PENTA-STATE PROBABILISTIC MODEL

Nabil Islam
Independent Researcher
w3nabil@gmail.com

October 19, 2026

ABSTRACT

Probability is usually represented either as a binary truth value or as a fuzzy extension of it, and neither form describes well how people hold a partial belief, such as "probably yes" or "probably no". We ask whether this kind of human belief and decision making can be written as a probabilistic model. We introduce the theoretical foundation of the Penta-State Probabilistic Model (PSPM), a five-valued probabilistic scoring and readout model for directional confidence estimation in AI reasoning, inspired by the human decision making process. PSPM describes a probability with five understandable states: True, False, Partial-True, Partial-False, and Undecided. It is designed for short evidence checklists (roughly 5–10 major evidences) and turns evidence counts into an evidence ratio in $[0, 1]$ through two stages. The Major stage chooses the positive, negative or neutral pole and how strong it is, while the Minor stage only breaks exact ties between major evidences, so partial evidence alone can never reach True or False. Unanswered evidence can only lower the strength of the result, which stops PSPM from overclaiming. We prove the structural properties of the score (state partition, closure, mirror symmetry, exact state thresholds, and monotonicity in a restricted case) and state clearly where they do not hold.

Keywords probabilistic reasoning, uncertainty modelling, cognitive modelling, human-aligned AI, AI confidence estimation

1 Introduction

Human reasoning often includes directional uncertainty such as "maybe-yes" or "maybe-no", a bias toward "yes" or "no", and also pure ignorance. These are hard to read out of a single probability value or a plain true/false output. This paper introduces the Penta-State Probabilistic Model (PSPM), a five-valued probabilistic scoring and readout model designed to address these gaps. PSPM turns a short checklist of evidences into one evidence ratio $ER \in [0, 1]$ and reads it as one of five states: True (T), Partial True (PT), Undecided (U), Partial False (PF), and False (F).

PSPM works with two types of evidence: major evidence and minor (partial) evidence. Minor evidence is rarely decisive on its own, but it still affects how people reach a decision. A core example is a loan decision. Whether the applicant passed the required checks (identity, income, existing debt, and so on) is major evidence, while an optional reference letter is minor evidence. A person does not approve a loan only because of the letter, but the letter can tip the decision when the required checks are split.

Accordingly, PSPM works in two stages. The Major stage uses the major evidences. It chooses the positive, negative or neutral pole and how strong that pole is, so it carries most of the score and can end in any of the five states. The Minor stage uses the minor evidences and only breaks exact ties of the Major stage. It can move the result to PT or PF, but never to T or F. Consequently, once the Major stage has a lead, the minor evidences do not change the result.

Some major evidences may still be unanswered. PSPM counts them as unknown evidence: unknowns do not lean to either side, they only reduce the strength of the result. When fewer than half of the major evidences are answered (the default threshold is $c = 0.5$), PSPM stays in U but still gives a small directional lean, so the result is not flat ignorance. For example, take eight required checks where six pass, one fails and one is still unanswered. With $\lambda = 5$, PSPM gives $ER = 0.9625$ and reads it as True, with the sub-state T_{neu}, which says that this is neither a weak nor a strong True.

The results of PSPM depend on the gain parameter λ: through the gain function $g(\lambda)$, it sets how large the normalised difference between true and false evidence must be before the directional component reaches its maximum contribution.

The model is defined in section 3. Its structural properties are proved in Appendix B, Appendix A lists the design principles, Appendix D explains why a lead starts from a raised base, and Appendix E gives an example of separating major from minor evidence.

The main contributions of this work are:

- A five-valued probabilistic scoring and readout model that treats a positive lean, a negative lean, and pure ignorance as different outcomes.
- A separation of required (major) evidence from optional (minor) evidence: major evidence chooses the pole and its strength, while minor evidence only breaks exact ties, so partial evidence alone can reach PT or PF but never T or F (Proposition 4).
- A human-readable distinction between ignorance (U), biased uncertainty (PT, PF), and strong evidence (T, F), with explicit rules for leaning when the evidence is incomplete.
- A range-based parameterisation $V_i = R_{i_{max}} - R_{i_{min}}$ whose final evidence ratio stays within $[0, 1]$ without clipping, whenever the state ranges are contiguous and symmetric around 0.5, as the default ranges are (Proposition 3).
- A treatment of unanswered evidence as a strength suppressor rather than a directional input: adding an unanswered major evidence never raises the strength of the output state (Proposition 7).
- A directional sub-state within U, so that cases with too few answered evidences still yield a lean (negative, neutral, positive) instead of flat ignorance, while the output state remains U.
- An optional graded confidence readout (three sub-bands per decided state) beside the five-state output, as a finer view of how strong a decision is inside its state (Appendix C).
- A negation: swapping true and false evidences maps every state to its mirror and the score to $1 - ER$ (Propositions 5 and 6).
- Proofs of the structural properties of the score (closed forms, state partition, closure, mirror symmetry, unknowns only suppress, exact state thresholds, and monotonicity in the restricted case), together with counterexamples showing where monotonicity fails in general (Appendix B).

2 Background & Related Work

2.1 Multi-Valued Logics

Multi-valued logics extend classical binary reasoning to represent partial or uncertain truth. Łukasiewicz introduced three-valued logic to codify indeterminism, where the third value violates both the law of excluded middle and the principle of contradiction [1]. Kleene later developed strong three-valued logic for reasoning about partial recursive predicates, treating the third value as "unknown" rather than "possible" [2]. Belnap's four-valued logic was designed for reasoning under incomplete or contradictory information, with a bilattice structure that supports both truth and information orderings [3]. These foundational systems, however, treat truth values as discrete qualitative categories rather than graded probabilistic measures. PSPM keeps their idea of a small set of named values, but it computes the value from counted evidences instead of assigning it by hand.

2.2 Probabilistic Logics

A parallel line of inquiry connects many-valued logic with probability. Rašković, Ognjanović and Marković establish completeness results for logics with approximate conditional probabilities, showing that probability models over logical formulae admit Henkin-style constructions satisfying axioms of well-definedness, extensionality, and finite additivity [4]. These results show that logics over probability models can be given a formal footing. PSPM is related to this line of work in motivation, but it does not define formulas, connectives or a proof system, so it does not inherit those completeness results. The only logical operation defined here is a negation (Propositions 5 and 6). Conjunction, disjunction and a full logic are left for future work.

Markov Logic Networks combine first-order logic with probabilistic graphical models, allowing weighted formulas to express uncertain knowledge [5]. While MLNs provide a rigorous probabilistic semantics, they operate in the classical two-valued domain and do not model directional uncertainty or partial belief states as first-class entities. They are also built for relational knowledge, whereas PSPM scores a short checklist of evidences for one question.

2.3 Evidence-Based and Subjective Approaches

Evidence-based approaches model uncertainty via belief mass functions. Dempster–Shafer (DS) theory assigns probability masses to sets of hypotheses, including ignorance as mass on the full frame [6]. With separate masses on {true}, {false} and the full frame, DS can already express a lean toward either side through its belief and plausibility values. What DS does not give is a small discrete readout of that lean, nor a rule that treats required and optional evidence differently.

Jøsang's Subjective Logic (SL) represents belief, disbelief, and uncertainty as a triple (b, d, u) with $b + d + u = 1$ [7]. This is the closest competitor to PSPM. An SL opinion with $b > d$ and a large u is already an uncertain opinion with a positive lean, so PSPM does not claim to represent something that SL cannot. SL also maps counts of positive and negative evidence to an opinion, and its projected probability $P = b + \alpha u$ (with base rate $\alpha = 0.5$) plays a role similar to the evidence ratio of PSPM. PSPM differs in three practical ways. First, it reads the result as one of five named states with fixed ranges, instead of leaving it as a continuous opinion. Second, it separates required (major) evidence from optional (minor) evidence, where minor evidence only breaks ties. Third, it counts unanswered required evidence explicitly in Q_n, so missing answers reduce the strength of the result directly. PSPM can therefore be seen as a discrete readout for checklist-style evidence, and not as a replacement for SL.

Other formalisms cover similar ground. Beta and Dirichlet posteriors turn counts of positive and negative evidence into a mean and a spread [8]. Imprecise probability describes belief with lower and upper probabilities, where the width of the interval shows ignorance [9]. Intuitionistic fuzzy sets give a membership degree, a non-membership degree and a hesitation degree [10], and neutrosophic logic treats truth, indeterminacy and falsity as independent components [11]. Three-way decision theory is the closest in spirit for the readout: it divides outcomes into accept, reject and defer regions using thresholds [12]. PSPM can be seen as a graded version of this idea, where the accept and reject regions are each split into a strong and a partial state. Table 1 summarises these differences. It is a rough qualitative comparison and not a measured one.

Table 1: Rough comparison of PSPM with related formalisms.

Framework	Directional lean	Ignorance	Discrete readout	Required vs. optional evidence
Dempster–Shafer	Yes (Bel/Pl)	Yes (full-frame mass)	No	No
Subjective Logic	Yes (b vs. d)	Yes (u)	No	No
Beta/Dirichlet	Yes (mean)	Yes (spread)	No	No
Imprecise probability	Yes (interval position)	Yes (interval width)	No	No
Intuitionistic fuzzy sets	Yes (μ vs. ν)	Yes (hesitation)	No	No
Neutrosophic logic	Yes	Yes (indeterminacy)	No	No
Three-way decisions	Accept/reject only	Yes (defer)	Yes (three levels)	No
PSPM	Yes (PT, PF)	Yes (U)	Yes (five levels)	Yes

2.4 Uncertainty in Machine Learning

Modern machine learning distinguishes *aleatoric uncertainty* (irreducible noise in the data-generating process) from *epistemic uncertainty* (reducible uncertainty due to limited model knowledge or data) [13]. Recent work has criticised this dichotomy as insufficiently expressive, arguing that a decision-theoretic perspective better captures the distinct quantities researchers care about [14]. Evidential deep learning goes a step further by letting a network output evidence counts, which are turned into a Dirichlet distribution and an SL-style opinion [15]. This is close to the evidence-counting view used here, and it is one possible place where a PSPM readout could be applied.

PSPM does not try to separate aleatoric from epistemic uncertainty. Its U state mostly reflects missing or balanced evidence, and its T and F states reflect a strong lead of the required evidence. The PT and PF states mean that a lean exists but is not strong enough to be True or False. At the moment, U covers both unanswered evidence and evenly split evidence, so it does not tell the two apart. This is a limitation of the current version. In return, PSPM stays simple and only needs the evidences to be counted explicitly.

2.5 Five-Valued Logics in the Literature

Several five-valued logics have been introduced. Ferreira proposed a fragment of @-logic for mobile agent systems, with truth values *unknown*, *possibly known but consistent*, *false*, *true*, and *inconsistent* [16]. Zehr introduced ST5, a five-valued framework for modelling linguistic phenomena such as vagueness and presupposition, with formal semantics and experimental backing [17]. Patrascu's FP5 and bifuzzy set representations introduced multiple truth values to model imprecision, inconsistency, and ambiguity [18]. The FiveASP paraconsistent logics of Osorio et al. explicitly handle contradictions and incomplete information [19].

Most of these frameworks are *symbolic* rather than *probabilistic*. They assign categorical truth values but do not provide a calculus for aggregating counted evidence into those values. (Patrascu's bifuzzy work attaches graded membership degrees, so it is closer to PSPM than the others.) We are not aware of one that separates required from optional evidence or reads a directional lean from evidence counts. The five states of PSPM are not merely additional truth values appended to a symbolic logic. They are *evidence-ratio regions* defined over a continuous score, with an explicit procedure that maps evidence counts to states. PSPM does not define a full logic, so this comparison concerns the truth values, not the proof systems.

In short, PSPM is positioned as a discrete five-level readout for short, checklist-style evidence, with an explicit split between required and optional evidence. It is not a new probability theory, and the comparison above is about what is practical for this setting, not about expressive power.

3 The Penta-State Probabilistic Model (PSPM)

3.1 Definition of Epistemic States

The Penta-State Probabilistic Model (PSPM) defines a finite set of epistemic states:

$$S = \{F,\ PF,\ U,\ PT,\ T\}$$

where each state S_i is associated with a dedicated range $R_i \subseteq [0, 1]$. The states are defined as follows:

- T: **True**
- F: **False**
- PT: **Partial True**
- PF: **Partial False**
- U: **Undecided**

3.2 Weight, Range and Evidence Ratio

Let $ER \in [0, 1]$ denote the total evidence ratio, and let ER_i denote the contribution of the evidence belonging to state i, where

$$ER_i \in \{ER_f,\ ER_{pf},\ ER_u,\ ER_{pt},\ ER_t\}.$$

3.2.1 State Range

For each state we define an evidence range:

$$R_i \in \{R_t, R_{pt}, R_u, R_{pf}, R_f\}, \tag{1}$$

with the default values

$$\begin{aligned} R_t &= [R_{t_{\min}}, R_{t_{\max}}] = [0.9, 1.0] \\ R_{pt} &= [R_{pt_{\min}}, R_{pt_{\max}}] = [0.7, 0.9] \\ R_u &= [R_{u_{\min}}, R_{u_{\max}}] = [0.3, 0.7] \\ R_{pf} &= [R_{pf_{\min}}, R_{pf_{\max}}] = [0.1, 0.3] \\ R_f &= [R_{f_{\min}}, R_{f_{\max}}] = [0.0, 0.1] \end{aligned}$$

These are the default ranges and can be customised for a given application, but the construction must satisfy

$$(R_{t_{\max}} - R_{t_{\min}}) = (R_{f_{\max}} - R_{f_{\min}}), \quad (R_{pt_{\max}} - R_{pt_{\min}}) = (R_{pf_{\max}} - R_{pf_{\min}}). \tag{2}$$

We also write $R_{u_{mid}} = (R_{u_{\min}} + R_{u_{\max}})/2 = 0.5$.

The ranges are divided into three poles. R_t and R_{pt} form the positive pole, R_f and R_{pf} form the negative pole, and R_u is the neutral pole. The neutral range is further divided into three sub-ranges (negative, neutral and positive), which tell us that even when the evidence is insufficient, the result can still carry a slight bias toward one pole. subsection 3.4 discusses how the Undecided state carries this bias.

Adjacent ranges share a boundary value (0.1, 0.3, 0.7 and 0.9). When ER falls exactly on a shared boundary, the state closer to U is chosen. For instance, $R_{t_{\min}} = R_{pt_{\max}} = 0.9$, so $ER = 0.9$ is read as PT and not T. On the negative side the rule is mirrored, so $ER = 0.1$ is read as PF and $ER = 0.3$ and $ER = 0.7$ are read as U. Figure 1 illustrates the ranges and this rule.

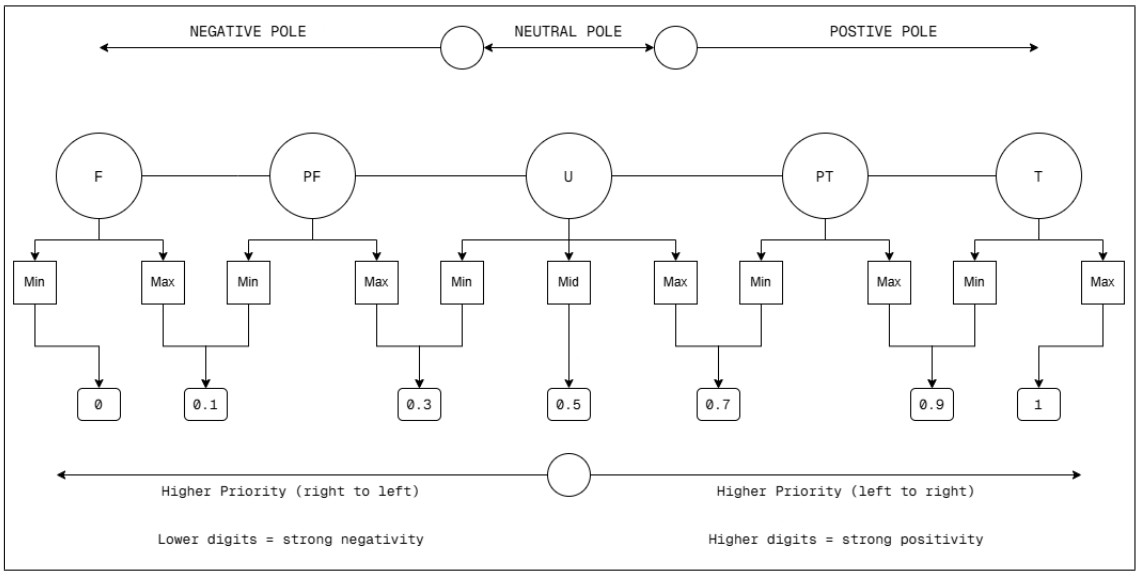


Figure 1: The PSPM range structure and directional bias. The upper axis shows the three semantic poles: negative (left), neutral (centre), and positive (right). The middle row shows the five epistemic states in order of increasing positivity: F, PF, U, PT, T. The lower axis gives the numerical Evidence Ratio (ER) at each boundary. Ranges are contiguous and non-overlapping; when ER falls on a shared boundary (0.1, 0.3, 0.7, 0.9), the state closer to U is chosen. The bottom arrows indicate that lower ER values correspond to stronger negative evidence and higher ER values to stronger positive evidence.

3.2.2 State Weight

The weight V of a state is the width of its range:

$$\begin{aligned} V_t &= R_{t_{max}} - R_{t_{min}} = 0.1, \qquad V_f = R_{f_{max}} - R_{f_{min}} = 0.1, \\ V_{pt} &= R_{pt_{max}} - R_{pt_{min}} = 0.2, \quad V_{pf} = R_{pf_{max}} - R_{pf_{min}} = 0.2, \\ V_u &= R_{u_{max}} - R_{u_{min}} = 0.4. \end{aligned} \tag{3}$$

3.2.3 Extended Weight and Range for Uncertainty

A natural question is what a human or an LLM can learn from an undecided state. We address this by dividing U into three sub-ranges with a directional bias. The bias indicates that, if more evidence for one pole were found, U would move toward that pole (see Appendix Appendix A). The neutral sub-range takes half of the weight of U, and the positive and negative sub-ranges take one quarter each. A wide neutral sub-range also makes sure that a lean is not declared because of noise or a very small difference from pure ignorance. Pure ignorance corresponds to $ER = 0.5$. This value is also the fallback when the coverage condition below is not satisfied and the evidence is exactly tied.

Therefore:

$$V_{u_{neg}} = \frac{V_u}{4}, \qquad V_{u_{neu}} = \frac{V_u}{2}, \qquad V_{u_{pos}} = \frac{V_u}{4},$$

and

$$R_{u_{neg}} = [R_{u_{min}}, R_{u_{min}} + V_{u_{neg}}] = [0.3, 0.4],$$
$$R_{u_{neu}} = [R_{u_{neg_{max}}}, R_{u_{neg_{max}}} + V_{u_{neu}}] = [0.4, 0.6],$$
$$R_{u_{pos}} = [R_{u_{neu_{max}}}, R_{u_{neu_{max}}} + V_{u_{pos}}] = [0.6, 0.7].$$

A value that lies on the boundary between two sub-ranges is assigned to the neutral sub-range.

3.3 Computation Procedure

PSPM distinguishes major evidence (Q_t, Q_f, Q_u) from minor evidence (Q_{pt}, Q_{pf}). Major evidence determines the primary direction and is counted in Q_n. Minor evidence is counted separately and is only used to break exact ties of the major evidence. Partial states do not carry their own uncertainty: the uncertainty of the system is fully captured by U and its directional sub-states. Appendix E gives an example of how to decide whether an evidence is major or minor.

3.3.1 Step 1: Evidence Quantification

Let:

$$Q_t = \text{number of major evidences supporting true},$$
$$Q_f = \text{number of major evidences supporting false},$$
$$Q_u = \text{number of major evidences left unanswered},$$
$$Q_{pt} = \text{number of minor evidences supporting true},$$
$$Q_{pf} = \text{number of minor evidences supporting false},$$

and

$$Q_n = Q_t + Q_f + Q_u, \qquad Q_{np} = Q_{pt} + Q_{pf}.$$

If $Q_n = 0$ (no major evidence at all), then $ER = R_{u_{mid}} = 0.5$ and the procedure stops.

3.3.2 Step 2: Coverage and Evidence Ratio

According to Appendix Appendix A, we assume that a person prefers to choose a side (positive or negative) only when enough of the evidence is known. We define the coverage

$$\text{cov} = \frac{Q_t + Q_f}{Q_n}, \tag{4}$$

and the minimum accepted coverage $c \in (0, 1]$. The default is $c = 0.5$, which requires that at least half of the major evidences are answered with true or false (Appendix Appendix A). It should be tuned for the domain. The case $\text{cov} \geq c$ is called the *main case* and the case $\text{cov} < c$ the *low-coverage case* (subsection 3.4).

In the main case, the contribution of each kind of evidence is

$$ER_t = \left(\frac{V_t}{Q_n}\right) Q_t, \qquad ER_f = \left(\frac{V_f}{Q_n}\right) Q_f, \tag{5}$$

$$ER_{pt} = \left(\frac{V_{pt}}{Q_{np}}\right) Q_{pt}, \qquad ER_{pf} = \left(\frac{V_{pf}}{Q_{np}}\right) Q_{pf}, \tag{6}$$

and if $Q_{np} = 0$, then $ER_{pt} = V_{pt}$ and $ER_{pf} = V_{pf}$. In this way, the absence of minor evidence treats the two minor sides equally. Appendix D explains why a lead in the major evidence is given the full minor weight V_{pt} as its starting base.

3.3.3 Step 3: State Bias Adjustment

In the main case, the pole is chosen by comparing ER_t and ER_f. When they are equal, the minor evidence decides. The tie-break terms (used only when $ER_t = ER_f$ and $Q_{np} > 0$) are

$$r_{(+)} = (ER_{pt} - ER_{pf}) + (R_{u_{max}} - R_{u_{mid}}) \frac{Q_{pt} - Q_{pf}}{Q_{np}}, \tag{7}$$

$$r_{(-)} = (ER_{pf} - ER_{pt}) + (R_{u_{mid}} - R_{u_{min}}) \frac{Q_{pf} - Q_{pt}}{Q_{np}}, \tag{8}$$

and the lead terms are

$$q_{(+)} = \{(R_{u_{max}} - R_{u_{mid}}) + V_{pt}\} \min\Big(1,\ g(\lambda)\,\frac{Q_t - Q_f}{Q_t + Q_f}\Big) + (ER_t - ER_f)\,, \tag{9}$$

$$q_{(-)} = \{(R_{u_{mid}} - R_{u_{min}}) + V_{pf}\} \min\Big(1,\ g(\lambda)\,\frac{Q_f - Q_t}{Q_t + Q_f}\Big) + (ER_f - ER_t)\,. \tag{10}$$

The gain function is $g(\lambda) = 10/\lambda$ with $\lambda > 0$. It controls how large the normalised difference $|Q_t - Q_f|/(Q_t + Q_f)$ must be for the directional component to reach its maximum contribution. With $\lambda = 1$ a difference of 10% of the answered evidence is enough, and with the default $\lambda = 5$ the difference must reach 50%. A larger λ therefore makes the model less sensitive to an imbalance between the two poles. The final evidence ratio is

$$\text{If } ER_t > ER_f, \quad ER = R_{u_{mid}} + q_{(+)}, \tag{11}$$

$$\text{If } ER_f > ER_t, \quad ER = R_{u_{mid}} - q_{(-)}, \tag{12}$$

$$\text{If } ER_t = ER_f \text{ and } ER_{pt} > ER_{pf}, \quad ER = R_{u_{mid}} + r_{(+)}, \tag{13}$$

$$\text{If } ER_t = ER_f \text{ and } ER_{pf} > ER_{pt}, \quad ER = R_{u_{mid}} - r_{(-)}, \tag{14}$$

$$\text{If } ER_t = ER_f \text{ and } ER_{pt} = ER_{pf}, \quad ER = R_{u_{mid}}. \tag{15}$$

Note that Q_{np} appears in a denominator only when $Q_{np} > 0$, since otherwise $ER_{pt} = ER_{pf}$ and the last rule applies. Under the default ranges, these rules reduce to a single closed form (Proposition 1): for $Q_t \neq Q_f$,

$$ER = 0.5 + 0.4\,\kappa + 0.1\,\frac{Q_t - Q_f}{Q_n}, \qquad \kappa = \max\Big(-1, \min\Big(1, g(\lambda)\frac{Q_t - Q_f}{Q_t + Q_f}\Big)\Big). \tag{16}$$

3.4 The Low-Coverage Case and Directional Uncertainty

When $\mathrm{cov} < c$, PSPM does not choose a pole, but it should still try to show a side. The result therefore stays in U and uses the three directional sub-states (negative, neutral, positive) defined above. If $Q_t > Q_f$:

$$ER = R_{u_{pos_{min}}} + \frac{V_{u_{pos}}}{Q_n}, \tag{17}$$

if $Q_f > Q_t$:

$$ER = R_{u_{neg_{max}}} - \frac{V_{u_{neg}}}{Q_n}, \tag{18}$$

and if $Q_f = Q_t$:

$$ER = R_{u_{mid}}. \tag{19}$$

By default these give $0.6 + 0.1/Q_n$, $0.4 - 0.1/Q_n$ and 0.5. In this case the margin $|Q_t - Q_f|$ is not used (see Limitation 4).

3.5 Interpretation

The state of a PSPM calculation is the state S_i whose range R_i contains the final ER, with shared boundaries resolved toward U as described in Equation 3.2.1.

3.6 Worked Examples

Table 2 shows the output of PSPM for some small checklists with the default parameters ($c = 0.5$, $\lambda = 5$). The first row is the example from the Introduction. Rows 4 and 5 show two different ways of reaching U (low coverage and an even split), and rows 6 and 7 show the Minor stage acting only on an exact tie of the major evidence.

4 Limitations and Future Work

Although PSPM gives a clear and checkable way to turn a short list of evidences into five states, several limitations remain. We list them here together with the work that would address them.

Table 2: Worked examples with the default parameters. Rows with a tie in the major evidence use the minor counts.

Q_t	Q_f	Q_u	Q_{pt}	Q_{pf}	ER	State	Comment
6	1	1	0	0	0.9625	T (T_{neu})	Clear lead, one unanswered
4	2	0	0	0	0.800	PT	Lead is not strong enough for T
3	2	0	0	0	0.680	U	Small lead stays undecided
1	0	2	0	0	0.633	U (U_{pos})	Coverage $1/3 < c$
5	5	0	0	0	0.500	U	Evenly split
1	1	0	2	0	0.900	PT	Tie broken by minor evidence
4	4	0	0	3	0.100	PF	Tie broken by minor evidence
0	3	1	0	0	0.025	F	Unanimous false, one unanswered

4.1 Limitations

1. **No empirical validation.**
   PSPM is motivated by how humans seem to decide, but this paper has no human study and no benchmark. "Human-aligned" is therefore a design goal and not a result. The state ranges, the weights V, the default $c = 0.5$, the default $\lambda = 5$, and the half-coverage principle in the Design Principles are chosen by hand and are not fitted or tested.

2. **Evidences are only counted.**
   Every major evidence has the same weight. There is no strength, reliability, or independence, so ten copies of one source count as ten evidences. Whether an evidence is major, minor, or unanswered is decided by the user, and this paper gives no protocol for it.

3. **Short checklists only, and no minimum evidence count.**
   PSPM is meant for short checklists of roughly 5–10 major evidences. The formulas use ratios, so the amount of evidence is not used, and nothing asks for a minimum number of evidences. A single True major evidence gives $ER = 1$ (T), and a 3:1 lead with no unanswered evidence already gives $ER = 0.95$ (T) under the default λ.

4. **The coverage threshold is a cliff.**
   With $c = 0.5$, one True and one unanswered evidence give $ER = 0.95$ (T), but one True and two unanswered give 0.633 (U). When coverage is below c, ER depends only on Q_n and not on the margin, so 3 True with 7 unanswered and 2 True with 1 False and 7 unanswered both give 0.61.

5. **Monotonicity holds only in a restricted case.**
   Proposition 9 covers inputs where both points meet the coverage threshold and there is no minor evidence. In general, adding a True evidence can lower ER and adding a False evidence can raise it, because of the low-coverage branch, the change of branch at c, and the tie branch (Remark 1).

6. **Ignorance and conflict look the same.**
   An evenly split case (5 True and 5 False) and a case with all evidences unanswered both give $ER = 0.5$ (U), so the state does not say which one happened. The directional sub-states of U give a lean only in the low-coverage case, and they do not change the output state.

7. **PSPM is not a full logic.**
   Only a negation is defined: swapping the true and false evidences maps every state to its mirror and ER to $1 - ER$ (Propositions 5 and 6). There is no conjunction or disjunction, so no logical completeness result is claimed.

8. **The properties depend on the parameters.**
   Closure and mirror symmetry hold when the ranges are contiguous and centred at 0.5, as the default ranges are (Propositions 3 and 5). Customised ranges must be checked against these conditions. Unanimity always gives T (Proposition 8(a)) only when $g(\lambda) \geq 1$, that is $\lambda \leq 10$. Resolving boundaries toward U, so that T needs $ER > 0.9$, is a modelling choice.

9. **Comparison and integration are not tested.**
   The comparison with Subjective Logic, Beta posteriors, and three-way decisions is written, not measured on data. PSPM is also not differentiable, since it uses branches and thresholds, so it cannot be trained end to end inside a neural network. Its computing cost is not an issue, because it is constant once the evidences are counted.

4.2 Future Work

1. **Validation.**
   Run a small human study to check whether people agree with the five states, and test PSPM on a concrete task such as checklist-based verification of AI answers. Compare it with Subjective Logic and Beta baselines on the same checklists.
2. **Evidence protocol and weights.**
   Write a protocol for deciding major, minor, and unanswered evidences. Add strength and reliability weights, and handle evidences that depend on each other.
3. **Smoother coverage handling.**
   Replace the hard threshold c with a smooth ramp so that ER changes gradually, and aim for monotonicity in general. Add a minimum evidence count, so that a single evidence cannot give True.
4. **Separate ignorance from conflict.**
   Add a tag for the U state that tells whether the result comes from unanswered evidences or from evenly split evidences.
5. **Connectives.**
   Define conjunction and disjunction over evidence lists and test which laws they satisfy. Only then should PSPM be called a logic.
6. **Parameter study.**
   Study how sensitive the states are to c, λ, the ranges, and the weights, and fit them from data where labelled examples exist.
7. **Hybrid and differentiable versions.**
   Build a smooth variant of PSPM that can be used as a readout layer in learning systems. We also plan to extend the state logic with further sub-states to improve the confidence readout, and to introduce a universal parameter in a later version (PSPM2).

5 Conclusion

The Penta-State Probabilistic Model (PSPM) introduces a five-valued framework for representing directional uncertainty in probabilistic reasoning, in a form that is meant to be readable by humans. Compared to binary or fuzzy readouts, PSPM separates strong evidence, partial evidence and ignorance, and it separates required from optional evidence. We proved that the score is well defined, stays in $[0, 1]$, is mirror symmetric, and is only lowered in strength by unanswered evidence, and we showed by counterexample where monotonicity fails. The human-alignment claim, however, remains a design goal. Real-world deployment requires validation on human judgements and on concrete tasks, and careful tuning of the ranges, c and λ. We hope that PSPM can serve as a simple and interpretable readout for checklist-style evidence.

Acknowledgement

The author expresses gratitude to the process of independent research and exploration that enabled the development of the Penta-State Probabilistic Model (PSPM). The author also confirms that no external funding or support was received for this work.

References

[1] J. Łukasiewicz, *Selected works*. North-Holland Pub. Co, 1970.

[2] S. C. Kleene, *Introduction to Metamathematics*. New York: Van Nostrand, 1952.

[3] N. D. Belnap, "A useful four-valued logic," *Springer eBooks*, pp. 5–37, 01 1977.

[4] M. Rašković, Z. Marković, and Z. Ognjanović, "A logic with approximate conditional probabilities that can model default reasoning," *International Journal of Approximate Reasoning*, vol. 49, no. 1, pp. 52–66, 2008.

[5] M. Richardson and P. Domingos, "Markov logic networks," *Machine Learning*, vol. 62, pp. 107–136, 01 2006.

[6] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *The Annals of Mathematical Statistics*, vol. 38, pp. 325–339, 1967. [Online]. Available: https://www.jstor.org/stable/2239146

[7] A. Jøsang, *Subjective Logic*. Springer International Publishing, 01 2016.

[8] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. Boca Raton, FL: Chapman and Hall/CRC, 2013.

[9] P. Walley, *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall, 1991.

[10] K. T. Atanassov, "Intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 20, no. 1, pp. 87–96, 1986.

[11] F. Smarandache, *Neutrosophy: Neutrosophic Probability, Set, and Logic*. Rehoboth: American Research Press, 1998.

[12] Y. Yao, "Three-way decisions with probabilistic rough sets," *Information Sciences*, vol. 180, no. 3, pp. 341–353, 2010.

[13] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods," *Machine Learning*, vol. 110, no. 3, pp. 457–506, 3 2021, accessed 2026-10-06. [Online]. Available: https://arxiv.org/pdf/1910.09457.pdf

[14] F. Bickford Smith, J. Kossen, E. Trollope, M. Van Der Wilk, and A. Foster, "Rethinking aleatoric and epistemic uncertainty," in *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, 2025, also available at ORA: https://ora.ox.ac.uk/objects/uuid:a56858e4-4ead-4ece-a909-36fdb80461a3.

[15] M. Sensoy, L. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," in *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*, 2018.

[16] U. Ferreira, "View of a five-valued logic and a system," JCS&T, 10 2004. [Online]. Available: https://journal.info.unlp.edu.ar/JCST/article/view/891/423

[17] J. Zehr, "St5: A 5-valued logic for truth-value judgments involving vagueness and presuppositions," *Lecture notes in computer science*, pp. 247–265, 01 2014.

[18] V. Patrascu, "A new penta-valued logic based knowledge representation," arXiv.org, 2015. [Online]. Available: https://arxiv.org/abs/1502.05562

[19] M. Osorio and C. Zepeda, "Three new genuine five-valued logics," 2019. [Online]. Available: https://ceur-ws.org/Vol-2585/paper8.pdf

Appendix A Design Principles

Design Principle 1 (Human Evidence Principle). Human reasoning relies on evidence when estimating a probability. PSPM therefore does not produce a lean out of nothing: in the absence of evidence, it returns U (Undecided). We assume, as a modelling choice and not as an empirical finding, that a person tends to choose a side (a positive or negative pole) only when at least half of the evidences are known. This assumption is the origin of the default coverage threshold $c = 0.5$.

Design Principle 2 (Human Probability Categorisation). We assume that human probability judgements can be categorised within the same five epistemic states as PSPM:

$$S = \{F,\ PF,\ U,\ PT,\ T\}. \tag{20}$$

This is a design assumption. The set of states may be expanded or contracted into more or fewer states as needed.

Appendix B Structural Properties and Proofs

This appendix proves the properties that the main text relies on. Nothing here changes the model. It only checks what the definitions in section 3 already imply. Each result states the conditions it needs, because some properties do not hold in general (see Remark 1 and the discussion after Proposition 7).

Appendix B.1 Setting and Notation

Write $a = Q_t$, $b = Q_f$, $u = Q_u$ and $n = Q_n = a + b + u$. Let $h = R_{u_{\max}} - R_{u_{mid}} = R_{u_{mid}} - R_{u_{\min}} = V_u/2$, and define

$$A = h + V_{pt}, \qquad B = V_t.$$

Let $\mathrm{cov} = (a + b)/n$, and assume $c > 0$ and $g(\lambda) = 10/\lambda$ as in the main text.

Two conditions on the ranges are used below.

- **(C1) Contiguity.** The five ranges are consecutive intervals of $[0, 1]$: $R_{f_{\min}} = 0$, $R_{f_{\max}} = R_{pf_{\min}}$, $R_{pf_{\max}} = R_{u_{\min}}$, $R_{u_{\max}} = R_{pt_{\min}}$, $R_{pt_{\max}} = R_{t_{\min}}$, and $R_{t_{\max}} = 1$.
- **(C2) Symmetry.** Equation (2) holds (so $V_t = V_f$ and $V_{pt} = V_{pf}$), $R_{u_{\min}} + R_{u_{\max}} = 1$ (so $R_{u_{mid}} = \frac{1}{2}$), and $V_{u_{pos}} = V_{u_{neg}} = V_u/4$.

The default ranges satisfy both, with $h = 0.2$, $A = 0.4$ and $B = 0.1$. Under (C1) and (C2),

$$A + B = (R_{u_{\max}} - \tfrac{1}{2}) + (V_{pt} + V_t) = (R_{u_{\max}} - \tfrac{1}{2}) + (R_{t_{\max}} - R_{pt_{\min}}) = \tfrac{1}{2}, \tag{21}$$

since $V_{pt} + V_t = R_{t_{\max}} - R_{pt_{\min}}$ when $R_{pt_{\max}} = R_{t_{\min}}$, and $R_{pt_{\min}} = R_{u_{\max}}$.

Boundary convention. When ER lies on a shared boundary, the state closer to U is chosen (Equation 3.2.1). It is written as the map $\sigma : [0, 1] \to S$:

$$\sigma(x) = \begin{cases} T & 0.9 < x \le 1 \\ PT & 0.7 < x \le 0.9 \\ U & 0.3 \le x \le 0.7 \\ PF & 0.1 \le x < 0.3 \\ F & 0 \le x < 0.1. \end{cases}$$

Let μ be the mirror map $T \leftrightarrow F$, $PT \leftrightarrow PF$, $U \leftrightarrow U$.

Branches. The procedure of section 3 has three cases: the empty case $n = 0$, the *low-coverage* case $\mathrm{cov} < c$, and the *main* case $\mathrm{cov} \ge c$. Let $\mathrm{clip}(x) = \max(-1, \min(1, x))$.

Proposition 1 (Closed forms)**.** *Under (C2):*

(a) If $n = 0$, then $ER = \frac{1}{2}$.

(b) If $\mathrm{cov} < c$: $ER = R_{u_{pos_{\min}}} + V_{u_{pos}}/n$ when $a > b$; $ER = R_{u_{neg_{\max}}} - V_{u_{neg}}/n$ when $a < b$; $ER = \frac{1}{2}$ when $a = b$. By default these are $0.6 + 0.1/n$, $0.4 - 0.1/n$ and 0.5.

(c) If $\mathrm{cov} \geq c$ *and* $a \neq b$*:*

$$ER = \tfrac{1}{2} + A\,\kappa + B\,\frac{a-b}{n}, \qquad \kappa = \mathrm{clip}\Big(g(\lambda)\,\frac{a-b}{a+b}\Big). \tag{22}$$

(d) If $\mathrm{cov} \geq c$ *and* $a = b$*:* $ER = \frac{1}{2} + A\,\theta$*, where* $\theta = (Q_{pt} - Q_{pf})/Q_{np}$ *if* $Q_{np} > 0$ *and* $\theta = 0$ *if* $Q_{np} = 0$*.*

Proof. (a) and (b) are the definitions. For (b), (C2) gives $R_{u_{neg_{\max}}} + R_{u_{pos_{\min}}} = R_{u_{\min}} + R_{u_{\max}} = 1$, so the two lead cases are mirror images.

(c) Since $V_t = V_f$, $ER_t > ER_f$ holds exactly when $a > b$, and $ER_t - ER_f = B(a-b)/n$. For $a > b$, the definition gives $ER = \frac{1}{2} + q_{(+)}$ with $(R_{u_{\max}} - R_{u_{mid}}) + V_{pt} = A$, which is (22). For $a < b$, $ER = \frac{1}{2} - q_{(-)} = \frac{1}{2} - A\min\big(1, g\frac{b-a}{a+b}\big) - B\frac{b-a}{n}$, and $\min(1,-x) = -\max(-1,x)$ turns this into (22). The denominator $a+b$ is positive here, because $a \neq b$ means at least one of them is positive.

(d) Here $ER_t = ER_f$. If $Q_{np} = 0$, then $ER_{pt} = V_{pt} = V_{pf} = ER_{pf}$, so the last tie rule gives $\frac{1}{2}$. If $Q_{np} > 0$, then $ER_{pt} - ER_{pf} = V_{pt}\theta$ because $V_{pt} = V_{pf}$. For $\theta > 0$, $r_{(+)} = V_{pt}\theta + h\theta = A\theta$ and $ER = \frac{1}{2} + A\theta$. For $\theta < 0$, $r_{(-)} = -A\theta$ and $ER = \frac{1}{2} - r_{(-)} = \frac{1}{2} + A\theta$. For $\theta = 0$ the last tie rule gives $\frac{1}{2}$. □

Appendix B.2 State Partition and Closure

Proposition 2 (State partition)**.** *The map* σ *assigns exactly one state to every* $x \in [0,1]$*. Hence the output state of PSPM is well defined for any value of* ER*.*

Proof. The five intervals in the definition of σ are pairwise disjoint and their union is $[0,1]$. Each shared boundary $(0.1, 0.3, 0.7, 0.9)$ appears in exactly one of its two neighbours. □

Proposition 3 (Closure)**.** *Under (C1) and (C2),* $ER \in [0,1]$ *for every admissible input, with no clipping of the final score. Moreover, in the low-coverage case* $ER \in [R_{u_{\min}}, R_{u_{\max}}]$*, so the output is always* U*.*

Proof. We go through the cases of Proposition 1. In (a), $ER = \frac{1}{2}$. In (b), ER stays between $R_{u_{\min}}$ and $R_{u_{\max}}$: for $a > b$ it lies in $(R_{u_{pos_{\min}}}, R_{u_{pos_{\max}}}]$ because $n \geq 1$, and for $a < b$ it lies in the mirrored interval. In (c), $|\kappa| \leq 1$ and $|a-b|/n \leq 1$, so $|ER - \frac{1}{2}| \leq A + B = \frac{1}{2}$ by (21). In (d), $|\theta| \leq 1$ and $A \leq A + B = \frac{1}{2}$, so $|ER - \frac{1}{2}| \leq \frac{1}{2}$. □

Appendix B.3 Minor Evidence Only Breaks Ties

Proposition 4 (Minor evidence as a tie-breaker)**.** *Under (C1) and (C2):*

(i) If $Q_t \neq Q_f$*, or if* $\mathrm{cov} < c$*, then* ER *does not depend on* Q_{pt} *or* Q_{pf}*.*

(ii) If $Q_t = Q_f$ *and* $\mathrm{cov} \geq c$*, then* $ER \in [\frac{1}{2} - A, \frac{1}{2} + A] = [R_{pf_{\min}}, R_{pt_{\max}}]$*, so* $\sigma(ER) \in \{PF, U, PT\}$*. Partial evidence alone can therefore never produce* T *or* F*.*

(iii) With the default ranges, in case (ii), $\sigma(ER) = PT$ *if and only if* $Q_{pt} > 3Q_{pf}$*, and* $\sigma(ER) = PF$ *if and only if* $Q_{pf} > 3Q_{pt}$*. Otherwise it is* U*.*

Proof. (i) The expressions in Proposition 1(a)–(c) contain no minor counts.

(ii) By Proposition 1(d), $ER = \frac{1}{2} + A\theta$ with $|\theta| \leq 1$. By (21), $\frac{1}{2} + A = R_{u_{mid}} + h + V_{pt} = R_{u_{\max}} + V_{pt} = R_{pt_{\max}} = R_{t_{\min}}$, and by symmetry $\frac{1}{2} - A = R_{pf_{\min}}$. The value $R_{t_{\min}}$ itself is assigned to PT by the boundary convention, so T is never reached. The same argument shows that F is never reached.

(iii) By default, $ER = 0.5 + 0.4\theta$. PT requires $0.4\theta > 0.2$, that is $\theta > \frac{1}{2}$. With $p = Q_{pt}$ and $q = Q_{pf}$, $\theta > \frac{1}{2} \iff 2(p-q) > p + q \iff p > 3q$. The case for PF is symmetric. □

Appendix B.4 Mirror Symmetry and Negation

Proposition 5 (Mirror symmetry)**.** *Under (C2), let* E' *be obtained from an input* E *by swapping* $Q_t \leftrightarrow Q_f$ *and* $Q_{pt} \leftrightarrow Q_{pf}$ *(and leaving* Q_u *unchanged). Then*

$$ER(E') = 1 - ER(E), \qquad \sigma\big(ER(E')\big) = \mu\big(\sigma(ER(E))\big).$$

In the low-coverage case, U_{pos} and U_{neg} are exchanged and U_{neu} is fixed.

Proof. The swap leaves n and cov unchanged, so E and E' fall in the same branch. We check each case of Proposition 1. In (a) and in the tie cases, $ER = \frac{1}{2} = 1-\frac{1}{2}$. In (b), $R_{u_{neg_{\max}}} - V_{u_{neg}}/n = 1-(R_{u_{pos_{\min}}} + V_{u_{pos}}/n)$ because of (C2). In (c), $a-b$ changes sign, and κ and $(a-b)/n$ are odd in $a-b$ (the function clip is odd), so $ER(E') - \frac{1}{2} = -(ER(E) - \frac{1}{2})$. In (d), $\theta \mapsto -\theta$ gives the same conclusion. For the state, $\sigma(1-x) = \mu(\sigma(x))$ holds because the intervals defining σ are symmetric around $\frac{1}{2}$ and the boundary convention is symmetric (for example, $0.7 \mapsto U$ and $0.3 \mapsto U$, $0.9 \mapsto PT$ and $0.1 \mapsto PF$). The sub-state claim follows from $[0.6, 0.7] \mapsto [0.3, 0.4]$. □

Proposition 6 (Negation)**.** *Define the negation $\neg E$ of an evidence list as the swap in Proposition 5. Under (C2):*

(i) $\neg\neg E = E$.

(ii) $ER(\neg E) = 1 - ER(E)$ *and* $\sigma(ER(\neg E)) = \mu(\sigma(ER(E)))$.

(iii) If $\neg E = E$ (that is, $Q_t = Q_f$ and $Q_{pt} = Q_{pf}$), then $ER(E) = \frac{1}{2}$ and the state is U. At the state level, U is the only fixed point of μ.

Proof. (i) Swapping twice returns the original counts. (ii) is Proposition 5. (iii) By (ii), $ER = 1 - ER$, so $ER = \frac{1}{2}$ and $\sigma(\frac{1}{2}) = U$. □

The converse of (iii) is false. A state of U does not mean that $E = \neg E$, and this is the same "ignorance and conflict look the same" limitation listed in section 4.

Appendix B.5 Unanswered Evidence Only Lowers Strength

Proposition 7 (Unknowns only suppress)**.** *Fix Q_t, Q_f, Q_{pt}, Q_{pf} and raise Q_u by one. Let ER and ER' be the scores before and after (if $n = 0$, both are $\frac{1}{2}$). Under (C1) and (C2):*

(i) If both inputs are in the main case, then $|ER' - \frac{1}{2}| \le |ER - \frac{1}{2}|$, with strict inequality when $a \neq b$.

(ii) If both inputs are in the low-coverage case, then $|ER' - \frac{1}{2}| \le |ER - \frac{1}{2}|$, with strict inequality when $a \neq b$.

(iii) If the extra unknown moves the input from the main case to the low-coverage case, then $\sigma(ER') = U$.

No other transition is possible, because cov *can only decrease.*

Proof. (i) If $a \neq b$, the value κ in (22) depends only on a and b, and $|a-b|/n$ strictly decreases when n grows. If $a = b$, then $ER = \frac{1}{2} + A\theta$ does not involve u, so it is unchanged.

(ii) For $a > b$, $ER - \frac{1}{2} = (R_{u_{pos_{\min}}} - \frac{1}{2}) + V_{u_{pos}}/n$ is positive and strictly decreasing in n, so it moves toward $\frac{1}{2}$. The case $a < b$ is symmetric, and for $a = b$ both scores equal $\frac{1}{2}$.

(iii) This is Proposition 3. □

What this does and does not say. With the default ranges, let the *strength level* be $\ell = 0$ for U, 1 for PT or PF, and 2 for T or F. Then ℓ is non-decreasing in $|ER - \frac{1}{2}|$ (U is exactly $|ER - \frac{1}{2}| \le 0.2$ and T/F is exactly $|ER - \frac{1}{2}| > 0.4$). Proposition 7 therefore gives that adding an unanswered evidence never raises the strength level, and an input that is already in U stays in U. The *score* itself, however, can jump at the coverage threshold, so the statement "the score never moves away from the middle" needs the condition that coverage stays on one side of c. For example, with $c = 0.5$ and the default ranges, $(Q_t, Q_f, Q_u) = (5, 4, 9)$ has $\text{cov} = 0.5$ and $ER \approx 0.594$, while $(5, 4, 10)$ falls below c and has $ER = 0.6 + 0.1/19 \approx 0.605$. The score moves away from the middle here, but the state remains U in both cases, so the claim about states is unaffected.

Appendix B.6 Exact State Thresholds

Proposition 8 (State thresholds, default parameters)**.** *Take the default ranges, $c > 0$ and $\lambda = 5$ (so $g = 2$). Write $\delta = (a-b)/n$.*

(a) ***Unanimity.*** *If* $\text{cov} \ge c$*, $b = 0$ and $a \ge 1$, then $ER = 0.9 + 0.1\, a/n \in (0.9, 1]$, so the state is T, and $ER = 1$ exactly when $u = 0$. Symmetrically, $a = 0$ and $b \ge 1$ gives F. The same holds for any $\lambda \le 10$.*

(b) ***Main case with a lead.*** *If* $\mathrm{cov} \geq c$ *and* $a > b$*, then* $ER = \frac{1}{2} + (4\kappa + \delta)/10$*, and*

$$\sigma(ER) = \begin{cases} T & 4\kappa + \delta > 4 \\ PT & 2 < 4\kappa + \delta \leq 4 \\ U & 4\kappa + \delta \leq 2. \end{cases}$$

The case $a < b$ *is the mirror image.*

(c) ***Low coverage.*** *If* $\mathrm{cov} < c$*, then* $ER \in [0.3, 0.4) \cup \{0.5\} \cup (0.6, 0.7]$ *and the state is* U*.*

(d) ***Tie with minor evidence.*** *See Proposition 4(iii).*

Proof. (a) With $b = 0$ we have $\kappa = \min(1, g) = 1$ and $\delta = a/n$, so (22) gives $0.5 + 0.4 + 0.1\, a/n$. (b) From (22), $ER - \frac{1}{2} = 0.4\kappa + 0.1\delta = (4\kappa + \delta)/10$. The thresholds 0.7 and 0.9 translate to $4\kappa + \delta = 2$ and $4\kappa + \delta = 4$, and the boundary convention puts equality in the lower state. (c) This is Proposition 1(b) with $n \geq 1$. □

As a check, the example in the Introduction (six passes, one failure, one unanswered, with $c = 0.5$) has $\mathrm{cov} = 7/8$, $\kappa = \min(1, 2 \cdot 5/7) = 1$ and $\delta = 5/8$. Hence $ER = 0.5 + 4.625/10 = 0.9625$ and the state is T.

Appendix B.7 Monotonicity in the Restricted Case

Proposition 9 (Restricted monotonicity)**.** *Assume (C1), (C2),* $c > 0$*,* $Q_{np} = 0$*, and that the input is in the main case (*$\mathrm{cov} \geq c$*). Then adding one True major evidence gives* $ER' \geq ER$*, and adding one False major evidence gives* $ER' \leq ER$*. In both cases the new input is also in the main case.*

Proof. Since $Q_{np} = 0$, Proposition 1(c) and (d) combine into $ER = \frac{1}{2} + A\ \mathrm{clip}\big(g\frac{a-b}{a+b}\big) + B\frac{a-b}{n}$ for all main-case inputs (when $a = b$ both terms are 0). Also $a + b > 0$ because $\mathrm{cov} \geq c > 0$.

Coverage. Adding a True or False evidence changes $(a + b)/n$ to $(a + b + 1)/(n + 1) \geq (a + b)/n$, because $a + b \leq n$. So cov does not decrease and the new input stays in the main case.

Adding True ($a \to a + 1$, $n \to n + 1$). First, $\frac{a+1-b}{a+b+1} - \frac{a-b}{a+b} = \frac{2b}{(a+b)(a+b+1)} \geq 0$, and clip is non-decreasing, so the first term does not decrease. Second, $\frac{a+1-b}{n+1} - \frac{a-b}{n} = \frac{n-(a-b)}{n(n+1)} = \frac{2b+u}{n(n+1)} \geq 0$. Since $A, B \geq 0$, $ER' \geq ER$.

Adding False ($b \to b + 1$, $n \to n + 1$). By the same computation, $\frac{a-b-1}{a+b+1} - \frac{a-b}{a+b} = -\frac{2a}{(a+b)(a+b+1)} \leq 0$, and $\frac{a-b-1}{n+1} - \frac{a-b}{n} = -\frac{n+(a-b)}{n(n+1)} = -\frac{2a+u}{n(n+1)} \leq 0$. So $ER' \leq ER$. □

Remark 1. Each of the three conditions of Proposition 9 is needed. Below, all examples use the default ranges, $c = 0.5$ and $\lambda = 5$.

- *Low coverage.* $(Q_t, Q_f, Q_u) = (2, 1, 7)$ gives $ER = 0.61$, and adding a True evidence gives $(3, 1, 7)$ with $ER = 0.6 + 0.1/11 \approx 0.609$, which is lower. Likewise $(1, 2, 7)$ gives 0.39 and adding a False evidence gives $(1, 3, 7)$ with ≈ 0.391, which is higher.
- *Crossing c.* $(10, 9, 20)$ has $\mathrm{cov} = 19/39 < c$ and $ER \approx 0.6026$. Adding a True evidence gives $(11, 9, 20)$ with $\mathrm{cov} = 0.5$, and $ER = 0.5 + 0.4 \cdot 0.2 + 0.1 \cdot 2/40 = 0.585$, which is lower.
- *Minor evidence (tie branch).* $(Q_t, Q_f, Q_u) = (1, 1, 0)$ with $Q_{pt} = 2$ and $Q_{pf} = 0$ gives $ER = 0.9$ (PT). Adding a True evidence gives $(2, 1, 0)$, where the major lead now decides and the minor evidence is ignored, so $ER = 0.5 + 0.4 \cdot \frac{2}{3} + 0.1 \cdot \frac{1}{3} = 0.8$, which is lower.

These are the cases mentioned in Limitation 5 of section 4. They all come from the hard threshold c and from the rule that minor evidence only matters on an exact tie, which are the two design choices that Future Work item 3 proposes to smooth.

Appendix C State Confidence

The confidence sub-ranges apply only to decided states (T, PT, PF, F). The Undecided state instead uses the directional sub-states (U_{pos}, U_{neg}, U_{neu}) to express which pole the residual uncertainty leans toward. These are separate mechanisms operating on different ranges.

For a decided state S_i, the range R_i is split into a lower, a middle and an upper sub-band, denoted by the subscripts neg, neu and pos (this is the position inside the range, not a direction of belief). To keep the confidence labels stable, the middle band gets half of the range width, and the remaining half is split equally between the lower and upper bands. With

$$\gamma = \frac{V_i}{4}, \tag{23}$$

we have, for any S_i except U:

$$R_{i_{neg}} = [R_{i_{min}}, R_{i_{min}} + \gamma], \tag{24}$$

$$R_{i_{neu}} = [R_{i_{min}} + \gamma, R_{i_{max}} - \gamma], \tag{25}$$

$$R_{i_{pos}} = [R_{i_{max}} - \gamma, R_{i_{max}}]. \tag{26}$$

For example, $T = [0.9, 1.0]$ has $\gamma = 0.025$, so $T_{neg} = [0.9, 0.925]$, $T_{neu} = [0.925, 0.975]$ and $T_{pos} = [0.975, 1.0]$. The score 0.9625 in the Introduction therefore lies in T_{neu}. Because R_i is ordered by ER, the strongest band lies at the upper end for T and PT and at the lower end for F and PF. As with the sub-states of U, a value on the boundary between two bands is assigned to the middle band, and on the boundary between two states it is assigned following Equation 3.2.1. The confidence readout is computed last, once a calculation has reached a state.

Appendix D Why a Lead Starts from a Raised Base

A common question is whether the minor weight V_{pt} should be included in the score when no minor evidence is present, and whether it distorts the result. In the lead case the starting base is $R_{u_{mid}} + (R_{u_{max}} - R_{u_{mid}}) + V_{pt}$, that is, $A = h + V_{pt}$ in (22), so V_{pt} is always included. The reason is the following. Without it, a lead could only lift the score to the top of U plus the small contribution of V_t, and for the default ranges the maximum would be $0.5 + 0.2 + 0.1 = 0.8$. Even when every major evidence supports True and no minor evidence exists, PSPM would then end in PT, although strong evidence should give T. Because $A + B = \frac{1}{2}$ under (C1) and (C2) ((21)), including V_{pt} makes the maximum score exactly 1, which falls under T. In this way the minor weight acts as the base that a lead has to climb through PT, and it is consistent with Step 2, where the absence of minor evidence gives both minor sides their full weight so that they cancel. When minor evidence is present, it is only used on an exact tie of the major evidence (Proposition 4).

Appendix E Major Evidence Versus Minor Evidence

This appendix gives an illustrative (hypothetical) example of how to classify evidence. Assume that the question is "Is this text AI-written?" and that two evidences are available: "Does the text keep nearly the same entropy rate throughout?" and "Does the text contain em dashes?". Suppose an analyst judges the first to be a strong signal, because a human writer would rarely keep such a pattern. The analyst would then treat it as a major evidence. The second is a weak signal: em dashes are common in academic and news writing, so a human author can easily produce them, and the same signal may vary with writing style and era. A text with em dashes should therefore not immediately change the overall decision to T or F. PSPM models this through partial belief: the evidence is minor and can only move the result between PT and PF when the major evidence is tied. The values in this example are assumptions for illustration only. They are not empirical claims about AI-written text, and deciding which evidences are major is the responsibility of the user (see Limitation 2).