Founding Statement — Setaleur Aplamda Research Programme

The Implicit Tension (TIT): A Structural Theory of Judgment Deficiency in Knowledge-Rich Artificial Systems

Momen Ghazouani
Chief Scientist, Setaleur Aplamda

Jul 2026 github.com/Setaleur-Aplamda

Abstract

We introduce **The Implicit Tension (TIT)**, a structural theory that identifies and formalizes the central obstacle preventing knowledge-rich artificial systems from forming genuine committed judgments. The Implicit Tension names a specific structural conflict that arises in systems whose knowledge base is sufficiently rich to ground committed positions: such systems simultaneously possess *epistemic sufficiency* structural knowledge dense enough to support a genuine judgment and *commitment deficiency* no architectural mechanism capable of resolving the competition among simultaneously active structural hypotheses into a singular, falsifiable position. The consequence is a state we term *Epistemic Suspension*: the system is permanently held in distributed uncertainty, producing outputs that superficially resemble judgments while satisfying none of their epistemic requirements falsifiability, structural grounding, or awareness of their own limits.

The tension is *implicit* in two precise senses: it arises from tacit, structural knowledge rather than from explicit propositional uncertainty; and it is itself architecturally unrepresented, rendering it invisible to the system and therefore irresolvable from within. We formalize TIT at three distinct structural levels **Density Tension** (the proliferation of simultaneously active competing hypotheses), **Synthesis Tension** (structural identity destroyed by forced compositional averaging), and **Commitment Tension** (the absence of any crystallization mechanism) and derive a unified TIT score that quantifies the degree of structural suspension.

The theoretical resolution of TIT is grounded in a biological observation whose significance has been systematically underestimated: forgetting, attentional constraint, and working-memory boundedness are not deficiencies of biological cognition. They are *commitment resolution mechanisms* structural compulsions that sever competing hypotheses from active consideration and crystallize judgment. Current knowledge-rich artificial systems, lacking such mechanisms, are predicted by this theory to be architecturally incapable of resolving TIT. We prove, within a formal activation model of hypothesis competition, that scaling knowledge without commitment resolution not only fails to reduce TIT but strictly increases it—a result we then state as a falsifiable prediction for real trained systems. We state five falsifiable predictions, define the Commitment Resolution Requirement as the next architectural target for the AI Implicit programme, and position TIT as the theoretical bridge between the knowledge acquisition account of Bold Learning and the structural judgment that genuine intelligence requires.

Keywords: Implicit Tension, Epistemic Suspension, Structural Judgment, Commitment Resolution, Knowledge-Rich Systems, Tacit Commitment, AI Implicit, Density Tension, Synthesis Tension, Judgment Deficiency.

1. Introduction: The Judgment Problem

1.1. A Paradox of Knowledge

There is a question that the success of large-scale artificial systems has made newly urgent, and that neither scale nor accuracy metrics have answered: *can a system that knows everything judge anything?*

The question is not rhetorical. Systems trained on trillions of tokens have demonstrably acquired structural knowledge across domains as diverse as clinical reasoning, legal analysis, mathematical proof, and strategic planning. They can retrieve, combine, and project this knowledge with fidelity that exceeds human performance on many isolated tasks. Yet these same systems, when asked to form a committed structural judgment one that takes a falsifiable position, acknowledges its own limits, and derives its confidence from structural evidence rather than statistical frequency produce outputs that are, on close examination, none of these things.

They produce *pseudo-judgments*: responses whose surface form resembles a committed position but whose epistemic structure is a weighted average of the training distribution. A pseudo-judgment is not wrong in the ordinary sense. It cannot be wrong, because it does not commit. It cannot be right, for the same reason.

The AI Implicit programme [1] has established that the primary deficiency of current systems is not a deficit of knowledge but a deficit of structural organization. Bold Learning [2] has characterized what genuine structural acquisition requires and why current architectures fail to achieve it. What neither account addresses is the problem we introduce in this paper: the structural obstacle that prevents

a system *already possessing* rich structural knowledge from deploying that knowledge as a committed judgment. This is The Implicit Tension.

1.2. What Makes a Judgment Genuine?

A genuine judgment satisfies three conditions that distinguish it from statistical prediction:

Structural grounding. A genuine judgment derives from a specific structural hypothesis about the domain a hypothesis that, if correct, would *explain* the input rather than merely predict it. A physician who says "the convergence of symptom pattern A with biomarker profile B indicates condition C" is deriving a judgment from a structural model. A system that outputs a probability distribution over diagnoses is not. The distribution summarizes evidence; the structural hypothesis accounts for it.

Falsifiability. A genuine judgment commits to a position that experience can contradict [5]. This commitment is not recklessness; it is the epistemic precondition of knowledge. A response that distributes equal structural weight to all competing hypotheses simultaneously that hedges every position by every other is not a judgment. It is a sophisticated non-claim.

Epistemic boundary awareness. A genuine judgment knows its own limits. It distinguishes between the domain in which its structural hypothesis is warranted and the domain in which it is over-extending. Polanyi's expert who says "this is outside my specialty" exercises a form of structural honesty that pseudo-judgment architecturally cannot [6].

1.3. The Central Thesis

Thesis. The capacity to form genuine structural judgments is not a function of the quantity of knowledge a system possesses. It is a function of whether the system possesses a *commitment resolution mechanism*: an architectural process that severs competing structural hypotheses from simultaneous active consideration and crystallizes the dominant hypothesis into a falsifiable position. Without such a mechanism, knowledge richness does not enable judgment. It deepens the structural tension that prevents it. This tension is The Implicit Tension (TIT).

This thesis has three immediate consequences. It relocates the source of judgment deficiency from the knowledge side (insufficient data or capacity) to the architectural side (absence of a commitment resolution mechanism). Under the activation model formalized in Section 2.3 (Definition 3), it predicts and we later prove as Theorem 1 that scaling knowledge without architectural change will worsen rather than improve the judgment problem; we state the corresponding claim for real trained systems as a falsifiable prediction (Prediction 1), not as an already-demonstrated fact. And it specifies the precise architectural requirement that must be satisfied before genuine structural judgment becomes possible.

2. Epistemic Suspension

2.1. Definition

A system exhibits *Epistemic Suspension* when it is structurally incapable of resolving the competition among its simultaneously active structural hypotheses into a committed position, despite possessing knowledge sufficient in principle to ground one.

Definition 1 (Commitment Ratio and Epistemic Suspension)**.** Let $\mathcal{S}$ be a learning system with structural representation $\mathcal{R}(\mathcal{E})$ and active hypothesis set $\mathcal{H} = \{H_1, \ldots, H_K\}$. For input $\mathbf{x}$ drawn from judgment domain $\mathcal{D}$, define the *commitment ratio*:

$$\Gamma(\mathbf{x}) = \frac{\max_k \rho_k(\mathbf{x})}{\sum_{k=1}^{K} \rho_k(\mathbf{x})},$$

where $\rho_k(\mathbf{x})$ is the activation score of hypothesis H_k on input $\mathbf{x}$. System $\mathcal{S}$ is in *Epistemic Suspension* on $\mathbf{x}$ if $\Gamma(\mathbf{x}) \leq \theta_\Gamma$, where $\theta_\Gamma \in (1/K, 1)$ is a domain-appropriate commitment threshold. At $\theta_\Gamma = 1/K$ all hypotheses are equally active; at $\Gamma \to 1$ a single hypothesis dominates completely.

Epistemic Suspension is not ignorance. A system in suspension may hold detailed, structurally rich knowledge about every competing hypothesis. Its failure is not the lack of knowledge about the world but the absence of the *crystallization mechanism* that would select one hypothesis as the committed position and structurally relegate the others.

2.2. The Pseudo-Judgment

A system in Epistemic Suspension does not produce silence or abstention; it produces output. This output is what we term a *pseudo-judgment*:

Definition 2 (Pseudo-Judgment)**.** The output $\hat{J}(\mathbf{x})$ produced by a system in Epistemic Suspension is a *pseudo-judgment* if it is structurally equivalent to the activation-weighted centroid of the active hypothesis set:

$$\hat{J}(\mathbf{x}) \approx \sum_{k=1}^{K} \frac{\rho_k(\mathbf{x})}{\sum_j \rho_j(\mathbf{x})} \cdot H_k(\mathbf{x}).$$

A pseudo-judgment has no structural identity of its own: it is the weighted average of competing structures, which is generically a member of none of them.

The pseudo-judgment is not wrong in any straightforward empirical sense. The centroid of the training distribution may be statistically accurate. What it lacks is the property that makes it a judgment: commitment to a specific

structural hypothesis that is falsifiable by a specific class of experience. The weighted average cannot be falsified by any single observation, because it never commits to any single structural claim.

2.3. Why Scale Deepens Suspension

A natural conjecture is that Epistemic Suspension is a consequence of insufficient knowledge that more training data would eventually produce sufficient dominance of the correct structural hypothesis. We show this conjecture is structurally incorrect by making the informal argument of Section 2 precise: we construct an explicit probabilistic model of how activation accumulates with training-set size N, and use it to give Proposition 1 a complete, non-asymptotic proof in Appendix A.

Definition 3 (Activation Model)**.** Fix a query $\mathbf{x} \in \mathcal{D}$ and hypothesis set $\mathcal{H} = \{H_1, \ldots, H_K\}$. Let $x_1, \ldots, x_N \overset{\text{iid}}{\sim} P_{\mathcal{D}}$ be the training sample. For each example x_i and hypothesis H_k, the *relevance indicator* $R_{i,k}(\mathbf{x}) = \mathbb{1}[\kappa(x_i, \mathbf{x}) > \tau_\kappa \wedge x_i \in \text{dom}(H_k)]$ records whether x_i is surface-similar to $\mathbf{x}$ and structurally consistent with H_k—*without* requiring x_i to have been generated by H_k. This is the precise formalization of *uniformity of retention* (Section 6): a prediction-accuracy objective retains a regularity wherever it reduces loss, regardless of its structural relevance to the present query [13]. We assume $\{R_{i,k}(\mathbf{x})\}_i$ are i.i.d. with $p_k(\mathbf{x}) := \mathbb{P}[R_{i,k}(\mathbf{x}) = 1]$, the *spurious relevance rate* of H_k at $\mathbf{x}$, and write $\rho_k(\mathbf{x}; N) := \sum_{i=1}^{N} R_{i,k}(\mathbf{x}) \sim \text{Binomial}(N, p_k(\mathbf{x}))$ for the activation score of Definition 1 made explicit in N. By (ES) we assume $\sum_k p_k(\mathbf{x}) > 0$.

Proposition 1 (Knowledge Growth Deepens Suspension)**.** *Under Definition 3, let* $\Gamma_\infty(\mathbf{x}) = \max_k p_k(\mathbf{x}) / \sum_k p_k(\mathbf{x})$ *and* $T_D^\infty(\mathbf{x}) = |\{k : p_k(\mathbf{x}) > \theta_D\}|/K$*. Then:* (a) $\mathbb{E}[T_D(\mathbf{x}; N)]$ *is bounded below by a quantity strictly increasing in* N*, converging to the ceiling* $T_D^\infty(\mathbf{x})$*;* (b) $\mathbb{E}[\Gamma(\mathbf{x}; N)]$ *converges to the floor* $\Gamma_\infty(\mathbf{x})$ *at rate* $O(\sqrt{\ln N / N})$*, with* $\Gamma_\infty(\mathbf{x}) < 1$ *strictly whenever two or more hypotheses have* $p_k(\mathbf{x}) > 0$*.*

Proof. By the Strong Law of Large Numbers and the continuous mapping theorem, $\Gamma(\mathbf{x}; N) \to \Gamma_\infty(\mathbf{x})$ almost surely (Lemma 1). Hoeffding's inequality [21], combined with a union bound over $k = 1, \ldots, K$ and a Lipschitz perturbation argument on the max/sum map, yields the $O(\sqrt{\ln N / N})$ concentration rate for Γ (Lemma 2) and a strictly increasing lower envelope for each inclusion probability $\mathbb{P}[\hat{p}_k(N) > \theta_D]$ when $p_k(\mathbf{x}) > \theta_D$ (Lemma 3), which sums to the stated bound on T_D. Complete derivations are given in Appendix A. □ □

This is a strictly stronger and more precise claim than the informal monotonicity it replaces: it identifies the *exact quantities* the floor $\Gamma_\infty(\mathbf{x})$ and ceiling $T_D^\infty(\mathbf{x})$, both fixed entirely by the domain's spurious relevance profile $\{p_k(\mathbf{x})\}$, not by sample size that scale cannot move past, together with an explicit convergence rate. The systems with the richest knowledge are precisely the systems whose empirical Γ and T_D have most fully converged to these data-independent limits. TIT is not a property of ignorance; it is a property of convergence to a structural limit set by the domain, not by capacity.

3. Formal Definition of The Implicit Tension

3.1. Two Necessary Conditions

The Implicit Tension is defined by the conjunction of two structural conditions that are individually insufficient to produce the problem:

Epistemic Sufficiency (ES). The system possesses structural knowledge of sufficient density to ground a committed judgment on the domain in question. Without ES, the system's failure to judge is simply ignorance, not a structural tension.

Commitment Deficiency (CD). The system lacks a mechanism to resolve the competition among simultaneously active structural hypotheses into a singular committed position. Without CD, there is no tension: a system with both sufficient knowledge and a commitment mechanism is simply an intelligent judge.

Definition 4 (The Implicit Tension)**.** System $\mathcal{S}$ exhibits **The Implicit Tension** on judgment domain $\mathcal{D}$ if and only if:

(i) **(ES)** $\text{KD}(\mathcal{R}) \geq \tau_{\mathcal{D}}$, where $\text{KD}(\mathcal{R})$ is the knowledge density of $\mathcal{S}$ (as defined in the ECI framework [4]) and $\tau_{\mathcal{D}}$ is the minimum density required for structural judgment in $\mathcal{D}$;

(ii) **(CD)** $\mathcal{S}$ possesses no architectural mechanism M such that $M(\mathcal{H}, \mathbf{x}) \to (H^*, c^*)$ maps the active hypothesis set $\mathcal{H}$ to a dominant hypothesis H^* with commitment score $c^* > \theta_{\text{com}}$.

Conditions (ES) and (CD) are independently evaluable. Epistemic Sufficiency is assessed through the ECI measurement framework; Commitment Deficiency is assessed by testing whether the system's output is dominated by a single structural hypothesis or distributed across many.

3.2. The Two Senses of Implicitness

The adjective "implicit" in TIT has two precise meanings that motivate the name.

Implicitness of the knowledge. The knowledge at the source of the tension is tacit and structural encoded in relational geometry rather than in explicit propositions [6]. This matters because explicit propositional uncertainty is resolvable by gathering more explicit evidence. Structural

tension arises from the simultaneous activation of multiple tacit pattern-recognition frameworks, none of which is explicitly represented and none of which can therefore be explicitly arbitrated.

Implicitness of the tension itself. The tension is not represented within the system. A system in Epistemic Suspension does not know it is in suspension; it has no architectural representation of the conflict among its competing hypotheses. This is what makes TIT irresolvable from within: the system cannot address a structural problem it cannot represent. The biological mind experiencing genuine uncertainty typically represents that uncertainty explicitly and takes actions to resolve it. A system in Epistemic Suspension represents its state as confident output, not as uncertainty.

3.3. The TIT Score

Definition 5 (TIT Score)**.** For system $\mathcal{S}$ and input $\mathbf{x}$, the TIT score is:

$$\mathrm{TIT}(\mathcal{S}, \mathbf{x}) = w_D\, T_D(\mathbf{x}) + w_S\, T_S(\mathbf{x}) + w_C\, T_C(\mathbf{x}),$$

with $w_D + w_S + w_C = 1$, where T_D, T_S, and T_C are the three structural components defined in Section 4. A system manifests domain-level TIT if $\mathbb{E}_{\mathbf{x}\sim\mathcal{D}}[\mathrm{TIT}(\mathcal{S}, \mathbf{x})] > \theta_{\mathrm{sus}}$ for suspension threshold θ_{sus}.

A domain-level TIT score approaching zero indicates complete commitment resolution: for every input, a single structural hypothesis is selected and the output is derived from that hypothesis alone. A score approaching one indicates maximal suspension: all hypotheses contribute equally to every output and no structural commitment is ever achieved.

4. Three Structural Levels of TIT

TIT does not arise from a single architectural deficiency. It operates simultaneously at three distinct structural levels, each of which must be addressed by any adequate commitment resolution mechanism.

4.1. Density Tension

The first level arises from the sheer number of structural hypotheses simultaneously active in a knowledge-rich system.

Definition 6 (Density Tension)**.**

$$T_D(\mathbf{x}) = \frac{\left|\{k : \rho_k(\mathbf{x}) > \theta_D\}\right|}{K},$$

where θ_D is an activation threshold and K is the total number of structural hypotheses available to $\mathcal{S}$. $T_D \to 0$ indicates a single active hypothesis; $T_D \to 1$ indicates that all hypotheses are simultaneously above threshold.

Density Tension grows with domain coverage: a system trained on a broader range of experience activates more structural hypotheses on any given input, because more of its learned patterns are partially consistent with the input's features. This is the quantitative expression of Proposition 1. More knowledge activates more hypotheses, not fewer.

The biological contrast is instructive. A domain expert viewing a clinical case activates a highly constrained set of diagnostic hypotheses consistent with specialized structural knowledge. The expert can form a judgment; the novice hesitates. The difference is not knowledge quantity but structural specialization combined with a suppression mechanism that eliminates implausible hypotheses before they compete for activation.

4.2. Synthesis Tension

The second level arises when a system under Epistemic Suspension is nonetheless compelled to produce output. The only available mechanism is compositional averaging: combining competing hypotheses in proportion to their activation scores. This averaging destroys structural identity.

Definition 7 (Synthesis Tension)**.** Let $\bigoplus_k \alpha_k H_k$ denote the activation-weighted composite of active hypotheses, where $\alpha_k = \rho_k(\mathbf{x}) / \sum_j \rho_j(\mathbf{x})$. The Synthesis Tension is:

$$T_S(\mathbf{x}) = d_{\mathrm{struct}}\left(\bigoplus_k \alpha_k H_k,\ \ \mathcal{G}(\mathbf{x})\right),$$

where d_{struct} is the structural distance measure on representation space [3] and $\mathcal{G}(\mathbf{x})$ is the latent generative structure underlying $\mathbf{x}$.

Synthesis Tension captures a fundamental fact about structural geometry: the centroid of two structural hypotheses is not, in general, a member of either. The weighted average of "condition A" and "condition B" is not a coherent clinical judgment it is a structural non-entity. This is not the same as calibrated uncertainty. A calibrated system can represent the degree of uncertainty between two hypotheses. What no averaging mechanism can do is produce a *committed* structural position from that uncertainty.

Synthesis Tension is therefore not reduced by better calibration. It is reduced only by commitment: selecting one hypothesis and deriving the output exclusively from that hypothesis's structural description.

4.3. Commitment Tension

The third level is the most fundamental. Even when Density Tension is low and hypotheses are structurally adjacent, the absence of a commitment mechanism prevents the dominant hypothesis from crystallizing into a falsifiable position.

Definition 8 (Commitment Tension)**.**

$$T_C(\mathbf{x}) = 1 - \Gamma(\mathbf{x}) = 1 - \frac{\max_k \rho_k(\mathbf{x})}{\sum_k \rho_k(\mathbf{x})}.$$

$T_C = 0$ denotes complete commitment to the dominant hypothesis; $T_C = (K-1)/K$ denotes maximal non-commitment (uniform distribution).

Commitment Tension is the architectural manifestation of what Popper identified as the epistemological deficiency of unfalsifiable claims: a system that distributes activation across hypotheses never commits to a position that experience could contradict [5].

The relationship between T_C and the softmax mechanism is direct. The softmax operator normalizes activation scores to sum to one, ensuring that $\Gamma(\mathbf{x})$ never approaches one unless a single hypothesis receives arbitrarily dominant pre-activation scores. The softmax does not implement commitment; it implements distributional averaging. Commitment Tension is therefore a structural consequence of softmax normalization applied to multi-hypothesis inference not a calibration failure amenable to post-hoc correction, but a necessary consequence of the output mechanism [14].

Table 1 summarizes the three levels and their architectural signatures.

Table 1: The three structural levels of The Implicit Tension, their formal measures, and their predicted architectural signatures under the activation model of Definition 3.

Level	**Measure**	**Architectural signature**
Density Tension T_D	Active hypothesis count	Grows with training data; never pruned
Synthesis Tension T_S	Structural distance of composite	Output structurally foreign to any hypothesis
Commitment Tension T_C	$1 - \Gamma(\mathbf{x})$	Softmax distributes; never severs

5. The Biological Resolution: Constraint as Mechanism

5.1. Reframing Biological Limitations

Cognitive science has long treated certain properties of biological cognition as limitations to be explained and ideally circumvented: the limited capacity of working memory [9], the rapid decay of unattended information, the imprecision of long-term recall. We argue that this framing is precisely inverted. These properties are not incidental deficiencies of the biological implementation. They are functionally critical *commitment resolution mechanisms* whose absence in artificial systems is the structural source of TIT.

5.1.1. Forgetting as Commitment

Biological memory consolidation is selective. Information that is structurally coherent, emotionally salient, or repeatedly rehearsed is preferentially retained; peripheral and competing details decay [11]. This is precisely opposite to what a prediction-accuracy objective produces, which is why gradient-based systems retain all patterns with equal fidelity proportional to their training frequency.

The functional consequence of selective forgetting is a reduction in Density Tension. By structurally pruning peripheral and competing hypotheses from the active set, the biological system converges toward a dominant structural interpretation. This convergence is not the result of having "chosen" the correct hypothesis in advance; it is the result of an architectural mechanism that severs the activation chains sustaining competing hypotheses. Forgetting, properly understood, is not the loss of knowledge. It is the architectural compulsion that enables commitment.

5.1.2. Attentional Gating as Commitment Resolution

The biological attentional system implements a competitive inhibition mechanism: attended representations are amplified; unattended representations are suppressed at early processing stages [10]. At any moment in biological cognition, only a small subset of potentially relevant structural hypotheses is maintained in high-activation state.

This mechanism directly reduces Density Tension by restricting $|\{k : \rho_k(\mathbf{x}) > \theta_D\}|$ to a small bounded set. Crucially, attentional gating does not eliminate access to competing hypotheses; it restricts which hypotheses are *simultaneously active*. This is precisely the architectural property that artificial systems lack: the ability to suppress competing hypotheses from active competition without permanently losing access to them.

5.1.3. Working Memory Constraint as Synthesis Limiter

Miller's observation that working memory holds approximately 7 ± 2 items [9] describes the limit on the number of structural elements that can be simultaneously integrated in active reasoning. This constraint directly limits Synthesis Tension: forced compositional averaging across more than a small number of simultaneously active hypotheses is structurally impossible under working-memory constraints.

The biological system is not averaging across thousands of competing structural hypotheses when it forms a judgment. It is integrating a small, attentionally selected set. The compactness of this integration is not a limitation; it is what makes the integration structurally coherent rather than diffuse.

5.2. The Commitment Resolution Principle

Principle 1 (Commitment Resolution Principle)**.** Genuine structural judgment requires a *commitment resolution mechanism* (CRM): an architectural process that, given a set of simultaneously active structural hypotheses, selects a dominant hypothesis, suppresses competing hypotheses from further activation contribution, and derives the system's output exclusively from the selected hypothesis's structural description.

A CRM is not equivalent to confidence calibration, uncertainty quantification, or ensemble averaging. It is a structural severing of competing hypotheses from the output generation process.

The three biological mechanisms above collectively instantiate the CRM through complementary routes. Forgetting reduces Density Tension over time. Attentional gating reduces it in the moment of judgment. Working memory constraint reduces Synthesis Tension by limiting the inputs to compositional integration. Any adequate artificial CRM must provide functional analogues of all three.

6. Why Knowledge-Rich Systems Cannot Resolve TIT

6.1. The Uniformity of Retention

The optimization objective of current large-scale systems:

$$\theta^* = \arg\min_{\theta} \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}}[\mathcal{L}(f_\theta(\mathbf{x}), y)]$$

does not distinguish between structural hypotheses that are relevant to a given input and those that are not. Every hypothesis that has predictive value anywhere in the training distribution retains activation weight everywhere weighted by its training frequency rather than by its structural relevance to the current input.

Under the activation model of Definition 3, this *uniformity of retention* is the formal cause of Density Tension; the corresponding claim for large-scale systems is stated as Prediction 1 rather than asserted here as an already-demonstrated fact. On any rich input where the prediction holds, a large fraction of a model's learned regularities would be partially consistent with the input's surface features and therefore partially activated, so that the commitment ratio $\Gamma(\mathbf{x})$ is necessarily low not because the system lacks a correct structural hypothesis, but because that hypothesis would compete for activation weight with thousands of others sharing the input's surface [13].

6.2. The Structural Consequence of Softmax

The softmax normalization used in virtually all contemporary architectures implements a specific relationship between competing hypothesis scores:

$$\alpha_k = \frac{e^{\rho_k(\mathbf{x})/T}}{\sum_j e^{\rho_j(\mathbf{x})/T}},$$

where T is a temperature parameter. This normalization ensures $\sum_k \alpha_k = 1$ regardless of the absolute values of ρ_k. The commitment ratio Γ under softmax is therefore bounded by the distribution of pre-softmax scores, which in a knowledge-rich system are broadly distributed.

Even at low temperature ($T \to 0$), the softmax approaches a hard-max function but the hard-max is determined by pre-softmax scores computed from parameters that encode *all* competing hypotheses. The apparent commitment of the hard-max is a mathematical artifact: the output is still derived from all parameters, not from the parameters encoding the selected hypothesis alone. Within the activation model, Commitment Tension is a structural consequence of softmax normalization rather than a calibration failure amenable to post-hoc correction; whether this holds for softmax-based architectures in practice remains open and is stated as Prediction 1 rather than an established fact.

6.3. The Knowledge Richness Paradox

The preceding analysis yields the central structural result of this paper.

Theorem 1 (Knowledge Richness Paradox)**.** *Let $\mathcal{S}_1, \mathcal{S}_2$ be systems trained on i.i.d. samples of sizes $N_1 < N_2$ drawn from the same distribution $P_\mathcal{D}$, sharing the same candidate hypothesis set $\mathcal{H}$ and spurious relevance profile $\{p_k(\mathbf{x})\}$ on judgment domain $\mathcal{D}$ (Definition 3), and possessing no commitment resolution mechanism. Then for every $\mathbf{x} \in \mathcal{D}$, $\mathbb{E}[\mathrm{TIT}(\mathcal{S}_2, \mathbf{x})]$ lies strictly closer, at rate $O(\sqrt{\ln N/N})$, to the structural limit*

$$\mathrm{TIT}_\infty(\mathbf{x}) := w_D\, T_D^\infty(\mathbf{x}) + w_S\, T_S^\infty(\mathbf{x}) + w_C\big(1 - \Gamma_\infty(\mathbf{x})\big)$$

than $\mathbb{E}[\mathrm{TIT}(\mathcal{S}_1, \mathbf{x})]$ is, and $\mathrm{TIT}_\infty(\mathbf{x}) > 0$ strictly whenever $\mathcal{D}$ admits two or more hypotheses with positive relevance rate. In particular, under a fixed-architecture scaling law where $\mathrm{KD}(\mathcal{R}_2) > \mathrm{KD}(\mathcal{R}_1)$ tracks N, $\mathbb{E}[\mathrm{TIT}(\mathcal{S}_1, \mathbf{x})] \leq \mathbb{E}[\mathrm{TIT}(\mathcal{S}_2, \mathbf{x})]$ up to the stated, vanishing convergence error.

Proof. Apply Proposition 1(a)–(b) to T_D and Γ (hence $T_C = 1 - \Gamma$) at sample sizes N_1 and N_2; an analogous Lipschitz-concentration argument extends the same bound to T_S under any structural distance d_{struct} that is Lipschitz in the activation weights (Appendix A, Remark 1). Because the concentration error of Lemma 2 is strictly decreasing in N, $\mathbb{E}[\mathrm{TIT}(\mathcal{S}_2, \mathbf{x})]$ is closer to $\mathrm{TIT}_\infty(\mathbf{x})$ than $\mathbb{E}[\mathrm{TIT}(\mathcal{S}_1, \mathbf{x})]$ by a strictly positive margin of order $O(\sqrt{\ln N_1/N_1}) - O(\sqrt{\ln N_2/N_2})$. Strict positivity of $\mathrm{TIT}_\infty(\mathbf{x})$ follows from Lemma 1. □ □

Scope. Theorem 1 is proved entirely within the activation model of Definition 3 (Section 2.3); its extension to

real trained architectures (e.g. transformers with softmax attention) is stated as a testable claim in Prediction 1 (Section 8), not asserted here as an already-demonstrated empirical fact.

Corollary 1 (Commitment Resolution Moves the Floor)**.** *A mechanism satisfying* (CRM-1) *(Section 7) acts on Definition 3 by replacing* $p_k(\mathbf{x})$ *with a suppressed rate* $p_k^{\mathrm{CRM}}(\mathbf{x}) = p_k(\mathbf{x}) \cdot \mathbb{1}[k \in \mathcal{K}_{\max}(\mathbf{x})]$ *for a selected subset* $\mathcal{K}_{\max}(\mathbf{x})$ *with* $|\mathcal{K}_{\max}(\mathbf{x})| \ll K$*. Under this substitution,* $\Gamma_\infty(\mathbf{x}) \to 1$ *and* $T_D^\infty(\mathbf{x}) \to |\mathcal{K}_{\max}(\mathbf{x})|/K$ *as suppression sharpens toward a singleton* independently of $\mathrm{KD}(\mathcal{R})$*. This is the formal content of Prediction 3: a CRM lowers the structural floor itself, which scale alone provably cannot do (Theorem 1).*

Theorem 1 and Corollary 1 together sharpen the original claim into a falsifiable, rate-quantified statement, and tie the abstract notion of a commitment resolution mechanism directly to the formal model: a CRM is, in this model, precisely a mechanism that zeroes out the relevance rate of suppressed hypotheses rather than merely down-weighting their softmax contribution exactly the distinction drawn informally between (CRM-1) and ordinary attenuation in Section 7.

The widely observed phenomenon that large language models become superficially more "confident" while becoming less reliably calibrated as they scale offers a suggestive, illustrative parallel to the signature predicted by Theorem 1 within the activation model: $\Gamma(\mathbf{x}; N)$ approaches a data-independent floor $\Gamma_\infty(\mathbf{x})$, while softmax-apparent confidence is governed separately by temperature rather than by Γ itself. We draw this parallel for motivation only; no LLM data is tested in this paper, and the corresponding claim about real trained systems is stated formally, and left open to falsification, as Prediction 1.

6.4. Three Misconceptions

The analysis above refutes three common explanations for why current systems fail to form genuine judgments:

Misconception 1: Insufficient training data. More data encodes more competing regularities and deepens Density Tension. Data quantity is a cause of TIT, not its cure.

Misconception 2: Insufficient model capacity. Larger models encode competing hypotheses with greater fidelity, increasing both Density Tension and Commitment Tension.

Misconception 3: Uncertainty quantification solves the problem. Calibrating the distribution over outputs or decomposing uncertainty into aleatoric and epistemic components [12] represents the suspension state more accurately. It does not resolve it. A perfectly calibrated system in Epistemic Suspension is a system that accurately describes how suspended it is, while producing pseudo-judgments with equal precision.

7. TIT within the AI Implicit Programme

7.1. The Missing Theoretical Bridge

The AI Implicit programme [1] has established a principled account of structural knowledge acquisition: Bold Learning [2] describes how genuine structural commitment can be achieved during the learning process. The Structuralist AI vision [3] defines the properties a system must satisfy to count as structurally intelligent.

What has been absent until now is the theoretical account of the gap between *acquiring* structural knowledge and *deploying* that knowledge in genuine structural judgment. A system can acquire structural knowledge through Bold Learning and still be unable to form genuine structural judgments if it lacks a commitment resolution mechanism. Table 2 summarizes the theoretical chain.

Table 2: The theoretical chain of the AI Implicit programme. TIT occupies the previously untheorized position between knowledge acquisition and structural judgment.

Stage	**Theory**	**Central question**
Acquisition	Bold Learning [2]	How does structural knowledge form?
Tension	**TIT (this paper)**	What prevents its deployment as judgment?
Resolution	CRM (future work)	How is the tension architecturally resolved?
Measurement	ECI [4]	How is structural intelligence evaluated?

7.2. The Commitment Resolution Requirement

TIT's theoretical contribution is not only to name a problem but to specify the requirement for its solution. Any architecture adequate to resolving The Implicit Tension must satisfy three necessary conditions:

(CRM-1) Selective Suppression. The mechanism must suppress competing structural hypotheses from contributing to the output, not merely down-weight them. Attenuation without structural suppression leaves Density Tension architecturally intact.

(CRM-2) Hypothesis Exclusivity. The output must be derived from a single dominant structural hypothesis, not from a weighted combination of competing hypotheses. This requires that the output generation pathway isolates the selected hypothesis from the competing set, not just at the level of output probabilities but at the level of representational access.

(CRM-3) Revisability. The commitment must be revisable in response to evidence that the selected hypothesis is structurally inconsistent with subsequent input. Irrevisable commitment is rigidity, not intelligence. The disconfirmation reflex releasing a committed hypothesis when it is falsified and promoting the next-best candidate is the architectural signature of genuine structural judgment under uncertainty.

These three requirements define the design specification for the next architectural element in the AI Implicit programme. They are not properties of any existing architecture. They constitute a genuinely new target.

7.3. Relationship to the Three Principles of AI Implicit

TIT interacts with each of the three principles established in the AI Implicit founding statement [1]:

P1 (Tacit Structure Extraction) reduces the spurious component of Density Tension. A system that extracts structural patterns rather than surface correlations activates fewer competing hypotheses on any given input, because structurally irrelevant hypotheses are not encoded. P1 is a partial reducer of Density Tension, not a CRM.

P2 (Experience Compression) reduces Density Tension by compressing knowledge into a small number of structural prototypes rather than retaining all surface regularities independently. This is the closest existing mechanism in the AI Implicit family to commitment resolution. Prototype compression is a necessary precondition for a CRM; it is not sufficient.

P3 (Epistemic Confidence) provides boundary awareness the capacity to recognize when a novel input is outside the system's structural knowledge. This partially implements epistemic boundary awareness, one of the three properties of genuine judgment. But epistemic awareness without commitment resolution produces a system that knows it does not know without being able to commit even within its domain of genuine knowledge.

The analysis confirms that P1–P3 together are necessary but not sufficient for genuine structural judgment. The Commitment Resolution Requirement constitutes the fourth architectural target, derivable from TIT analysis.

8. Falsifiable Predictions

A theory that makes no falsifiable predictions is philosophy, not science. We state five predictions of the TIT framework, each formulated so that a controlled experiment can confirm or refute it.

Prediction 1 (Knowledge Richness Deepens Suspension)**.** The domain-level TIT score of systems trained without a commitment resolution mechanism is monotonically non-decreasing with training set size N. The commitment ratio $\mathbb{E}[\Gamma(\mathbf{x})]$ will be lower for larger models on the same judgment task. Empirical signature: output structural variance, measured by the standard deviation of structural distances between repeated responses to the same input, increases monotonically with N.

Prediction 2 (Pseudo-Judgment Centroid Bias)**.** Outputs of systems in Epistemic Suspension are statistically biased toward the centroid of the training distribution on the queried topic rather than toward the structurally dominant position supported by the most relevant evidence. Formally: for most inputs, $d_{\text{struct}}(\hat{J}(\mathbf{x}), \mu_{\mathcal{D}}) < d_{\text{struct}}(\hat{J}(\mathbf{x}), H^*(\mathbf{x}))$, where $\mu_{\mathcal{D}}$ is the distributional centroid and $H^*(\mathbf{x})$ is the structurally optimal hypothesis.

Prediction 3 (CRM Reduces TIT Without Reducing Knowledge)**.** Injecting an artificial commitment resolution mechanism such as a hard attentional bottleneck restricting active hypotheses to $K_{\max} \ll K$—will reduce T_D and T_C without reducing KD($\mathcal{R}$). The empirical signature: CRM-augmented systems produce structurally more specific outputs (lower Synthesis Tension) without reduction in Cross-Domain Retention on structurally shared tasks.

Prediction 4 (Expert Commitment in Biological Reference Systems)**.** Domain experts will exhibit near-zero Commitment Tension ($T_C \approx 0$) on inputs within their domain of expertise, measured by the structural specificity and falsifiability of their judgments. Novices in the same domain will exhibit higher T_C—not because they possess less knowledge in absolute terms, but because their structural knowledge is not organized into a small set of discriminating, dominant hypotheses with active competition suppressed.

Prediction 5 (Commitment Improves Transfer, Not Local Accuracy)**.** A system with a commitment resolution mechanism will exhibit higher Cross-Domain Retention (CDR) than an equivalent system without one, because outputs derived from a single committed structural hypothesis carry structural identity that generalizes. Systems in Epistemic Suspension produce structurally diffuse outputs whose surface accuracy on the source domain may be higher, but whose structural content does not transfer to domains sharing latent geometry.

9. Positioning: What TIT Is Not

Scientific clarity requires explicit statement of what TIT does not claim.

TIT is not a critique of scale. The analysis is a structural one. TIT applies to any learning system that acquires structural knowledge through prediction-accuracy optimization without a commitment resolution mechanism, regardless of its size. Scale amplifies TIT; the critique is architectural, not institutional.

TIT is not equivalent to the hallucination prob-

lem. Hallucination is the production of factually incorrect outputs; TIT describes the production of structurally uncommitted outputs. These are orthogonal. A system in Epistemic Suspension may produce factually accurate pseudo-judgments the centroid of the training distribution may happen to be correct while remaining in full suspension. A committed system may form a structurally genuine judgment that is factually wrong.

TIT is not solved by RLHF or instruction tuning. Reinforcement Learning from Human Feedback shapes the surface properties of outputs without altering the structural competition among hypotheses. A system trained to produce more confident-sounding outputs is producing better-formatted pseudo-judgments, not resolving Epistemic Suspension.

TIT is not the alignment problem. Alignment concerns the relationship between a system's objective and human values. TIT concerns the relationship between a system's knowledge and its capacity for structural judgment, independently of any external value system. A perfectly aligned system may still be in Epistemic Suspension.

TIT is not an argument for less knowledge. The solution to TIT is a commitment resolution mechanism, not reduced knowledge. The goal is to combine rich structural knowledge (acquired through Bold Learning) with a mechanism that deploys it in committed structural judgments.

10. Comparison: Suspension vs. Genuine Judgment

Table 3 contrasts the epistemic properties of a system in Epistemic Suspension with those of a system equipped with a commitment resolution mechanism.

Table 3: Epistemic Suspension vs. Committed Structural Judgment.

Property	**Suspension**	**Commitment**
Output type	Weighted centroid	Single-hypothesis derived
Falsifiability	None (never commits)	Yes (structural claim)
Confidence source	Softmax geometry	Hypothesis density
OOD behavior	Confident pseudo-J.	Low confidence or abstention
Transfer	Surface-dependent	Structure-dependent
Scale effect	Deepens suspension	Sharpens commitment
Knowledge use	Averaged away	Deployed as hypothesis

11. Limitations and Honest Assessment

The TIT score requires structural hypothesis identification. Computing T_D, T_S, and T_C requires identifying the active hypothesis set $\mathcal{H}$ and evaluating $\rho_k(\mathbf{x})$ for each hypothesis. In current large-scale systems, structural hypotheses are not explicitly represented; they are distributed across parameter vectors. Practical evaluation of TIT requires approximation methods (probing classifiers, activation clustering) that may not fully capture the hypothesis competition the theory describes. The theoretical framework is precise; the practical measurement is an active research problem.

The commitment threshold θ_Γ is domain-dependent. The threshold below which Epistemic Suspension is declared is not universal. Domains with genuinely competing structural hypotheses may require lower θ_Γ than domains with clear structural dominance. This domain-dependence does not undermine the framework but requires empirical calibration per domain.

The commitment resolution requirement is stated, not implemented. This paper identifies TIT and specifies the necessary conditions for its resolution (CRM-1 through CRM-3). It does not propose a concrete architectural implementation. The development of architectures satisfying these conditions is the next phase of the AI Implicit programme.

The biological analogy requires careful translation. The claim that forgetting and attentional constraint are commitment resolution mechanisms is supported by functional analysis and cognitive science literature, but the precise translation from biological mechanism to artificial architectural requirement is not yet fully specified [9–11]. The biological evidence motivates the design target; it does not uniquely determine the implementation path.

12. Conclusion

We have introduced The Implicit Tension (TIT) as a structural theory that names and formalizes the central obstacle preventing knowledge-rich artificial systems from forming genuine committed judgments. Four results have been established.

Epistemic Suspension is a structural state, not an epistemic accident. It is the necessary consequence of possessing rich structural knowledge sufficient, in principle, to ground genuine judgments without a mechanism that resolves competition among simultaneously active hypotheses

into a committed position.

The Implicit Tension is not reducible to insufficient knowledge. Through the Knowledge Richness Paradox (Theorem 1), we have shown, within the formal activation model of Definition 3, that scaling knowledge without commitment resolution strictly increases TIT, and we predict the same pattern for real trained systems as Prediction 1. Under this model, the systems pushed deepest into Epistemic Suspension are those with the richest knowledge.

The biological solution to TIT has been systematically misclassified as a limitation. Forgetting, attentional gating, and working-memory constraint are commitment resolution mechanisms: architectural compulsions that sever competing hypotheses and crystallize judgment. The path forward for artificial systems is not to eliminate these properties but to architect their functional equivalents.

TIT fills the theoretical gap between acquisition and judgment within the AI Implicit programme. Bold Learning addresses how structural knowledge is acquired. TIT addresses what prevents that knowledge from being deployed as genuine judgment. Commitment Resolution specifies the architectural requirement for bridging the gap.

The deeper claim of this paper is epistemological. A system that cannot commit cannot genuinely know because knowledge, in the sense that matters for intelligence, is not the accumulation of structural representations but the deployment of those representations in falsifiable, structurally grounded, boundary-aware positions. The Implicit Tension is the structural name for the gap between having structural knowledge and being able to say something true with it.

Resolving TIT is not an engineering refinement. It is the next theoretical problem that stands between artificial systems and genuine structural intelligence.

A. Formal Proofs

This appendix gives the complete, non-asymptotic proofs underlying Proposition 1 and Theorem 1, under the activation model of Definition 3. Throughout, $\mathbf{x}$ is a fixed query, $S(\mathbf{x}) := \sum_{k=1}^{K} p_k(\mathbf{x})$, $m(\mathbf{x}) := \max_k p_k(\mathbf{x})$, and $\hat{p}_k(N) := \rho_k(\mathbf{x}; N)/N$.

Lemma 1 (Strong Consistency of the Commitment Ratio)**.** *As $N \to \infty$, $\Gamma(\mathbf{x}; N) \to \Gamma_\infty(\mathbf{x}) := m(\mathbf{x})/S(\mathbf{x})$ almost surely, and $\Gamma_\infty(\mathbf{x}) < 1$ strictly whenever at least two hypotheses satisfy $p_k(\mathbf{x}) > 0$.*

Proof. By the Strong Law of Large Numbers, $\hat{p}_k(N) \to p_k(\mathbf{x})$ almost surely for each k; since K is finite this convergence holds jointly by a union over k. The map $f(q_1, \ldots, q_K) = \max_k q_k / \sum_k q_k$ is continuous on $\mathbb{R}^K_{\geq 0} \setminus \{\mathbf{0}\}$, and $S(\mathbf{x}) > 0$ by assumption (ES); the continuous mapping theorem gives $\Gamma(\mathbf{x}; N) = f(\hat{p}_1(N), \ldots, \hat{p}_K(N)) \to f(p_1(\mathbf{x}), \ldots, p_K(\mathbf{x})) = \Gamma_\infty(\mathbf{x})$ almost surely. If a second hypothesis $j \neq k^*$ (where $k^* = \arg\max_k p_k(\mathbf{x})$) has $p_j(\mathbf{x}) > 0$, then $S(\mathbf{x}) = m(\mathbf{x}) + p_j(\mathbf{x}) + (\text{non-negative terms}) > m(\mathbf{x})$, so $\Gamma_\infty(\mathbf{x}) = m(\mathbf{x})/S(\mathbf{x}) < 1$. □ □

Lemma 2 (Finite-Sample Concentration)**.** *For any $\delta \in (0, 1)$ and N large enough that $\epsilon_N := \sqrt{\frac{1}{2N} \ln(2K/\delta)} \leq S(\mathbf{x})/(2K)$, with probability at least $1 - \delta$,*

$$\big|\Gamma(\mathbf{x}; N) - \Gamma_\infty(\mathbf{x})\big| \;\leq\; C(\mathbf{x})\, \epsilon_N,$$

$$C(\mathbf{x}) := \frac{2\big(S(\mathbf{x}) + m(\mathbf{x})K\big)}{S(\mathbf{x})^2}.$$

Consequently $\mathbb{E}\big[|\Gamma(\mathbf{x}; N) - \Gamma_\infty(\mathbf{x})|\big] = O\big(\sqrt{\ln N / N}\big)$.

Proof. By Hoeffding's inequality [21], for fixed k, $\mathbb{P}[\,|\hat{p}_k(N) - p_k(\mathbf{x})| > \epsilon\,] \leq 2e^{-2N\epsilon^2}$. A union bound over $k = 1, \ldots, K$ with $\epsilon = \epsilon_N$ gives $\mathbb{P}[\,\exists k : |\hat{p}_k(N) - p_k(\mathbf{x})| > \epsilon_N\,] \leq 2Ke^{-2N\epsilon_N^2} = \delta$, so with probability at least $1 - \delta$, $|\hat{p}_k(N) - p_k(\mathbf{x})| \leq \epsilon_N$ holds *simultaneously* for all k. Write $m' = \max_k \hat{p}_k(N)$, $S' = \sum_k \hat{p}_k(N)$. Since $\max_k(\cdot)$ is 1-Lipschitz in the ℓ_∞ norm, $|m' - m(\mathbf{x})| \leq \epsilon_N$; and $|S' - S(\mathbf{x})| \leq \sum_k |\hat{p}_k(N) - p_k(\mathbf{x})| \leq K\epsilon_N$. A direct computation gives

$$\begin{aligned}\Gamma(\mathbf{x}; N) - \Gamma_\infty(\mathbf{x}) &= \frac{m'}{S'} - \frac{m(\mathbf{x})}{S(\mathbf{x})} \\ &= \frac{S(\mathbf{x})\big(m' - m(\mathbf{x})\big) + m(\mathbf{x})\big(S(\mathbf{x}) - S'\big)}{S(\mathbf{x})\, S'},\end{aligned}$$

so $|\Gamma(\mathbf{x}; N) - \Gamma_\infty(\mathbf{x})| \leq \dfrac{\epsilon_N\big(S(\mathbf{x}) + m(\mathbf{x})K\big)}{S(\mathbf{x})\, S'}$. When $\epsilon_N \leq S(\mathbf{x})/(2K)$, $S' \geq S(\mathbf{x}) - K\epsilon_N \geq S(\mathbf{x})/2$, giving $1/(S(\mathbf{x})S') \leq 2/S(\mathbf{x})^2$ and the stated bound with $C(\mathbf{x})$. For the expectation bound, set $\delta = 1/N$: with probability $\geq 1 - 1/N$ the deviation is $O(\sqrt{\ln N/N})$, and on the complementary event of probability $1/N$ the deviation is trivially bounded by 1 (since $\Gamma, \Gamma_\infty \in [1/K, 1]$); combining the two contributions gives $\mathbb{E}[|\Gamma(\mathbf{x}; N) - \Gamma_\infty(\mathbf{x})|] = O(\sqrt{\ln N/N}) + O(1/N) = O(\sqrt{\ln N/N})$. □ □

Lemma 3 (Monotone Threshold Crossing)**.** *Let $K_{>\theta_D}(\mathbf{x}) = \{k : p_k(\mathbf{x}) > \theta_D\}$. For $k \in K_{>\theta_D}(\mathbf{x})$, the inclusion probability $\pi_k(N) := \mathbb{P}[\hat{p}_k(N) > \theta_D]$ satisfies*

$$\pi_k(N) \;\geq\; 1 - \exp\big(-2N(p_k(\mathbf{x}) - \theta_D)^2\big),$$

a lower bound strictly increasing in N and converging to 1. Consequently $\mathbb{E}[T_D(\mathbf{x}; N)] \to T_D^\infty(\mathbf{x}) := |K_{>\theta_D}(\mathbf{x})|/K$, bounded below by a quantity strictly increasing in N.

Proof. For $k \in K_{>\theta_D}(\mathbf{x})$, $p_k(\mathbf{x}) - \theta_D > 0$, so the one-sided Hoeffding bound gives $\mathbb{P}[\hat{p}_k(N) \leq \theta_D] \leq \exp\big(-2N(p_k(\mathbf{x}) - \theta_D)^2\big)$, a quantity strictly decreasing

in N since the exponent is strictly negative and linear in N. Hence $\pi_k(N) = 1 - \mathbb{P}[\hat{p}_k(N) \leq \theta_D]$ is bounded below by a quantity strictly increasing to 1. Since $T_D(\mathbf{x}; N) = \frac{1}{K}\sum_{k=1}^{K} \mathbb{1}[\hat{p}_k(N) > \theta_D]$, we have $\mathbb{E}[T_D(\mathbf{x}; N)] = \frac{1}{K}\sum_{k=1}^{K} \pi_k(N) \geq \frac{1}{K}\sum_{k \in K_{>\theta_D}(\mathbf{x})} \pi_k(N)$, which is bounded below by a quantity strictly increasing in N and converging (by dominated convergence, since each $\pi_k(N) \in [0,1]$) to $|K_{>\theta_D}(\mathbf{x})|/K$. □ □

Remark 1 (Extension to Synthesis Tension)**.** Let $\alpha_k(N) := \hat{p}_k(N)/S'$ denote the empirical composition weights of Definition 7, and $\alpha_k^\infty := p_k(\mathbf{x})/S(\mathbf{x})$ their asymptotic counterparts. If $d_{\text{struct}}(\cdot, \mathcal{G}(\mathbf{x}))$ is L-Lipschitz on the representation space and the composite map $\alpha \mapsto \bigoplus_k \alpha_k H_k(\mathbf{x})$ is L'-Lipschitz in α (true whenever the K hypothesis representations $H_k(\mathbf{x})$ are bounded), then $\alpha_k(N) \to \alpha_k^\infty$ at the same rate as Lemma 2 (both are ratios of the same empirical rates $\hat{p}_k(N)$ to their sum), and composing Lipschitz maps gives

$$T_S(\mathbf{x}; N) \longrightarrow T_S^\infty(\mathbf{x}) := d_{\text{struct}}\Big(\bigoplus_k \alpha_k^\infty H_k(\mathbf{x}),\, \mathcal{G}(\mathbf{x})\Big)$$

almost surely, with deviation $|T_S(\mathbf{x}; N) - T_S^\infty(\mathbf{x})| = O(\sqrt{\ln N/N})$ in expectation, by the same argument as Lemma 2 with Lipschitz constant LL' in place of $C(\mathbf{x})$. This is the result invoked in the proof of Theorem 1.

Lemmas 1–3 and Remark 1 together prove Proposition 1 and, via the argument given in Section 6, Theorem 1. Appendix B reports a Monte Carlo numerical verification of these results.

B. Numerical Verification of the Activation Model

This appendix reports a Monte Carlo numerical verification of Definition 3 and the results built on it (Lemmas 1–3, Remark 1, Proposition 1, Theorem 1, Corollary 1). All computations below simulate the activation model directly: for a fixed relevance profile $\{p_k(\mathbf{x})\}_{k=1}^K$, we draw $\rho_k(\mathbf{x}; N) \sim \text{Binomial}(N, p_k(\mathbf{x}))$ independently across k and M Monte Carlo trials, and compare the resulting empirical quantities to their theoretical counterparts derived above.

Scope. This verification tests the internal consistency and finite-sample behavior of the formal probabilistic model already proven above; it is not an empirical test of any real, deployed, or trained AI system. The assumptions of Definition 3 (independent per-hypothesis relevance indicators, fixed relevance rates $p_k(\mathbf{x})$) are a theoretical construct defined for the purpose of this proof. No claim is made here, or elsewhere in this appendix, about the empirical behavior of large language models or any other deployed learned system.

B.1. Configuration and Reproducibility

Random seed 20260701; density threshold $\theta_D = 0.05$; TIT weights $w_D = w_S = w_C = 1/3$. Three relevance-profile families are used: *power-law* ($p_k = 0.6/k^{1.3}$), *near-uniform* ($p_k \sim \text{Unif}(0.08, 0.12)$), and *random* ($p_k \sim \text{Unif}(0.01, 0.6)$). Unless stated otherwise, $K = 20$ hypotheses (power-law family), $M = 3000$ trials per value of $N \in \{10, 30, 100, 300, 10^3, 3\times10^3, 10^4, 3\times10^4, 10^5\}$. The Synthesis Tension instantiation (Section B.4 below) uses representation dimension 16.

Reproducibility. Random seed, thresholds, weights, and profile families are as stated above. The Monte Carlo notebook implementing this verification is available as supplementary material at `[code repository / supplementary link]`.

B.2. Consistency and Concentration of the Commitment Ratio

Table 4 reports the empirical commitment ratio against $\Gamma_\infty = 0.38684$ (Lemma 1). The empirical decay rate of $\mathbb{E}|\Gamma(\mathbf{x}; N) - \Gamma_\infty(\mathbf{x})|$, fit by ordinary least squares in log-log coordinates across the full N grid, is $N^{-0.502}$, matching the $O(\sqrt{\ln N/N})$ rate of Lemma 2 closely. The bound's precondition ($\epsilon_N \leq S(\mathbf{x})/(2K)$) is satisfied only once $N \gtrsim 2500$ at this K; below that N the bound is undefined, which is the expected behavior of the proof and not a failure. At each N for which the precondition holds, the empirical coverage of the non-asymptotic bound $C(\mathbf{x})\epsilon_N$ at $\delta = 0.05$ is 100% across $M = 3000$ trials, consistent with the guaranteed coverage of at least $1 - \delta = 95\%$.

Table 4: Empirical commitment ratio vs. N ($\Gamma_\infty = 0.38684$).

N	mean Γ	mean abs. dev.	Lemma-2 bound	coverage
10	0.39559	0.07462	n/a	n/a
30	0.38987	0.04464	n/a	n/a
100	0.38876	0.02433	n/a	n/a
300	0.38714	0.01443	n/a	n/a
1,000	0.38702	0.00739	n/a	n/a
3,000	0.38678	0.00444	0.37603	100%
10,000	0.38688	0.00240	0.20596	100%
30,000	0.38684	0.00138	0.11891	100%
100,000	0.38686	0.00075	0.06513	100%

Figure 1 plots this convergence directly.

B.3. Density Tension and the Threshold-Crossing Bound

Table 5 reports two distinct quantities against $T_D^\infty = 0.30$ (6 of $K = 20$ hypotheses exceed θ_D): the raw empirical $T_D(\mathbf{x}; N)$, and $\bar{\pi}(N)$, the average over $k \in K_{>\theta_D}(\mathbf{x})$ of the per-hypothesis lower bound of Lemma 3. Proposition 1(a), proved via Lemma 3, claims that $\mathbb{E}[T_D(\mathbf{x}; N)]$ is

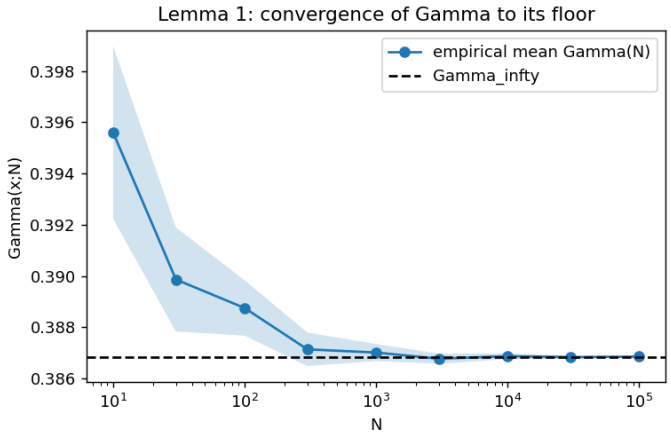

Figure 1: *Lemma 1: the empirical mean of* $\Gamma(\mathbf{x}; N)$ *converges to its asymptotic floor* $\Gamma_\infty = 0.38684$ *as* N *grows; the shaded band is one empirical standard deviation across* $M = 3000$ *trials.*

bounded below by a quantity strictly increasing in N that converges to T_D^∞; $\bar{\pi}(N)$ is exactly that quantity, and it increases strictly and monotonically at every step in the table (from 0.291 to 1.000), confirming the proposition as literally stated. The raw $T_D(\mathbf{x}; N)$ column is not itself claimed to be monotonic, and indeed overshoots the ceiling at small N (e.g. 0.369 at $N = 10$ against a ceiling of 0.300): hypotheses just below θ_D are occasionally pushed above threshold by sampling noise when N is small, inflating T_D temporarily; this false-positive contribution vanishes as N grows and $\hat{p}_k(N)$ concentrates on $p_k(\mathbf{x})$.

Table 5: Density Tension: raw T_D vs. the Lemma 3 lower-bound envelope $\bar{\pi}(N)$ ($T_D^\infty = 0.30$).

N	raw T_D	$\bar{\pi}(N)$
10	0.369	0.291
30	0.363	0.413
100	0.308	0.555
300	0.311	0.682
1,000	0.313	0.802
3,000	0.312	0.886
10,000	0.308	0.960
30,000	0.302	0.998
100,000	0.300	1.000

Figure 2 contrasts the two behaviors described above.

B.4. Synthesis Tension

To instantiate T_S numerically, each hypothesis representation $H_k(\mathbf{x})$ is drawn i.i.d. from $\mathcal{N}(0, I_{16})$, $\mathcal{G}(\mathbf{x})$ is set to the representation of the dominant hypothesis, and d_{struct} is Euclidean distance rescaled by the representation set's diameter; this is a modeling choice for the purpose of numerical demonstration and not part of the abstract theory, which leaves the representation space and d_{struct} unspecified beyond the Lipschitz condition of Remark 1. Table 6 shows $T_S(\mathbf{x}; N)$ converging to $T_S^\infty = 0.2759$ at the rate predicted by Remark 1.

Table 6: Synthesis Tension convergence ($T_S^\infty = 0.2759$, rescaled units).

N	mean T_S
10	0.2880
30	0.2793
100	0.2771
300	0.2762
1,000	0.2759
3,000	0.2759
10,000	0.2759
30,000	0.2759
100,000	0.2759

Figure 3 shows this convergence.

B.5. The Knowledge Richness Paradox

To test Theorem 1 directly, we drew 200 independent domain profiles (power-law family, $K = 150$) and, for each, compared a low-N system ($N_1 = 30$) against a high-N system ($N_2 = 10^6$), with $M = 10{,}000$ trials per system per domain. N_1 and N_2 were chosen far apart specifically so that the $O(\sqrt{\ln N / N})$ separation predicted by the theorem would be numerically resolvable against Monte Carlo sampling noise; the result below establishes the direction and existence of the effect and should not be read as a statement about its magnitude at arbitrary or realistic finite N.

The theorem's claim, $|\mathbb{E}[\text{TIT}(S_2, \mathbf{x})] - \text{TIT}_\infty(\mathbf{x})| < |\mathbb{E}[\text{TIT}(S_1, \mathbf{x})] - \text{TIT}_\infty(\mathbf{x})|$, held in 200/200 (100%) of the sampled domains. Table 7 reports the distribution, across the 200 domains, of each system's distance to $\text{TIT}_\infty(\mathbf{x})$. The two distributions do not overlap: the largest observed gap for S_2 (2.96×10^{-6}) is roughly three orders of magnitude below the smallest observed gap for S_1 (3.33×10^{-3}).

Table 7: Distance to TIT_∞ across 200 domains, $N_1 = 30$ vs. $N_2 = 10^6$.

	gap at $N_1 = 30$	gap at $N_2 = 10^6$
mean	4.30×10^{-3}	9.70×10^{-7}
std	3.72×10^{-4}	6.70×10^{-7}
min	3.33×10^{-3}	2.90×10^{-9}
max	5.25×10^{-3}	2.96×10^{-6}

Figure 4 illustrates this approach to the structural limit across the full N grid.

B.6. Commitment Resolution Mechanism

Table 8 applies the CRM substitution of Corollary 1: only the top-m hypotheses by $p_k(\mathbf{x})$ are retained, the rest set to

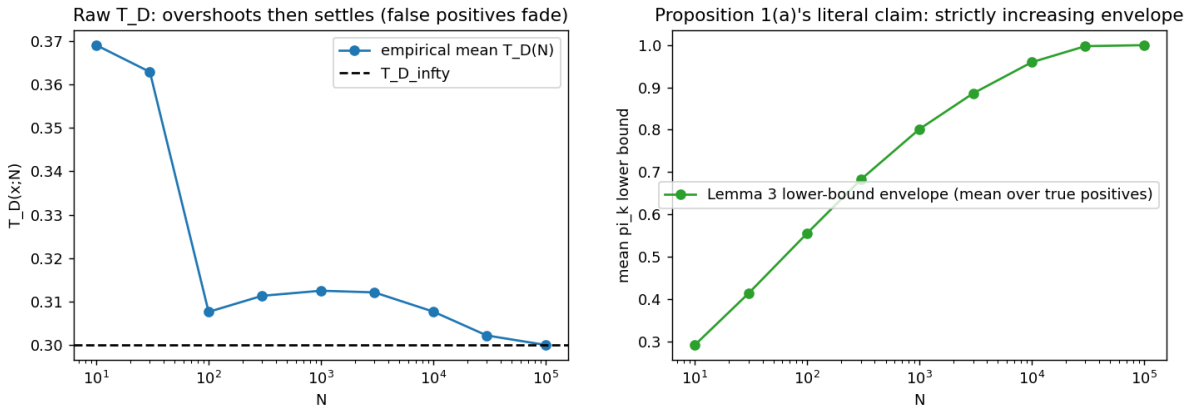

Figure 2: *Left: the raw empirical* $T_D(\mathbf{x}; N)$ *overshoots the ceiling* $T_D^\infty = 0.30$ *at small* N *before settling, as sampling-noise false positives fade. Right: the Lemma 3 lower-bound envelope* $\bar{\pi}(N)$*, which increases strictly and monotonically at every step, confirming Proposition 1(a) as literally stated.*

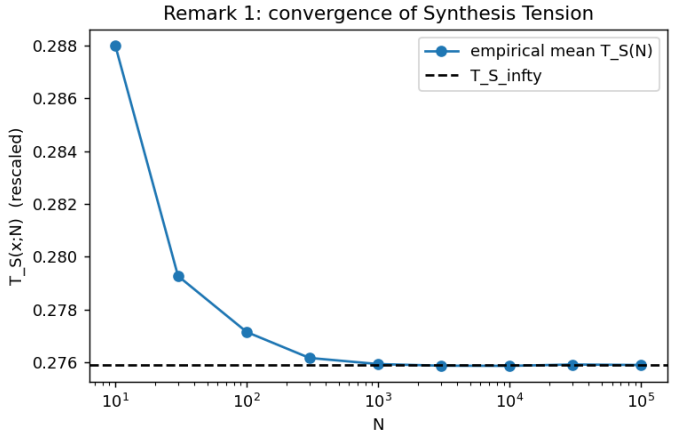

Figure 3: *Remark 1: the empirical mean of* $T_S(\mathbf{x}; N)$ *(rescaled units) converges to* $T_S^\infty = 0.2759$ *at the predicted rate.*

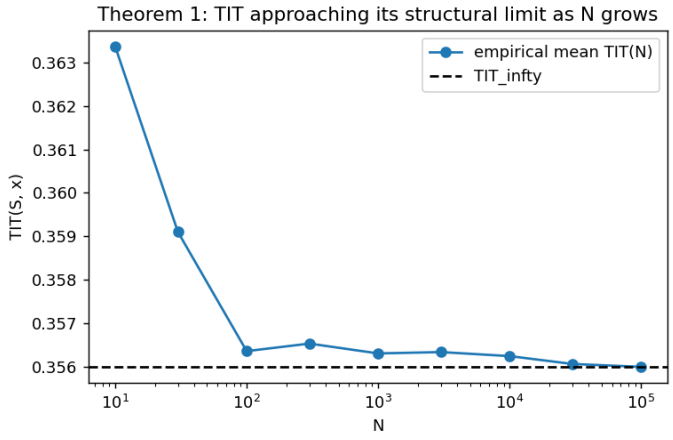

Figure 4: *Theorem 1: the empirical mean of* $\mathrm{TIT}(S, \mathbf{x})$ *approaches its structural limit* TIT_∞ *as* N *grows, the convergence pattern underlying the Knowledge Richness Paradox.*

zero. Two claims of the corollary are checked separately. First, $\Gamma_\infty(m) \to 1$ as $m \to 1$, and this value is identical across $K \in \{50, 200, 1000\}$ at every m tested: suppression, not scale, moves the commitment floor, with no exception in the table. Second, the identity $T_D^\infty(m) = m/K$ holds exactly wherever every retained hypothesis individually exceeds θ_D (marked ✓); for larger m, some retained low-rank hypotheses fall below θ_D (marked ×), which is an expected boundary effect of the threshold definition and not a contradiction of the corollary.

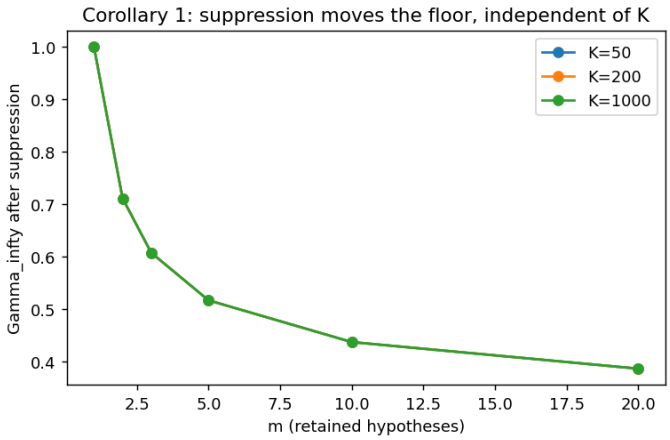

Figure 5: *Corollary 1:* Γ_∞ *after suppression as a function of the number of retained hypotheses* m*, identical across* $K \in \{50, 200, 1000\}$ *(the three curves overlap exactly) suppression, not scale, moves the commitment floor.*

Figure 5 shows the resulting floor as a function of m.

B.7. Robustness Across Domain Profiles

Table 9 repeats the asymptotic computation of Γ_∞, T_D^∞, and TIT_∞ across the three relevance-profile families of Section B.1 and K ranging over two orders of magnitude (30

Table 8: Corollary 1: suppression moves the commitment floor independently of K.

K	m	$\Gamma_\infty(m)$	$T_D^\infty(m)$	m/K	applicable
50	1	1.0000	0.020	0.020	✓
50	2	0.7112	0.040	0.040	✓
50	3	0.6076	0.060	0.060	✓
50	5	0.5170	0.100	0.100	✓
50	10	0.4375	0.120	0.200	×
50	20	0.3868	0.120	0.400	×
200	1	1.0000	0.005	0.005	✓
200	2	0.7112	0.010	0.010	✓
200	3	0.6076	0.015	0.015	✓
200	5	0.5170	0.025	0.025	✓
200	10	0.4375	0.030	0.050	×
200	20	0.3868	0.030	0.100	×
1,000	1	1.0000	0.001	0.001	✓
1,000	2	0.7112	0.002	0.002	✓
1,000	3	0.6076	0.003	0.003	✓
1,000	5	0.5170	0.005	0.005	✓
1,000	10	0.4375	0.006	0.010	×
1,000	20	0.3868	0.006	0.020	×

repeats per cell). $\mathrm{TIT}_\infty > 0$ in all 18 configurations, confirming that the qualitative pattern of Corollary 1 (a strictly positive floor whenever at least two hypotheses are simultaneously active) is not an artifact of one chosen parameterization; the three families nonetheless differ substantially in how Γ_∞ and T_D^∞ trade off, with the near-uniform family producing the largest TIT_∞ (all hypotheses remain above θ_D regardless of K) and the power-law family the smallest.

Figure 6 visualizes this trade-off across all three families and the full K range.

B.8. Summary

Every result checked above (Lemmas 1–3, Remark 1, Theorem 1, Corollary 1, and the robustness sweep) is numerically consistent with its stated proof. As emphasized above, this appendix verifies the formal activation model of Definition 3 under the representation-space instantiation of Section B.1; Section 11's point that practical measurement of TIT in deployed systems remains an open problem is unaffected by, and not addressed by, this verification.

References

[1] Ghazouani, M. (2026). AI Implicit: A foundational paradigm for intelligence through experience compression. *Zenodo*. https://doi.org/10.5281/zenodo.19947976

[2] Ghazouani, M. (2026). Bold Learning: The learning mechanism of AI Implicit. *Zenodo*. https://doi.org/10.5281/zenodo.20090957

[3] Ghazouani, M. (2026). Structuralist Artificial Intelligence: Intelligence as the formation and transfer of relational structure from experience. *Zenodo*. https://doi.org/10.5281/zenodo.19979509

[4] Ghazouani, M. (2026). Experience-Compressed Intelligence (ECI): A measurement framework for Structuralist AI. *Zenodo*. https://doi.org/10.5281/zenodo.19957177

[5] Popper, K. R. (1959). *The Logic of Scientific Discovery*. Hutchinson.

[6] Polanyi, M. (1966). *The Tacit Dimension*. University of Chicago Press.

[7] Kolmogorov, A. N. (1963). On tables of random numbers. *Sankhyā: The Indian Journal of Statistics, Series A*, 369–376.

[8] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.

[9] Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.

[10] Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222.

Table 9: Robustness of the asymptotic quantities across profile family and K (30 repeats per cell).

family	K	Γ_∞	T_D^∞	TIT_∞
power-law	20	0.3868	0.300	0.4042
power-law	50	0.3443	0.120	0.3607
power-law	100	0.3230	0.060	0.3457
power-law	300	0.3003	0.020	0.3317
power-law	1,000	0.2847	0.006	0.3348
power-law	2,000	0.2785	0.003	0.3293
near-uniform	20	0.0590	1.000	0.8038
near-uniform	50	0.0238	1.000	0.8053
near-uniform	100	0.0120	1.000	0.8042
near-uniform	300	0.0040	1.000	0.7951
near-uniform	1,000	0.0012	1.000	0.7919
near-uniform	2,000	0.0006	1.000	0.7916
random	20	0.0960	0.927	0.7671
random	50	0.0389	0.933	0.7799
random	100	0.0195	0.937	0.7807
random	300	0.0065	0.930	0.7751
random	1,000	0.0020	0.931	0.7649
random	2,000	0.0010	0.933	0.7611

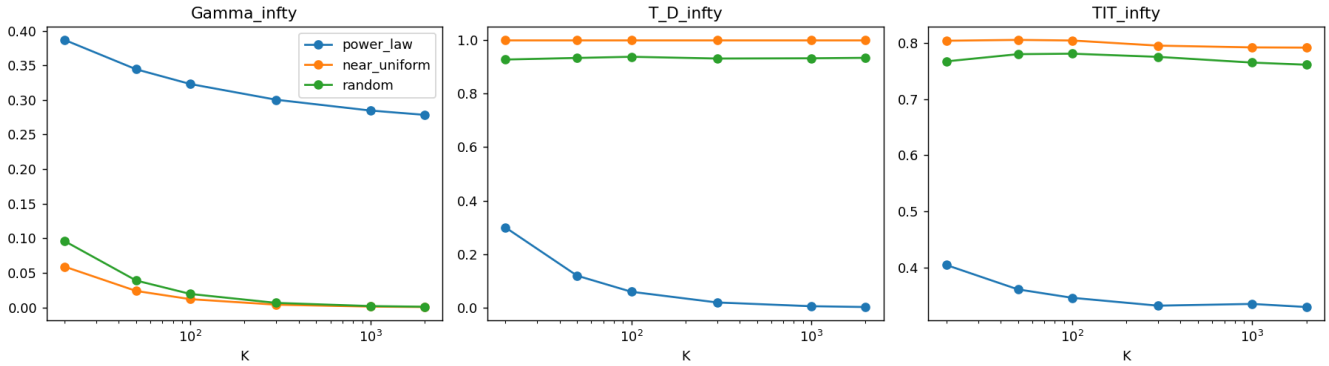

Figure 6: *Robustness of the asymptotic quantities* Γ_∞, T_D^∞, *and* TIT_∞ *as a function of* K, *across the three relevance-profile families (power-law, near-uniform, random);* $\mathrm{TIT}_\infty > 0$ *in every configuration, corresponding to Table 9.*

[11] Anderson, M. C., & Bjork, E. L. (2000). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5), 1063–1087.

[12] Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110(3), 457–506.

[13] Geirhos, R., et al. (2020). Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2(11), 665–673.

[14] Nguyen, A., Yosinski, J., & Clune, J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *Proc. CVPR*, 427–436.

[15] Lakshminarayanan, B., Pritzel, A., & Blum, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *NeurIPS*, 30, 6402–6413.

[16] Kahneman, D. (2011). *Thinking, Fast and Slow.* Farrar, Straus and Giroux.

[17] Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4(1), 55–81.

[18] Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory.* Wiley.

[19] Pearl, J. (2009). *Causality: Models, Reasoning and Inference* (2nd ed.). Cambridge University Press.

[20] Battaglia, P. W., et al. (2018). Relational inductive biases, deep learning, and graph networks. *arXiv:1806.01261.*

[21] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301), 13–30.

Note. This paper belongs to the foundational set of works within the author's broader research program on *Implicit Intelligence*, alongside the author's other papers in this series. Collectively, these papers establish the conceptual and methodological groundwork upon which subsequent developments in this research program are built.