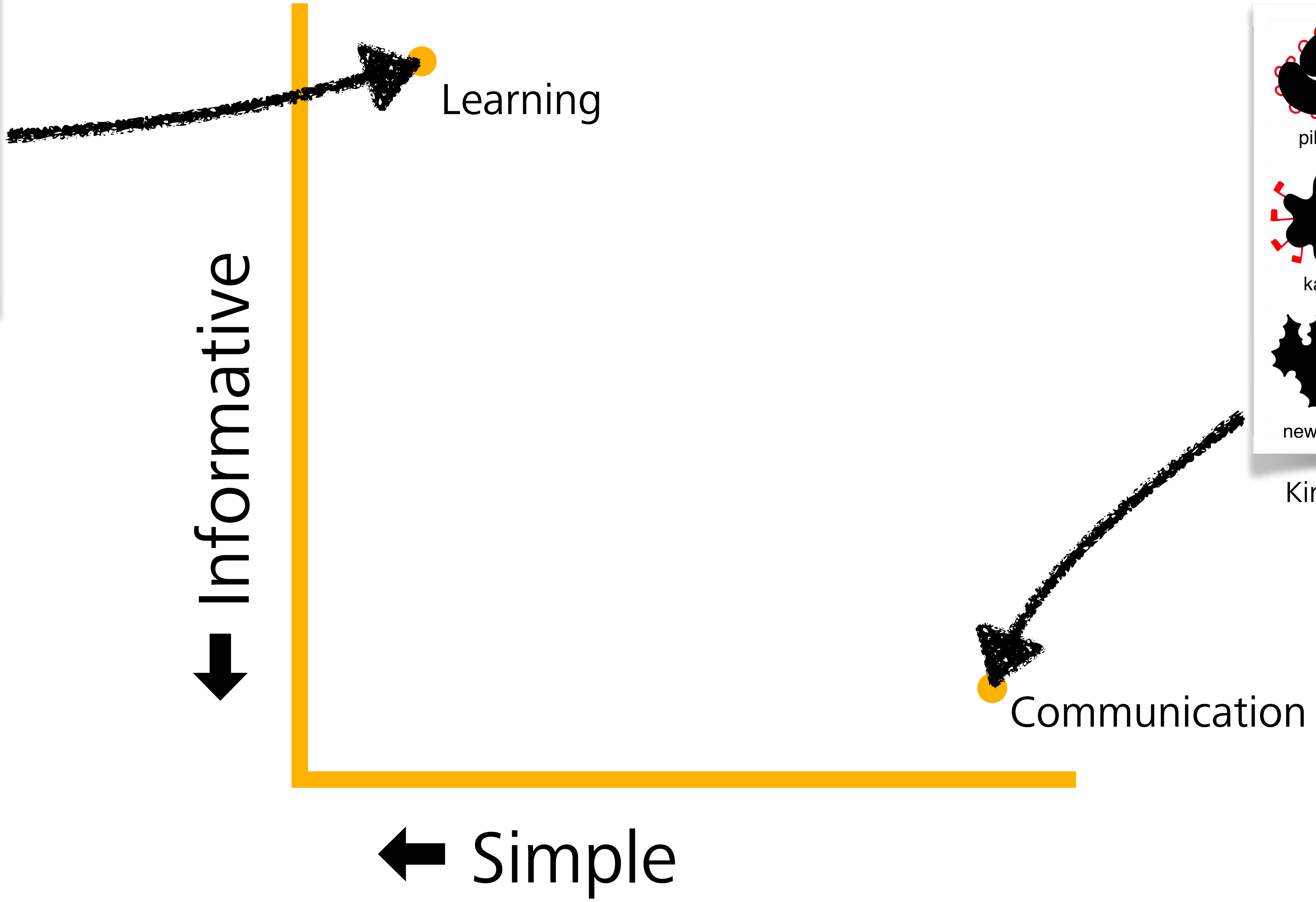
Kirby, Cornish, & Smith (2008)

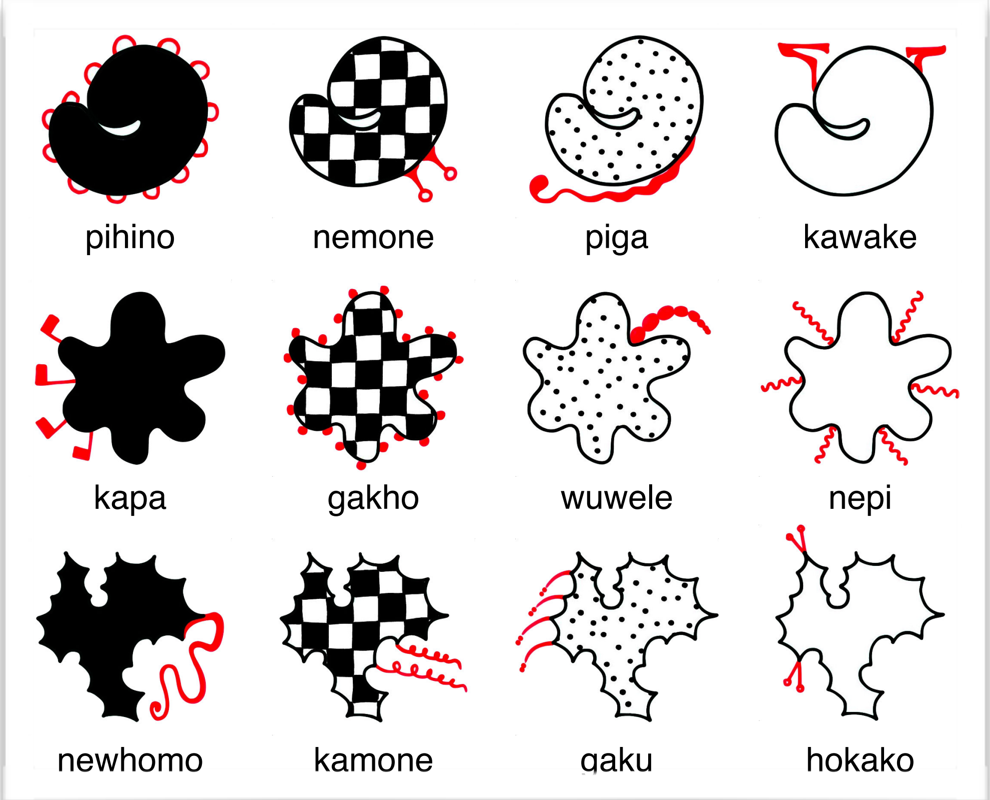


Learning and communication in the CLE framework



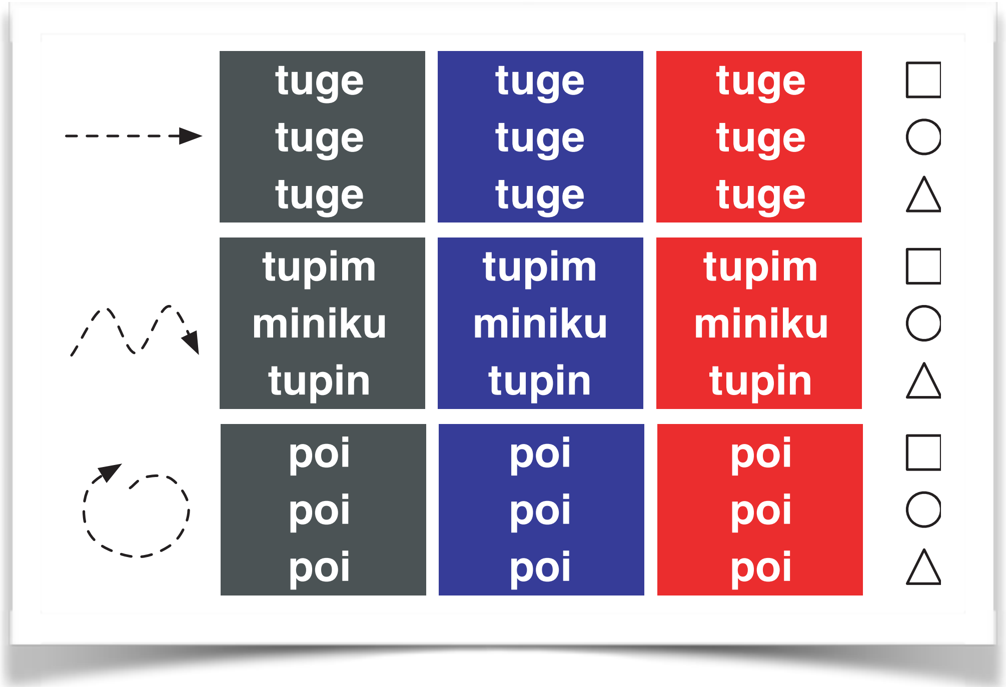
Kirby, Cornish, & Smith (2008)



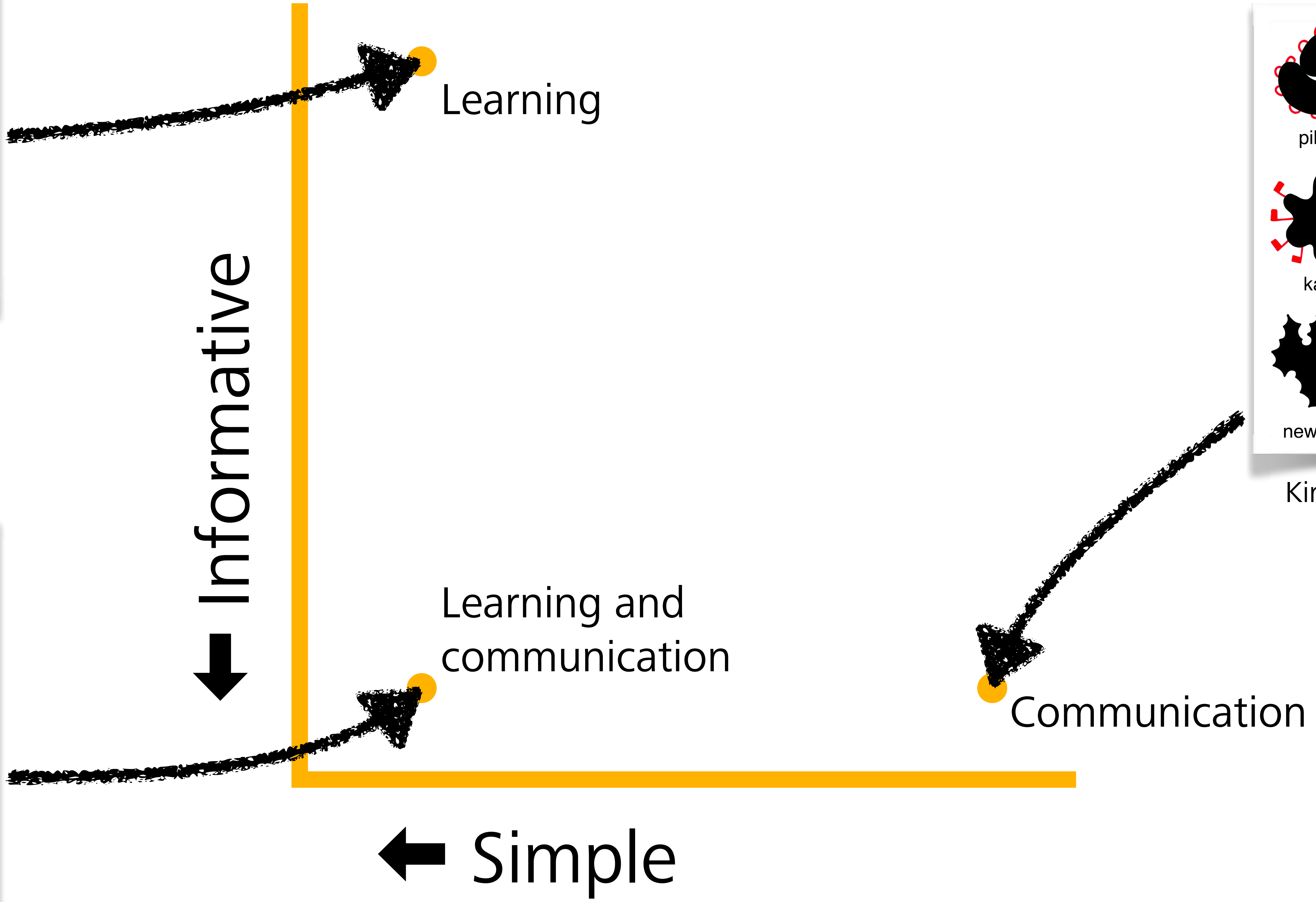


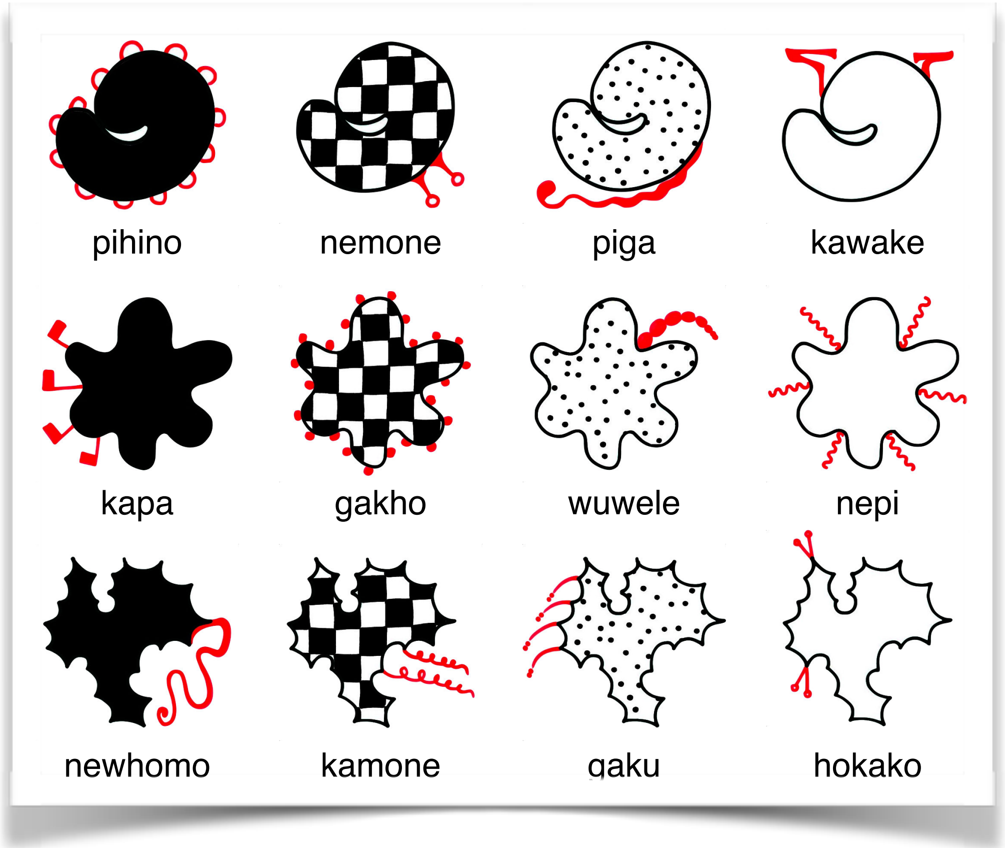
Kirby, Tamariz, Cornish, & Smith (2015)

Learning and communication in the CLE framework



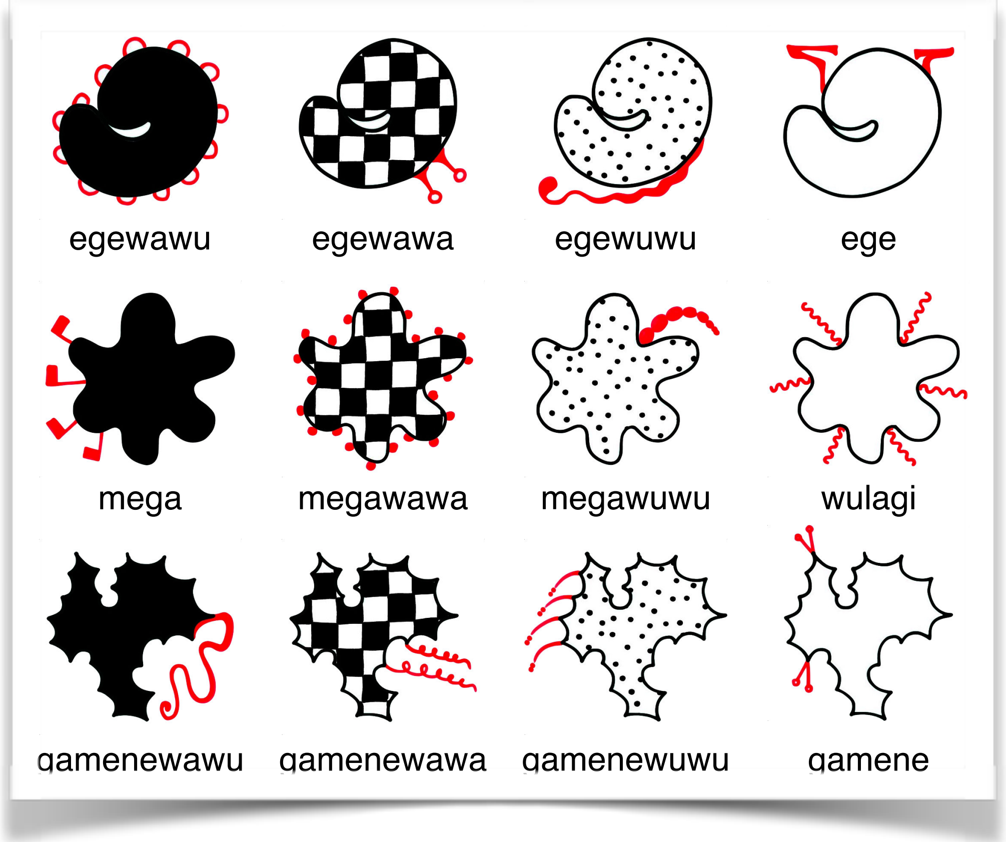
Kirby, Cornish, & Smith (2008)





Kirby, Tamariz, Cornish, & Smith (2015)

Kirby, Tamariz, Cornish, & Smith (2015)



Summary

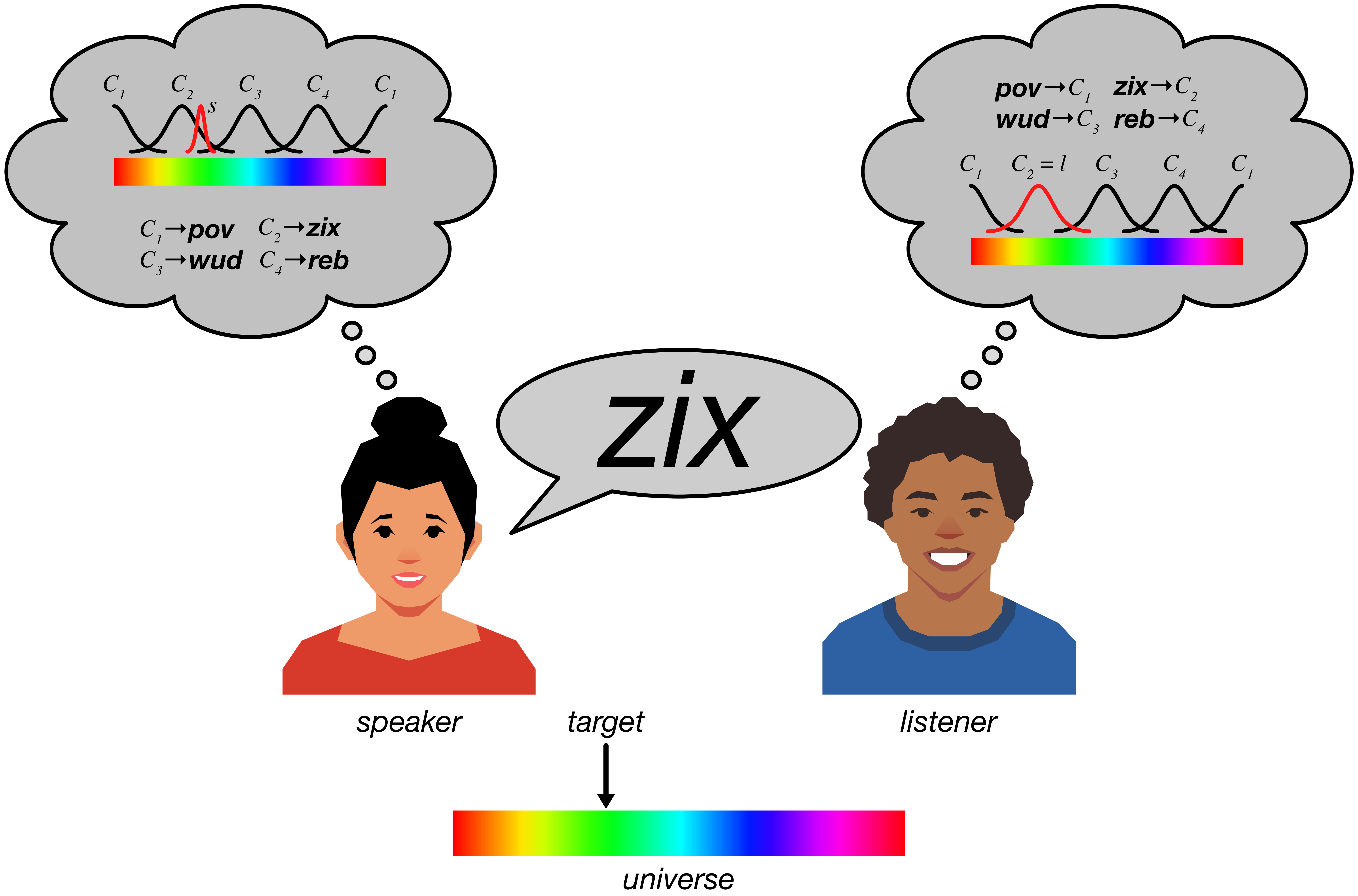
	Pressure from learning	Pressure from communication
CLE	**Compressibility:** To what extent can the language be compressed? **Measure:** MDL, gzip, entropy	**Expressivity:** How many meaning distinctions does the language allow? **Measure:** Number of words
Regier	**Simplicity:** How many words does an individual need to remember? **Measure:** Number of words, number of rules	**Informativeness:** How effectively can a meaning be transmitted? **Measure:** Communicative cost

Summary

Pressure from learning	Pressure from communication
Compressibility: To what extent can the language be compressed? **Measure:** MDL, gzip, entropy	**Informativeness:** How effectively can a meaning be transmitted? **Measure:** Communicative cost
bits required to represent the language	*bits lost during communication*

Communicative cost

Communicative cost: High-level overview



Communicative cost: Low-level details

To compute the cost of a category partition, we start by considering a individual target meaning and compute how much error would be incurred in trying to reconstruct that target

Reconstruction error is defined as the Kullback-Leibler divergence between *s* and *l*:

$$D_{\mathrm{KL}}(s||l) = \sum_{i \in \mathcal{U}} s(i) \log_2 \frac{s(i)}{l(i)} = \log_2 \frac{1}{l(t)}$$

Summing the divergences for all targets yields the communicative cost for the partition:

$$k = \sum_{t \in \mathcal{U}} p(t) D_{\mathrm{KL}}(s||l)$$

$$k = \sum_{t \in \mathcal{U}} p(t) \log_2 \frac{1}{l(t)}$$

Communicative cost: Example of a discrete categorizer

universe $\mathcal{U} = \{i_1, i_2, ..., i_{16}\}$

category partition $\mathcal{P} = \{C_1, C_2, C_3, C_4\}$
$= \{\{i_1, i_2, i_3, i_4\}, \{i_5, i_6, i_7, i_8\}, \{i_9, i_{10}, i_{11}, i_{12}\}, \{i_{13}, i_{14}, i_{15}, i_{16}\}\}$

speaker's lexicon $\mathcal{S} = \{C_1 \to \mathtt{00}, C_2 \to \mathtt{01}, C_3 \to \mathtt{10}, C_4 \to \mathtt{11}\}$

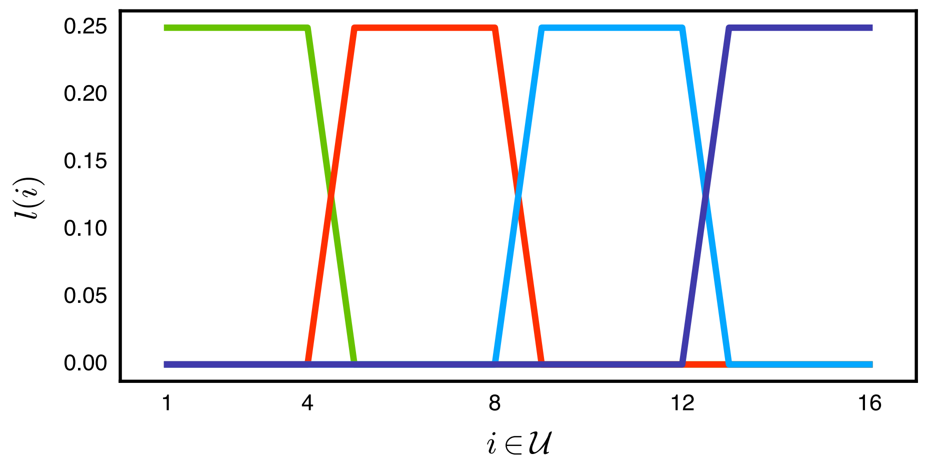
listener's lexicon $\mathcal{L} = \{\mathtt{00} \to C_1, \mathtt{01} \to C_2, \mathtt{10} \to C_3, \mathtt{11} \to C_4\}$

need probabilities $p = [\frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}]$

speaker distributions (for each meaning)
$s_1 = [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$
$s_2 = [0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$
...
$s_{16} = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]$

listener distributions (for each category)
$l_{C_1} = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$
$l_{C_2} = [0, 0, 0, 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, 0, 0, 0, 0, 0]$
$l_{C_3} = [0, 0, 0, 0, 0, 0, 0, 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, 0]$
$l_{C_4} = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}]$

$$\begin{aligned}
k &= \sum_{t \in \mathcal{U}} p(t) \log_2 \frac{1}{l(t)} \\
&= \sum_{t \in \mathcal{U}} \frac{1}{16} \log_2 \frac{1}{1/4} \\
&= 16(\frac{1}{16} \log_2 \frac{1}{1/4}) \\
&= \log_2 \frac{1}{1/4} \\
&= \log_2 4 \\
&= 2 \text{ bits}
\end{aligned}$$



Communicative cost: Example of a discrete categorizer

universe $\mathcal{U} = \{i_1, i_2, ..., i_{16}\}$

category partition $\mathcal{P} =$

$=$

speaker's lexicon $\mathcal{S} =$

listener's lexicon $\mathcal{L} =$

need probabilities $p =$

speaker distributions (for each meaning) $s_1 =$

$s_2 =$

...

$s_{16} =$

listener distributions (for each category)

$l_{C_1} = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$

$l_{C_2} = [0, 0, 0, 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, 0, 0, 0, 0, 0]$

$l_{C_3} = [0, 0, 0, 0, 0, 0, 0, 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, 0]$

$l_{C_4} = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}]$

Why 2 bits?

Ideal system:

0000	0100	1000	1100
0001	0101	1001	1101
0010	0110	1010	1110
0011	0111	1011	1111

4-bit signals

(1 signal for every meaning)

Actual system: 00 01 10 11 2-bit signals

(Pressure from leaning prefers more compressed system)

Loss of information on every communicative episode:
4 bits – 2 bits = 2 bits

$$k = \sum_{t \in \mathcal{U}} p(t) \log_2 \frac{1}{l(t)}$$

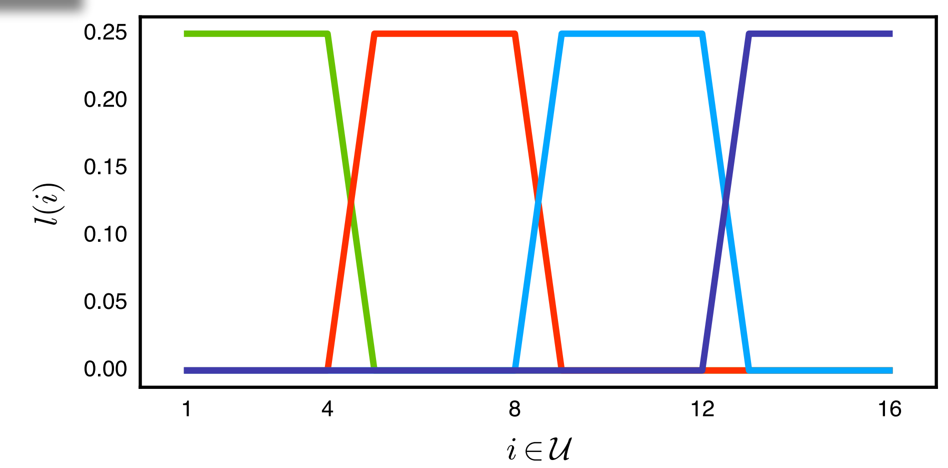
$$= \sum_{t \in \mathcal{U}} \frac{1}{16} \log_2 \frac{1}{1/4}$$

$$= 16(\frac{1}{16} \log_2 \frac{1}{1/4})$$

$$= \log_2 \frac{1}{1/4}$$

$$= \log_2 4$$

$$= 2 \text{ bits}$$



Communicative cost: Listener distributions

Humans aren't discrete categorizers; in human cognition, we see two effects:

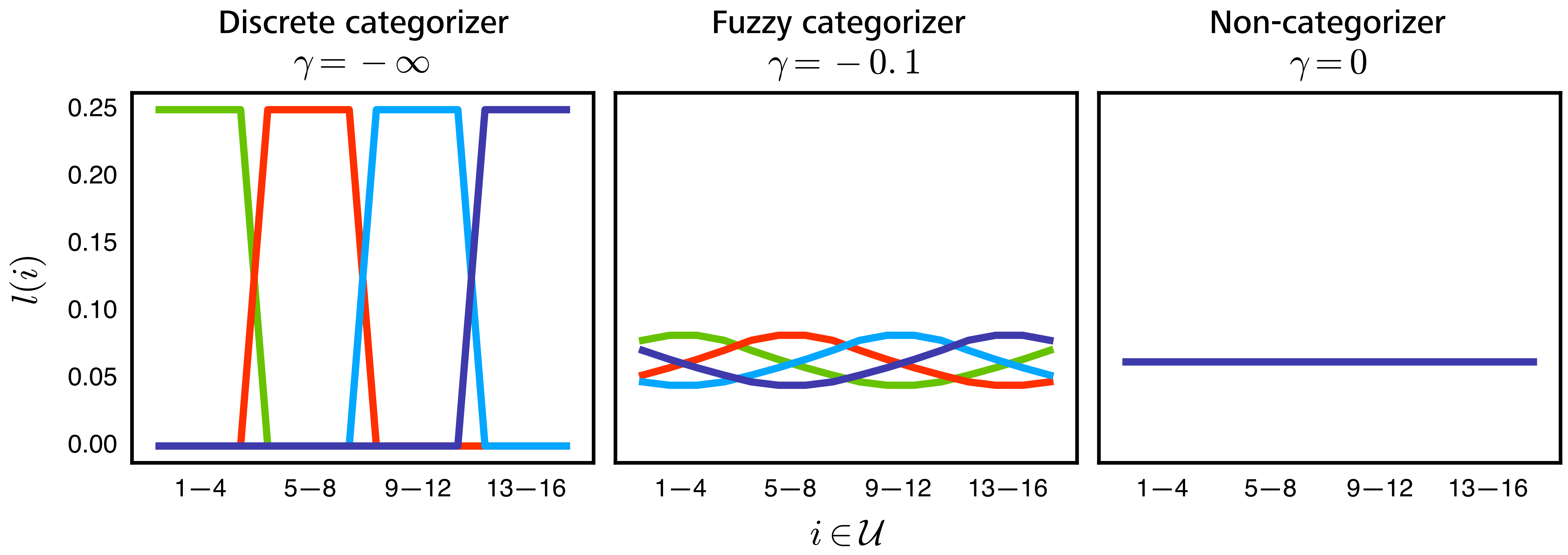
(a) within-category prototypicality

(b) across-category fuzziness

Instead, the listener distributions can be modelled as Gaussians:

$$l_C(i) \propto \sum_{j \in \mathcal{U}} e^{\gamma d(i,j)}$$

where γ allows you to model various types of categorizer



Communicative cost: Example of a fuzzy categorizer

$$k = \sum_{t \in \mathcal{U}} p(t) \log_2 \frac{1}{l(t)}$$
$$= 3.636 \text{ bits}$$

universe $\mathcal{U} = \{i_1, i_2, ..., i_{16}\}$

category partition $\mathcal{P} = \{C_1, C_2, C_3, C_4\}$
$= \{\{i_1, i_2, i_3, i_4\}, \{i_5, i_6, i_7, i_8\}, \{i_9, i_{10}, i_{11}, i_{12}\}, \{i_{13}, i_{14}, i_{15}, i_{16}\}\}$

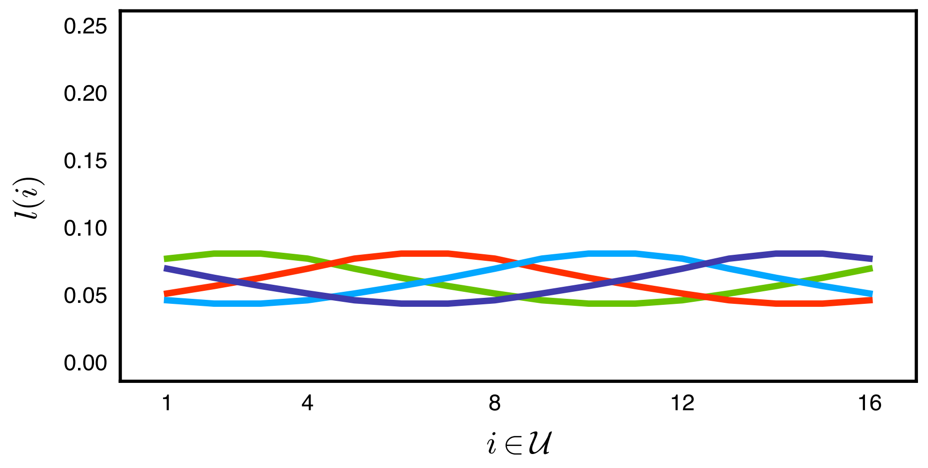
speaker's lexicon $\mathcal{S} = \{C_1 \rightarrow \texttt{00}, C_2 \rightarrow \texttt{01}, C_3 \rightarrow \texttt{10}, C_4 \rightarrow \texttt{11}\}$

listener's lexicon $\mathcal{L} = \{\texttt{00} \rightarrow C_1, \texttt{01} \rightarrow C_2, \texttt{10} \rightarrow C_3, \texttt{11} \rightarrow C_4\}$

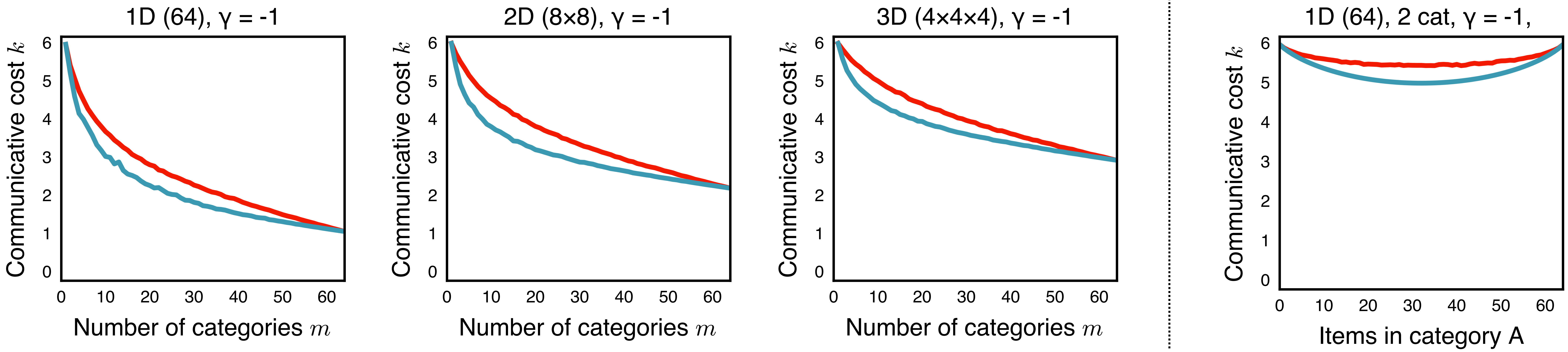
need probabilities $p = [\frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}]$

speaker distributions (for each meaning)
$s_1 = [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$
$s_2 = [0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$
$\ldots$
$s_{16} = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]$

listener distributions (for each category)
$l_{C_1} = [.079, .082, .082, .079, .071, .064, .058, .053, .048, .045, .045, .048, .053, .058, .064, .071]$
$l_{C_2} = [.053, .058, .064, .071, .079, .082, .082, .079, .071, .064, .058, .053, .048, .045, .045, .048]$
$l_{C_3} = [.048, .045, .045, .048, .053, .058, .064, .071, .079, .082, .082, .079, .071, .064, .058, .053]$
$l_{C_4} = [.071, .064, .058, .053, .048, .045, .045, .048, .053, .058, .064, .071, .079, .082, .082, .079]$



Communicative cost: Six predictions



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

Convexity A system of convex categories (blue) is more informative than a system of nonconvex categories (red)

Balanced categories A system of equally sized categories is more informative than a system of unequally sized categories

Discreteness A system of discrete categories is more informative than a system of fuzzy categories

Dimensionality A system that uses many dimensions is less (?) informative than a system that uses few dimensions

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

Communicative cost: Summary

When communicating, interlocutors want to align as closely as possible on the same meaning in the face of:

(a) the speaker’s uncertainty about the true meaning

(b) the lossy information conveyed to the listener by a general category

Communicative cost tells us how ‘good’ a partition is in the context of using it for communication

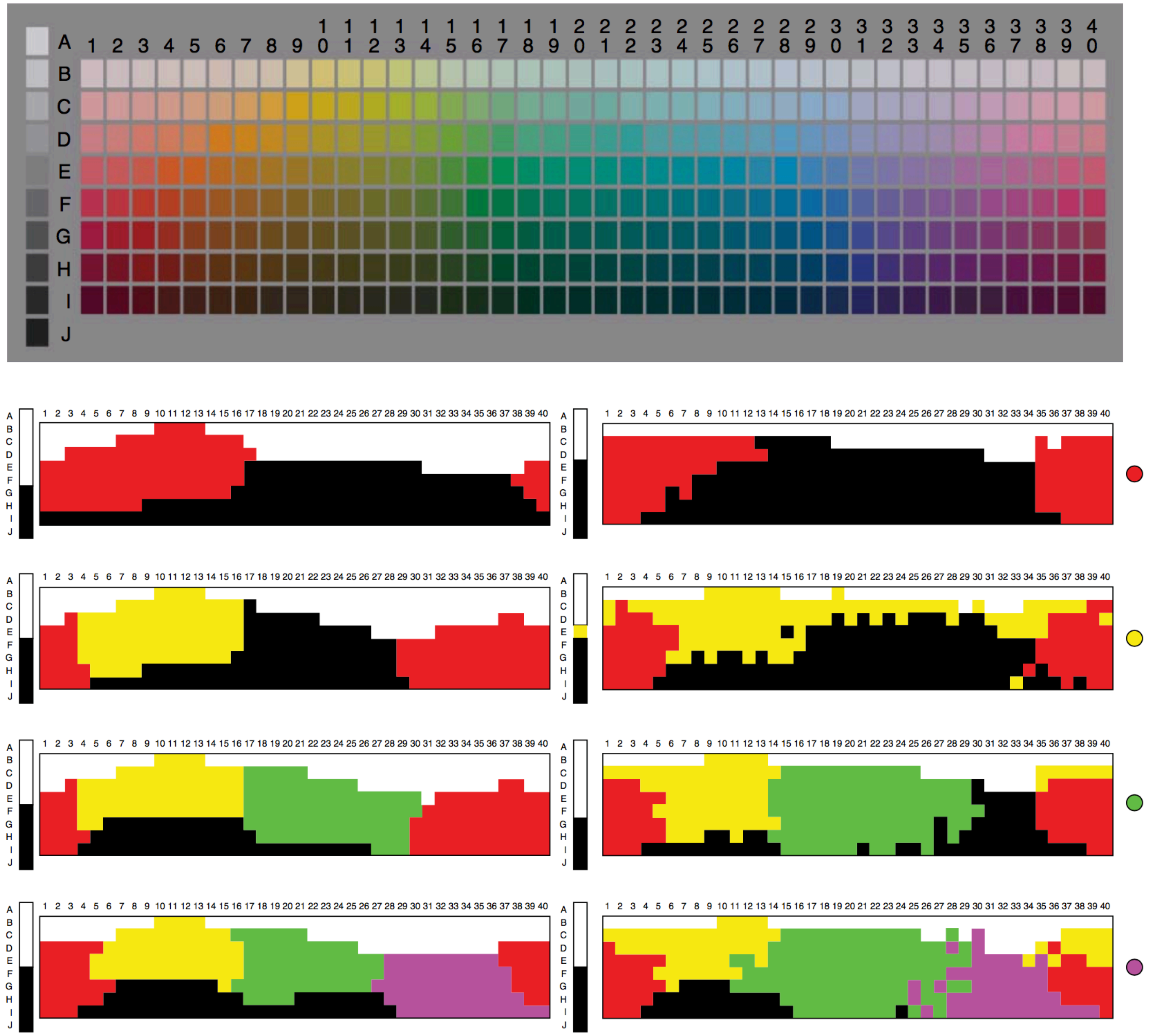
A good partition results, on average, in low information loss (it has low communicative cost)

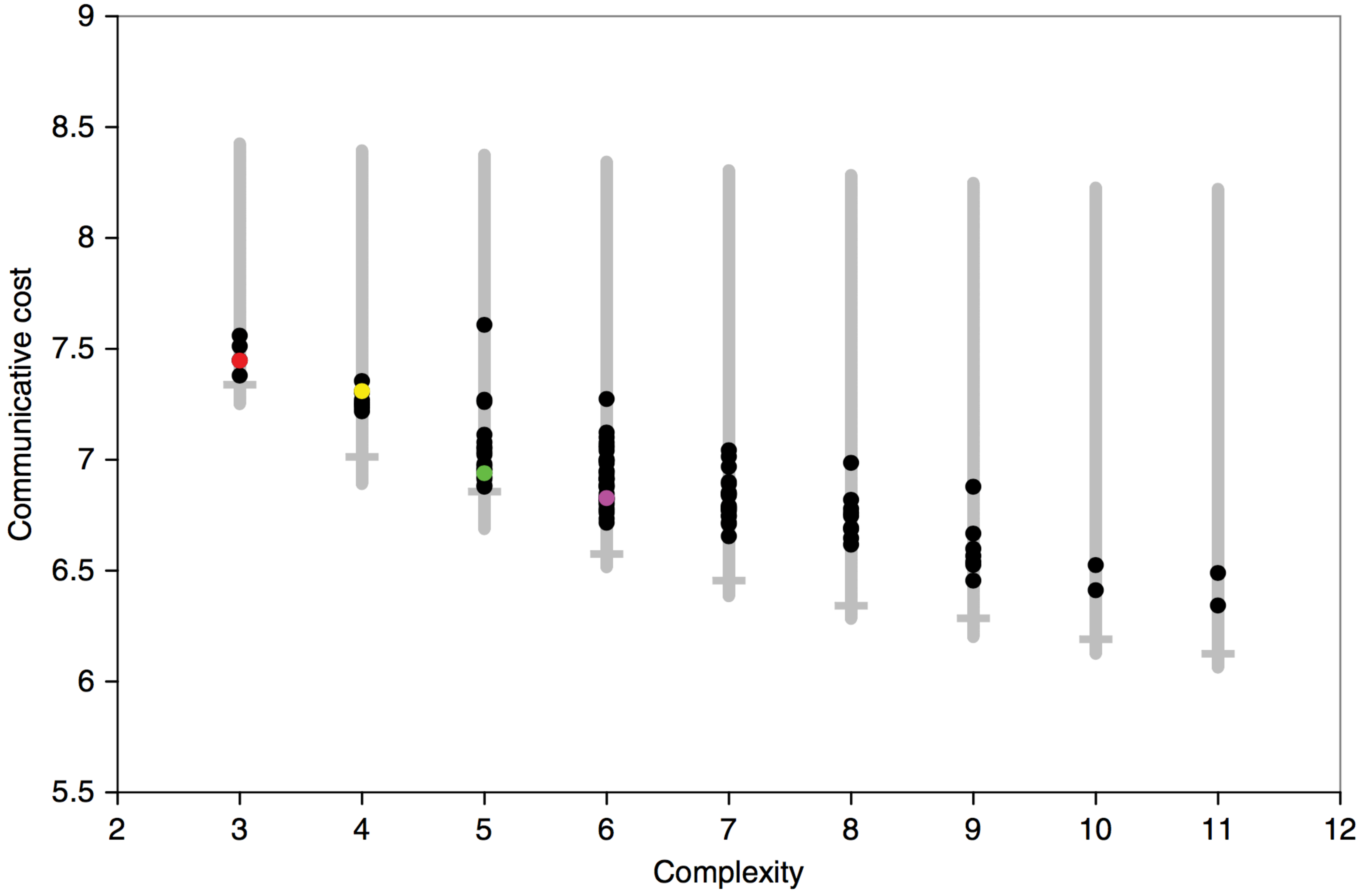
This model makes various predictions about what makes a language informative

Studies of informativeness

Colour categories are informative for given complexity

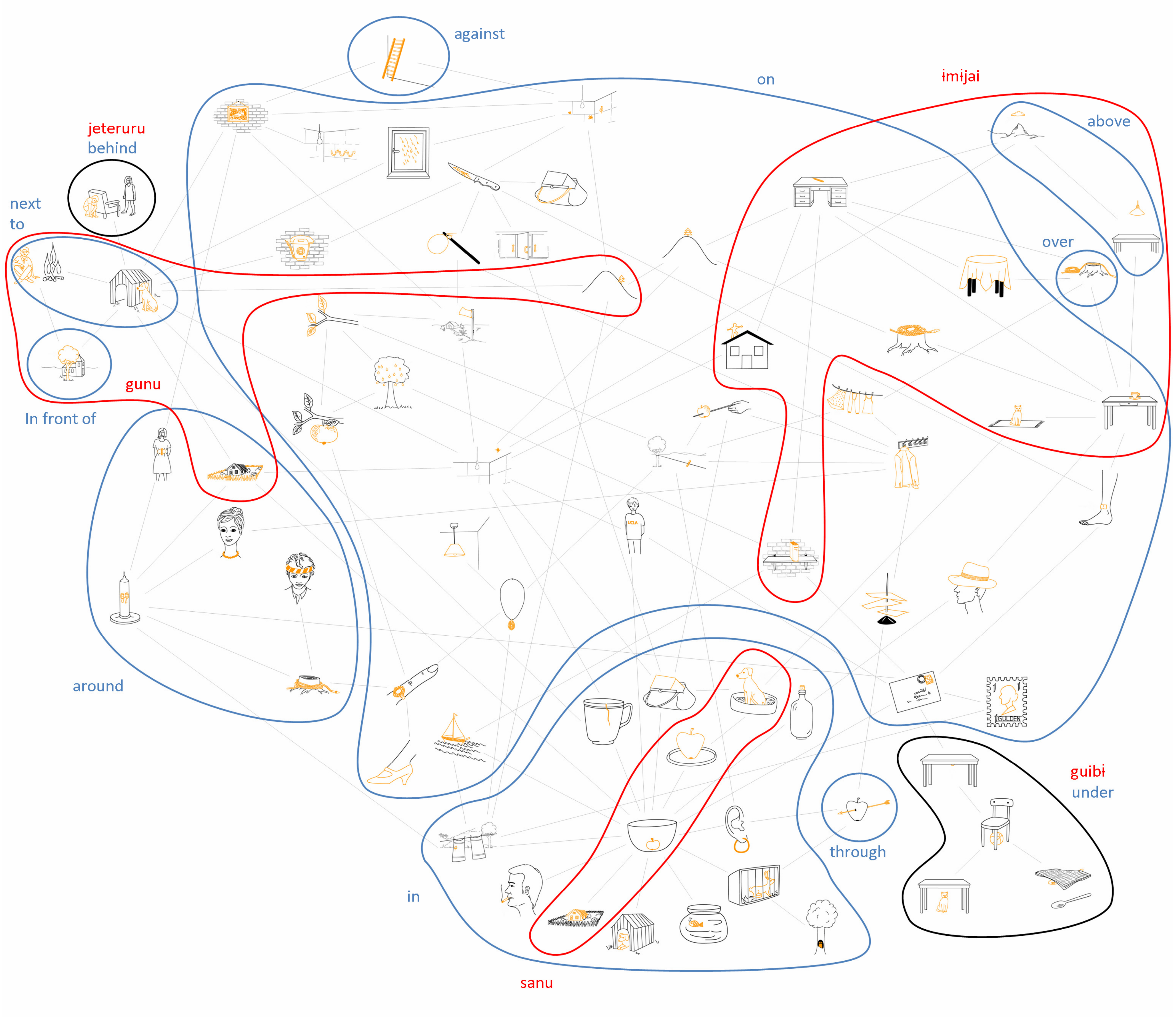
Regier, Kemp, & Kay (2015); reanalysed from Regier, Kay, & Khetarpal (2007)

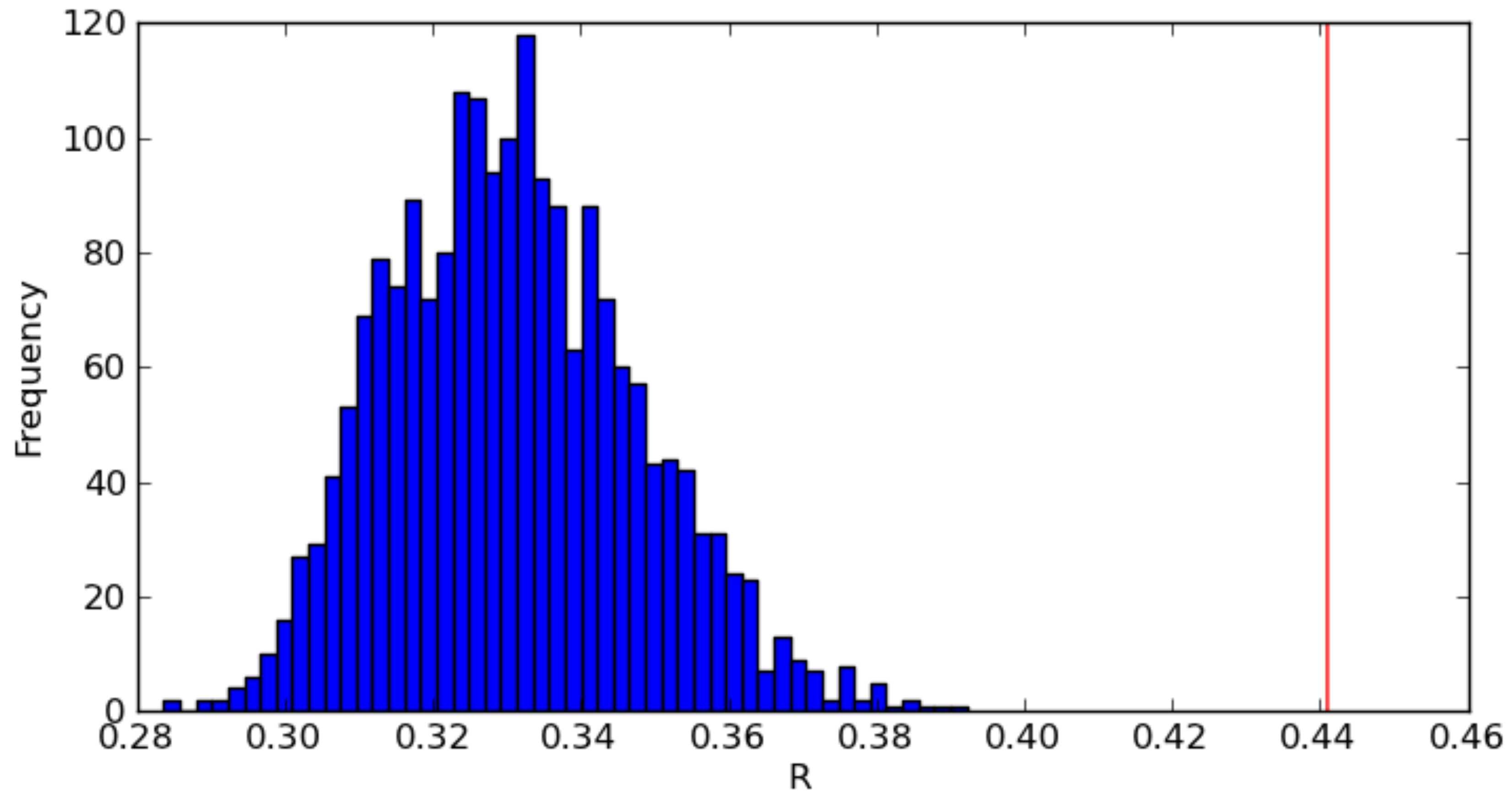




Spatial terms are more informative than chance

Khetarpal, Neveu, Majid, Michael, & Regier (2013); data from Levinson et al. (2003)

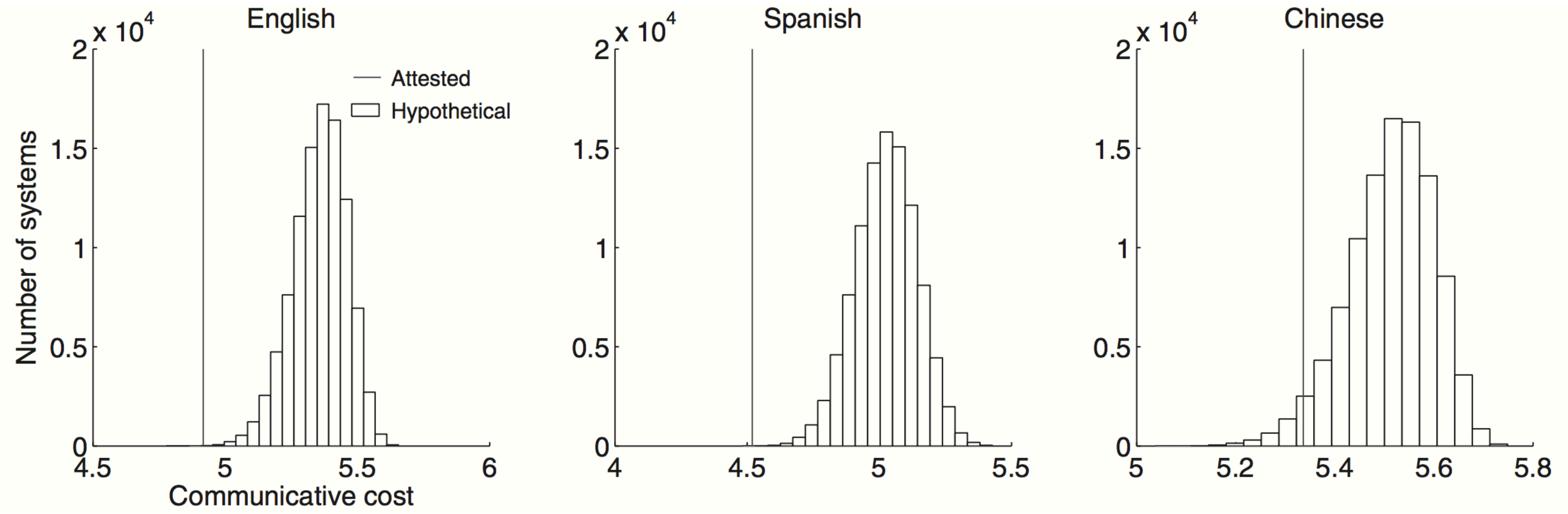




Language	Result
Basque	> 99.95%
Dutch	> 100.00%
English	> 100.00%
Ewe	> 99.95%
Lao	> 96.20%
Lavukaleve	> 99.75%
Maijɨki	> 100.00%
Tiriyó	> 100.00%
Trumai	> 100.00%
Yélî-Dnye	> 97.35%
Yukatek	> 99.95%

Container names are more informative than chance

Xu, Regier, & Malt (2016); data from Malt et al. (1999)



Iterated learning & informativeness

Iterated leaning and informativeness

Carstensen, Xu, Smith, & Regier (2015, p. 303):

> [Our] prior work has also left an important question unaddressed. In a commentary on Kemp and Regier's (2012) kinship study, Levinson (2012) pointed out that although [our] 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.

Although their model of informativeness is framed in terms of the communicative benefit, in this paragraph they appear to be open to the idea that there could be an explanation from learning

Iterated leaning and informativeness

If true, this doesn’t sit well with our (post-2015?) framework which says that:

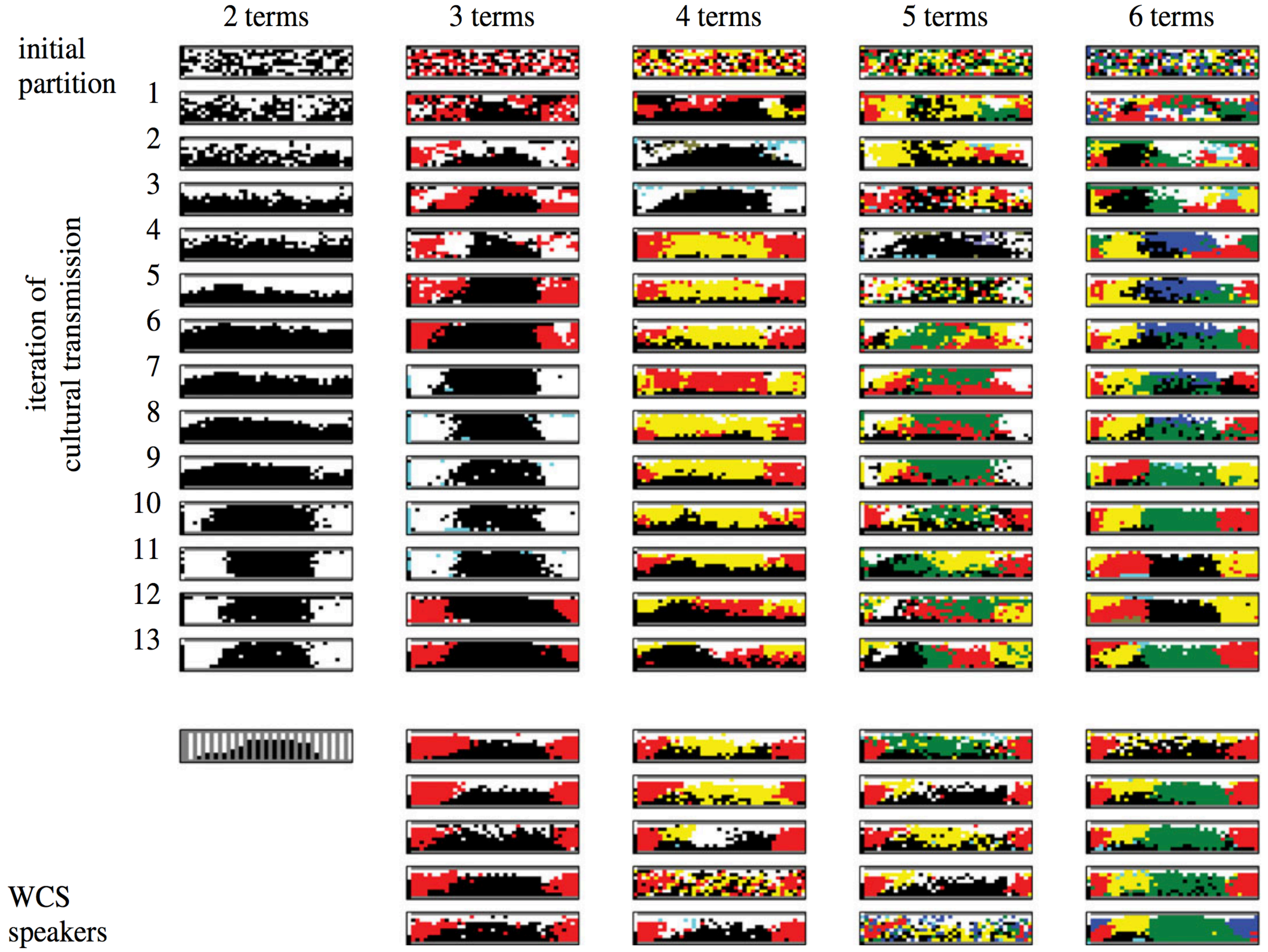
(a) communication promotes informativeness/expressivity, and

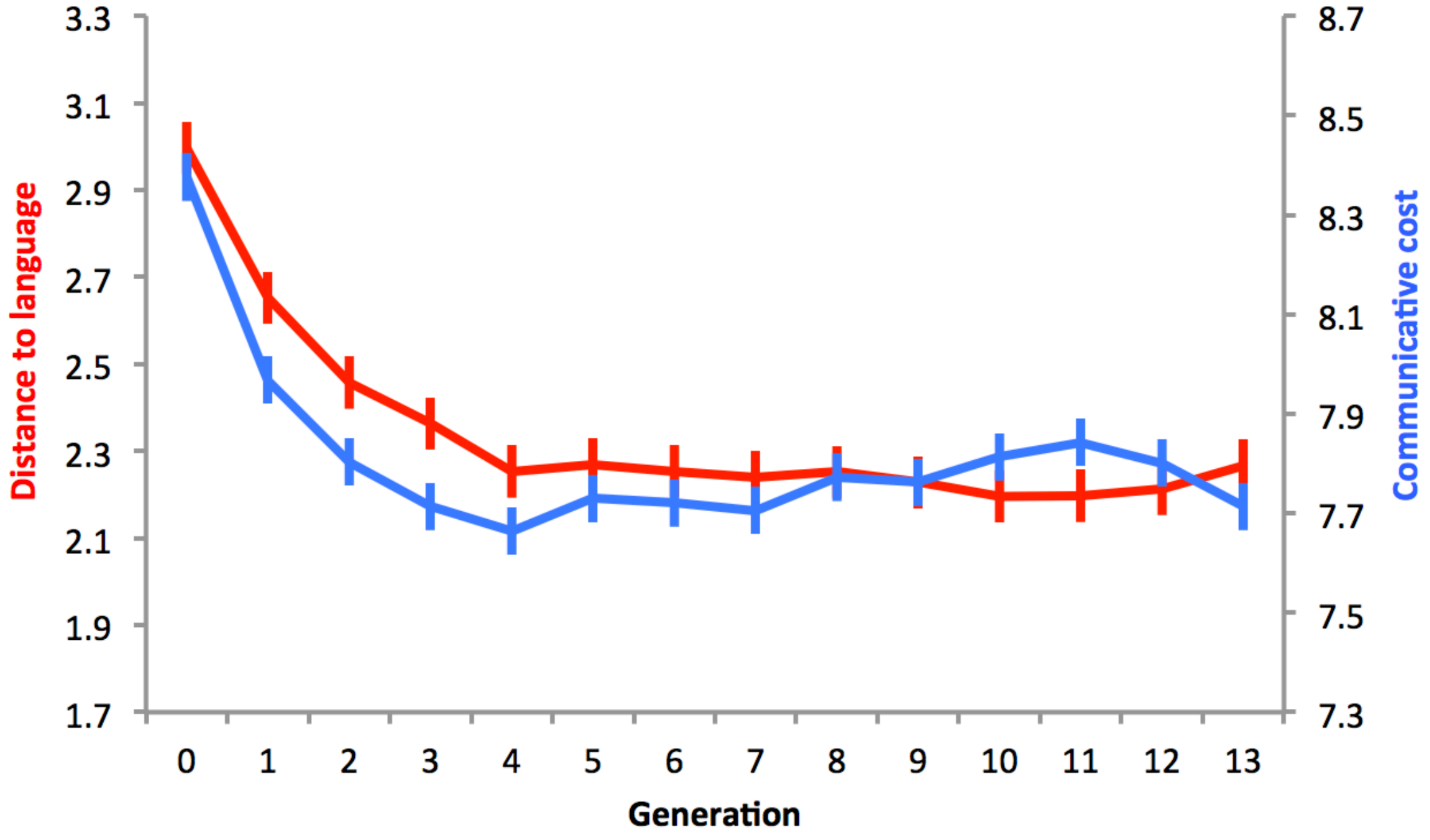
(b) (iterated) learning promotes simplicity/compressibility

However, they present two iterated learning studies in support of this idea

Study 1: Iterated learning gives rise to informative colour categories

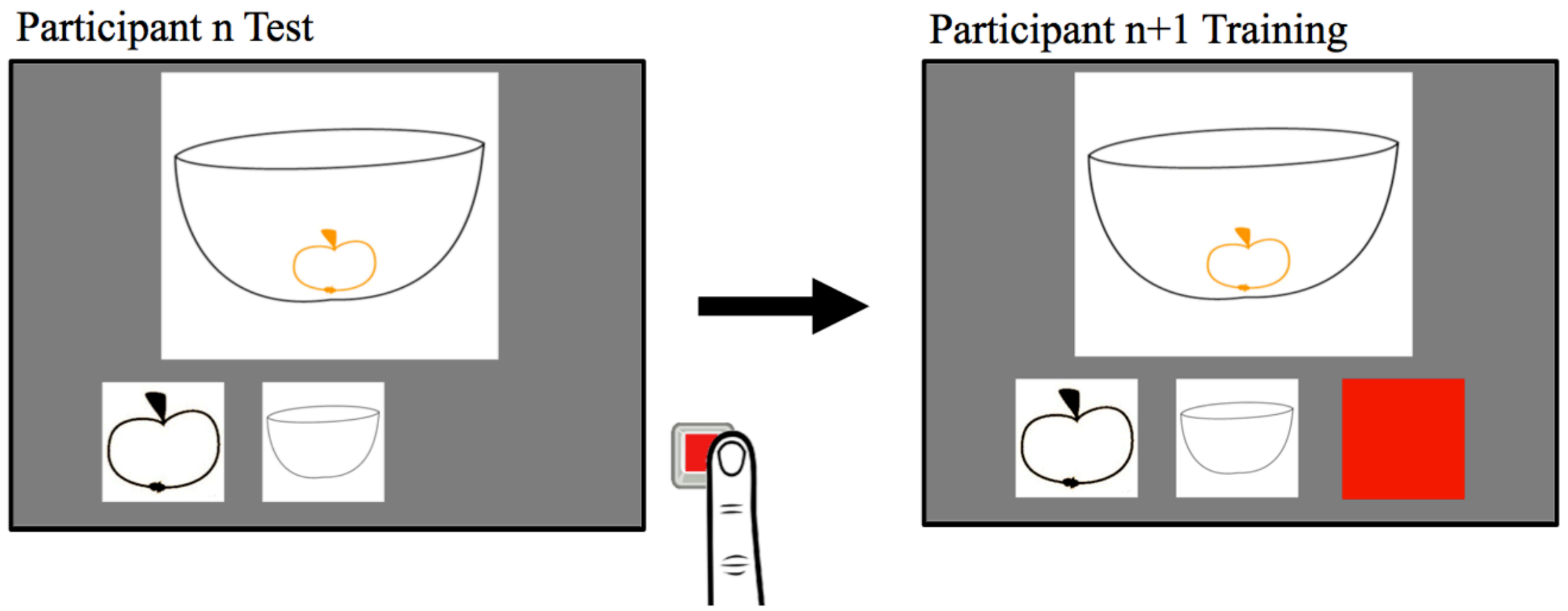
Carstensen, Xu, Smith, & Regier (2015); data from Xu, Dowman, & Griffiths (2013)

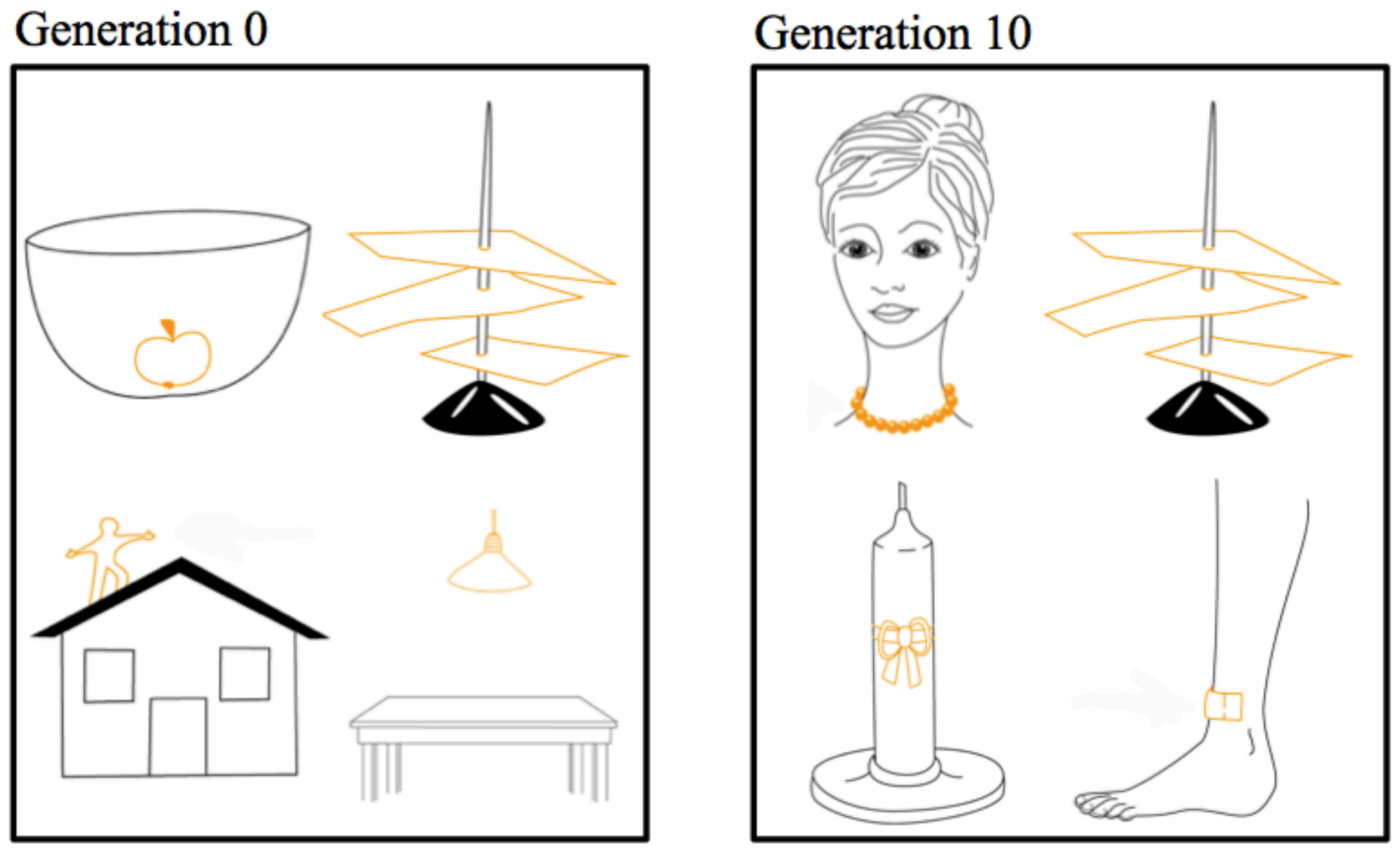


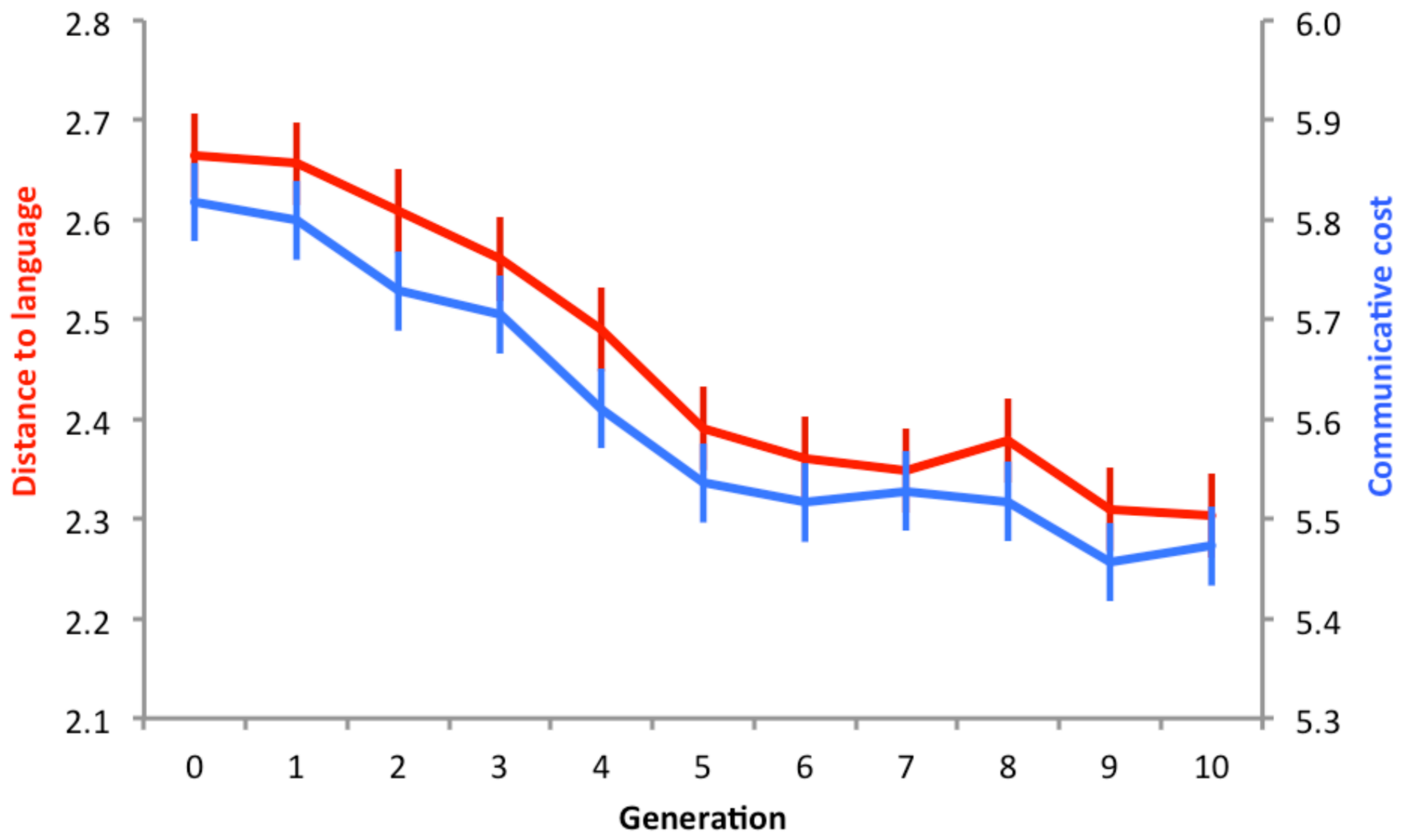


Study 2: Iterated learning gives rise to informative spatial terms

Carstensen, Xu, Smith, & Regier (2015)







Iterated learning promotes informativeness?

The paper sets out to establish what process gives rise to informative categories

Their results suggest that informative categories may arise cumulatively through iterated learning

The effect can’t be driven by expressivity, since the number of categories is fixed

Problem 1: What’s the mechanism? Why should learning care about informativeness?

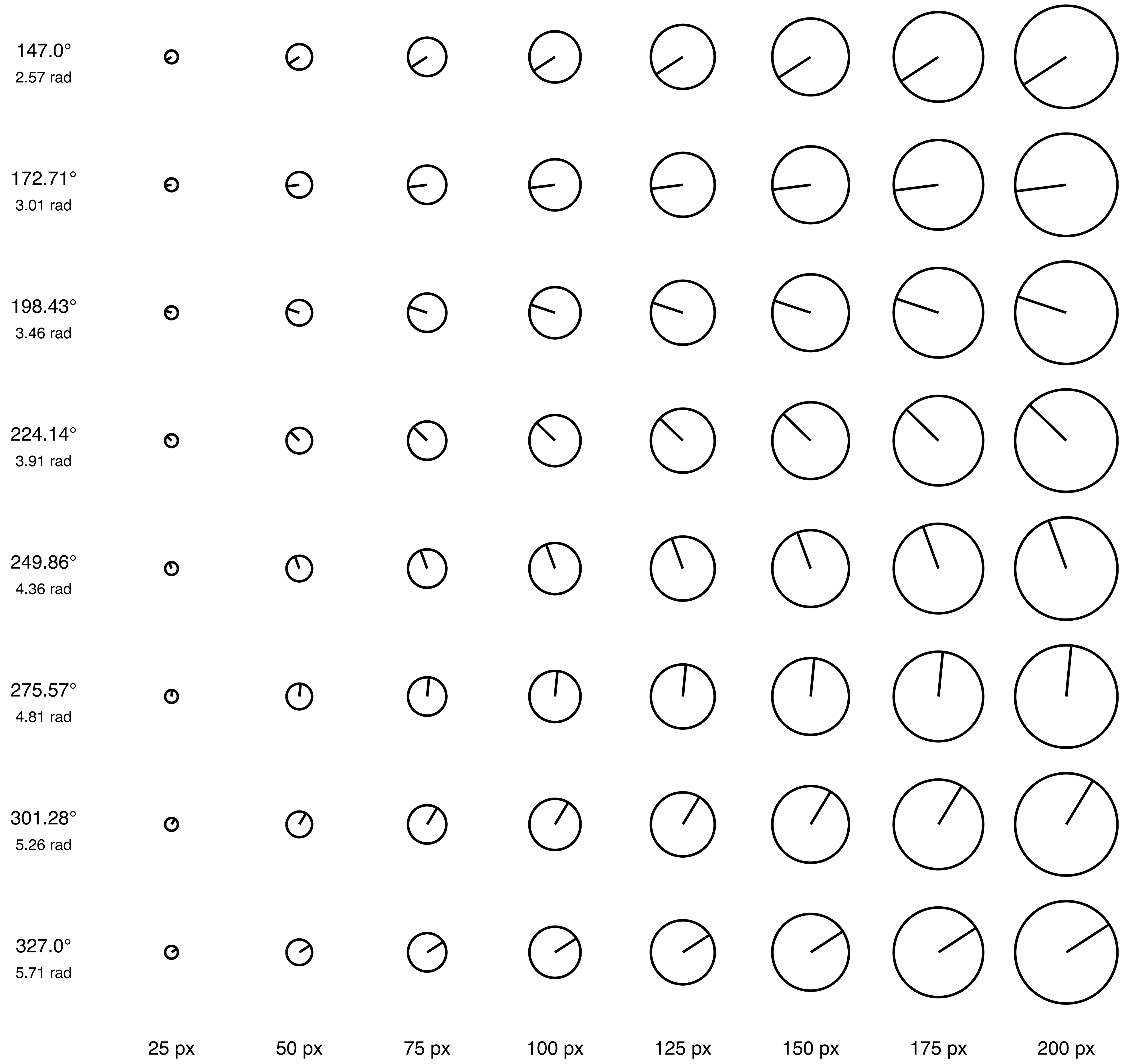
Problem 2: Both experiments only test iterated learning; there is no experiment testing the effect of communication alone

Problem 3: Both experiments force participants to use a certain number of categories, so our prediction that learning should lead to simplicity can’t be observed

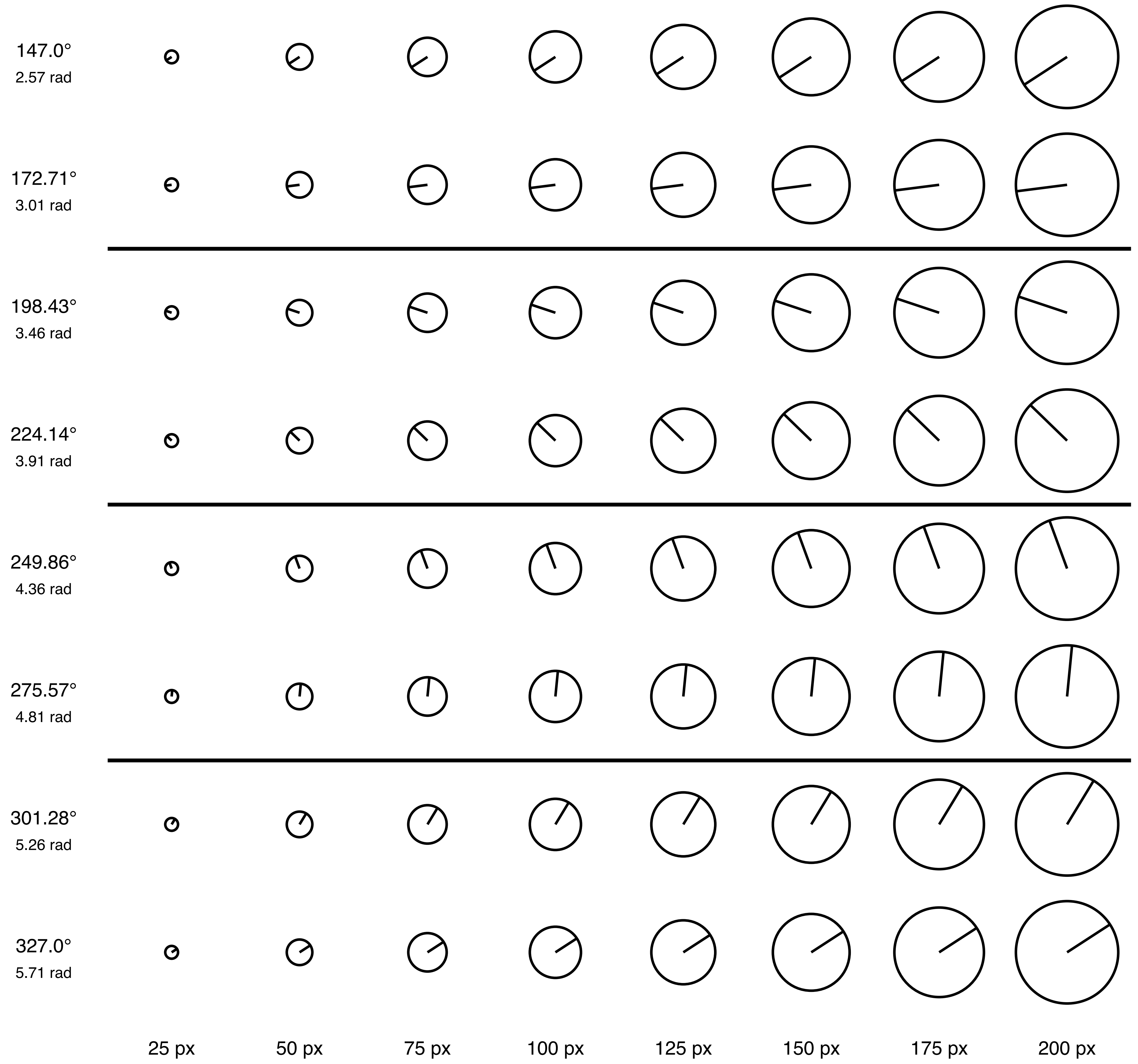
Solution? Since the languages can’t simplify, the only effect a participant can have is to introduce a more sensible structuring of the space; over time, these effects add up to more informative systems

Experiment 1

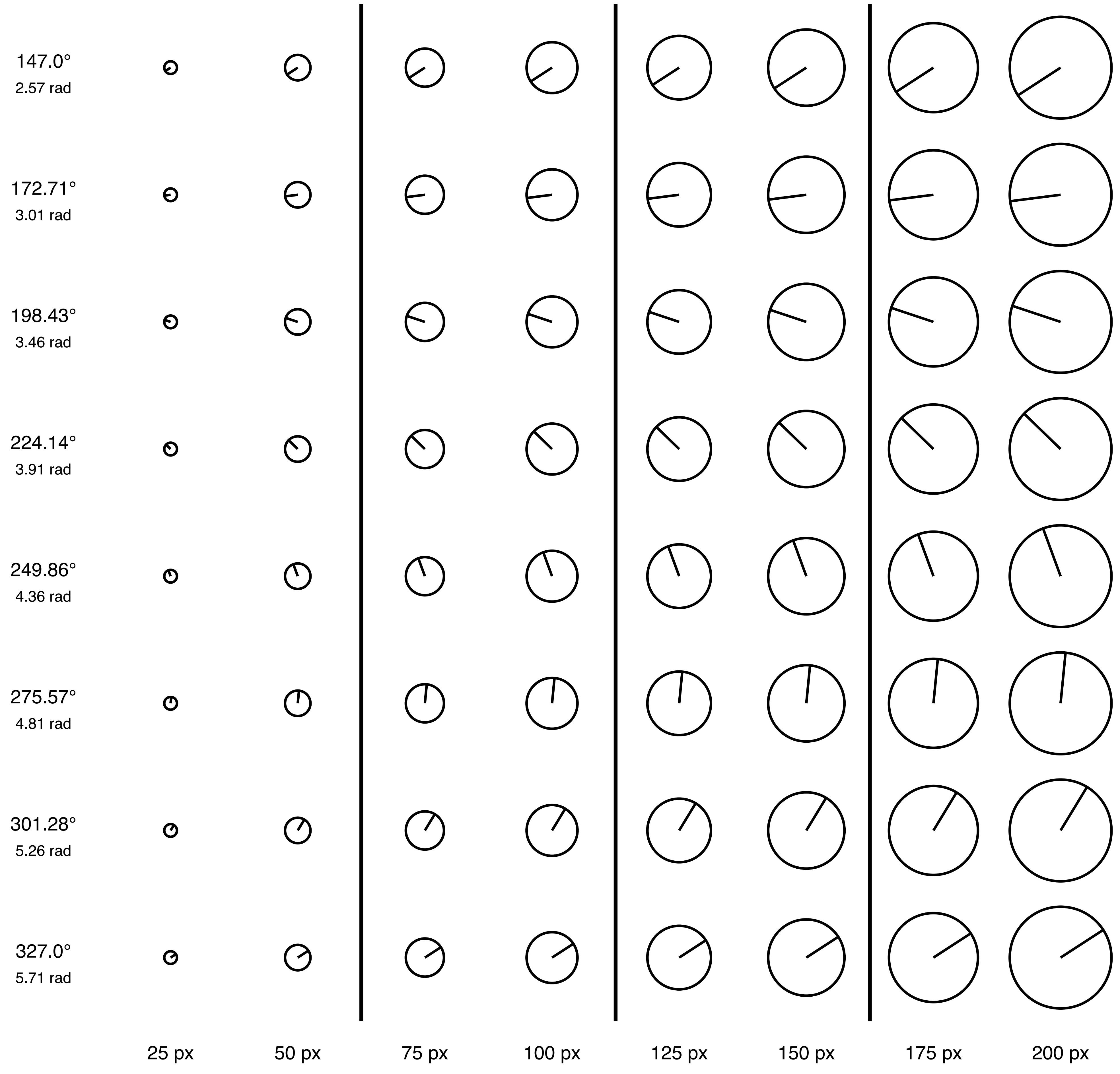
Shepard circles



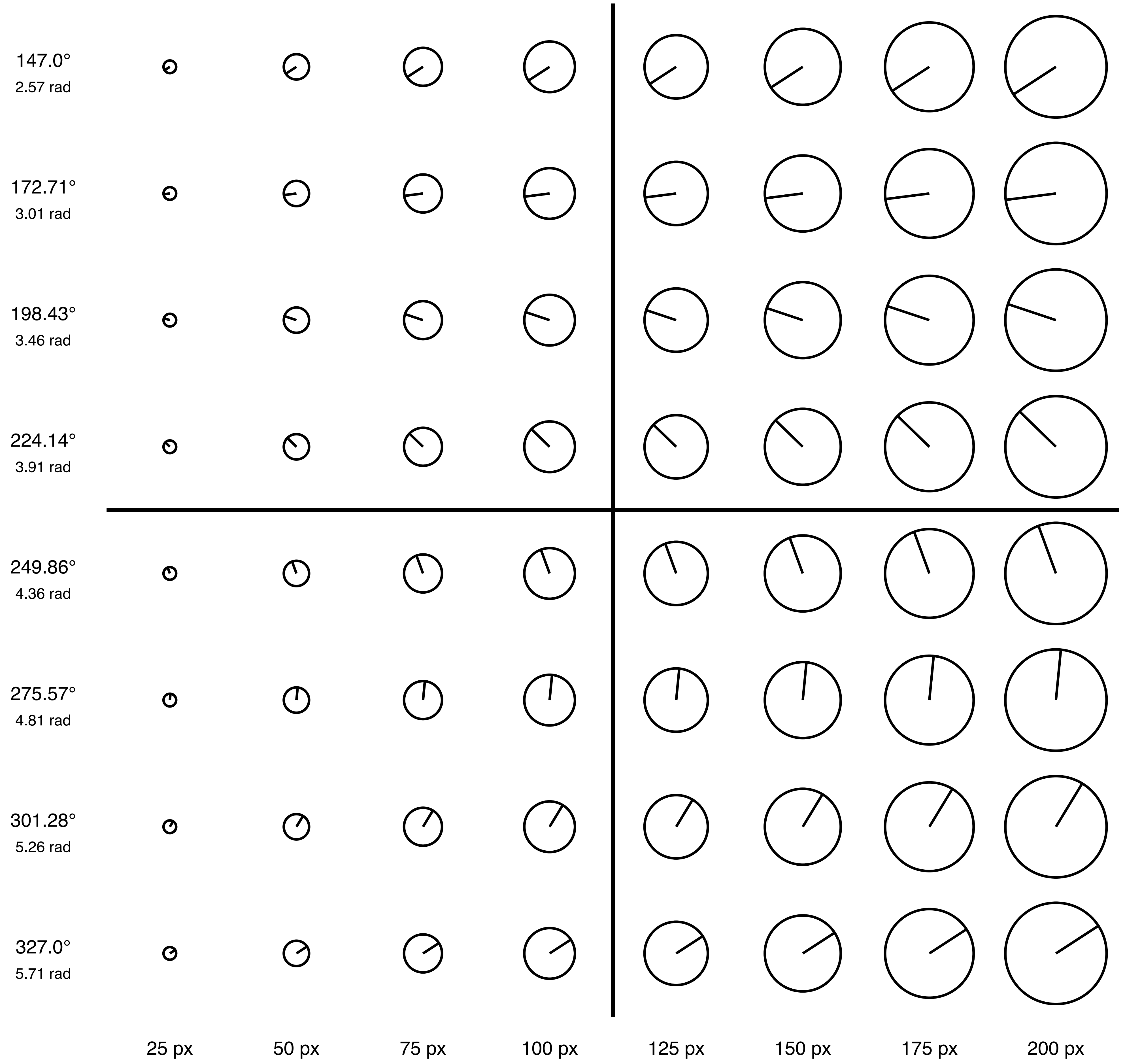
Shepard circles



Shepard circles

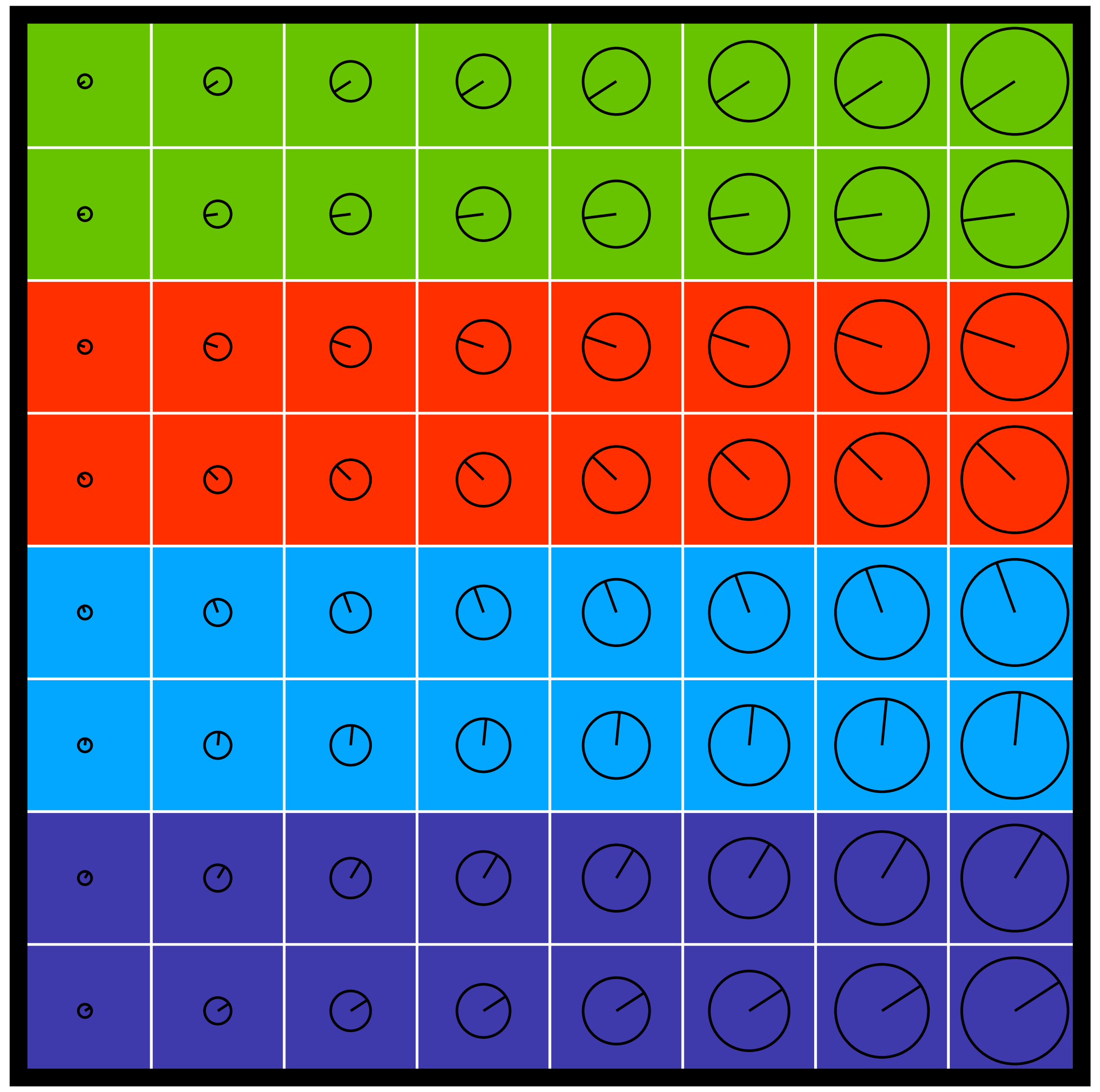


Shepard circles

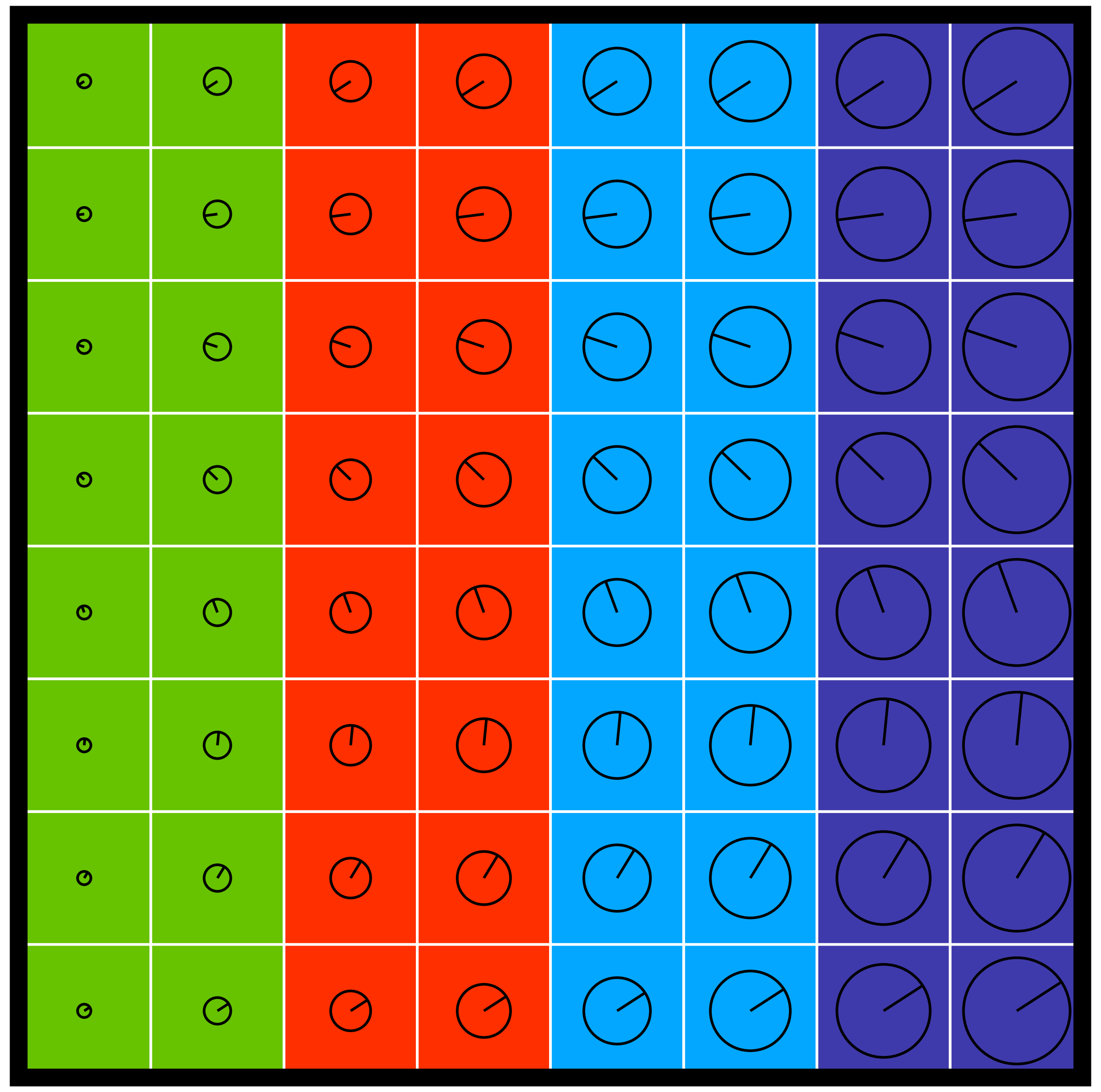


Squares and stripes: Predictions

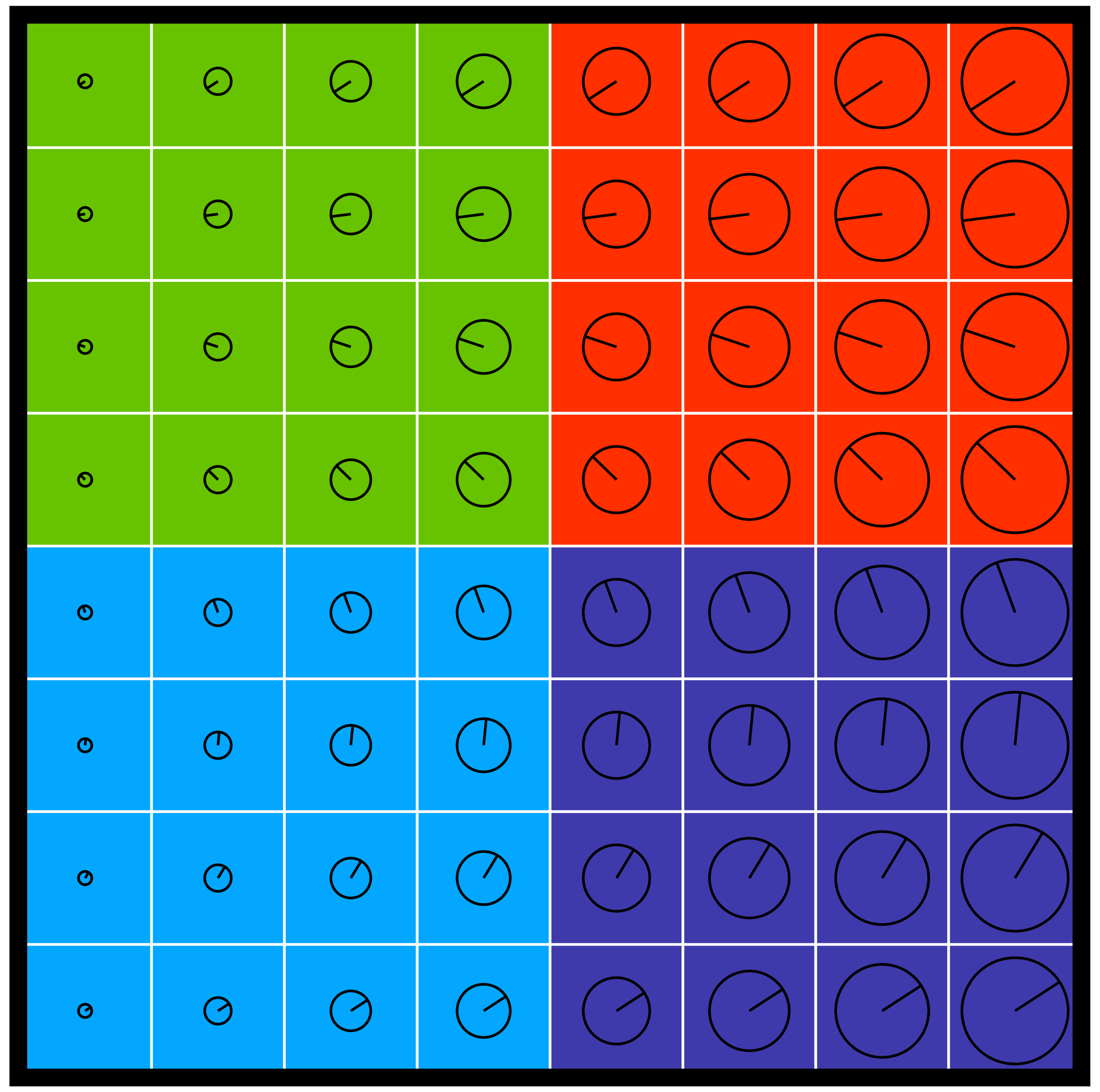
Angle-only



Size-only



Angle & Size



Easy to learn but low informativeness

Informative but hard to learn

Experimental design

20-minute online experiment run on CrowdFlower

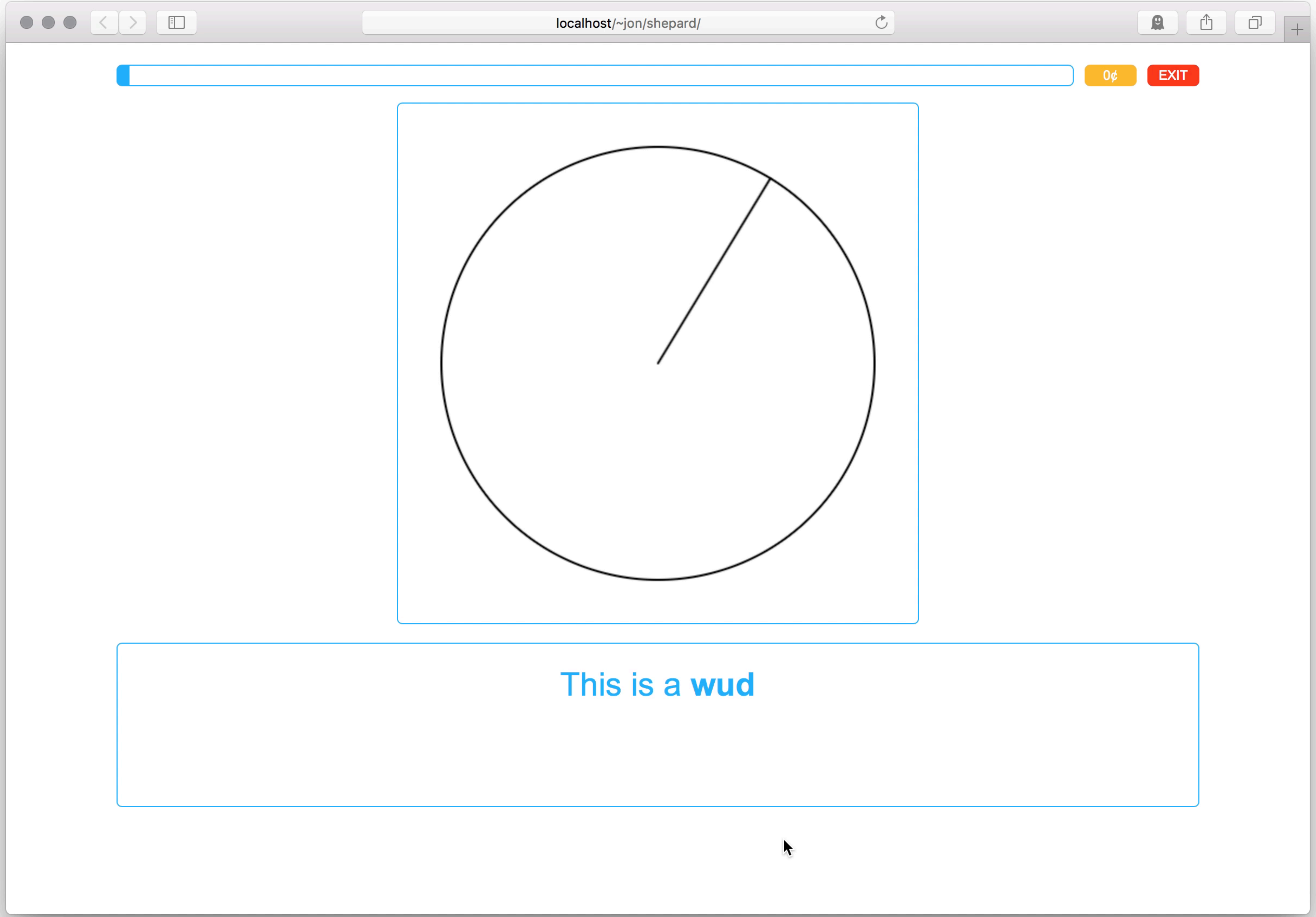
40 participants per condition

Paid \$3 + bonuses for getting answers correct (potentially up to \$4.92)

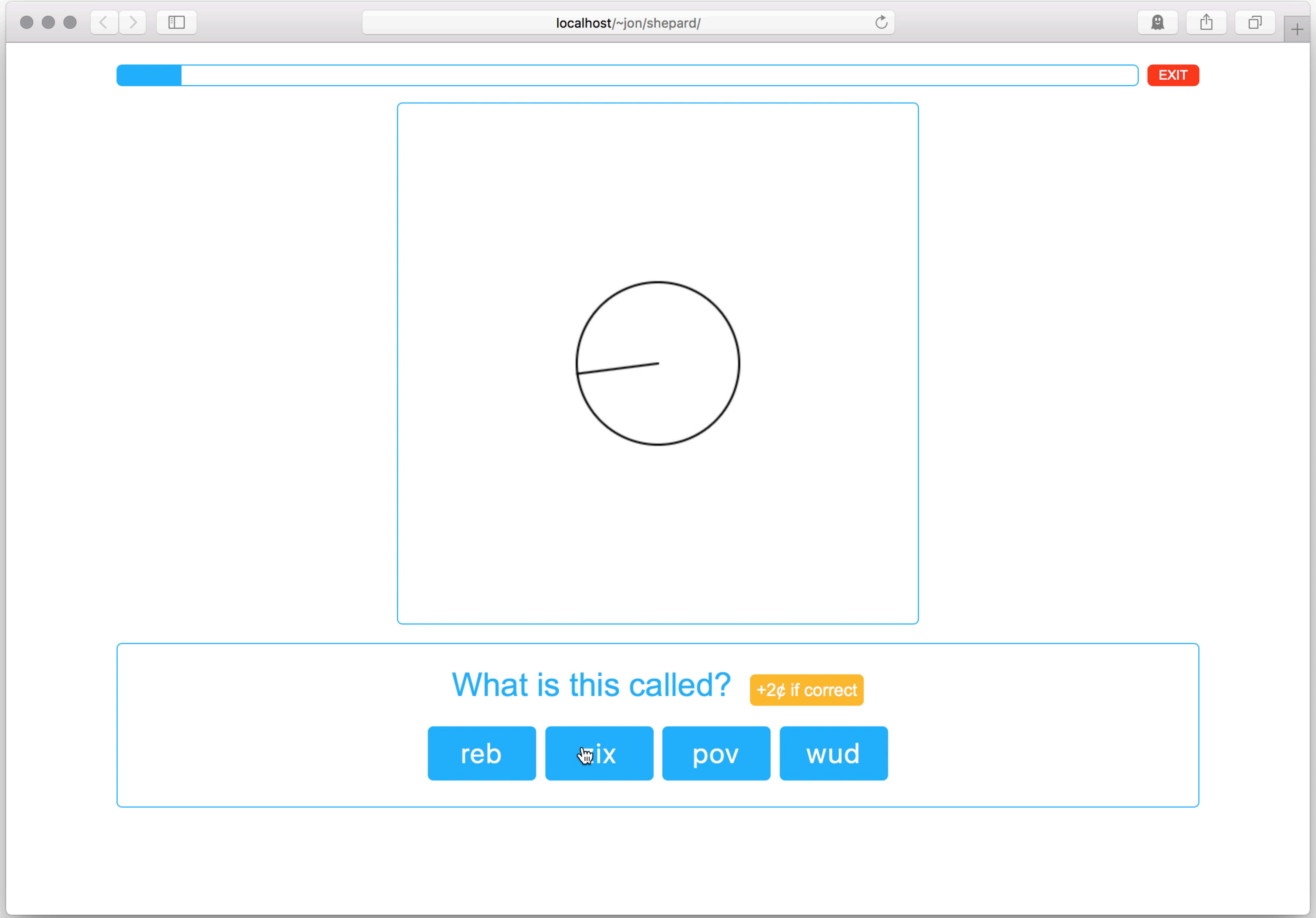
Training phase in which they learn an artificial language

Test phase in which they produce a word for each meaning

Training



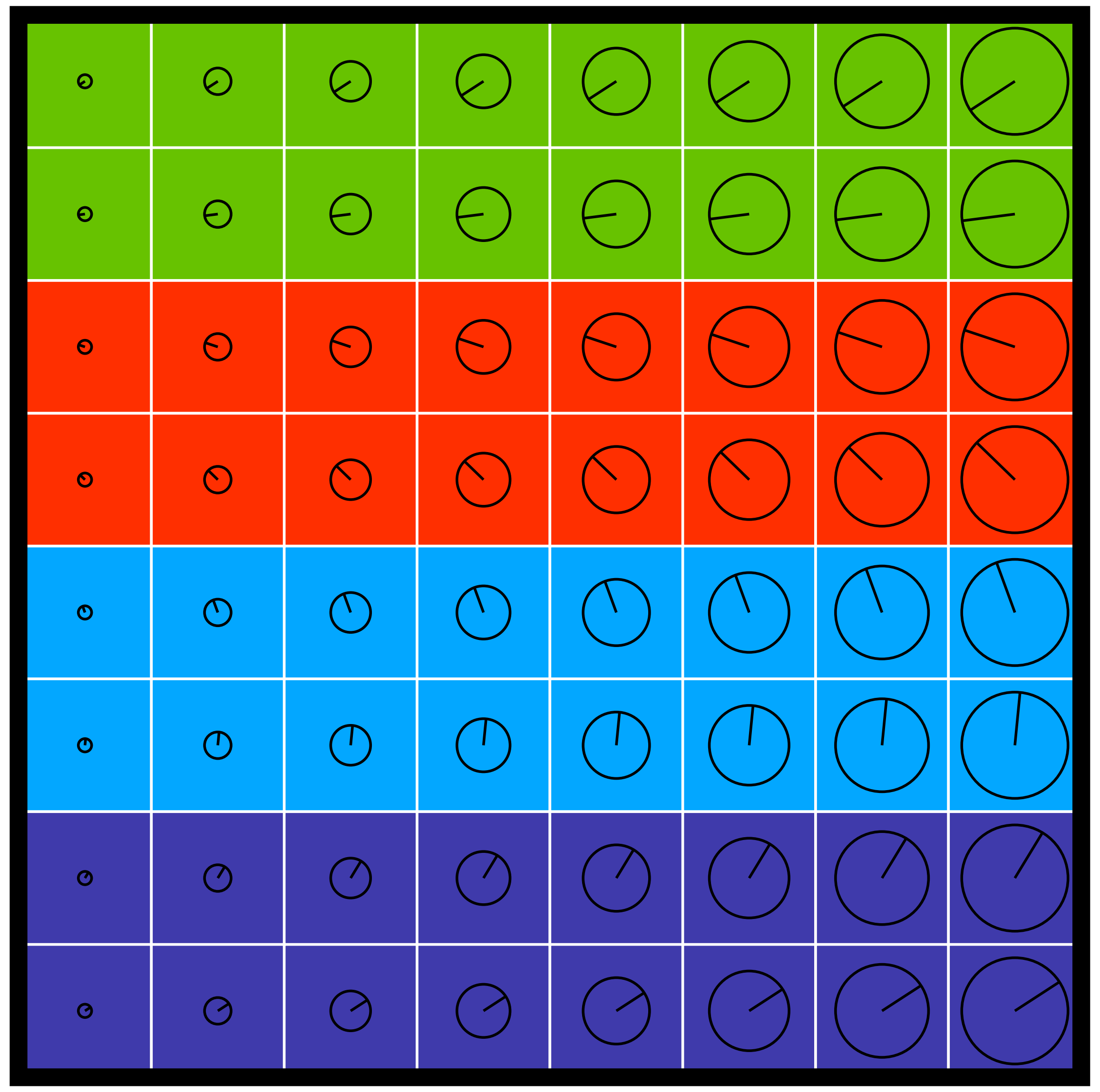
Test

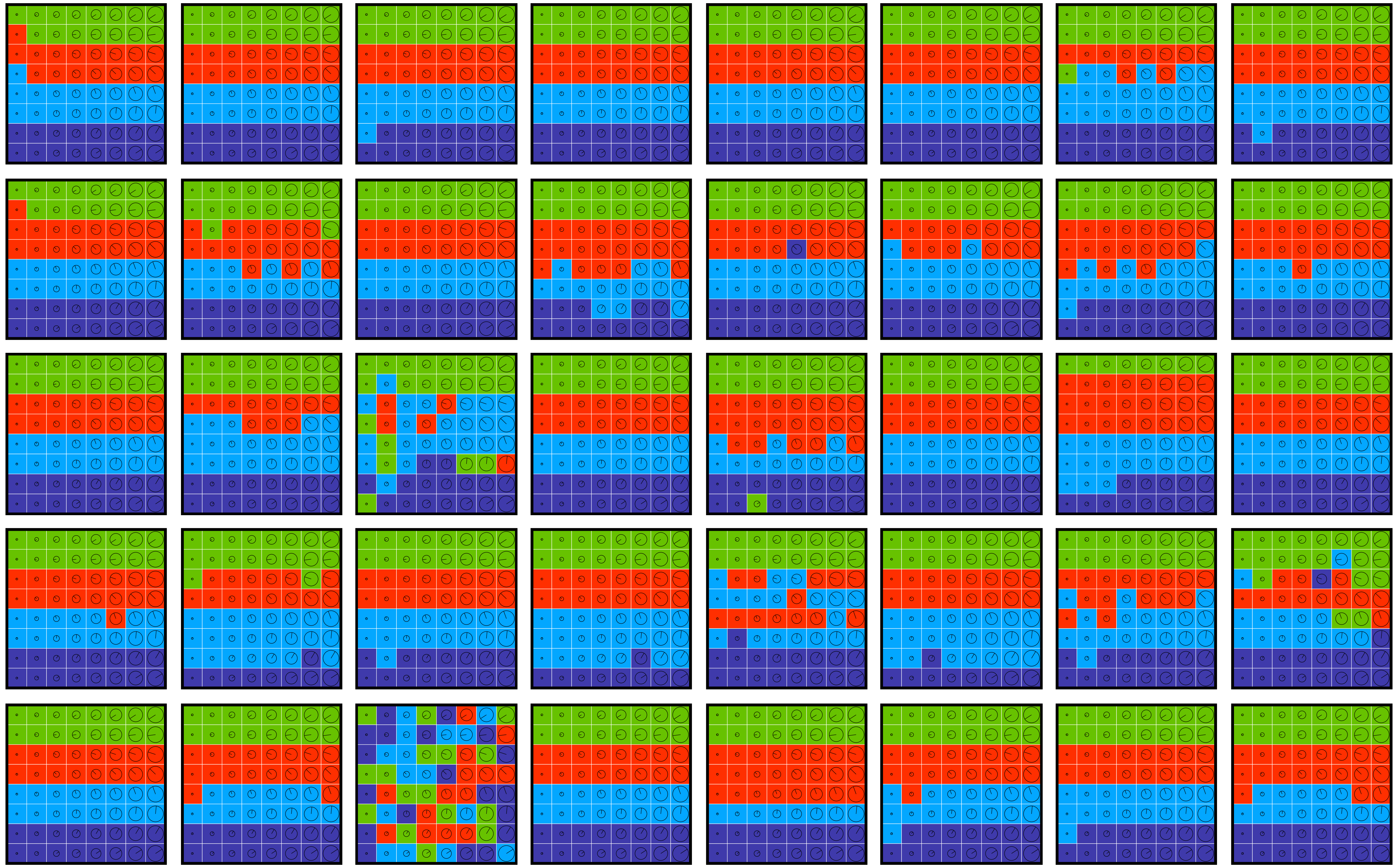


localhost/~jon/shepard/
EXIT
What is this called?
+2¢ if correct
reb
pov
wud

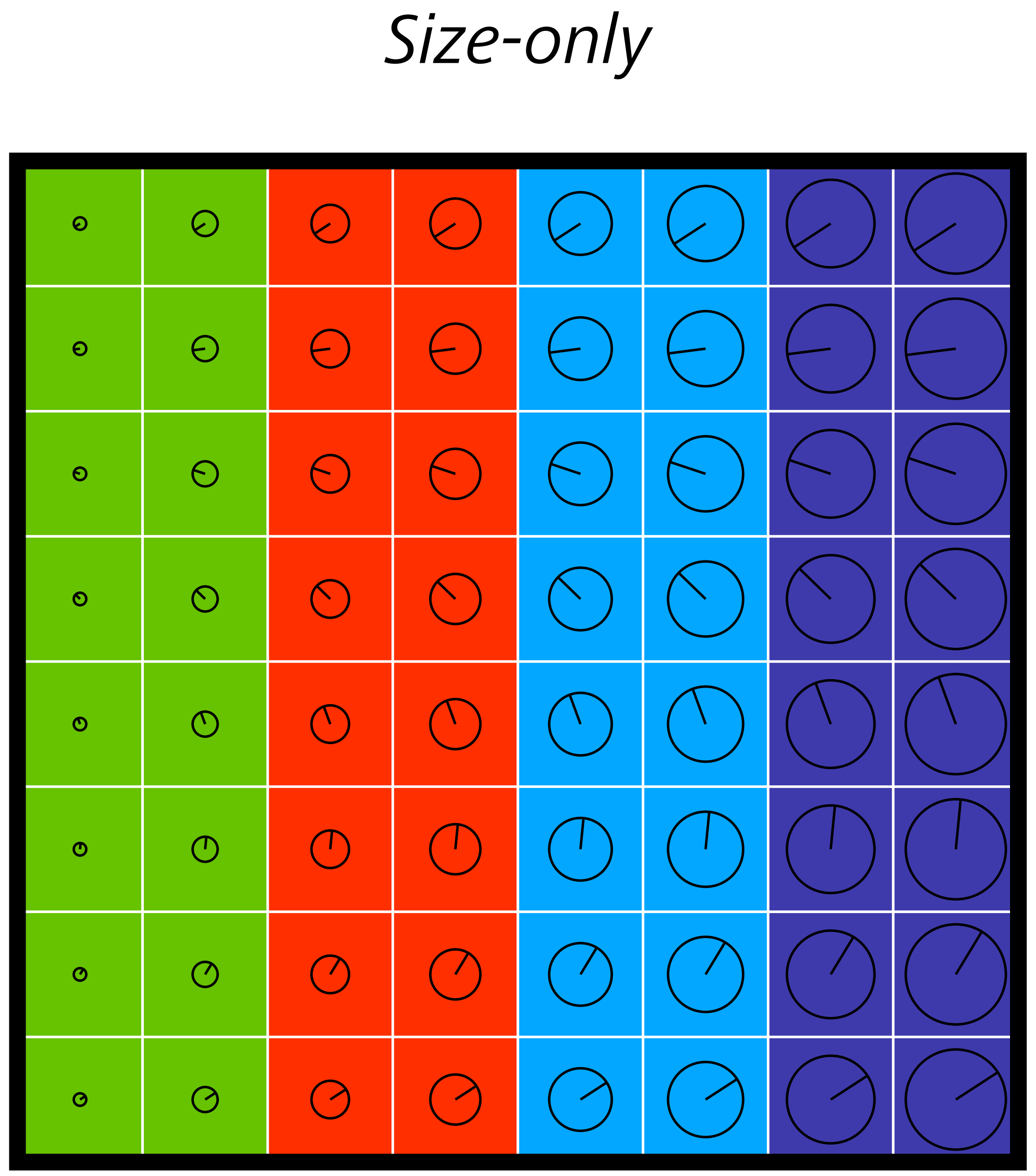
Results

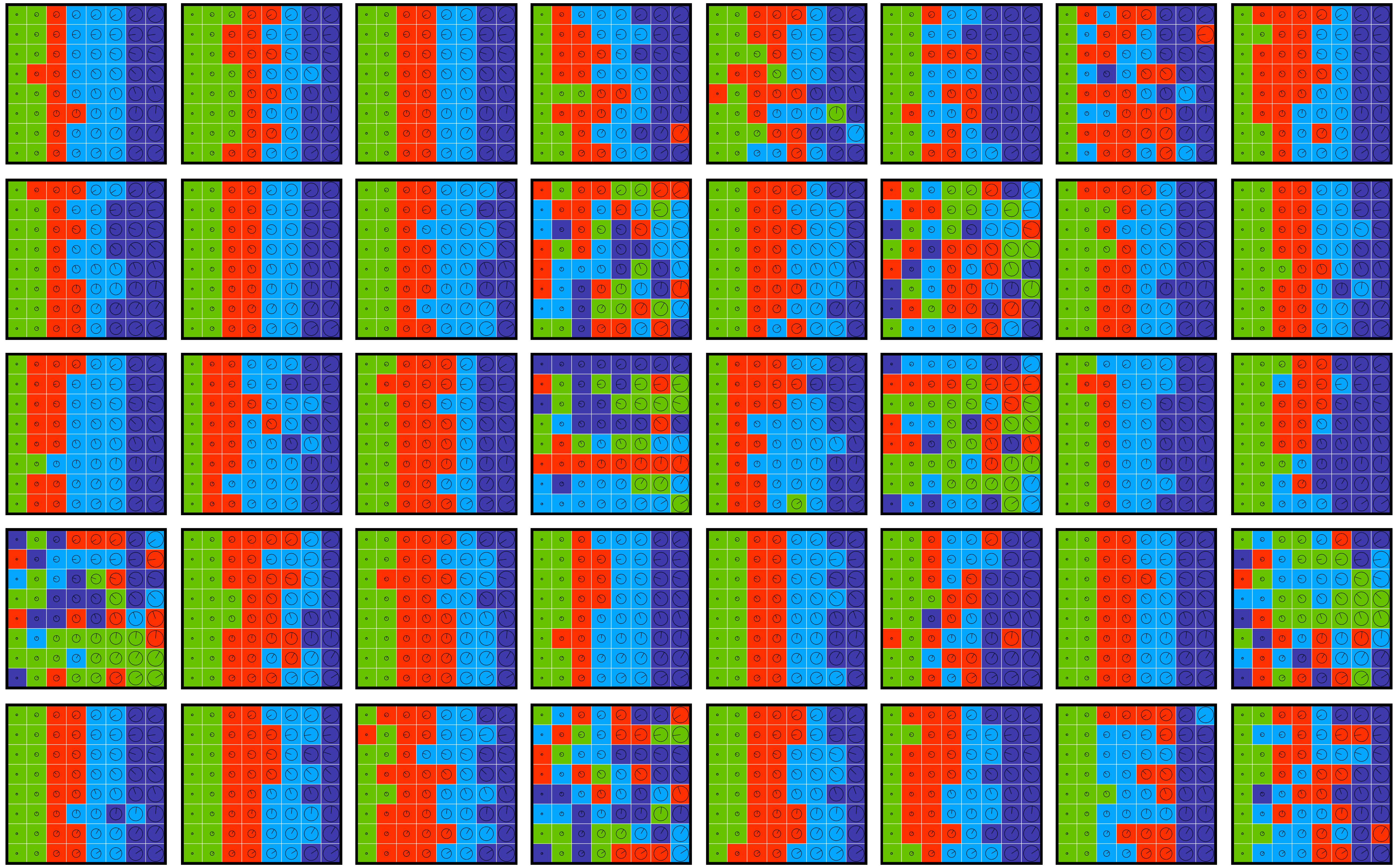
Angle-only





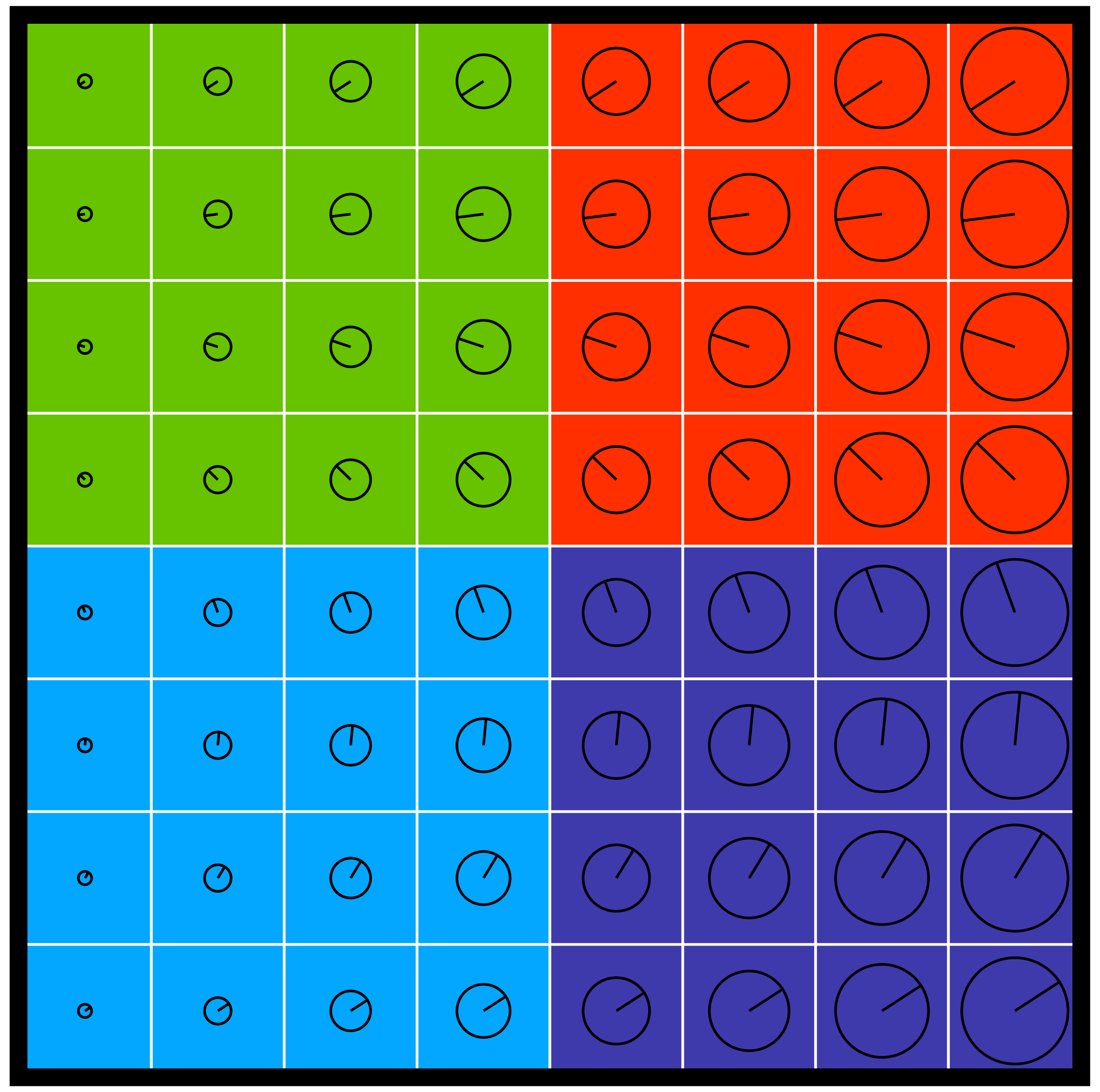
Results

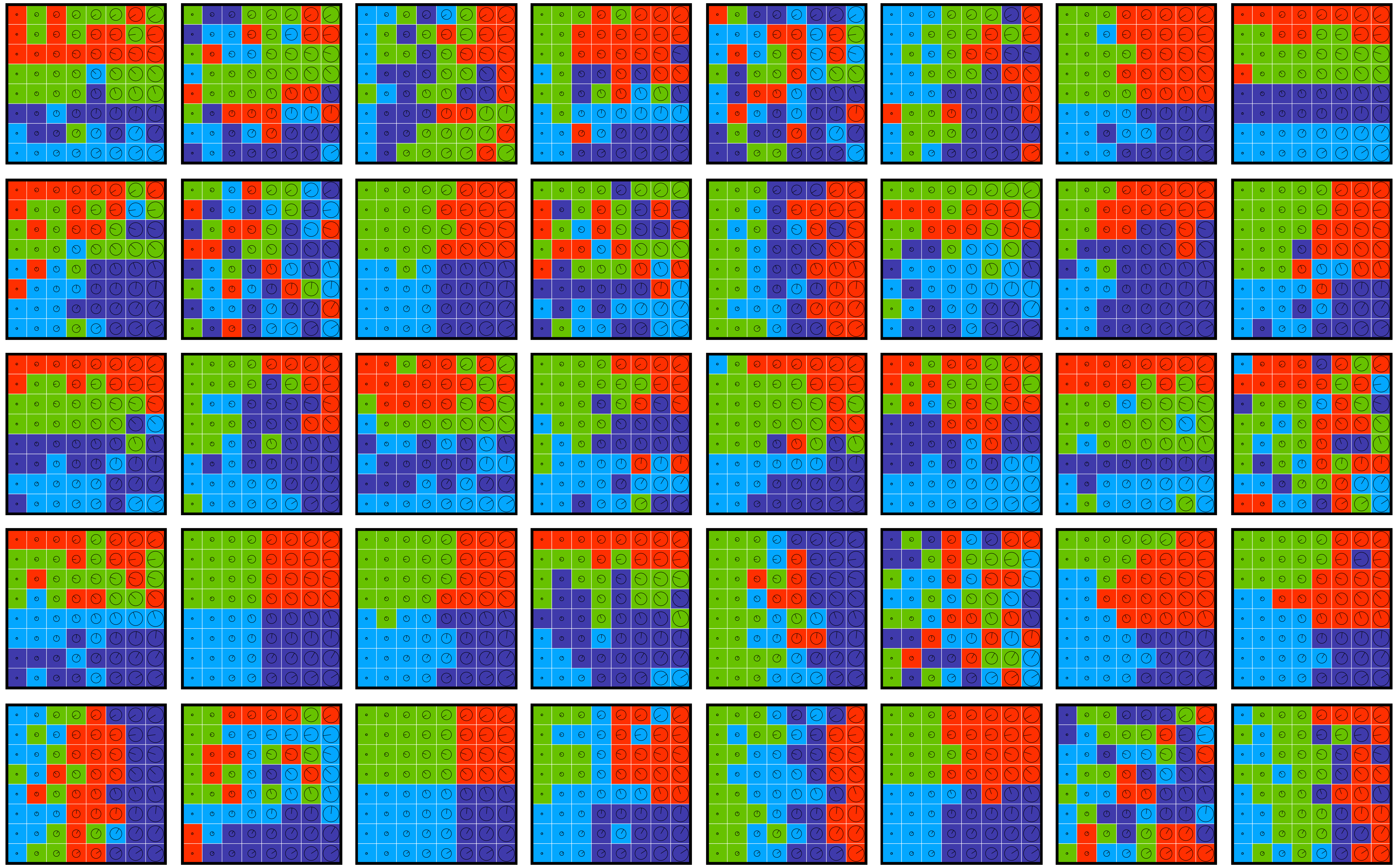




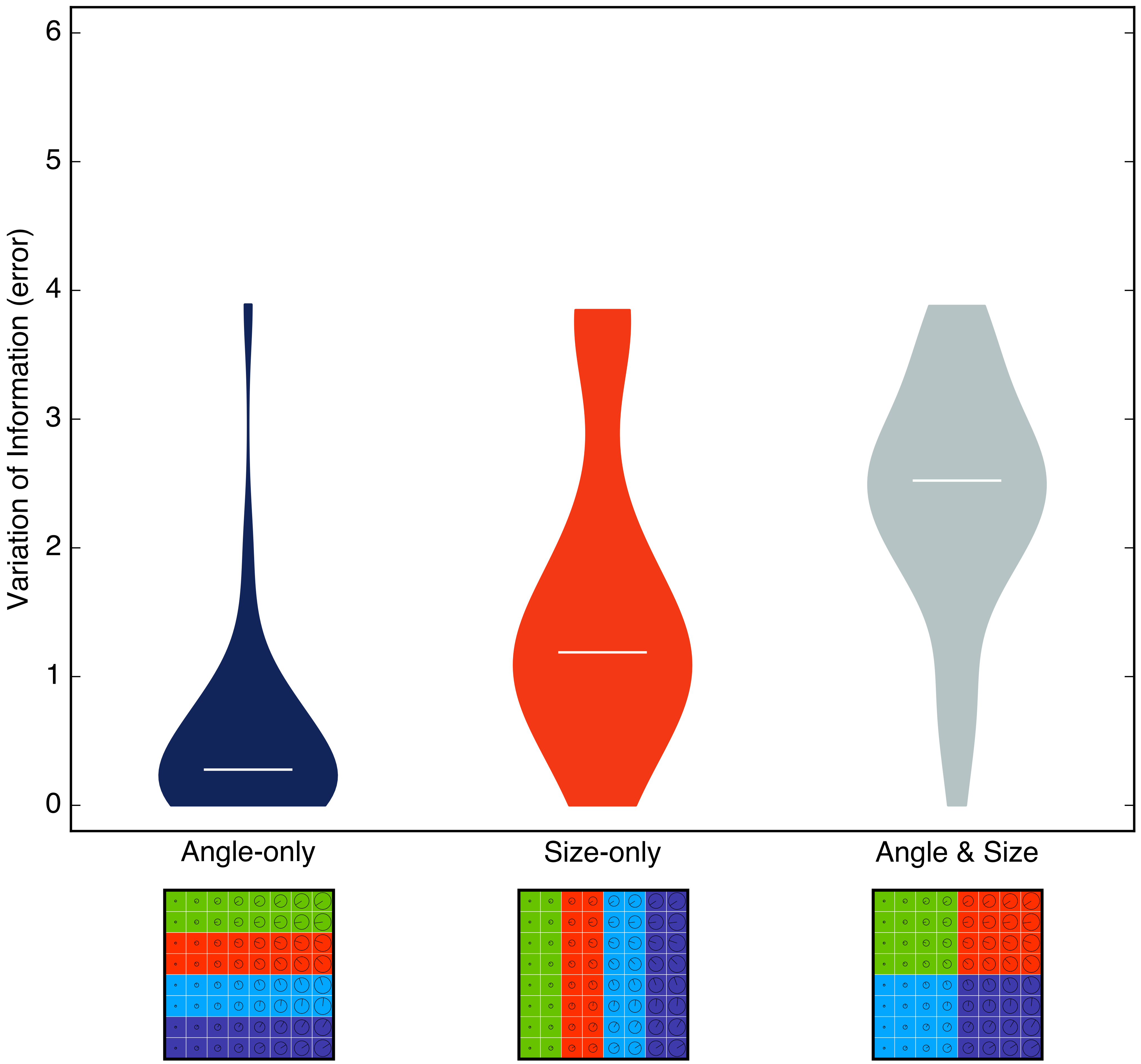
Results

Angle & Size



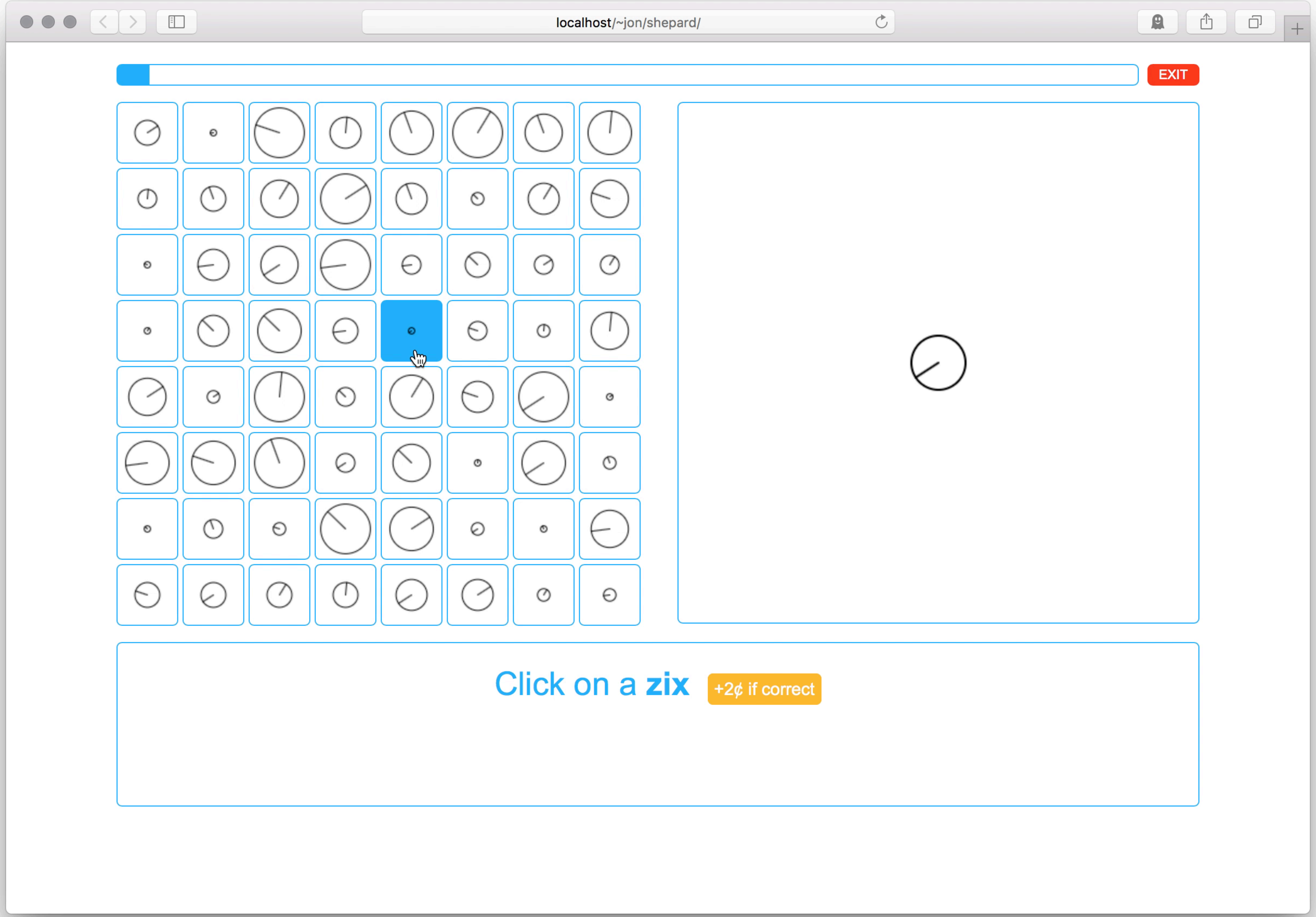


Result: Learnability advantage for the less informative systems

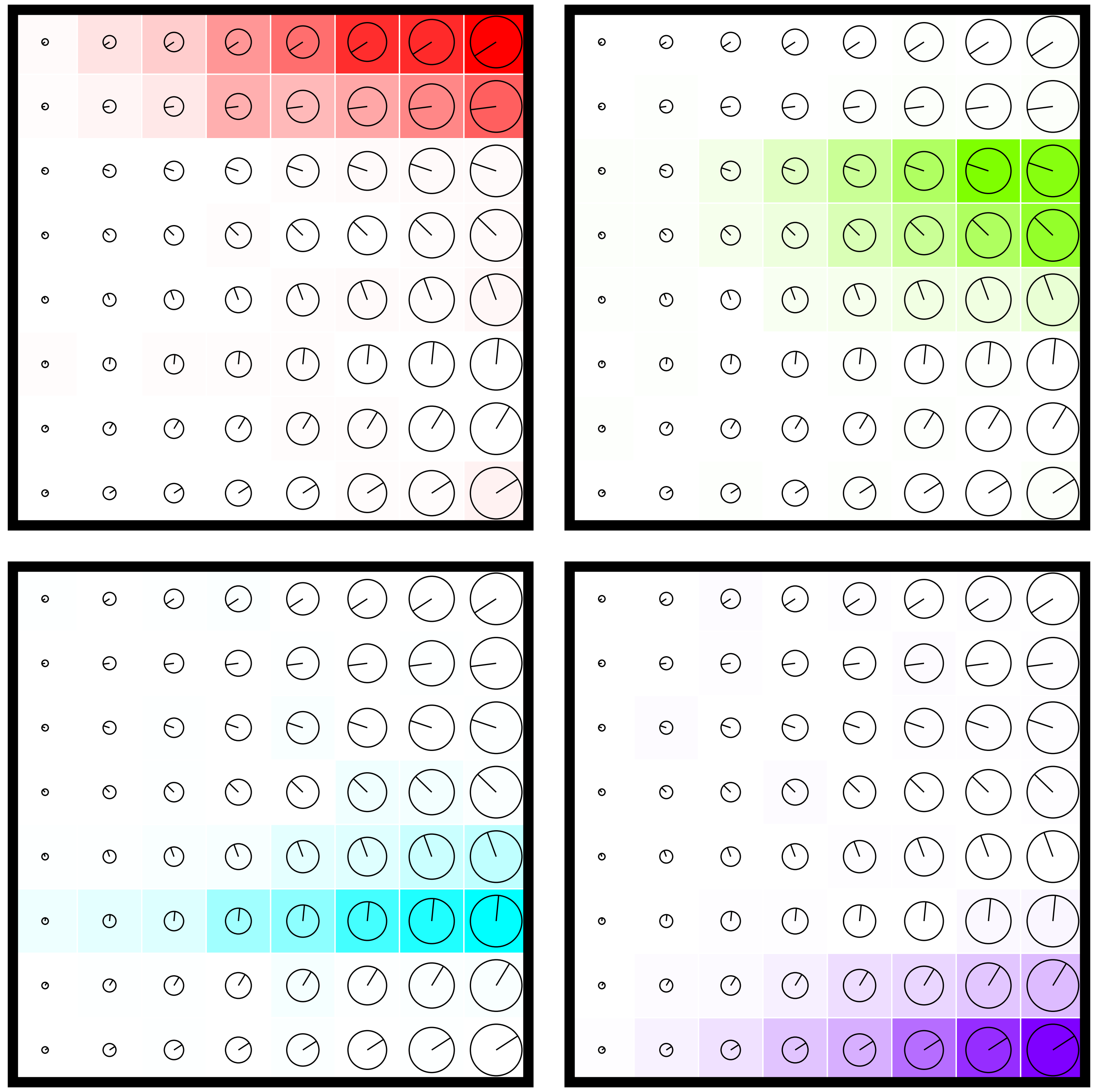


Experiment 2

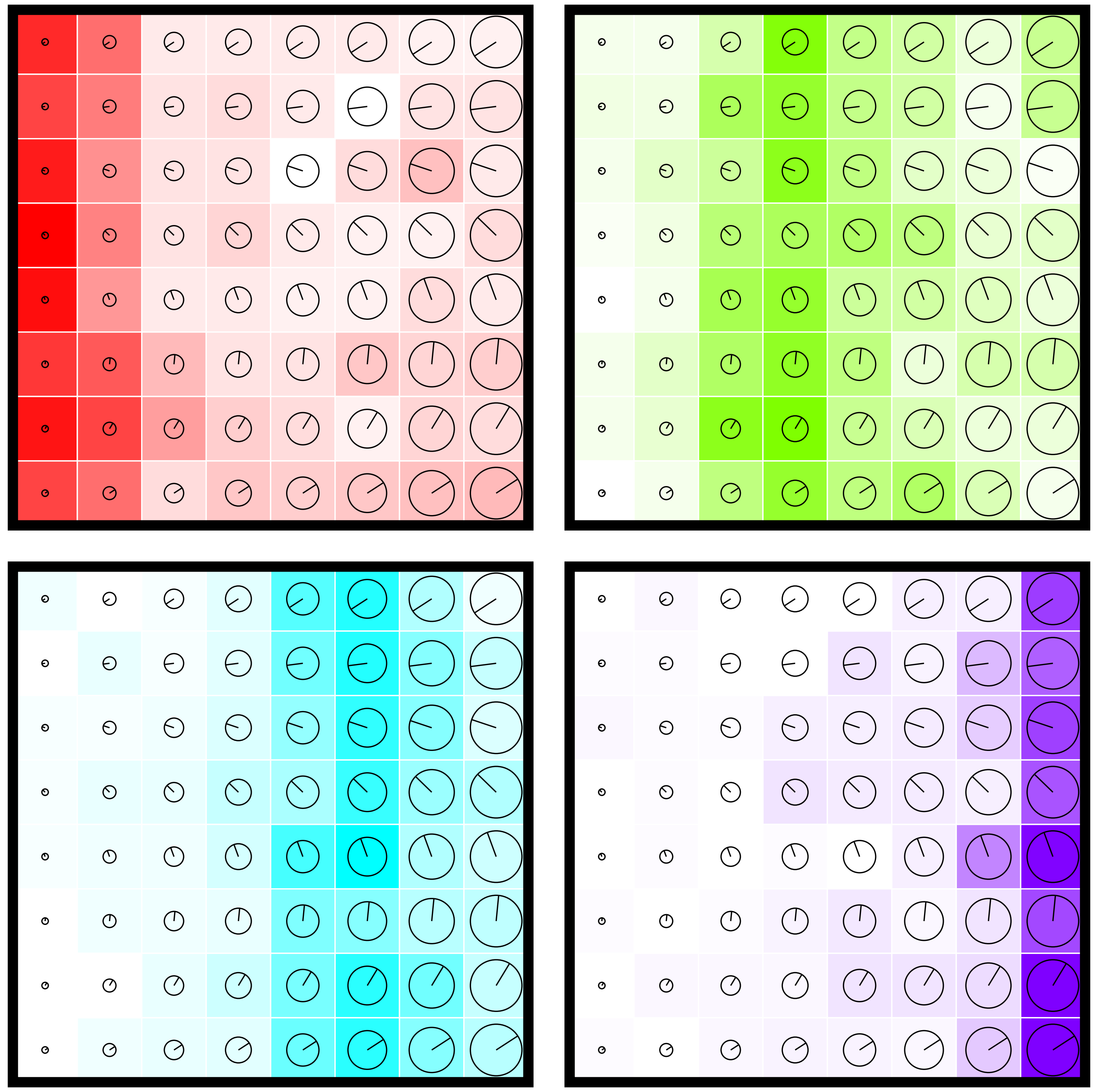
Comprehension test



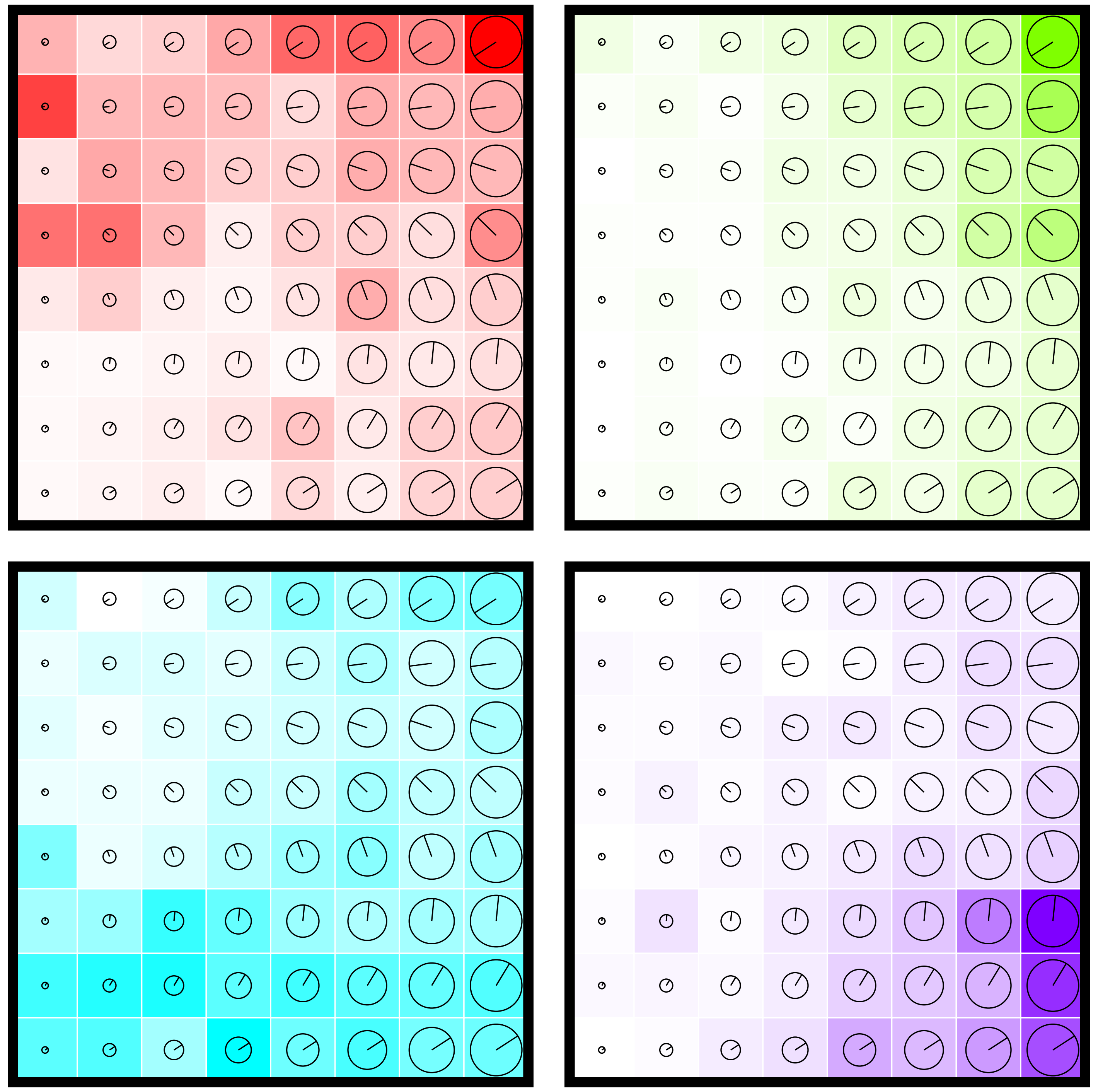
Experiment 2 results



Angle-only



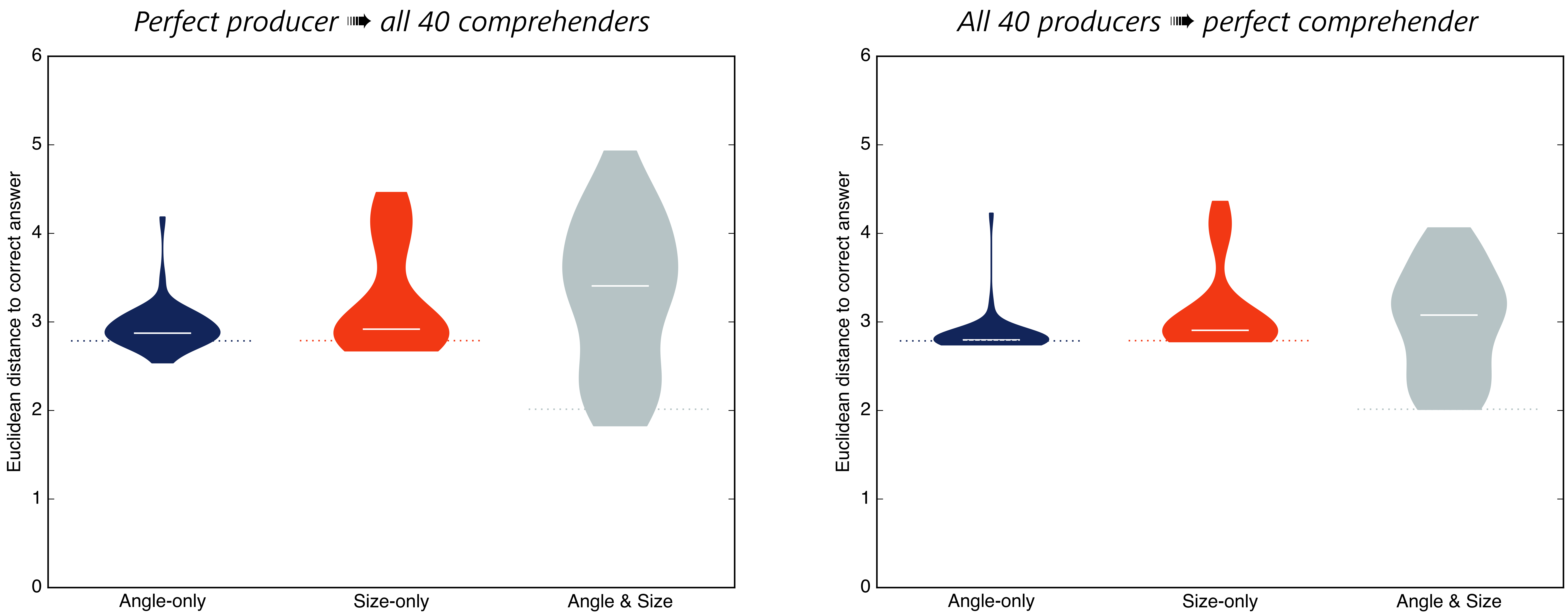
Size-only



Angle & Size

Simulated communication

Simulating communication



Conclusions

Conclusions

Regier's lab has shown that real languages are at the optimal frontier of informativeness and simplicity

Meanwhile, we've been interested in explaining which pressures explain informativeness and simplicity by using artificial languages

Both frameworks share many commonalities and may be amenable to a unifying information-theoretic model

Their first work with iterated learning suggests that communication is not required for informative languages; learning alone may be enough

However, our initial experiments suggest that informativeness is driven by communication

Perhaps the result would be stronger with a genuine communicative task

References

Carstensen, A., Xu, J., Smith, C. T., & Regier, T. (2015). Language evolution in the lab tends toward informative communication. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 303–308). Austin, TX: Cognitive Science Society.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*, 1049–1054.

Khetarpal, N., Neveu, G., Majid, A., Michael, L., & Regier, T. (2013). Spatial terms across languages support near-optimal communication: Evidence from Peruvian Amazonia, and computational analyses. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 764–769). Austin, TX: Cognitive Science Society.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the USA*, *105*, 10681–10686.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Levinson, S., Meira, S., & the Language and Cognition group (2003). ‘Natural concepts’ in the spatial topological domain—adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language*, *79*, 485-516.

Malt, B. C., Sloman, S. A., Gennari, S. P., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, *40*, 230–262.

Regier, T., Carstensen, A., & Kemp, C. (2016). Languages support efficient communication about the environment: Words for snow revisited. *PLOS ONE*, *11*, e0151138–17.

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences of the USA*, *104*, 1436–1441.

Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. In B. MacWhinney & W. O’Grady (Eds.), *The handbook of language emergence* (pp. 237–263). Hoboken, NJ: John Wiley & Sons.

Xu, J., Dowman, M., & Griffiths, T. L. (2013). Cultural transmission results in convergence towards colour term universals. *Proceedings of the Royal Society B: Biological Sciences*, *280*, 1–8.

Xu, Y., Regier, T., & Malt, B. C. (2016). Historical semantic chaining and efficient communication: The case of container names. *Cognitive Science*, *40*, 2081–2094.