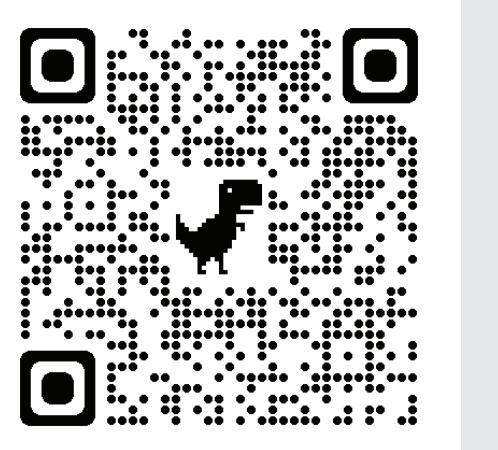




Audiovisual Semantic Relatedness of Real World Objects

Kira Wegner-Clemens¹, George L. Malcolm², Sarah Shomstein¹

Department of Psychological & Brain Sciences, George Washington University¹; School of Psychology, University of East Anglia²



Questions? kira@gwu.edu

How does semantics shape AV attention?

Semantics is important, yet difficult to quantify. Existing methods include:

- Shared-category (e.g., fruits, kitchen items, animate)^{1,2}
- Shared-source (e.g., a meow & a cat, same voice)^{3,4}
- Distributional semantic models based on text corpora⁵

Shared-source sounds make it easier to find a visual target

Is that due to semantics in general or due to shared-source?

To answer that, we needed to create a measure that is:

- Multisensory
- Continuous
- Based on human judgements

Across the two studies described here:

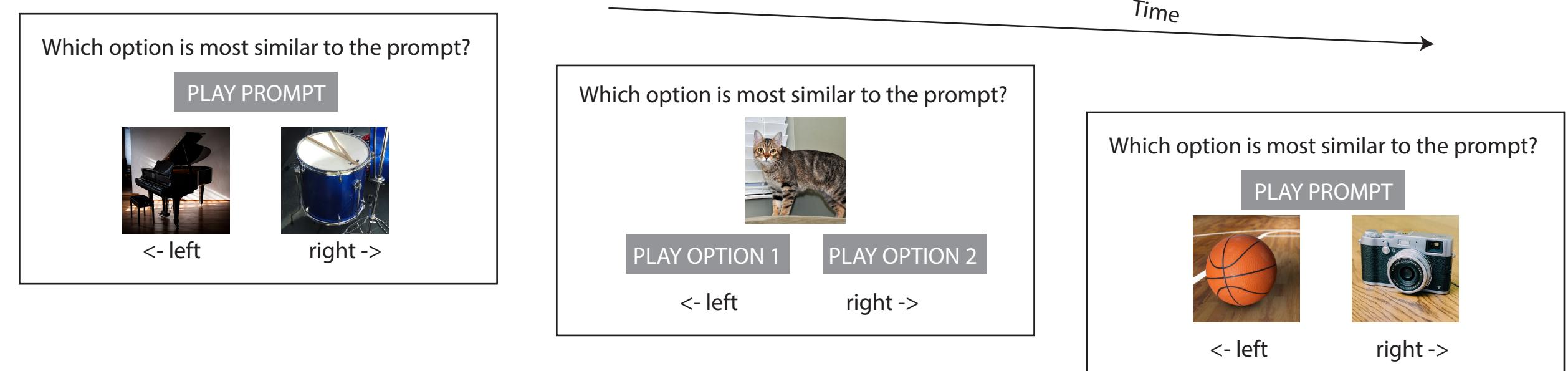
- (1) we created the Sight-Sound Semantics Database⁶
- (2) investigated how relatedness modulates audiovisual attention

Methods

Stimuli: Recognizable images and sounds

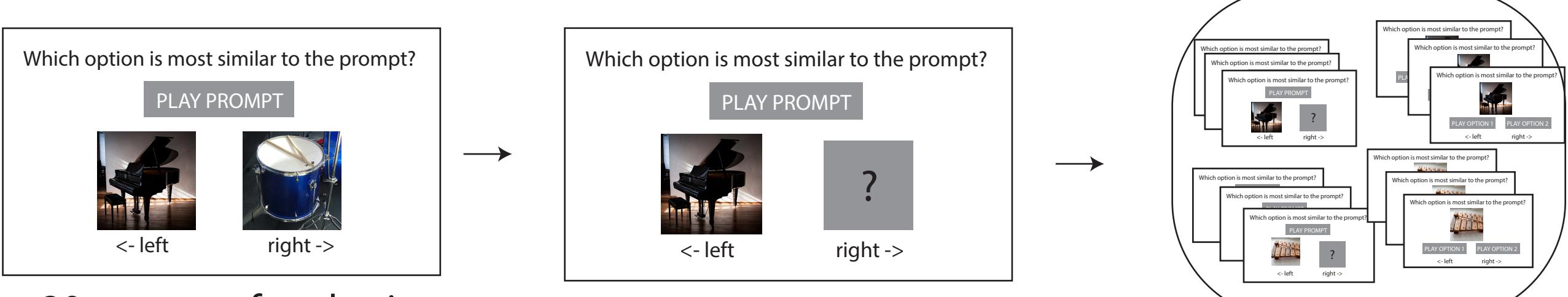
Semantic judgement task: Participants (n=140) made similarity judgements.

- A sound played & participants picked between 2 pictures
- An image appeared & participants picked between 2 sounds



Calculating semantic relatedness values

1. Collect judgements for trios of stimuli
2. For a given prompt & option pair, calculate the likelihood that option will be picked
3. Average over trial, participant, prompt modality



Audiovisual search task: Participants (n=123) searched for target image while a sound in the same category played



Sight-Sound Semantics Database

ANIMALS

| | .1-.2 | .2-.3 | .3-.4 | .4-.5 | .5-.6 | .6-.7 | .7-.8 | .8-.9 | 1 |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| cat - | 0.38 | 0.75 | 0.43 | 0.42 | 0.34 | 0.36 | 0.44 | 0.54 | 0.50 |
| cow - | 0.52 | 0.39 | 0.35 | 0.78 | 0.81 | 0.57 | 0.80 | 0.24 | 0.9 |
| dog - | 0.34 | 0.29 | 0.53 | 0.54 | 0.36 | 0.40 | 0.50 | 0.59 | 0.30 |
| duck - | 0.81 | 0.26 | 0.39 | 0.74 | 0.51 | 0.75 | 0.8 | 0.7 | 0.6 |
| frog - | 0.27 | 0.56 | 0.48 | 0.43 | 0.72 | 0.6 | 0.5 | 0.4 | 0.5 |
| horse - | 0.65 | 0.41 | 0.67 | 0.26 | 0.4 | 0.5 | 0.5 | 0.3 | 0.4 |
| pig - | 0.45 | 0.74 | 0.18 | 0.45 | 0.73 | 0.28 | 0.3 | 0.2 | 0.3 |
| rooster - | 0.57 | 0.73 | 0.28 | 0.57 | 0.73 | 0.28 | 0.3 | 0.2 | 0.2 |
| sheep - | 0.57 | 0.73 | 0.28 | 0.57 | 0.73 | 0.28 | 0.3 | 0.2 | 0.2 |
| songbird - | 0.57 | 0.73 | 0.28 | 0.57 | 0.73 | 0.28 | 0.3 | 0.2 | 0.2 |

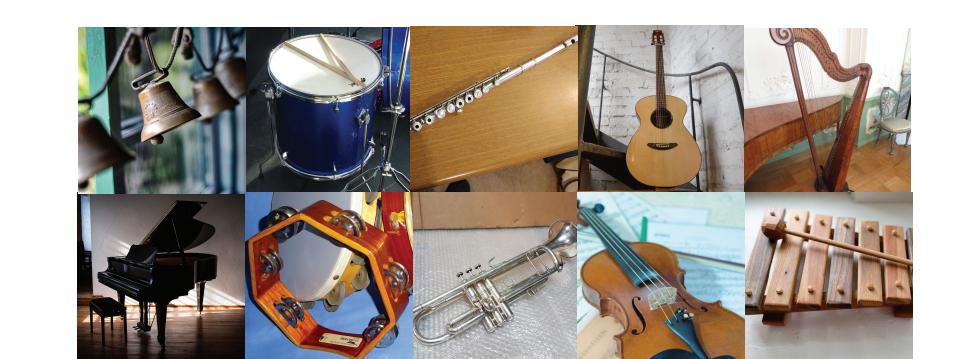
range: 0.18-0.81



INSTRUMENTS

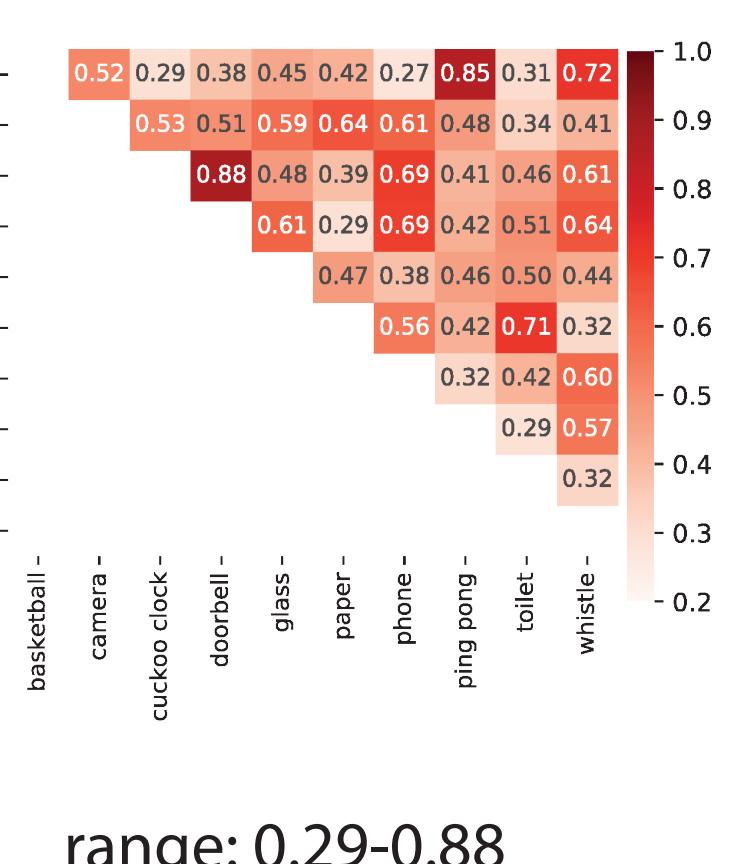
| | .1-.2 | .2-.3 | .3-.4 | .4-.5 | .5-.6 | .6-.7 | .7-.8 | .8-.9 | 1 |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| bell - | 0.51 | 0.47 | 0.32 | 0.49 | 0.36 | 0.76 | 0.42 | 0.27 | 0.76 |
| drum - | 0.23 | 0.39 | 0.16 | 0.34 | 0.75 | 0.61 | 0.21 | 0.68 | 0.9 |
| flute - | 0.46 | 0.60 | 0.50 | 0.37 | 0.83 | 0.63 | 0.55 | 0.8 | 0.7 |
| guitar - | 0.82 | 0.72 | 0.42 | 0.44 | 0.81 | 0.42 | 0.37 | 0.5 | 0.6 |
| harp - | 0.76 | 0.34 | 0.37 | 0.80 | 0.54 | 0.33 | 0.5 | 0.4 | 0.5 |
| piano - | 0.25 | 0.42 | 0.66 | 0.57 | 0.37 | 0.27 | 0.61 | 0.5 | 0.3 |
| tambourine - | 0.37 | 0.42 | 0.61 | 0.33 | 0.56 | 0.31 | 0.3 | 0.2 | 0.2 |
| trumpet - | 0.42 | 0.71 | 0.32 | 0.42 | 0.71 | 0.32 | 0.3 | 0.2 | 0.2 |
| violin - | 0.56 | 0.42 | 0.71 | 0.32 | 0.56 | 0.31 | 0.3 | 0.2 | 0.2 |
| xylophone - | 0.33 | 0.42 | 0.61 | 0.33 | 0.56 | 0.31 | 0.3 | 0.2 | 0.2 |

range: 0.16-0.83

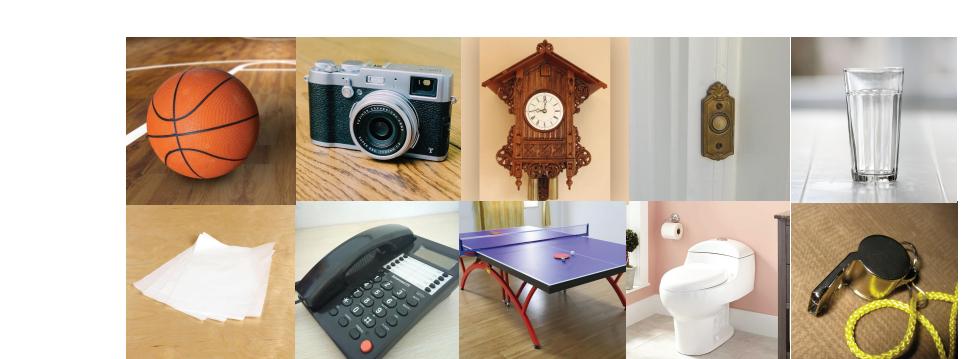


HOUSEHOLD ITEMS

| | .1-.2 | .2-.3 | .3-.4 | .4-.5 | .5-.6 | .6-.7 | .7-.8 | .8-.9 | 1 |
|----------------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| basketball - | 0.52 | 0.29 | 0.38 | 0.45 | 0.42 | 0.27 | 0.85 | 0.31 | 0.72 |
| camera - | 0.53 | 0.51 | 0.59 | 0.64 | 0.61 | 0.48 | 0.34 | 0.41 | 0.9 |
| cuckoo clock - | 0.88 | 0.48 | 0.39 | 0.61 | 0.41 | 0.46 | 0.61 | 0.8 | 0.7 |
| doorbell - | 0.61 | 0.29 | 0.61 | 0.42 | 0.51 | 0.64 | 0.7 | 0.5 | 0.6 |
| glass - | 0.47 | 0.39 | 0.46 | 0.50 | 0.46 | 0.32 | 0.41 | 0.32 | 0.4 |
| paper - | 0.56 | 0.42 | 0.60 | 0.32 | 0.42 | 0.60 | 0.32 | 0.4 | 0.5 |
| phone - | 0.29 | 0.57 | 0.32 | 0.29 | 0.57 | 0.32 | 0.3 | 0.2 | 0.3 |
| ping pong - | 0.42 | 0.60 | 0.32 | 0.42 | 0.60 | 0.32 | 0.3 | 0.2 | 0.3 |
| toilet - | 0.29 | 0.57 | 0.32 | 0.29 | 0.57 | 0.32 | 0.3 | 0.2 | 0.3 |
| whistle - | 0.52 | 0.42 | 0.60 | 0.32 | 0.42 | 0.60 | 0.32 | 0.4 | 0.5 |



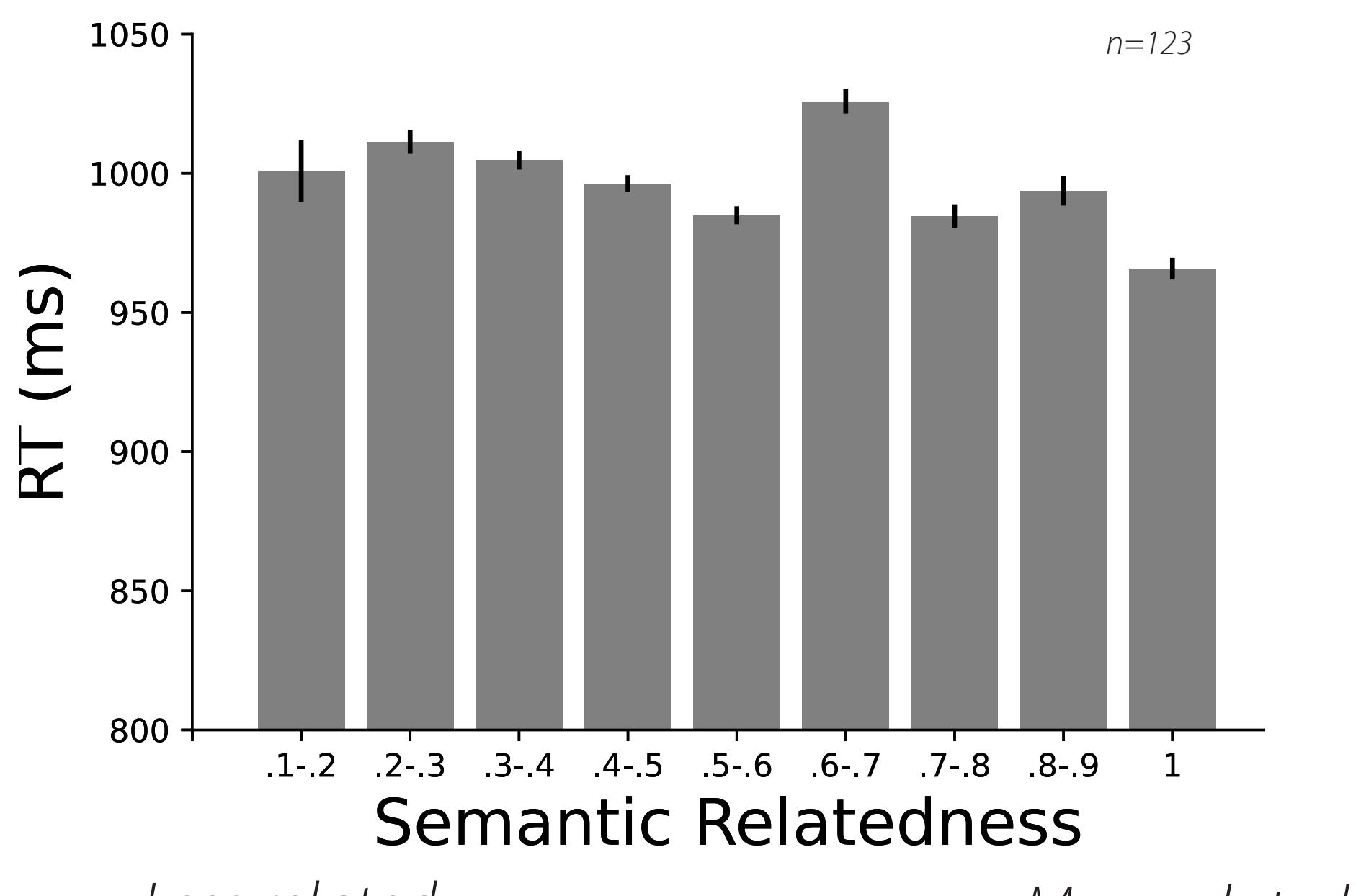
range: 0.29-0.88



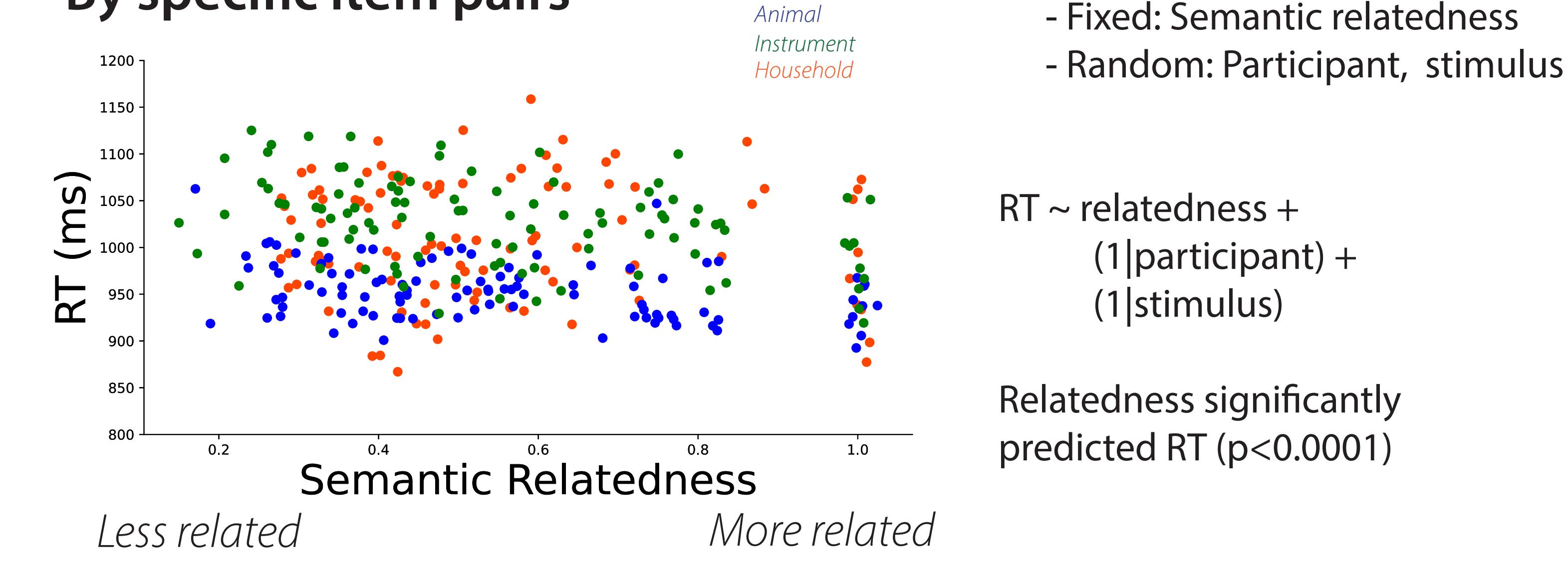
- Each category had pairs that were more related & pairs that were less related
- Participants largely agreed on how related a pair was

A related sound helps you find a visual target

By binned semantic relatedness value



By specific item pairs



Search efficiency increases with semantic relatedness

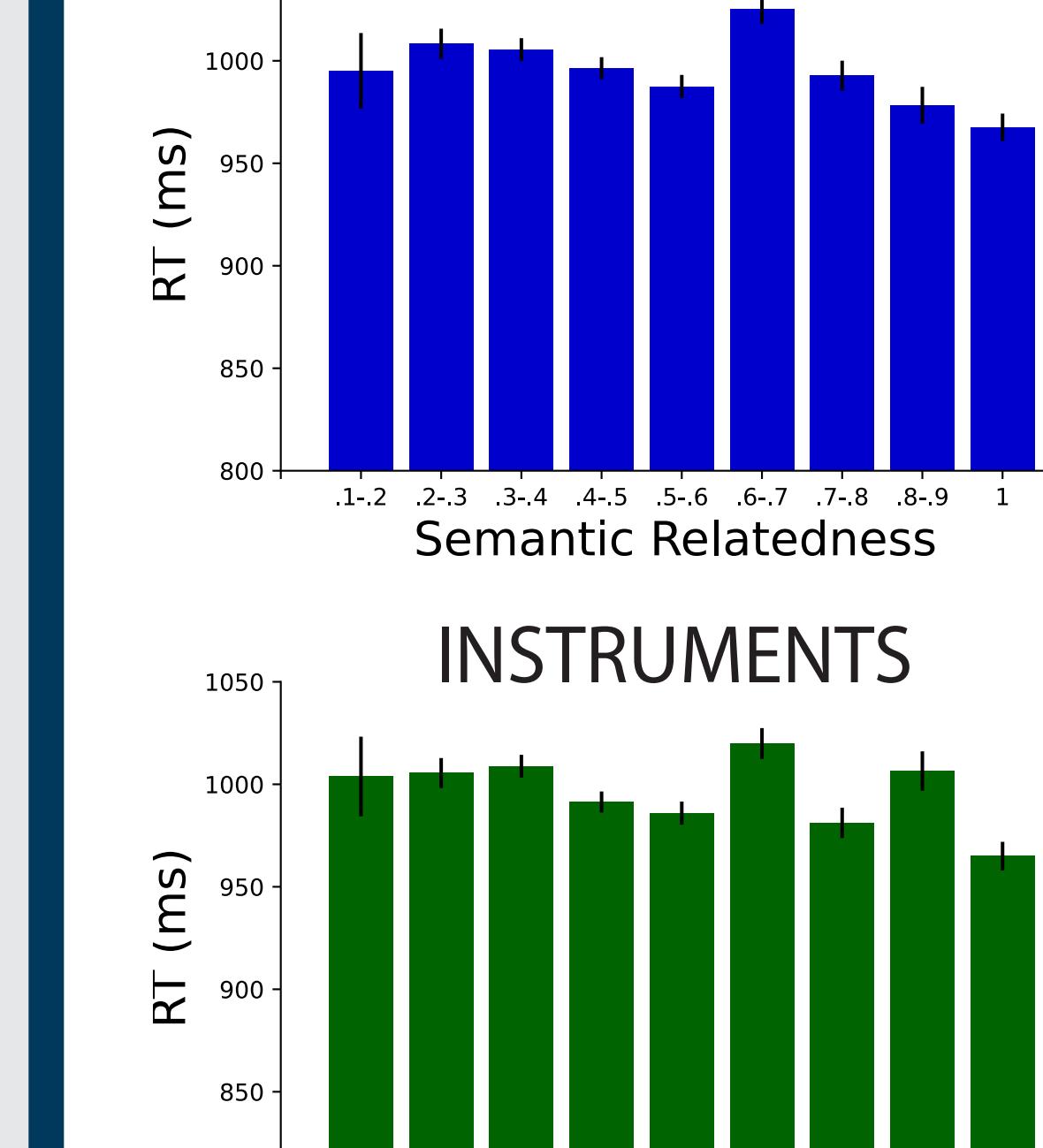
1 way within subject rmANOVA:
p<0.0001

Linear mixed effect model
- Fixed: Semantic relatedness
- Random: Participant, stimulus

RT ~ relatedness + (1|participant) + (1|stimulus)

Relatedness significantly predicted RT (p<0.0001)

Is the search benefit category specific?



Each trial was within category
-- is the effect specific to a subset of trials in one category?

Nope! Each category shows a similar pattern

1 way within subject rmANOVA
- Animals: p<0.0001
- Instruments: p<0.0001
- Household: p<0.0001

Linear mixed effect model
- Fixed: category
- Random: participant, item

Category did not predict RT

Conclusion

The Sight-Sound Database provides measures of semantics that are
- Multisensory (including directionality: auditory & visual prompts)
- Continuous, capturing a range of relatedness
- Based on human judgements (140 participants!)
- Publicly available on OSF

Semantic relatedness predicts audiovisual search speeds

More related sounds lead to better search efficiency

Future work will investigate various remaining questions:

- Do the images need to be task relevant?
- Is there a "near match" repulsion effect?
- What are the underlying neural mechanisms?

References

- (1) Moores, et al 2003 (2) Malcolm, et al 2016 (3) Iordanescu, et al 2008
- (4) Kvasova, et al 2019 (5) Bhatia, et al 2019 (6) Wegner-Clemens, et al, 2022 (7) Mikolov, et al 2017

Acknowledgments

Research supported by NSF BCS-1921415 & BCS-2022572 to SS.