Do Large Language Models Agree with Humans on Color Harmony?

Laboratory name: EconophysiX Lab
Internship location: EconophysiX Lab, CFM (Paris)
Supervision: Ortensia Forni, Alexandre Darmon, Michael Benzaquen
Email: michael.benzaquen@polytechnique.edu

Motivation and Background

Recent evidence suggests that large language models (LLMs) can recover nontrivial aspects of human perception, including color, and can predict human similarity judgments across modalities [1]. In parallel, quantitative models and cross-linguistic analyses show that color naming reflects both perceptual constraints and communicative pressures [2, 3]. This internship will test whether LLMs can also reproduce human judgments of *pleasant color pairing* and *palette completion*, using a harmonious pairing framework we developed in [4] as primary target. The work will also connect to recent evaluations of LLMs on color word associations [5] and to emerging LLM-based approaches for multimodal color recommendation [6].

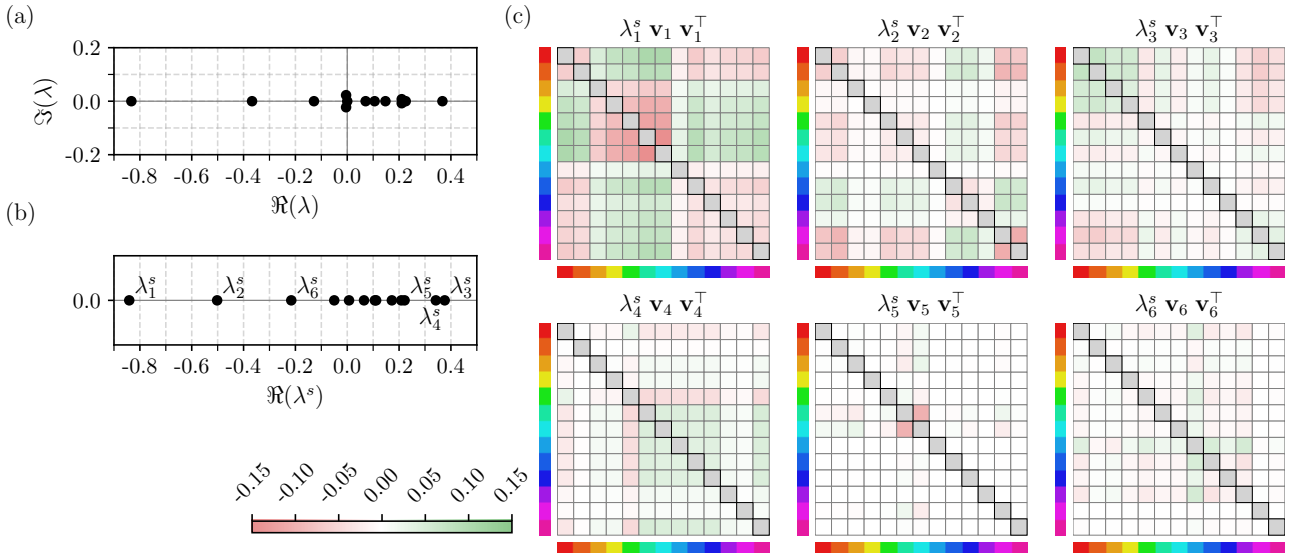

Figure 1: Figure from [4], summarizing the results of a color combination preference survey on humans. (a) Eigenvalues of the color pairing preference score matrix S on the complex plane. (b) Eigenvalues of the symmetrized score matrix $\mathrm{S}^s = \frac{1}{2}(S + S^\top)$. (c) Outer products $\lambda_i^s v_i v_i^\top$ for the six eigenvalues λ_i^s with the highest absolute values, where v_i denotes the eigenvector associated with the i^{th} eigenvalue.

Objectives and Methodology

We will start by defining a small set of controlled tasks around pleasant color pairing and palette completion, together with a clean evaluation protocol and scoring criteria based on human data [4]. We will then benchmark several LLMs using text-based color encodings (HEX, HSL, or CIELAB) and, when available, multimodal color patches, and quantify agreement with human judgments [1, 5]. Finally, we will analyze where agreement succeeds or fails by relating model outputs to perceptual structure and to cross-linguistic or usage-driven regularities reported in the literature [2, 3].

Expected Outcomes and Impact

We expect to obtain quantitative evidence on whether LLMs can reproduce human judgments of pleasant color pairing and, more generally, whether they capture the structure of human color preferences beyond color naming and semantic associations. A key outcome will be a clear characterization of the regimes where LLM predictions are reliable (e.g., for specific hue regions or pairing types) and where systematic deviations occur. The broader impact is twofold. First, from a cognitive-science perspective, the project will contribute to the ongoing debate on what aspects of human perception and preference can be inferred from language statistics alone, extending recent findings on sensory judgments to the domain of aesthetic evaluation [1]. Second, from an applied perspective, the results will inform principled, controllable color recommendation strategies

for palette completion and design assistance, complementing recent LLM-based multimodal recommendation approaches [6].

References

[1] Raja Marjieh, Ilia Sucholutsky, Pol van Rijn, Nori Jacoby, and Thomas L Griffiths. Large language models predict human sensory judgments across six modalities. *Scientific Reports*, 14(1):21445, 2024.

[2] Andrea Baronchelli, Tao Gong, Andrea Puglisi, and Vittorio Loreto. Modeling the emergence of universality in color naming patterns. *Proceedings of the National Academy of Sciences*, 107(6):2403–2407, 2010.

[3] Edward Gibson, Richard Futrell, Julian Jara-Ettinger, Kyle Mahowald, Leon Bergen, Sivalogeswaran Ratnasingam, Mitchell Gibson, Steven T Piantadosi, and Bevil R Conway. Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, 114(40):10785–10790, 2017.

[4] Ortensia Forni, Alexandre Darmon, and Michael Benzaquen. Harmonious color pairings: Insights from human preference and natural hue statistics. *arXiv preprint arXiv:2508.15777*, 2025.

[5] Makoto Fukushima, Shusuke Eshita, and Hiroshige Fukuhara. Advancements and limitations of llms in replicating human color-word associations. *Discover Artificial Intelligence*, 5(1):64, 2025.

[6] Ding Xia, Naoto Inoue, Qianru Qiu, and Kotaro Kikuchi. Colorgpt: Leveraging large language models for multimodal color recommendation. 2025.