Decoding Predictive Inference in Visual Language Processing via Spatiotemporal Neural Coherence

Sean C. Borneman
Department of Physics
Carnegie-Mellon University
Pittsburgh, PA 15213
sbornema@andrew.cmu.edu

Julia Krebs
Department of Linguistics
University of Salzburg
Salzburg, Austria 5020
julia.krebs@plus.ac.at

Ronnie Wilbur
Department of Linguistics
Purdue University
West Lafayette, IN 47904
wilbur@purdue.edu

Evie A. Malaia
Department of Communicative Disorders
University of Alabama
Tuscaloosa, AL 35406
eamalaia@ua.edu

Abstract

Human language processing relies on the brain's capacity for predictive inference. We present a machine learning framework for decoding neural (EEG) responses to dynamic visual language stimuli in Deaf signers. Using coherence between neural signals and optical flow-derived motion features, we construct spatiotemporal representations of predictive neural dynamics. Through entropy-based feature selection, we identify frequency-specific neural signatures that differentiate interpretable linguistic input from linguistically disrupted (time-reversed) stimuli. Our results reveal distributed left-hemispheric and frontal low-frequency coherence as key features in language comprehension, with experience-dependent neural signatures correlating with age. This work demonstrates a novel multimodal approach for probing experience-driven generative models of perception in the brain.

1 Introduction

The brain implements hierarchical generative models that minimize prediction errors by aligning top-down predictions with sensory evidence [1, 2]. According to predictive coding theory, neural networks maintain internal models of environmental structure, continuously updating predictions to minimize free energy—an information-theoretic measure of surprise [3]. While predictive coding is established in auditory speech comprehension [4], its mechanisms in visual language remain less understood.

Sign languages offer a compelling testbed for visual predictive coding because they involve rich spatiotemporal dynamics requiring real-time hierarchical inference. Signed stimuli exhibit greater Shannon entropy and information density than non-linguistic actions [5, 6], reflecting structured linguistic content embedded in continuous motion. Deaf signers demonstrate enhanced sensitivity to dynamic visual patterns underlying comprehension [7], engaging language-network regions when viewing signs, whereas non-signers rely predominantly on visual processing areas [8]. This experience-dependent neural specialization suggests that lifelong exposure shapes internal generative models for visual language.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Foundation Models for the Brain and Body.

Problem Formulation: We formalize sign language comprehension as hierarchical temporal inference under uncertainty. A fluent signer's brain maintains multi-scale generative models predicting upcoming visual input based on internalized linguistic structure. Successful comprehension corresponds to minimizing prediction error across hierarchical levels: high-level expectations (syntactic/semantic transitions) constrain lower-level perceptual predictions (kinematic features), consistent with theories of hierarchical predictive coding.

Approach: We introduce a coherence-based multimodal fusion approach combining EEG and computer vision to probe hierarchical predictive coding in sign language. We recorded EEG from Deaf signers viewing normal versus time-reversed sign language videos (preserving motion energy but destroying temporal linguistic structure), extracted optical flow motion features, and computed frequency-resolved coherence between neural and visual signals. Using machine learning with entropy-based feature selection, we decode brain states and identify neural mechanisms underlying structured visual language processing.

Contribution: Our approach demonstrates that: (1) neural coherence features discriminate structured vs. unstructured visual language inputs, revealing frequency-specific oscillatory mechanisms consistent with predictive coding; (2) language experience modulates these dynamics, with greater exposure correlating with stronger low-frequency neural entrainment; (3) multimodal fusion enables reverse-engineering of hierarchical internal models from neural-stimulus coupling patterns.

1.1 Related Work

Predictive coding frames perception as Bayesian inference with hierarchical generative models [9]. Brain oscillations track temporal structure in sensory input [10], with low-frequency dynamics carrying high-level predictions about structured events. Recent studies demonstrate cortical tracking of biological motion: Shen et al. showed MEG signals entrain to hierarchical kinematic patterns with stronger low-frequency coherence for structured movements [11]. In sign language, low-frequency neural entrainment is critical for comprehension [7], similar to delta-band oscillations in speech parsing.

Multimodal fusion approaches combining neural data with stimulus features have proven effective for decoding cognitive states [12]. Our work extends this by fusing EEG with optical flow to quantify neural tracking of complex visual motion, using entropy-based methods to characterize linguistic complexity [6, 13]. This represents the first application of EEG-optical flow coherence to sign language comprehension with age-dependent analysis.

2 Methods

2.1 Participants & Stimuli

24 Deaf native signers of Austrian Sign Language (ÖGS), ages 20-60s (M=42, SD=12.27), with normal vision and ÖGS as primary language. All participants acquired ÖGS from birth, ensuring age serves as a valid proxy for cumulative language exposure in this population. Study was approved by local ethics board with informed consent, with 30 Euro/hour remuneration for participants. Prior to data collection, participants received comprehensive instructions in Austrian Sign Language (ÖGS) with German translation available. Instructions specified the experimental task: watch sign language videos and evaluate sentence acceptability using mouse responses. A practice session familiarized participants with procedures before the main experiment comprising ten 5-minute blocks with self-paced breaks between blocks. Experimenters remained present throughout to address questions during breaks.

40 signed sentences in ÖGS plus time-reversed versions created by temporal reversal. This manipulation preserves local motion statistics but destroys linguistic temporal predictability, creating visually matched but linguistically meaningless controls analogous to reversed speech. Participants performed attention judgments confirming structured vs. unstructured stimulus interpretability.

2.2 EEG Recording & Preprocessing

26 electrodes (10/20 layout), 500Hz sampling, 1-100Hz band-pass filtering. Signals preprocessed with artifact removal, mastoid re-referencing, and epochs time-locked to video presentations. This provided robust neural activity measures across widespread cortical regions.

2.3 Optical Flow & Neural Coherence Analysis

Motion Feature Extraction: Horn-Schunck algorithm computed motion vectors between consecutive video frames [14]:

$$\mathbf{u}(x,y,t) = \arg\min_{\mathbf{u}} \left[\|\nabla I \cdot \mathbf{u} + I_t\|^2 + \alpha^2\|\nabla \mathbf{u}\|^2\right]$$

where $I(x,y,t)$ is image intensity, $\mathbf{u} = (u_x, u_y)$ are velocity components, and α controls smoothness. We derived univariate motion signals by spatially aggregating flow magnitudes at 30Hz:

$$M(t) = \frac{1}{N}\sum_{x,y}\|\mathbf{u}(x,y,t)\|$$

capturing overall motion dynamics while preserving temporal linguistic structure.

EEG-Motion Coherence: Prior to coherence computation, EEG and optical flow time-series were aligned and resampled to common 30Hz rate. Coherence at frequency f was computed as:

$$C_{xy}(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)}$$

where $S_{xy}(f)$ is the cross-power spectral density between EEG channel x and motion signal y, and $S_{xx}(f)$, $S_{yy}(f)$ are auto-power spectral densities. We focused on 0.5-12Hz range encompassing dominant temporal modulations in sign language.

Electrodes were grouped into four regional clusters (Left/Right × Anterior/Posterior) motivated by known language lateralization differences [15]. For each region and frequency bin, we obtained coherence magnitude and optimal temporal lag, resulting in spatiotemporal coherence profiles characterizing neural-stimulus coupling.

2.4 Feature Selection & Machine Learning

The full coherence feature set yielded 496 features (magnitude/timing × 62 frequency bins × 4 regions) for 24 participants. We employed unsupervised feature selection based on information-theoretic criteria, Shannon entropy (1) and mutual information (2):

$$H(X) = -\sum_{i} p(x_i)\log p(x_i) \tag{1}$$

$$I(X;Y) = H(X) - H(X|Y) \tag{2}$$

Features maximizing mutual information $I(X;Y)$ with target variables were selected, ensuring interpretable dimensionality reduction while preserving predictive signal for both age-related changes and stimulus-specific neural entrainment.

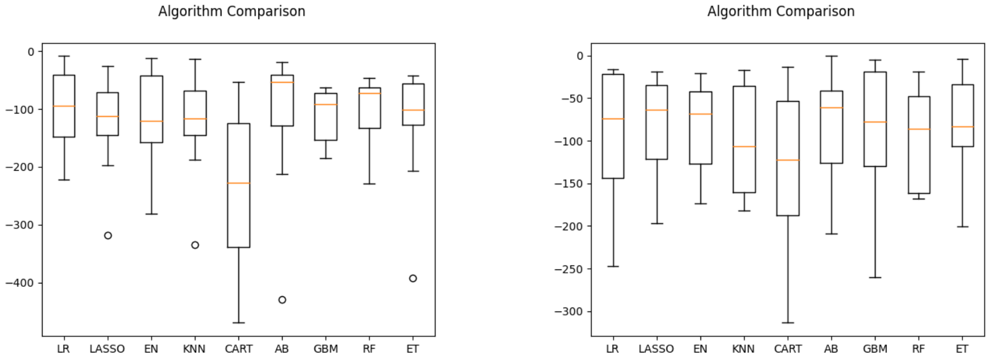
Modeling Pipelines: Two complementary approaches: (A) Age-targeted regression identifying features predictive of experience, comparing performance across algorithms (Linear Regression, LASSO, Elastic Net, kNN, Random Forest, Gradient Boosting); (B) Stimulus-driven classification identifying features discriminating structured vs. unstructured input, refined via recursive feature elimination for age correlation analysis. Cross-validation ensured robust performance estimates.

Note that feature selection in pipeline B was optimized for stimulus discrimination (structured vs. unstructured), not age prediction. This approach avoids circularity in subsequent age-correlation analyses.

3 Results

3.1 Age Prediction from Neural Coherence

We predicted participant age from EEG-optical flow coherence features using multiple regression algorithms. Model performance quantified by mean squared error: $\mathrm{MSE}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2$.



(a) Full feature set.

(b) Top 10 entropy-selected features.

Figure 1: Age prediction performance. Boxplots show cross-validated MSE distributions (lower = better). Feature selection improved consistency for tree-based models while maintaining performance for regularized approaches. Baseline (mean age prediction): MSE = 151 years^2; best models achieved 85-100 years^2 (RMSE 9-10 years).

With full features, best models achieved MSE 85-100 years^2 (RMSE 9-10 years), substantially outperforming baseline mean-age prediction (MSE = 151 years^2). LASSO, Elastic Net, and Gradient Boosting performed best. Using top 10 entropy-selected features maintained or improved performance, indicating most predictive signal captured by key features. Low-frequency timeshift features in frontal/posterior regions and high-frequency correlations in lateral areas emerged as strongest age predictors, suggesting experience modulates predictive neural timing across multiple temporal scales.

3.2 Stimulus Classification Performance

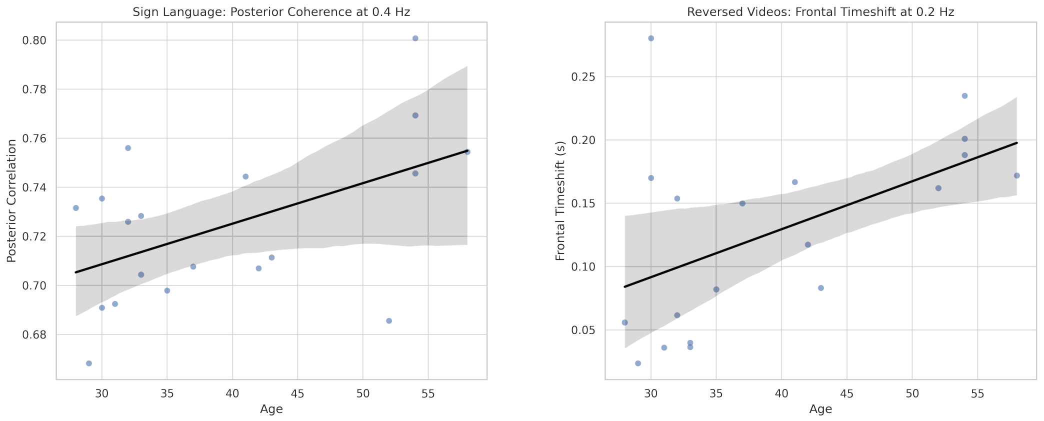
Neural coherence features achieved near-perfect classification of structured vs. unstructured input (cross-validated accuracy = 94.2 ± 3.1%), demonstrating categorically distinct brain states during linguistic vs. non-linguistic visual processing. This supports predictive coding theory: structured input conforming to internal models elicits strong prediction signals, while violations result in breakdown of neural-stimulus alignment.

3.3 Experience-Dependent Neural Signatures

Two specific coherence features showed significant age correlations, as depicted in Figure 2. In structured sign language, posterior coherence at ~ 0.4 Hz increased with age ($r \approx 0.51, p < 0.03$), suggesting older signers exhibit stronger neural synchronization over 2–3 s intervals during linguistic processing. In reversed videos, frontal temporal lag at 0.2 Hz correlated with age ($r \approx 0.52, p < 0.03$), indicating older participants showed longer neural delays to unpredictable motion. These patterns indicate experience refines internal models: older signers show enhanced prediction over longer timescales for structured input but increased processing costs when predictions are violated.

4 Discussion

Our findings demonstrate how experience tunes hierarchical generative models in visual language processing. The coherence-based approach reveals that Deaf signers' brains exhibit categorically different states when viewing structured versus unstructured visual language, with measurable neural coupling at timescales of linguistic units. This aligns with predictive coding theory: structured input



(a) Posterior coherence (0.4Hz) during structured signing increases with age.

(b) Frontal delays (0.2Hz) during reversed input increase with age.

Figure 2: Age-related changes in predictive dynamics. (a) Enhanced low-frequency coherence suggests improved higher-order predictions with experience. (b) Increased delays to unstructured input reflect greater reliance on learned generative models that incur larger prediction error costs when violated.

conforming to internal models elicits strong prediction signals, while violations result in breakdown of neural-stimulus alignment.

Hierarchical Inference Mechanisms: Age-related patterns suggest experience optimizes internal representations across temporal scales. Younger signers may rely primarily on local motion cues and short-range predictions, while older signers demonstrate neural evidence of integration over longer timescales corresponding to syntactic/pragmatic levels. The positive correlation between age and 0.4Hz coherence indicates experienced signers synchronize with multi-sign units, consistent with hierarchical inference where higher-level predictions involve slower neural dynamics [16].

Computational Interpretation: The trade-off between enhanced structured processing and increased costs for random input reflects optimization for typical linguistic environments. From a computational perspective, this supports viewing the brain as an adaptive generative model that improves inference under familiar conditions at the expense of flexibility with anomalous inputs—a hallmark of experience-dependent specialization.

Free Energy Minimization: Our results provide empirical evidence for free energy minimization in visual language. The coherence differences between structured and unstructured conditions can be interpreted as neural signatures of prediction error: reduced coherence to reversed stimuli reflects increased sensory surprise when internal models fail to predict input structure. The age-dependent increase in this effect suggests refined models generate stronger predictions (and thus larger prediction errors when violated).

Methodological Innovation: Our entropy-based feature selection identified interpretable neural signatures linked to specific frequency bands and brain regions. The multimodal fusion approach demonstrates how combining neural and stimulus-derived features can reveal meaningful brain-behavior relationships while maintaining interpretability through careful dimensionality reduction. This provides a generalizable framework for studying experience-driven neural adaptation in complex natural tasks.

5 Limitations & Future Work

Key limitations include small sample size (N=24) requiring replication in larger cohorts, and population specificity (Austrian Sign Language users) necessitating cross-linguistic validation. While age serves as a valid proxy for cumulative language exposure in native signers who acquired ÖGS from birth, future studies could incorporate direct measures of language proficiency for finer-grained modeling. EEG provides excellent temporal but limited spatial resolution; future work combining with fMRI/MEG could improve anatomical precision of predictive coding localization. Additional controls beyond time-reversal (e.g., spatially scrambled or semantically violated signs) would strengthen claims about linguistic specificity. The current approach focuses on motion-based features; incorporating hand shape and spatial configuration could reveal complementary aspects of sign language neural processing.

6 Broader Impact

This work has implications for energy-efficient AI systems by demonstrating predictive strategies that treat predictable inputs as requiring minimal processing. Computer vision and sign language recognition systems could incorporate multi-timescale predictive mechanisms, focusing computational resources on unexpected movements while auto-completing predictable sequences. The experience-dependent optimization observed here suggests AI systems could similarly adapt their internal models based on domain-specific exposure. Neural coherence patterns could serve as biomarkers for healthy brain aging or cognitive assessment, given the systematic relationship between experience and predictive neural signatures. The approach provides tools for studying neural plasticity and specialization across diverse populations and sensory modalities.

References

[1] Karl Friston. Hierarchical models in the brain. *PLoS computational biology*, 4(11):e1000211., 2008.

[2] Ryszard Auksztulewicz and Karl Friston. Attentional enhancement of auditory mismatch responses: a DCM/MEG study. *Cerebral cortex*, 25(11):4273–4283., 2015.

[3] Karl Friston. Does predictive coding have a future? *Nature neuroscience*, 21(8):1019–1021., 2018.

[4] Anne-Lise Giraud and David Poeppel. Cortical oscillations and speech processing: emerging computational principles and operations. *Nature neuroscience*, 15(4):511–517, 2012.

[5] Evie Malaia, Joshua D. Borneman, and Ronnie B. Wilbur. Assessment of information content in visual signal: Analysis of optical flow fractal complexity. *Visual Cognition*, 24(3):246–251., 2016.

[6] Joshua D. Borneman, Evie A. Malaia, and Ronnie B. Wilbur. Motion characterization using optical flow and fractal complexity. *Journal of Electronic Imaging*, 27(05):1, 2018. doi: 10.1117/1.JEI.27.5.051229.

[7] Evie Malaia, Sean C Borneman, Julia Krebs, and Ronnie B Wilbur. Low-frequency entrainment to visual motion underlies sign language comprehension. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:2456–2463., 2021.

[8] Evie Malaia, Ruwan Ranaweera, Ronnie B Wilbur, and Thomas M Talavage. Event segmentation in a visual language: Neural bases of processing american sign language predicates. *NeuroImage*, 59(4):4094–4101, 2012.

[9] Maxwell JD Ramstead, Casper Hesp, Alexander Tschantz, Ryan Smith, Axel Constant, and Karl Friston. Neural and phenotypic representation under the free-energy principle. *Neuroscience & Biobehavioral Reviews*, 120:109–122, 2021.

[10] Nai Ding, Lucia Melloni, Hang Zhang, Xing Tian, and David Poeppel. Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, 19(1):158–164., 2016. doi: 10.1038/nn.4186.

[11] Li Shen, Xiqian Lu, Xiangyong Yuan, Ruichen Hu, Ying Wang, and Yi Jiang. Cortical encoding of rhythmic kinematic structures in biological motion. *NeuroImage*, page 119893., 2023.

[12] Alain de Cheveigné, Daniel DE Wong, Giovanni M Di Liberto, Jens Hjortkjaer, Malcolm Slaney, and Edmund Lalor. Decoding the auditory brain with canonical component analysis. *NeuroImage*, 172:206–216., 2018.

[13] Saúl Solorio-Fernández, J Ariel Carrasco-Ochoa, and José Fco Martínez-Trinidad. A review of unsupervised feature selection methods. *Artificial Intelligence Review*, 53(2):907–948., 2020.

[14] Berthold K.P. Horn and Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, 17 (1):185–203, 1981. doi: 10.1016/0004-3702(81)90024-2.

[15] Evie Malaia, Katie Ford, Sean Borneman, and Brendan Ames. Salience of low-frequency entrainment to visual signal for classification points to predictive processing in sign language. In Proceedings of 30th Annual Computational Neuroscience meeting: CNS* 2021. *Journal of Computational Neuroscience*, 49(S1):3., 2021.

[16] Alice Blumenthal-Dramé and Evie Malaia. Shared neural and cognitive mechanisms in action and language: The multiscale information transfer framework. *Wiley Interdisciplinary Reviews: Cognitive Science*, 10(2):1484., 2019.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Yes; the abstract and introduction clearly state the study's goals and methods, which are consistently reflected in the results and discussion.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Yes; the discussion section acknowledges limitations in sample size and generalizability.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: This is an empirical study; it does not include theoretical results or formal proofs.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper provides detailed descriptions of the stimuli, EEG recording protocols, preprocessing steps, feature extraction, and analysis pipelines, allowing the experiments to be reproduced with comparable datasets.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The analysis code is described in detail and can be replicated with similar EEG data. However, due to IRB restrictions, the dataset itself cannot be shared because it involves a small, protected population.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The Methods section provides detailed information on data collection, preprocessing, feature extraction, and analysis pipelines, sufficient to understand and replicate the experimental setup.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The paper includes standard evaluation metrics across multiple models to assess performance variation.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: The experiments use small-scale EEG data and standard statistical/machine learning analyses that do not require specialized computing resources.

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research followed all ethical guidelines, including obtaining IRB approval, informed consent, data privacy safeguards, and responsible reporting. No ethical concerns or conflicts of interest are present.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The study highlights the benefits of modeling predictive neural mechanisms in aging and language, which could inform inclusive design in communication and cognition-related tools. It also acknowledges ethical concerns around privacy and profiling in small, identifiable populations, and safeguards data accordingly.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper does not release models or datasets with high misuse potential. It focuses on human EEG data analysis with IRB-constrained access and no general-purpose deployable models.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The paper uses established software packages (e.g., MATLAB Vision Toolbox, NoiseTools) with proper citation and under appropriate institutional licenses. No third-party datasets or models requiring separate licensing were used.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not release any new datasets, models, or code artifacts. Analysis was performed on existing, research-restricted data under IRB constraints.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [Yes]

Justification: The paper involves research with human subjects and includes full instructions given to participants and compensation details in Supplementary Materials.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [Yes]

    Justification: The paper clearly states that all procedures were approved by the Institutional Review Board of the University of Salzburg and conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all participants.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: The core methodology does not involve large language models in any way. LLM use was limited to spellcheck and formatting.