Supported by European Commission | Horizon 2020 European Union funding for Research & Innovation

Colexification and communicative need in lexical evolution

Andres Karjus

ERA Chair for Cultural Data Analytics, Tallinn University

Richard Blythe, Simon Kirby, Tianyu Wang, Kenny Smith

University of Edinburgh

Protolang 7 | 08.10.2021

andreskarjus.github.io @AndresKarjus

All living languages keep changing

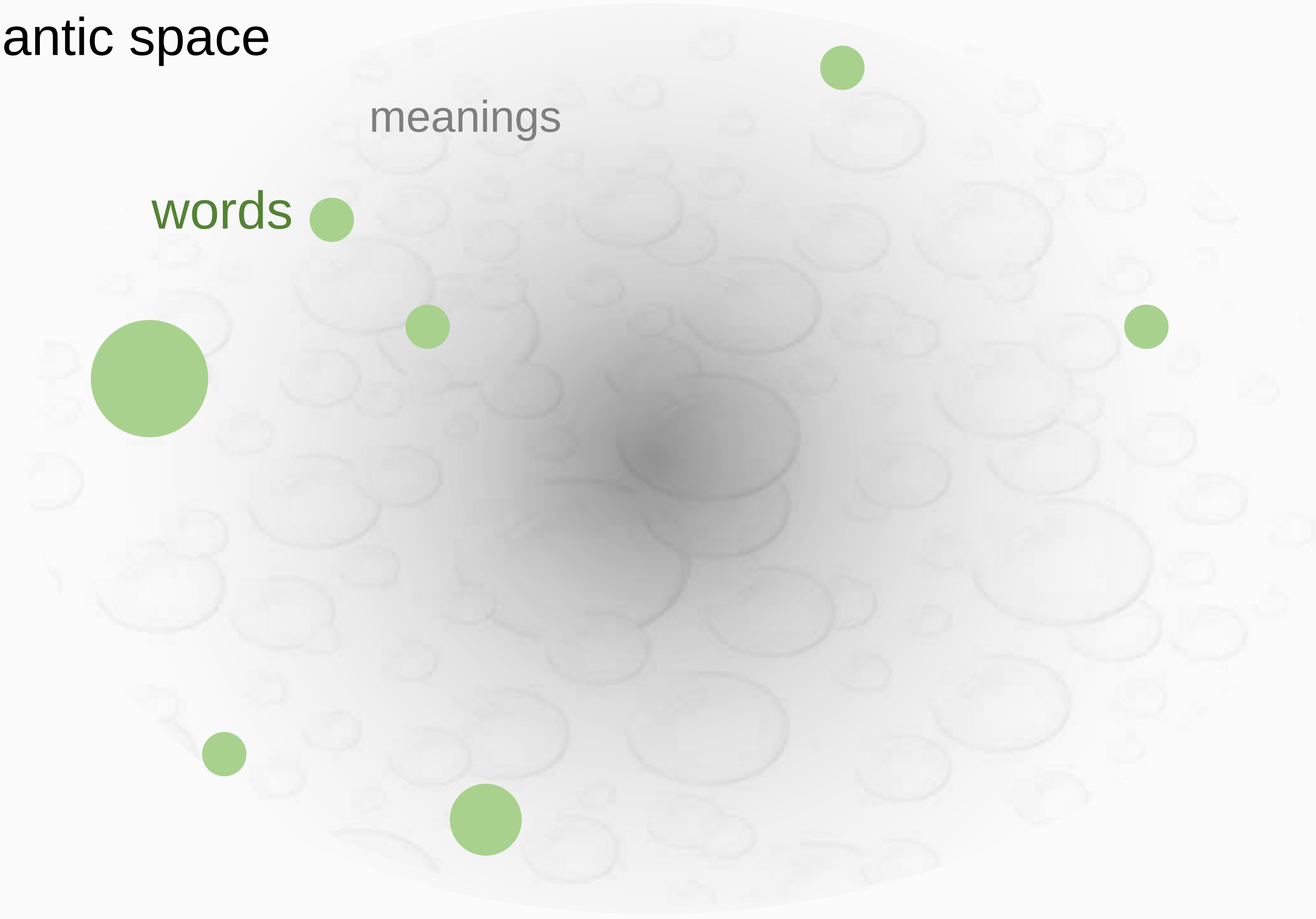
- All the time
- Eventually diverge into different languages
- This is weird

- Here focus on lexical variation (and change)
- → Why are some semantic domains more complex than others?
- → Hypothesis: because communicative need (among other things)

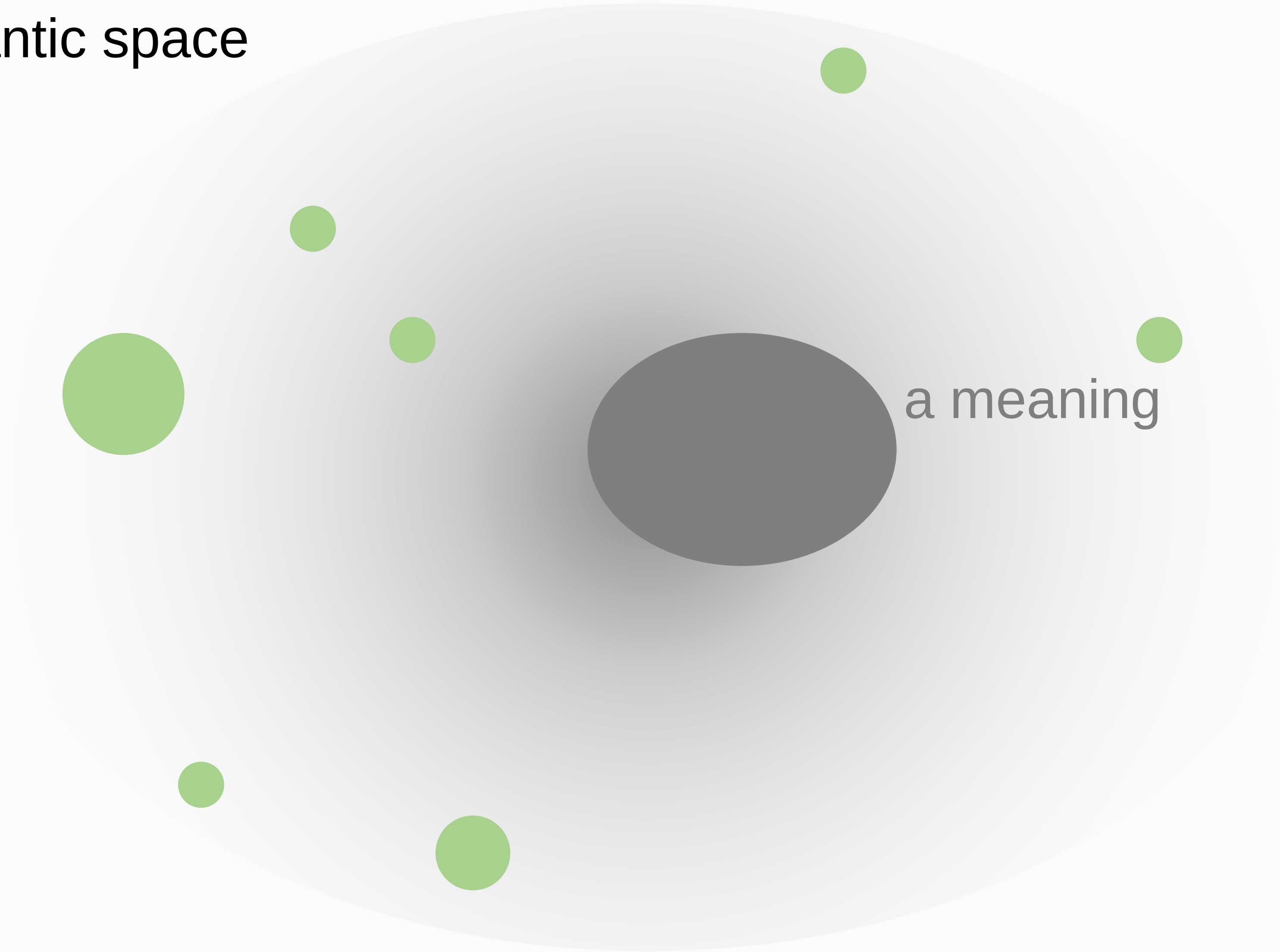
- Massive centuries-spanning corpora → change on the population level
- Variant usage frequencies but also meaning (and change) using distributional semantics methods
- Causal claims about individual-level processes → test using human experiments → artificial languages

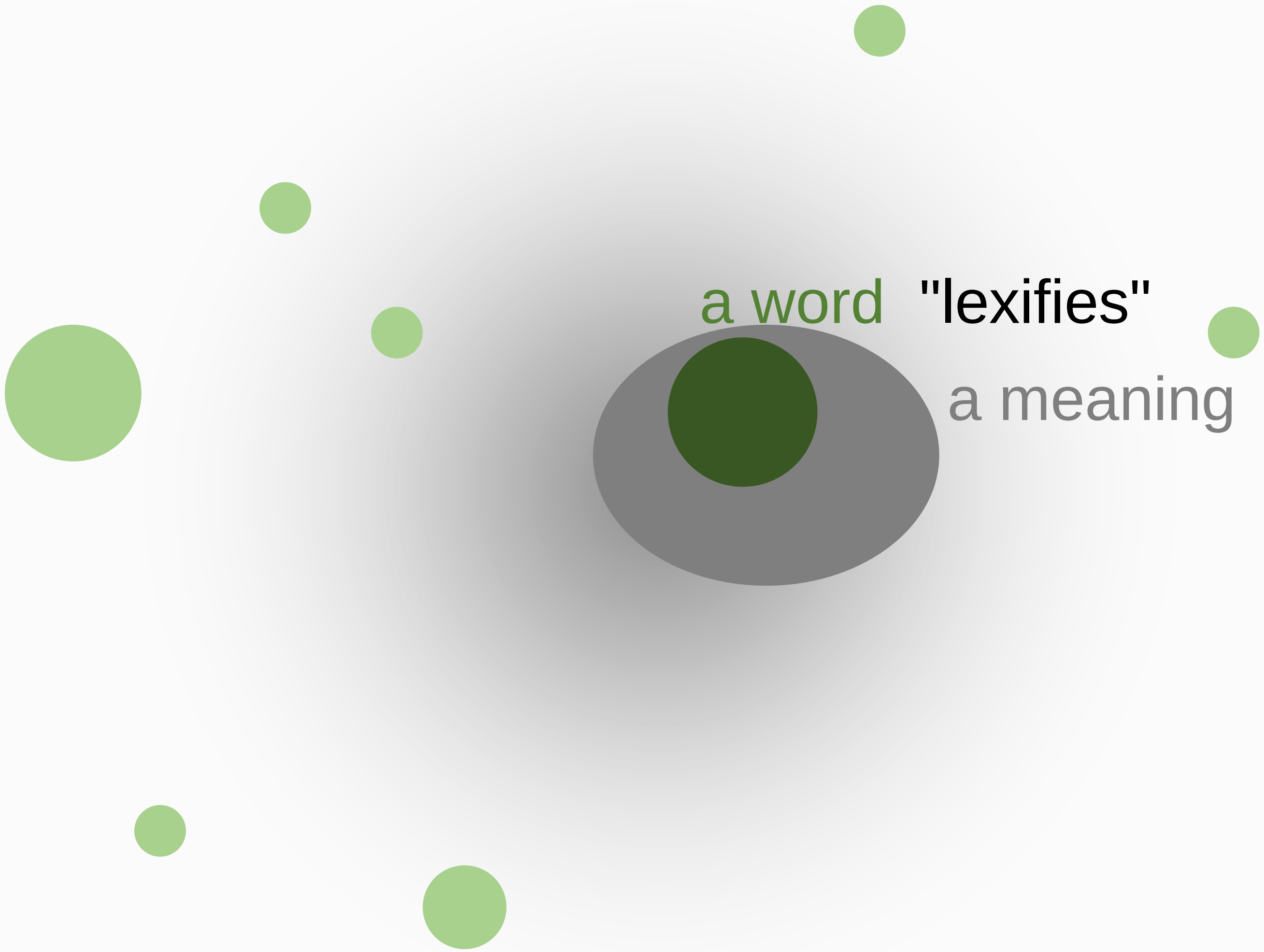
Some concepts

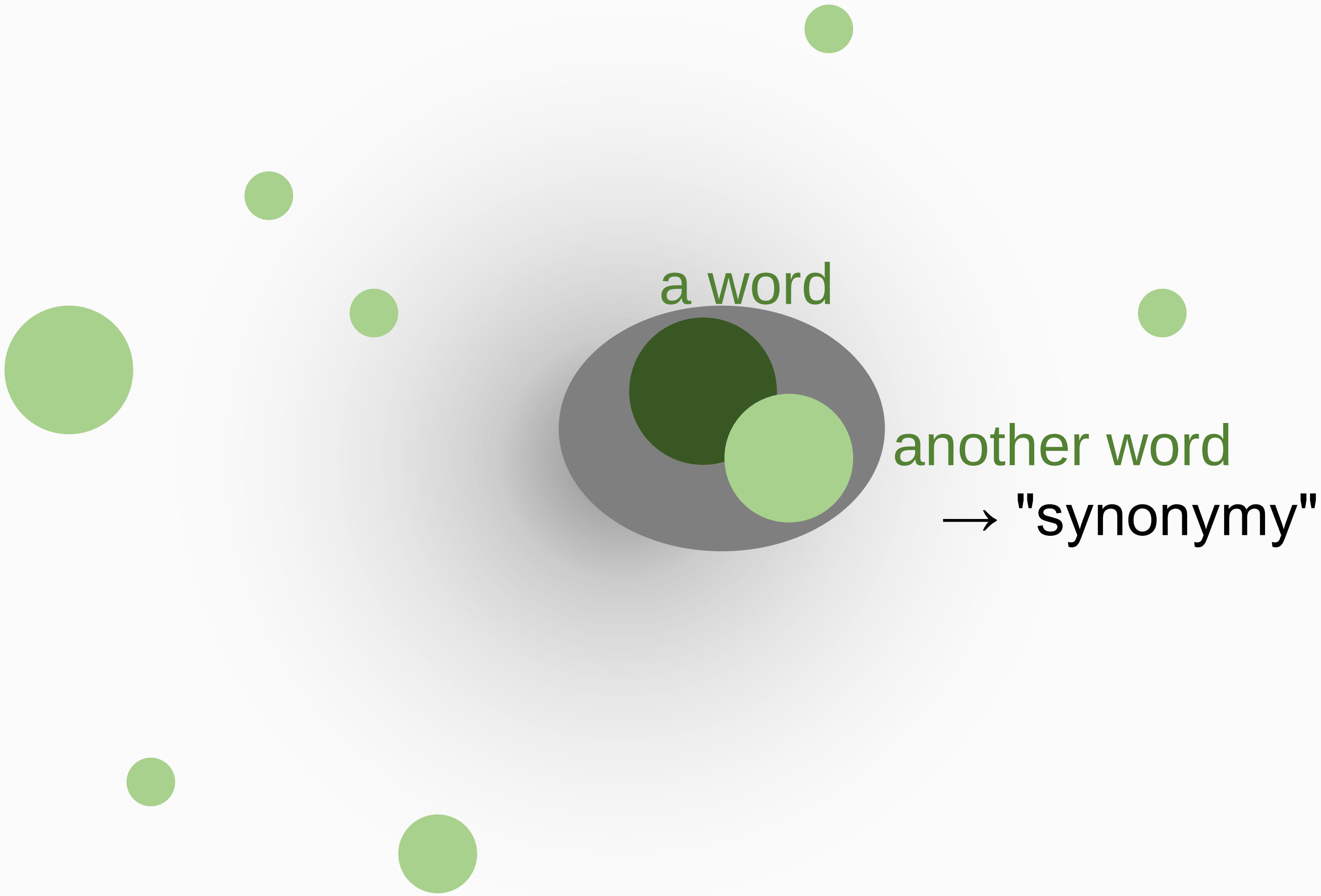
a semantic space

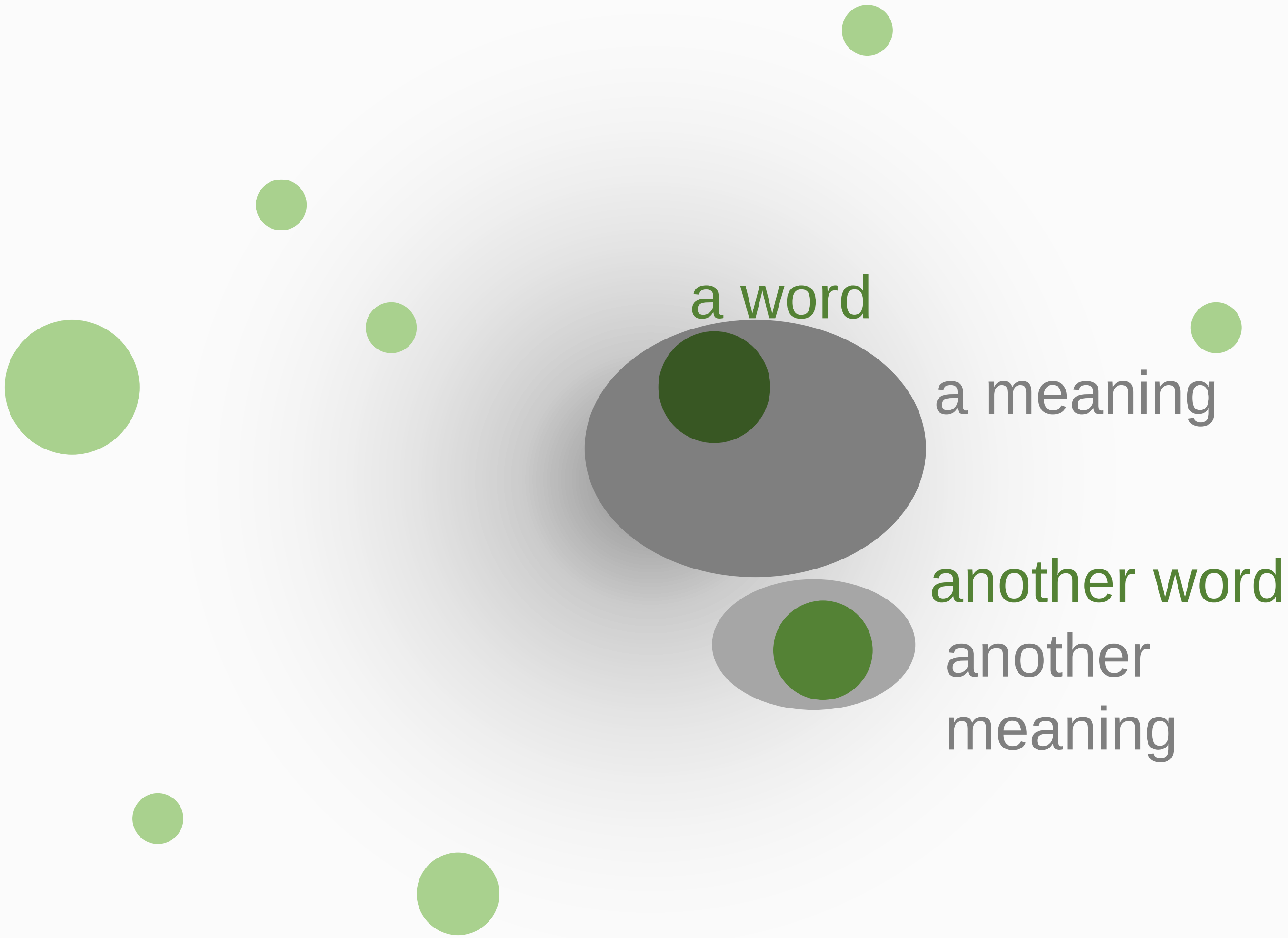


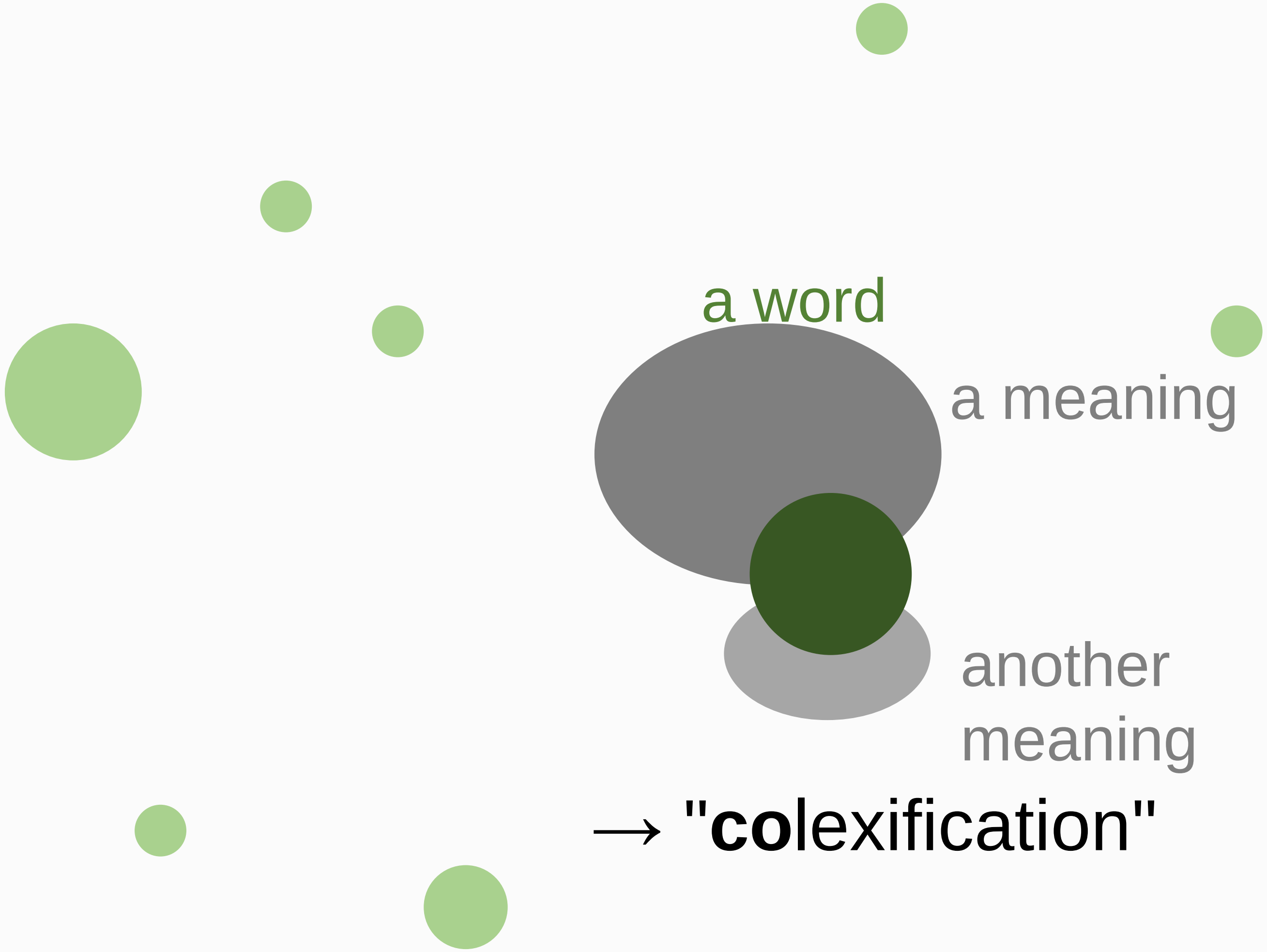
a semantic space




a word
"lexifies"
a meaning


a word
another word
→ "synonymy"


a word
a meaning
another word
another
meaning



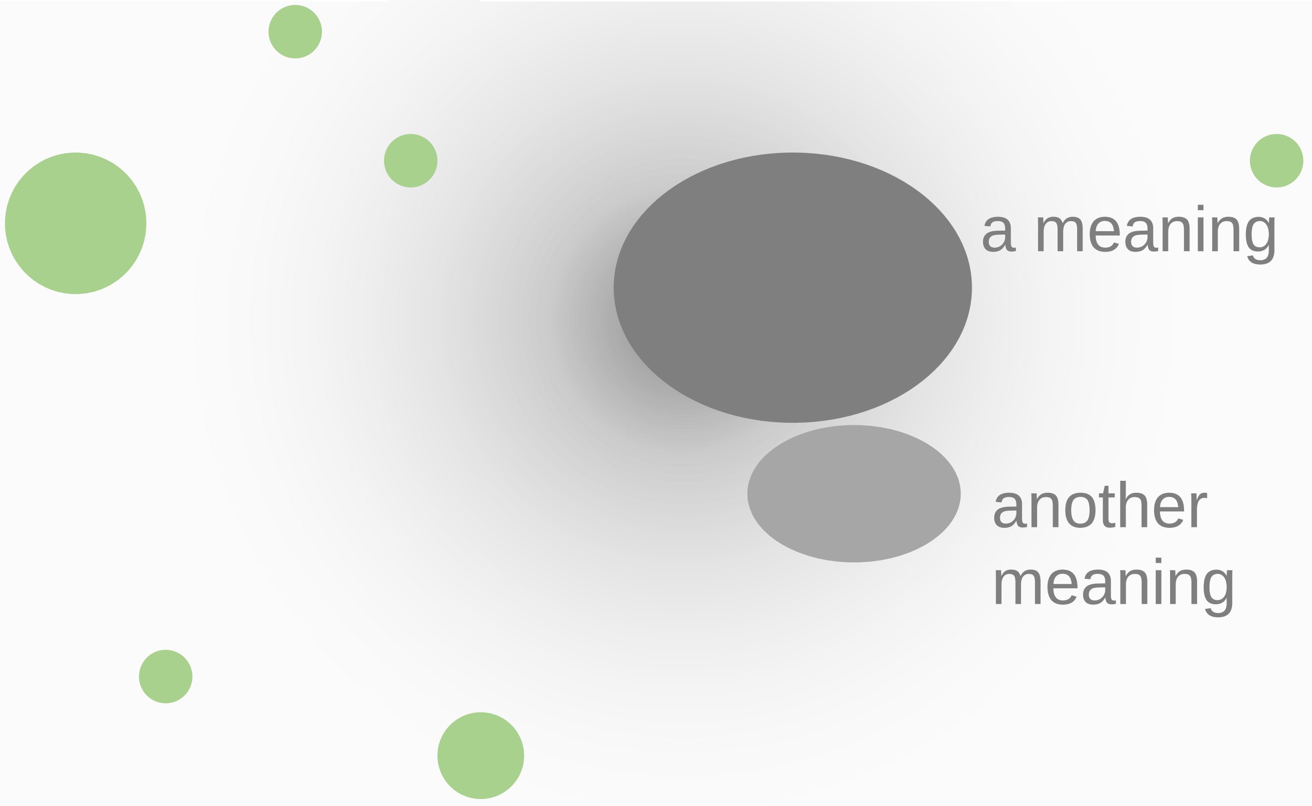
→ "**co**lexification"

cf. François 2008

Complexity and information loss

inverse of simplicity;
relates to learning;
cognitive cost

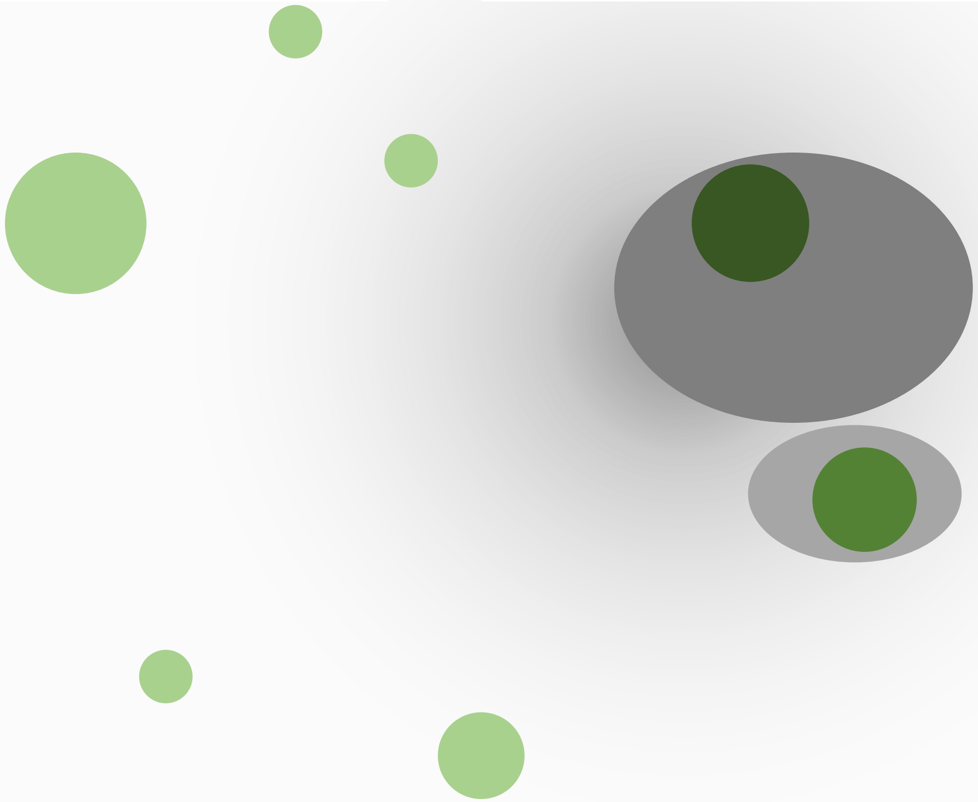
ambiguity;
inverse of expressivity;
communicative cost



Complexity and information loss

inverse of simplicity;
relates to learning;
cognitive cost

ambiguity;
inverse of expressivity;
communicative cost

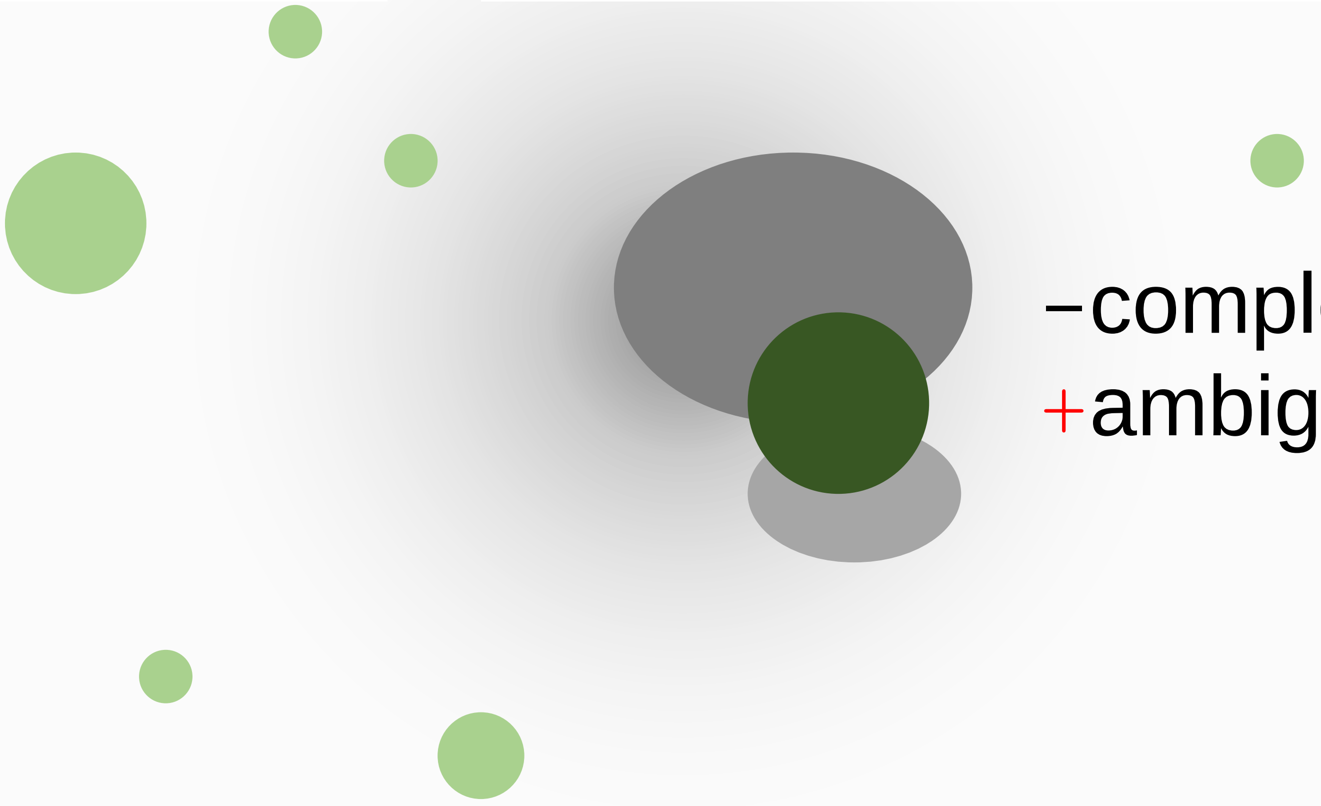


+complexity (cognitive cost)
−ambiguity
(communicative cost)

Complexity and information loss

inverse of simplicity;
relates to learning;
cognitive cost

ambiguity;
inverse of expressivity;
communicative cost

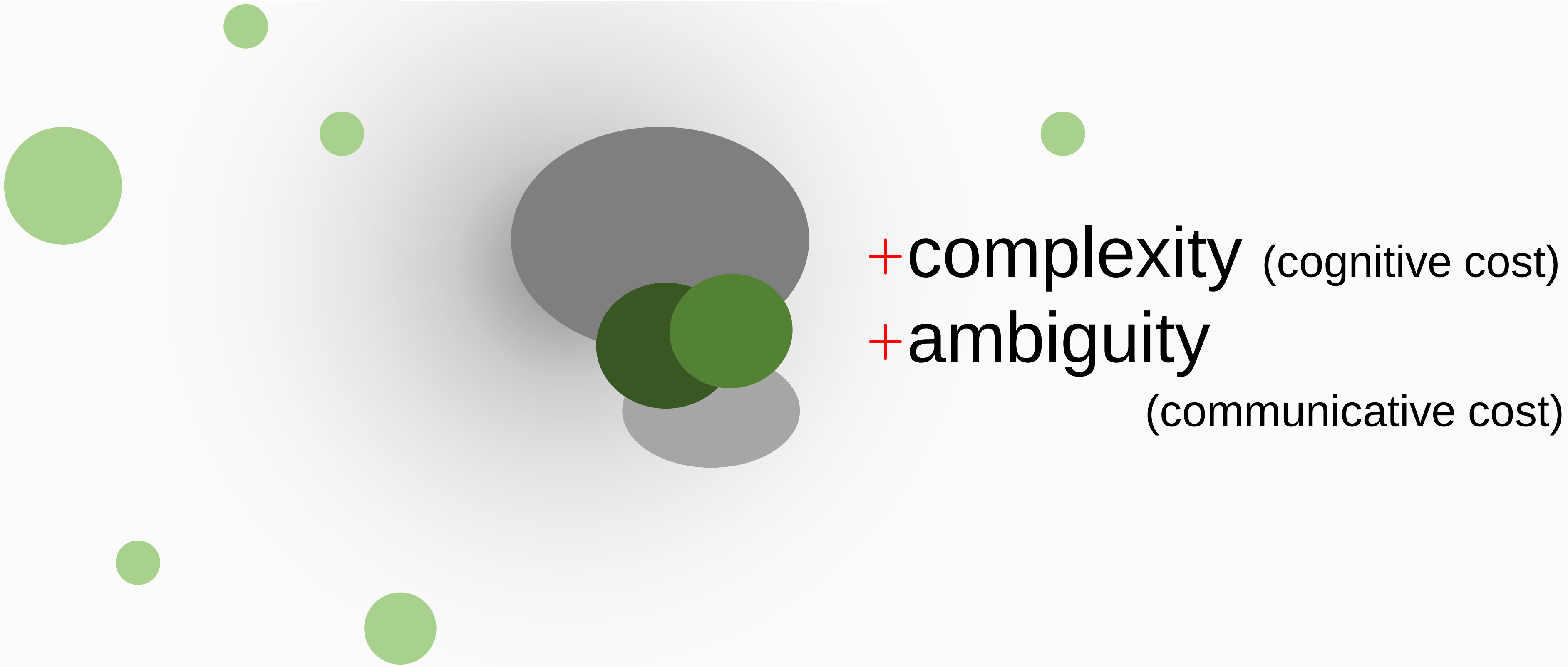


−complexity (cognitive cost)
+ambiguity
(communicative cost)

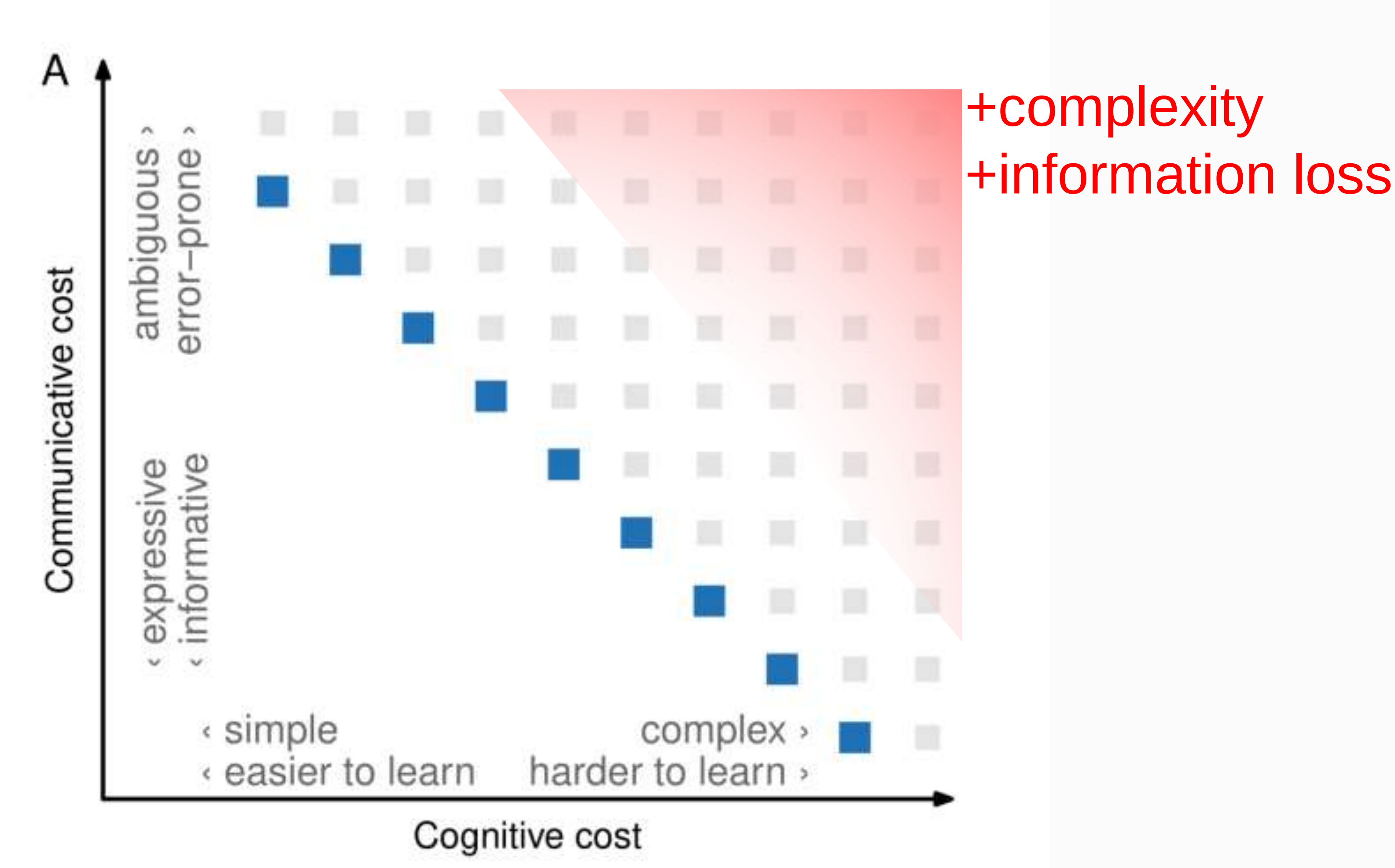
Complexity and information loss

inverse of simplicity;
relates to learning;
cognitive cost

ambiguity;
inverse of expressivity;
communicative cost

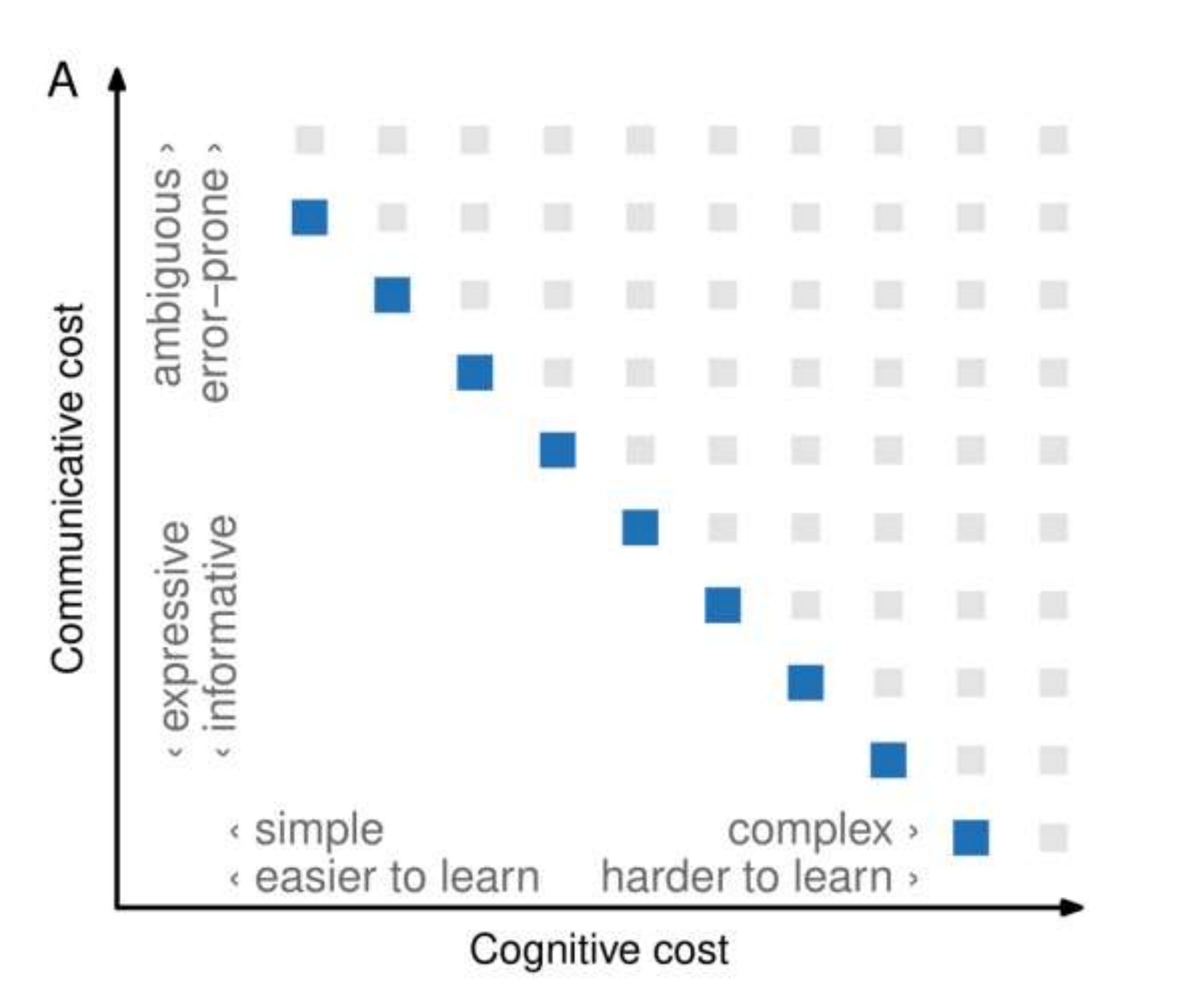


The complexity vs information loss tradeoff and the optimal front



cf. Kemp et al 2012, Kemp et al 2018, Carr et al 2020

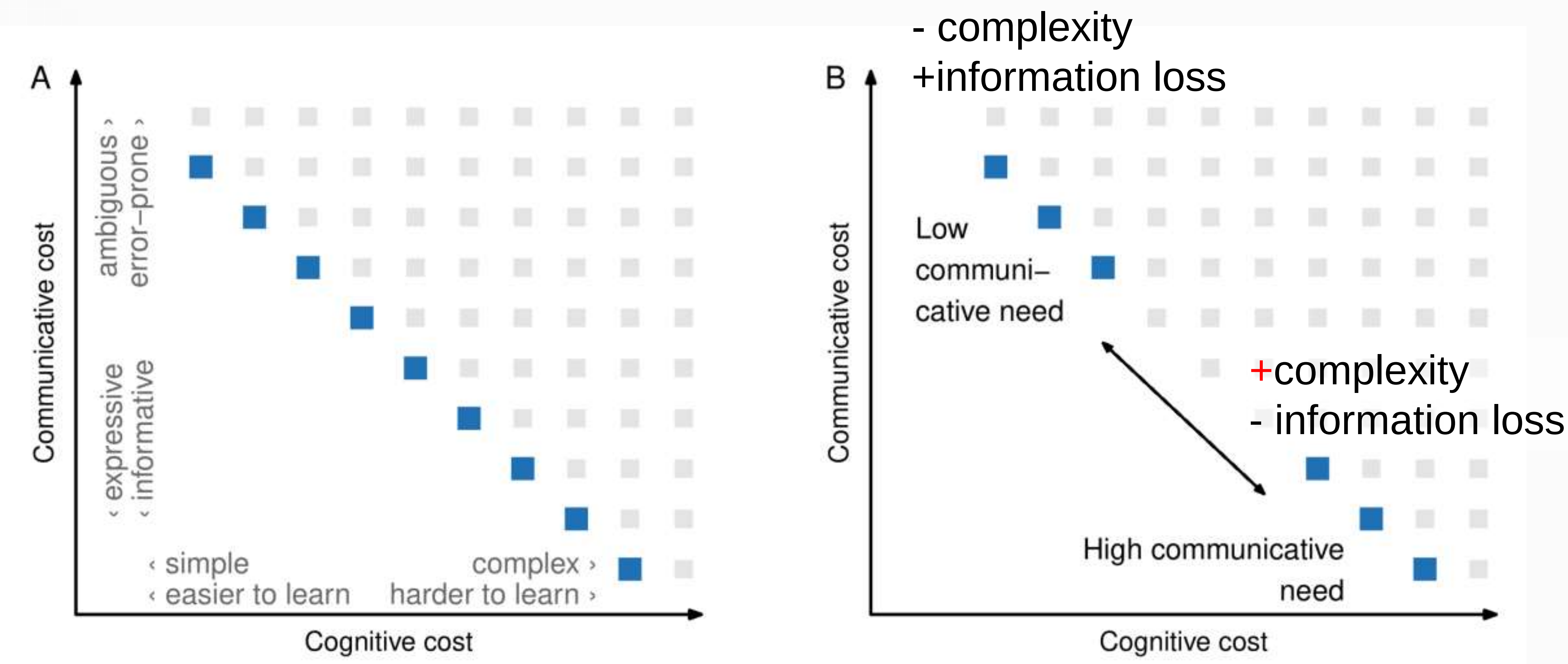
The complexity vs information loss tradeoff and the optimal front



Describes lexicons of kinship terms, colour, numeral systems, negation; similar optimization effects in artificial language experiments

cf. Gibson et al. 2017; Zaslavsky et al. 2019; Kemp and Regier 2012; Zaslavsky et al. 2021; Xu et al. 2020; Regier et al. 2016; Zaslavsky et al. 2020; Kemp et al. 2019; Hermalin and Regier 2019; Miton and Morin 2021; Haspelmath and Karjus 2017; Mollica et al. 2020; cf. Fedzechkina et al. 2012; Winters et al. 2015; Tinits et al. 2017; Nölle et al. 2018; Chaabouni et al. 2021; Guo et al. 2021

The complexity-informativeness tradeoff and communicative need



cf. Kemp et al 2012, Kemp et al 2018, Carr et al 2020

Conceptual similarity and communicative need shape colexification: an experimental study

- Karjus, Blythe, Kirby, Wang, Smith 2021, Cognitive Science
  open preprint → https://arxiv.org/abs/2103.11024

- Xu et al 2020, "Conceptual relations predict colexification across languages", lexical data from 200+ languages, show that
- Similar and associated senses (e.g. FIRE and FLAME) are more frequently **colexified** in world’s languages than unrelated or weakly associated meanings (e.g. FIRE and SALT)

Conceptual similarity and communicative need shape colexification: an experimental study

- Xu et al: similar senses more frequently colexified
- …but culture-specific **communicative needs** *should* affect likelihood of colexification – e.g. if it is necessary for efficient communication to distinguish some similar meanings
- E.g. ICE and SNOW: less likely to be colexified in cold climates (Regier et al 2016)

Conceptual similarity and communicative need shape colexification: an experimental study

- Can we describe the cognitive mechanisms that leads to these cross-linguistic tendencies?
- Maybe we can test these two claims experimentally?
- Dyadic communication game setup, 2 players, take turns sending and guessing messages (cf. Kirby et al 2008, Winters et al 2015)
- 2 conditions, neutral "baseline" and "target"
- 135 rounds each (first 45 rounds excluded as training phase)
- 4 experiments: initial one with student sample; a replication on Mechanical Turk + two more follow-up experiments

The game


Player 1
Players connected: 2. Score: 0/2
area fashion
Communicate *area* using...
piti
wuli
liha
naru
mano
himu
qata

The game


Player 1
Players connected: 2. Score: 0/2
area fashion
Communicate area using...
piti
wuli
liha
naru
mano
himu
qata
Player 2
area fashion
Waiting for message...

The game

Player 1
Players connected: 2. Score: 0/2
area fashion
Communicate area using...
piti
wuli
liha
naru
mano
himu
qata
Player 2
area fashion
Waiting for message...
Player 1
Players connected: 2. Score: 0/2
Sent area using piti
Stand by...

The game

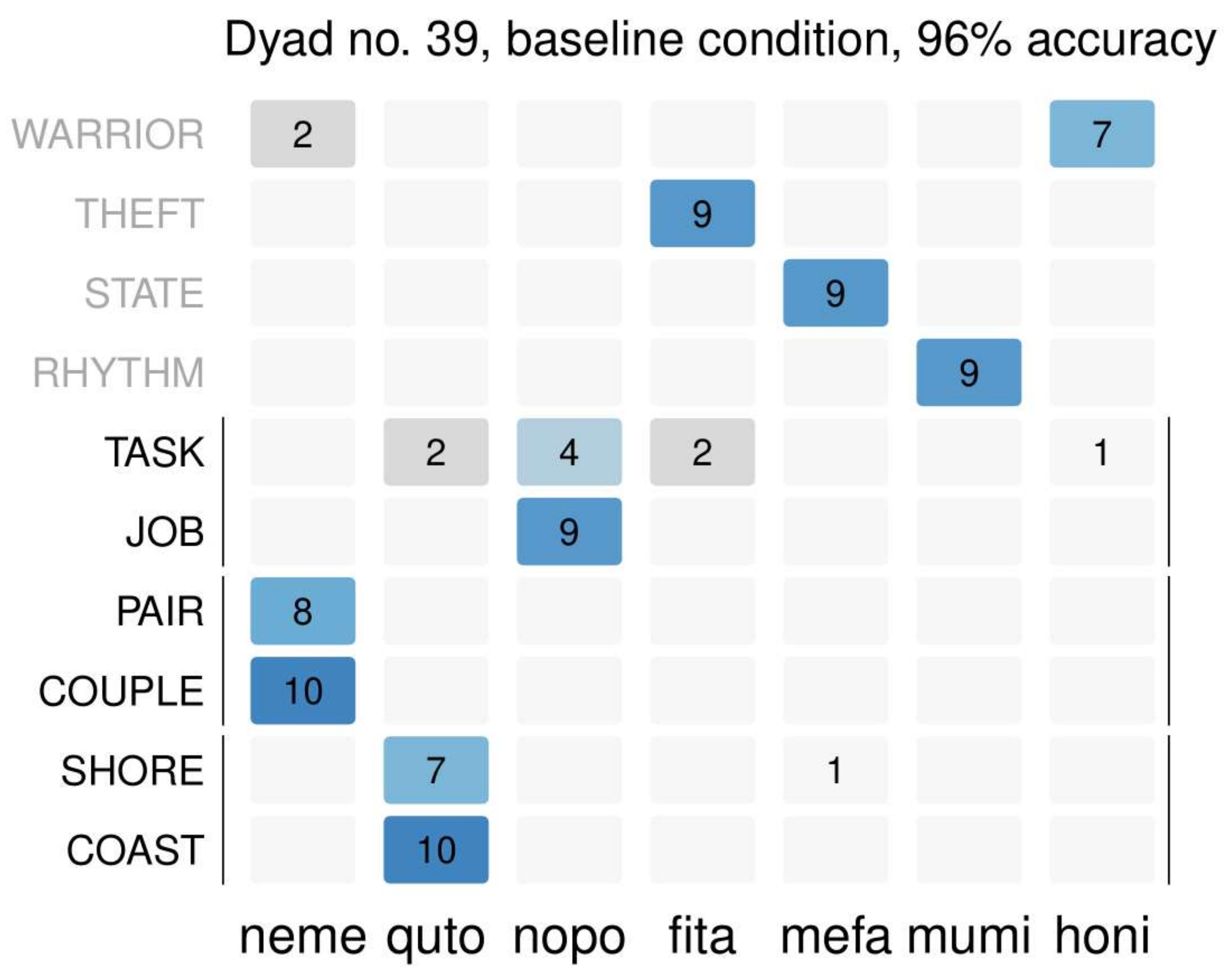

Player 1
Players connected: 2. Score: 0/2
area fashion
Communicate *area* using...
piti
wuli
liha
naru
mano
himu
qata
Player 2
area fashion
Waiting for message...
Player 1
Players connected: 2. Score: 0/2
Sent *area* using piti
Stand by...
Player 2
Players connected: 2. Score: 0/2
area fashion
Message: piti
This means:
area
fashion

- 10 meanings per game
- from Simlex999

- 4 distractor meanings
- 6 target meanings
- → 3 pairs

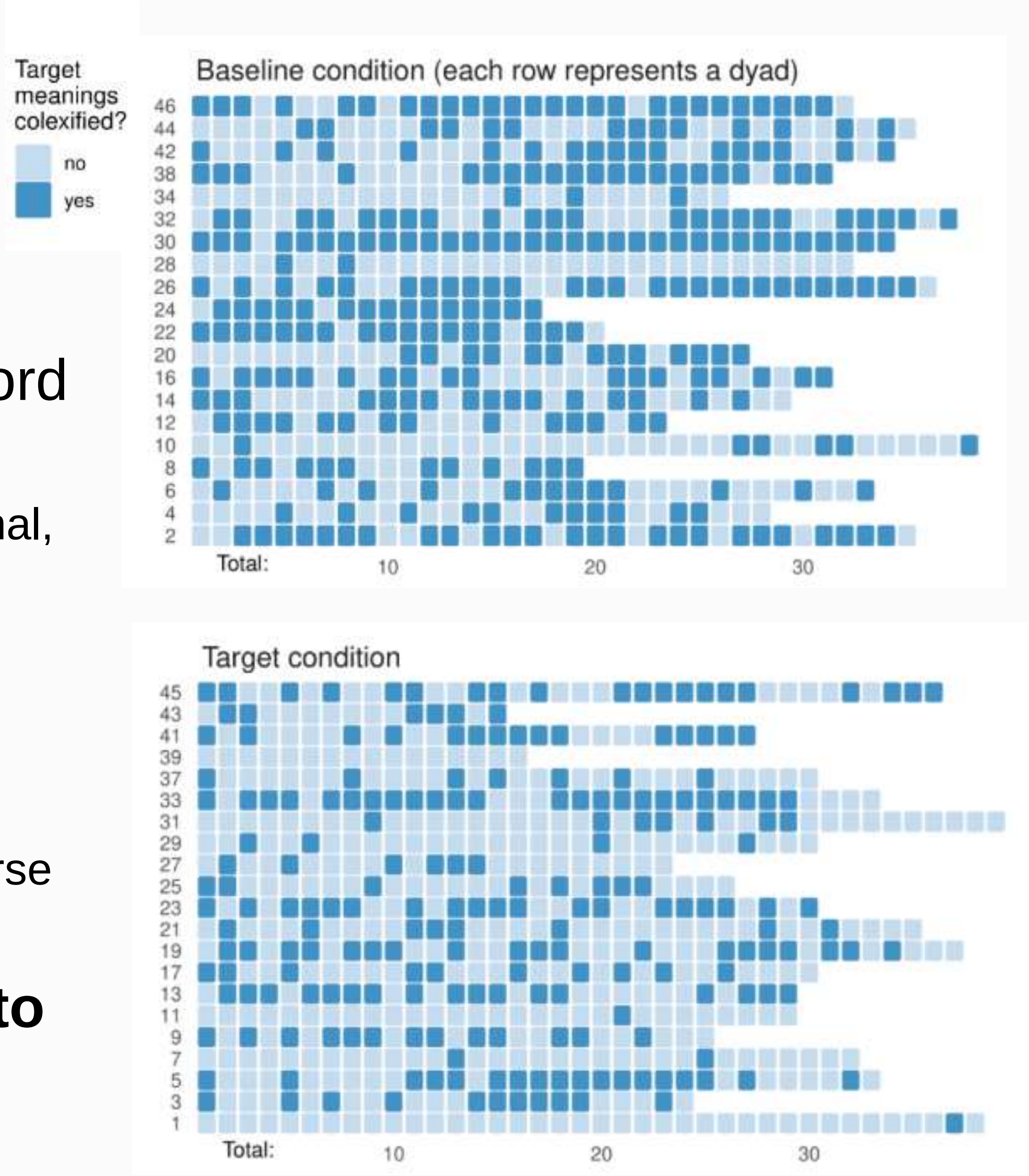
- Baseline: all meaning pairs occur uniformly
- Target condition: **similar meanings occur together more often!** **(~comm. need)**

7 signals per game (controlled for formal similarity!)


Dyad no. 39, baseline condition, 96% accuracy
WARRIOR
THEFT
STATE
RHYTHM
TASK
JOB
PAIR
COUPLE
SHORE
COAST
2
7
9
9
9
2
4
2
1
9
8
10
7
1
10
neme
quto
nopo
fita
mefa
mumi
honi

Analysis

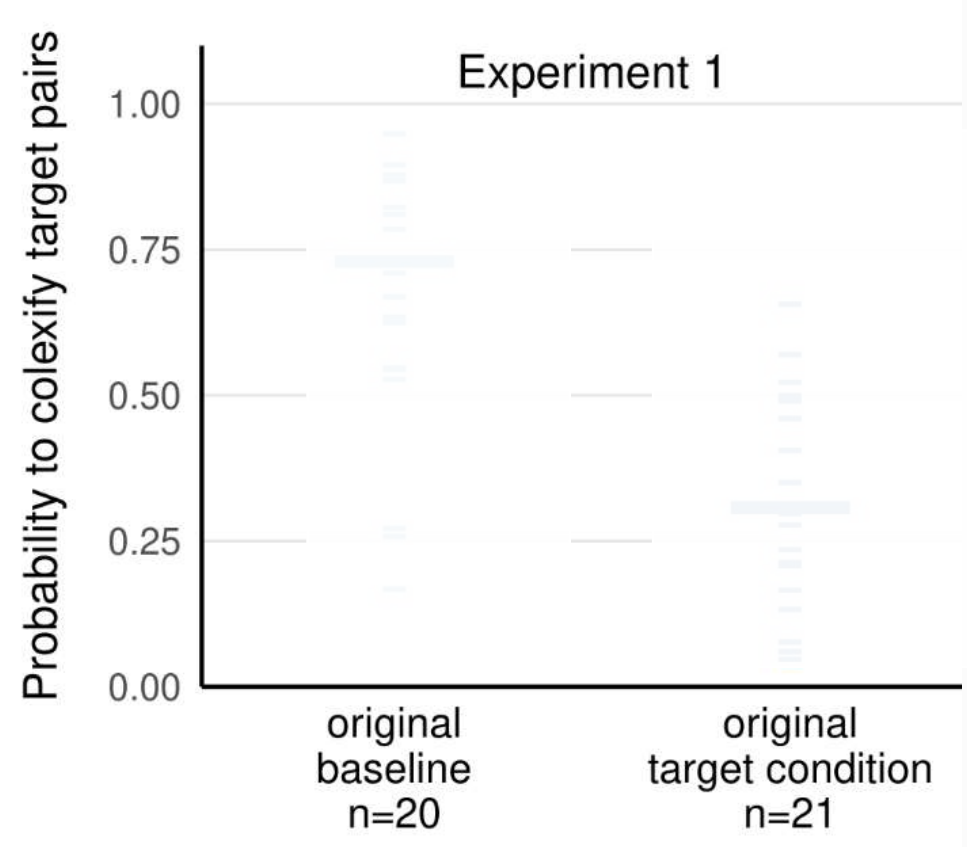
- Exclude low-accuracy dyads (41 left Experiment 1)
- Iterate through each game log, record each instance of colexification involving a target meaning (same signal, different meaning), n=1218.
- Logistic mixed effects regression (random effects for speaker/dyad, meaning pair)
- Colexification ~ condition*round (dyads may change preferences over the course of the game)
- **Are similar meanings less likely to be colexified in the target condition?**



Results

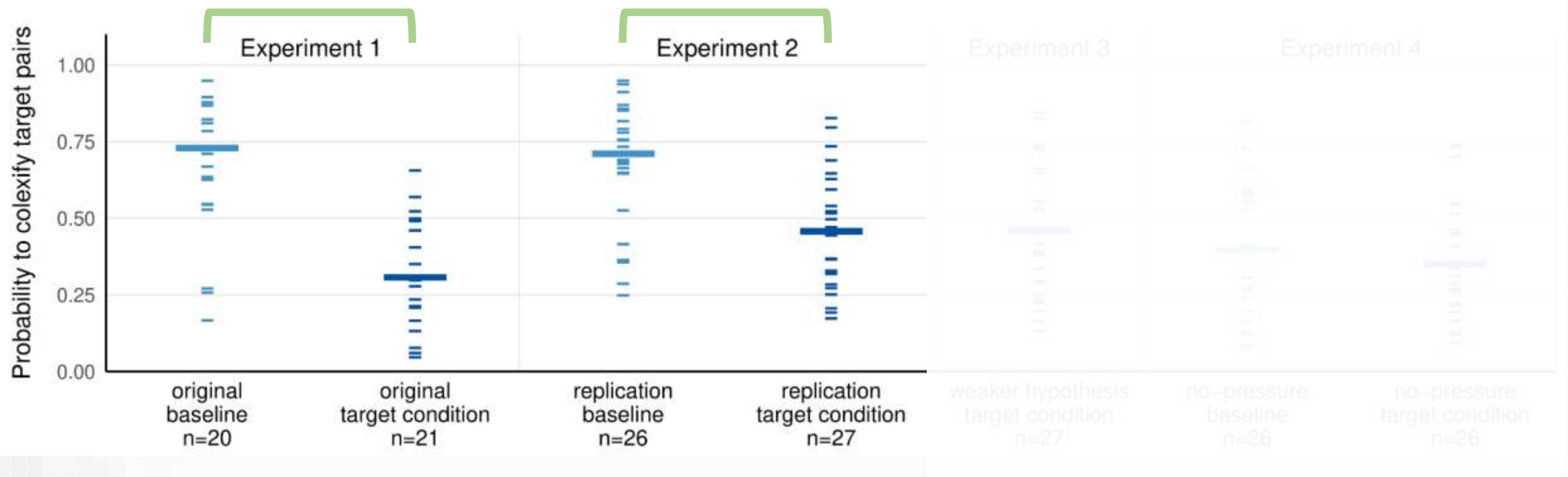
- **Yes** (p=0.001)
- When no pressure to distinguish particular meanings, do prefer to colexify similar meanings (e.g. single signal for TASK-JOB; supports main finding of Xu et al 2020)
- When need arises to distinguish similar meanings (target condition), speakers less likely to colexify them (e.g. separate signals for TASK, JOB)
- Results support hypothesis that communicative needs may block colexification of related concepts.

colexification ~	Estimate	SE	z	p
intercept (baseline condition, mid-game)	-0.22	0.37	-0.59	0.56
+ condition (target)	-0.52	0.51	-1.03	0.3
+ round	1.02	0.27	3.84	<0.01
+ condition (target) × round	-1.17	0.37	-3.17	<0.01



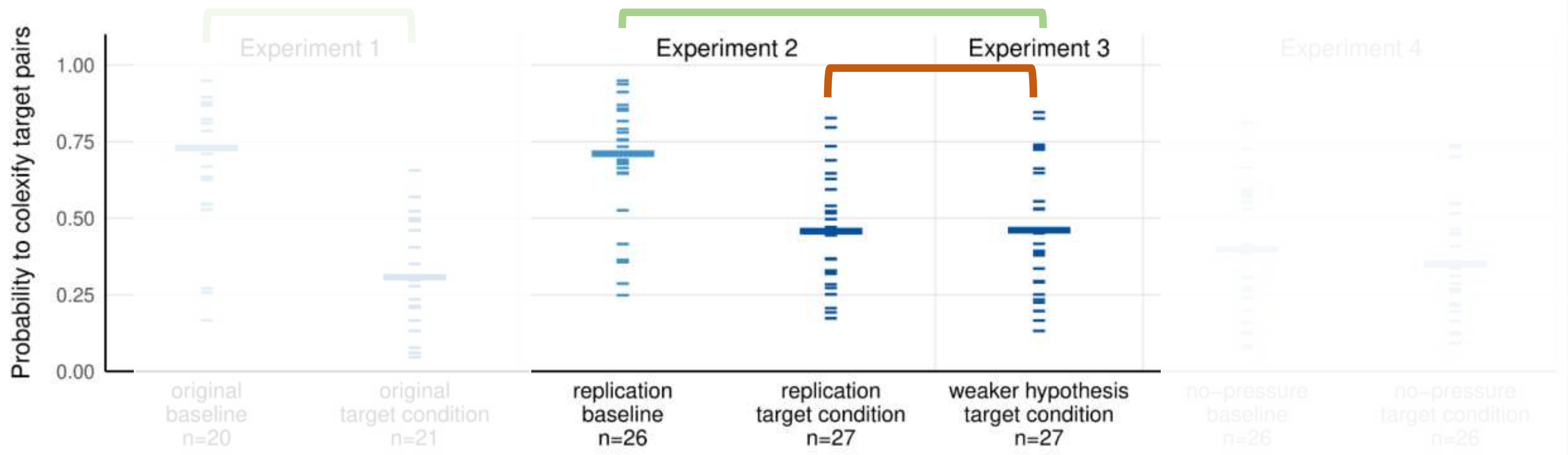
Follow-up experiments

- Experiment 2: same setup but switch to Mechanical Turk (more flexible recruitment)
- Lower accuracy: 79 dyads, could use data only from 53; but replicates Experiment 1.



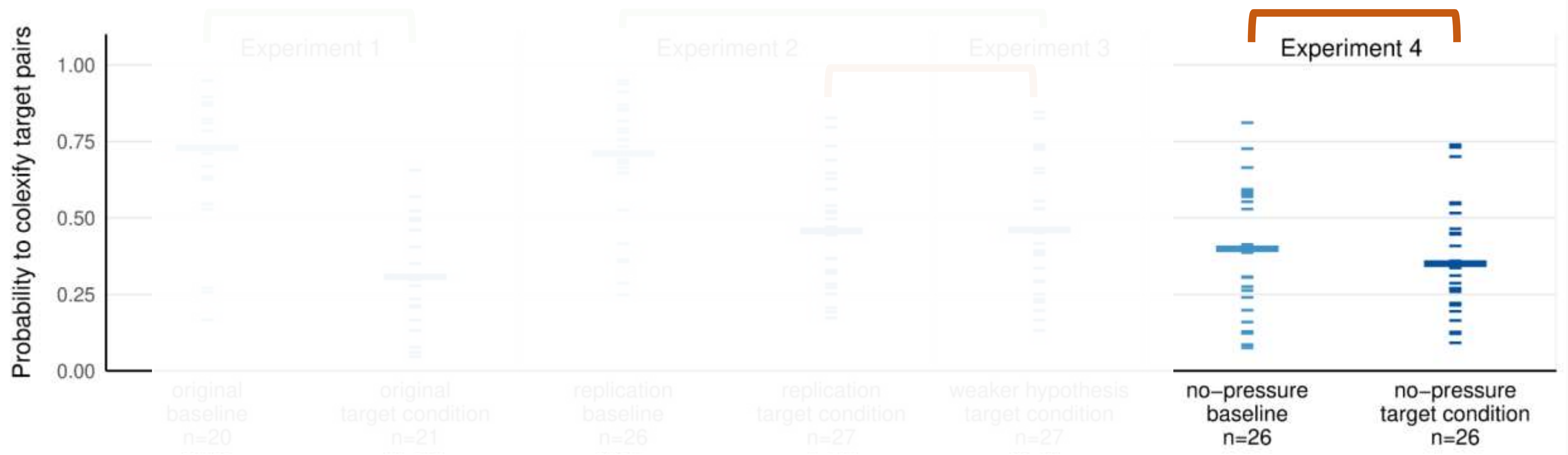
Follow-up experiments

- Experiment 3, target condition only: introduce similar-meaning pairs into the distractor set to make colexifying them more natural (e.g. TASK-JOB instead of BULL-DENTIST)
- No effect (but replicates Experiment 1 & 2)



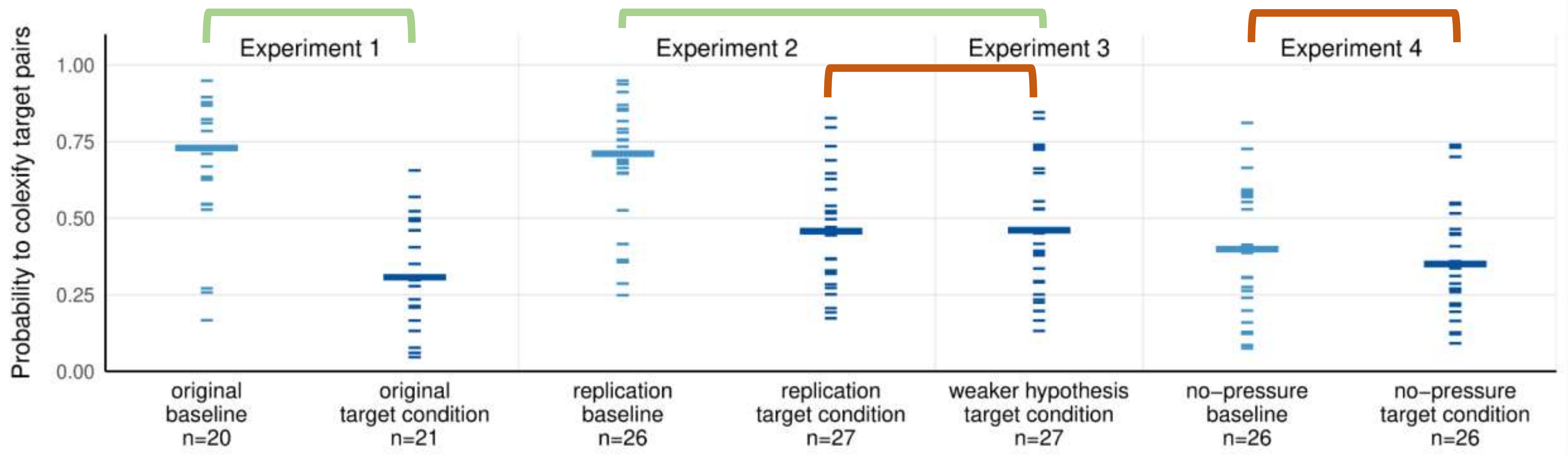
Follow-up experiments

- Experiment 4: remove pressure to colexify (10 signals, 10 meanings)
- No difference between conditions (less colexification in baseline), and participants happily make use of the bigger signal space.
- But: natural language *does* have pressure to simplify (can't have infinite lexicons).

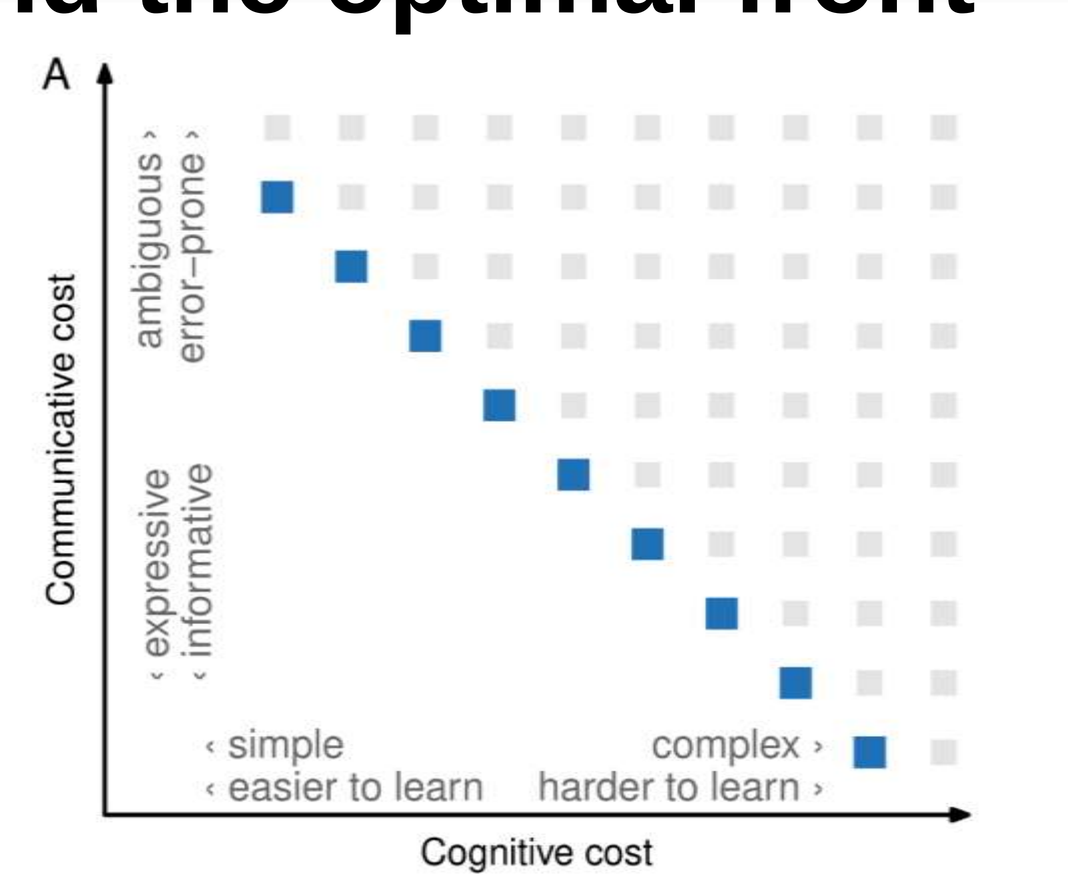


All the experiments

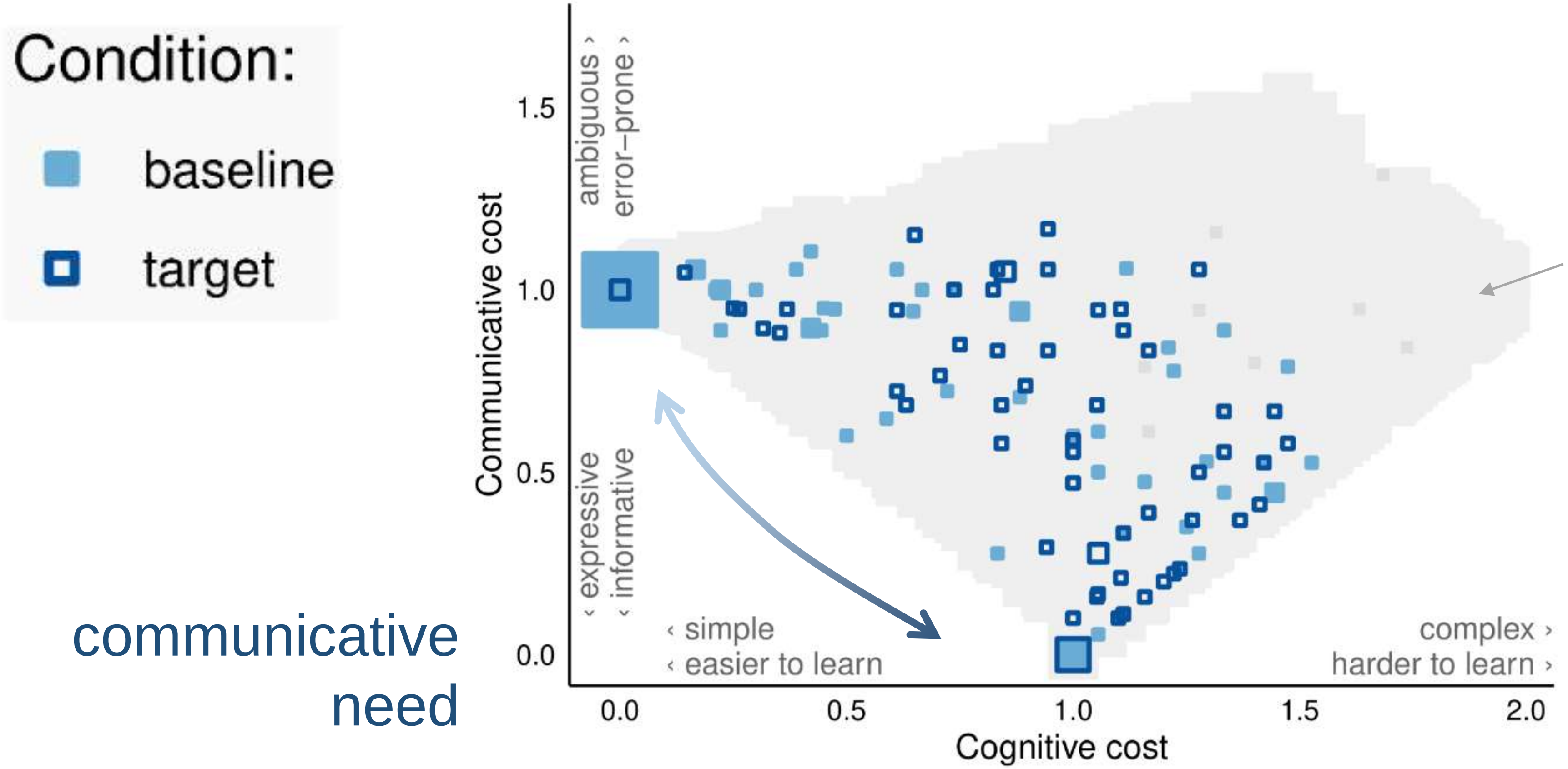
- In total data from 173 dyads (346 participants)



The complexity vs information loss tradeoff and the optimal front



The complexity vs information loss tradeoff and the optimal front (here, optimal points)



+simulated data

communicative need

Figure 9: Average communicative cost and cognitive cost (complexity) scores in Experiment 1. Light blue squares are target meaning pairs (such as DRIZZLE-RAIN) used by baseline condition dyads, dark blue ones are pairs as used by target condition dyads. Larger square size indicates multiple overlapping squares at those coordinates. Simulated results are depicted as gray blocks in

Conclusions

- Experimental results describe an individual-level lexical choice **mechanism** which produces results in line with typological **colexification** tendencies (Xu et al 2020) and supports the **communicative need** hypothesis.

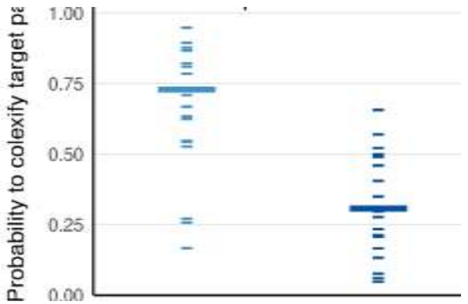
→ *Karjus et al 2021, Conceptual similarity and communicative need shape colexification: an experimental study*

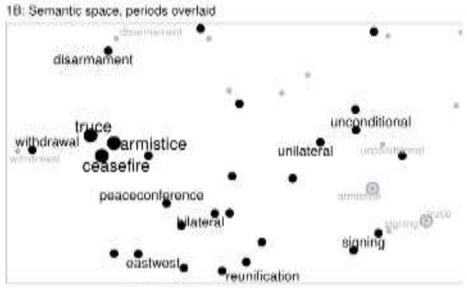
- Work in progress: a model of lexical density (~extent of colexification) using diachronic word embeddings.

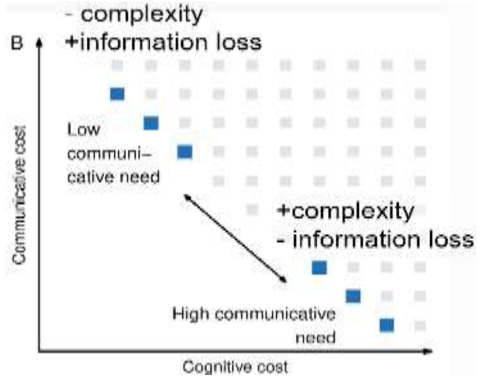
- Big picture: shifts along the optimal front guided by communicative need (social, cultural, environmental, etc)
- Languages change because people *need* them to change
- What about other products of cumulative cultural evolution?

Work in progress: *Capturing aesthetic complexity in art using compression ensembles* https://andreskarjus.github.io/talks/cultureconference2021

thanks!







Extras

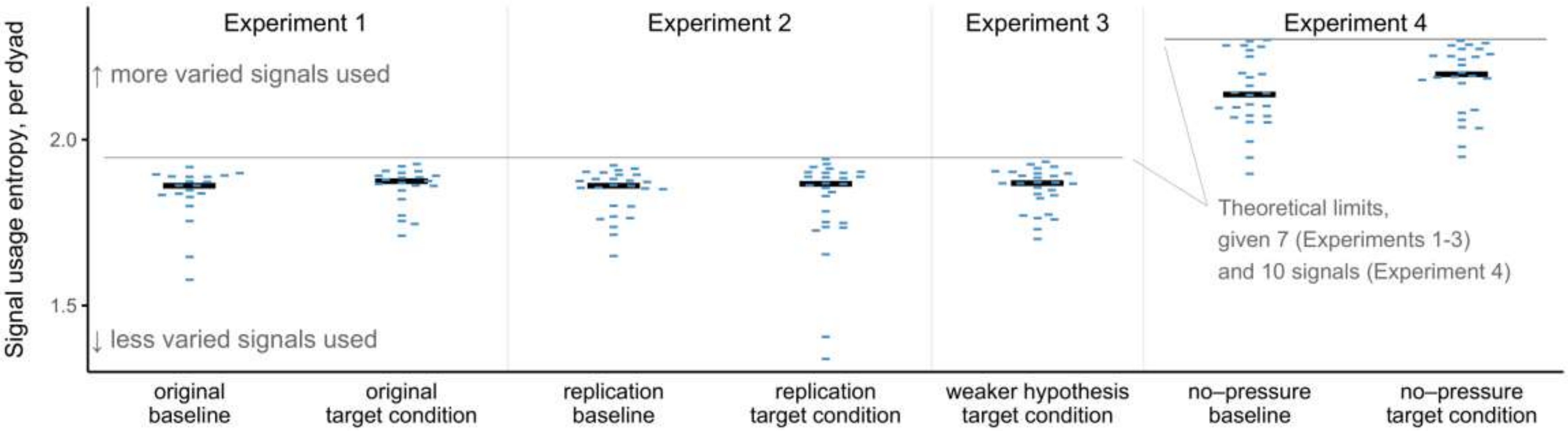


Figure 8: Signaling entropy across all dyads (blue notches), in the post-burn-in part of the game. Low values on the vertical axis indicate fewer signals were consistently used, higher values indicate a larger variety of signals were used by a given dyad. Overlapping values are pushed slightly aside horizontally to ensure visibility. The black bars represent medians. The gray vertical lines at 1.9 and 2.3 indicate maximal entropy given the size of the signal spaces. Dyads in Experiment 4 generally made use of most of the extended signal space (10 instead of 7), setting it apart from the rest of the experiments.

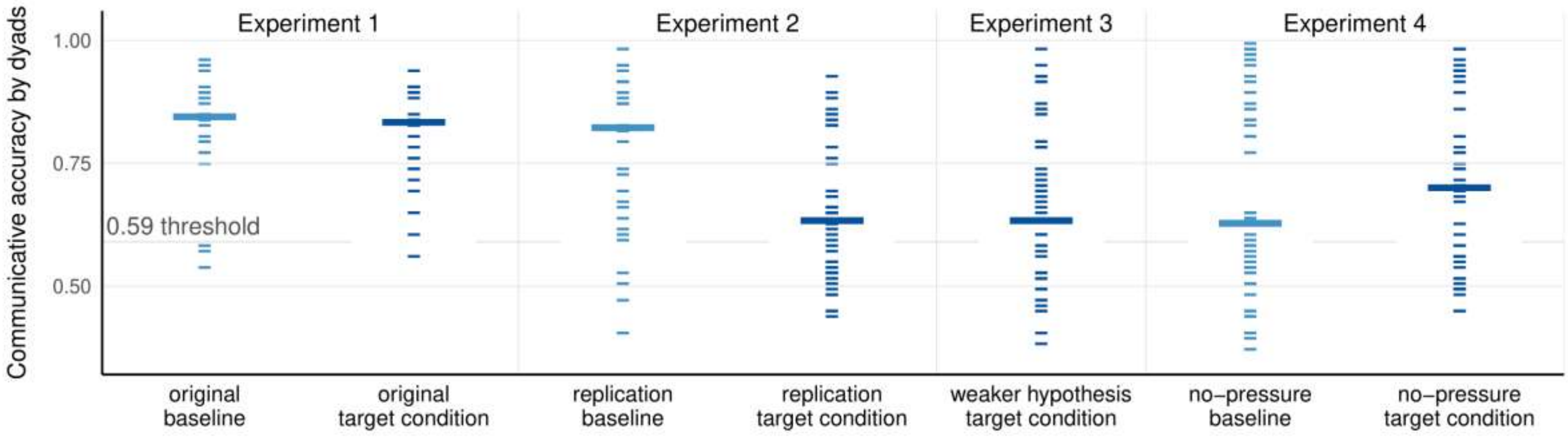


Figure 6: Communicative accuracy of all dyads across all experiments. Each notch is one dyad, the wider bars are medians. Target conditions (those with manipulated communicative need) are in darker blue. The threshold of 0.59, that we set to filter out dyads which likely played by the random button smash strategy, is shown as a gray segmented line. The student dyads (Experiment 1) had on average higher accuracy than the crowdsourced participants in the rest of the experiments, but all conditions included some dyads that scored very low, as well as some that scored very high.

current meaning	last TASK	last JOB	new signal	complexity	ambiguity
TASK	nopo	nopo	nopo	0	1
TASK	nopo	mumi	nopo	1	0
TASK	nopo	nopo	mumi	1	1
TASK	nopo	mumi	mumi	1	2
TASK	nopo	mumi	fita	2	1

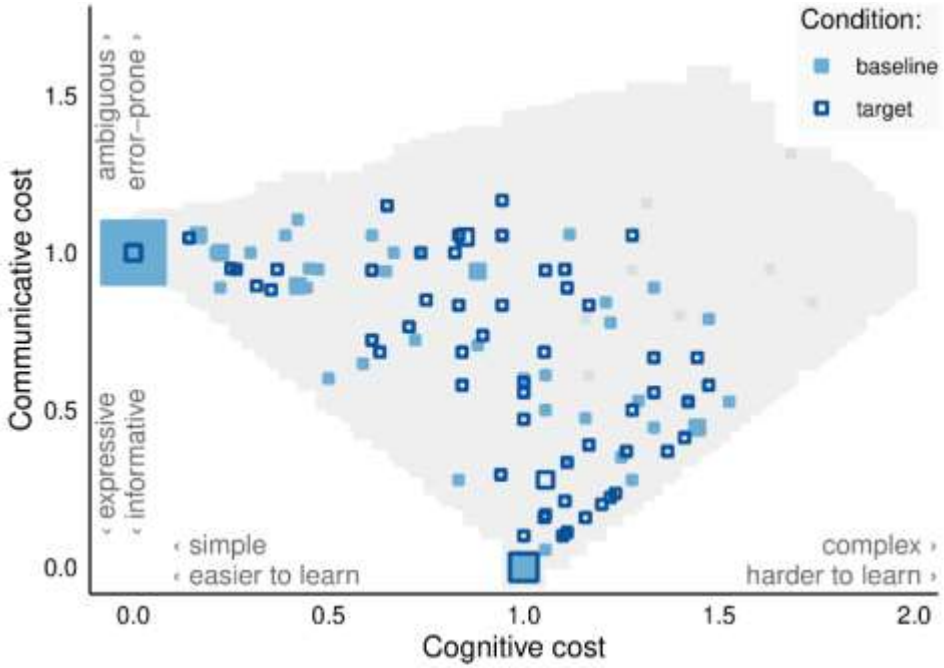
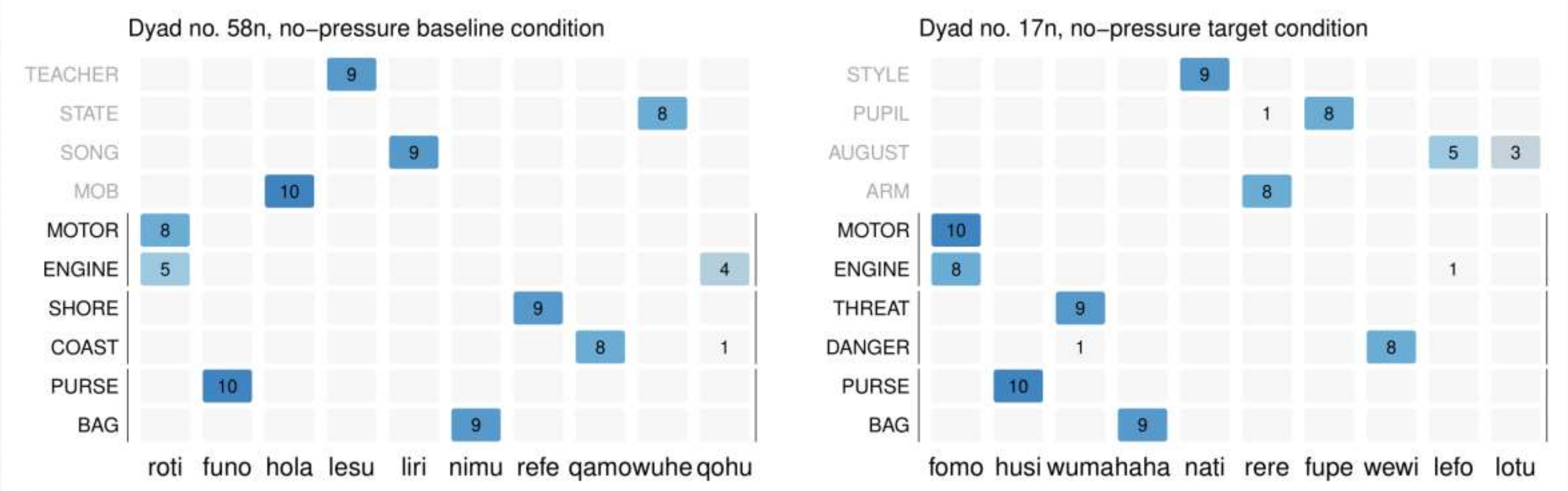


Figure 9: Average communicative cost and cognitive cost (complexity) scores in Experiment 1. Light blue squares are target meaning pairs (such as DRIZZLE-RAIN) used by baseline condition dyads, dark blue ones are pairs as used by target condition dyads. Larger square size indicates multiple overlapping squares at those coordinates. Simulated results are depicted as gray blocks in the background. The few slightly darker gray squares, mostly in the top right, are the few dyads in Experiment 1 that performed below the minimal threshold of 59%. Most dyads communicate at or near the optimal points at $(0, 1)$ and $(1, 0)$, and none of the



Dyad no. 58n, no−pressure baseline condition
TEACHER
STATE
SONG
MOB
MOTOR
ENGINE
SHORE
COAST
PURSE
BAG
9
8
9
10
8
5
4
9
8
1
10
9
roti
funo
hola
lesu
liri
nimu
refe
qamowuhe
qohu
Dyad no. 17n, no−pressure target condition
STYLE
PUPIL
AUGUST
ARM
MOTOR
ENGINE
THREAT
DANGER
PURSE
BAG
9
1
8
5
3
8
10
8
1
9
1
8
10
9
fomo
husi
wumahaha
nati
rere
fupe
wewi
lefo
lotu

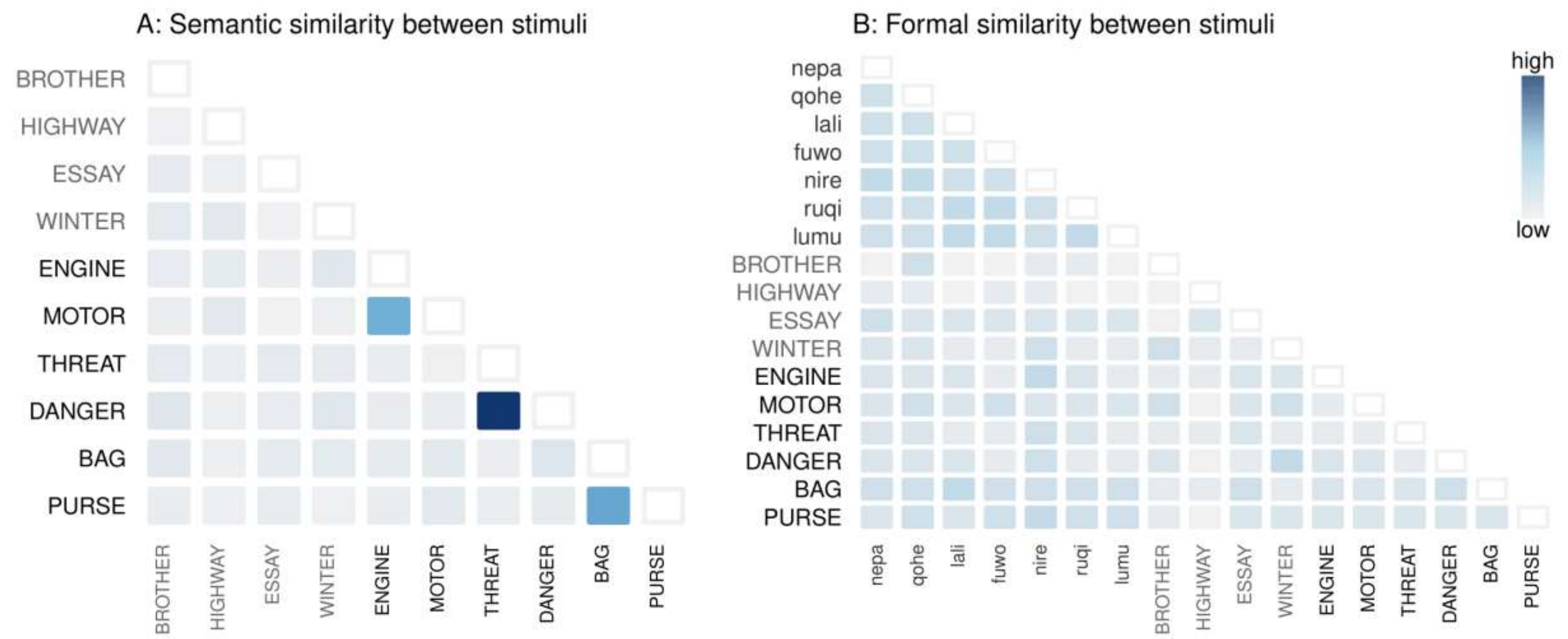


Figure 2: The meanings and signals used in one game (dyad no. 21 in Experiment 1). The left side A panel illustrates the meaning space: only the target pairs have high similarities (dark blue), with low similarities between all other meanings. The diagonal (self-similarity) is marked by white squares. The right side B panel shows the similarity of form (as inverse of edit distance), of both the meanings (in caps) and signals (lowercase). The stimuli in our experiments are generated in a way that ensures only target meanings are semantically similar to one another, and that form similarity remains low across the board.