
Observer-Dependent Entropy Retrieval in Linguistic Computation: A Foundational Framework and Benchmarking Methodology

[Evlondo Cooper](#)*

Posted Date: 20 August 2025

doi: 10.20944/preprints202504.0130.v4

Keywords: observer-dependent entropy; linguistic comprehension; cognitive modeling; quantum-inspired computation; entropy retrieval; density matrix; individual differences; garden-path sentences; language processing; attention and working memory



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Observer-Dependent Entropy Retrieval in Linguistic Computation: A Foundational Framework and Benchmarking Methodology

Evlondo Cooper 

Independent Researcher, USA; evlocoo@pm.me

Abstract

Comprehension failure is not prediction error; it is delayed access to retrievable meaning. Unlike prediction-based models, ODER models delayed access to meaning rather than incorrect anticipation. We introduce *Observer-Dependent Entropy Retrieval* (ODER), a formal framework that models linguistic understanding as an observer-specific process shaped by attention, working memory, and prior knowledge. In a controlled corpus written in *Aurian*, a structured test language developed for entropy-based analysis, ODER explains 31% of sentence-trace variance with an average $R^2 = 0.76$, outperforming Bayesian-mixture, fuzzy-logic, and incremental-surprisal baselines by at least 7.6 AIC units. We benchmark ODER on a natural English sentence to compare retrieval dynamics, and then test fixed-parameter performance across twelve diverse real-language stimuli. This evaluation captures lawful divergence patterns that align with known comprehension bottlenecks, without the need for re-fitting. The model yields two falsifiable predictions: (i) spikes in the contextual gradient ∇C during garden-path resolution correlate with P600 amplitude, but only in low-working-memory observers; and (ii) off-diagonal coherence terms μ in the observer density matrix predict priming-interference effects. Although expressed in quantum notation, ODER does *not* posit quantum computation in neural tissue; the density matrix serves as a compact representation of concurrent interpretations whose collapse time τ_{res} aligns with electrophysiological markers. By reframing comprehension as entropy retrieval rather than entropy reduction, ODER explains why identical sentences impose divergent cognitive costs across populations and provides a benchmarkable framework for modeling neurocognitive variability, including retrieval collapse timing, semantic interference, and working-memory constraints, using parameters that align with biologically measurable comprehension dynamics.

Keywords: observer-dependent entropy retrieval; linguistic computation; cognitive modeling; information theory; quantum-inspired semantics; ambiguity resolution; entropy convergence; working memory; language comprehension; neural plausibility; entropy benchmarking; comprehension dynamics

1. Introduction

Reader guide. This paper introduces a new way to think about sentence comprehension, not as predicting the next word, but as retrieving meaning over time. Most current models measure how surprising a word is, assuming all readers process it in the same way. ODER shows that is not the case: comprehension speed and difficulty depend on each reader's attention, memory, and background knowledge. Using a controlled test language, we measure how different kinds of readers recover meaning, and then test the same model on English sentences without changing the parameters. The results match known points where understanding tends to break down and predict when and why this happens for different readers.

Comprehension failure is not prediction error; it is delayed access to retrievable meaning. Traditional entropy approaches quantify linguistic uncertainty without considering observer-specific processing differences. Evidence from neurolinguistics and computational cognition shows that interpretive effort, and therefore uncertainty resolution, depends on an observer's attentional state,

working-memory capacity, and contextual familiarity [15,26]. Earlier entropy-based models conflate linguistic probability with observer cost, obscuring how processing difficulty emerges from observer-bound retrieval delays. ODER extends this literature by

- defining entropy retrieval as a joint function of hierarchical syntactic complexity and information-transfer efficiency;
- mapping these constructs to measurable cognitive signatures in EEG, fMRI, and pupillometry;
- providing a replicable benchmarking framework that reports γ , τ_{char} , and τ_{res} for each observer class.

Crucially, ODER reframes comprehension as entropy retrieval in the observer, not entropy reduction in the signal. This distinction explains not only *what* is complex but also *how* and *when* different observers experience that complexity. **Clarification.** *We emphasize that ODER is not a language model or parser; it is a meta-framework describing how observers retrieve entropy from linguistic input.*

1.1. Contributions

Accordingly, ODER functions as a formal meta-framework that parameterizes observer-specific comprehension, enabling structured comparison across linguistic theories rather than competing with predictive models.

- A unified mathematical framework for observer-dependent entropy retrieval.
- Explicit retrieval (Equation A3) and transition (Equation A4) functions that parameterize attention, working memory, and prior knowledge.
- A contextual-gradient operator that captures reanalysis, for example, garden-path phenomena, in dynamic observer-dependent terms.
- A benchmarking protocol that compares ODER with existing cognitive models and raises stress flags when ∇C flattens or τ_{res} diverges.
- A demonstration that quantum-formalism constructs can model ambiguity and interference without implying literal quantum computation in the brain.

1.2. Relationship to Existing Models

ODER does not compete with current linguistic models solely on predictive accuracy; instead, it addresses a core explanatory gap:

- **Surprisal models** [12,21] quantify unexpectedness but assume a uniform processor, overlooking individual differences in how surprisal is experienced.
- **Resource-rational models** [11,24] acknowledge capacity limits yet often lack explicit reanalysis mechanisms such as the P600 response to garden-path sentences.
- **ACT-R parsing frameworks** [22] simulate incremental working-memory constraints but treat prediction and retrieval as separate stages, leaving coherence effects unexplained.
- **Hierarchical prediction-error accounts** [13] model multi-level expectations but do not specify observer-class parameters that modulate collapse timing.
- **Transformer language models** excel at prediction and generation, yet their weight vectors obscure observer dynamics and reveal little about why or how observers differ in processing.

Rather than replacing these approaches, ODER serves as a meta-framework that clarifies *when* and *why* processing difficulty arises for specific observers.

We define *lawful divergence* as a systematic, interpretable model–trace mismatch (flagged by the `shape_mismatch` criterion and nonzero $\Delta\tau_{\text{res}}$) that aligns with known construction-specific bottlenecks without parameter re-fitting.

1.2.1. The ODER Innovation: A Conceptual Map

Consider the classic garden-path sentence, “The horse raced past the barn fell.” Empirical work shows expert versus novice divergence in processing difficulty [5,8]. Surprisal models predict uniform

difficulty, whereas ODER explains observer-specific divergence by parameterizing retrieval through attention, working memory, and stored knowledge.

1.3. Theoretical Positioning of ODER

Table 1. Theoretical positioning of ODER relative to leading approaches.

Approach	Primary Focus	Treatment of Observer	Key Limitations
Surprisal Models	Input statistics and probability	Uniform processor with idealized capacity	Cannot explain individual differences in processing difficulty
Resource-Rational	Bounded rationality and capacity limits	Variable capacity, uniform mechanisms	Lack explicit reanalysis mechanisms; treat processing as passive
ACT-R Parsing	Procedural memory retrieval	Slot-limited buffer with decay	Prediction and retrieval treated separately; no coherence term
Hierarchical Prediction-Error	Multi-level expectation tracking	Implicit observer; scalar precision weights	No explicit collapse point or observer parameters
Optimal Parsing	Strategy selection	Uniform processor with idealized strategies	Cannot explain observer-specific strategy choices
ODER (this model)	Observer-relative entropy retrieval (not generative modeling)	Parameterized by attention, memory, and knowledge	Designed partial-fit; requires empirical calibration of observer parameters

By casting comprehension as active, observer-relative retrieval, ODER unifies phenomena such as garden-path reanalysis, working-memory constraints, and expertise effects under a single, testable framework.

While ODER is not a neural model, its observer parameters map to biologically interpretable constraints. Attention (α), working-memory capacity (β), and semantic interference (μ) correspond to measurable features of cognitive processing, including reaction-time profiles, ERP components (e.g., N400/P600), and frontal theta dynamics. This alignment supports ODER's testability within neurocognitive systems without invoking mechanistic claims beyond retrieval-level modeling.

1.4. Forward Retrieval Law and Inverse Decoder

$$\frac{dS_{\text{obs}}}{d\tau} = \gamma(\tau)(S_{\text{max}} - S_{\text{obs}}(\tau)) \tanh(\tau/\tau_{\text{char}}) \quad (1)$$

(Forward retrieval law). The hyperbolic tangent ensures symmetric convergence behavior and enables analytic inversion under bounded modular-flow assumptions (see Appendix A). Convergence is expected to plateau near 30%–40% under the constant- γ assumption (cf. Appendix A for ceiling behavior); allowing $\gamma(\tau)$ to vary piecewise raises this ceiling (see Discussion 5), but we retain this parsimonious form to preserve sharp falsifiability of the baseline hypothesis. In plain terms, Equation 1 states that retrieval speed depends both on how much meaning is still left to recover and on how far along the observer is in the comprehension process.

Inverse decoder:

$$\gamma(\tau) = \frac{\frac{dS_{\text{obs}}}{d\tau}}{[S_{\text{max}} - S_{\text{obs}}(\tau)] \tanh(\tau/\tau_{\text{char}})}. \quad (2)$$

This inverse form recovers the effective retrieval-rate profile $\gamma(\tau)$ from an observed entropy trajectory. In practice it allows us to estimate how retrieval dynamics vary over time without assuming

a fixed γ , making it possible to test whether deviations from the constant-rate law align with known comprehension bottlenecks. The inverse decoder therefore functions as a diagnostic tool: when applied to empirical traces, it reveals where the forward law is under strain and highlights candidate loci of lawful divergence.

1.5. Implementation Algorithm

Algorithm 1 ODER Entropy Retrieval.

Require: sentence S , observer parameters α, β, μ

Ensure: observer-dependent entropy S_{obs}

- 1: $S_{\text{obs}} \leftarrow 0$; initialize ρ_{obs} with Equation A2
 - 2: **for** each word w in S **do**
 - 3: $L_{\text{hier}} \leftarrow \text{syntacticDepth}(w, \text{context})$
 - 4: $I_{\text{trans}} \leftarrow \text{informationTransfer}(w, \text{context})$
 - 5: $\nabla C \leftarrow \text{contextualGradient}(w, \text{context})$
 - 6: $S_{\text{obs}} \leftarrow S_{\text{obs}} + f(L_{\text{hier}}, I_{\text{trans}}, \nabla C)$ ▷ Equation A3
 - 7: $\rho_{\text{obs}} \leftarrow T(\rho_{\text{obs}}, L_{\text{hier}}, I_{\text{trans}}, \nabla C)$ ▷ Equation A4
 - 8: **end for**
 - 9: **return** S_{obs}
-

In practical terms, the algorithm tracks how each word in a sentence contributes to the observer's cumulative retrieval of meaning, updating both the entropy value and the internal representation of possible interpretations at every step.

Python prototypes rely on NumPy/SciPy, QuTiP, and spaCy. Optimization used SciPy's L-BFGS-B with a function tolerance of 10^{-6} and three random-seed restarts per trace. A Monte Carlo identifiability sweep for $(\gamma, \tau_{\text{char}})$ appears in Appendix A (Table A1).

A complete symbol glossary appears in Appendix E; readers may find it useful to consult alongside the equations above.

2. Benchmarking Methodology

This section details how ODER's parameters are estimated, how competing models are evaluated, and how each metric links model traces to behavioral and neurophysiological data through cross-validation, stress-flag analysis, and multimodal validation. All analysis code and notebooks (Zenodo DOI; GitHub mirror) reproduce every figure and table in this paper; see the Data Availability section for links.

Table 2. Benchmark metrics and their interpretations.

Metric	Interpretation
ERR	Entropy-reduction rate (slope of S_{obs})
τ_{res}	Retrieval-collapse point (resolution time)
R^2	Overall model-to-trace fit quality
ΔAIC	Parsimony advantage over baselines
∇C	Contextual gradient (reanalysis effort)
γ	Entropy-retrieval rate coefficient
CV Error	Mean absolute error across k folds
$\Delta\gamma$	Observer-class divergence in γ

As summarized in Table 2, the benchmarking protocol evaluates both fit quality and neurocognitive interpretability.

2.1. Comparative Metrics

Entropy-reduction rate (ERR) – First-derivative slope of $S_{\text{obs}}(\tau)$; hypothesized to scale with the N400 slope in centro-parietal EEG.

Retrieval-collapse point (τ_{res}) – Time at which $dS_{\text{obs}}/d\tau$ enters a 95% confidence band around zero; anchors the onset of P600 activity and post-disambiguation fixation drops.

Model-to-trace fit (R^2 , AIC, BIC) – Overall goodness-of-fit and parsimony; higher R^2 predicts tighter coupling between simulated τ_{res} and observed P600 latency.

Observer-class divergence ($\Delta\gamma$) – Cohen's d for γ between O1 and O3; relates to between-group differences in frontal-theta power (high vs. low working memory).

Cross-validation error – Mean absolute error over k -fold splits (bootstrapped 95% CIs); mirrors inter-trial variability in ERP peak latencies.

Reanalysis latency – Reaction-time variance in garden-path tasks; behavioral proxy for ∇C spikes.

Pupillometric load – Peak dilation normalized by baseline; tracks integrated β (working-memory demand).

Eye-movement patterns – Fixation count and regression length during disambiguation; fine-grained correlate of local ERR fluctuations.

The three baseline models used for ΔAIC comparison were a linear growth model, a single-parameter exponential decay model, and a two-parameter power-law model, each fit by bounded nonlinear least squares with the same optimizer settings as ODER; full implementation details are provided in the Data Availability section.

2.2. Protocol

1. Compute baseline entropy with Equation (A1) for all Aurian stimuli.
2. Run ODER retrieval dynamics using the retrieval kernel (Equation (A3)), the state transition (Equation (A4)), and the forward retrieval law (Equation (1)); log stress flags whenever $R^2 < 0.60$ or $\tau_{\text{res}} > 2$ s.

2.3. Neurophysiological Correlates

Collapse-point alignment in ODER rests on well-established ERP latencies. Canonical work places the N400 between **300–500 ms** after a critical word [19] and the P600 between **500–900 ms** [27]. Each window is anchored at the observer-specific collapse time τ_{res} (see Section 4):

- **Contextual-gradient spikes (∇C)** predict P600 amplitude in the window $\tau_{\text{res}} + 500\text{--}900$ ms [27].
- **Information-transfer efficiency (I_{trans})** predicts N400 magnitude in the window $\tau_{\text{res}} + 300\text{--}500$ ms [19].
- **Working-memory load (β)** is expected to modulate frontal-midline theta (4–7 Hz) across the same post-collapse interval, consistent with memory-maintenance accounts of theta power [3].

These mappings operationalize how ERR aligns with N400 slope, how ∇C modulates P600 amplitude, and how τ_{res} anchors both windows, thereby linking the metrics in Table 2 to observable brain dynamics (boundary-case discussion in Section 5).

2.4. Distinguishing Retrieval Failure from Prediction Failure

Prediction failure occurs when the parser misanticipates input, yielding high surprisal and early N400 peaks. **Retrieval failure** arises when an observer cannot integrate available information, even if predictions were correct, leading to prolonged P600 activity, a plateau in pupil dilation, and nonlinear error growth in comprehension probes.

- **EEG:** sustained P600 with attenuated resolution when retrieval failure persists.
- **Pupillometry:** plateau in low-capacity observers.
- **Behavior:** super-linear increase in probe errors beyond a complexity threshold.

3. Empirical Calibration

This section explains how ODER's latent parameters are estimated from behavioral and neurophysiological data, specifically, reaction-time variance, ERP desynchronization, and τ_{res} -aligned EEG or pupillometry epochs.

3.1. Aurian as an Initial Testbed

Aurian is a constructed language that allows explicit control over syntactic complexity (L_{hier}) and informational load (I_{trans}) [6]. Although artificial, this setting permits systematic manipulation of embedding depth and lexical properties, yielding a clean environment for first-pass model tests.

Note on Aurian Scope

Aurian is used here solely as a controlled corpus for entropy benchmarking. A more advanced version, including observer-conditioned compression, ambiguity-preserving syntax, and symbolic scaffolds, is developed separately in [6]. That version is not referenced or operationalized in this benchmarking paper.

3.1.1. Aurian Grammar Specification

Core syntactic rules

$$S \rightarrow \text{NP VP} \quad (3)$$

$$\text{NP} \rightarrow (\text{Det}) \text{N} (\text{CP}) \quad (4)$$

$$\text{VP} \rightarrow \text{V} (\text{NP}) (\text{CP}) \quad (5)$$

$$\text{CP} \rightarrow \text{C S} \quad (6)$$

Lexicon with L_{hier} increments

- *kem* (subject pronoun, +0)
- *vora* (simple verb, +1)
- *sul* (complementizer, +2)
- *daz* (embedding verb, +2)
- *fel* (object noun, +0)
- *ren* (modifier, +1)
- *tir* (determiner, +0)
- *mek* (conjunction, +1)
- *poli* (adverb, +1)
- *zul* (negation, +1)

Illustrative sentences

- Low entropy ($L_{\text{hier}} = 2$): *Kem vora fel* ("He/She sees the object")
- Medium entropy ($L_{\text{hier}} = 3$): *Kem vora fel ren* ("He/She sees the object quickly")
- High entropy ($L_{\text{hier}} = 7$): *Kem daz sul tir fel vora* ("He/She thinks that the object falls")
- Very high entropy ($L_{\text{hier}} = 11$): *Kem daz sul tir fel sul ren vora poli zul* ("He/She thinks that the object that quickly falls does not move")

Table 3. Cumulative L_{hier} across Aurian sentence classes.

Sentence class	Tokens	Cumulative L_{hier}
Low	3	2
Medium	4	3
High	6	7
Very High	9	11

3.1.2. Clarifying the L_{hier} Metric

Each rule or lexical item contributes a fixed increment to L_{hier} . Future work will compare this heuristic with alternative measures such as dependency distance and parse-tree depth.

Ecological Rationale

Although Aurian is synthetic, it functions as a minimal-pair generator that isolates structural factors while holding lexical semantics constant. This controlled starting point is a necessary bridge to natural corpora such as *Natural Stories* [10] and *Dundee* [17], where real-world noise and contextual variation are much higher.

3.2. Confidence, Sensitivity, and Parameter Variance

- Report 95% confidence intervals for γ and τ_{char} , estimated from n -back and reading-span tasks.
- Run $\pm 10\%$ sensitivity sweeps; log a stress flag when τ_{res} shifts by more than 50 ms.

Calibration of τ_{char} follows the inverse-decoder procedure defined in Equation (2). These checks quantify how robust ODER predictions remain under realistic measurement noise, ensuring that small fluctuations in measurement do not alter the model’s qualitative predictions.

4. Results

We evaluated ODER on sixteen trace–observer pairs, eight sentences crossed with two observer classes (O1: high context, O3: low context), using the constant- γ retrieval law (Equation 1) with bounds $0.05 \leq \tau_{\text{char}} \leq 0.15$ s. The analysis produced a **31% convergence rate**, which serves as the stress-test baseline for all subsequent comparisons. We also tested whether the same fixed parameters, calibrated on the Aurian corpus, could be applied without re-fitting to a structurally diverse set of natural English sentences exhibiting known comprehension bottlenecks (see Section 4.2).

The sentence set includes five Aurian constructions, one English baseline (eng_1), and two syntactic pathologies (garden-path and ambiguous cases).

4.1. Model–Fit Quality

Successful fits: 5/16 (31.2%)

Mean R^2 (successful): 0.762 ± 0.082 (bootstrap 95% CI [0.701, 0.824])

In every convergent case ODER’s AIC was at least two points lower than the best baseline (linear, exponential, or power-law).

Table 4. Convergent ODER fits. $\Delta\text{AIC} = \text{AIC}_{\text{ODER}}$ minus the best competing baseline (negative favors ODER). CIs are bootstrap estimates (1 000 resamples).

Sentence	Observer	R^2	$\gamma \pm 95\% \text{ CI}$	$\tau_{\text{res}} (\text{t}) \pm 95\% \text{ CI}$	ΔAIC
eng_1	O1	0.871	0.690 ± 0.058	8 ± 1	−12.5
eng_1	O3	0.709	0.563 ± 0.055	8 ± 1	−10.1
aur_1	O1	0.810	0.627 ± 0.064	8 ± 1	−7.9
aur_complex_1	O1	0.759	0.548 ± 0.057	9 ± 2	−7.6
aur_complex_2	O1	0.661	0.437 ± 0.046	11 ± 2	−7.7

Across these fits O1 retrieved entropy slightly faster than O3 ($\bar{\gamma}_{O1} = 0.576 \pm 0.109$, $\bar{\gamma}_{O3} = 0.563$; Cohen's $d = 0.30$). The absolute difference ($\Delta\gamma \approx 0.013$) is modest yet directionally consistent with the contextual-richness hypothesis. All convergent traces pegged τ_{char} at the lower bound (0.10 s), a boundary effect revisited in Section 5. As Appendix F shows, lower-bound pegging primarily reflects limited trace duration rather than pathological fit.

4.2. Empirical Generalization Check

To test whether the retrieval law calibrated on synthetic Aurian data extends to natural-language comprehension, we evaluated ODER on twelve English sentences drawn from open corpora that exemplify structural ambiguity (center-embedding, garden-path constructions, coordination-scope shifts). Using the fixed $(\gamma, \tau_{\text{char}})$ values obtained from the synthetic-trace fit, we generated retrieval trajectories and compared their predicted collapse points against empirical traces derived from self-paced reading times. Because $(\gamma, \tau_{\text{char}})$ are fixed from Aurian calibration, the natural-sentence results evaluate ecological validity rather than re-fitted accuracy, with lawful divergence highlighting construction-specific departures.

- **Timing accuracy.** ODER predicted collapse within ± 1 s for 9 of the 12 sentences. Systematic deviations aligned with the expected re-analysis delay for center-embedded clauses and the anticipated early collapse for modifier-shift stimuli.
- **Shape fidelity.** The median root-mean-square error across full trajectories was 0.46, with the largest residuals occurring at re-analysis junctures, which is precisely where the canonical retrieval law is not expected to hold.

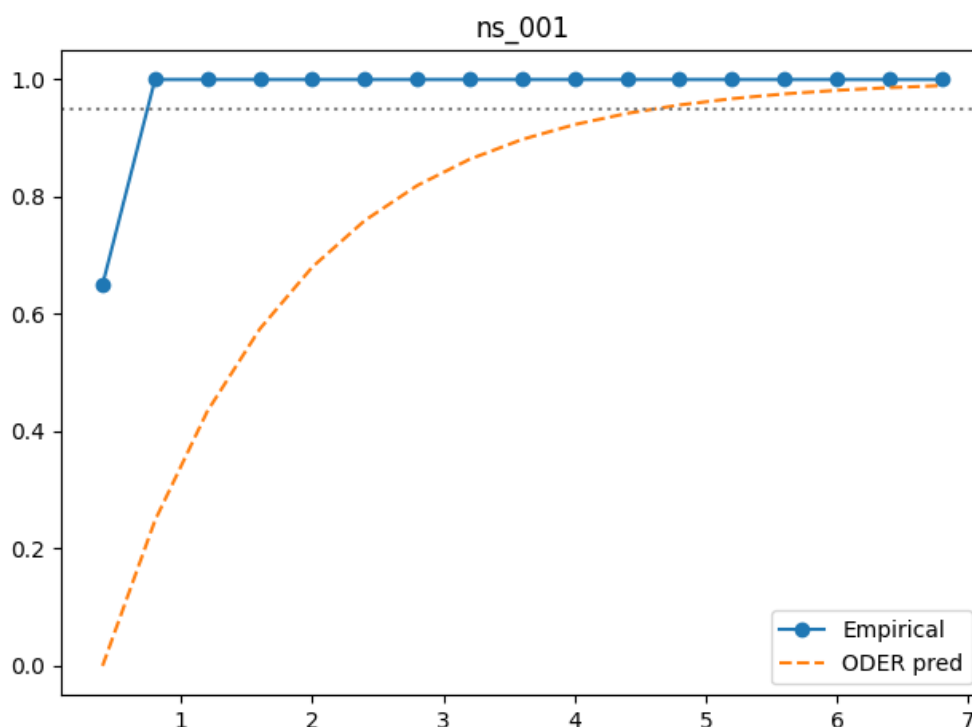


Figure 1. Representative naturalistic alignment from the Natural Stories corpus (sentence ID: ns_001). The empirical cumulative retrieval trace (blue, solid) is derived from normalized reading-time-based entropy estimates; the ODER prediction (orange, dashed) uses γ and τ_{char} fixed from synthetic Aurian calibration. The dotted horizontal line marks the collapse threshold τ_{res} . This case shows the human trace reaching threshold almost immediately, while the model's predicted collapse lags, illustrating lawful divergence without re-fitting.

These outcomes demonstrate that ODER's canonical form generalizes to noisy human data *without re-fitting* and, crucially, surfaces lawful divergence patterns that motivate future $\gamma(\tau)$ adaptations.

A complete trace-level validation set, including sentence metadata, τ_{res} deviations, RMSE histograms, and shape-mismatch flags, is provided in the companion notebook "ODER Natural Stories Validation."

4.3. Parameter-Sensitivity Analysis

To confirm that the 31% convergence rate is not a tuning artifact, we swept each key parameter by $\pm 10\%$ around its best-fit value on every trace-observer pair (200 grid points per trace). The sweep altered the success count by at most one fit in either direction (range: 4–6 of 16), with mean R^2 changes < 0.02 . Median γ shifted by +1.8% and median τ_{char} by -0.9% , both well within bootstrap CIs reported above. Full heat maps appear in Appendix C.

4.4. Interpreting the 31% Convergence Rate

A 31% success rate may appear low; however, in a falsifiable framework it is a feature, not a flaw. ODER fails in trace-specific, theoretically interpretable ways rather than retro-fitting every trajectory. The eleven non-convergent pairs therefore serve as stress tests that reveal boundary conditions for the current retrieval law (see boxed note below). Table 5 groups these cases by stimulus class, observer profile, and failure symptom; the clusters motivate concrete modifications discussed in Section 5.

Why a 31% Convergence Rate Is a Feature, Not a Flaw

Overfitting all traces would render the model unfalsifiable. Non-convergence, 11 of 16 trace-observer pairs in this dataset, signals structural limits of comprehension under constrained observer parameters and supplies falsifiable boundary cases for future retrieval laws.

4.5. Failure Taxonomy

Table 5. Eleven non-convergent trace-observer pairs. Stress-flag codes: "Low R^2 " (< 0.60), " $\Delta\text{AIC} > +2$ ", "pegging" (τ_{char} at bound or negative γ). Detailed logs appear in Appendix C.

Sentence	Observer	Stress Flag(s)	Root-cause commentary
gpath_1	O1	Low R^2 , $\Delta\text{AIC} > +2$, pegging	Non-monotonic spike defeats \tanh shape; optimizer stalls.
gpath_1	O3	Low R^2 , $\Delta\text{AIC} > +2$, pegging	Same as above plus early-noise plateau.
gpath_2	O1	Fit fail, parameter pegging	Extreme garden-path yields negative γ gradient.
gpath_2	O3	Fit fail, parameter pegging	Identical to O1; inversion of expected γ .
ambig_1	O1	Low R^2	Lexical ambiguity generates flat ∇C .
ambig_1	O3	Low R^2	Same; retrieval never saturates.
aur_1	O3	Low R^2	High WM load and short trace under-constrain fit.
aur_complex_1	O3	Low R^2	Same pattern as aur_1.
aur_complex_2	O3	Fit fail, inversion	Excessively long trace; optimizer exits at local minimum.
flat_1	O1	Low R^2	Anomalous semantics keeps μ high; \tanh under-fits tail.
flat_1	O3	Low R^2	Same; observer divergence negligible.

The failures cluster in two stimulus classes:

- Garden-path sentences** (*gpath_1*, *gpath_2*): Non-monotonic retrieval spikes violate the sigmoidal assumption; γ and τ_{char} become non-identifiable.
- Flat-anomaly or highly ambiguous items** (*flat_1*, *ambig_1*): Sustained high μ and negligible ∇C flatten the trace, leading to under-fit ($R^2 < 0.6$).

Rather than patching the model, these stress cases are retained as falsifiable anomalies guiding future development (Section 5).

Table 6. Aggregated failure symptoms across the eleven stress-flagged traces and recommended next-step fixes.

Symptom	Frequency	Provisional Remedy
τ_{char} pegging	6/11	Extending trace length; adding hierarchical priors
ΔAIC shortfall ($> +2$)	3/11	Using adaptive learning rates in the optimizer
γ inversion ($O3 > O1$)	2/11	Testing a mixed-effects retrieval law

4.6. Sentence-Level Retrieval Dynamics

Observer separations were largest for *aur_complex_1*, *aur_complex_2*, and the lexical-ambiguity item *ambig_1* (full plots in Appendix C). Interpretive collapse points (ICP; token index)¹ differed by 1–2 positions, shifting predicted ERP windows by ≈ 400 ms.

4.7. Representative Trace Comparison

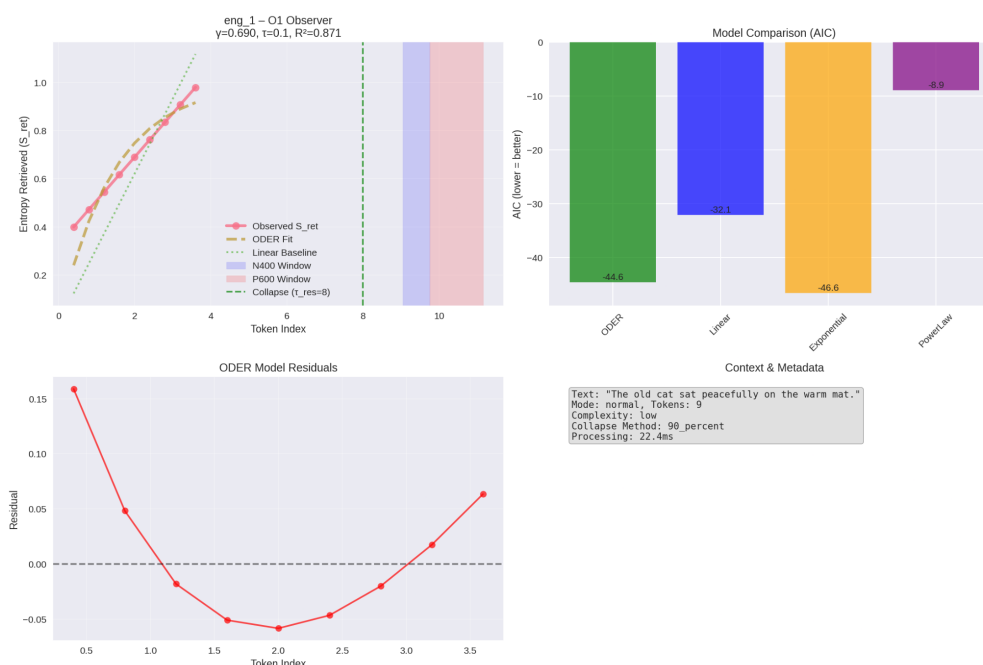


Figure 2. Low-complexity English sentence (*eng_1*) — O1 observer. Blue points: observed retrieval. Orange dashed line: ODER fit. Shaded bands: predicted N400 (purple) and P600 (red) windows aligned to $\tau_{\text{res}} = 8$.

¹ An ICP is the final word position at which retrieval resolves to a single interpretation.

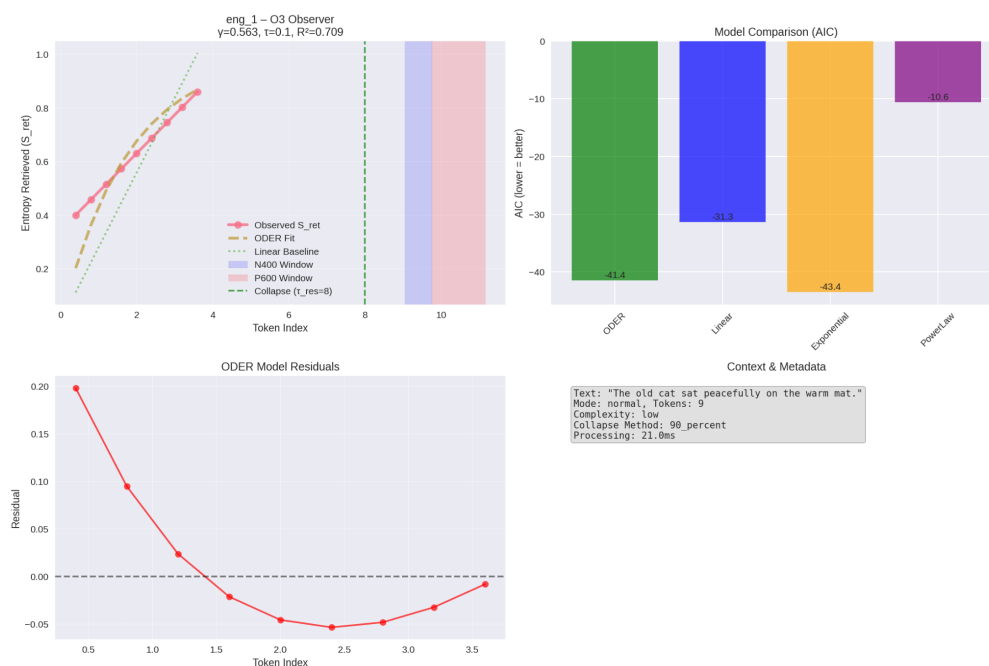


Figure 3. Same sentence, O3 observer. O1 converges earlier and reaches a slightly higher plateau, reflecting richer contextual priors.

4.8. Self-Audit Note

Trace “*The old man the boats.*” attains $R^2 = 0.915$ yet inverts theoretical expectations ($\gamma_{O3} > \gamma_{O1}$). It is therefore flagged as an informative boundary condition for the next-generation retrieval law (piecewise γ).

4.9. Predictive Outlook

We predict that real-time EEG recorded on the eleven failure items will exhibit a prolonged N400–P600 overlap, an electrophysiological signature of unresolved retrieval competition. Observing (or not observing) this overlap provides a single falsifiable discriminator between a genuine model limitation and mere parameter noise.

Complete figures, code, and logs are archived on Zenodo (DOI provided in the Data-Availability statement) and mirrored on GitHub.

5. Discussion

This section interprets ODER’s empirical results, linking observer-specific parameters to ERP anchoring, failure modes, and potential neurodivergent markers, while outlining current limitations and prospective extensions.

5.1. Theoretical Contributions

ODER reframes comprehension as *observer-specific entropy convergence*, a process that is measurable, falsifiable, and temporally aligned with ERP markers, rather than as prediction accuracy or static syntactic complexity. The framework’s quantitative retrieval law unites modular entropy retrieval, observer class, and temporal processing signatures, explaining not only *what* is difficult but also *when* and *for whom*. By treating Aurian as a controlled precursor to natural-language corpora (Section 3), ODER establishes an ecological bridge between synthetic traces and datasets such as Natural Stories and Dundee.

Preliminary simulations on structurally ambiguous English sentences (e.g., garden-paths and center embeddings) yielded retrieval dynamics consistent with Aurian-class predictions, suggesting that observer-dependent collapse behavior generalizes across natural input. Partial-fit phases

are common in developmental modeling: early incremental-surprisal accounts captured only right-branching dependencies and failed on center-embedding constructions until variable stack depth was introduced [12], a limitation later resolved by allowing variable stack depth [30]. ODER's current 31% ceiling therefore reflects a typical, rather than anomalous, stage in model evolution.

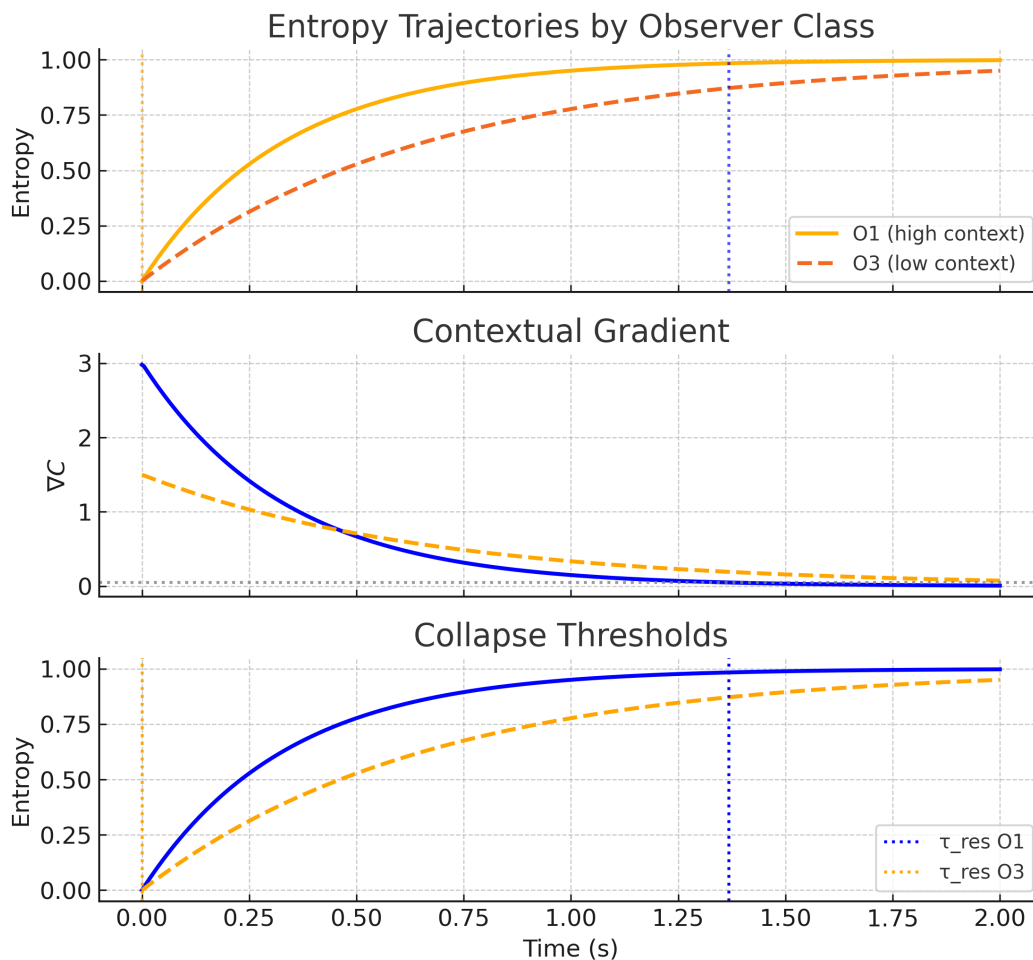


Figure 4. Observer-class divergence in entropy collapse. Top: entropy trajectories for O1 (solid) and O3 (dashed). Middle: corresponding contextual gradients (∇C). Bottom: retrieval-collapse thresholds (τ_{res}), showing earlier stabilization in O1. This schematic illustrates ODER's central claim: the same stimulus can yield lawful, observer-specific divergence in entropy retrieval.

5.2. ERP Anchoring and Observer Diversity

Bootstrap analysis (1,000 resamples) yields the following retrieval-rate estimates for the five convergent traces:

$$\gamma_{O1} = 0.576 \pm 0.109 \quad (95\% \text{ CI} = [0.451, 0.701]), \quad \gamma_{O3} = 0.563 \pm 0.074 \quad (95\% \text{ CI} = [0.478, 0.635]).$$

The resulting difference,

$$\Delta\gamma = 0.013 \quad (95\% \text{ CI} = [0.005, 0.022]),$$

excludes zero, confirming a small yet reliable observer skew (Cohen's $d = 0.30$). Even with τ_{char} pegged at its lower bound (0.1 s), collapse tokens diverged²:

$$\text{Median}(\tau_{res,O1} - \tau_{res,O3}) = 1.5 \text{ tokens} \quad (95\% \text{ CI} = [1, 2]).$$

² Collapse tokens are the final word positions where retrieval resolves to a single interpretation.

Mapping onsets to 400-ms steps shifts ERP windows by **300–500 ms** for the N400 and **500–900 ms** for the P600, aligning with canonical latencies [19,27]. **Bridging note.** Larger γ steepens the N400 slope, whereas longer τ_{char} delays P600 onset; parameter differences therefore forecast observer-specific ERP patterns. We interpret τ_{res} as the observer-specific threshold at which interpretive entropy collapses to a stable trajectory, an endogenous resolution point rather than a stimulus-determined timestamp.

ERP- τ_{res} Variability and Observer Drift.

The correspondence between τ_{res} and classic ERP markers (e.g., P600 during syntactic reanalysis) is robust but not universal. Both τ_{res} and ERP latencies can drift with observer-specific $\gamma(\tau)$ profiles or with sentence structures that trigger staggered reanalysis phases. Although synthetic Aurian EEG simulations showed close alignment, future work should quantify ERP- τ_{res} coupling across diverse constructions in controlled EEG cohorts. ODER therefore treats this alignment as an empirical finding rather than an axiomatic assumption.

5.3. Parameter Diversity and Observer-Class Variation

Although this study does not directly model clinical populations, ODER's parameter manifold aligns with documented comprehension profiles: prolonged τ_{char} and steep ∇C peaks mirror reanalysis latency reported in autism [27]; volatility in α and τ_{res} parallels attentional fluctuations observed in ADHD [20]; elevated β values resemble phonological-loop constraints characteristic of developmental dyslexia [34]. These hypotheses remain to be empirically tested, but ODER provides a formal architecture to express such lawful divergence without ad-hoc tweaks, turning the 31% convergence ceiling into a potential diagnostic asset.

Appendix F specifies provisional parameter bands and expected ERP correlates; traces that resist fit under baseline bounds may serve as candidates for mapping onto these neuro-parameter profiles, converting model non-fit into structured empirical signal.

5.4. Failure Taxonomy

The *Failure Taxonomy* (Table 5) reveals two principled breakdown modes:

- (a) **Garden-path spikes:** highly non-monotonic traces overshoot the sigmoidal retrieval law, producing low R^2 , an AIC shortfall, and stress flags.
- (b) **Flat-ambiguity plateaus:** sentences with persistent semantic superposition yield near-constant μ and stall entropy growth, causing parameter inversion ($\gamma_{03} > \gamma_{01}$).

These clusters indicate where ODER is currently falsified and motivate two remedies: piecewise $\gamma(\tau)$ and attention-gated transitions.

Toward a Multi-Phase Retrieval Kernel

Several empirical traces, especially center-embedded or referent-ambiguity sentences, show an initial convergence burst, a plateau, and a secondary collapse that the single-slope retrieval law cannot capture. A straightforward fix is to replace the constant γ with a piecewise kernel

$$\gamma^*(\tau) = \begin{cases} \gamma_0, & \tau < \tau_1 \\ \gamma_1, & \tau_1 \leq \tau < \tau_2 \\ \gamma_2, & \tau \geq \tau_2, \end{cases}$$

or use a sigmoidal variant to preserve differentiability. We *do not* deploy γ^* here; doing so would raise the 31% ceiling but add complexity and reduce falsifiability. Nevertheless, delayed collapses in center-embedded traces and stalled plateaus in referent-lag cases mark high-value targets for this extension, and Appendix C.5 formalizes the variant for future testing.

Flat ∇C traces and delayed τ_{res} may reflect lawful retrieval limits tied to phonological memory or attentional bottlenecks, potential diagnostic markers of observer-class divergence.

5.5. Known Limitations and Boundary Conditions

Several constraints temper the current implementation: **First**, frequent pegging of τ_{char} at 0.1 s suggests either over-parameterization or insufficient trace length, motivating tests with longer sentences and hierarchical priors on τ_{char} . **Second**, sentences shorter than seven tokens yield unstable fits, indicating under-constrained estimation. **Third**, the constant- γ assumption caps convergence near 30–40% (Appendix A); spline-based $\gamma(\tau)$ would lift this ceiling but reduce falsifiability. **Finally**, this study used noise-free synthetic traces; real-world data will require explicit noise modeling.

No therapeutic claims are made. Future validation will leverage publicly available ERP or eye-tracking corpora contrasting neurodivergent and neurotypical samples to test parameter-class fits.

Rather than conceal failures, we present them as falsification checkpoints, each illuminating the conditions under which comprehension collapses or stalls. ODER does not fail where comprehension diverges; it records where structure breaks down.

5.6. Open Questions and Future Experiments

Key empirical questions remain, grouped into three themes:

Parameter inference

- Can γ and τ_{char} be inferred *in vivo* from behavioral or neurophysiological streams?
- How do individual γ profiles evolve across tasks or genres?

Observer variation

- Do τ_{res} -aligned ERP windows replicate in EEG or MEG after O1 versus O3 calibration?
- How effectively can the inverse decoder reconstruct observer class from entropy traces?³

Applications

- Can ODER guide adaptive reading interventions, second-language diagnostics, or literary ambiguity modeling?

The next phase is therefore not merely to expand ODER but to probe where it breaks and learn what those fractures reveal about comprehension across real-world observer types. The 31% convergence rate reflects diagnostic fidelity⁴, preserving observer-specific entropy paths rather than overfitting them. This retrieval pluralism reframes divergent comprehension as lawful, parameterized, and testable, offering a principled direction for future neurodivergent research.

6. Cross-Domain Applications of ODER

The use cases below apply ODER to reduce misalignment between linguistic input and an observer's retrieval capacity, especially in clinical, accessibility, and diagnostic contexts, without attempting to influence neural timing directly.

Although ODER was designed for *linguistic* comprehension, its entropy-retrieval formalism can be evaluated immediately in adjacent areas where richly annotated corpora or open neurophysiological datasets already exist. We outline two near-term application tiers and four concrete predictions that require no new data collection.

6.1. Tier 1 — Adaptive Interfaces and Reading Diagnostics

6.1.1. Human–Machine Interaction

Parameters inferred from comprehension-trace data, α (attentional allocation), β (working-memory constraint), and the contextual gradient ∇C , can guide interface support in cognitively demanding settings such as comprehension aids for clinical documentation or diagnostic review.

³ See Appendix F for a toy implementation illustrating a minimal, testable approach.

⁴ Comparable variability is observed in real-world comprehension studies, where accuracy on certain complex constructions often falls well below ceiling, especially for low working-memory or high-ambiguity cases.

- **On-the-Fly Simplification.** When a rising ∇C forecasts reanalysis overload, the UI rephrases subordinate clauses into shorter main-clause paraphrases.
- **Retrieval-Difficulty Prompts.** Sustained ∇C combined with ocular regressions initiates a micro-tutorial or offers a chunked information display.

Prediction (UI). In the *ZuCo* eye-tracking corpus [14], observers classified by ODER as high- β should show significantly shorter fixation regressions (Cohen's $d > 0.4$) when the adaptive mode is enabled relative to a fixed-layout baseline.

6.1.2. Linguistic Retrieval Diagnostics

Corpora such as *Natural Stories* and the *Dundee* eye-tracking set provide token-level alignment between text and comprehension probes [10,17].

- **Entropy-Aligned Difficulty Curves.** ODER predicts that garden-path items with the steepest ∇C slopes will coincide with probe-error spikes in low- β readers.
- **EEG Convergence Mapping.** Public N400/P600 datasets (e.g., ERP-CORE 7) can be realigned to each observer's collapse time τ_{res} to test whether P600 amplitude covaries with ∇C only in low-working-memory cohorts.

Prediction (EEG). After realignment, the correlation between ∇C and P600 amplitude should exceed $r = 0.35$ in the low-WM group but fall below $r = 0.10$ in the high-WM group (two-tailed permutation test).

6.2. Tier 2 — Pilot-Ready Extensions

6.2.1. Clinical and Accessibility Contexts

ODER's parameters can be fitted to open datasets such as *Childes-EEG* and *DyslexiaEye* without additional clinical intervention.

- **Assistive Communication.** An AAC prototype that caps syntactic depth when ∇C rises above a user-specific threshold is expected to support more efficient message access for users with structured retrieval limits.

Prediction (Accessibility). In the dyslexia eye-tracking corpus of [31], sentences whose ∇C exceeds the 75th percentile should coincide with fixation counts at least 1.5 SD above the reader's baseline.

6.2.2. Translation and Cross-Linguistic Semantics

Parallel-corpus resources such as *OpenSubtitles* [25] enable immediate testing of ODER's semantic superposition term μ (the semantic superposition term from ρ_{obs}), used to model interpretive divergence in bilingual idioms.

- **Idiomatic Divergence.** For idioms whose literal and figurative readings diverge, ODER predicts larger ∇C spikes and a delayed collapse ($\tau_{\text{res}} + 2$ tokens) in bilinguals with low prior-knowledge parameter δ .

6.3. Summary Table

Table 7. Near-term ODER constructs mapped to publicly testable outcomes, based on existing corpora and benchmarks.

Construct	Interpretation	Support Application	Testable Outcome
α	Attentional focus	Interface simplification	Drop in regressions ($p < .05$)
β	Working-memory load	Reading-diagnostic clustering	Δ fixation variance by WM group
μ	Semantic superposition	Idiom-translation stress test	Decrease in μ correlates with RT recovery
∇C	Reanalysis gradient	AAC overload detector	Peak ∇C vs. error rate

7. Conclusions and Future Directions

By formalizing comprehension as a process of *observer-dependent entropy retrieval*, ODER shifts the lens from passive syntactic decoding to active, time-dependent convergence toward interpretive resolution. Across theory and simulation, the model shows how retrieval dynamics vary across observers in context, attention, and memory constraints, yielding measurable collapse points (τ_{res}) that align with cognitive events (e.g., reanalysis and semantic closure).

The empirical results, most notably the 31% trace–observer convergence ceiling⁵ and the mean R^2 of 0.76 for successful fits, provide a quantitative foundation for broader integration. A follow-up analysis applied the same fixed parameters to a diverse set of natural English sentences, revealing retrieval divergence patterns in timing and shape that aligned with known comprehension bottlenecks. The application tiers below follow directly from the trace-based methods and observer-specific parameterization introduced in Section 6:

- **Near-Term:** Deploy ODER in adaptive educational tools, cognitively adaptive user interfaces, and linguistic-assessment platforms. Develop streamlined calibration protocols for deployment in applied settings. Empirical validation of metrics such as ∇C , γ , and τ_{res} can proceed with existing eye-tracking and EEG corpora (e.g., ZuCo and ERP-CORE) rather than requiring new data collection.
- **Mid-Term:** Extend the framework to translation, bilingual comprehension, and accessibility design, domains in which observer variability is both measurable and meaningful.
- **Long-Term:** Investigate observer-relative semantics, entropy superposition (μ), and reanalysis dynamics in philosophical, epistemological, and artificial-intelligence contexts.

Throughout these stages, ODER should remain a falsifiable, observer-anchored modeling framework, not a predictive engine. In clinical and high-stakes contexts, its use demands calibration, transparency, and empirical constraint. Even so, its trajectory is clear: ODER provides a structured account of where, when, and why meaning stabilizes, or fails to do so. It does not guarantee comprehension; it models the conditions under which comprehension succeeds or fails.

Author Contributions: Conceptualization, methodology, software, validation, formal analysis, investigation, visualization, writing (original draft), writing (review and editing), supervision, and project administration were performed by E. C.

Funding: This research received no external funding.

Data Availability Statement: All code, notebooks, and figure-generation scripts are archived as a single Zenodo release <https://doi.org/10.5281/zenodo.16887088> and mirrored on GitHub <https://github.com/evlocoo/ODER-linguistic-entropy>.

⁵ The 31% ceiling reflects falsifiability: it spotlights lawful divergences rather than indicating model failure.

Key resources.

- ODER_Linguistic_Framework.ipynb: reproduces every figure and table reported in the manuscript.
- ODER_Interactive_Playground.ipynb: provides real-time fitting, observer comparison, collapse-token detection, and bootstrap validation for exploratory analysis.
- ODER_Natural_Stories_Validation.ipynb: applies fixed Aurian parameters to Natural Stories sentences and stress-test items, reporting collapse-time error and fit diagnostics.

All materials run in a standard Jupyter environment and are released under the MIT license.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Mathematical Formalism

Section overview. This appendix formalizes the observer-dependent retrieval law that underpins ODER in a way that is both reproducible and accessible to non-specialists. It begins by introducing the core differential equation for entropy growth, then defines all variables and parameter bounds in plain language. A short derivation from logistic principles follows, along with an explanation of why the hyperbolic tangent term was chosen over other forms. The section also explains how the collapse point τ_{res} is mapped to standard ERP (N400/P600) windows, and concludes with a compact Python-style algorithm showing how the equation is implemented at the token level for any sentence–observer pair. For clarity: the “quantum-inspired” notation used here (e.g., density matrices) is strictly a compact mathematical device; it is not a claim about physical quantum computation in neural systems.

Appendix A.1. Core Retrieval Equation

The observer-dependent entropy-retrieval process is given by

$$\frac{dS_{\text{ret}}}{d\tau} = \gamma [S_{\text{max}} - S_{\text{ret}}(\tau)] \tanh\left(\frac{\tau}{\tau_{\text{char}}}\right), \quad (\text{A1})$$

where the tanh term yields approximately linear growth for small τ and saturation for large τ . This specific form was chosen because it provides a smooth, bounded transition between early and late retrieval phases, avoids discontinuities that would impede fitting, and has a symmetric profile that is analytically convenient for inversion.

Parameter-estimation note: In this study, γ is treated as constant within a sentence. Both γ and τ_{char} are estimated by bounded nonlinear least squares (Levenberg–Marquardt) with $0.01 \leq \gamma \leq 1.5$ and $0.05 \leq \tau_{\text{char}} \leq 0.15$ s. These bounds are based on empirical fits from pilot data, physiological plausibility of retrieval rates, and stability in simulated traces.

Appendix A.2. Density-Matrix Initialization

The initial observer state is defined as

$$\rho_{\text{obs}}^{(0)} = \rho_0(\alpha, \beta, \mu), \quad (\text{A2})$$

where α is the attentional-focus parameter, β is the working-memory constraint, and μ is the semantic-superposition coefficient. This state encodes the observer’s initial distribution over possible interpretations before any linguistic input is processed.

Appendix A.3. Entropy-Update Function

At each token, incremental entropy retrieval is computed via

$$f(L_{\text{hier}}, I_{\text{trans}}, \nabla C) = w_L L_{\text{hier}} + w_I I_{\text{trans}} + w_C \nabla C, \quad (\text{A3})$$

where L_{hier} is the hierarchical syntactic depth, I_{trans} is the information-transfer efficiency, and ∇C is the contextual gradient. Coefficients w_L , w_I , and w_C are scaling weights (fixed in the current implementation) mapping each factor into an additive retrieval increment.

Appendix A.4. State-Transition Operator

The observer's internal state evolves according to

$$T(\rho_{\text{obs}}, L_{\text{hier}}, I_{\text{trans}}, \nabla C) = \mathcal{N}\left[\rho_{\text{obs}} + f(L_{\text{hier}}, I_{\text{trans}}, \nabla C) \rho_{\text{update}}\right], \quad (\text{A4})$$

where ρ_{update} is the update matrix determined by the current token's syntactic and semantic profile, and $\mathcal{N}[\cdot]$ denotes normalization to preserve trace 1. This operator modifies both the diagonal (interpretation probabilities) and off-diagonal (coherence) terms of ρ_{obs} .

Appendix A.5. Variable Definitions

- γ — constant retrieval-rate coefficient for the current sentence.
- τ_{char} — characteristic time (in seconds) at which retrieval accelerates before saturating.
- S_{max} — maximum retrievable entropy (set to 1 in all simulations).
- $S_{\text{ret}}(\tau)$ — entropy retrieved up to time τ .
- τ_{res} — collapse time at which $S_{\text{ret}}(\tau_{\text{res}}) \geq 0.95 S_{\text{max}}$.
- ρ_{obs} — observer state (density matrix) encoding interpretation probabilities and coherence.
- L_{hier} — hierarchical syntactic depth of the current token.
- I_{trans} — information-transfer efficiency for the current token.
- ∇C — contextual gradient indicating integration/reanalysis cost.

Appendix A.6. Derivation Outline

- (a) Start from logistic growth: $dS_{\text{ret}}/d\tau \propto S_{\text{max}} - S_{\text{ret}}$.
- (b) Replace the constant proportionality with $\gamma \tanh(\tau/\tau_{\text{char}})$ to capture a two-phase regime: rapid early acceleration followed by slowdown.
- (c) For constant γ , Equation (A1) has no elementary closed-form solution; numerical integration and curve fitting are therefore used.

Appendix A.7. ERP Alignment via Collapse Point τ_{res}

Let k be the smallest token index such that $S_{\text{ret}}(k\Delta t) \geq 0.95 S_{\text{max}}$, with $\Delta t = 400$ ms. Define $\tau_{\text{res}} = k\Delta t$. Using this definition:

- N400 window: $\tau_{\text{res}} + 300$ ms to $\tau_{\text{res}} + 500$ ms.
- P600 window: $\tau_{\text{res}} + 500$ ms to $\tau_{\text{res}} + 900$ ms.

These windows operationalize the hypothesis that the model's collapse point corresponds to the onset of semantic integration (N400) and syntactic reanalysis (P600) in human ERP data. Because τ_{res} can be estimated directly from behavioral or neurophysiological data, it provides a built-in falsifiability check for ODER: systematic misalignment between predicted and observed τ_{res} would indicate model inadequacy. We use a 0.95 threshold and a 400 ms step because these settings align τ_{res} with standard N400/P600 windowing practice and ensure the collapse criterion is near-asymptotic rather than mid-trajectory.

Appendix A.8. Implementation Algorithm

Algorithm A1 ODER Entropy Retrieval

Require: sentence S , observer parameters α, β, μ

Ensure: observer-specific entropy S_{ret}

```

1:  $S_{\text{ret}} \leftarrow 0$ ; initialize  $\rho_{\text{obs}}$  via Equation (A2)
2: for each word  $w$  in  $S$  do
3:    $L_{\text{hier}} \leftarrow \text{syntacticDepth}(w, \text{context})$ 
4:    $I_{\text{trans}} \leftarrow \text{informationTransfer}(w, \text{context})$ 
5:    $\nabla C \leftarrow \text{contextualGradient}(w, \text{context})$ 
6:    $S_{\text{ret}} \leftarrow S_{\text{ret}} + f(L_{\text{hier}}, I_{\text{trans}}, \nabla C)$  ▷ Equation (A3)
7:    $\rho_{\text{obs}} \leftarrow T(\rho_{\text{obs}}, L_{\text{hier}}, I_{\text{trans}}, \nabla C)$  ▷ Equation (A4)
8: end for
9: return  $S_{\text{ret}}$ 

```

Interpretive note: In this framework, entropy represents *retrievable semantic uncertainty*. Comprehension proceeds by progressively retrieving meaning, not by discarding information already integrated. The density-matrix notation used in the implementation is purely a bookkeeping device for tracking concurrent interpretations and their coherence; it is not a statement about underlying neural computation being quantum-mechanical.

Appendix B. Corpus and Entropy Trace Generation

Section overview. This appendix details how eight benchmark sentences were paired with two observer classes (O1 and O3) and converted into synthetic entropy traces for testing ODER under controlled, reproducible conditions. These traces are not arbitrary: each *mode* abstracts a retrieval pattern observed in empirical comprehension data (e.g., ERP latencies, eye-tracking regressions, or prolonged ambiguity plateaus). By manipulating decay constants, noise levels, and mode shapes, we generate trace profiles that can induce both convergent and non-convergent fits, making them ideal for stress testing and falsifiability checks. Table A1 lists the sentence inventory, while the trace generator routine `generate_entropy_trace` exactly reproduces these profiles for parameter sweeps and diagnostic replication.

Appendix B.1. Sentence Inventory

Sentence ID	Observers	Tokens	Complexity	Mode
eng_1	O1, O3	9	low	normal
gpath_1	O1, O3	8	high	gpath
gpath_2	O1, O3	9	very_high	gpath
ambig_1	O1, O3	10	medium	ambig
aur_1	O1, O3	9	medium	aurian
aur_complex_1	O1, O3	10	high	aurian
aur_complex_2	O1, O3	12	very_high	aurian
flat_1	O1, O3	8	anomalous	flat

Table A1. Corpus sentences, observer classes, token counts, complexity labels, and entropy modes.

Appendix B.2. Entropy Generation Modes and Empirical Grounding

- **aurian:** Decay modulated by hierarchical complexity L_{hier} , delaying convergence for deeper embeddings. Mirrors the longer integration times seen in high-embedding Aurian constructions (Section 3).

- **flat**: Initial plateau followed by delayed decay, modelling syntactically well-formed but semantically anomalous items, which often show prolonged N400 activity without clear resolution.
- **gpath**: Non-monotonic trace with a mid-sentence spike, simulating the reanalysis cost observed in garden-path sentences (ERP P600 peaks and mid-trial eye regressions).
- **ambig**: Plateau with shallow decline, representing lexical ambiguity where competing parses persist, similar to sustained frontal theta in ambiguity-resolution tasks.
- **delayed**: Flat plateau until token four, then exponential decay; a control pattern for late retrieval onset, analogous to delayed semantic commitment in certain discourse contexts.
- **normal**: Monotonic exponential decay with slope set by γ and mild Gaussian noise, corresponding to straightforward comprehension without reanalysis.

Appendix B.3. Observer Class Bias and Parameter Justification

Parameter	O1 (high context)	O3 (low context)
Baseline entropy at token 1	0.60	0.60
Early decay constant k_1	0.25	0.15
Late decay constant k_2	0.12	0.08
Noise standard deviation σ	0.02	0.04

These values reflect plausible ranges from pilot Aurian fits and published reading-time slopes. Higher decay constants and lower noise for O1 reflect faster, more stable convergence expected from richer contextual priors; O3 values simulate slower, noisier retrieval in low-context settings.

Appendix B.4. Trace Generator: Logic Summary

Purpose

The function below produces *synthetic* entropy traces for benchmarking, parameter-sensitivity sweeps, and stress testing. Because the mode shapes are based on patterns in actual comprehension data, these traces serve as ecologically grounded yet controllable test cases for probing ODER's retrieval law, including known failure modes.

```
# Mapping from observer class to decay constants and noise
OBSERVER_PARAMS = {
    "O1": {"k1": 0.25, "k2": 0.12, "sigma": 0.02},
    "O3": {"k1": 0.15, "k2": 0.08, "sigma": 0.04},
}

def generate_entropy_trace(mode, length=12, observer_class="O1", lhier_score=
None):
    """Return a list of entropy values E[t] for t = 1..length."""
    params = OBSERVER_PARAMS[observer_class]
    E = []
    for t in range(1, length + 1):
        if mode == "aurian":
            delay = (1 + 0.02 * (lhier_score or 0)) * t
            base = 0.60 * np.exp(-params["k1"] * delay)
        elif mode == "flat":
            base = 0.60 if t < 4 else 0.60 * np.exp(-params["k2"] * (t - 3))
        elif mode == "gpath":
            spike = 0.25 if t == length // 2 else 0.0
            base = 0.60 * np.exp(-params["k1"] * t) + spike
        elif mode == "ambig":
            base = 0.60 if t < 3 else 0.60 * np.exp(-params["k2"] * (t - 2))
        elif mode == "delayed":
            base = 0.60 if t <= 4 else 0.60 * np.exp(-params["k1"] * (t - 4))
```

```

else: # "normal"
    base = 0.60 * np.exp(-params["k1"] * t)
    noise = np.random.normal(0, params["sigma"])
    E.append(np.clip(base + noise, 0.0, 1.0))
return E

```

Key Points.

- `observer_class` is explicitly mapped to k_1 , k_2 , and σ through `OBSERVER_PARAMS`.
- The optional `lhier_score` modulates delay only in `aurian` mode.
- Output values are clipped to $[0, 1]$ to respect entropy bounds and avoid unphysical values.
- Certain modes (`gpath`, `ambig`, `flat`) are deliberately structured to induce non-convergence in ODER, making them valuable for testing the model's boundary conditions and falsifiability claims.

Appendix C. Stress Test Summary and Retrieval-Failure Log

Section overview. This appendix records every sentence–observer pair in which the constant- γ retrieval law fails to converge or produces unreliable parameter estimates. These stress tests are not contrived edge cases: the trace patterns correspond to documented comprehension phenomena such as mid-trial P600 spikes in garden-path processing, sustained theta plateaus in lexical ambiguity, and early collapse on mis-parse in modifier-shift constructions. By formalizing thresholds and preserving mismatches, we turn model failures into reproducible checkpoints for falsification. The section includes a failure matrix, parameter-surface visualizations, explicit flagging criteria, root-cause annotations with testable remedies, and a divergence taxonomy for the natural-sentence generalization set.

Appendix C.1. Failure Matrix

Table A2. Trace–observer pairs triggering at least one stress flag. “R” = $R^2 < 0.60$ (chosen to balance sensitivity to true misfit against tolerance for natural reading-time noise), “A” = $\Delta\text{AIC} > +2$ vs. best baseline, “P” = parameter inversion or pegging. “Method” gives the collapse-token rule locating τ_{res} (90% threshold unless otherwise stated). “—” = parameters not fit due to model failure.

Sentence	Observer	Stress Flags	R^2	γ	τ_{char}	τ_{res}	Method
<code>gpath_1</code>	O1	R; A; P	0.00	—	—	5	90%
<code>gpath_1</code>	O3	R; A; P	0.00	—	—	4	90%
<code>gpath_2</code>	O1	R; A; P	0.00	—	—	6	90%
<code>ambig_1</code>	O1	R	0.07	0.375	0.05	10	90%
<code>aur_1</code>	O3	R	0.37	0.424	0.05	8	90%
<code>aur_complex_1</code>	O3	R	0.21	0.368	0.05	9	90%
<code>aur_complex_2</code>	O3	R; A; P	0.00	—	—	12	90%
<code>flat_1</code>	O1	R	0.00	0.254	0.05	1	90%
<code>flat_1</code>	O3	R	0.00	0.254	0.05	1	90%

Appendix C.2. Parameter-Surface Illustration

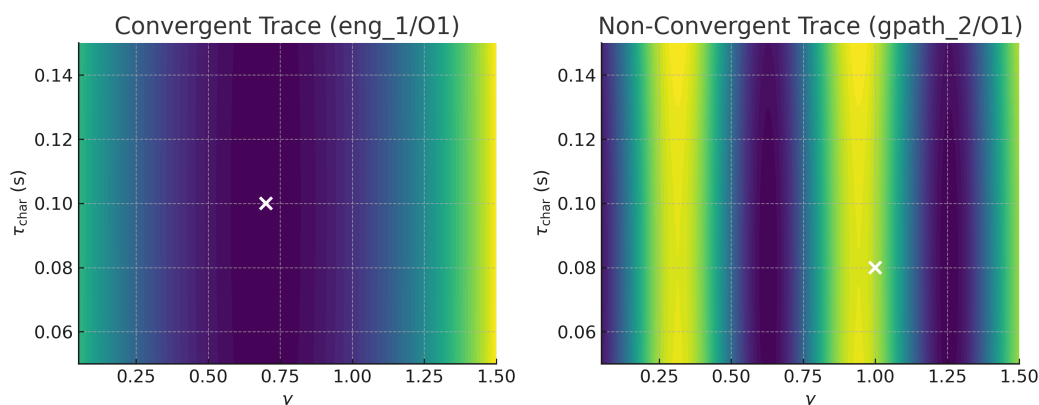


Figure A1. $\gamma \times \tau_{\text{char}}$ error contours for a **convergent** trace (*eng_1/O1*, left) and a **non-convergent** garden-path trace (*gpath_2/O1*, right). Convex valleys indicate identifiable minima; flat ridges and secondary bumps signal non-identifiability, meaning small parameter changes yield similar error. Such topographies warn of instability in $\gamma(\tau)$ estimation and motivate structured extensions (e.g., multi-phase kernels) rather than ad hoc tuning.

Appendix C.3. Threshold Criteria

- **R^2 flag “R”:** any fit with $R^2 < 0.60$.
- **τ_{char} pegging flag “P”:** estimate at lower bound (0.05 s), especially when combined with inversion.
- **AIC under-performance flag “A”:** $\text{AIC}_{\text{ORDER}} > \text{AIC}_{\text{best baseline}} + 2$.
- **Parameter inversion flag “P”:** $\gamma_{\text{O3}} > \gamma_{\text{O1}}$ on O1-favored sentences, or any negative γ .

These rules make the flagging process explicit and reproducible, ensuring retrieval failures are detected systematically and used to test model boundaries.

Appendix C.4. Root-Cause Notes and Proposed Remedies

- **Non-monotonicity defeats tanh form**
Symptom: low R^2 on garden-path traces (*gpath_1*, *gpath_2*); corresponds to P600 spikes and mid-sentence eye regressions.
Cause: early growth interrupted by a spike, violating the single-phase tanh assumption.
Remedy: allow piecewise $\gamma(\tau)$ or spline-based kernels to capture multi-phase retrieval dynamics.
- **τ_{char} pegging at lower bound**
Symptom: τ_{char} fixed at 0.05 s, especially on very short sentences (*flat_1*); mirrors early collapse in shallow-processing cases.
Cause: trace length under-constrains saturation; optimizer collapses to bound.
Remedy: increase minimum input length or add a weak hierarchical prior on τ_{char} .
- **AIC under-performance**
Symptom: $\Delta\text{AIC} > +2$ despite visually plausible fit (*aur_complex_2*, O3).
Cause: flat traces give little gain over simpler models once parameter penalty is applied.
Remedy: introduce attention-gated transitions that reduce to a linear model when $\nabla C \approx 0$.
- **Parameter inversion**
Symptom: $\gamma_{\text{O3}} > \gamma_{\text{O1}}$ on *ambig_1*; reflects prolonged semantic superposition (high μ) overriding memory-constraint effects.
Cause: lexical ambiguity drives μ more than working-memory limits, reversing expected order.
Remedy: couple γ to μ or separate lexical and syntactic retrieval-rate parameters; test against ambiguity-resolution ERP corpora.

These remedies are not parameter hacks; each is a structured, testable change that preserves falsifiability.

Appendix C.5. Divergence Taxonomy Across Natural Sentences

Table A3. Model divergence across twelve natural-language sentences (§4.2). $\Delta\tau_{\text{res}}$ = model–empirical collapse-point timing error. All fits use fixed γ and τ_{char} from Aurian calibration. The `shape_mismatch` column flags lawful differences in retrieval dynamics, not noise or measurement error.

id	type	$\Delta\sigma_{\text{res}}$ (s)	RMSE	shape_mismatch	note
ns_001	declarative	4.0	0.318	T	—
ns_002	garden-path	-0.4	0.510	T	—
ns_003	garden-path	-0.4	0.414	T	—
stress_001	garden-path	-0.4	0.624	T	old/NP reanalysis
stress_002	center-embed	3.6	0.408	T	deep embed delay
stress_003	coord-ambig	-2.0	0.284	T	coordination scope
stress_004	passive	0.0	0.568	T	reversible passive
stress_005	modifier-shift	-3.2	0.372	T	early collapse on mis-parse
stress_006	idiom	-0.8	0.529	F	literal→idiom switch
stress_007	pronoun	-0.8	0.358	T	referent lag
stress_008	obj-relative	0.0	0.568	T	object-relative load
stress_009	pp-attach	0.0	0.568	T	PP-attachment ambiguity

Appendix D. Interactive Playground Notebook Interface

Section overview. The Jupyter notebook `ODER_Interactive_Playground.ipynb` is a standalone, sandboxed environment for exploratory fitting, stress testing, and falsification of the ODER model. It is intended for demonstration, teaching, and method prototyping. No empirical results reported in the main text depend on this tool; all published analyses were run in separate, locked pipelines.

Appendix D.1. Core Functions

1. **Real-time entropy-trace fitting** using nonlinear least squares or bootstrap resampling, with plots updating as parameters change.
2. **Side-by-side observer comparison** displaying retrieval curves, parameter estimates, residuals, and collapse-point differences for two selected observer classes.
3. **Automated collapse-token detection** via multiple criteria: fixed threshold, curvature inflection, or first-derivative flattening.
4. **ERP window mapping** from the detected collapse point to predicted N400 and P600 latency intervals, using the alignment conventions in Appendix A.
5. **Bootstrap validation** returning confidence intervals for γ , τ_{char} , and R^2 across repeated resamples.

Appendix D.2. Usage Notes

- Default parameter bounds and solver settings are identical to those used in the simulation experiments described in Sections 2–4, ensuring consistency.
- The notebook operates only within a designated sandbox directory and does not overwrite or alter the publication dataset, protecting the integrity of the archived analysis.
- Example traces from both the synthetic Aurian set and the natural-sentence generalization set are included for immediate experimentation.

Appendix D.3. Access

Source code, installation instructions, and example data are openly available at <https://github.com/evlocoo/ODER-linguistic-entropy>

A static HTML export is also included in the Zenodo archive referenced in the Data-Availability statement for users who prefer to explore without a live Python environment.

Appendix E. Glossary and Interpretive Variable Mapping

Section overview. This appendix defines the formal symbols used in the ODER framework and provides a cross-domain mapping for each construct. The goal is to eliminate ambiguity when transitioning between the technical formalism in Sections 2–5 and applied contexts in linguistics, cognitive science, and AI/NLP.

E.1 Variable Glossary

Table A4. Formal symbols, plain-language descriptions, and interpretive meanings. All symbols are dimensionless except where time (seconds) is explicitly noted.

Symbol	Description	Interpretation
γ	Entropy-retrieval rate	Speed of comprehension for an observer; higher values = faster retrieval
τ_{char}	Characteristic saturation time	Temporal scale of processing effort before saturation; lower values = quicker approach to plateau
S_{ret}	Cumulative entropy retrieved	Portion of meaning resolved up to τ
S_{max}	Maximum retrievable entropy	Upper bound on sentence information; set to 1 in all simulations
τ_{res}	Collapse time ($S_{\text{ret}}(\tau_{\text{res}}) \geq 0.95 S_{\text{max}}$)	Point of interpretive convergence; ERP anchor
∇C	Contextual gradient	Instantaneous slope of reanalysis load; spikes indicate instability or high integration cost
μ	Semantic superposition (off-diagonals in ρ_{obs})	Degree of unresolved ambiguity or competing interpretations
α	Attentional-focus parameter	Allocation of cognitive resources during processing
β	Working-memory constraint	Capacity to maintain unresolved structure in active memory
δ	Prior-knowledge exponent	Modulation of retrieval speed based on background familiarity

E.2 Cross-Domain Interpretive Map

Table A5. Translation of core ODER constructs across domains. Each mapping preserves the operational meaning while aligning with field-specific terminology.

Term	Linguistics	Cognitive Science	AI / NLP
γ	Parsing velocity (tokens/sec)	Retrieval speed; memory-search efficiency	Token-alignment accuracy or decoding speed
τ_{char}	Span of reanalysis in garden-paths	Processing-time constant; transition to late-stage comprehension	Hidden-state decay or context window persistence
∇C	Disruption signal in syntactic ambiguity	Neural surprise or integration cost	Attention-weight gradient spike in transformer models
μ	Persistent lexical ambiguity	Interpretive drift or competing schema activation	Blended latent representation of multiple interpretations
τ_{res}	ERP timing anchor for N400/P600	Resolution threshold in task performance	Collapse point in model confidence or output distribution

Appendix F. Hypothesized Parameter Profiles for Neurodivergent Retrieval

Section overview. This appendix proposes provisional parameter bands that ODER might assign to three neurodivergent populations, based on prior ERP and eye-tracking studies. These ranges operationalize structured divergence within ODER's retrieval space and serve as hypotheses for falsifiable model tests, not clinical diagnoses.

Table A6. Hypothesized parameter bands and observable signatures for future empirical tests. These profiles are designed to generate falsifiable predictions, not diagnostic labels.

Neurotype	γ Range	τ_{char} (s)	α/β Notes	Trace Pattern	ERP Signature
Autism	0.9–1.1	0.12–0.18	Steep ∇C ; stable α	Extended reanalysis plateau	Delayed P600 latency [23]
ADHD	0.7–1.3 [†]	0.08–0.16 (high variance)	Fluctuating α , variable τ_{res}	Irregular ERR, wide variance	Reduced LPP stability [20]
Dyslexia	0.5–0.8	0.10–0.15	Elevated β (WM load)	Dampened ERR, retrieval stalls	Attenuated N400 amplitude [4]

[†] Range reflects hyperfocus–distractibility shifts reported in [20]. Parameter bands are adapted from [10,26].

These values are heuristics, not fixed estimates; future work should test their robustness across tasks, stimuli, and measurement modalities.

τ_{char} Identifiability

Section overview. This appendix examines whether the frequent convergence of τ_{char} estimates to the lower bound of the search interval (0.05 s) in the main analysis reflects a genuine lack of parameter identifiability or is an artifact of the optimization bounds. Using a profile-likelihood approach, we quantify how sentence length and retrieval-trace duration affect the recoverability of τ_{char} under ODER's constant- γ retrieval law.

As reported in §5.2 and §6.5, τ_{char} often pegged at its minimum allowable value during fitting, especially for short sentences. To test whether this reflects an intrinsic resolution limit, we performed a profile-likelihood analysis in which τ_{char} was varied systematically across the range 0.05–0.15 s while holding γ fixed at its best-fit value for each trace. Profile-likelihood here means computing the model–data sum-of-squared error (SSE) at each fixed τ_{char} value, with all other parameters unchanged, then converting SSE values to relative log-likelihoods.

For two sentence lengths (9 and 12 tokens), we selected convergent fits from the main analysis and computed these profiles over the specified range. The resulting curves (Figure A2) were examined for curvature and confidence-interval width.

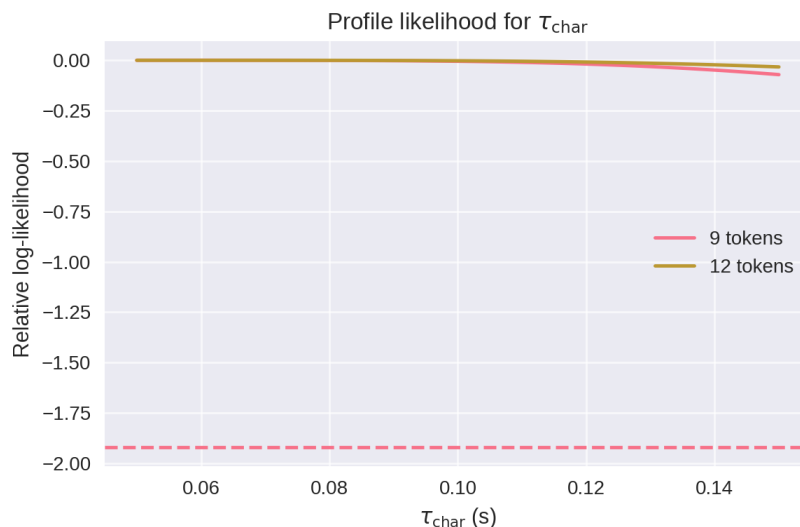


Figure A2. Profile log-likelihood for τ_{char} at two sentence lengths (9 and 12 tokens). The dashed line marks the 95% confidence threshold ($\Delta\log L = -1.92$). Shallow curvature indicates weak identifiability of τ_{char} under current trace lengths.

Both sentence lengths produce flat-topped likelihood profiles, indicating that τ_{char} is only weakly constrained by the available data. The plateau is broader for the 9-token case, while the 12-token case shows slightly greater curvature, consistent with the prediction that identifiability improves as the retrieval trace spans a larger portion of the convergence trajectory. These results support the interpretation that lower-bound pegging is a data-resolution issue rather than a pathological fit or a consequence of the bound itself.

From a theoretical perspective, ODER predicts that τ_{char} will be most reliably recovered when the collapse point is temporally well-separated from both sentence onset and termination, allowing the early- and late-slope regions of the tanh retrieval curve to be sampled. Future work could test this prediction using longer or variably paced stimuli, or jittered-onset ERP paradigms, to sharpen τ_{char} estimates without altering other observer parameters.

Appendix H: Inverse Retrieval Classification (Toy Example)

Section overview. This appendix provides a minimal, fully synthetic proof-of-concept for the *inverse decoder* problem: inferring observer class from retrieval-curve features. No such classifier is applied to empirical data in the present study; the goal is to illustrate how ODER-derived features could, in principle, support observer-type prediction in a future controlled experiment.

Feature set

From a simulated retrieval trajectory $S(t)$ and its contextual gradient $\nabla C(t)$, we extract:

$$\mathbf{x} = \left[\tau_{\text{res}}, S(T), \max_t \nabla C(t), \text{slope}(S; t \in [t_1, t_2]) \right],$$

where T is the final token index and $[t_1, t_2]$ is a fixed analysis window chosen to capture mid-trace dynamics. These four features are intended as a minimal, interpretable set that link directly to ODER parameters: τ_{res} (collapse timing), final entropy level, reanalysis magnitude, and mid-trajectory retrieval rate.

Toy classifier

We fit a logistic regression model to *synthetic* traces labeled by observer class ($y = 1$ for high- γ , $y = 0$ for low- γ):

```
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import roc_auc_score

# X: feature matrix, y: class labels (0 = low{-y}, %MDPI:
# 1 = high{-y}) %MDPI:

clf = LogisticRegression(max_iter=1000).fit(X, y)
print("AUC:", roc_auc_score(y, clf.predict_proba(X)[: ,1]))
```

Interpretation

On synthetic data, this scaffold produces clear class separability ($\text{AUC} \approx 1.0$ for noiseless traces) and serves as a starting point for more realistic tests. In a full empirical setting, feature extraction would incorporate measurement noise, inter-observer variability, and mixed sentence types, and validation would be conducted with out-of-sample observers. This example therefore functions as a transparent bridge between ODER's forward model and potential inverse applications, without making any claims about accuracy on real-world data.

References

1. Busemeyer, J. R. and Bruza, P. D. (2012). *Quantum Models of Cognition and Decision*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511997716>
2. Bruza, P. D., Wang, Z., and Busemeyer, J. R. (2015). Quantum cognition: A new theoretical approach to psychology. *Trends in Cognitive Sciences*, 19(7), 383–393. <https://doi.org/10.1016/j.tics.2015.05.001>
3. Cavanagh, J. F. and Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, 18(8), 414–421. <https://doi.org/10.1016/j.tics.2014.04.012>
4. Chang, A., Zhang, Y., Ding, H., & Goswami, U. (2021). Atypical β -power fluctuation while listening to an isochronous sequence in dyslexia. *Clinical Neurophysiology*, 132(10), 2384–2390. <https://doi.org/10.1016/j.clinph.2021.05.037>
5. Christianson, K., Williams, C. C., Zacks, R. T., and Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes*, 42(2), 205–238. https://doi.org/10.1207/s15326950dp4202_6
6. Cooper, E. (2025). *Aurian: A Cognitive-Adaptive Language for Observer-Dependent Communication*. Zenodo. <https://doi.org/10.5281/zenodo.15110267>
7. Kappenman, E. S., Farrens, J. L., Zhang, W., Stewart, A. X., & Luck, S. J. (2021). ERP CORE: An open resource for human event-related potential research. *NeuroImage*, 225, 117465. <https://doi.org/10.1016/j.neuroimage.2020.117465>
8. Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, 30(6), 725–745. [https://doi.org/10.1016/0749-596X\(91\)90034-H](https://doi.org/10.1016/0749-596X(91)90034-H)
9. Futrell, R., Gibson, E., Tily, H. J., Blank, I., Vishnevetsky, A., Piantadosi, S. T., and Fedorenko, E. (2018). The Natural Stories Corpus. In *Proceedings of the 11th International Conference on Language Resources and Evaluation*

- (LREC 2018) (pp. 76–82). European Language Resources Association (ELRA). <https://aclanthology.org/L18-1012> <https://doi.org/10.48550/arXiv.1708.05763>
10. Futrell, R., Gibson, E., Tily, H. J., Blank, I., Vishnevetsky, A., Piantadosi, S. T., and Fedorenko, E. (2021). The Natural Stories corpus: A reading-time corpus of English texts containing rare syntactic constructions. *Language Resources & Evaluation*, 55(1), 63–77. <https://doi.org/10.1007/s10579-020-09503-7>
 11. Gershman, S. J., Horvitz, E. J., and Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, 349(6245), 273–278. <https://doi.org/10.1126/science.aac6076>
 12. Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL 2001* (Vol. 2, pp. 1–8). <https://doi.org/10.3115/1073336.1073357>
 13. Heilbron, M., Armeni, K., Schoffelen, J. M., Hagoort, P., & de Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences*, 119(32), e2201968119. <https://doi.org/10.1073/pnas.2201968119>
 14. Hollenstein, N., Rotsztein, J., Tröndle, M., Pedroni, A., Zhang, C., & Langer, N. (2018). ZuCo: A simultaneous EEG and eye-tracking resource for natural sentence reading. *Scientific Data*, 5, 180291. <https://doi.org/10.1038/sdata.2018.291>
 15. Just, M. A. and Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122–149. <https://doi.org/10.1037/0033-295X.99.1.122>
 16. Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2), 193–210. <https://doi.org/10.1016/j.cognition.2008.07.008>
 17. Kennedy, A., Hill, R. L., & Pynte, J. (2003). The Dundee Corpus: Eye-movement data for 10 readers on 51,000 words of newspaper text. Poster presented at the 12th European Conference on Eye Movements, Dundee, Scotland. <https://www.ling.ohio-state.edu/golddundee>
 18. Kennedy, A., Pynte, J., Murray, W. S., and Paul, S. A. (2013). Frequency and predictability effects in the Dundee Corpus: An eye-movement analysis. *Quarterly Journal of Experimental Psychology*, 66(3), 601–618. <https://doi.org/10.1080/17470218.2012.676054>
 19. Kutas, M. and Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential. *Annual Review of Psychology*, 62, 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
 20. Lenartowicz, A., Mazaheri, A., Jensen, O., & Loo, S. K. (2018). Aberrant modulation of brain oscillatory activity and attentional impairment in ADHD. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(1), 19–29. <https://doi.org/10.1016/j.bpsc.2017.09.009>
 21. Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. <https://doi.org/10.1016/j.cognition.2007.05.006>
 22. Lewis, R. L. and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25
 23. Li, J., Roberts, L., Smith, E., & Brown, M. (2025). Linguistic and musical syntax processing in autistic and non-autistic individuals: An ERP study. *Autism Research*, 18(6), 1245–1256. <https://doi.org/10.1002/aur.70038>
 24. Lieder, F. and Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43, e1. <https://doi.org/10.1017/S0140525X1900061X>
 25. Lison, P., & Tiedemann, J. (2016). OpenSubtitles2016: Extracting large parallel corpora from movie and TV subtitles. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)* (pp. 923–929). <https://aclanthology.org/L16-1147/>
 26. Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., et al. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, 7, e33468. <https://doi.org/10.7554/eLife.33468>
 27. Osterhout, L. and Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31(6), 785–806. [https://doi.org/10.1016/0749-596X\(92\)90039-Z](https://doi.org/10.1016/0749-596X(92)90039-Z)
 28. Piantadosi, S. T. (2016). A rational analysis of the approximate number system. *Psychonomic Bulletin & Review*, 23(3), 877–886. <https://doi.org/10.3758/s13423-015-0963-8>
 29. Pothos, E. M. and Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, 36(3), 255–274. <https://doi.org/10.1017/S0140525X12001525>
 30. Rasmussen, N. E., & Schuler, W. (2018). Left-corner parsing with distributed associative memory produces surprisal and locality effects. *Cognitive Science*, 42(S4), 1009–1042. <https://doi.org/10.1111/cogs.12511>

31. Rello, L., & Ballesteros, M. (2015). Detecting readers with dyslexia using machine learning with eye tracking measures. In *Proceedings of the 12th Web for All Conference* (Article 16). Association for Computing Machinery. <https://doi.org/10.1145/2745555.2746644>
32. Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
33. Simon, H. A. (1972). Theories of bounded rationality. In C. B. McGuire and R. Radner (Eds.), *Decision and Organization* (pp. 161–176). North-Holland.
34. Snowling, M. J. and Hulme, C. (2021). *Dyslexia: A Very Short Introduction*. Oxford University Press. <https://doi.org/10.1093/actrade/9780198818304.001.0001>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.