

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

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:

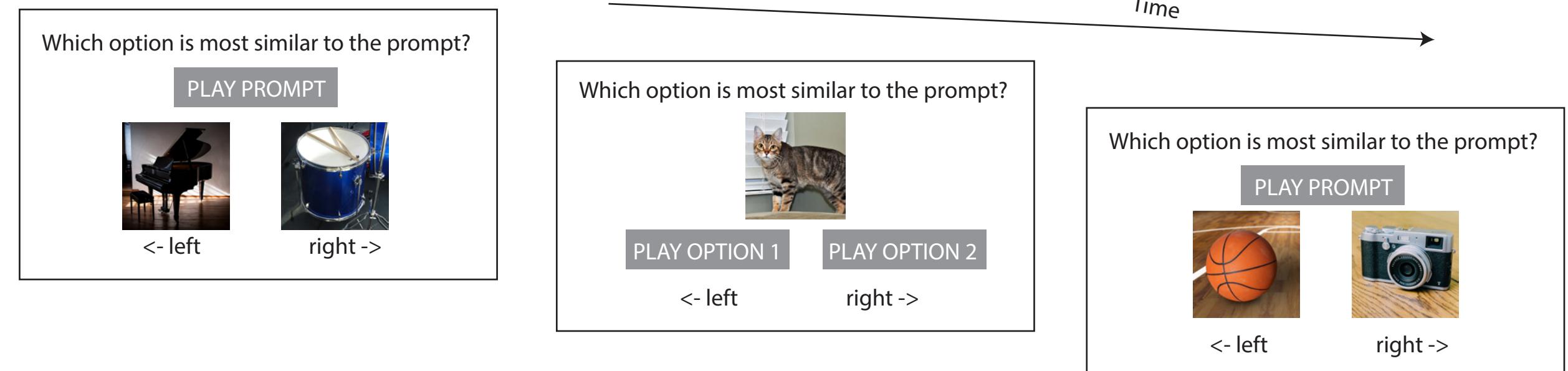
- (1) we created the Sight-Sound Semantics Database⁶
- (2) investigated how relatedness modulates audiovisual attention

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

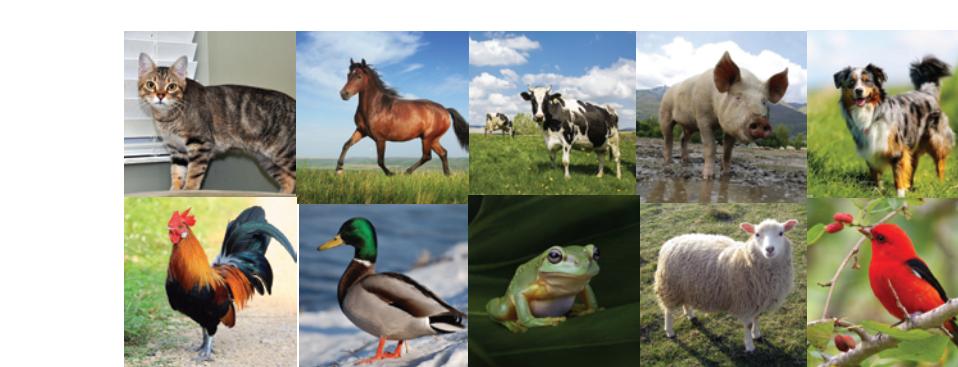
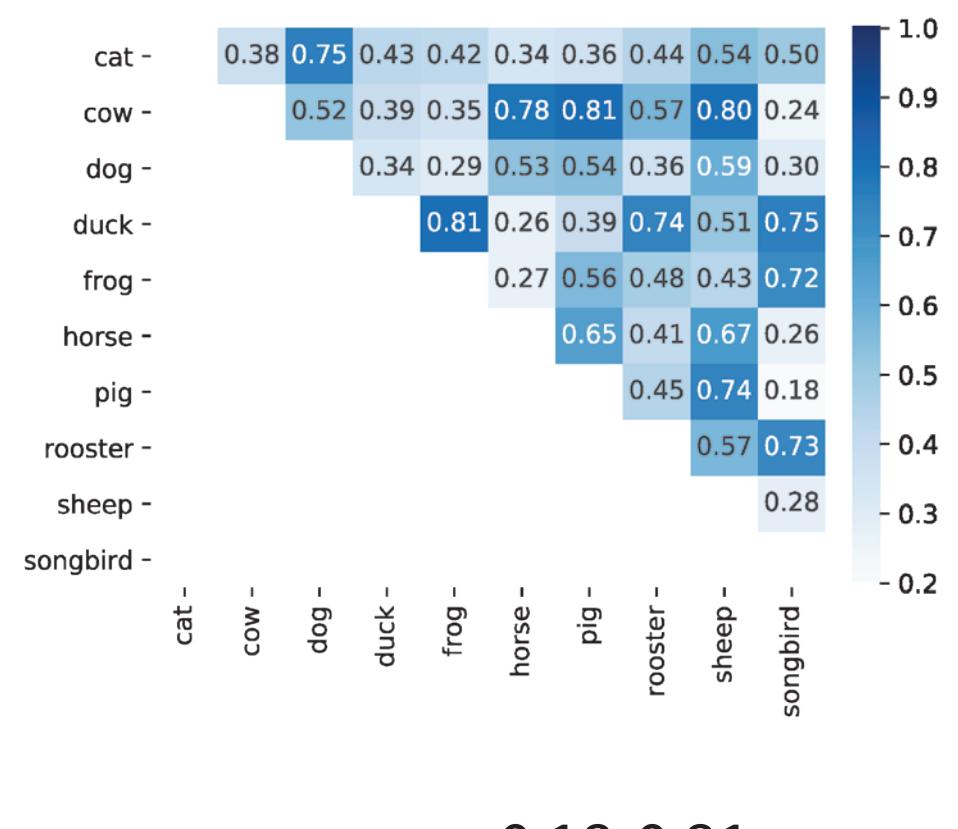
1. Collect judgements for trios of stimuli
 2. For a given prompt & option pair, calculate the likelihood that option will be picked
 3. Average over trial, participant, prompt modality
-
- x 20 repeats of each trio

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

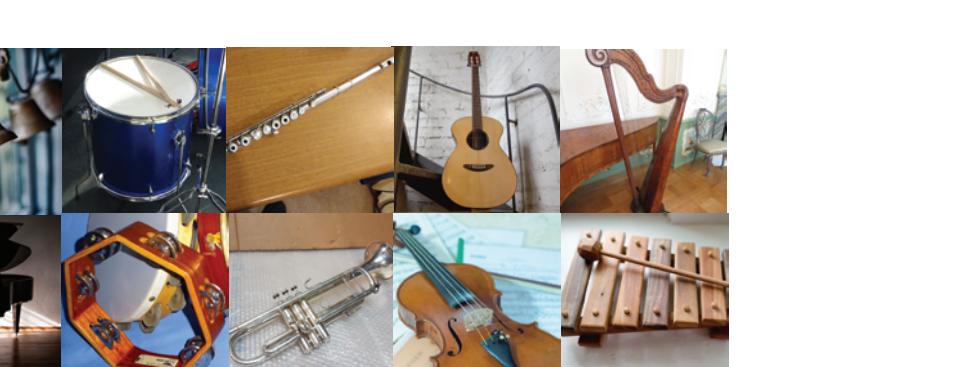
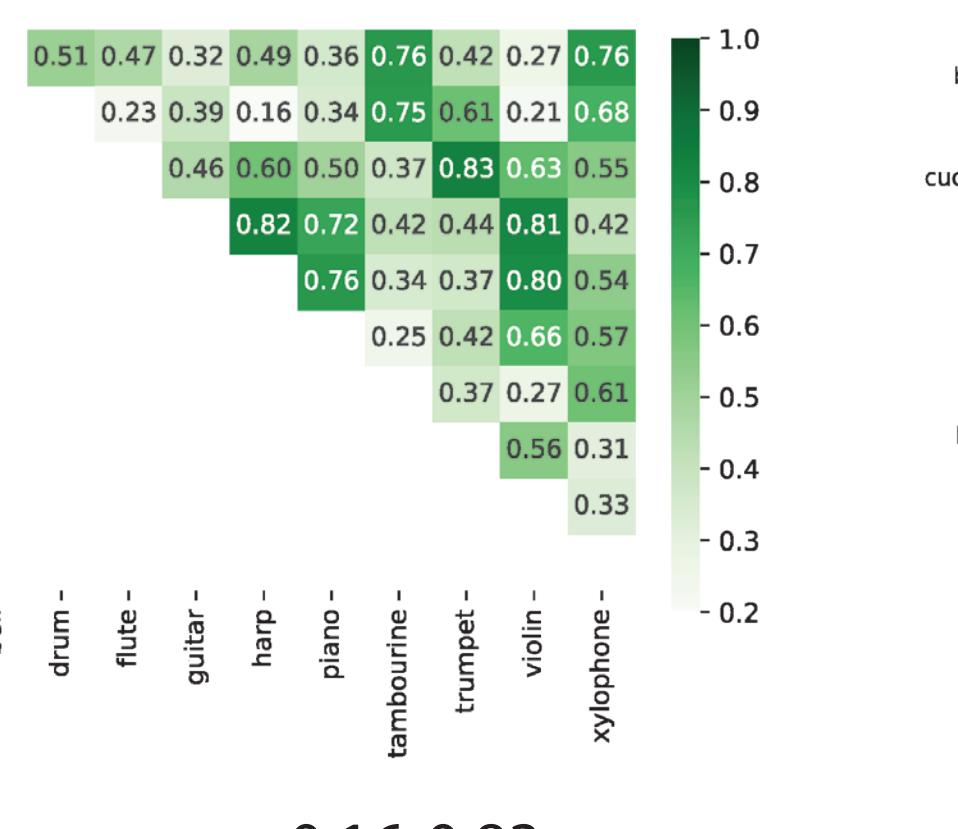


Sight-Sound Semantics Database

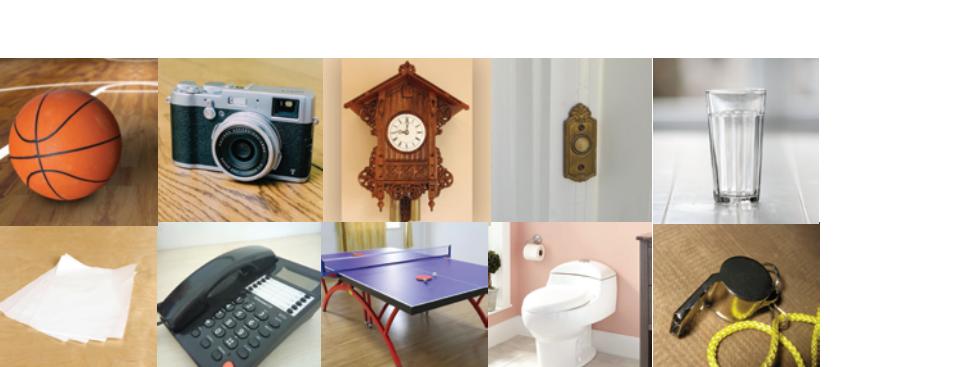
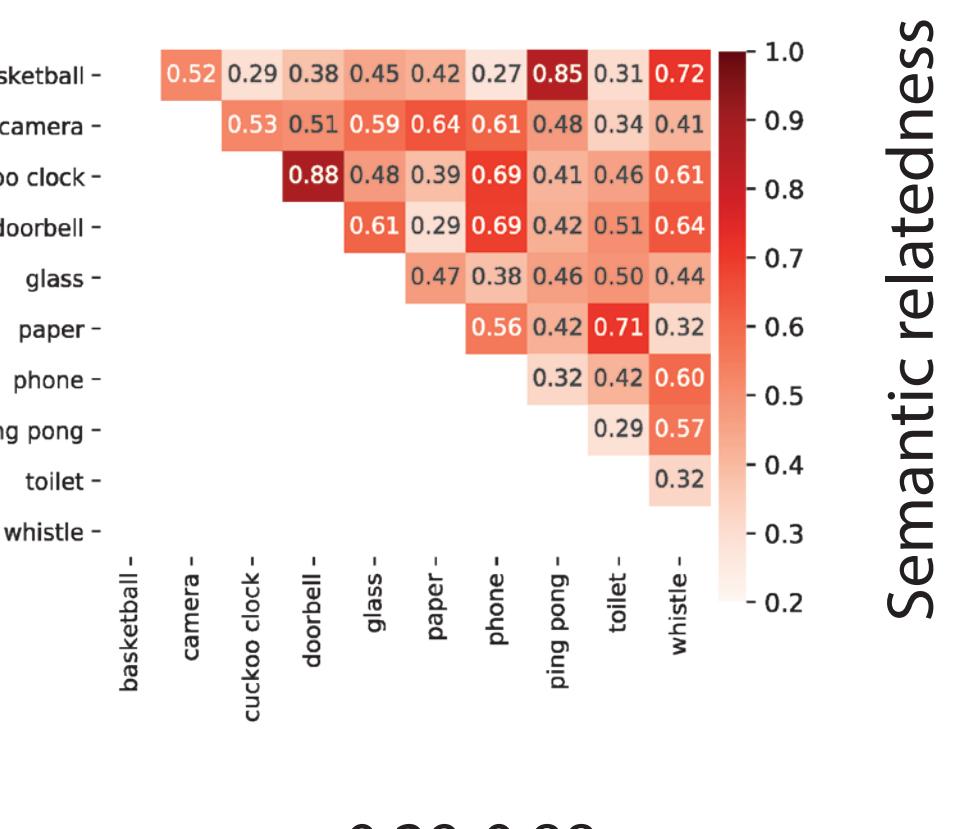
ANIMALS



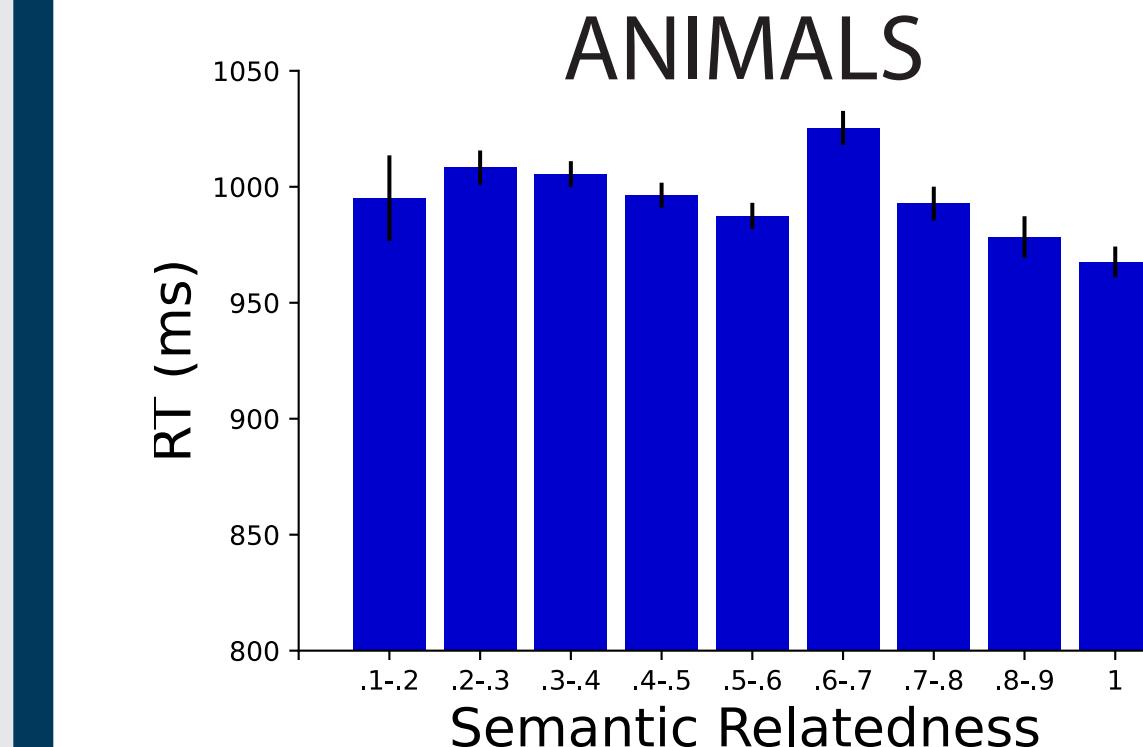
INSTRUMENTS



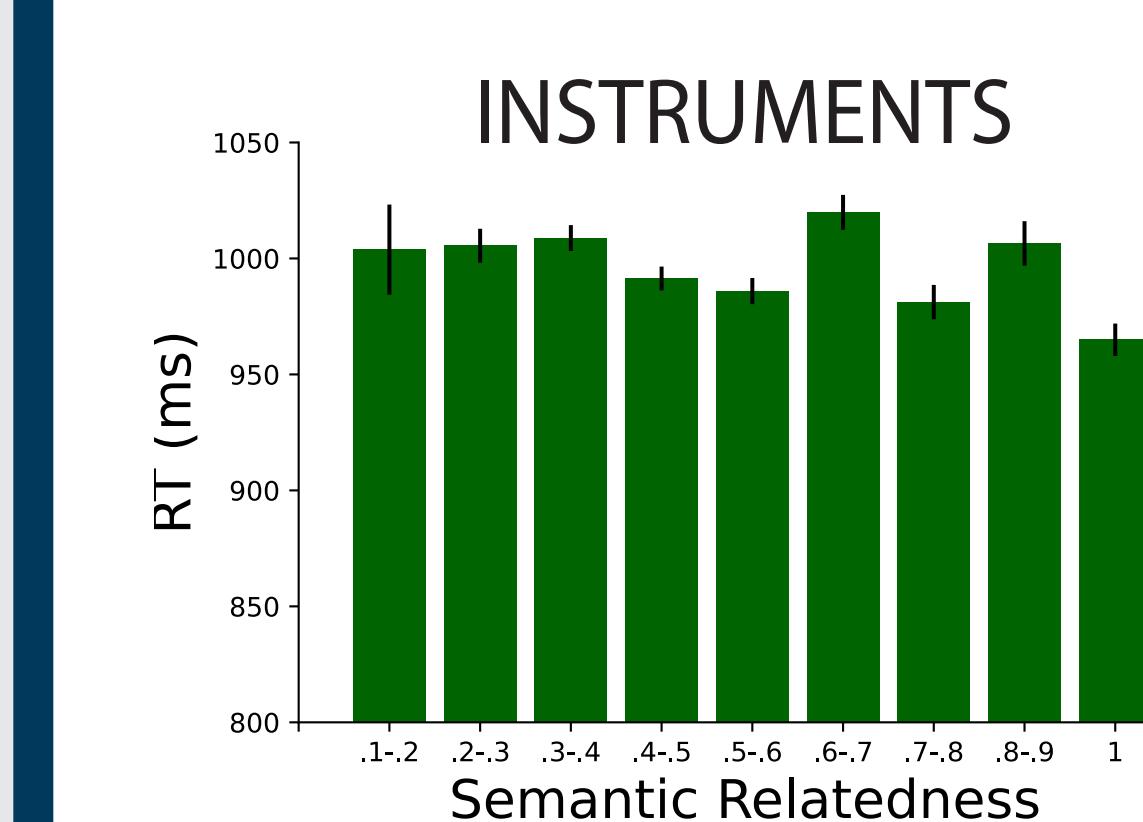
HOUSEHOLD ITEMS



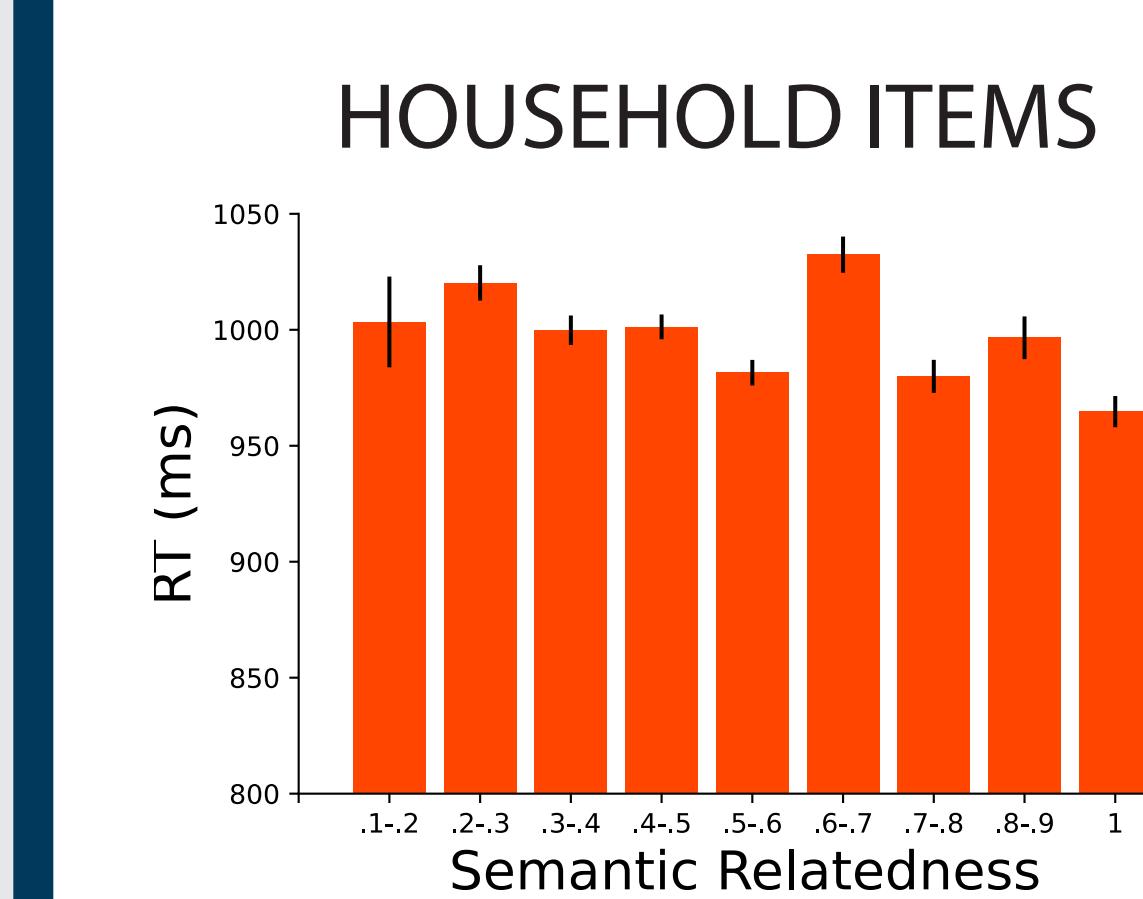
Is the search benefit category specific?



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



Nope! Each category shows a similar pattern



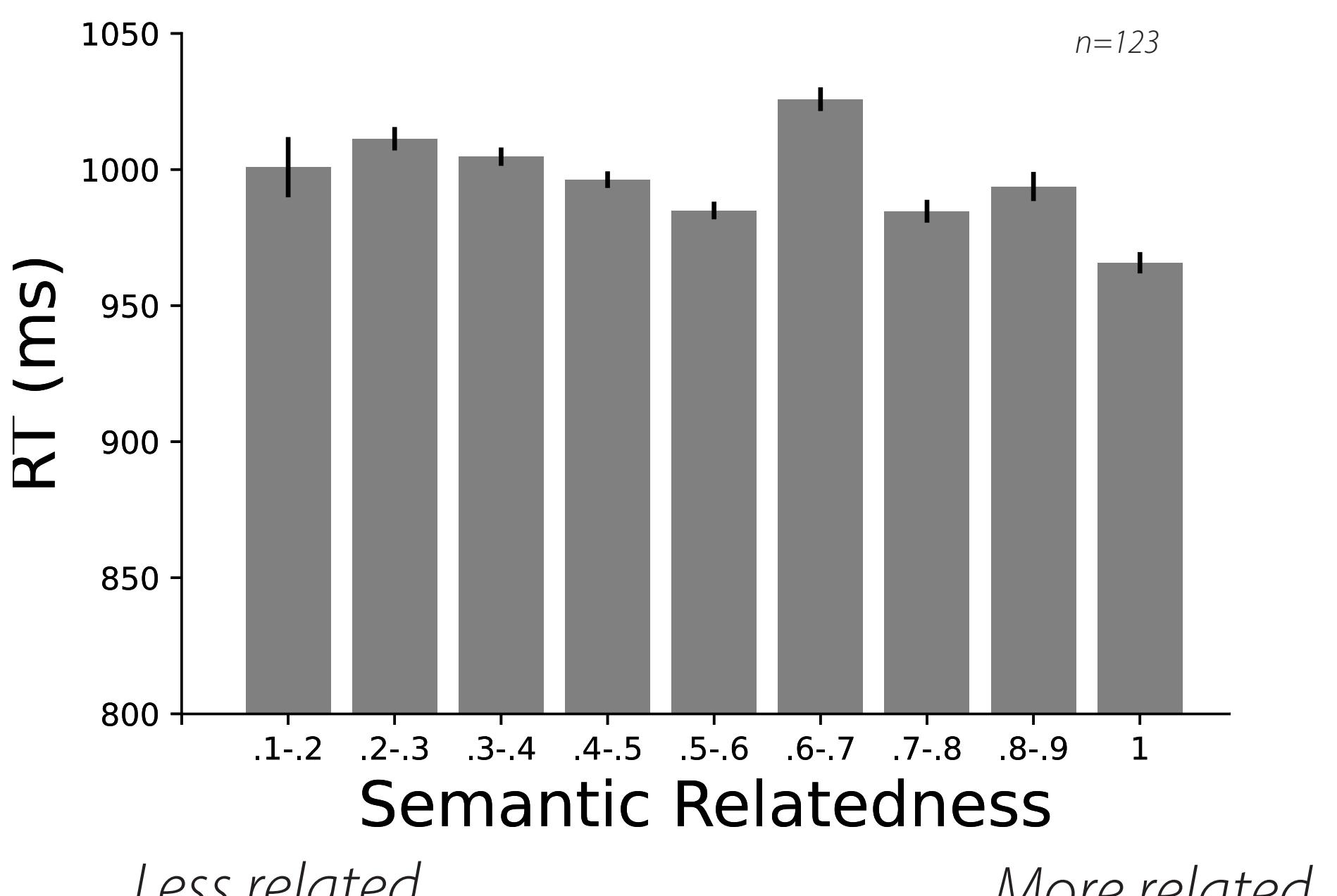
1 way within subject rmANOVA
- Animals: p<0.0001
- Instruments: p<0.0001
- Household: p<0.0001

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

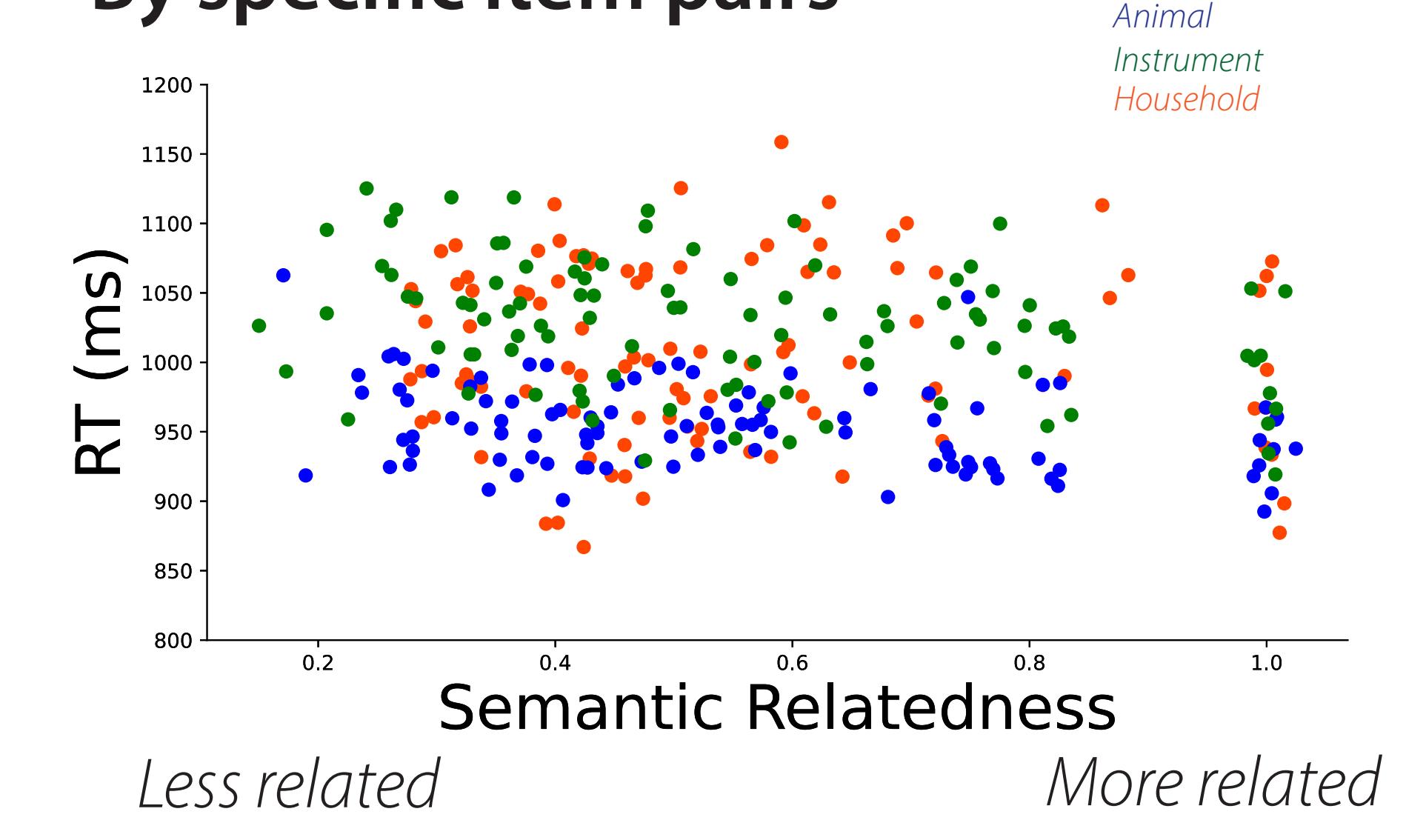
Category did not predict RT

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

$$RT \sim \text{relatedness} + (1|\text{participant}) + (1|\text{stimulus})$$

Relatedness significantly predicted RT (p<0.0001)

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 mechanism?

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.