Sensorimotor and linguistic distributional knowledge in semantic category production: An empirical study and model

Briony Banks
b.banks@lancaster.ac.uk

Cai Wingfield
c.wingfield@lancaster.ac.uk

Louise Connell
l.connell@lancaster.ac.uk

Department of Psychology
Lancaster University, UK

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 682848.

Background and overview

- The importance of sensorimotor and linguistic information in conceptual processing is debated.

- **Linguistic Shortcut Hypothesis:**
  In conceptual processing, linguistic associations may be computationally cheaper than deeper sensorimotor simulations, providing a linguistic shortcut.

- What is the role of sensorimotor and linguistic information in semantic category production?

  1. Behavioural Studies

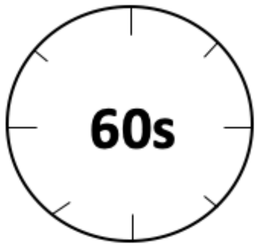
  2. Modelling Study

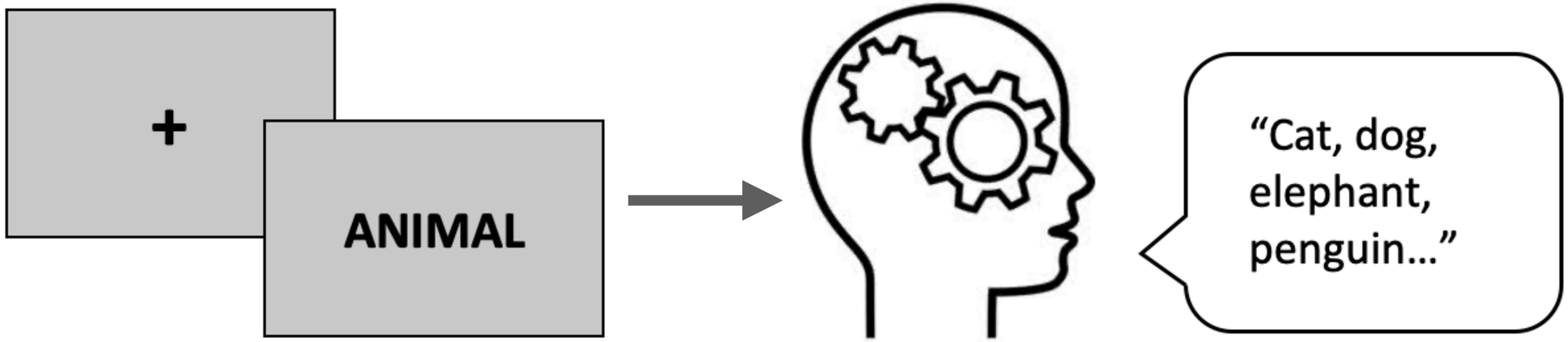
Behavioural studies

Study 1A

Semantic category production

NAME AS MANY CATEGORY MEMBERS AS POSSIBLE



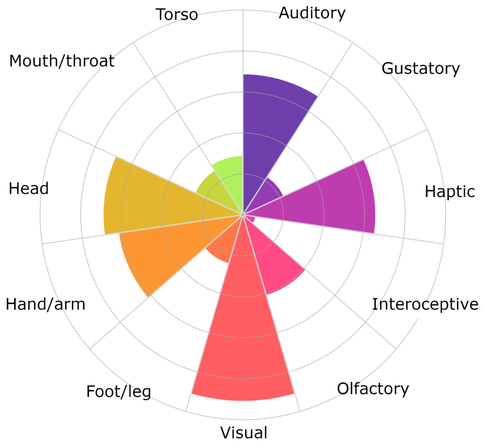


117 categories

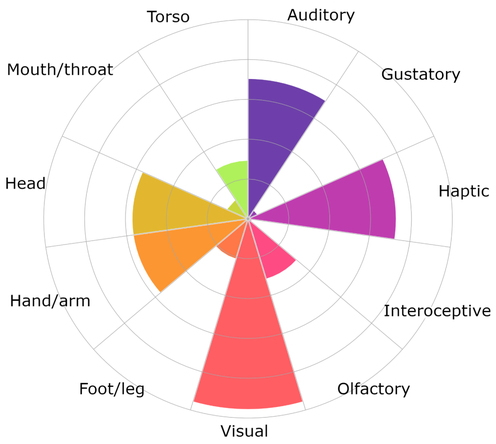
20 participants per category

Study 1A

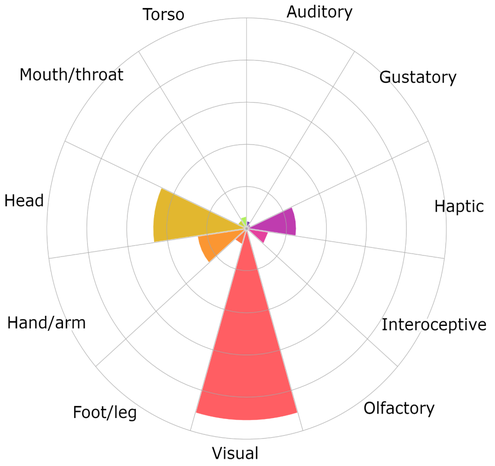
Predictor 1: sensorimotor similarity



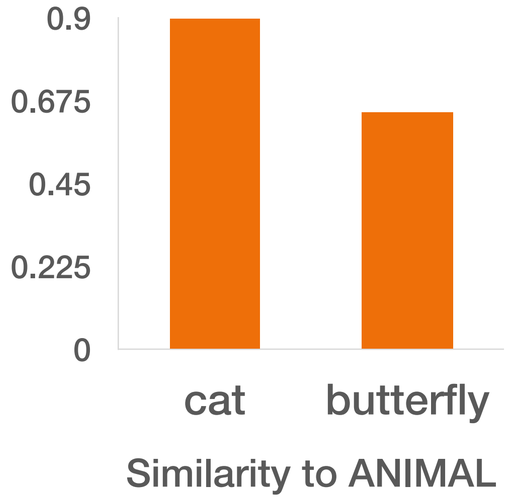
ANIMAL



CAT



BUTTERFLY



- Vector distance (Minkowski-3)
- Higher values = greater similarity

11 dimensions of sensorimotor experience

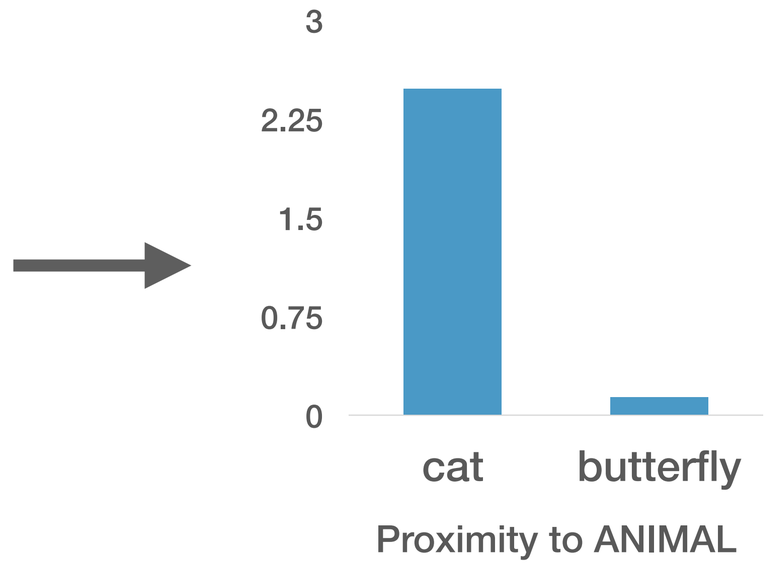
Lynott et al. (2020) doi:10.3758/s13428-019-01316-z

Study 1A

Predictor 2: linguistic proximity

My **CAT** is a funny **ANIMAL**

She took the stray **CAT** to an **ANIMAL** shelter



Co-occurrence (PPMI n-gram, r = 5) of each category + concept in a corpus of 200 million subtitles.

- Higher values = greater proximity

Study 1A

Bayesian hierarchical regressions

Dependent variables

1. Production Frequency
2. Mean Rank
3. First Rank Frequency
4. Response Time (first responses)

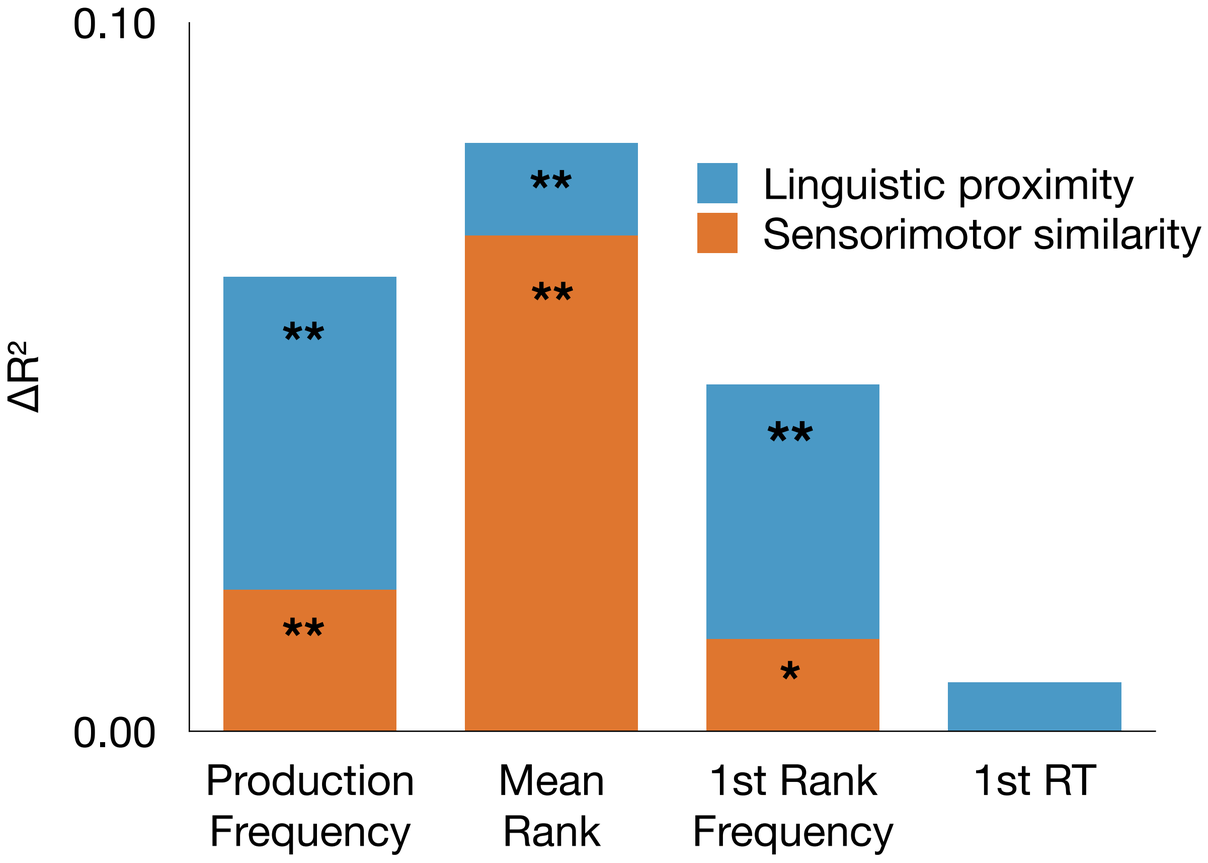
Baseline lexical predictors

Word frequency (LgSUBTLWF)

Word frequency, number of syllables, PLD20 (+ random effects of participant and category)

Study 1A

Results



** BF ≥ 20 (strong evidence for H1)
* BF ≥ 3 (positive evidence for H1)

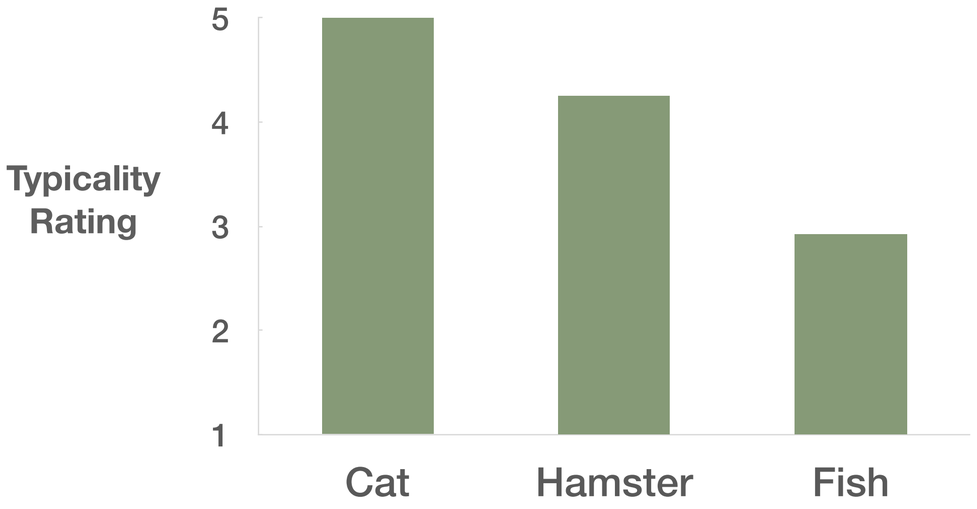
- Concepts with similar sensorimotor strength and that appear in the same linguistic contexts to the category, were named more frequently and earlier.
- Linguistic proximity contributed over and above sensorimotor similarity.
- Only equivocal evidence for RTs.

Banks, Wingfield & Connell (2020). PsyArXiv. https://doi.org/10.31234/osf.io/7eucw

Study 1B

Typicality ratings

How good an example of an ANIMAL is a…?



- Known to predict order and frequency of category production responses[1]
- Do sensorimotor similarity and linguistic proximity predict category production *over and above* typicality?

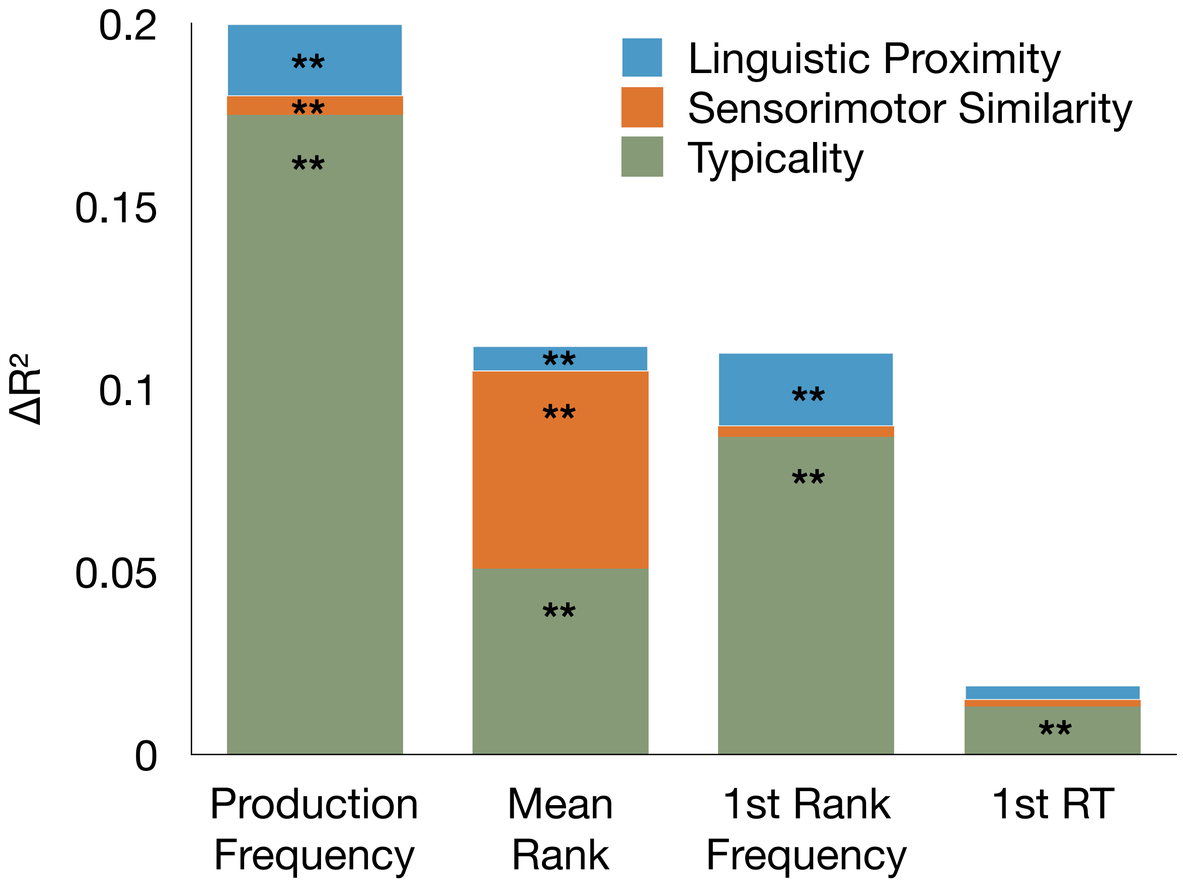
2234 word pairs (category + concept)
N = 12 ratings per word pair

PREREGISTERED

1 E.g., Rosch 1975, 1976, 1978; Barsalou & Sewell, 1985

Study 1B

Results



** BF ≥ 20 (strong evidence for H1)
* BF ≥ 3 (positive evidence for H1)

- Typicality predicted all four measures

Over and above typicality:

- Linguistic proximity predicted the frequency of responses
- Sensorimotor similarity predicted mean rank

Banks, Wingfield & Connell (2020). PsyArXiv. https://doi.org/10.31234/osf.io/7eucw

Behavioural studies

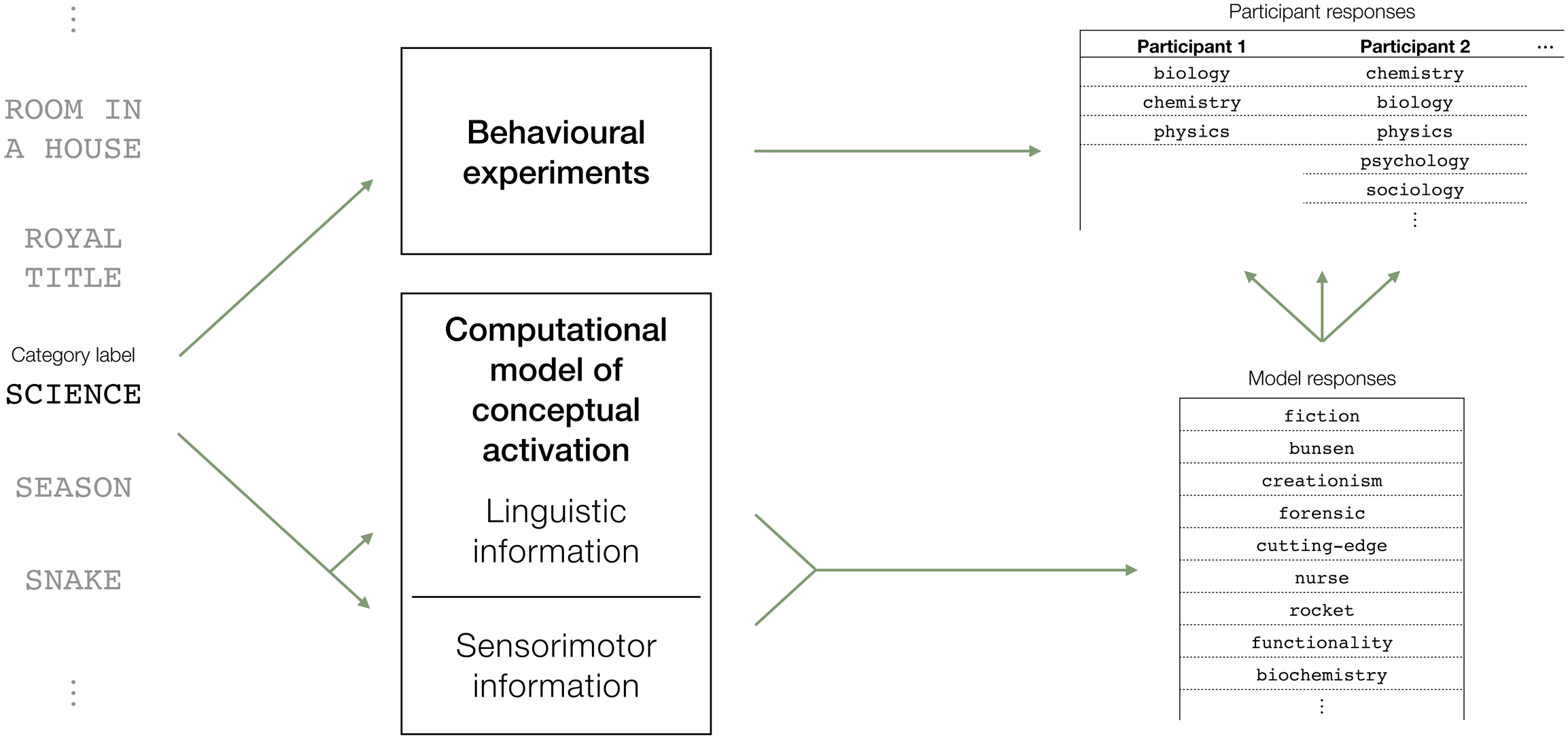
Summary

- Sensorimotor experience and linguistic associations predicted the frequency and rank order of responses.
- Linguistic information contributed over and above sensorimotor information, supporting the Linguistic Shortcut Hypothesis.
- Typicality of concepts was important, but sensorimotor and linguistic information contributed over and above typicality.

Computational modelling study

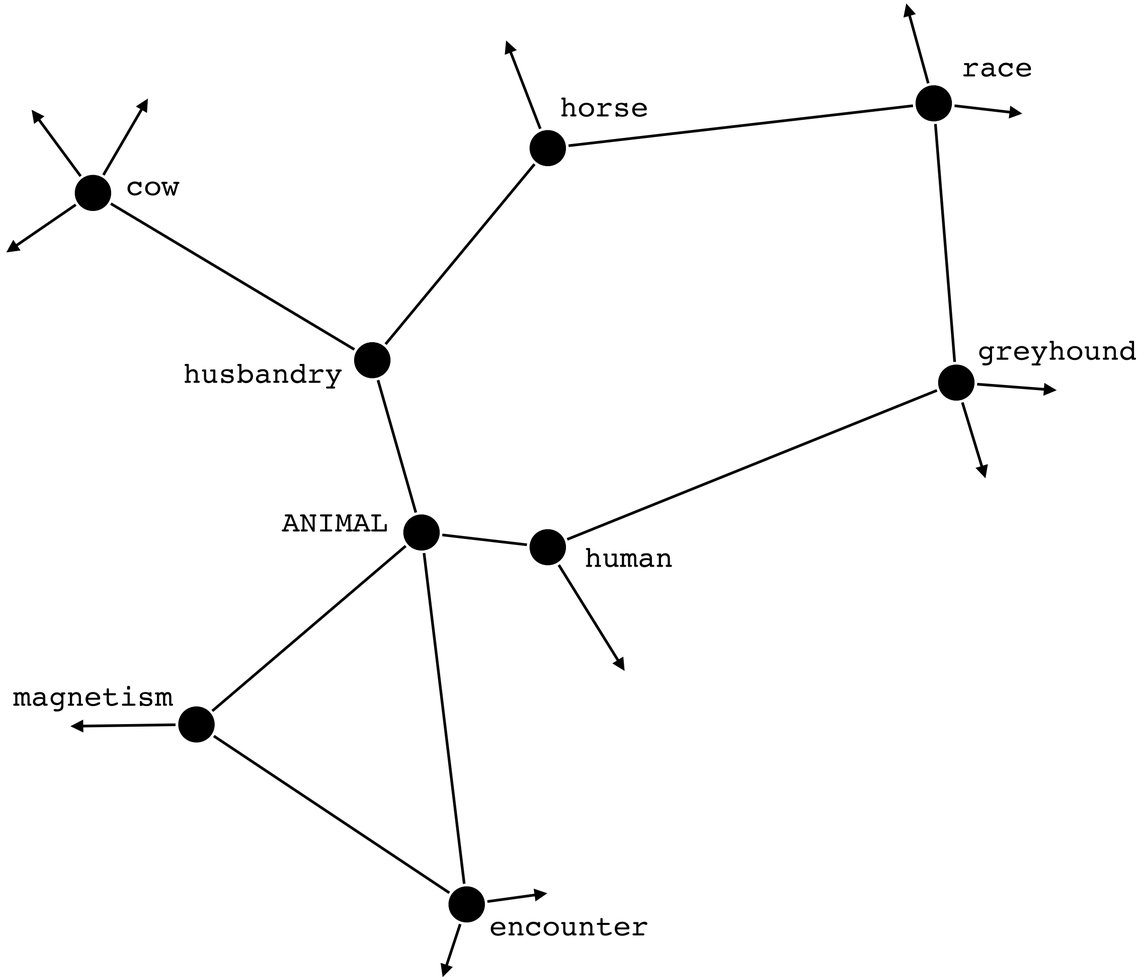
Cognitive model

Overview



Cognitive model

Linguistic component



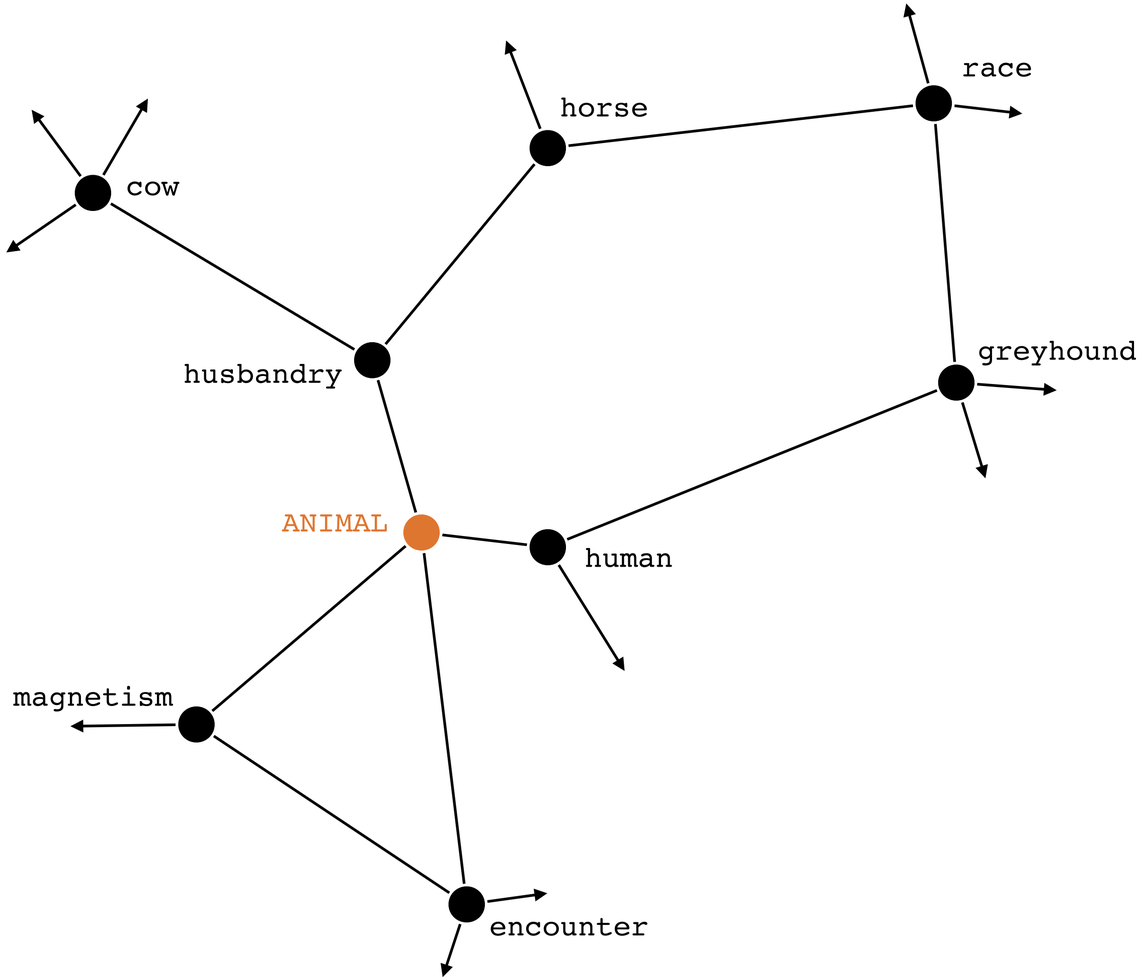
- 40,000 words from subtitles corpus*.
- Co-occurrence PPMI between all pairs of words.
- Graph of nodes (words) and edges (distances; PPMI > 0).
- Activation spreads along edges to neighbouring nodes.

*van Heuven et al. (2014) doi:10.1080/17470218.2013.850521

Cognitive model

Linguistic component

ANIMAL



Category label

ANIMAL

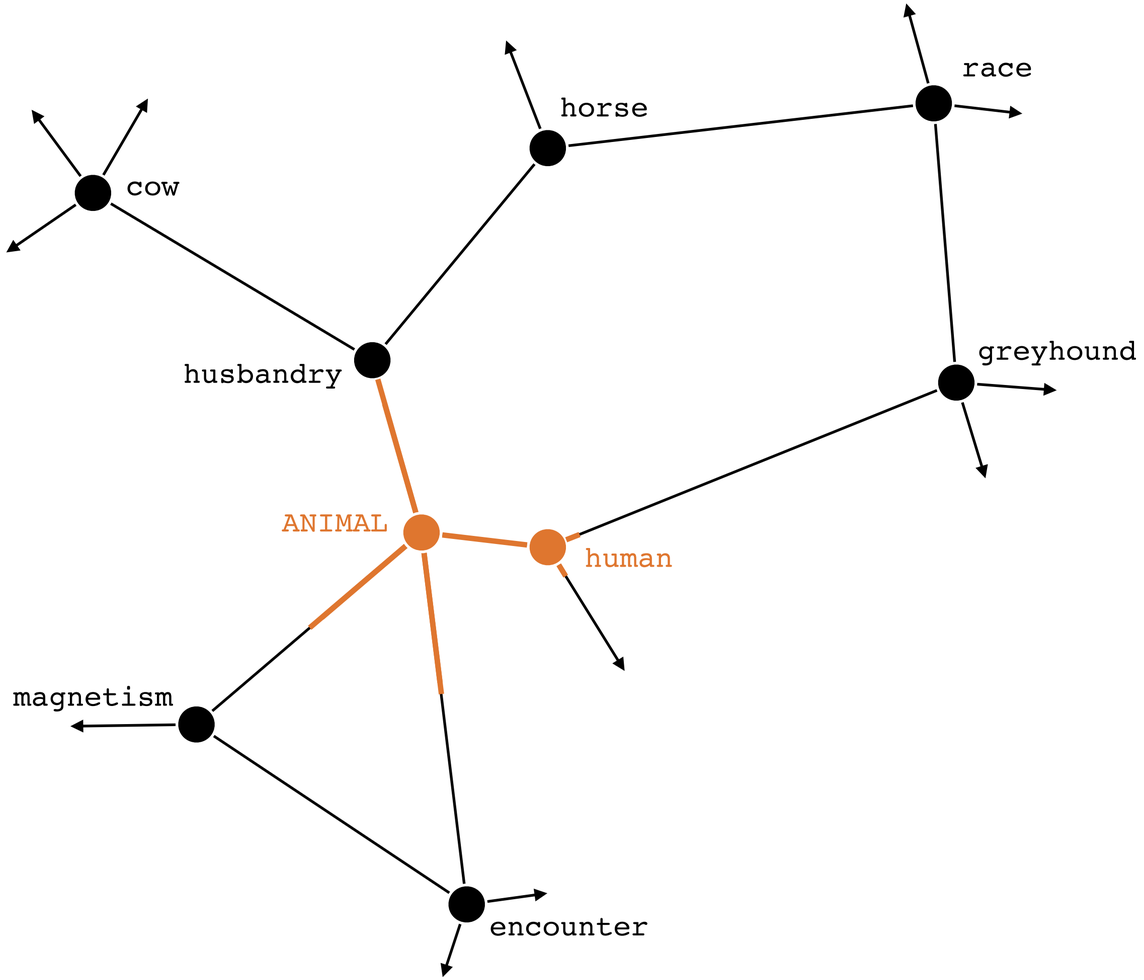
Cognitive model

Linguistic component

ANIMAL

↓

human

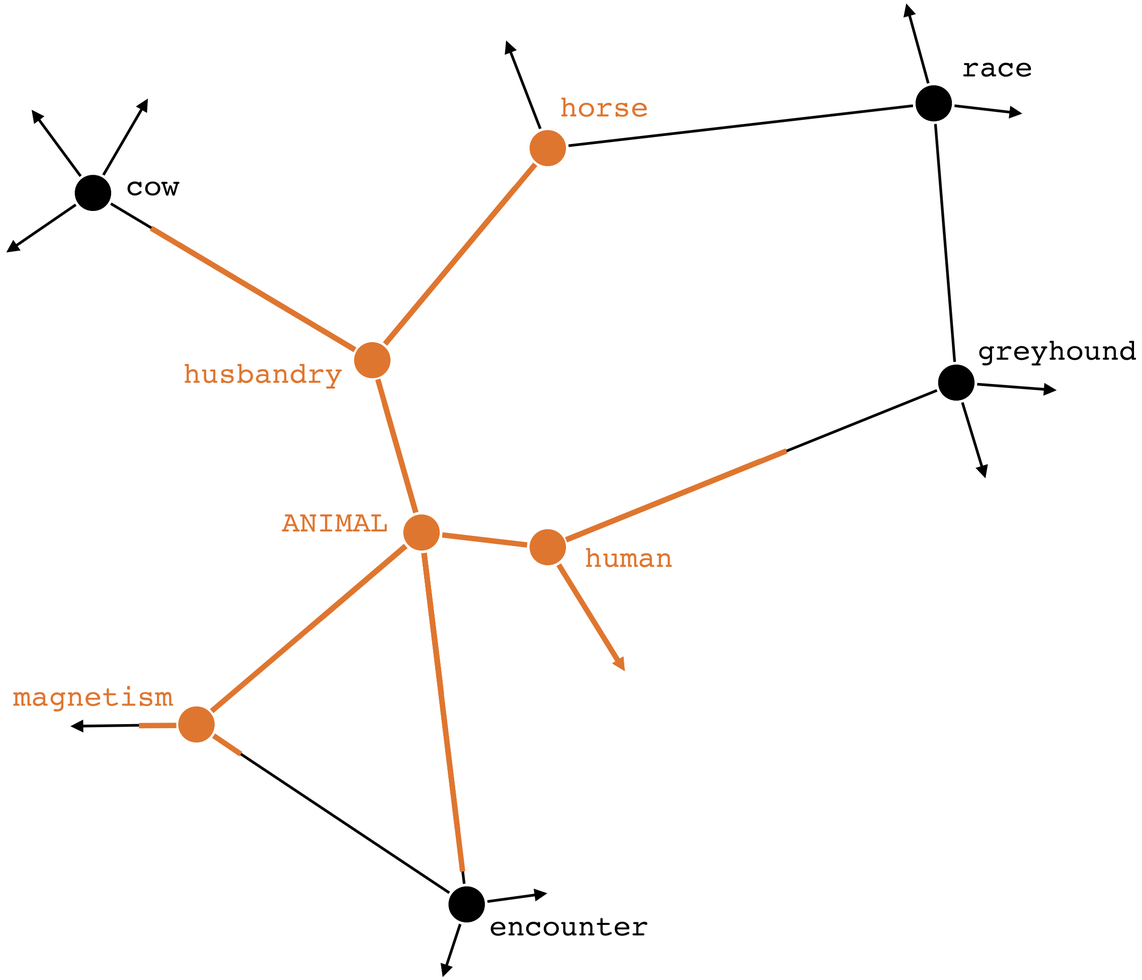


Category label

ANIMAL

Cognitive model

Linguistic component



ANIMAL

↓

human

husbandry

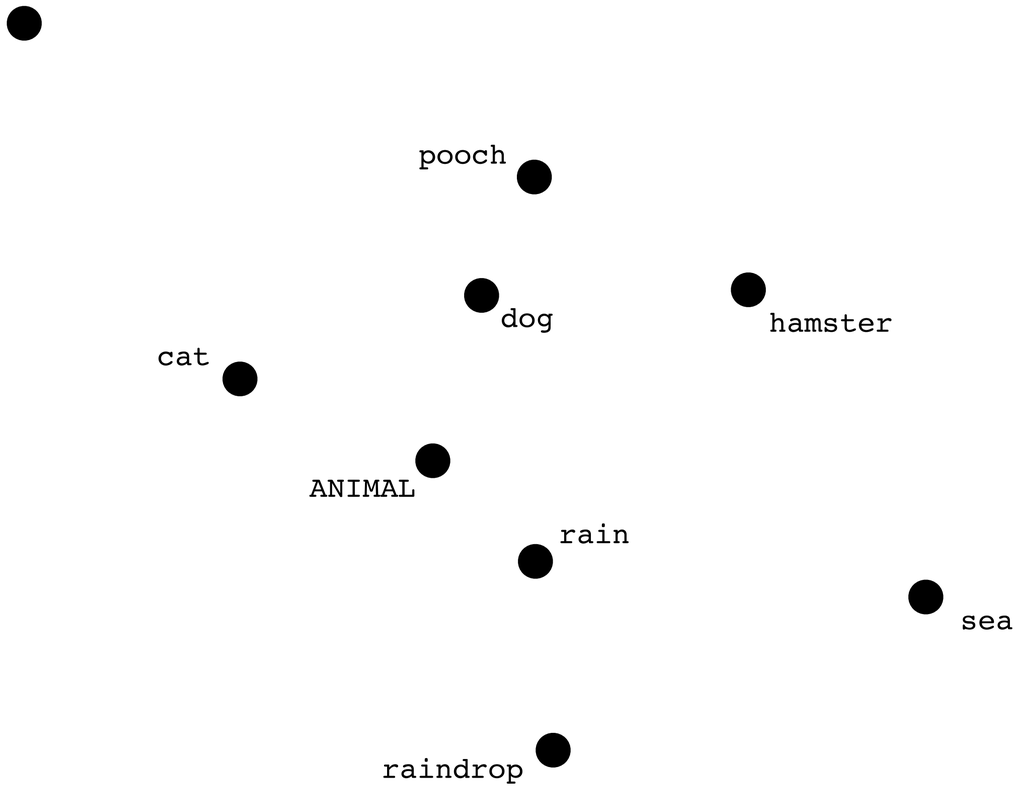
magnetism

horse

Category label

ANIMAL

Cognitive model



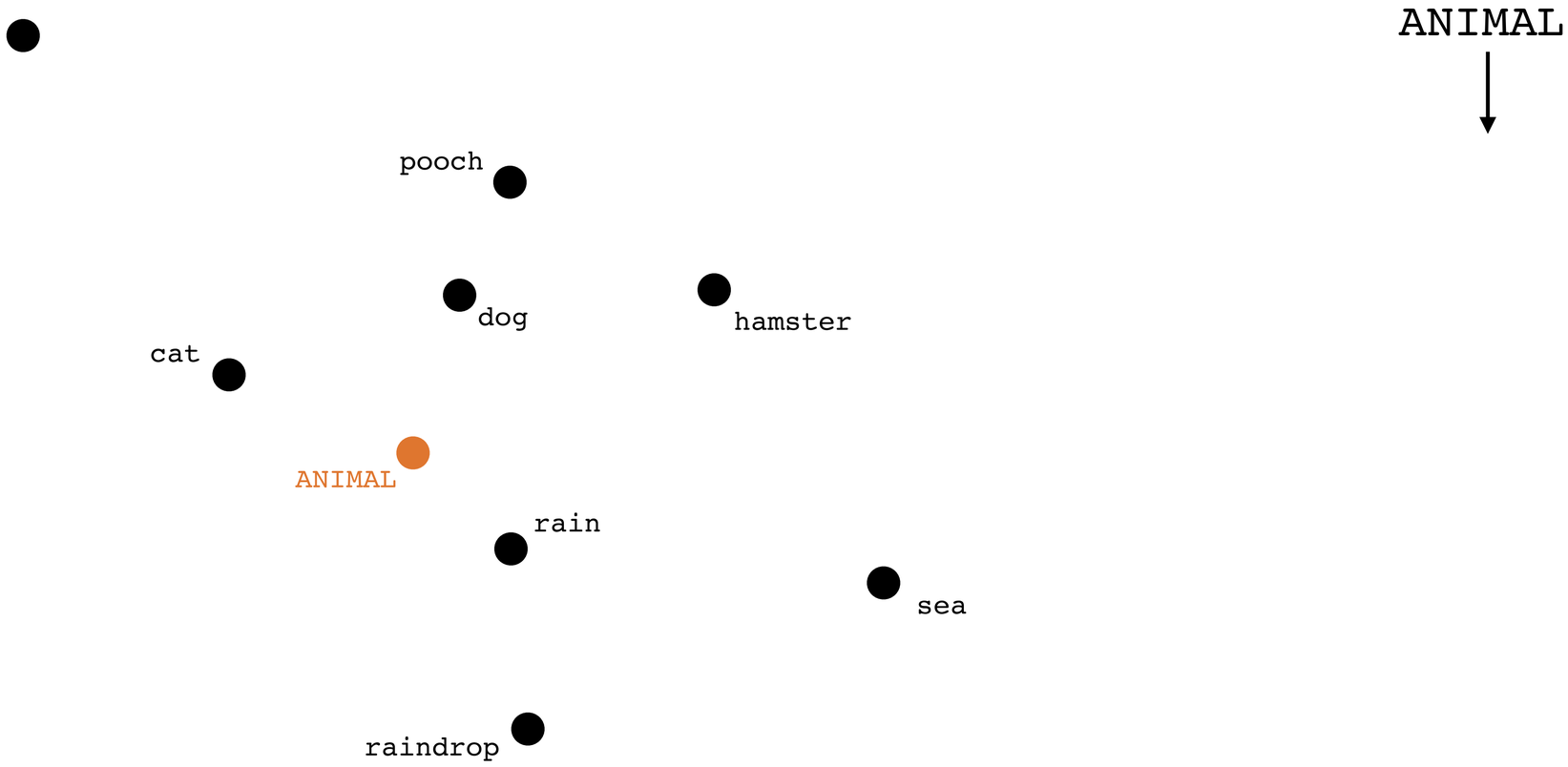
Sensorimotor component

- Approx. 40,000 concepts from the Lancaster Sensorimotor Norms*.
- Concepts are located in 11-dimensional space.
- Activation spreads uniformly in space to neighbouring concepts.
    - Distance between sensorimotor vectors.

*Lynott et al. (2020) doi:10.3758/s13428-019-01316-z

Cognitive model

Sensorimotor component

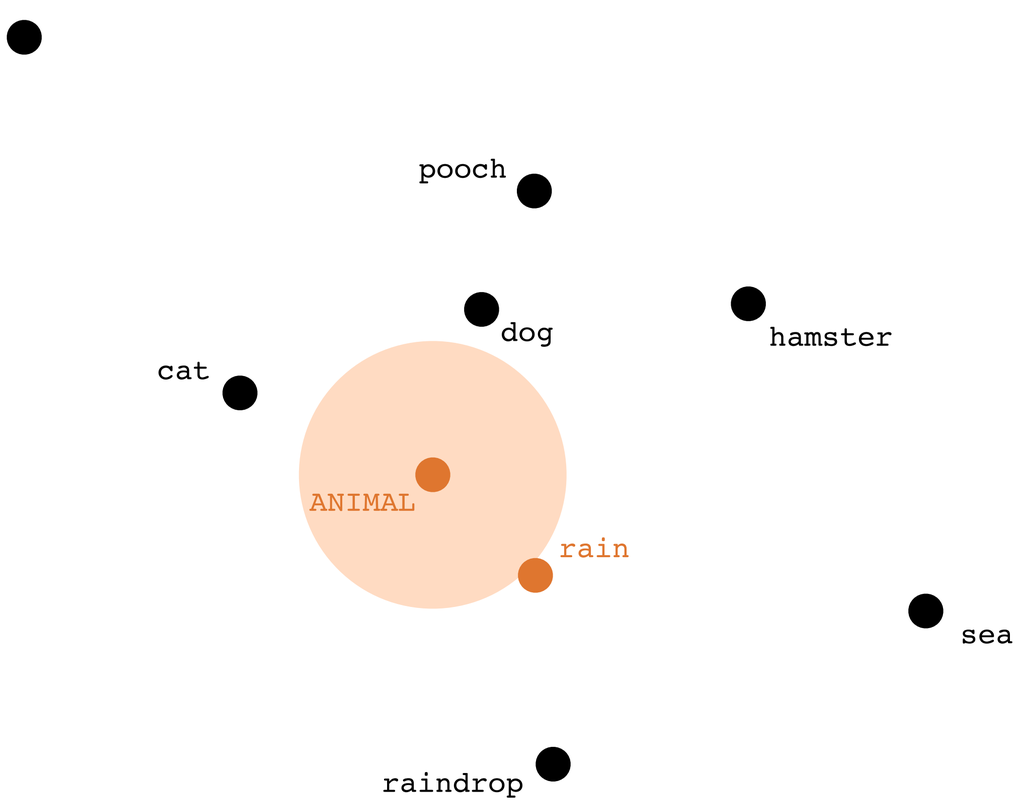


Category label

ANIMAL

Cognitive model

Sensorimotor component



ANIMAL

↓

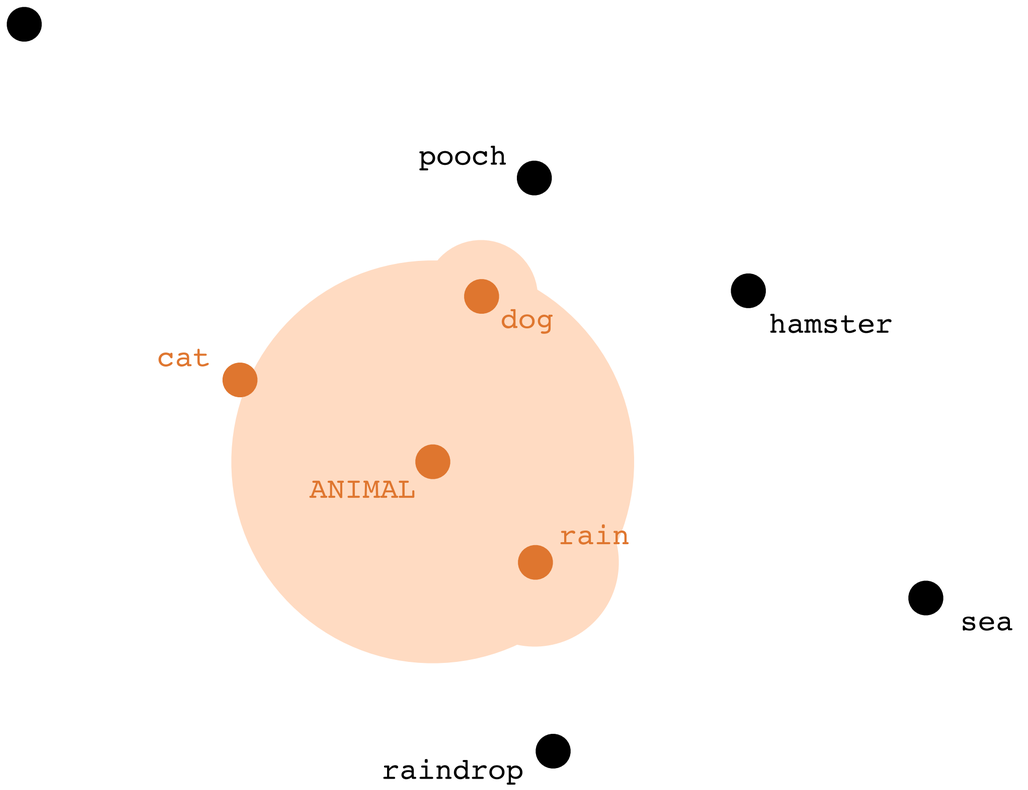
rain

Category label

ANIMAL

Cognitive model

Sensorimotor component



ANIMAL

↓

rain

dog

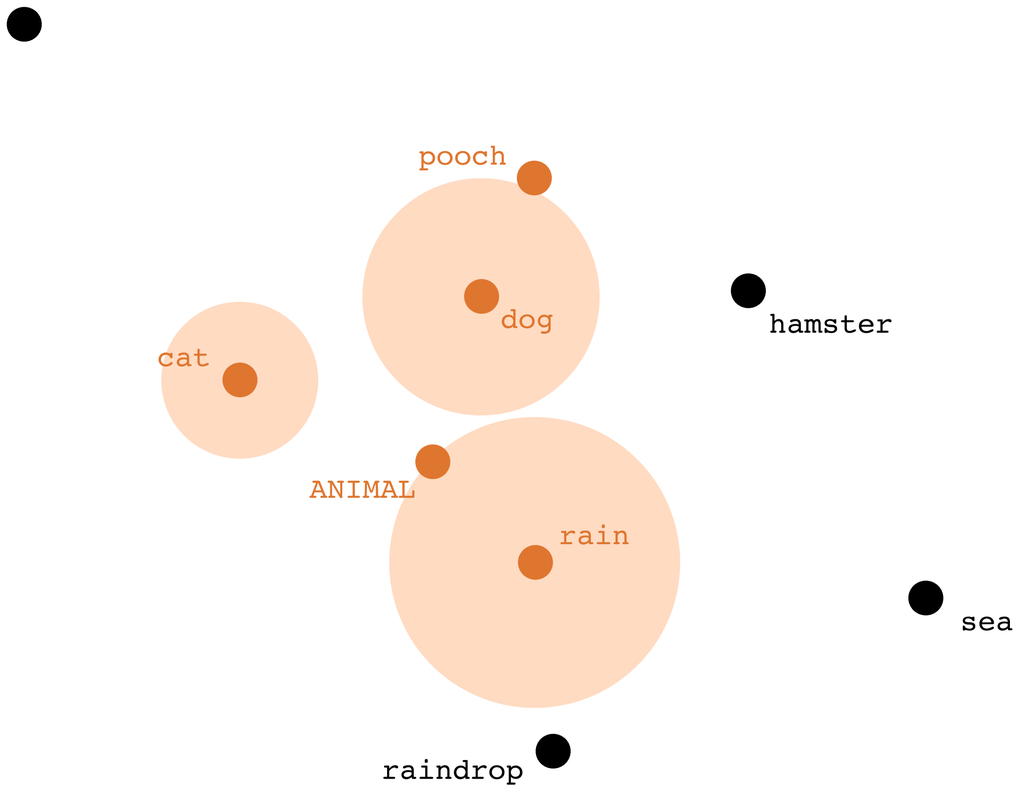
cat

Category label

ANIMAL

Cognitive model

Sensorimotor component



ANIMAL

↓

rain

dog

cat

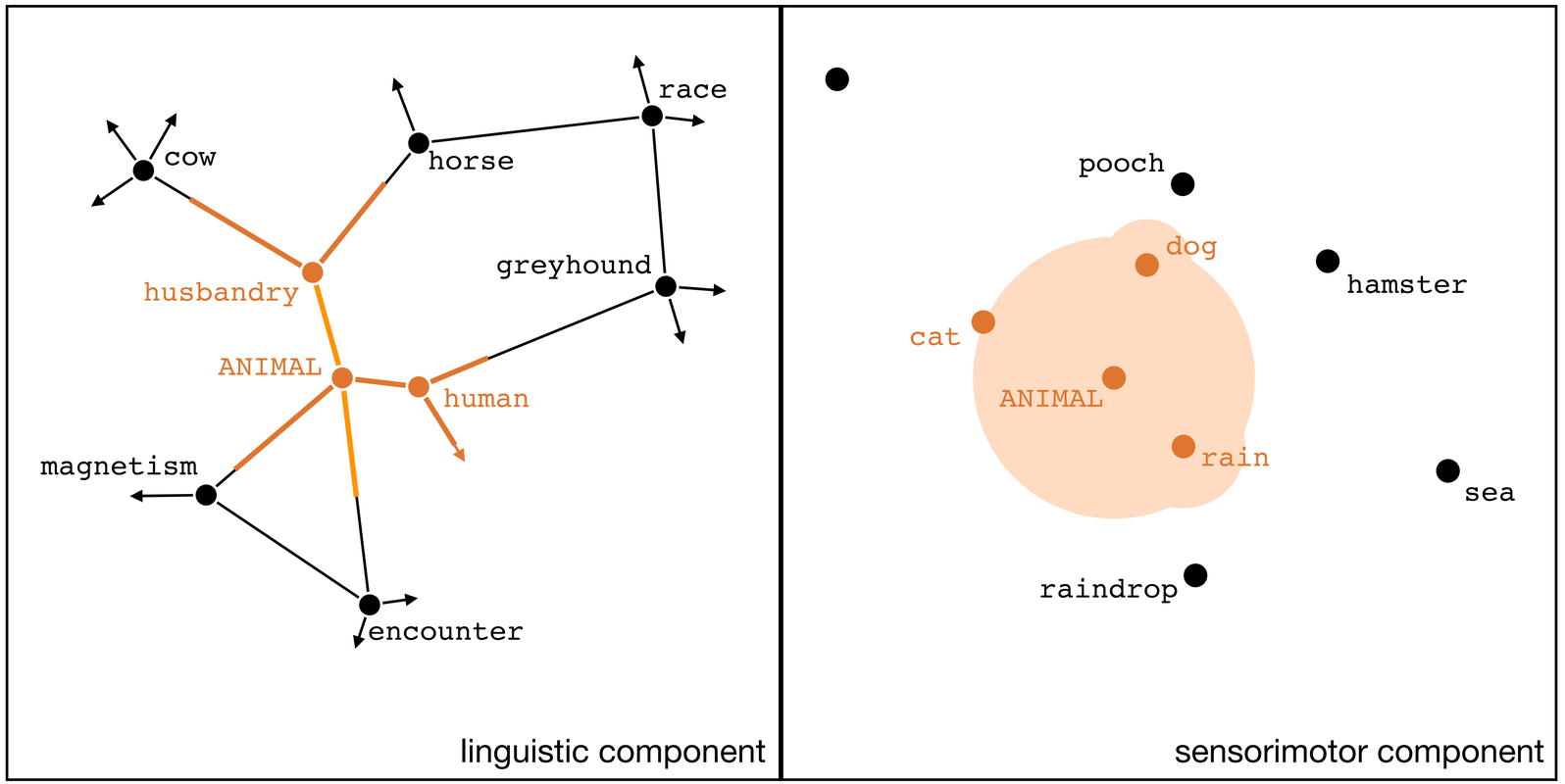
pooch

Category label

ANIMAL

Category label

ANIMAL



rain
cat
dog

human
husbandry

pooch

magnetism

raindrop

⋮ ⋮

Evaluating the model

- How often does the model match participant responses in each category?
- Evaluate model against *the performance of a typical participant*.

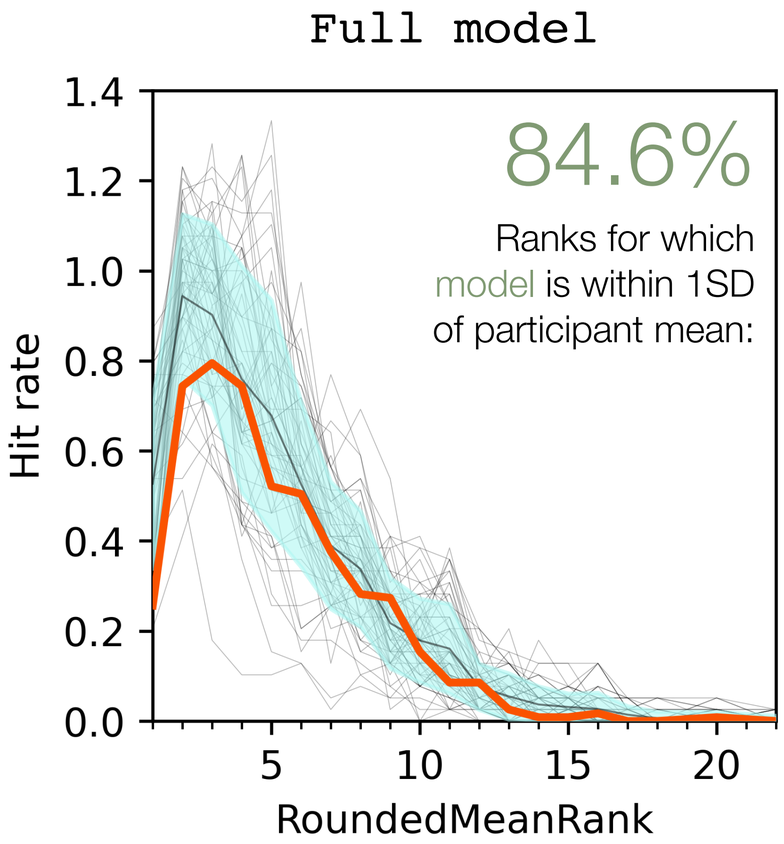
Category label	Participant responses	Mean Rank* (rounded)
CRIME	stealing	1
	burglary	2
	littering	9
	⋮	
FOUR-LEGGED ANIMAL	dog	1
	cat	2
	hippo	11
	⋮	
SNAKE	python	1
	water	3
	boa constrictor	5
⋮	⋮	

Mean Rank*	How many of these responses were produced (per category) by…				
	…the model?	…participant…			
		1?	2?	⋯	Mean
1	26%	64%	87%		53%
2	78%	95%	97%		94%
3	79%	74%	90%		90%
4	74%	67%	62%		76%
5	55%	46%	64%		68%
6	49%	54%	44%		52%
7	38%	26%	33%		39%
8	28%	28%	31%		34%
9	28%	15%	18%		22%
10	16%	10%	26%		18%
⋮	⋮	⋮	⋮		⋮

*Production Frequency results are similar.

Results

- Individual participants
- Participant average
- ±1 SD
- Model performance



Ranks for which
participants were within
1SD of the mean:

68%

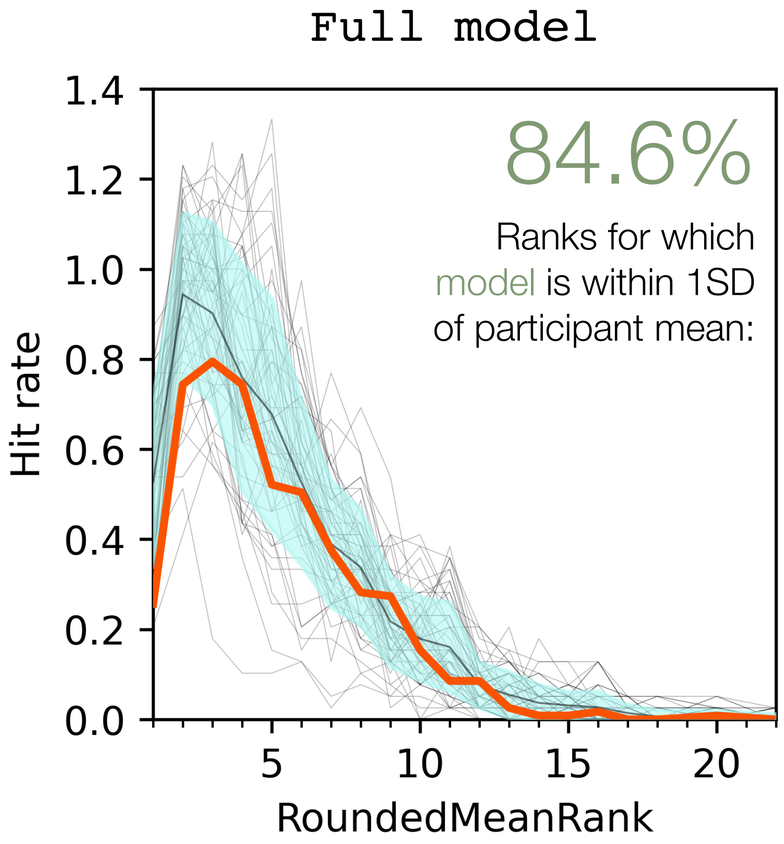
* Hit rates over 1.0 are just due to how we resolve tied ranks: it's not important.

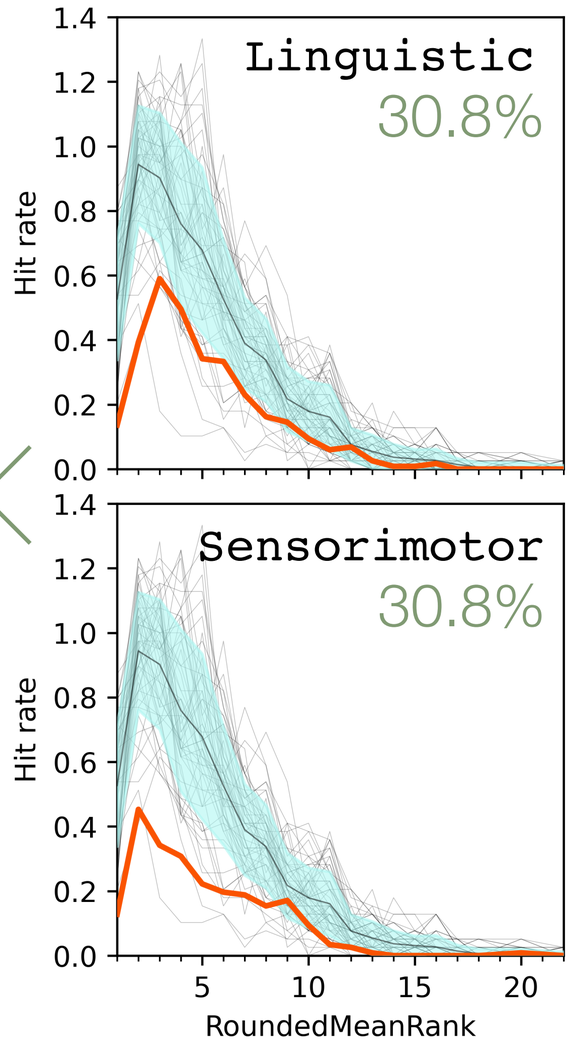
Results

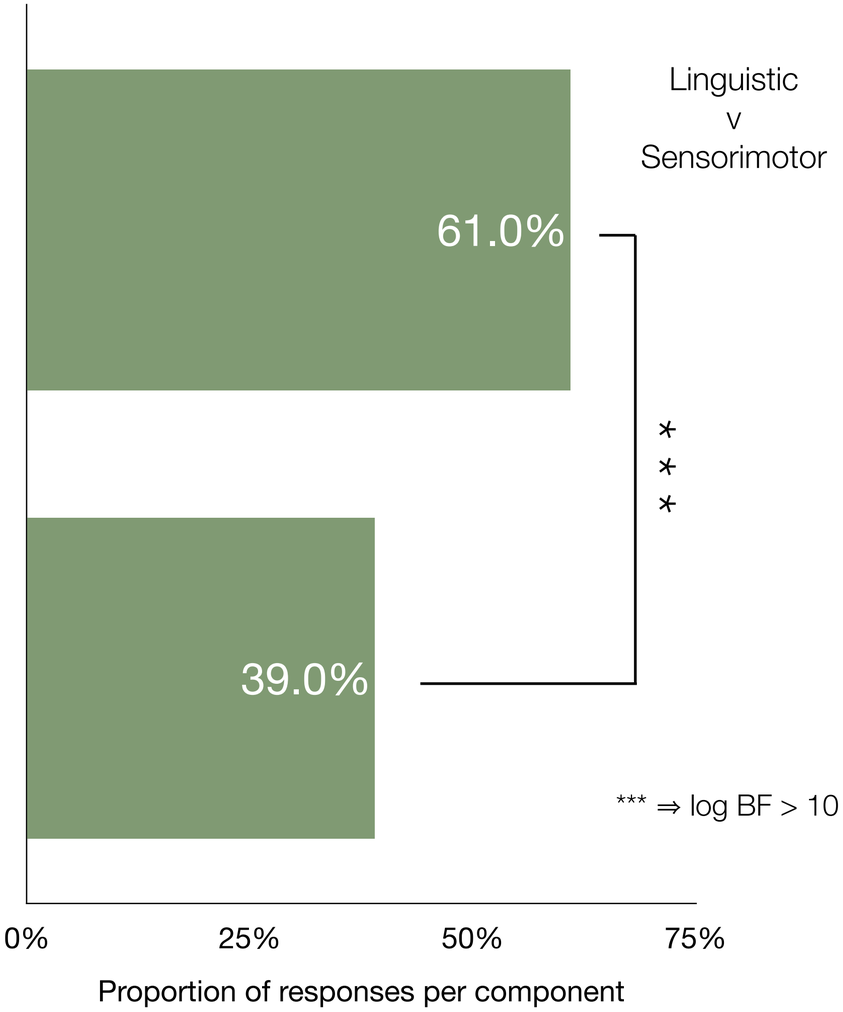
Ranks for which participants were within 1SD of the mean:

68%

- Individual participants
- Participant average
- ±1 SD
- Model performance







Results

Ranks for which participants were within 1SD of the mean:

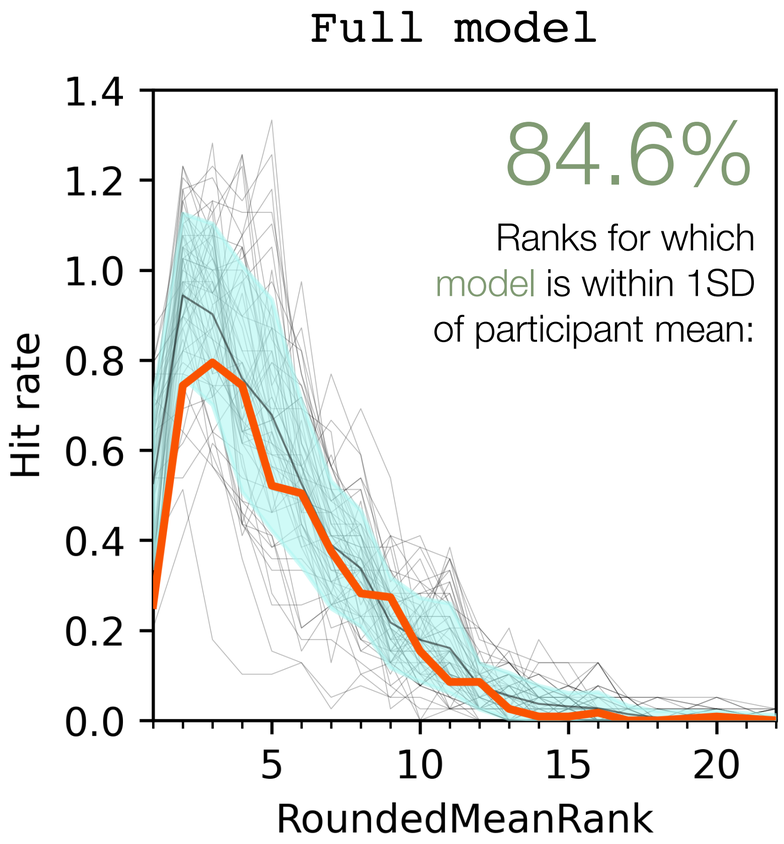
68%

Individual participants

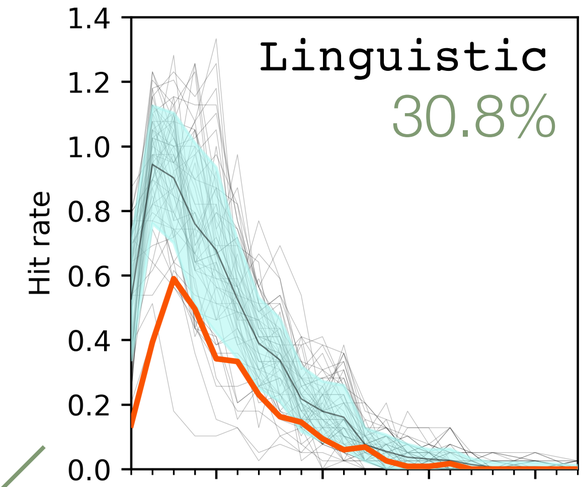
Participant average

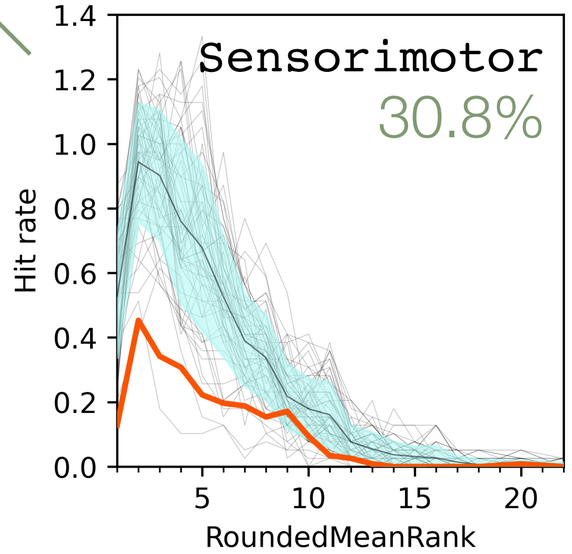
±1 SD

Model performance

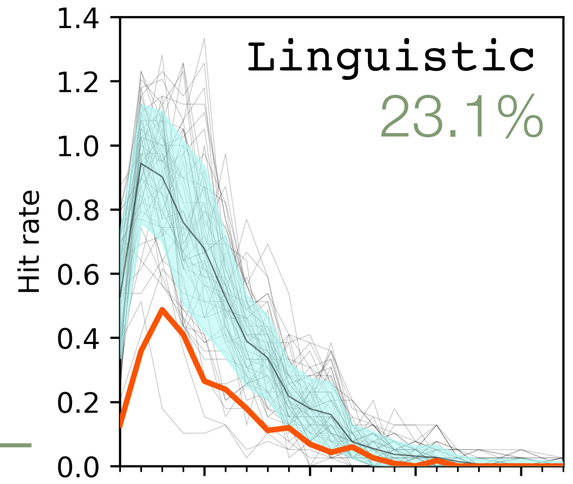


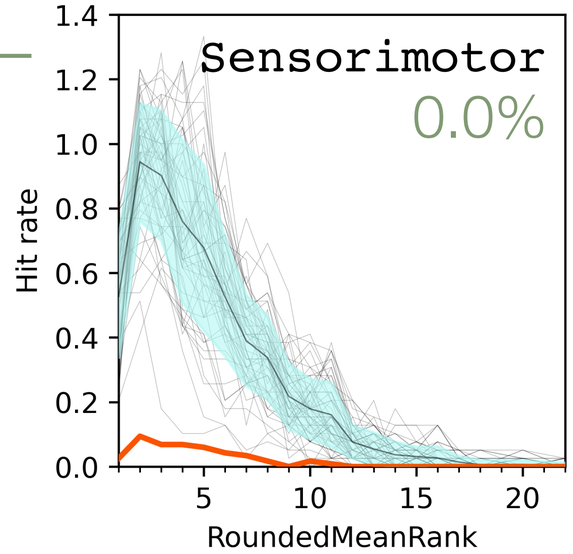
+ indirect activations





- indirect activations





Computational model

Summary

- Computational model incorporated linguistic and sensorimotor knowledge.
- Linguistic and sensorimotor information contributed independently.
- Allowing activations to spread indirectly greatly improved model performance (e.g. ANIMAL → dog → pooch).
- Linguistic component contributed most responses.

Banks, Wingfield & Connell (2020) doi:10.31234/osf.io/7eucw

Conclusions

- Examined the role of linguistic distributional and sensorimotor knowledge in category production.
  - Two pre-registered behavioural experiments and a grounded computational model.
- Results:
  - Showed linguistic and sensorimotor knowledge play independent roles in category production.
  - Showed indirect associations were important to explain participant responses.
  - Provide evidence in support of the linguistic shortcut hypothesis for category production.

Banks, Wingfield & Connell (2020) doi:10.31234/osf.io/7eucw

Lancaster University

Thank you

Briony Banks
b.banks@lancaster.ac.uk

Cai Wingfield
c.wingfield@lancaster.ac.uk

Louise Connell
l.connell@lancaster.ac.uk

erc
European Research Council
Established by the European Commission

Results

- Individual participants
- Participant average
- ±1 SD
- Model performance

Ranks for which average participant was within 1SD of the mean:

67%