

# The impact of semantic descriptions on learning object-to-object relationships in a scene

Poster Presentation 43.443: Monday, May 20, 2024, 8:30 am – 12:30 pm, Pavilion  
Session: [Scene Perception: Categorization](#)

[Schedule of Events](#) | [Search Abstracts](#) | [Symposia](#) | [Talk Sessions](#) | [Poster Sessions](#)

Victoria Nicholls<sup>1</sup> ([nicholls@psych.uni-frankfurt.de](mailto:nicholls@psych.uni-frankfurt.de)), Kim Kira Philippi<sup>1</sup>, Mohamed Abdelrahman<sup>1</sup>, Jo Ackerman<sup>1</sup>, Franka Schulze<sup>1</sup>, Melissa Le-Hoa Vo<sup>1</sup>; <sup>1</sup>Goethe University Frankfurt

Our knowledge of scenes is thought to have a hierarchical structure, at the lowest level are local objects, often smaller objects such as soap. Followed by anchor objects, often larger objects, e.g. sinks. Local and anchor objects together e.g. soap on a sink, form a phrase. Phrases can have multiple local objects (co-locals). Multiple phrases combined form a scene. What is not clear is how we learn this structure, can this be learned with visual associations alone or is semantic object information required? To examine this, we performed two experiments. In the learning phase of the first experiment participants were presented with objects in isolation with audio descriptions of the object functions, or with non-descriptive audio. This was followed by two recall phases, the first where participants were presented with two objects which they rated on a scale from 1-9 how likely the objects would be grouped together based on the descriptions they received in the learning phase. In the second recall phase participants were shown a scene image containing all the objects and participants grouped the objects into phrases based on the object descriptions received. In the learning phase of the second experiment participants viewed videos of phrases in scenes, where each object was highlighted, along with descriptive or non-descriptive audio. This was followed by the same rating recall phase as in the first experiment. In the video conditions we found that participants learned the anchor-local relationships even with non-descriptive audio, while descriptive audio boosted learning the local- to-local relationships. This suggests that hierarchical scene knowledge can be learned through visual associations but the detail of the knowledge can be improved with the inclusion of semantic information such as descriptions of functions the objects perform together.

Acknowledgements: This work was supported by the Deutsche Forschungsgemeinschaft (DFG-FOR German Research Foundation;Arena)