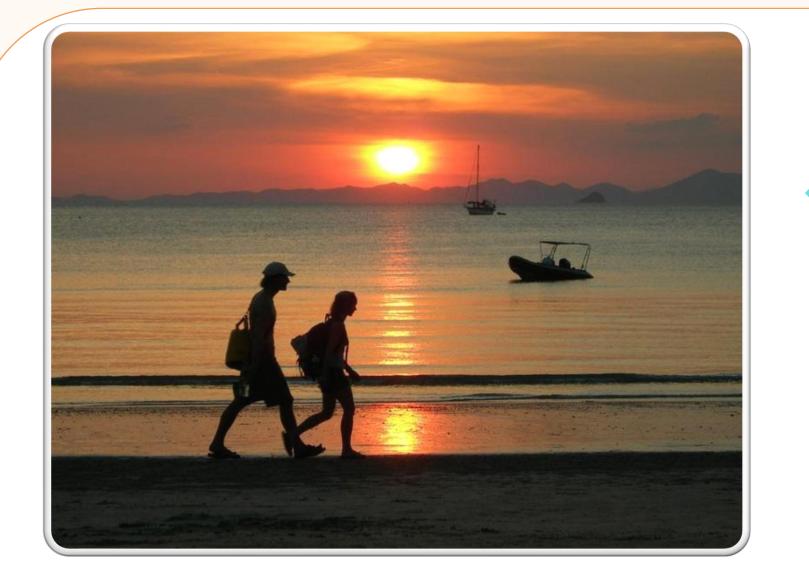
Semantic memories: Image memorability transfers to scene descriptions

Diana Kollenda*, Sophie Schwenger, Sophie Halstenberg & Ben de Haas

*Diana.Kollenda@gmail.com | INDIVISUAL lab: https://www.individual-perception.com/



"Two people walk along the beach while the sun sets in bright orange on the horizon over the sea."

LOW

INTRODUCTION

high

The intrinsic image property of Image Memorability ... (cf. Bainbridge, 2019)

- is defined as the likelihood that a given individual will later recall a particular stimulus.
- can be predicted using the residual neural network model *ResMem* (Needell & Bainbrige. 2023).
- may be influenced by semantic features (e.g., categories like animals or food) and visual attributes (e.g., color, patterns; cf. Kramer et al., 2023; Lin et al., 2021),



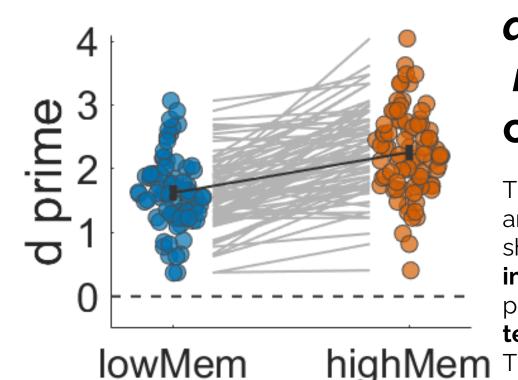
"An adult throws a child into the air, while the silhouettes of both can be seen against the sunset."

EXPERIMENTS

All experiments were conducted online and followed the same general procedure:

- 1. Encoding phase: "Is this an indoor or outdoor scene?"
- Participants viewed 40 scenes each with high-(M = .91, SD = .02) and low- (M = .63, SD = .05)
 ResMem memorability scores or their corresponding descriptions.
- Participants with accuracies below M-2SD were excluded from further analysis.
- 2. Decoding phase: "Do you remember this?"
- Participants responded using a four-point scale: yes, rather yes, rather no, no.
- The task included 80 old and 80 new scenes/descriptions, matched for image memorability scores.
- 3. Memory anchor task: "What helped you decide?"
- Free text responses (Exp. 1); single-word selection (Exp. 2a) or multiple-word selection (Exp. 2b)
- Responses were categorized as reflecting either semantic, visual, mixed (cf. Kramer et al., 2023) or unknown dimensions.

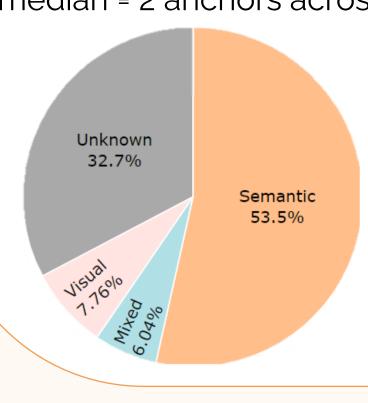
Experiment 1: Scene images & free text responses



d': t(80) = 11.3; *r* <.001, d = 1.25; criterion: n.s.

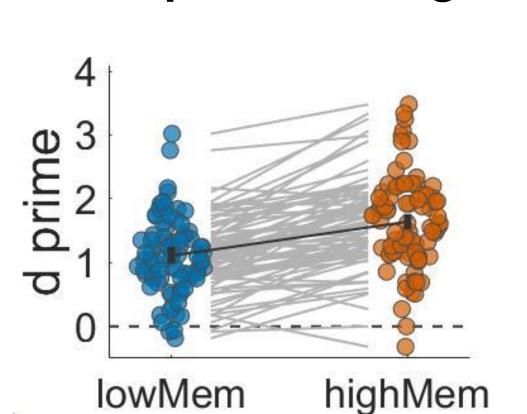
Trial-wise GLME models (with image and participant ID as random effects) showed independent effects of image ResMem score (t = 10.6), presence of humans (t = 5.4), and text (t = 2.2), but no effect of animals. The most parsimonious model included only ResMem scores as a fixed effect

Anchor analysis: We excluded self-referenced descriptions; median = 2 anchors across all trials and participants



Top 5 selected dimensions
Animal/animals (8.8%)
Body/body parts (7.2%)
Sports/sport (6.5%)
Baby/children (6.1%)
Text/signs (4.5%)

Experiment 2a: Human scene descriptions & single-word selection

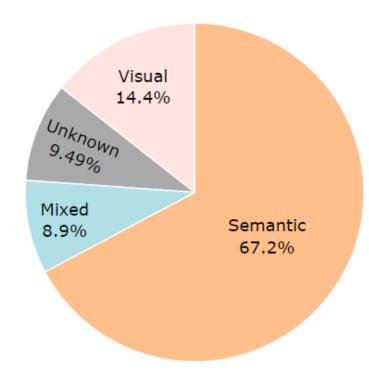


d': t(80) = 10.8; r <.001, d = 1.2; criterion: n.s.

RESULTS In all experiments

Tial-wise GLME models revealed independent effects of **ResMem score** (**t** = **7.9**), **human** references (**t** = 4.0), and **description brevity** (**t** = **-3.3**), with no significant effects for animals, color, or text. The most parsimonious model included only ResMem score.

Anchor analysis:



Top 5 selected dimensions

Text/signs (10.3%)

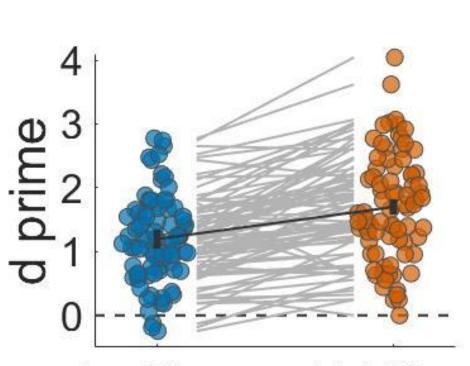
Furniture/bland to colorful (8.8%)

Sports/sport (8.7%)

Animal/animals (8%)

Red/color (5.3%)

Experiment 2b: Chat-GPT scene descriptions & multiple-word selection



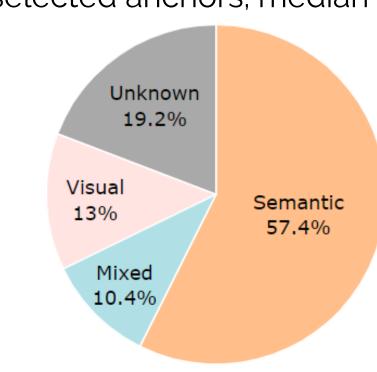
wMem highMem

d': t(80) = 8.67; r <.001, d = 0.96;

criterion: n.s.

Trial-wise GLME models showed independent effects of **ResMem** score (t = 3.3), human references (t = 7.1), brevity (t = -5.8), and animal references (t = 2.1), with no effects for color or text. The best-fitting model included only ResMem and

Anchor analysis: We excluded 13 participants with > Mean+3*SD selected anchors; median = 3 anchors across all trials and participants



Top 5 selected dimensions

<u>human references.</u>

Sports/sport (8.8 %)
Technology/electronic (6.3 %)

Body/body parts (6 %)

Red/color (5.4 %)
Text/signs (5 %)

DISCUSSION & CONCLUSION

Closed-loop AI: We could produce memorable scenes from memorable descriptions, ResMem predictions for original images and Gemini-Imagen 3 reproductions: *r* = .55, *p* <.001, and Chat-GPT-DALL.E, *r* = .58, *p* <.001







- > Scene descriptions capture the memorability of an image
- > Selected memory anchors were mostly semantic
- ➤ BUT: Is memorability truly modality-independent and driven by semantics over visual features? Testing this in congenitally blind or aphantasic individuals would be especially informative. If you work with these groups, we'd love to connect—please get in touch!





European Research Council

