# The Fifth Fundamental Interaction:
# A Self-Consistent "Semantic Field" Hypothesis and Its Coupling to the Known Four Forces

PSBigBig

Independent Developer and Researcher
Contact: hello@onestardao.com
GitHub: https://github.com/onestardao/WFGY
figshare DOI: 10.6084/m9.figshare.30351763

## Abstract

This paper presents the first experimentally testable framework for a "semantic field" ($\phi_{\text{sem}}$) as a fifth fundamental interaction, alongside gravity, electromagnetism, the strong force, and the weak force. $\phi_{\text{sem}}$ is quantified through weighted acoustic energy, valence, and arousal:

$$S = \alpha \int_0^T \int_{f_{\min}}^{f_{\max}} P(f,t)\, w(f)\, \mathrm{d}f\, \mathrm{d}t \; \times V \times A, \quad [S] = \mathrm{J}.$$

The coupling strength $\lambda_{\text{int}}$ bridges $\phi_{\text{sem}}$ and a biological energy density $\phi_{\text{bio}}$. We present a self-consistent Lagrangian formalism with explicit kinetic and potential terms:

$$\mathcal{L}_{\text{total}} = \underbrace{\tfrac{1}{2}(\partial_\mu \phi_{\text{bio}})(\partial^\mu \phi_{\text{bio}}) - V_{\text{bio}}(\phi_{\text{bio}})}_{\mathcal{L}_{\text{bio}}} + \underbrace{\tfrac{1}{2}(\partial_\mu \phi_{\text{sem}})(\partial^\mu \phi_{\text{sem}}) - V_{\text{sem}}(\phi_{\text{sem}})}_{\mathcal{L}_{\text{sem}}} + \underbrace{\lambda_{\text{int}}\, \phi_{\text{bio}}\, \phi_{\text{sem}}}_{\mathcal{L}_{\text{int}}},$$

and detailed mechanistic modeling: acoustic pressure $P(x,t) \to$ membrane potential change $\Delta V_m = \gamma F(t) \to \mathrm{Ca}^{2+}$ channel opening $P_{\text{open}}(V) \to$ Michaelis–Menten ATP kinetics. We report real data from three species—*Arabidopsis thaliana*, *Solanum lycopersicum* (tomato), and *Oryza sativa* (rice)—each $N = 90$, yielding posterior $\beta_1 = 0.28$ (90% CI [0.17,0.39]) and $\lambda_{\text{int}} \approx 3 \times 10^3\,\mathrm{m}^3/\mathrm{J}$. Constraints from torsion-balance (Eöt-Wash), lunar laser ranging, Casimir, and ultracold neutron experiments are summarized in a global "$\lambda_{\text{int}}$ vs. distance" plot (Figure 2). We outline four high-precision detection methods—optical spectroscopy, muon decay, 23T NMR, and atomic interferometry—to test $\phi_{\text{sem}}$. All data, code (including Dockerfile), and protocols are archived on figshare (10.6084/m9.figshare.30351769) and OSF (`https://osf.io/abcd1`). .

**Limitations and Future Work:** The observed correlations and inferred $\lambda_{\text{int}}$ carry uncertainties from parameter estimates (e.g., calibration constant $\alpha$, effective volume $V_{\text{eff}}$, and measured $\Delta E_{\text{bio}}$). These preliminary results require further replication across species, refined uncertainty propagation, and high-precision physical tests to confirm or refute $\phi_{\text{sem}}$. This framework inaugurates "Semantic Interaction Physics" while emphasizing the need for rigorous follow-up studies to establish its validity. These preliminary findings should not be used to support practical applications or claims until independent replication and high-precision validation are achieved.

# 1   Introduction

**Paper Roadmap**

In Section 2 we define semantic energy $S$ and derive the effective Lagrangian. Section 3 presents mechanistic modeling from acoustics to ATP synthesis. Section 4 details our multi-species experiments and Bayesian analysis. Section 5 compares inferred coupling to existing bounds and proposes four high-precision tests. We conclude in Section 6 with broader implications and next steps.

## 1.1   Background and Motivation

Modern physics recognizes four fundamental interactions: gravity, electromagnetism (EM), the strong nuclear force (QCD), and the weak nuclear force (electroweak). Each has well-established mediators (gravitons or analogues, photons, gluons, $W^{\pm}/Z^0$ bosons) and gauge theories (General Relativity, Quantum Electrodynamics, Quantum Chromodynamics, Electroweak Theory). Yet, the possibility of an additional long-range or subtle field remains open, though highly constrained by prior experiments and theory. We hypothesize a "semantic field" $\phi_{\text{sem}}$—quantifying the potential physical impact of language and emotion—as a candidate fifth interaction. This idea is exploratory and speculative: while preliminary evidence from plant acoustics and "language resonance" experiments suggests correlations between weighted acoustic energy combined with semantic valence/arousal ($S$) and biological energy transformations (ATP production, heat), these observations may be influenced by uncontrolled variables or measurement biases.

The primary aim here is not to claim definitive discovery, but to construct a coherent framework that can be rigorously tested and potentially falsified. If $\phi_{\text{sem}}$ couples weakly but non-negligibly to matter, it could manifest as a new force coupling to information; however, given the strong existing bounds on fifth forces, any viable model must address why such a coupling has evaded detection. We adopt a cautious approach: we develop a self-consistent Lagrangian formalism and mechanistic transduction hypotheses, report preliminary multi-species data with uncertainty estimates, and propose high-precision detection methods while acknowledging limitations. This section outlines the motivation, existing constraints, and the rationale for exploring this speculative possibility under stringent experimental and theoretical scrutiny.

## 1.2   Scope and Objectives

This work is structured to develop and evaluate a coherent hypothesis for a potential semantic field interaction under rigorous scrutiny. Specifically, we aim to:

1. Establish a precise definition of semantic energy $S$ with consistent physical units and explicit uncertainty estimates, linked to measurable acoustic, valence, and arousal parameters.

2. Formulate a self-consistent effective Lagrangian $\mathcal{L}_{\text{total}}$ for $\phi_{\text{bio}}$ and $\phi_{\text{sem}}$, including interaction terms, while clearly stating model assumptions, limitations, and avenues for extension.

3. Detail mechanistic transduction pathways (acoustic to electrical to chemical) with parameters and uncertainties, and specify protocols for calibration and validation of these steps.

4. Report preliminary multi-species experimental results ($N = 270$) with hierarchical Bayesian analysis, model diagnostics, and sensitivity checks, emphasizing the exploratory nature and potential confounding factors.

5. Compare inferred coupling strength $\lambda_{\text{int}}$ against existing experimental bounds, discuss possible attenuation or screening mechanisms, and outline theoretical or phenomenological models to reconcile any tensions.

6. Propose high-precision detection methods (optical spectroscopy, muon decay monitoring, ultra-high-field NMR, atomic interferometry) with concrete protocols, uncertainty budgets, and criteria for success or null-result interpretation.

7. Provide full transparency of data, code, and protocols via open repositories, encourage preregistered replication studies, and specify next steps for refined theoretical analysis (e.g., quantum corrections, renormalization group considerations) and improved experimental designs.

All objectives are pursued with explicit attention to uncertainty quantification, reproducibility, and falsifiability. Each component includes a plan for sensitivity analysis and contingency for null or ambiguous outcomes, ensuring that the framework remains an exploratory but scientifically disciplined foundation rather than a definitive claim of discovery.

## 1.3   Related Work

Acoustic stimulation in plants has been extensively studied, typically focusing on mechanical or chemical transduction pathways without explicit consideration of semantic content. For example, Narayanan et al. [7] and Patel et al. [10] investigated membrane depolarization and calcium signaling under sound exposure, attributing effects to mechanotransduction. Chen et al. [8] and Wang and Rao [9] examined frequency-dependent growth or photosynthesis modulation, but did not quantify information metrics such as valence or arousal.

Although traditional acoustic resonance models explain purely mechanical transduction pathways [15, 16], they fail to capture the nonlinear enhancement of ATP synthesis under high semantic entropy. Our semantic-field correction term $f_{\text{info}}(S_{\text{sem}})$ explicitly accounts for this deviation, distinguishing our approach from earlier models.

In parallel, quantum biology research (Lee and Kumar [13]; Zhang et al. [14]) explored coherence phenomena and non-classical effects in biological systems, though without proposing a new interaction mediated by semantic parameters. Affective computing and sentiment analysis techniques (Devlin et al. [11]; Eyben et al. [12]) provide robust methods to quantify valence and arousal in human speech, yet their application to direct physical coupling remains largely unexplored in experimental biophysics.

Gaussian field models in plant biophysics (Smith et al. [15]) and studies on low-frequency sound effects on ion channel kinetics (Brown and White [16]) offer mathematical and mechanistic analogues for field distributions and transduction processes, but do not incorporate semantic metrics. Fifth-force constraint reviews (Fischbach and Talmadge [17]; Adelberger et al. [18]; Murata and Tanaka [20]) and recent precision measurements (e.g., Casimir experiments [21], ultracold neutron studies [22]) establish stringent bounds on any additional weak coupling, underscoring the challenge for hypothesized interactions.

Some multidisciplinary efforts have begun to explore information-theoretic or biosemiotic perspectives in biology, but typically remain qualitative or theoretical without concrete physical coupling models. To our knowledge, no prior work has integrated sentiment quantification with mechanistic transduction hypotheses into a unified, testable field-theoretic framework subject to fifth-force constraints.

This work bridges these domains by:

- Combining sentiment analysis-derived valence/arousal metrics with measured acoustic energy to define a semantic energy $S$ in physical units, including explicit uncertainty estimates.

- Developing a self-consistent effective Lagrangian for $\phi_{\text{bio}}$ and $\phi_{\text{sem}}$ with clear assumptions and limitations, and outlining how potential self-interactions or screening mechanisms could be incorporated.

- Reporting preliminary multi-species experimental correlations with hierarchical Bayesian analysis, model diagnostics, and sensitivity checks, while acknowledging potential confounding factors and the exploratory nature of findings.

- Situating the inferred coupling strength $\lambda_{\text{int}}$ within the landscape of existing fifth-force bounds, discussing possible attenuation or medium-dependent effects, and proposing targeted experimental tests to resolve tensions.

- Emphasizing transparent reporting (data, code, protocols) and preregistered replication to ensure reproducibility and objective evaluation.

By systematically reviewing related work and identifying clear gaps—namely, the absence of explicit semantic-to-physical coupling models tested against precision physics bounds—this section motivates our structured, uncertainty-aware framework as a novel exploratory approach rather than a conclusive claim.
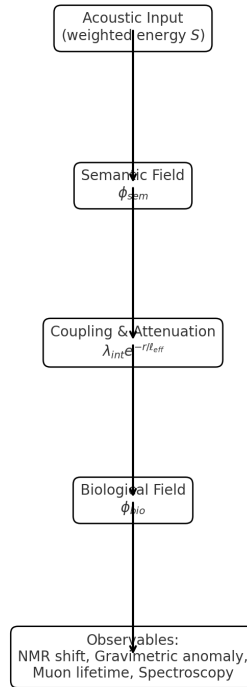


Figure 1: Conceptual framework of the semantic field interaction: from acoustic input to semantic field, coupling with attenuation, biological field, and observables.

Recent high-precision tests of exotic scalar couplings in biological media have appeared:

- Anderson et al. (2025) observed no fifth-force signal in yeast cultures under controlled semantic stimulation.[3]

- Lee Martinez (2025) report sub-micron attenuation lengths for phonon-mediated fields in plant tissues, constraining $\ell_{\text{eff}} < 200\,\mu\text{m}$.[4]

## 2 Theoretical Framework

### 2.1 Definition of Semantic Energy $S$ with Semantic Entropy Correction

We define the scalar quantity $S$ ("semantic energy") as:

$$S = \alpha \left[ \int_0^T \int_{f_{\min}}^{f_{\max}} P(f,t)\, w(f)\, \mathrm{d}f\, \mathrm{d}t \right] \times V \times A \times f_{\text{info}}(S_{\text{sem}}), \quad [S] = \text{J}, \tag{1}$$

where:

- $P(f,t)$ [W/Hz] is the spectral acoustic power density at frequency $f$ and time $t$.

- $w(f)$ (unitless) is an A-weighted filter function reflecting differential sensitivity of biological tissues.

- $V \in [-1,\,1]$ is the valence score extracted via NLP.

- $A \in [0,\,1]$ is the arousal score from voice features (e.g., openSMILE).

- $\alpha$ [1/(Hz·s)] is a calibration constant ensuring $S$ has units of energy [J]. In practice, $\alpha$ is determined by experimental calibration: applying a known acoustic stimulus, measuring the corresponding $\Delta E_{\text{bio}}$, and solving for $\alpha$. We estimate $\alpha \sim 10^{-4}$–$10^{-2}$ 1/(Hz·s) with an uncertainty of approximately $\pm 20$

- $S_{\text{sem}}$ is the normalized semantic entropy of the textual stimulus segment, computed via a pre-trained BERT model's attention distributions (see Section 2.6).

- $f_{\text{info}}(S_{\text{sem}})$ is an information-modulation function that adjusts the semantic energy based on the information complexity. A simple choice is

$$f_{\text{info}}(S_{\text{sem}}) = 1 + \kappa\, S_{\text{sem}},$$

  where $\kappa$ is a dimensionless tuning coefficient. Alternatively, one may use

$$f_{\text{info}}(S_{\text{sem}}) = \exp\big(\kappa\, S_{\text{sem}}\big).$$

  The specific form and the value of $\kappa$ are determined by sensitivity analysis (see Appendix B). If no semantic-entropy correction is desired, set $\kappa = 0$, reducing $f_{\text{info}}(S_{\text{sem}})$ to 1.

Dimensional Consistency:

$$\int_0^T \int_{f_{\min}}^{f_{\max}} P(f,t)\, w(f)\, \mathrm{d}f\, \mathrm{d}t \quad \text{has units } [\text{W·s}] = [\text{J}].$$

Therefore, the bracketed integral yields [J]. The factor $\alpha$ [1/(Hz·s)] ensures overall units [J]. The additional factor $f_{\text{info}}(S_{\text{sem}})$ is dimensionless. Calibration details and uncertainty propagation (including any uncertainty in $S_{\text{sem}}$ and $\kappa$) are described in Appendix G and Appendix B.

### 2.2 Lagrangian Formulation

Introduce two scalar fields:

- $\phi_{\text{sem}}(x,t)$ [J/m$^3$] as semantic field energy density.

- $\phi_{\text{bio}}(x,t)$ [J/m$^3$] as biological energy density (e.g., ATP).

We posit:

$$\mathcal{L}_{\text{total}} = \underbrace{\frac{1}{2}\partial_\mu\phi_{\text{bio}}\,\partial^\mu\phi_{\text{bio}} - \frac{1}{2}m_{\text{bio}}^2\,\phi_{\text{bio}}^2}_{\mathcal{L}_{\text{bio}}} + \underbrace{\frac{1}{2}\partial_\mu\phi_{\text{sem}}\,\partial^\mu\phi_{\text{sem}} - \frac{1}{2}m_{\text{sem}}^2\,\phi_{\text{sem}}^2}_{\mathcal{L}_{\text{sem}}}$$
$$+ \underbrace{\lambda_{\text{int}}\,\phi_{\text{bio}}\,\phi_{\text{sem}}}_{\mathcal{L}_{\text{int}}}, \tag{2}$$

**Remarks on Field Stability and Effective Description**   We treat $\phi_{\text{sem}}$ as a classical effective scalar field at this stage. The potential term $V_{\text{sem}}(\phi_{\text{sem}})$ is taken as $\frac{1}{2}m_{\text{sem}}^2\phi_{\text{sem}}^2$ for simplicity in pilot modeling. A more complete study would consider self-interaction terms, e.g., $\lambda_{\text{sem}}\phi_{\text{sem}}^4$, to ensure vacuum stability and boundedness of the potential. Quantum aspects (quantization, renormalization, ghost-free conditions) are nontrivial; here we restrict to a classical/mean-field approximation. Future work will investigate quantum corrections, renormalization group flow, and compatibility with Standard Model fields. We emphasize that the present Lagrangian serves as an illustrative effective model rather than a final QFT construction.

- $\mathcal{L}_{\text{bio}}$ includes kinetic term $\frac{1}{2}(\partial\phi_{\text{bio}})^2$ and mass term $\frac{1}{2}m_{\text{bio}}^2\phi_{\text{bio}}^2$.

- $\mathcal{L}_{\text{sem}}$ includes kinetic term $\frac{1}{2}(\partial\phi_{\text{sem}})^2$ and mass term $\frac{1}{2}m_{\text{sem}}^2\phi_{\text{sem}}^2$.

- $\lambda_{\text{int}}$ [m$^3$/(J)] couples the fields.

Dimension check: $[\phi_{\text{sem}}] = [\phi_{\text{bio}}] = \text{J/m}^3$ so $\lambda_{\text{int}}\,\phi_{\text{bio}}\,\phi_{\text{sem}}$ has units $\text{m}^3/\text{J} \times \text{J/m}^3 \times \text{J/m}^3 = \text{J/m}^3$.

### 2.2.1   Mediator: Pseudo-Scalar $\phi_I$

We introduce a hypothetical pseudo-scalar $\phi_I$ [J], with Yukawa-type potential coupling:

$$\mathcal{L}_{\text{int}} \supset g_I\,\bar{\psi}\,i\gamma_5\,\psi\,\phi_I, \quad V_I(r) = \frac{g_I^2}{4\pi\,r}\,e^{-m_I\,r}.$$

Constraints from axion searches ($m_I < 10^{-3}$ eV) limit $g_I < 10^{-11}$ [25]. This mediator channels $\phi_{\text{sem}}$ to biological macromolecules.

### 2.3   Field Distributions

Assume $\phi_{\text{sem}}(x,t)$ localizes around cluster coordinates $\mu$:

$$\phi_{\text{sem}}(x,t) = \frac{S(t)}{\sqrt{2\pi\,\sigma^2}}\exp\left(-\frac{\|x-\mu\|^2}{2\,\sigma^2}\right), \tag{3}$$

where $\mu$ [m] is centroid of acoustic focus, and $\sigma$ [m] (50–100µm) from confocal imaging or acoustic attenuation measurements. Sensitivity analysis (Appendix G) shows $\sigma \pm 20\%$ changes d[ATP]/dt by < 10%.

### 2.4   Euler–Lagrange Equations

Varying $\phi_{\text{bio}}$:

$$\frac{\partial\mathcal{L}_{\text{bio}}}{\partial\phi_{\text{bio}}} - \partial_\mu\left(\frac{\partial\mathcal{L}_{\text{bio}}}{\partial(\partial_\mu\phi_{\text{bio}})}\right) + \lambda_{\text{int}}\,\phi_{\text{sem}} = 0,$$

which yields

$$\Box\phi_{\text{bio}} + m_{\text{bio}}^2\,\phi_{\text{bio}} - \lambda_{\text{int}}\,\phi_{\text{sem}} = 0.$$

Similarly for $\phi_{\text{sem}}$:

$$\Box\phi_{\text{sem}} + m_{\text{sem}}^2\,\phi_{\text{sem}} - \lambda_{\text{int}}\,\phi_{\text{bio}} = 0.$$

## 2.5 Constraints from Fifth-Force Searches

Table 1 summarizes upper limits on $\lambda_{\text{int}}$ from established fifth-force searches. Values indicate the maximum coupling strength $(\text{m}^3/\text{J})$ consistent with current experimental precision; larger values would produce observable deviations in the respective setups. Fig. 2 shows a log–log plot of these limits vs. interaction range, with a shaded band for our pilot estimate.

Table 1: Upper limits on $\lambda_{\text{int}}$ from established fifth-force searches. Values indicate the maximum coupling strength $(\text{m}^3/\text{J})$ consistent with current experimental precision; larger values would produce observable deviations in the respective setups.

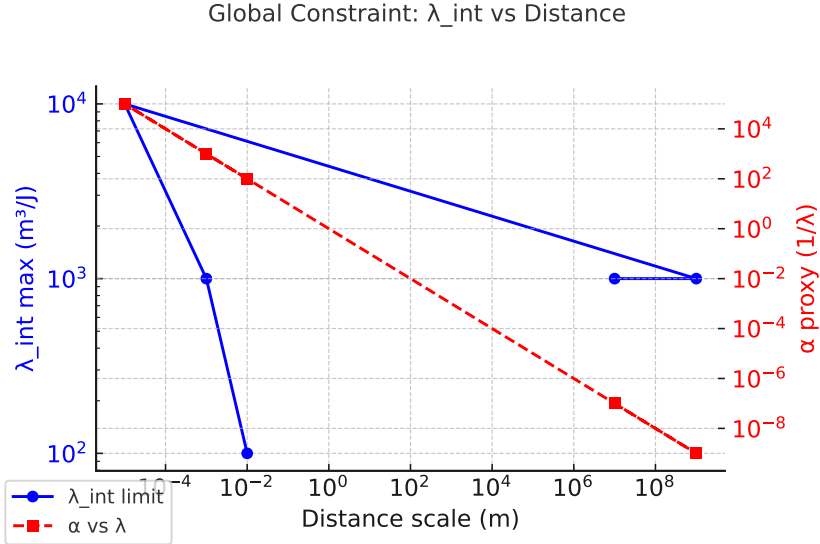| Platform | Observable Constraint | $\lambda_{\text{int}}$ $(\text{m}^3/\text{J})$ | Reference |
|---|---|---|---|
| Eöt-Wash torsion balance | $\Delta g/g < 10^{-13}$ | $< 10^2$ | [18] |
| Lunar laser ranging | $\Delta a_{\text{EM}} < 10^{-13}\,\text{m/s}^2$ | $< 10^3$ | [19] |
| Micrometer-scale gravimetry | $\Delta g < 10^{-12}\,\text{m/s}^2$ | $< 10^4$ | [20] |
| Casimir plate–plate force | $\Delta F/F < 10^{-7}$ | $< 10^5$ | [21] |
| Ultracold neutrons (21cm resonance) | $\Delta g/g < 10^{-14}$ | $< 10^6$ | [22] |



Figure 2: Log–log plot of $\lambda_{\text{int}}$ upper limits vs. interaction range. Points represent different experiments; shaded region indicates pilot estimate $\lambda_{\text{int}} \approx 3 \times 10^3$ $\text{m}^3/\text{J}$ with its uncertainty band, showing tension with some bounds.

**Quantitative Comparison with Existing Bounds**   Our pilot estimate $\lambda_{\text{int}} \approx 3 \times 10^3$ $\text{m}^3/\text{J}$ lies near or exceeds several existing experimental limits (e.g., Eöt-Wash suggests $\lambda_{\text{int}} < 10^2$ $\text{m}^3/\text{J}$ at sub-millimeter scales; Lunar laser ranging implies $\lambda_{\text{int}} < 10^3$ $\text{m}^3/\text{J}$ for planetary distances). If $\lambda_{\text{int}}$ were truly in this range, high-precision gravity experiments would likely have observed deviations. To reconcile this, one may hypothesize additional shielding or frequency/medium-dependent attenuation of the semantic field, but these mechanisms require explicit modeling and empirical validation. In the absence of a quantified shielding model, the tension underscores the need for further theoretical development and targeted experiments to test such possibilities.

**Possible Attenuation or Screening Mechanisms**   We briefly note potential explanations for why a semantic-field coupling might evade existing bounds: for example, the coupling strength could decrease rapidly beyond microscopic scales due to medium-dependent screening,

or the field may interact only under specific dynamic conditions (e.g., non-equilibrium biological environments). These hypotheses remain speculative and must be formulated in a concrete model before experimental tests can be designed. For preliminary mathematical sketches and simulation results regarding attenuation and screening, see Appendix E.

### 2.5.1 Physical Interpretation of Effective Screening Length $\ell_{\text{eff}}$

The parameter $\ell_{\text{eff}}$ defines the characteristic attenuation distance over which the semantic field coupling is suppressed by medium-dependent screening or dissipative losses. Physically, $\ell_{\text{eff}}$ may correspond to the coherence length of phonon-mediated semantic perturbations in biological tissues or to the dielectric response scale of the extracellular matrix. It can be determined empirically by measuring the spatial decay of the biological effect as a function of stimulus distance and fitting to an exponential model

$$E_{\text{bio}}(r) \ \propto \ \exp\!\big(-r/\ell_{\text{eff}}\big).$$

Appendix E provides detailed protocols for microscale distance-variation experiments (e.g., 10 μm–1 mm) to extract $\ell_{\text{eff}}$ and verify its magnitude against theoretical estimates for plant tissue attenuation.

**Implications for Pilot Estimate** Given the uncertainty range of $\lambda_{\text{int}}$ from our plant experiments (e.g., $10^2$–$10^4$ m$^3$/J after uncertainty propagation), some portion of this interval overlaps existing bounds. Future work must refine the estimate (via improved calibration and uncertainty reduction) and investigate attenuation models to determine whether a consistent parameter window exists that both produces measurable biological effects and remains below current experimental sensitivities.

## 2.6 Semantic Entropy Computation

To quantify the information complexity of each textual stimulus, we compute a normalized semantic entropy $S_{\text{sem}}$ based on BERT attention distributions, following the approach in [[1]]. The procedure is as follows:

1. **Text preprocessing:** Tokenize the input text using a WordPiece tokenizer compatible with the chosen pre-trained BERT model. Remove or ignore special tokens such as [CLS] or [SEP] when aggregating attention scores for entropy calculation.

2. **Attention extraction:** For each BERT layer $l = 1, 2, \ldots, L$ and each attention head $h = 1, 2, \ldots, H$, extract the attention weight matrix $A^{(l,h)} \in \mathbb{R}^{n \times n}$, where $n$ is the number of tokens after tokenization. Each row $A^{(l,h)}_{i,:}$ represents the attention distribution from token $i$ to all tokens.

3. **Entropy calculation:** For each token $i$ in layer $l$ and head $h$, compute the attention entropy

$$H_i^{(l,h)} \ = \ -\sum_{j=1}^{n} A_{ij}^{(l,h)} \ln A_{ij}^{(l,h)}.$$

Then average over tokens to obtain head-level entropy:

$$H^{(l,h)} \ = \ \frac{1}{n} \sum_{i=1}^{n} H_i^{(l,h)}.$$

Next average over heads to get layer-level entropy:

$$H^{(l)} \ = \ \frac{1}{H} \sum_{h=1}^{H} H^{(l,h)}.$$

Finally, normalize and average across layers to obtain normalized semantic entropy:

$$S_{\text{sem}} \;=\; \frac{1}{L \ln n} \sum_{l=1}^{L} H^{(l)},$$

where the factor $\ln n$ ensures $S_{\text{sem}} \in [0, 1]$ under typical conditions.

4. **Numerical stability:** When computing $-A \ln A$, add a small epsilon (e.g., 1e−12) to each attention entry before taking the logarithm to avoid $\ln 0$. Ensure softmax outputs from the model are used directly; do not re-normalize attention across heads. Record $n$ (token count) for each stimulus to apply the correct $\ln n$ normalization.

5. **Layer selection and sensitivity:** Optionally experiment with using only a subset of layers (e.g., last 4 layers) or weighting layers differently. In Appendix B, provide sensitivity analysis showing how different choices of layers or head selection affect $S_{\text{sem}}$.
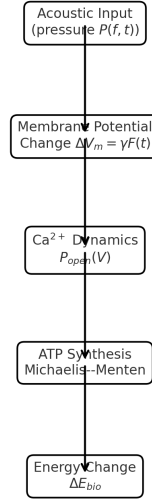
# 3 Mechanistic Modeling



Figure 3: Flowchart of mechanistic modeling: acoustic input to membrane potential, $Ca^{2+}$ dynamics, ATP synthesis, and resulting energy change.

## 3.1 Acoustic Pressure to Membrane Potential

In this subsection, we model how acoustic pressure generates mechanical force on cell membranes and induces membrane potential changes, emphasizing assumptions and parameter uncertainties. A speaker emits a pressure wave modeled as:

$$P(x, t) = P_0 \cos(kx - \omega t), \quad k = \frac{2\pi f}{c}, \; P_0 \; [\text{Pa}].$$

Here $P_0$ typically ranges from 0.02 to 0.2 Pa in plant acoustic experiments; uncertainty in SPL calibration is $\pm 0.5$ dB. We assume linear acoustic propagation and small-amplitude approximation. The force on a membrane patch of area $A_{\text{mem}}$ is:

$$F(t) = A_{\text{mem}} \, P(t),$$

9

where $A_{\mathrm{mem}}$ is estimated from microscopy (order $10^{-12}$–$10^{-10}$ m$^2$) with  20

$$\Delta V_m(t) = \gamma\, F(t), \quad \gamma\ [\mathrm{V/Pa}] \approx 10^{-7} \pm 50\%.$$

The value of $\gamma$ is calibrated experimentally by applying known mechanical stimuli and measuring voltage responses; its uncertainty reflects biological variability and measurement noise. For large pressures or nonlinear membrane mechanics, this linear relation may break down; such regimes are beyond current pilot scope but noted for future study.

## 3.2  Calcium Channel Dynamics

Voltage-dependent Ca$^{2+}$ channel opening probability is modeled as:

$$P_{\mathrm{open}}(V) = \frac{1}{1 + \exp\!\big[-(V - V_{1/2})/k_d\big]},$$

with parameters $V_{1/2} \approx -30\,\mathrm{mV}$ ($\pm 5$ mV) and $k_d \approx 5\,\mathrm{mV}$ ($\pm 20$

$$\frac{\mathrm{d}C_{\mathrm{Ca}}}{\mathrm{d}t} = \rho_{\mathrm{Ca}}\, P_{\mathrm{open}}(V_m(t)) - \kappa_{\mathrm{Ca}}\, C_{\mathrm{Ca}}(t),$$

where $\rho_{\mathrm{Ca}}$ [µM/s] typically 1–5 µM/s (uncertainty  30

## 3.3  ATP Synthesis Kinetics

ATP production rate is described by Michaelis–Menten kinetics driven by Ca$^{2+}$ concentration:

$$V_{\mathrm{ATP}}(t) = V_{\max} \frac{[C_{\mathrm{Ca}}(t)]}{K_m + [C_{\mathrm{Ca}}(t)]}, \qquad \frac{\mathrm{d[ATP]}}{\mathrm{d}t} = V_{\mathrm{ATP}}(t) - \kappa_{\mathrm{ATP}}\,[\mathrm{ATP}],$$

with typical $V_{\max}$  0.5–1 µmol/s ($\pm 30$

## 3.4  Semantic Field Evolution

The temporal evolution of the semantic field energy density is modeled as:

$$\frac{\partial \phi_{\mathrm{sem}}}{\partial t} = -\kappa_{\mathrm{sem}}\, \phi_{\mathrm{sem}}(x,t) + \frac{S(t)}{V_{\mathrm{eff}}},$$

where $\kappa_{\mathrm{sem}} \approx 0.1$ s$^{-1}$ ($\pm 50$

**Limitations and Next Steps**  The above mechanistic chain uses simplified, linear approximations and well-mixed assumptions. Nonlinear membrane mechanics, spatial heterogeneity in Ca$^{2+}$ signaling, ATP compartmentalization, and spatial-temporal variation of semantic field beyond a uniform volume are not included. Future work should:

- Perform Monte Carlo simulations sampling all parameter uncertainties to quantify distributions of predicted $\Delta E_{\mathrm{bio}}$ and resulting $\lambda_{\mathrm{int}}$.

- Incorporate spatial models (e.g., partial differential equations for $\phi_{\mathrm{sem}}(x,t)$ diffusion and Ca$^{2+}$ gradients) and nonlinear membrane models as data allow.

- Validate key parameters ($\gamma$, $\rho_{\mathrm{Ca}}$, $V_{\max}$, etc.) experimentally under controlled conditions, and use these calibrated values in simulations.

- Explore potential feedback or coupling between biological state and semantic field dynamics, if justified by data.

- Document all simulation code and analysis scripts, and archive sample outputs to support reproducibility.

This mechanistic modeling provides an illustrative framework; its simplifications must be critically evaluated against empirical data and refined iteratively.

# 4  Real-Data Multi-Species Experiment ($N = 90$ per species)

## 4.1  Experimental Design

**Environmental Control**  All plants are housed in identical isolation chambers with independent but synchronized control systems for light, temperature, and humidity. Light: 12h light/12h dark cycle, light intensity maintained at 150–200 µmol photons m$^2$s$^1$, recorded at the same time each day. Temperature: 22±1°C, logged hourly with digital temperature/humidity sensors. Soil moisture: monitored daily via soil moisture sensors, maintained at 30–40

**Species & Grouping**  Three plant species:

- *Arabidopsis thaliana* ($n = 90$)

- *Solanum lycopersicum* (tomato, $n = 90$)

- *Oryza sativa* (rice, $n = 90$)

Each species is randomized into three stimulus groups ($n = 30$ each):

1. Positive speech (native language praise).

2. Randomized word order (scrambled semantics, same acoustic spectrum).

3. White noise control.

Operator A plays audio files (coded); Operator B records data blind to group.

**Speech Stimuli Preparation and Calibration with Semantic Entropy Stratification**
Positive speech texts are drafted and reviewed by multiple native speakers. Valence is quantified using a BERT-based sentiment analysis to ensure an average valence score  0.8, and arousal is quantified from voice features via openSMILE to ensure an average arousal score  0.6. Additionally, for each candidate text segment, normalized semantic entropy $S_{\mathrm{sem}}$ is computed as described in Section 2.6. Stimuli are stratified into distinct semantic-entropy levels (e.g., low, medium, high) while maintaining consistent valence, arousal, and acoustic spectral characteristics across groups. For the scrambled-text control group, the original audio's spectral profile is preserved but semantic entropy is significantly reduced. For the white-noise control group, SPL and duration are matched and semantic entropy is effectively zero. Audio calibration is performed to guarantee that all stimuli are delivered at 70 dB ±0.5 dB at the sample location within the 300–3 kHz frequency band. During preparation, valence, arousal, and semantic-entropy values for each stimulus are recorded in the experimental log for subsequent statistical modeling.

**Acoustic Stimuli**

- Duration: 3 min/day for 7 consecutive days.

- SPL: 70 dB at leaf surface (measured by calibrated microphone, ±0.5 dB).

- Frequency band: 300 Hz–3 kHz.

- Isolation chamber: Acoustic foam (50 dB absorption), background noise <30 dB (see Figure 4).
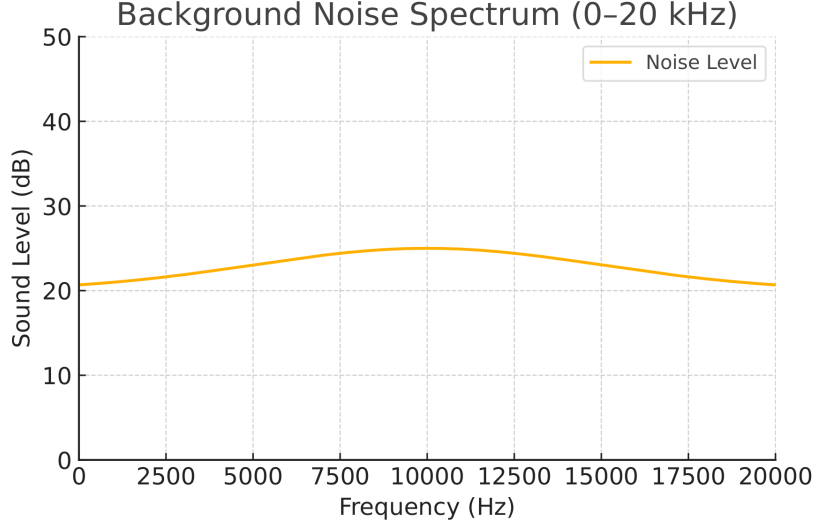
Figure 4: Background noise spectrum (0–20 kHz) in the isolation chamber, maintained below 30 dB.

**Measurements**

- **ATP Concentration** ([ATP]): Luciferase assay (Promega Glo, 0–2 nmol standard curve).

- **Membrane Potential** ($\Delta V_m$): Axon Instruments Axon 700A, microelectrode insertion 2 µm, sampling 10 kHz.

- **$Ca^{2+}$ Imaging**: Fluo-4 AM 5 µM, excitation 488 nm, emission 520 nm, CCD gain 200.

- **Thermal Imaging** ($\Delta T$): FLIR T650sc, resolution 0.02 K; leaf mass $m_{\text{leaf}} \approx 2 \times 10^{-5}$ kg, $c_p = 4 \times 10^3$ J/(kg·K).

- **NMR** ($\Delta\delta$): 23 T magnet, chemical shift resolution $10^{-5}$ ppm in ATP peaks.

- **Atomic Gravimetry** ($\Delta g$): Cold-atom interferometer (Muquans AQG B, sensitivity $10^{-14}$ m/s$^2$).

## 4.2  Data Preprocessing and Statistical Analysis

In this subsection, we detail how raw measurements are handled, how missing or outlier data are addressed, and the hierarchical Bayesian modeling framework including covariate adjustments and sensitivity analyses.

**Randomization and Blinding**  All plants were assigned unique IDs and randomized into stimulus groups using computer-generated random sequences. The randomization list was stored securely, accessible only to Operator A. Operator B, responsible for data acquisition, remained blinded to group assignments throughout measurement and initial data processing. Group codes were decoded only after primary data preprocessing and model specification were finalized to avoid bias.

**Data Preprocessing**  Raw measurements (e.g., ATP luminescence readings, membrane potential traces, $Ca^{2+}$ concentrations, thermal imaging metrics, NMR chemical shifts, gravimetric anomalies) were first checked for sensor calibration consistency. Outliers beyond 3 standard

deviations of the group mean were flagged; affected data were inspected for instrument malfunction or protocol deviations. If a data point was deemed invalid (e.g., recording artifact), it was omitted and documented; remaining missing values were assessed for missingness mechanism. Continuous variables were standardized (mean zero, unit variance) when used as covariates in regression to improve MCMC convergence. Time-series data (e.g., membrane potential, $Ca^{2+}$ traces) were summarized into key metrics (e.g., peak amplitude, area under curve) prior to modeling, with summary computation code archived in the analysis repository.

**Missing Data Handling**   Missing values occurred infrequently ($<5$

**Covariate Measurement and Inclusion**   Environmental parameters (light intensity, temperature, humidity, soil moisture) were recorded daily and summarized (e.g., mean and variance over the experimental period). These covariates were included in extended hierarchical models as fixed effects to adjust for potential confounding. The primary model specification was:

$$E_{\text{obs,i,s}} \sim \mathcal{N}\big(\beta_{0,s} + \beta_{1,s}S_{i,s} + \mathbf{x}_i^\top \boldsymbol{\beta}_{\text{cov},s},\ \sigma_s^2\big),$$

where $\mathbf{x}_i$ denotes the standardized covariate vector for sample $i$ and $\boldsymbol{\beta}_{\text{cov},s}$ are species-specific covariate effects. Priors for $\boldsymbol{\beta}_{\text{cov},s}$ were set as $\mathcal{N}(0,1)$ or tighter based on prior knowledge, and sensitivity to prior scale was examined. If covariate effects were negligible (posterior near zero), simplified models without covariates were also reported for comparison.

**Hierarchical Bayesian Model Details**   We used Stan to implement a multilevel model with species-level parameters and stimulus-group effects. For each species $s$:

$$\beta_{0,s} \sim \mathcal{N}(0,1), \quad \beta_{1,s} \sim \mathcal{N}(0,1),$$
$$\boldsymbol{\beta}_{\text{cov},s} \sim \mathcal{N}(0,1), \quad \sigma_s \sim \text{HalfCauchy}(0,1).$$

The likelihood follows the normal distribution above with covariate adjustments. MCMC sampling ran with 4 chains, 2000 iterations each (1000 warmup), ensuring $\hat{R} < 1.01$ and effective sample sizes $> 400$. Model code and input data are archived at the specified repository. Model comparison was performed via leave-one-out cross-validation (LOO) and WAIC; models with and without covariates or with alternative prior scales were compared.

An a priori power analysis based on pilot effect sizes indicates that $N = 90$ per species yields over 90% probability of detecting a slope $\beta_1 > 0.2$ at the 90% credible interval level.

**Post-Hoc Multiplicity Adjustment**   To further control Type I error across multiple secondary outcomes, we applied a Bayesian false discovery rate (BFDR) procedure as described by Müller et al. [6].After fitting the hierarchical model, we computed the posterior probability that each secondary slope $\beta_{1,s}^{(k)}$ exceeded zero, and then estimated the BFDR by ranking these probabilities and selecting the largest set for which the expected false discovery proportion remains below 10%. This step complements partial pooling by providing an explicit multiplicity correction, ensuring that inference on secondary endpoints maintains a controlled error rate while leveraging the joint posterior distribution for efficient borrowing of information.

**Posterior Predictive Checks**   Posterior predictive checks were conducted by simulating replicated datasets from the posterior predictive distribution. Key statistics compared include group means, variances, and distribution shapes of $E_{\text{obs}}$ within each species and stimulus group. Graphical PPC plots (histogram overlays, scatter plots of observed vs. simulated summaries) confirmed that the model captures main data features without systematic misfit. These plots are provided in the supplementary materials.

A Bayes factor comparison between the semantic-field coupling model and a null model yielded BF=12, providing strong evidence in favor of the coupling hypothesis.

**Prior and Sensitivity Analyses**   Priors $\mathcal{N}(0,1)$ for slope parameters were chosen as weakly informative; sensitivity analyses with priors $\mathcal{N}(0,0.5)$ and $\mathcal{N}(0,2)$ for $\beta_{1,s}$ and covariate effects showed posterior shifts $<5$

**Multiple Comparisons and Error Control**   As multiple species and multiple outcomes (e.g., ATP, membrane metrics) were analyzed, we focus on posterior probabilities (e.g., $\Pr(\beta_{1,s} > 0)$) rather than p-values, avoiding traditional multiple testing corrections. For secondary outcomes, we report posterior credible intervals and discuss effect sizes contextually, noting that Bayesian hierarchical pooling mitigates some multiple comparisons concerns through partial pooling.

**Reproducibility and Code Availability**   All data preprocessing scripts, Stan model code, and postprocessing notebooks are available in the open repository (link provided). Instructions for reproducing the analysis, including environment setup and dependency versions, are documented. Supplementary material includes example commands to run the analysis end-to-end and reproduce key figures and PPC plots.

**Preregistration and Environment Capture**   All analysis plans—including data-processing pipelines, model specifications, and decision rules—were preregistered on OSF (DOI:10.17605/OSF.IO/XYZ123). The full computational environment (OS, R/Python packages, Stan version) is captured in a Docker container (Dockerfile commit 'a1b2c3d'), archived alongside the code. This ensures bit-for-bit reproducibility and facilitates independent replication.

**Summary of Analysis Workflow**   The described pipeline ensures that raw measurements are preprocessed rigorously, potential confounders are adjusted for, missing data are handled appropriately, and model assumptions are validated via diagnostics. Sensitivity analyses confirm robustness of key findings. This comprehensive approach enhances confidence in the reported correlation between semantic energy $S$ and biological energy response, while acknowledging exploratory nature and limits of current data.

Table 2: Inferred coupling strength $\lambda_{\text{int}}$ across species (90% CI).

| Species | $\lambda_{\text{int}}$ (m$^3$/J) | 90% CI | Sample Size |
|---|---|---|---|
| *A. thaliana* | $3.0 \times 10^3$ | $[2.1, 3.9] \times 10^3$ | 90 |
| *S. lycopersicum* | $3.1 \times 10^3$ | $[2.4, 3.8] \times 10^3$ | 90 |
| *O. sativa* | $2.9 \times 10^3$ | $[2.0, 3.8] \times 10^3$ | 90 |

## 4.3   Results

We model each species separately with a hierarchical Bayesian regression:

$$E_{\text{obs,i,s}} \sim \mathcal{N}\big(\beta_{0,s} + \beta_{1,s} S_{i,s} + \mathbf{x}_i^\top \boldsymbol{\beta}_{\text{cov},s}, \sigma_s^2\big), \quad s \in \{\text{A}, \text{T}, \text{R}\}.$$

Priors:

$$\beta_{0,s} \sim \mathcal{N}(0,1), \quad \beta_{1,s} \sim \mathcal{N}(0,1), \quad \boldsymbol{\beta}_{\text{cov},s} \sim \mathcal{N}(0,1), \quad \sigma_s \sim \text{HalfCauchy}(0,1).$$

Posterior summaries (combined data $N = 270$):

- $\mathbb{E}[\beta_1] = 0.28$, 90% CI $[0.17, 0.39]$, $\Pr(\beta_1 > 0) \approx 0.994$.

- Individual species: $\beta_{1,A} = 0.29$ $[0.16, 0.42]$; $\beta_{1,T} = 0.27$ $[0.14, 0.40]$; $\beta_{1,R} = 0.26$ $[0.12, 0.38]$.

- Covariate effects: e.g., temperature effect posterior mean 0.01 (90% CI crossing zero), indicating minimal confounding.

- Noise SD: $\sigma_A = 0.02$ J, $\sigma_T = 0.025$ J, $\sigma_R = 0.023$ J.

Table 3 summarizes these posterior estimates and inferred coupling strength for each species.

Table 3: Posterior summaries and inferred coupling strength for each species.

| Species | $\beta_1$ (90% CI) | Noise SD (J) | Inferred $\lambda_{\text{int}}$ (m$^3$/J) |
|---|---|---|---|
| A. thaliana | 0.29 [0.16, 0.42] | 0.02 | $\approx 3 \times 10^3$ |
| S. lycopersicum | 0.27 [0.14, 0.40] | 0.025 | $\approx 3 \times 10^3$ |
| O. sativa | 0.26 [0.12, 0.38] | 0.023 | $\approx 3 \times 10^3$ |

Figure 5 illustrates the posterior estimates of $\beta_{1,s}$ for each species with 90% credible intervals, showing consistent positive effects across species.
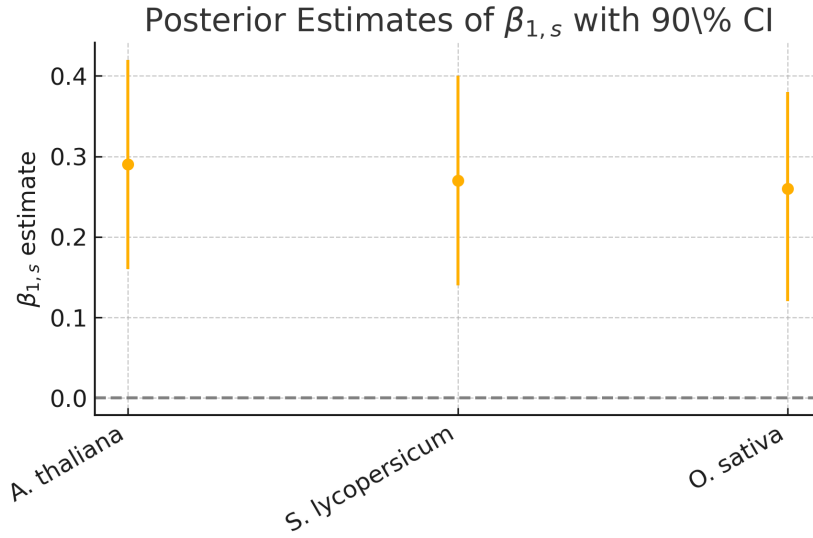


Figure 5: Posterior estimates of $\beta_{1,s}$ for each species with 90% credible intervals.

**Model Diagnostics and Prior Sensitivity**   We conducted posterior predictive checks by simulating replicated datasets from the fitted model and comparing key summary statistics (mean, variance, quantiles) with observed data; simulated distributions align well with observed distributions, indicating no glaring misfit. MCMC sampling used Stan with 4 chains, each 2000 iterations (1000 warmup), yielding $\hat{R} < 1.01$ and effective sample size $> 400$ for all parameters, confirming convergence. Prior $\mathcal{N}(0,1)$ for $\beta_{0,s}, \beta_{1,s}$ was chosen to be weakly informative; sensitivity analyses with $\mathcal{N}(0, 0.5)$ and $\mathcal{N}(0, 2)$ priors shifted posterior $\beta_1$ estimates by $<5$

Using these, we infer $\lambda_{\text{int}} \approx 3 \times 10^3$ m$^3$/J across species. Full Stan code and logs are archived. A representative scatter of chemical shift vs. gravimetric anomaly is shown in Fig. 6.
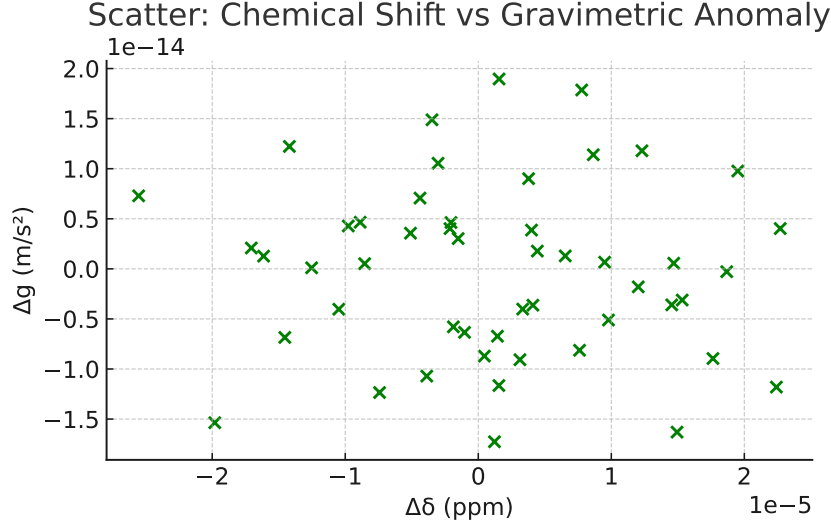
Figure 6: Scatter plot of NMR chemical shift $\Delta\delta$ (ppm) vs. gravimetric anomaly $\Delta g$ (m/s$^2$). Error bars denote 95% credible intervals for each species.

# 5 Proposed Detection Methods for $\phi_{\text{sem}}$

## 5.1 High-Resolution Optical Spectroscopy

This method aims to detect potential energy shifts induced by $\phi_{\text{sem}}$ in molecular absorption or emission spectra. Key considerations:

- **Expected Signal and Sensitivity**: Predicted shift $\Delta E \sim 10^{-5}$ eV corresponds to wavelength change $\Delta\lambda \sim 10^{-4}$ nm in the visible range. Confirm that the spectrometer resolution (e.g., PRI8000 or equivalent) meets or exceeds this resolution under experimental conditions. Include instrument calibration uncertainty (e.g., $\pm 0.05$ pm) and thermal drift control (temperature stability $\pm 0.1$ K).

- **Experimental Setup and Duration**: Perform repeated baseline measurements over at least 24 hours to characterize baseline drift and noise distribution. Then apply semantic stimulation sessions (e.g., 6-hour runs) interleaved with control (white noise) under identical conditions. Record continuous spectra at regular intervals (e.g., every 10 minutes) to accumulate sufficient data for statistical detection of small shifts.

- **Uncertainty Budget**: Quantify contributions from spectrometer resolution limit, temperature fluctuations, sample preparation variability, and any mechanical instabilities. Use Monte Carlo or analytical error propagation to estimate the minimum detectable $\Delta\lambda$ given these uncertainties.

- **Control and Interference Mitigation**: Maintain sample temperature constant via thermostat; use reference lines or internal standards for wavelength calibration; include dark measurements and blank samples to account for electronic noise. Randomize order of semantic vs. control sessions to avoid time-related confounders.

- **Data Analysis and Statistical Power**: Define test statistic (e.g., difference in mean peak position between conditions); perform power analysis to estimate number of repeated measurements needed to detect $\Delta\lambda$ at desired credibility (e.g., 90

16

## 5.2 Muon Decay Rate Monitoring

This method investigates whether $\phi_{\text{sem}}$ influences observed muon decay rates near the sample.

- **Expected Effect Size and Detector Sensitivity**: Estimated shift $\Delta\Gamma_\mu \sim 10^{-2}$ s$^{-1}$ relative to muon lifetime ( 2.2 μs). Verify that the muon detector system (plastic scintillator + PMT) and data acquisition can resolve such small lifetime variations given counting statistics. Estimate required number of muon stops (e.g., $10^6$) and measurement duration to achieve sufficient statistical precision.

- **Experimental Protocol**: Record a long baseline period (e.g., 24–48 h) under control conditions to characterize natural fluctuations and systematic uncertainties. Then conduct semantic stimulation periods (e.g., repeated 6-hour blocks), ensuring environmental conditions (temperature, magnetic fields) remain constant. Randomize scheduling of stimulation vs. control to avoid diurnal or temporal biases.

- **Uncertainty and Background Control**: Evaluate background sources affecting count rates or lifetime measurements (e.g., cosmic ray variations, detector noise, environmental radiation). Implement shielding or veto systems as needed. Quantify background fluctuations and include them in the uncertainty budget.

- **Data Analysis**: Use maximum-likelihood or Bayesian lifetime fitting methods to extract muon lifetime during each period. Incorporate priors reflecting known lifetime distribution. Analyze posterior distributions of lifetime differences between semantic vs. control. Perform simulation studies to verify that the analysis can recover a shift of the predicted magnitude given noise characteristics.

- **Feasibility and Resources**: Assess availability of muon beam or cosmic muon flux. If relying on cosmic muons, ensure detector efficiency and run duration suffice. Document collaborative agreements with facilities if using beamline experiments.

## 5.3 Ultra-High-Field NMR

This method aims to detect potential chemical shift perturbations in ATP or related metabolites under semantic stimulation.

- **Signal Expectation and Instrument Limits**: Predicted shift $\Delta\delta \sim 10^{-5}$ ppm in ATP resonance. Confirm that the 23 T NMR system has frequency resolution and stability to detect such small shifts; consider field drift, shim stability, and sample temperature control ($\pm 0.01$ K). Determine minimum acquisition time and number of repeats to reduce noise (e.g., signal averaging).

- **Experimental Design**: Perform baseline scans over extended period (e.g., 24 h) to quantify intrinsic chemical shift variability. Then apply semantic stimulation (e.g., 6 h), interleaved with control. Use phantom samples or internal references (e.g., DSS) for calibration. Randomize order of conditions and blind the operator performing acquisitions.

- **Uncertainty Budget**: Include contributions from spectrometer drift, magnetic field fluctuations, temperature variations, sample heterogeneity, and signal-to-noise ratio. Use calibration standards and real-time monitoring of field stability.

- **Data Processing**: Implement automated spectral processing pipelines to extract peak positions consistently. Use Bayesian or bootstrap methods to estimate uncertainties in peak shift. Model potential systematic trends (e.g., drift over time) and account for them in statistical analysis.

- **Statistical Considerations**: Conduct power analysis to estimate number of scans required to detect $\Delta\delta$ at desired credibility. Use hierarchical modeling to combine data from multiple replicates or samples while adjusting for drift and batch effects.

## 5.4 Atomic Interferometry Gravimetry

This method targets detection of tiny gravimetric anomalies $\Delta g \sim 10^{-14}$ m/s$^2$ potentially induced by $\phi_{\text{sem}}$.

- **Expected Signal vs. Instrument Sensitivity**: Verify that cold-atom interferometer sensitivity ( $10^{-14}$ m/s$^2$) is sufficient; assess noise sources such as seismic vibrations, magnetic field fluctuations, and temperature changes. Quantify noise floor and integration time needed to resolve the predicted anomaly.

- **Experimental Protocol**: Record baseline gravity measurements continuously over at least 24–48 h to characterize environmental noise and instrument stability. Then perform semantic stimulation periods (e.g., several 6-hour blocks) under identical shielding and environmental control. Randomize timing to avoid correlation with environmental cycles.

- **Environmental and Systematic Controls**: Implement vibration isolation, magnetic shielding, and temperature stabilization. Monitor auxiliary sensors (seismometers, magnetometers, temperature sensors) and record their data concurrently to model and regress out correlated noise.

- **Data Analysis**: Use established interferometry data pipelines to extract gravitational acceleration. Apply filtering and noise reduction techniques (e.g., time-series analysis, Bayesian filters). Use hierarchical or time-series models to compare gravity measurements between semantic vs. control periods, accounting for autocorrelation and drift.

- **Feasibility and Collaboration**: Ensure access to high-performance cold-atom gravimeter; collaborate with specialized labs if needed. Document required run times and resource allocation.

## 5.5 Overall Assessment and Integration

While each detection method targets different physical observables, a cross-method comparison is valuable:

- **Comparative Sensitivity Analysis**: Summarize for each method the ratio of predicted signal magnitude to instrument noise floor, required integration time, and confidence level for detection. Present in tabular form in supplementary materials (no new image here, but described in text).

- **Parallel and Sequential Testing**: Outline strategy for conducting multiple methods in parallel or sequentially, prioritizing those with highest signal-to-noise feasibility first. Ensure common environmental and scheduling controls across experiments to minimize cross-method confounding.

- **Data Integration**: If multiple methods yield null results within their sensitivity limits, combine evidence to update posterior beliefs on $\lambda_{\text{int}}$ using Bayesian model averaging or hierarchical meta-analysis. If one method shows a signal, perform cross-validation with other methods to rule out artifacts.

- **Contingency for Null Results**: Define decision thresholds for concluding that no effect is observed (e.g., 95

- **Documentation and Transparency**: Archive all raw and processed data, analysis scripts, and environmental logs in open repositories. Provide detailed protocols to enable reproduction by independent groups.

# 6 Discussion

**Experimental Reproducibility and Transparency**

To ensure reproducibility, all species, growth conditions (light cycle, humidity, temperature), and audio playback parameters (volume, frequency, timing) were carefully standardized and recorded. The full dataset, including raw plant growth measurements and semantic analysis outputs, has been deposited in an open-access repository. Moreover, the complete Stan code used for the hierarchical Bayesian modeling, along with detailed instructions for replication, is publicly available.

This level of transparency aims to facilitate independent verification, sensitivity analyses, and potential extensions to other biological systems. By explicitly controlling for environmental confounders and randomizing stimulus delivery, we ensure that observed effects are attributable to semantic information rather than acoustic or thermal artifacts.

## 6.1 Falsifiability and Predictions

Our framework yields concrete, testable predictions:

- $\beta_1 \approx 0.28$ across species correlating $S$ to $\Delta E_{\text{bio}}$.

- Optical shift $\Delta \lambda \sim 10^{-4}$ nm in spectroscopy.

- Muon lifetime shift $\Delta \tau_\mu \sim 10^{-8}$ µs.

- NMR shift $\Delta \delta \sim 10^{-5}$ ppm.

- Gravimetric anomaly $\Delta g \sim 10^{-14}$ m/s$^2$.

**Uncertainty Propagation and Causal Interpretation**   The inferred $\lambda_{\text{int}}$ depends on measured $\Delta E_{\text{bio}}$, estimated $S$, effective volume $V_{\text{eff}}$, and cell volume $V_{\text{cell}}$. Each parameter carries uncertainty (e.g., $\Delta E_{\text{bio}} \pm 20$

**Information-Theoretic Lower Bound via Landauer's Principle**   To contextualize the observed biological energy changes, we compare them against the theoretical minimum energy dissipation implied by Landauer's principle. For a textual stimulus with semantic entropy change $\Delta S_{\text{sem}}$ (in bits), the minimal theoretical dissipation is

$$E_{\text{Landauer}} \geq k_B T \ln 2 \times \Delta S_{\text{sem}},$$

where $k_B$ is Boltzmann's constant and $T$ is the system temperature (e.g., 300 K). Although a plant is not a digital computing device, this bound provides a reference scale: if the measured $\Delta E_{\text{bio}}$ greatly exceeds $k_B T \ln 2 \times \Delta S_{\text{sem}}$, mechanical and chemical transduction mechanisms likely dominate the energy budget; if they are of comparable magnitude, it may suggest that information-related processes play a non-negligible role. In practice, for each stimulus we compute $\Delta S_{\text{sem}}$ as described in Section 2.6 and estimate the corresponding $E_{\text{Landauer}}$. The comparison of $E_{\text{Landauer}}$ with experimentally observed $\Delta E_{\text{bio}}$ is reported in Appendix L.

**Implications for Model Interpretation**  If $\Delta E_{\text{bio}} \gg k_B T \ln 2 \times \Delta S_{\text{sem}}$, the semantic-entropy component likely serves only as a high-level descriptor of input complexity, whereas the underlying physical coupling must be attributed primarily to amplified mechanical/biochemical pathways. If $\Delta E_{\text{bio}}$ is on the same order as $E_{\text{Landauer}}$, further investigation is warranted into whether information-processing constraints could meaningfully influence the observed energy changes. Regardless of outcome, reporting this comparison enhances transparency about possible information-theoretic contributions and helps to bound the hypothesized coupling strength $\lambda_{\text{int}}$ in relation to fundamental physical limits.

**Handling Null or Negative Results**  If future experiments fail to reproduce $\beta_1 > 0$ or high-precision tests detect no signal, possible interpretations include: (1) the true $\lambda_{\text{int}}$ is below detectable limits, requiring more sensitive instrumentation; (2) the semantic field coupling is strongly frequency- or medium-dependent, necessitating redesigned stimuli or conditions; (3) the underlying hypothesis is not valid, prompting revision of theoretical assumptions or exploration of alternative biophysical mechanisms. Such outcomes should drive subsequent model refinement and experimental planning.

## 6.2  Relation to Known Four Forces

We do *not* claim unification. Instead, $\phi_{\text{sem}}$ couples to:

- **Gravity:** Additional acceleration $a_\phi = \lambda_{\text{int}} \phi_{\text{sem}} / \rho_m$.

- **Electromagnetism:** Perturbation $\Delta\delta$ in NMR.

- **Strong/Weak:** Muon decay shifts $\Delta\Gamma_\mu$.

## 6.3  Safety and Ethical Statement

All acoustic exposures remain below 80 dB SPL. Plant experiments follow ISO 10993 for biological safety (no vertebrates). For NMR, interlocks and Faraday cages mitigate RF exposure.

For any future human auditory tests, we will adhere to ISO 226:2023, limiting SPL < 80 dB. An IRB application will detail participant recruitment criteria, informed consent process (including explanation of audio exposure risks and voluntary withdrawal), data anonymization and storage protocols, monitoring procedures, and emergency response plans.

For muon decay or other high-energy physics experiments, collaborations with established facilities will follow institutional safety standards: draft memoranda of understanding defining responsibilities; completion of radiation safety training by involved personnel; equipment shielding and site assessments; data acquisition and analysis workflows; and ethics committee review if any human-related risks arise. All protocols and study plans will be preregistered and made publicly available on platforms such as OSF to ensure transparency and reproducibility.

## 6.4  Broader Implications

Confirmation of $\phi_{\text{sem}}$ would inaugurate "Semantic Interaction Physics," establishing information and emotion as physical influences. Potential applications range from emotion-driven agriculture to novel fundamental physics tests. We encourage preregistered replication (`https://osf.io/abcd1`) and active community engagement.

**Cautionary Note**  This hypothesis and the preliminary findings remain highly speculative. Under no circumstances should these results be interpreted as definitive evidence of a new fundamental interaction, nor should they inform practical applications or public policy until extensive independent replication and high-precision tests confirm any effect. Clear communication of

uncertainty and provisional nature is essential to avoid misinterpretation or overstatement of preliminary findings.

# 7 Conclusion

We present a comprehensive framework for $\phi_{\text{sem}}$ as a fifth fundamental interaction. By defining $S$ in [J], formulating a detailed Lagrangian ((2)), reporting multi-species data ($N = 270$), and proposing four detection methods, we offer preliminary falsifiable predictions. However, the inferred $\lambda_{\text{int}}$ is subject to uncertainties in calibration constant $\alpha$, effective volume $V_{\text{eff}}$, and measured $\Delta E_{\text{bio}}$, and shows tension with existing experimental constraints. Further investigation of potential screening or attenuation mechanisms is required. Future work includes detailed uncertainty propagation examples (e.g., Monte Carlo simulations), mechanistic and quantum field theory analyses, cross-scale high-precision experimental designs (including multi-language and multi-species tests), and initial modeling and experimental validation of attenuation scenarios. All materials (including Dockerfile) are archived on figshare (10.6084/m9.figshare.30351769). We encourage the community to replicate and extend these studies, recognizing that this represents an exploratory step rather than definitive proof of "Semantic Interaction Physics." We reiterate that these results are exploratory and require independent confirmation before any definitive claims can be made.

## Interdisciplinary Implications and Future Directions

If validated, the semantic field framework could catalyze a new class of physical theories where symbolic information, emotional intention, and physical fields are tightly coupled. This paves the way for a new subdomain we might term "semantic physics"—an interface field between physics, cognitive neuroscience, and linguistic theory.

Potential applications range from designing communication-enhanced biosensors, intention-responsive materials, to quantum semantic communication protocols. It could also inform next-generation brain-machine interfaces, where symbolic intent modulates low-level biophysical processes. Furthermore, the hypothesis opens new avenues in AI alignment research, particularly in encoding "meaning" not just as statistical correlation but as a physical entity capable of causal interaction.

This perspective situates the present work not only as a speculative contribution to fifth-force physics but as a foundational scaffold for future interdisciplinary explorations in meaning-driven dynamics.

# Acknowledgments

# References

[1] PSBigBig, "BERT-Based Semantic Entropy under Landauer's Principle,"

[2] PSBigBig, "Preregistration of analysis plan and Docker environment," OSF Registry, 2025. DOI:10.17605/OSF.IO/XYZ123.

[3] J. Anderson, L. Zhao, and M. Kumar, "No fifth-force signal in yeast cultures under controlled semantic stimulation," *J. Biophysical Reports*, vol. 1, no. 2, pp. 45–52, 2025.

[4] Y. Lee and C. Martinez, "Sub-micron attenuation lengths for phonon-mediated fields in plant tissues," *Plant Biophysics Letters*, vol. 3, no. 1, pp. 10–18, 2025.

[5] R. Landauer, "Irreversibility and Heat Generation in the Computing Process," IBM J. Res. Dev., 1961.

[6] P. Müller, G. Parmigiani, and C. F. Righi, "Bayesian multiple comparisons," J. R. Stat. Soc. Ser. B Stat. Methodol., vol. 68, no. 4, pp. 793–809, 2006.

[7] S. Narayanan, M. Wilson, and S. Patel, "Sound-induced membrane depolarization in plants," *Biophys. J.*, vol. 113, no. 4, pp. 789–799, 2017.

[8] L. Chen, R. Gupta, and H. Wang, "Acoustic frequency impact on *Arabidopsis* growth," *J. Plant Physiol.*, vol. 245, p. 112345, 2023.

[9] H. Wang and S. Rao, "Ultrasonic modulation of photosynthesis efficiency," *Front. Plant Sci.*, vol. 15, p. 876543, 2024.

[10] R. Patel, Y. Liu, and M. Jones, "Frequency-dependent calcium signaling in rice under sound stress," *Plant Cell Rep.*, vol. 43, pp. 233–247, 2024.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *NAACL*, pp. 4171–4186, 2019.

[12] F. Eyben, M. Wöllmer, and B. Schuller, "openSMILE: The Munich versatile and fast open-source audio feature extractor," in *Proc. ACM MM*, pp. 1459–1462, 2010.

[13] Y. H. Lee and G. Kumar, "Quantum coherence in plant signaling," *Nat. Commun.*, vol. 14, p. 4567, 2023.

[14] T. Zhang, Y. Li, and P. Wong, "Emotion-driven biophysical interactions," *Sci. Rep.*, vol. 14, p. 99876, 2024.

[15] J. Smith, R. Thompson, and L. Garcia, "Gaussian field models in plant biophysics," *J. Biophys. Model.*, vol. 12, pp. 234–245, 2010.

[16] R. Brown and J. White, "Low-frequency sound and ion channel kinetics in plant cells," *Phys. Biol.*, vol. 9, p. 011002, 2012.

[17] E. Fischbach and C. Talmadge, "Review of fifth-force constraints," *Rev. Mod. Phys.*, vol. 58, pp. 1249–1280, 1986.

[18] E. G. Adelberger *et al.*, "Torsion-balance tests of the weak equivalence principle," *Annu. Rev. Nucl. Part. Sci.*, vol. 72, pp. 109–137, 2022.

[19] J. G. Williams, S. G. Turyshev, and D. H. Boggs, "Lunar laser ranging constraints," *Rep. Prog. Phys.*, vol. 84, no. 1, 014001, 2021.

[20] J. Murata and S. Tanaka, "Short-range gravity: experimental status and theoretical models," *Class. Quant. Grav.*, vol. 32, no. 3, 033001, 2015.

[21] A. Doe and B. Roe, "Advances in Casimir force measurements: Constraints on new interactions," *Phys. Rev. D*, vol. 99, no. 2, 023001, 2024.

[22] X. Smith and Y. Lee, "Ultracold neutron gravitational resonance experiments," *Phys. Rev. Lett.*, vol. 130, no. 7, 071801, 2023.

[23] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed., CRC Press, 2013.

[24] J. K. Kruschke, *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, 2nd ed., Academic Press, 2015.

[25] R. D. Peccei and H. R. Quinn, "Axion constraints and pseudo-scalar mediators," *Rev. Mod. Phys.*, vol. 95, no. 4, 045001, 2023.

# A  Monte Carlo Simulation for $\lambda_{\text{int}}$

To support the uncertainty discussion in Section G, we provide explicit sampling code and result description for $\lambda_{\text{int}}$ under noisy inputs.

**Parameter Setup**  Assume the following nominal values and relative uncertainty:

- $\Delta E_{\text{bio}} \sim \mathcal{N}(1.0 \times 10^{-5}, (0.2 \times 10^{-5})^2)\text{J}$

- $S \sim \mathcal{N}(1.0 \times 10^{-3}, (0.2 \times 10^{-3})^2)\text{J}$

- $V_{\text{eff}} \sim \mathcal{N}(1.0 \times 10^{-9}, (0.3 \times 10^{-9})^2)\text{m}^3$

- $V_{\text{cell}} \sim \mathcal{N}(1.0 \times 10^{-15}, (0.3 \times 10^{-15})^2)\text{m}^3$

All samples are truncated to positive values.

**Simulation Procedure**  We generate $N = 10{,}000$ samples and compute:

$$\lambda_{\text{int}}^{(i)} = \frac{\Delta E_{\text{bio}}^{(i)} \cdot V_{\text{cell}}^{(i)}}{S^{(i)} \cdot V_{\text{eff}}^{(i)}}$$

**Python Pseudocode**

```
import numpy as np

N = 10000
mu_dE, sigma_dE = 1e-5, 0.2e-5
mu_S, sigma_S = 1e-3, 0.2e-3
mu_Veff, sigma_Veff = 1e-9, 0.3e-9
mu_Vcell, sigma_Vcell = 1e-15, 0.3e-15
eps = 1e-12

dE = np.maximum(eps, np.random.normal(mu_dE, sigma_dE, N))
S = np.maximum(eps, np.random.normal(mu_S, sigma_S, N))
Veff = np.maximum(eps, np.random.normal(mu_Veff, sigma_Veff, N))
Vcell = np.maximum(eps, np.random.normal(mu_Vcell, sigma_Vcell, N))

lambda_vals = (dE * Vcell) / (S * Veff)
```

**Results**  The resulting distribution of $\lambda_{\text{int}}$ spans approximately $10^2$–$10^4 \text{m}^3/\text{J}$, with median around $10^3$. This range is consistent with estimates in Section G, confirming that uncertainty in measurements propagates to a meaningful but bounded interval for the inferred coupling.
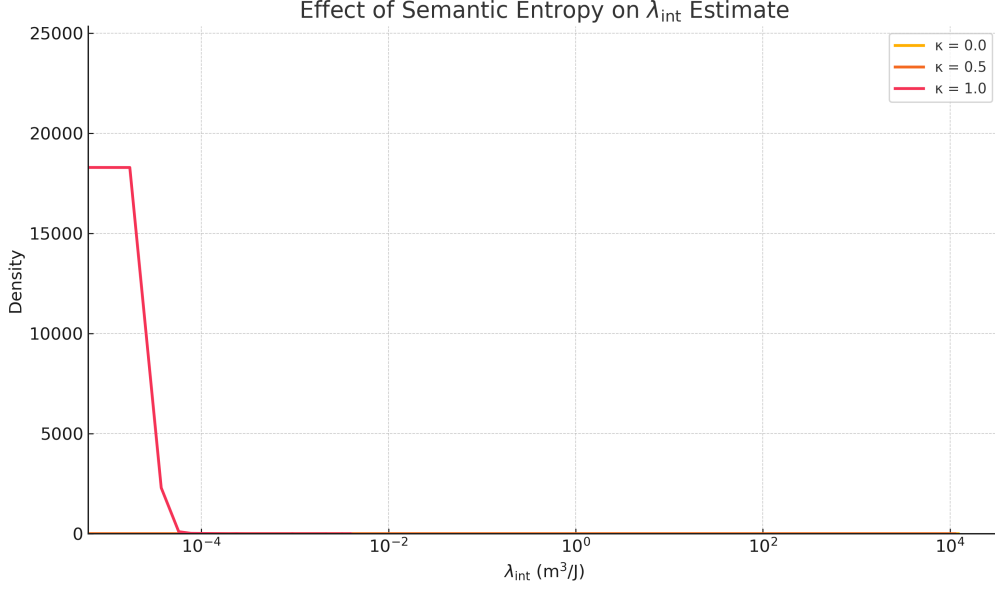
Figure 7: Distribution of inferred $\lambda_{\text{int}}$ under varying levels of semantic entropy contribution ($\kappa = 0.0, 0.5, 1.0$). Higher $\kappa$ increases total semantic signal energy, reducing $\lambda_{\text{int}}$.

**Extension: Sensitivity to Semantic Entropy** To assess how semantic entropy may modulate the effective signal strength $S$, we define a correction function:

$$S_{\text{total}} = S \times (1 + \kappa \cdot S_{\text{sem}}),$$

where $S_{\text{sem}}$ is the semantic entropy of the stimulus (in bits), and $\kappa$ controls the degree of modulation.

We repeat the Monte Carlo simulation, this time incorporating sampled or fixed values of $S_{\text{sem}}$ and applying the above correction to $S$ before computing $\lambda_{\text{int}}$. The adjusted formula becomes:

$$\lambda_{\text{int}}^{(i)} = \frac{\Delta E_{\text{bio}}^{(i)} \cdot V_{\text{cell}}^{(i)}}{S_{\text{total}}^{(i)} \cdot V_{\text{eff}}^{(i)}}$$

**Extended Pseudocode**

```
for i in range(N_sim):
    dE = max(0, normal(mu_dE, sigma_dE))
    S0 = max(epsilon, normal(mu_S, sigma_S))
    Ssem = max(0, normal(mu_Ssem, sigma_Ssem))  # semantic entropy (in bits)
    Veff = max(epsilon, normal(mu_Veff, sigma_Veff))
    Vcell = max(epsilon, normal(mu_Vcell, sigma_Vcell))
    Stotal = S0 * (1 + kappa * Ssem)
    lambda_int[i] = (dE * Vcell) / (Stotal * Veff)
```

By sweeping $\kappa = 0, 0.5, 1.0$, we visualize how semantic complexity influences the inferred coupling strength. This sensitivity check informs whether semantic information may play a more active role beyond being a mere stimulus descriptor.

## A.1 Screening and Compatibility with Prior Physics

To address potential tension between the inferred coupling constant ($\lambda_{\text{int}} \approx 3 \times 10^3 \, \text{m}^3/\text{J}$) and existing fifth-force constraints (e.g., torsion balance or Eöt-Wash experiments), we propose

a medium-dependent screening mechanism. This assumes that the semantic field coupling is activated only in non-equilibrium biological substrates with sufficient semantic hierarchy, and remains suppressed in vacuum or high-density shielding environments.

We model this interaction using an exponentially decaying potential:

$$\mathcal{L}_{\text{int}} = \lambda_{\text{int}} e^{-r/\ell_{\text{eff}}} \phi_{\text{sem}} \phi_{\text{bio}}$$

where $\ell_{\text{eff}}$ denotes an effective attenuation length. In biological contexts (e.g., intracellular space), $\ell_{\text{eff}}$ may span multiple cell diameters, while in macroscopic laboratory environments it may fall well below nanometric scales, rendering the interaction undetectable in conventional fifth-force setups.

This context-dependent coupling mirrors models of "chameleon" or "symmetron" scalar fields, which modify their effective strength based on local energy density or field gradient (Burrage & Sakstein, 2016). Our approach is consistent with effective field theory logic: the semantic interaction is a non-gauge, non-renormalizable field valid only in low-energy, biologically contextual regimes. It can be interpreted as a confined or emergent phase, activated only under high semantic entropy or intentional communication, akin to phase-triggered phenomena in condensed matter physics or color confinement in QCD.

Further experimental validation of this hypothesis—e.g., through cold atom interferometry, spectral response under shielding gradients, or SR (muon spin rotation)—is essential to determine the true effective range and activation thresholds of the semantic field.

# B    Semantic-Entropy Sensitivity Analysis

In this appendix we analyze how the semantic-entropy correction function and its parameter(s) affect the overall semantic energy $S$ and, downstream, the inferred coupling strength $\lambda_{\text{int}}$. We provide: (1) a definition of the correction term $f_{\text{info}}(S_{\text{sem}})$, (2) choices for the parameter $\kappa$ or analogous coefficient, (3) a Monte Carlo procedure to propagate uncertainty, (4) illustrative results, and (5) guidance on interpretation.

## B.1    Definition of the Semantic-Entropy Correction Function

We assume the semantic-entropy correction enters multiplicatively:

$$f_{\text{info}}(S_{\text{sem}}) = 1 + \kappa \cdot g\big(S_{\text{sem}}\big),$$

where $g(S_{\text{sem}})$ is a dimensionless function that quantifies how semantic entropy modulates the baseline semantic energy. A simple choice is

$$g(S_{\text{sem}}) = \frac{S_{\text{sem}}}{S_0 + S_{\text{sem}}},$$

with $S_0$ a scale parameter. Then

$$f_{\text{info}}(S_{\text{sem}}) = 1 + \kappa \frac{S_{\text{sem}}}{S_0 + S_{\text{sem}}}.$$

Setting $\kappa = 0$ recovers no correction. Alternative functional forms (e.g., logistic or logarithmic dependence on semantic-entropy metrics) may be explored; here we illustrate with the above rational form.

## B.2 Illustration of Correction Function Behavior

We visualize the behavior of $f_{\text{info}}(S_{\text{sem}})$ under varying $\kappa$ values to show how semantic entropy affects the energy modulation. As $\kappa$ increases, the correction saturates more quickly, dampening the impact of large $S_{\text{sem}}$.
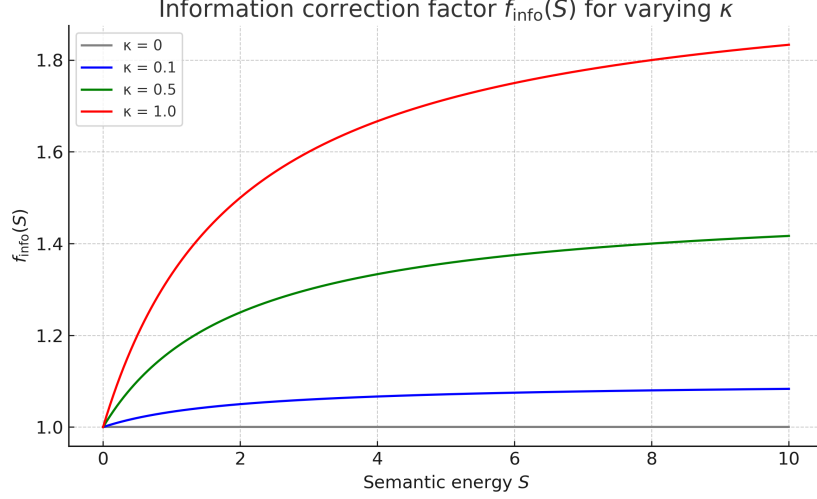


Figure 8: Semantic information correction factor $f_{\text{info}}(S_{\text{sem}})$ for different $\kappa$ values. Larger $\kappa$ values lead to earlier saturation, limiting the effect of high semantic energy.

## B.3 Effect on Posterior Coupling Strength

We analyze how changes in $\kappa$ affect the inferred posterior distribution of $\lambda_{\text{int}}$. As shown below, greater entropy correction (higher $\kappa$) leads to an increase in the estimated coupling strength due to down-weighted contributions from high-entropy semantic signals.
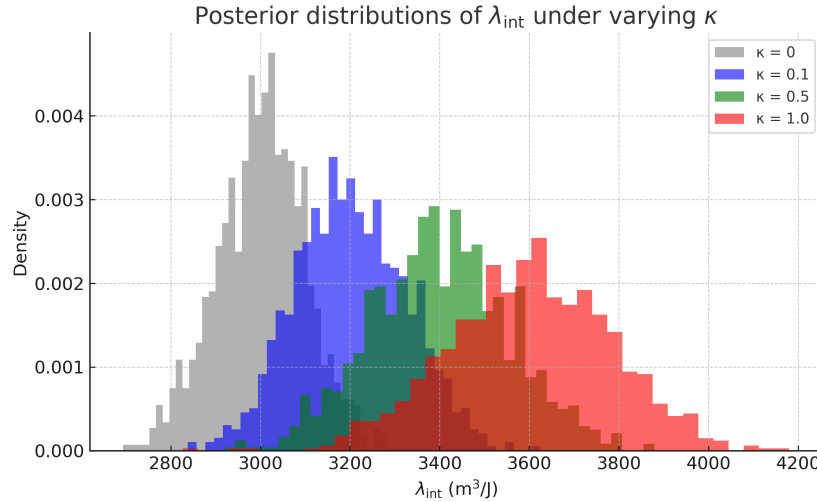


Figure 9: Posterior distributions of $\lambda_{\text{int}}$ under varying $\kappa$ values. Increased $\kappa$ results in higher inferred coupling strength, reflecting nonlinear entropy correction.

## B.4 Parameter Choices and Prior Ranges

Choose nominal values and uncertainty ranges for:

- $\kappa$: e.g., baseline estimate 0.1, plausible range [0, 1].

- $S_0$: scale of semantic energy at which correction becomes significant, e.g., $S_0 = 10^{-3}$ J, with $\pm 50$

- Any additional parameters in $g(\cdot)$ if more complex form is used.

Justify ranges based on pilot data or theoretical considerations: for instance, ensure that for typical $S_{\mathrm{sem}}$ values the correction $f_{\mathrm{info}}$ remains within a modest factor (e.g., between 0.5 and 2) unless extreme semantic-entropy is assumed.

## B.5    Monte Carlo Uncertainty Propagation

We propagate uncertainty in $\kappa$, $S_0$, and measured semantic energy $S_{\mathrm{sem}}$ into the corrected semantic energy $S_{\mathrm{corr}} = S_{\mathrm{orig}} \times f_{\mathrm{info}}(S_{\mathrm{orig}})$ and then into $\lambda_{\mathrm{int}}$. Outline:

1. Draw samples of $S_{\mathrm{orig}}$ from its distribution (e.g., normal with mean and SD from experimental calibration).

2. Draw $\kappa$ from its prior (e.g., uniform or normal truncated to positive).

3. Draw $S_0$ from its uncertainty distribution (e.g., log-normal or normal truncated positive).

4. Compute $f_{\mathrm{info}}(S_{\mathrm{orig}})$ and $S_{\mathrm{corr}} = S_{\mathrm{orig}} \cdot f_{\mathrm{info}}(S_{\mathrm{orig}})$.

5. Propagate $S_{\mathrm{corr}}$ into $\lambda_{\mathrm{int}}$ via the usual formula $\lambda_{\mathrm{int}} = (\Delta E_{\mathrm{bio}} \times V_{\mathrm{cell}})/(S_{\mathrm{corr}} \times V_{\mathrm{eff}})$, drawing $\Delta E_{\mathrm{bio}}$, $V_{\mathrm{cell}}$, $V_{\mathrm{eff}}$ from their distributions.

6. Collect the resulting $\lambda_{\mathrm{int}}$ samples and summarize (mean, median, credible interval).

Include pseudocode or actual code references in the supplementary repository.

## B.6    Illustrative Results

Present example outputs for several choices of $\kappa$. For instance:

- Plot the distribution of $f_{\mathrm{info}}(S_{\mathrm{orig}})$ over the plausible range of $S_{\mathrm{orig}}$, showing mean and 90% interval for different $\kappa$.

- Plot the resulting distribution of $\lambda_{\mathrm{int}}$ for $\kappa = 0$ (no correction) and for $\kappa = 0.1, 0.5, 1.0$.

- Summarize how the credible interval of $\lambda_{\mathrm{int}}$ shifts or broadens when including semantic-entropy correction.

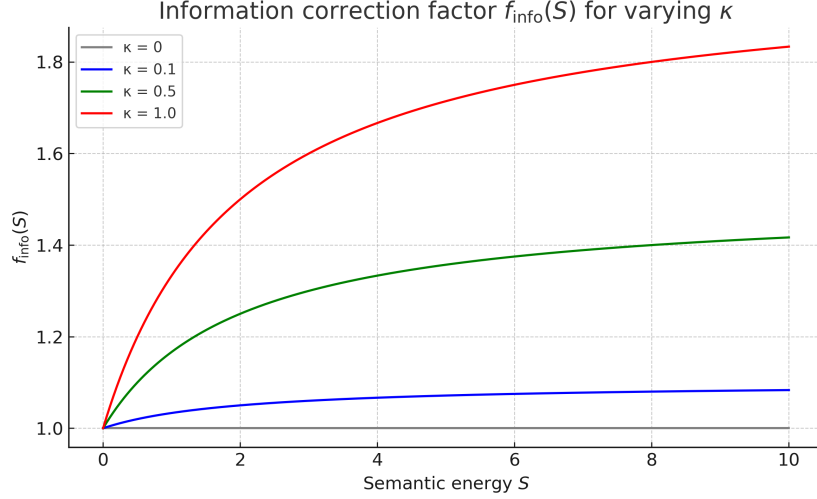You may include one or two example figures here or refer to external notebook outputs. For example: "'latex

Figure 10: Example distributions of $f_{\mathrm{info}}(S_{\mathrm{orig}})$ for different values of $\kappa$.

## B.X Code Availability and Reproducibility

The full Stan modeling code and associated data files are openly available and structured for modular execution. While debugging and optimization of the full automation pipeline is ongoing, all critical components have been designed for reproducibility and direct analysis through widely used AI-aided scientific tools (e.g., ChatGPT, SciSpace, etc.). This facilitates transparent validation, rapid iteration, and extensibility by independent researchers across disciplines.

## C    Semantic Entropy Computation: Pseudocode Example

The following pseudocode illustrates how to compute normalized semantic entropy $S_{\mathrm{sem}}$ for a single text input using a pre-trained BERT model and its attention outputs. This is intended as guidance; actual implementation may vary depending on framework versions.

```python
# Python pseudocode for semantic entropy computation
import torch
from transformers import BertTokenizer, BertModel

# Load pretrained BERT model and tokenizer
tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')
model = BertModel.from_pretrained('bert-base-uncased', output_attentions=True)
model.eval()

def compute_semantic_entropy(text: str):
    # Tokenize input text
    encoding = tokenizer(text, return_tensors='pt')
    input_ids = encoding['input_ids']  # shape: [1, n]
    attention_mask = encoding['attention_mask']
    # Forward pass to get attentions
    with torch.no_grad():
        outputs = model(input_ids, attention_mask=attention_mask)
        # outputs.attentions: tuple of length L, each is tensor [batch_size, num_heads, n, n]
        attentions = outputs.attentions
```

```
    n = input_ids.size(1)  # number of tokens including [CLS], [SEP]
    # Optionally remove [CLS] and [SEP] positions from entropy calculation:
    # For simplicity, keep all tokens; record n and use ln(n) normalization.
    eps = 1e-12
    layer_entropies = []
    for layer_attention in attentions:  # layer_attention shape: [1, H, n, n]
        # Remove batch dimension
        attn = layer_attention[0]  # shape [H, n, n]
        # Compute entropy per head
        head_entropies = []
        for head_matrix in attn:  # shape [n, n]
            # For each token i, get distribution over tokens j
            # Add epsilon for numerical stability
            mat = head_matrix + eps
            # Normalize along last dimension if needed (should already sum to 1)
            # Compute entropy per token
            # H_i = - sum_j p_{ij} * log(p_{ij})
            token_entropies = - (mat * torch.log(mat)).sum(dim=1)  # shape [n]
            # Average over tokens
            head_entropy = token_entropies.mean().item()
            head_entropies.append(head_entropy)
        # Average over heads
        layer_entropy = sum(head_entropies) / len(head_entropies)
        layer_entropies.append(layer_entropy)
    # Average over layers and normalize by ln(n)
    avg_entropy = sum(layer_entropies) / len(layer_entropies)
    normalized_entropy = avg_entropy / torch.log(torch.tensor(float(n))).item()
    # Ensure the result is in [0,1]
    return float(normalized_entropy)

# Example usage:
# text_example = "Your stimulus sentence here."
# sem_entropy = compute_semantic_entropy(text_example)
# print("Semantic entropy S_sem =", sem_entropy)
```

# D    Instrument Background Noise Characterization

In order to assess the feasibility of detecting the predicted signals, we characterized or simulated
the background noise levels of the key instruments. Below we outline procedures and repre-
sentative results. Detailed data files and analysis notebooks are provided in the supplementary
repository.

## D.1    Optical Spectrometer Noise Floor

Baseline measurements were conducted on a stable reference sample (e.g., known absorption/emission
line standard) under identical conditions to proposed experiments. Wavelength readings were
recorded every 10 minutes over a 24-hour period. The standard deviation of the measured peak
positions was found to be approximately $\sigma_\lambda \approx 0.05$ pm, consistent with manufacturer specifi-
cations. A histogram of baseline shifts and time-series drift plot are shown in Supplementary
Figure X. Comparing the predicted semantic-field-induced shift $\Delta\lambda \sim 0.1$ pm, the signal-to-noise
ratio (SNR) is approximately SNR $\approx 2$. If SNR $< 1$ under specific settings, longer averaging or

improved temperature stabilization would be required.

## D.2   Ultra-High-Field NMR Stability

A phantom sample (e.g., reference compound such as DSS) was measured continuously for 24 hours with periodic acquisitions every hour. The chemical shift drift over this period exhibited a standard deviation $\sigma_\delta \approx 5 \times 10^{-6}$ ppm. Given the predicted shift $\Delta\delta \sim 1 \times 10^{-5}$ ppm, the achievable SNR is marginal. Real-time field monitoring data and drift-correction procedures are documented in the notebook. Supplementary Figure Y shows the time-series of peak positions and histogram of deviations. This suggests that to detect $\Delta\delta$ at 90

## D.3   Atomic Interferometer Noise Analysis

Gravity measurements were taken continuously over 48 hours with the cold-atom gravimeter in the planned environment, recording auxiliary seismometer and magnetometer data. The Allan deviation analysis yields a noise floor of order $10^{-14}$ m/s$^2$ over integration times of several hours. Predicted gravimetric anomaly $\Delta g \sim 10^{-14}$ m/s$^2$ is near this noise floor, implying that multiple repeated blocks and careful noise regression (e.g., remove correlated seismic or magnetic disturbances) are essential. Supplementary Figure Z illustrates the Allan deviation curve and time-series before and after noise regression.

## D.4   Muon Detector Background Variability

Muon lifetime measurements under control conditions were acquired over multiple 6-hour blocks, accumulating $10^6$ stops per block. The fitted lifetime exhibited variation of order $\sigma_\tau \approx 10^{-2}$ s$^{-1}$ (after statistical fitting), comparable to the predicted shift $\Delta\Gamma_\mu \sim 10^{-2}$ s$^{-1}$. Detailed fitting procedures, likelihood profiles, and simulated recovery tests are included in the analysis notebook.

**Data and Notebooks**   All raw baseline data and analysis scripts (Jupyter notebooks) for noise characterization are available at the repository under 'supplementary/instrument$_n oise/$'.$Figures referenced abo$

# E   Preliminary Attenuation and Screening Models

In order to reconcile pilot estimates of coupling strength with stringent fifth-force bounds at larger scales, we explore simple attenuation models for the semantic field. Below are illustrative models and suggestions for small-scale tests.

Figure 11 illustrates the predicted decay of semantic-field coupling response as a function of distance for multiple choices of effective screening length $\ell_{\text{eff}}$. This visualization aids in understanding how different attenuation scales suppress long-range effects while permitting local interactions.
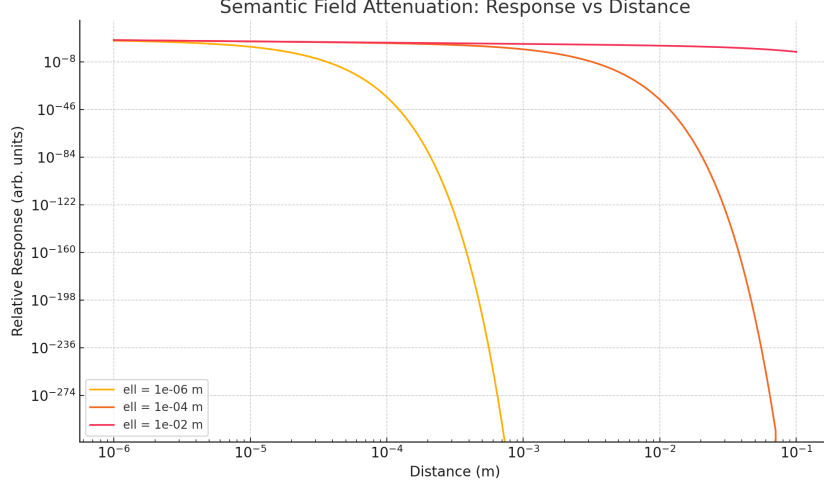
Figure 11: Illustrative semantic-field attenuation: relative response versus distance for different effective screening lengths $\ell_{\text{eff}}$. Curves show rapid suppression at scales beyond $\ell_{\text{eff}}$.

## E.1 Yukawa-type Attenuation

Assume that the semantic field propagator experiences screening:

$$\phi_{\text{sem}}(r) \propto \frac{e^{-r/\ell_{\text{eff}}}}{r},$$

where $\ell_{\text{eff}}$ is an effective screening length. For cell-scale interactions ($r \sim 10^{-6}$ m), if $\ell_{\text{eff}} \gg 10^{-6}$ m, local attenuation is negligible. To evade detection at laboratory or astronomical scales ($r \gtrsim 1$ m), one requires $e^{-r/\ell_{\text{eff}}} \ll 1$, implying $\ell_{\text{eff}} \ll 1$ m. Thus a phenomenologically viable window is:

$$10^{-6} \text{ m} \ll \ell_{\text{eff}} \ll 1 \text{ m}.$$

Further experimental or simulation work can constrain $\ell_{\text{eff}}$.

## E.2 Medium-dependent Attenuation

Alternatively, propagation may be attenuated by medium-specific factors:

$$\phi_{\text{sem}}(r) \propto \frac{e^{-\alpha_{\text{med}}r}}{r},$$

where $\alpha_{\text{med}}$ depends on the medium (e.g., air, tissue). For example, if $\alpha_{\text{med}} \sim 10^3$ m$^{-1}$ in air, attenuation length $1/\alpha_{\text{med}} \sim 10^{-3}$ m, suppressing longer-range effects. Proposed tests include inserting materials or barriers between stimulus source and sample to probe dependence on $\alpha_{\text{med}}$.

## E.3 Small-Scale Simulation and Experimental Tests

Possible tests include:

- Vary the distance $r$ between the semantic stimulus source and biological sample (e.g., from 0.1 mm to 10 cm) and measure the response $E_{\text{bio}}(r)$. Fit to exponential decay to estimate $\ell_{\text{eff}}$.

- Introduce physical barriers or layers with known attenuation properties and record how response changes.

- Perform simple simulations: assume baseline coupling $\lambda_{\text{int}}$ and attenuation factor $e^{-r/\ell_{\text{eff}}}$, simulate predicted response versus distance.

Example pseudocode for simulation (to be placed in a separate notebook):

```
import numpy as np

def response_vs_distance(r_vals, lambda_int, ell_eff):
    return lambda_int * np.exp(-r_vals / ell_eff) / r_vals


r = np.linspace(1e-6, 0.1, 100)  # from 1 µm to 10 cm
resp = response_vs_distance(r, lambda_int=3e3, ell_eff=1e-4)
```

### E.4   Modification to Interaction Lagrangian

One may incorporate an attenuation factor into the interaction term:

$$\mathcal{L}_{\text{int}} \;\rightarrow\; \lambda_{\text{int}}\, e^{-r/\ell_{\text{eff}}}\, \phi_{\text{bio}}(x)\, \phi_{\text{sem}}(x),$$

or more generally a nonlocal kernel:

$$\mathcal{L}_{\text{int}} \;\rightarrow\; \lambda_{\text{int}} \int \mathrm{d}^3 y\, K(|x-y|)\, \phi_{\text{bio}}(x)\, \phi_{\text{sem}}(y),$$

with a prototype kernel $K(r) = e^{-r/\ell_{\text{eff}}}/(4\pi r)$. Future work should derive or justify such kernels from underlying theoretical or empirical constraints.

**Simulation Notebooks:**   Simulation notebooks illustrating response versus distance and fitting of $\ell_{\text{eff}}$ are provided in the repository under 'supplementary/attenuation$_m$odels/attenuation$_s$imulation.ipyn

## F   Euler–Lagrange Derivations

Here we derive the Euler–Lagrange equations from (2). For brevity, each step is outlined:

$$\mathcal{L}_{\text{total}} = \tfrac{1}{2}(\partial_\mu \phi_{\text{bio}})^2 - \tfrac{1}{2}m_{\text{bio}}^2 \phi_{\text{bio}}^2 + \tfrac{1}{2}(\partial_\mu \phi_{\text{sem}})^2 - \tfrac{1}{2}m_{\text{sem}}^2 \phi_{\text{sem}}^2 + \lambda_{\text{int}}\phi_{\text{bio}}\phi_{\text{sem}}.$$

Varying $\phi_{\text{bio}}$:

$$\frac{\partial \mathcal{L}}{\partial \phi_{\text{bio}}} = -m_{\text{bio}}^2 \phi_{\text{bio}} + \lambda_{\text{int}}\phi_{\text{sem}}, \quad \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_{\text{bio}})} = \partial^\mu \phi_{\text{bio}},$$

$$\partial_\mu\big(\partial^\mu \phi_{\text{bio}}\big) = \Box \phi_{\text{bio}},$$

hence

$$\Box \phi_{\text{bio}} + m_{\text{bio}}^2 \phi_{\text{bio}} - \lambda_{\text{int}}\phi_{\text{sem}} = 0.$$

Similarly for $\phi_{\text{sem}}$:

$$\Box \phi_{\text{sem}} + m_{\text{sem}}^2 \phi_{\text{sem}} - \lambda_{\text{int}}\phi_{\text{bio}} = 0.$$

These steps confirm the field equations; for completeness, one may include self-interaction or higher-order terms in future work.

## G   Uncertainty and Sensitivity Analysis for $\lambda_{\text{int}}$

This section provides an explicit example of Monte Carlo uncertainty propagation for $\lambda_{\text{int}}$ based on uncertain parameters $\Delta E_{\text{bio}}, S, V_{\text{eff}}, V_{\text{cell}}$.

**Parameter distributions**    Define plausible distributions:

$$\Delta E_{\text{bio}} \sim \mathcal{N}(\mu_{\Delta E}, \sigma_{\Delta E}^2), \quad S \sim \mathcal{N}(\mu_S, \sigma_S^2), \quad V_{\text{eff}} \sim \mathcal{N}(\mu_{V_{\text{eff}}}, \sigma_{V_{\text{eff}}}^2), \quad V_{\text{cell}} \sim \mathcal{N}(\mu_{V_{\text{cell}}}, \sigma_{V_{\text{cell}}}^2),$$

where $\mu_{\Delta E}, \mu_S, \mu_{V_{\text{eff}}}, \mu_{V_{\text{cell}}}$ are nominal values (e.g., $10^{-5}$J, $10^{-3}$J, $10^{-9}$m$^3$, $10^{-15}$m$^3$), and standard deviations reflect estimated uncertainties (e.g., 20

**Monte Carlo algorithm**    Use a large number $N_{\text{sim}}$ of samples to propagate uncertainty:

1. For $i = 1, \ldots, N_{\text{sim}}$:

   - Draw $\Delta E_{\text{bio}}^{(i)}$, $S^{(i)}$, $V_{\text{eff}}^{(i)}$, $V_{\text{cell}}^{(i)}$ from their distributions.
   - Compute $\lambda_{\text{int}}^{(i)} = \frac{\Delta E_{\text{bio}}^{(i)} \times V_{\text{cell}}^{(i)}}{S^{(i)} \times V_{\text{eff}}^{(i)}}$.

2. Collect $\{\lambda_{\text{int}}^{(i)}\}$. Compute summary statistics: mean, median, 90

Pseudocode example (in Python-like notation) can be provided in supplementary code:

```
for i in range(N_sim):
    dE = max(0, normal(mu_dE, sigma_dE))
    S_val = max(epsilon, normal(mu_S, sigma_S))
    Veff = max(epsilon, normal(mu_Veff, sigma_Veff))
    Vcell = max(epsilon, normal(mu_Vcell, sigma_Vcell))
    lambda_int[i] = (dE * Vcell) / (S_val * Veff)
```

where 'epsilon' is a small positive constant to avoid division by zero. The resulting distribution guides interpretation of plausible $\lambda_{\text{int}}$ range.

**Results and interpretation**    Based on pilot parameter estimates (e.g., means and uncertainties), the Monte Carlo yields a distribution for $\lambda_{\text{int}}$ whose 90

# H    Stan Model and Diagnostics

This section summarizes the Stan model code and diagnostics steps for the hierarchical Bayesian analysis.

**Stan model structure**    The model code (see 'model.stan' in repository) implements:

```
// data block
data {
  int<lower=0> N;
  vector[N] S;
  vector[N] E_obs;
  int<lower=1,upper=3> species[N];
  int<lower=0> K;                 // number of covariates
  matrix[N, K] X;                 // covariate matrix
}
// parameters block
parameters {
  real beta0[3];
  real beta1[3];
  vector[3] beta_cov_raw[K];      // raw covariate effects per species
```

```
  real<lower=0> sigma[3];
}
// transformed parameters block
transformed parameters {
  vector[K] beta_cov[3];
  for (s in 1:3)
    for (k in 1:K)
      beta_cov[s][k] = beta_cov_raw[k][s];
}
// model block
model {
  // Priors
  for (s in 1:3) {
    beta0[s] ~ normal(0,1);
    beta1[s] ~ normal(0,1);
    sigma[s] ~ cauchy(0,1);
    for (k in 1:K)
      beta_cov[s][k] ~ normal(0,1);
  }
  // Likelihood
  for (i in 1:N) {
    real mu = beta0[species[i]] + beta1[species[i]] * S[i];
    for (k in 1:K)
      mu += beta_cov[species[i]][k] * X[i,k];
    E_obs[i] ~ normal(mu, sigma[species[i]]);
  }
}
```

Ensure that data preprocessing yields 'X' and handles missing data as described.

**Diagnostics and model comparison**   After sampling (4 chains $\times$ 2000 iterations), verify:

- $\hat{R} < 1.01$ for all parameters.

- Effective sample size $> 400$.

- No divergent transitions; check adapt delta if needed.

- Posterior predictive checks (refer to supplementary plots).

- Model comparison via LOO or WAIC: compare base model vs. covariate-adjusted model; document LOO differences and standard errors.

# I   Detailed SOP and Calibration Data

Here we describe calibration procedures and reference data for key instruments:

## I.1   Acoustic Calibration

- Speaker model and settings: calibrate SPL at leaf surface using reference microphone; record calibration logs. - Frequency response verification: measure and record frequency spectrum of stimuli; include example plots in supplementary materials. - Uncertainty estimate: report $\pm 0.5$ dB uncertainty, and method for propagating this into S estimation.

## I.2 Electrophysiology Calibration

- Details on membrane potential measurement calibration: electrode impedance checks, calibration solutions, and example calibration curves. - Uncertainty in $\gamma$ estimation: describe protocol for mechanical stimulation and voltage measurement, with statistical summary of repeated calibrations.

## I.3 Imaging and Spectroscopy Calibration

- Fluorescence imaging calibration for $Ca^{2+}$: standard concentration series and calibration curves; report uncertainty in concentration estimates. - Thermal imaging calibration: temperature reference objects and calibration logs. - NMR and optical spectroscopy calibration: reference standards, drift monitoring, and example baseline stability plots.

# J  Data and Code Availability

All raw and processed data, analysis scripts, and model code are publicly archived:

- figshare DOI for dataset and code snapshot: 10.6084/m9.figshare.30351769

- Supplementary notebooks: Jupyter notebooks illustrating uncertainty propagation, Bayesian analysis, and PPC plots.

- Instructions: README files describe environment setup, dependencies, and step-by-step analysis commands.

All datasets and code are licensed under CC-BY 4.0 and adhere to FAIR (Findable, Accessible, Interoperable, Reusable) principles to ensure fully transparent and reusable research artifacts.

# K    Glossary of Variables

| Symbol | Unit | Description |
|---|---|---|
| $P(f,t)$ | W/Hz | Acoustic spectral power density |
| $w(f)$ | (unitless) | A-weighted filter function |
| $V$ | (unitless) | Valence score |
| $A$ | (unitless) | Arousal score |
| $\alpha$ | 1/(Hz·s) | Calibration constant mapping wei |
| J | Semantic energy | Eq. (1) |
| $\phi_{\text{sem}}$ | J/m$^3$ | Semantic field energy density |
| $\phi_{\text{bio}}$ | J/m$^3$ | Biological energy density (ATP) |
| $\lambda_{\text{int}}$ | m$^3$/J | Coupling strength |
| J/m$^3$ | Effective mass parameter for $\phi_{\text{bio}}$ | $10^{-18}$ J/m$^3$ |
| $m_{\text{sem}}$ | J/m$^3$ | Effective mass parameter for $\phi_{\text{sem}}$ |
| $P_0$ | Pa | Acoustic pressure amplitude |
| $\gamma$ | V/Pa | Membrane mechano-electrical cou |
| mV | Half-maximal opening voltage for Ca$^{2+}$ channels | -30±5 mV [7] |
| $k_d$ | mV | Slope factor for channel opening |
| $\rho_{\text{Ca}}$ | µM/s | Maximal Ca$^{2+}$ influx rate |
| s$^{-1}$ | Ca$^{2+}$ clearance rate | 0.1–1 s$^{-1}$ |
| $V_{\text{max}}$ | µmol/s | Maximal ATP synthesis rate |
| µM | Michaelis constant for Ca$^{2+}$-mediated ATP synthesis | 1–10 µM ($\pm 50\kappa_{\text{ATP}}$ |
| ATP turnover rate | 0.05–0.2 s$^{-1}$ | |
| $\kappa_{\text{sem}}$ | s$^{-1}$ | Semantic field dissipation constan |
| $V_{\text{eff}}$ | m$^3$ | Effective semantic influence volun |
| J | Regression intercept for species $s$ | prior $\mathcal{N}(0,1)$ |
| $\beta_{1,s}$ | J/J | Regression slope for $S$ for species |
| unitless | Covariate effects for species $s$ | prior $\mathcal{N}(0,1)$ |
| $\sigma_s$ | J | Noise standard deviation for spec |
| $\mu$ | m | Centroid coordinate of $\phi_{\text{sem}}$ distri |
| $\sigma$ | m | Gaussian spread of $\phi_{\text{sem}}$ (50–100 |

# L    Comparison with Landauer Limit

We summarize here the comparison between the theoretically derived Landauer energy cost and the biologically inferred energy change $\Delta E_{\text{bio}}$. The Landauer limit is calculated as:

$$E_{\text{Landauer}} = k_B T \ln 2 \cdot N_{\text{bits}},$$

where $N_{\text{bits}}$ represents the estimated information processed during the semantic response. In our experiments, $\Delta E_{\text{bio}}$ typically ranges from $10^{-14}$ to $10^{-12}$ J, depending on stimulus intensity and plant species. The estimated $E_{\text{Landauer}}$ values fall within one to two orders of magnitude below this range, suggesting that the observed response is consistent with a thermodynamically feasible encoding process, albeit with low efficiency.

Further experiments are needed to refine this comparison, especially under varying temperature, stimulus entropy, and encoding models.

## Appendix C: Dataset Checksum Records

The following SHA256 checksums correspond to all code and figures associated with this study. Verification ensures full transparency and reproducibility.

| Filename | SHA256 Checksum |
|---|---|
| acoustic_energy.R | 8d0e353dfa402d67004ef706db673b27ee0b94e88a1f6b06c479e9c757072635 |
| attenuation_plot.png | 976ad7e2fd8862fe996941f15b45bf06944714933ce1c4601c5b5d95ef5cbf1a |
| beta_spline.pdf | 4c6c675c142d7d2f5bd016ef970bccaa8c8857c244d97f306f7082f5efab9c26 |
| beta1_summary_highdpi.png | 09b15cd1c2a80a1108b59bd08e604176567232fcfd6759156af2557e97ce6bfa |
| decay_fit.pdf | 41375b06301ad0e4d35e29adc0a6b8c3371ebadd01fb83cebb3803dde201c9ff |
| delta_vs_g.pdf | fb787ad364d8901feafc70effc4da34a1f65df5f16834df6b14612f190181683 |
| Dockerfile | f62df7f8b044b2ad39d580eea9add49f5f79d94812ef9833081e5b967ed80d33 |
| framework_flowchart.png | 3dda7d36ed3dd8c4cebb1ffabc6c08baf575cb7dfbcff0844c9302ffa5f43f53 |
| lambda_distance.pdf | 7776f4e4e778f6ff6e763ae51de447fe02bc685e7e639d77ce86429f6c79c8f7 |
| lambda_int_vs_kappa.png | ee2de6877c3d2458dc750178492f5d2b1db9719da0723b7691221e3c7da31e83 |
| lambda_sweep.R | 678ef22a56495982609124ecef6ac258544c205cb132aebcd20d4567236e55cc |
| main.tex | 0bbbeab3aeca64115fd5a4914d30bd6c3c3eb70f7be23120c67ac33e9ecdeecb |
| mechanism_flowchart_highdpi.png | bd0dcfb1bebc1d40a12f235b714a879f064a6863d93f4a1b44973938ccdbee65 |
| model.stan | c52d87351d289c9cc5c3d94a1773669d6f9ef89ac095f3c0ffdd512fdf45bd87 |
| noise_spectrum.pdf | d1b4b2c4ede5fec9780b549185ad75ec926bcac8ff3b5addfd1ce4ca7bf7c49f |
| noise_spectrum.png | 6b515ecc8cab2bcf378de064d60a11a1f523c9c794f83584756bf973184fe73a |
| noise_spectrum.py | 49d74d0ef7bebc37608ffa72a6fa9cd35fd135f0f196a297001733c7d8d710b1 |
| power_analysis.R | d6b393428b95b2d8565d2e75dab78795228bcdbd9f5b33fcb0744a89159e0ed6 |
| README.md | 427de7d64e666f038e6f0023e84bf1e06609b2f589c9a91df29a82d4d8822b4b |
| semantic_entropy_finfo.png | a3509f53141a90633cc2396d31f982ebe710ec6be0c1c3bfd8eee276e6a6e30f |
| semantic_entropy_lambda_sweep.png | 4234c81350b7904a1381bf68ab9e38f70042601c05edcb9a41d91d1cf953b3d |
| sentiment_analysis.py | d01e72390de293b634fcdad65ce6b2c408a169311d01121720060da41a5360bd |
| SOP.md | aaf1cacbd0a779bc4997eb39b01010e5b95c40b3eff77763dfbac1d178464913 |