Numerical Assessment

A Technical Justification of the Mathematics Behind i.a.T.'s
Truth Seeker, Scorekeeper, and Gatekeeper Modes

This document is intended for readers with an undergraduate background in probability theory. Its purpose is to justify — rather than merely describe — the mathematical operations the application performs when computing numerical assessments of argument trees.

Contents

1 Purpose and Scope

The i.a.T. application computes a numerical assessment for each node of a hierarchical argument tree. This assessment is intended to represent, in a mathematically principled way, the extent to which the argument at that node is supported by the evidence beneath it. Three assessment models are offered: *Truth Seeker*, which implements a hierarchical Bayesian update; *Scorekeeper*, a weighted averaging model (WAM) drawn from multi-attribute decision theory; and *Gatekeeper*, a WAM variant with hard veto conditions.

This document justifies the mathematics of all three models. It is written for a reader who wishes to verify that the application's claims to Bayesian rigour are defensible, and who is willing to work through equations and derivations rather than intuitive analogies. The intended audience is a technically qualified reviewer — for example, a statistically-literate user, an academic collaborator, or an internal contributor auditing the code.

1.1 What this document justifies

The mathematical operations described here are those actually executed by the application's assessment engine.[1] Every formula in this document corresponds to code that runs on every recalculation of a user's tree.

The Truth Seeker model receives the majority of the discussion (Sections 3–5) because it is the mathematically deepest of the three: it invokes Bayes' theorem directly, and its legitimacy as a Bayesian procedure requires the most careful defence. The Scorekeeper and Gatekeeper models are addressed more briefly, in Sections 6 and 7 respectively.

1.2 What this document defers

Two related pieces of the numerical assessment system are addressed in companion documents rather than here:

- **The Dependency system.** The application allows users to declare that arguments are correlated (share evidence) or causally related. When enabled, these declarations apply corrections to the baseline Truth Seeker output. The mathematics of these corrections — geometric-mean group aggregation, information-theoretic overlap discount, and causal dampening — is documented separately in the *Dependency Mathematics* reference.
- **User-interface tooltips and copy.** A short reference of the tooltip strings adopted by the application appears in Section 9 of the second part of this document (delivered separately). Their normative justification derives from the definitions established in Section 2 of this document.

1.3 Structure

Section 2 defines the terms and notation used throughout. Section 3 presents the formal Bayesian framework the Truth Seeker model claims to instantiate. Section 4 then justifies, step by step, how the application's parameterised procedure realises that framework, naming each modelling choice and the standard technique it invokes. Section 5 illustrates the resulting behaviour through worked numerical examples.

The second part of this document (Sections 6 onward) covers the Scorekeeper and Gatekeeper models, an honest audit of modelling choices and their limitations, tooltip and UI reference material, and a bibliography.

2 Terminology and Conventions

Throughout this document, we adopt the terminology used in the application's user interface, and formalise its mathematical meaning where relevant.

2.1 Tree structure

A *tree* is a rooted directed graph in which the root node is designated the *Central Argument* (CA). Each non-root node is a *sub-argument*, and immediate children of the CA are called *Main Arguments* (MA). Every sub-argument carries a *type* $\tau \in \{\text{support}, \text{oppose}\}$ indicating whether it advances or contradicts its parent. A *leaf* is a node with no children.

We denote the CA by CA, a generic node by n, and the set of children of n by $\text{ch}(n)$. Each node n carries the numerical fields defined below, together with a computed result $\pi_n \in [0, 1]$ (Truth Seeker) or $s_n \in [-1, 1]$ (Scorekeeper/Gatekeeper), stored on the node as `computedIndex`.

[1] Specifically, the pure functions in `iat_core/services/assessment_engine.dart`.

2.2 User-set parameters

Three user-adjustable parameters govern the assessment. Their intended semantics and range are as follows.

Importance ($I_n \in [0, 100]$). A parameter attached to every non-CA node. Importance encodes *how strong the argument would be as evidence for its parent claim, were the argument fully established*. It is calibrated so that $I = 100$ corresponds to the strongest evidence the system admits, and $I = 0$ to no evidential contribution. The interpretation of intermediate values is defined mechanically by the maximum Bayes factor formula introduced in Section 4.

The choice of the word *Importance* in preference to *Weight* avoids collision with I. J. Good's technical term *weight of evidence* (Good, 1950), which denotes log BF. Under our parameterisation the user-set Importance is monotonically related to Good's weight of evidence, but the two are not identical, and reserving distinct terms for each preserves clarity.

Confidence ($C_n \in [0, 1]$). A parameter attached to every non-CA node. Confidence encodes *how well-established the argument itself is*. It is presented in the interface as a percentage (0–100%) but treated internally as a real value in $[0, 1]$.

The role of Confidence differs between leaves and internal nodes, and this dual role is intentional:

- On a *leaf* node, C_n is the terminal assessed truth of that node; the computed result is $\pi_n = C_n$.
- On an *internal* node, C_n is a *local prior* which is then updated by the node's own children according to the Bayesian procedure of Section 4. The final π_n reflects both the prior C_n and the evidence contributed by descendants.

This construction is the essential feature that makes the model *hierarchical*: every proposition in the tree, at every depth, admits its own prior belief which subsequent evidence refines. We return to this point in Section 4.1.

Prior Probability ($P_{\mathrm{CA}} \in [0, 1]$). A parameter attached only to the Central Argument. It represents the user's belief in the CA before any argument in the tree has been considered. Like Confidence, it is presented as a percentage and stored as a real value.

Argument Style ($\sigma \in [0, 1]$). A single global parameter set on the CA. It governs the aggressiveness with which evidence updates belief across the entire tree. Argument Style scales the maximum Bayes factor a fully-established argument can produce, with $\sigma = 0$ (Conservative) producing the most cautious updating and $\sigma = 1$ (Aggressive) the most permissive. Its precise effect is given by Equation (6).

The term *Argument Style* replaces the earlier internal name *sensitivity*, which collides with the standard Bayesian term for the true-positive rate of a test, $P(\text{positive} \mid \text{condition})$. That collision is not merely cosmetic: a reader familiar with Bayesian statistics may parse "sensitivity" incorrectly at first glance.

2.3 Computed fields

Two fields are populated by the assessment engine on each node.

Computed Index (π_n or s_n). Stored as `computedIndex`. In Truth Seeker mode, this is a probability $\pi_n \in [0, 1]$ representing the posterior belief in the node's claim. In Scorekeeper and Gatekeeper modes, it is a score $s_n \in [-1, 1]$.

Vetoed Flag. A boolean used only by Gatekeeper mode. Set to `true` when a hard veto condition has been triggered on the node.

2.4 Notation

Throughout the remainder of the document we write:

$$\Omega(p) = \frac{p}{1-p} \qquad \text{(the odds function)}$$

$$\Omega^{-1}(o) = \frac{o}{1+o} \qquad \text{(the inverse odds function)}$$

$$\mathrm{BF} = \frac{P(D \mid H)}{P(D \mid \neg H)} \qquad \text{(a generic Bayes factor)}$$

and treat H ("the claim is true") and $\neg H$ ("the claim is false") as an exhaustive binary partition of possibilities.

3 Bayes' Theorem: The Formal Framework

We now state the formal Bayesian framework that the Truth Seeker model claims to instantiate. Nothing in this section is original — it is included so that subsequent sections may refer to specific equations when defending specific modelling choices.

3.1 The theorem

Let H be a proposition (the "hypothesis"), and D some observed evidence ("data"). Bayes' theorem, in its most familiar form, states:

$$P(H \mid D) = \frac{P(D \mid H)\, P(H)}{P(D)}. \tag{1}$$

The denominator $P(D) = P(D \mid H)P(H) + P(D \mid \neg H)P(\neg H)$ is a normalising constant. It is often more convenient to work with the ratio of $P(H \mid D)$ to $P(\neg H \mid D)$, which is the *posterior odds*.

3.2 The odds form

Dividing Equation (1) by the corresponding expression for $P(\neg H \mid D)$, the normaliser $P(D)$ cancels, yielding the odds form of Bayes' theorem:

$$\underbrace{\frac{P(H \mid D)}{P(\neg H \mid D)}}_{\text{posterior odds}} = \underbrace{\frac{P(D \mid H)}{P(D \mid \neg H)}}_{\text{Bayes factor}} \times \underbrace{\frac{P(H)}{P(\neg H)}}_{\text{prior odds}}. \tag{2}$$

The middle factor — the ratio of the likelihoods of the observed data under the two competing hypotheses — is the *Bayes factor*, henceforth BF. It is dimensionless, positive, and lies in $(0, \infty)$, with $\mathrm{BF} = 1$ representing evidentially neutral data. Equation (2) is the form the application implements.

3.3 Combining independent evidence

If $D_1, D_2, \ldots, D_k$ are pieces of evidence that are *conditionally independent given* H — i.e., $D_i \perp\!\!\!\perp D_j \mid H$ and $D_i \perp\!\!\!\perp D_j \mid \neg H$ for all $i \neq j$ — then their joint Bayes factor factorises:

$$\mathrm{BF}(D_1, \ldots, D_k) = \prod_{i=1}^{k} \mathrm{BF}(D_i). \tag{3}$$

The posterior odds after observing all k pieces of evidence is therefore the prior odds multiplied by the product of the individual Bayes factors. This is the standard formalism underlying every Bayesian network, and it is what licenses the multiplicative aggregation described in Section 4.5.

The conditional-independence assumption is substantive, and we discuss its role and limitations in Sections 4.6 and 8 (in the second part of this document).

3.4 Jeffrey conditioning on soft evidence

Standard Bayesian updating assumes evidence is observed with certainty. In many real settings — and certainly in argument mapping — evidence is itself uncertain: the user is not asked "did D happen?" but rather "to what degree do you believe D?".

The classical treatment of this case is *Jeffrey conditioning* (Jeffrey, 1965). Suppose the user's soft evidence about proposition D takes the form of a subjective probability $q = P^*(D)$. Then their updated belief in H is:

$$P^*(H) = q\,P(H \mid D) + (1 - q)\,P(H \mid \neg D). \tag{4}$$

Equation (4) is exact but requires the two conditional posteriors $P(H \mid D)$ and $P(H \mid \neg D)$. In practice, when q is treated as a modulator of an already-specified update strength, a standard approximation reduces Jeffrey conditioning to raising the Bayes factor to the power q:

$$\mathrm{BF}_{\mathrm{soft}} = \mathrm{BF}^{q}. \tag{5}$$

This form is exact under the simplifying assumption that at $q = 0$ the evidence provides no update, and at $q = 1$ the full Bayes factor applies — both of which the Truth Seeker model requires. Equation (5) is the mathematical basis for our treatment of Confidence as an exponent, developed in Section 4.3.

3.5 Subjective versus data-derived Bayes factors

A subtle but important point separates strict statistical practice from applied Bayesian tools. In classical statistical inference, a Bayes factor is *derived* from two fully-specified likelihood models: given $P(D \mid H)$ and $P(D \mid \neg H)$ as functions of the data, one computes their ratio. The value of BF is not a free parameter.

In subjective Bayesian decision support — which is the setting of the present application — the user rarely has, or wishes to construct, explicit likelihood models. Instead, they *declare* how strong a piece of evidence would be by directly setting a value that maps onto a Bayes factor. This is entirely licit within the subjective Bayesian tradition (Savage, 1954; de Finetti, 1974), provided that the user's declaration and the interpretation the system places on it are consistent.

The Truth Seeker model is a *parameterised* Bayesian procedure in exactly this sense. The user declares an Importance I; the system interprets it as $\mathrm{BF}_{\max} = 2^{I/d(\sigma)}$ for a divisor $d(\sigma)$ that depends on the Argument Style (Equation 6). This is a legitimate mode of Bayesian reasoning, but it is not the same operation as computing a Bayes factor from data. Section 4.2 addresses the calibration of this parameterisation to a recognised evidence scale.

4 The Truth Seeker Model

We now describe the Truth Seeker assessment procedure in detail. Each subsection identifies a specific modelling choice, states the operation the code performs, and defends that operation with reference to the framework of Section 3.

4.1 Architecture: a hierarchical Bayesian tree

The Truth Seeker model treats the argument tree as a hierarchical Bayesian network. The Central Argument is the top-level hypothesis, H_{CA}. Its Main Arguments are pieces of evidence bearing on H_{CA}. Each MA is itself a proposition with its own truth value, so it too has a prior — supplied by the user as its Confidence — and is in turn updated by its own children. This construction continues recursively to the leaves.

Formally, we associate with every non-leaf node n a Bernoulli random variable $T_n \in \{0, 1\}$ representing whether the claim at n is true. For the CA, the prior is $P(T_{\mathrm{CA}} = 1) = P_{\mathrm{CA}}$. For every other internal node n, the prior is $P(T_n = 1) = C_n$. For a leaf, we identify $\pi_n = C_n$ directly.

The computed index π_n at any internal node is the posterior probability $P(T_n = 1 \mid \text{evidence from ch}(n))$, computed via Equation (2) with the aggregation rules described below.

4.2 The maximum Bayes factor and its calibration

For a child node c of parent n, we define the *maximum Bayes factor* $\mathrm{BF}_{\max}(c)$ as the Bayes factor the child would contribute if the child's own claim were fully established (i.e., $\pi_c = 1$). The Truth Seeker model parameterises this as:

$$\mathrm{BF}_{\max}(c) = 2^{I_c / d(\sigma)}, \qquad d(\sigma) = 30 - 15\sigma. \tag{6}$$

The choice of exponential base 2 places our Importance parameter directly on the log-Bayes-factor scale used in the classical literature. Specifically, $\log_2 \mathrm{BF}_{\max}(c) = I_c / d(\sigma)$ is Good's *weight of evidence* (Good, 1950) in bits.

The divisor $d(\sigma)$ is calibrated so that at Importance $I = 100$, the maximum Bayes factor at the three canonical Argument Style values falls exactly on Jeffreys' evidence-strength scale (Jeffreys, 1961):

Argument Style	$d(\sigma)$	$\mathrm{BF}_{\max}$ at $I = 100$	Jeffreys' interpretation
$\sigma = 0.00$ (Conservative)	30.00	10.08	Strong
$\sigma = 0.67$ (Balanced)	19.95	32.28	Very strong
$\sigma = 1.00$ (Aggressive)	15.00	101.59	Decisive

This calibration is not incidental; the constants 30 and 15 were chosen so that $\mathrm{BF}_{\max}(I = 100)$ lands, at each canonical style, on a recognised category of the Jeffreys scale. The Argument Style parameter therefore has a defensible semantic meaning: it selects the strength that "the strongest possible argument in this tree" will be interpreted as producing.

Table 1 gives the full calibration reference at $I = 100$ across all six preset style values displayed in the user interface.

4.3 Geometric interpolation for partial confidence

The maximum Bayes factor $\mathrm{BF}_{\max}(c)$ is the strength the child contributes if fully established. When the child is only partially established — i.e., $\pi_c < 1$ — the Truth Seeker model scales its

Table 1: Argument Style calibration reference at Importance $I = 100$.

Preset	σ	$d(\sigma)$	$\mathrm{BF}_{\max}$	Jeffreys' interpretation
Conservative	0.00	30.00	10.08	Strong
Cautious	0.20	27.00	13.03	Strong
Pragmatic	0.40	24.00	17.96	Strong / Very strong
Balanced	0.67	19.95	32.28	Very strong
Assertive	0.85	17.25	55.60	Very strong
Aggressive	1.00	15.00	101.59	Decisive

evidential contribution accordingly. We use the *geometric* interpolation:

$$\mathrm{BF}_{\mathrm{act}}(c) = \mathrm{BF}_{\max}(c)^{\pi_c}. \tag{7}$$

This form is a direct application of the Jeffrey-conditioning approximation of Equation (5), with the child's computed probability π_c playing the role of the soft-evidence weight q. Its principal properties are:

- At $\pi_c = 0$: $\mathrm{BF}_{\mathrm{act}}(c) = 1$, i.e., no update. Correct: a completely unsupported child provides no evidence.
- At $\pi_c = 1$: $\mathrm{BF}_{\mathrm{act}}(c) = \mathrm{BF}_{\max}(c)$, the full declared strength. Correct: a fully-supported child contributes its full Bayes factor.
- Monotonic and continuous in π_c on $[0, 1]$.
- Multiplicatively factorable: $\log \mathrm{BF}_{\mathrm{act}}(c) = \pi_c \log \mathrm{BF}_{\max}(c)$, so the contribution to the log-odds scale is linear in π_c.

The final property is the reason the geometric form is preferred over the alternative *linear interpolation* $\mathrm{BF}_{\mathrm{act}} = 1 + \pi_c(\mathrm{BF}_{\max} - 1)$. Under geometric interpolation, if the Importance parameter I is understood as an "if fully established" evidence strength and Confidence C (equivalently π_c for leaves) is understood as "how well-established the argument is," then their combined effect on Good's weight-of-evidence scale is:

$$\log_2 \mathrm{BF}_{\mathrm{act}}(c) = \pi_c \cdot \log_2 \mathrm{BF}_{\max}(c) = \frac{I_c \cdot \pi_c}{d(\sigma)}. \tag{8}$$

That is, *effective evidence weight is proportional to $I \cdot C$*: exactly the product one would expect if Importance and Confidence played independent multiplicative roles. This matches the multi-attribute decision-theoretic structure of the Scorekeeper model (Section 6 in the second part of this document), where the two parameters also enter multiplicatively.

The linear alternative $\mathrm{BF}_{\mathrm{act}} = 1 + \pi_c(\mathrm{BF}_{\max} - 1)$ has an equally principled interpretation — it is the expected Bayes factor under the assumption that the child is either fully true (probability π_c) or fully false (probability $1 - \pi_c$) — but it does not respect the "effective weight equals $I \cdot C$" property. The application adopts the geometric form for this consistency reason.

4.4 Sign handling: support versus oppose

If a child c is of type *oppose*, its evidence bears against H_n rather than for it. The Bayes factor in favour of H_n is then the reciprocal of the factor computed for a support child:

$$\mathrm{BF}_{\mathrm{act}}^{\mathrm{oppose}}(c) = \frac{1}{\mathrm{BF}_{\mathrm{act}}(c)}. \tag{9}$$

This is the correct treatment: if the observation is evidence for $\neg H_n$ at strength $\mathrm{BF}_{\mathrm{act}}$, then by symmetry of Equation (2) it is evidence for H_n at strength $1/\mathrm{BF}_{\mathrm{act}}$. The application implements exactly this reciprocation.

4.5 Multiplicative aggregation across children

For a parent node n with children $\mathrm{ch}(n)$, the aggregate posterior odds are:

$$\Omega(\pi_n) = \Omega(P_n) \cdot \prod_{c \in \mathrm{ch}(n)} \mathrm{BF}_{\mathrm{act}}^{\tau(c)}(c), \tag{10}$$

where $P_n = P_{\mathrm{CA}}$ if n is the CA, $P_n = C_n$ otherwise, and $\tau(c) \in \{\mathrm{support}, \mathrm{oppose}\}$ selects the sign form of Equation (9).

This is a direct instantiation of Equations (2) and (3). The final posterior probability is recovered by $\pi_n = \Omega^{-1}(\Omega(\pi_n))$.

4.6 The conditional-independence assumption

The multiplicative aggregation of Equation (10) is licensed by the conditional-independence assumption of Equation (3). This is not a metaphor or an approximation but a substantive claim: the Truth Seeker model, in its baseline form, assumes that the truth values of a node's children are independent given the truth value of the parent.

This assumption is standard in Bayesian network modelling (Pearl, 1988) and is not unique to this application. It is the same assumption that licenses every naive Bayes classifier, every acyclic Bayesian graphical model with no lateral edges, and indeed every non-hierarchical Bayesian argument aggregation procedure in the argument-mapping literature.

In real argument trees, the assumption is frequently violated: two supporting arguments may cite overlapping evidence, or one argument may be true only if another is. When the assumption is violated, multiplicative aggregation double-counts shared evidence and overstates the posterior. This is a genuine limitation of the baseline model.

The Dependency system — documented separately — exists precisely to allow the user to declare such violations, and to apply information-theoretic corrections to the multiplicative aggregation that discount overlapping evidence appropriately. The Dependency corrections are heuristic, but they are honest heuristics with a legitimate mathematical motivation, and their design is a direct response to the substantive assumption stated here.

Users who do not declare any dependencies are, in effect, asserting that their argument tree respects conditional independence. This assertion is the user's responsibility, not the application's.

4.7 Boundary handling

The odds transformation $\Omega(p) = p/(1-p)$ is singular at $p = 1$ and vanishes at $p = 0$. To ensure numerical stability without materially altering computed results, the application clamps probabilities to the interval $[\varepsilon, 1-\varepsilon]$ with $\varepsilon = 10^{-9}$ before converting to odds. The clamp is defensive; users cannot in practice reach these bounds, and no computed result of interest is affected. Users should nonetheless be informed that Truth Seeker cannot output exactly 0.0 or exactly 1.0; the model expresses categorical claims as "essentially certain" rather than "certain," which is itself a defensible feature of a Bayesian system.

4.8 The recursive procedure

Bringing the preceding pieces together, the Truth Seeker assessment is a post-order traversal of the tree. At each leaf, we set $\pi_\ell = C_\ell$. At each internal node n, having already computed π_c for

all $c \in \mathrm{ch}(n)$, we:

1. Set the local prior $P_n \leftarrow P_{\mathrm{CA}}$ if $n = \mathrm{CA}$, else $P_n \leftarrow C_n$.
2. Convert to prior odds: $\Omega_n \leftarrow \Omega(P_n)$.
3. For each child $c \in \mathrm{ch}(n)$, compute:
$$\begin{aligned} \mathrm{BF}_{\max}(c) &\leftarrow 2^{I_c/d(\sigma)} \\ \mathrm{BF}_{\mathrm{act}}(c) &\leftarrow \mathrm{BF}_{\max}(c)^{\pi_c} \\ \mathrm{BF}^*_{\mathrm{act}}(c) &\leftarrow \begin{cases} \mathrm{BF}_{\mathrm{act}}(c) & \text{if } \tau(c) = \text{support} \\ 1/\mathrm{BF}_{\mathrm{act}}(c) & \text{if } \tau(c) = \text{oppose} \end{cases} \end{aligned}$$
4. Aggregate: $\Omega_n \leftarrow \Omega_n \cdot \prod_c \mathrm{BF}^*_{\mathrm{act}}(c)$.
5. Convert back: $\pi_n \leftarrow \Omega^{-1}(\Omega_n)$.

Every step of this procedure is either a direct application of Equation (2) or an explicitly-declared modelling choice defended above.

5 Worked Numerical Examples

We now illustrate the Truth Seeker procedure on a family of small trees. Each example uses a Central Argument with three Main Arguments, all of type *support*, all leaves, and all sharing the same Importance and Confidence values. The three Main Arguments are therefore evidentially identical.

We vary three inputs:

- The CA's Prior Probability: $P_{\mathrm{CA}} \in \{0.10, 0.50\}$.
- The Argument Style: $\sigma \in \{0.00, 0.67, 1.00\}$.
- The Importance/Confidence pair on the Main Arguments: $(I, C) \in \{(10, 0.10), (50, 0.50), (90, 0.90)\}$.

The three (I, C) combinations were chosen to illustrate different regions of the model's behaviour: low importance combined with low confidence (a weak argument the user is not sure of), moderate importance combined with moderate confidence (a plausibly-relevant argument the user has middling confidence in), and high importance combined with high confidence (a strong argument the user is sure of).

5.1 Computation

For each combination, we compute:

$$\begin{aligned} \mathrm{BF}_{\max} &= 2^{I/d(\sigma)} \\ \mathrm{BF}_{\mathrm{act}} &= \mathrm{BF}^{C}_{\max} \\ \text{posterior odds} &= \Omega(P_{\mathrm{CA}}) \cdot \mathrm{BF}^{3}_{\mathrm{act}} \\ \pi_{\mathrm{CA}} &= \Omega^{-1}(\text{posterior odds}). \end{aligned}$$

5.2 Results: Prior $P_{\mathrm{CA}} = 0.10$

5.3 Results: Prior $P_{\mathrm{CA}} = 0.50$

5.4 Discussion

Several features of these results are worth highlighting, as they illustrate both the model's intended behaviours and phenomena that will occasionally surprise users.

Table 2: Posterior probability of the Central Argument when three identical supporting Main Arguments update a prior of $P_{\text{CA}} = 0.10$.

(I, C)	σ	BF_{max}	BF_{act}	Prior odds	Posterior odds	π_{CA}
$(10, 0.10)$	0.00	1.260	1.023	0.111	0.119	0.1064
	0.67	1.415	1.035	0.111	0.123	0.1098
	1.00	1.587	1.047	0.111	0.128	0.1132
$(50, 0.50)$	0.00	3.175	1.782	0.111	0.629	0.3860
	0.67	5.681	2.384	0.111	1.505	0.6008
	1.00	10.079	3.175	0.111	3.556	0.7805
$(90, 0.90)$	0.00	8.000	6.498	0.111	30.474	0.9682
	0.67	22.805	16.681	0.111	515.535	0.9981
	1.00	64.000	42.224	0.111	8371.28	0.9999

Table 3: Posterior probability of the Central Argument when three identical supporting Main Arguments update a prior of $P_{\text{CA}} = 0.50$.

(I, C)	σ	BF_{max}	BF_{act}	Prior odds	Posterior odds	π_{CA}
$(10, 0.10)$	0.00	1.260	1.023	1.000	1.072	0.5173
	0.67	1.415	1.035	1.000	1.110	0.5260
	1.00	1.587	1.047	1.000	1.148	0.5346
$(50, 0.50)$	0.00	3.175	1.782	1.000	5.657	0.8498
	0.67	5.681	2.384	1.000	13.542	0.9312
	1.00	10.079	3.175	1.000	32.000	0.9697
$(90, 0.90)$	0.00	8.000	6.498	1.000	274.27	0.9964
	0.67	22.805	16.681	1.000	4639.81	0.9998
	1.00	64.000	42.224	1.000	75341.6	1.0000

Argument Style does what it advertises. At $(I, C) = (50, 0.50)$ with prior 0.10, the posterior varies from 0.386 (Conservative) to 0.780 (Aggressive) — a spread of nearly forty percentage points across the Argument Style range. Style has real, substantial effect on outcomes in the moderate-evidence regime where reasonable analysts most often disagree about how aggressively evidence should update belief.

Bayesian compounding of independent evidence. At $(90, 0.90)$, all three Argument Styles produce posteriors above 0.96 regardless of the prior. This is the correct Bayesian behaviour: three independent strong arguments compound to overwhelming evidence. The same phenomenon appears in more moderate form at $(50, 0.50)$, where three arguments a user has only middling confidence in nevertheless carry the posterior from 0.10 to 0.780 under Aggressive style. Some users will find this surprising — "three arguments I'm only half-confident in each shouldn't yield a 78% conclusion" is a natural reaction. It is nevertheless correct Bayes. The arithmetic is unambiguous: at $\sigma = 1$, each of the three MAs contributes a Bayes factor of $\mathrm{BF}_{\mathrm{act}} = 10.079^{0.5} \approx 3.175$; the three factors compound multiplicatively to $3.175^3 \approx 32.0$; applied to prior odds of 0.111 (corresponding to $P_{\mathrm{CA}} = 0.10$) this yields posterior odds of 3.556, which is a posterior probability of 0.780. This is the same phenomenon that underlies, for example, the classical result that multiple independent eyewitness testimonies rapidly converge to near-certainty in probabilistic jury models, even when each witness is individually only somewhat reliable. Users who find the compounded result too aggressive have three legitimate responses: reduce the Argument Style toward Conservative (which at $(50, 0.50)$ with prior 0.10 gives a much more modest 0.386); reduce the Importance or Confidence values on the individual MAs to reflect a more honest evidential assessment; or, most commonly, declare dependencies between the MAs to reflect that they are not in fact independent lines of evidence. The conditional-independence assumption of Section 4.6 is doing the work here, and the Dependency system is the appropriate correction when that assumption does not hold.

Weak evidence barely moves the needle. At $(10, 0.10)$, the posterior shifts by at most a few percentage points regardless of Argument Style. A weak argument the user is unsure of provides very little update, which is the correct answer.

Argument Style discriminates most where discrimination matters most. The behaviour of Style across the three regimes tells a coherent story. At $(10, 0.10)$ under prior 0.10, the posterior moves from 0.106 (Conservative) to 0.113 (Aggressive) — barely one percentage point of spread. At $(50, 0.50)$ under the same prior, the spread widens to nearly forty percentage points (0.386 to 0.780). At $(90, 0.90)$, the spread collapses again to about three percentage points (0.968 to 0.9999) because the posterior is saturating near certainty. The lesson: Argument Style dominates in the moderate-evidence regime where the user's disposition genuinely matters, and yields to the evidence at both extremes. In the weak-evidence extreme there is little to update, and in the strong-evidence extreme the update is overwhelming regardless of Style.

Confidence and Importance combine multiplicatively in log-Bayes-factor space. At $\sigma = 1$, examine the $\mathrm{BF}_{\mathrm{act}}$ values across the three (I, C) combinations: 1.047, 3.175, and 42.224. Taking base-2 logarithms yields log-Bayes-factors of 0.067, 1.667, and 5.400 bits respectively. The corresponding products $I \cdot C$ are 1, 25, and 81. The ratio of log-Bayes-factors between the first two cases is $1.667/0.067 = 25$, exactly matching the ratio of $I \cdot C$ values. The ratio between the first and third is $5.400/0.067 = 81$, again exact. This is Equation (8) in action: at fixed Argument Style, the log-Bayes-factor contribution of a child argument is directly proportional to $I \cdot C$. Doubling either the Importance or the Confidence doubles the log-Bayes-factor. This clean multiplicative property — shared with the Scorekeeper model's raw-score multiplication of

Importance and Confidence — is the principal reason the geometric interpolation of Section 4.3 is preferred over the linear alternative.

Prior matters, but its effect is bounded by the evidence. At $(90, 0.90)$, moving the prior from 0.10 to 0.50 moves the posterior from 0.9682 to 0.9964 (Conservative) — both essentially certain. At $(10, 0.10)$, moving the prior from 0.10 to 0.50 moves the posterior from 0.106 to 0.517 — an enormous shift, but the evidence itself contributed only fractional-percentage-point updates in either case. This is the correct Bayesian behaviour: when evidence is strong, it swamps the prior; when evidence is weak, the prior dominates.

These behaviours are not artefacts of the parameterisation or the calibration choices; they are direct consequences of Bayes' theorem applied through the aggregation rules of Section 4. A user who wishes to verify that the application is genuinely Bayesian can reproduce every entry of Tables 2 and 3 with a calculator, using only Equations (6)–(10).

6 The Scorekeeper Model

The Scorekeeper model is the application's implementation of a *weighted averaging model* (WAM) drawn from classical multi-attribute decision theory (Keeney & Raiffa, 1976; Edwards & Barron, 1994). It does not attempt to represent belief in a claim; instead, it produces a normalised aggregate score reflecting how well the totality of a user's arguments favour or disfavour the proposition at each node. It is offered for decisions in which a probabilistic interpretation is not the user's goal — most commonly, structured comparisons among options against a fixed set of criteria.

6.1 Motivation from multi-attribute decision theory

Consider a user deciding whether to purchase a particular car. They may care about several criteria: the car's colour, its price, its fuel economy, its safety rating. Each criterion has an *importance* to the decision (how much the user cares about it), and each option satisfies each criterion to some *degree* (how well this particular car meets that criterion). Multi-attribute utility theory prescribes that the overall score for an option is the weighted sum of criterion satisfaction levels, with the weights given by the criterion importances (Keeney & Raiffa, 1976).

This decomposition maps naturally onto an argument tree. Each argument beneath the Central Argument is a criterion; its Importance parameter $I \in [0, 100]$ encodes how much the criterion matters; its Confidence parameter $C \in [0, 1]$ encodes how well the particular option (the CA's claim) meets that criterion. The Scorekeeper model implements exactly this weighted average.

The car-buying example makes clear why Importance and Confidence must be separate parameters. A user might rate the colour *blue* as important to their satisfaction ($I = 90$) while acknowledging that the particular car they are considering does not quite match their preferred shade ($C = 0.4$). The overall contribution of this criterion is neither the importance alone nor the confidence alone but their product: $I \cdot C = 36$ points of a possible 90. The separation is not a UI convenience — it is the natural decomposition that multi-attribute decision theory requires. This is also why the geometric interpolation adopted for the Truth Seeker model in Section 4.3 is philosophically compatible: in log-Bayes-factor space, Truth Seeker's contribution is likewise proportional to $I \cdot C$ (Equation (7)). Both models honour the same underlying decomposition; they only differ in the space in which the aggregation is performed.

6.2 Formal definition

For a leaf node ℓ, the Scorekeeper computed score is:

$$s_\ell = C_\ell, \tag{11}$$

identifying the leaf's assessed contribution with its user-declared Confidence.

For an internal node n with children $\mathrm{ch}(n) = \{c_1, \ldots, c_k\}$, the aggregate score is:

$$s_n = \frac{\sum_{i=1}^{k} \varepsilon(c_i)\, I_{c_i}\, s_{c_i}}{\sum_{i=1}^{k} I_{c_i}}, \qquad \varepsilon(c) = \begin{cases} +1 & \text{if } \tau(c) = \text{support} \\ -1 & \text{if } \tau(c) = \text{oppose} \end{cases} \tag{12}$$

with the special case that if $\sum I_{c_i} = 0$ then $s_n = 0$. The result is clamped to the interval $[-1, +1]$, though for a well-formed tree with $C_\ell \in [0, 1]$ on every leaf, the clamp is defensive only.

Equation (12) is precisely the weighted mean of child scores, weighted by their Importance values, with the sign of each child's contribution flipped when the child is an *oppose* argument. The denominator normalises the sum so that a hypothetical tree in which every criterion is maximally satisfied ($s_{c_i} = 1$ for all support, $s_{c_i} = -1$ for all oppose) yields $s_n = 1$; conversely, a tree of unsatisfied criteria yields $s_n = -1$; a balanced tree yields $s_n = 0$.

6.3 Range and interpretation

The Scorekeeper's computed score s_n takes values in $[-1, +1]$. The user interface presents this as a number in $[-10.0, +10.0]$, obtained by multiplying by ten and formatting to one decimal place. The convention is that positive values favour the parent claim; negative values oppose it; zero indicates equipoise. The score has no probabilistic interpretation and should not be treated as one — a score of 0.7 does not mean "70% probability that the claim is true." It means "on the balance of the criteria and their satisfaction levels declared, the net position favours the claim with strength seven-tenths of the possible maximum."

This is a legitimate quantity for decision-making, and indeed for many practical problems — purchases, hires, feature prioritisation — it is the quantity the user actually wants. The Scorekeeper mode exists precisely for these cases, where the user's goal is normative deliberation rather than probabilistic assessment.

6.4 Interaction with Argument Style

The Scorekeeper model does not consult the Argument Style parameter σ. Its aggregation is a pure weighted average and admits no notion of "update aggressiveness." Users who switch from Truth Seeker to Scorekeeper mode will find that adjustments to the Argument Style slider have no effect on Scorekeeper results. This is by design, not an oversight: the Scorekeeper's mathematics does not have a slot for such a parameter.

6.5 Interaction with Confidence on internal nodes

The Scorekeeper model uses the Confidence parameter only on leaves. For an internal node, the node's own Confidence is not consulted; its score is entirely determined by its children's scores and their Importance weights. This contrasts with the Truth Seeker model, where internal-node Confidence plays the role of a local prior (Section 4.1). The dual role of Confidence discussed in Section 4.1 is therefore Truth-Seeker-specific; in Scorekeeper mode Confidence has a single, uniform meaning as "satisfaction level of this criterion" and is only meaningful at the leaves.

6.6 Worked example

To make the contrast with Truth Seeker concrete, we compute the Scorekeeper result for the same trees analysed in Section 5: a CA with three identical supporting Main Arguments (leaves), varying (I, C).

For a tree of k identical supporting leaves with parameters (I, C), Equation (12) collapses to $s_{\text{CA}} = C$, independent of I and independent of k. The three worked-example cases therefore yield:

(I, C)	Scorekeeper s_{CA}	UI display
$(10, 0.10)$	0.10	1.0
$(50, 0.50)$	0.50	5.0
$(90, 0.90)$	0.90	9.0

Two features of these results deserve comment.

The Scorekeeper does not compound. At $(90, 0.90)$, Scorekeeper reports a score of 0.90; Truth Seeker (Table 2) reports 0.9682–0.9999 depending on prior and Style. The Scorekeeper does not amplify the impact of multiple concurring criteria in the way Bayesian updating does. This is philosophically appropriate for its role: for decision-making purposes, three separate reasons that each satisfy the "blue car" criterion at 90% are still only 90%-satisfaction of that criterion — they do not somehow make the car more blue. The Bayesian compounding that appears in Truth Seeker is a specific feature of evidence-updating, not a general feature of preference aggregation.

The Scorekeeper does not consult Importance in this special case. For a tree of identical supporting leaves, I cancels from Equation (12) because it appears identically in numerator and denominator. This is not a bug: it reflects the mathematical fact that if every criterion has the same weight, the weighted mean equals the simple mean, and only the satisfaction levels remain. To see Importance play a role, the criteria must differ in Importance or the arguments must include a mix of support and oppose types.

A mixed-criterion illustration. Consider a CA with two supporting arguments (each $I = 50$, $C = 0.5$) and one opposing argument (also $I = 50$, $C = 0.5$). The numerator of Equation (12) is $(+1)(50)(0.5) + (+1)(50)(0.5) + (-1)(50)(0.5) = 12.5$; the denominator is 150; the score is 0.167 (UI display: 1.7). The two supporting arguments outweigh the one opposing argument, but only modestly, because they are the same strength on the same scale. This is exactly the behaviour a decision-maker would expect from a weighted average.

6.7 When Scorekeeper is the right choice

The Scorekeeper mode is appropriate whenever the user's question is normative (*how well does this option meet my criteria?*) rather than epistemic (*how confident should I be that this claim is true?*). Purchase decisions, hiring rubrics, product prioritisation, policy trade-off analysis, and multi-criteria comparison of alternatives are natural fits. Investigations of factual claims, historical hypotheses, scientific propositions, and evidential arguments are natural fits for Truth Seeker instead. The application does not enforce the choice — the user selects the model per Central Argument — but the mathematics of each model reflects a different intended purpose.

7 The Gatekeeper Model

The Gatekeeper model is a constrained variant of Scorekeeper: it performs the same weighted-average aggregation, but overlays a veto condition that forces a fail result whenever a designated *essential* argument is unsatisfied. It is the appropriate model when the user's decision has non-negotiable requirements — constraints so important that failing any one of them disqualifies the option, regardless of how well the remaining criteria are satisfied.

7.1 Motivation

Weighted-average models are appropriate for trade-offs but wrong for hard constraints. A car may score well on colour, price, and fuel economy, yet be immediately disqualified if it lacks side airbags. A candidate may have an excellent CV yet be disqualified if they lack the required work authorisation. In these cases, no amount of strength on the other criteria should compensate for failure on the essential one; the weighted-average model would incorrectly permit such compensation.

The formal literature refers to this pattern as a *lexicographic* or *conjunctive* decision rule (Tversky, 1972) — a rule that applies constraint-testing before any compensatory aggregation. The Gatekeeper model implements the simplest form of this pattern.

7.2 Formal definition

Each argument c carries a Boolean flag $E_c \in \{\text{false}, \text{true}\}$ (the `isEssential` field), user-declared, indicating whether the argument is essential to its parent claim. The Gatekeeper computation is:

1. For each child $c \in \text{ch}(n)$, check whether $E_c = \text{true}$ and the child's computed score $s_c \leq 0$.
2. If any such child exists, the veto condition is triggered: set $s_n = -1$ and mark n as vetoed.
3. Otherwise, compute s_n using the Scorekeeper formula, Equation (12).

The formal statement is:

$$s_n = \begin{cases} -1, \text{ vetoed} & \text{if } \exists\, c \in \text{ch}(n) \text{ with } E_c = \text{true and } s_c \leq 0 \\ \text{Equation (12)} & \text{otherwise} \end{cases} \tag{13}$$

7.3 Interpretation and UI representation

The Gatekeeper's output is displayed in the UI as either **PASS** or **FAIL**. FAIL is shown when the veto flag is set *or* when the underlying Scorekeeper score is negative; PASS is shown otherwise. This means Gatekeeper differs from Scorekeeper not only in its veto behaviour but in its final presentation — what Scorekeeper shows as a numerical score, Gatekeeper collapses to a categorical decision.

Note that the veto threshold is $s_c \leq 0$, not $s_c < 0$. An essential child that scores exactly zero triggers the veto. This is a deliberate choice: a criterion that is essential and unresolved (score of zero) should not silently pass, because "unresolved" is not the same as "satisfied."

7.4 Composition

Because each internal node in the Gatekeeper model computes its own veto check, vetos propagate upward through the tree in the natural way: a vetoed sub-argument itself has a score of -1, which will trigger its parent's veto if the sub-argument is itself marked essential to that

parent. The user therefore controls the scope of a veto by the placement of the `isEssential` flag: essential-at-the-CA means the whole assessment fails; essential-at-a-branch means only that branch's score is disqualified.

7.5 When Gatekeeper is the right choice

Gatekeeper is appropriate when the user's decision includes at least one non-negotiable requirement. It is not appropriate for pure trade-off decisions — if no criterion is truly disqualifying, the essential-flag mechanism will produce categorical results where a graded assessment would be more informative. In practice, users may reach for Scorekeeper by default and only escalate to Gatekeeper when they explicitly want to encode a constraint.

8 Modelling Choices and Their Limitations

This section addresses, in a single place, the modelling choices made throughout the assessment engine and the limitations they impose. The intent is to make explicit what a technically-literate reader could otherwise assemble from close reading of the preceding sections. It is written in the spirit that a reader who is genuinely evaluating the application's Bayesian claims deserves to know where the model makes strong assumptions, where it uses approximations, and where its answers should not be over-interpreted.

8.1 Geometric versus linear interpolation for soft evidence

The Truth Seeker model's treatment of partial confidence is $\mathrm{BF}_{\mathrm{act}} = \mathrm{BF}_{\mathrm{max}}^{C}$ (Equation (7)). An alternative, $\mathrm{BF}_{\mathrm{act}} = 1 + C \cdot (\mathrm{BF}_{\mathrm{max}} - 1)$, is equally common in the argument-mapping literature. Both share the correct limits ($\mathrm{BF}_{\mathrm{act}} = 1$ at $C = 0$ and $\mathrm{BF}_{\mathrm{act}} = \mathrm{BF}_{\mathrm{max}}$ at $C = 1$), and both are monotonic and continuous on $[0, 1]$. Neither corresponds exactly to strict Bayesian conditioning on uncertain evidence, which would require Jeffrey's rule and the full conditional posteriors — unavailable in this framework by design.

The geometric form was chosen for three reasons. First, it is the standard soft-evidence approximation to Jeffrey conditioning under the assumption that at $C = 0$ evidence provides no update. Second, it produces the clean multiplicative identity $\log \mathrm{BF}_{\mathrm{act}} = C \cdot \log \mathrm{BF}_{\mathrm{max}}$, so that effective log-Bayes-factor contribution is exactly $I \cdot C$ up to the divisor. Third, that multiplicativity aligns Truth Seeker with Scorekeeper's log-space semantics: both models treat the user's Importance-and-Confidence declaration as multiplying to produce the effective contribution. The linear form fails this alignment: it satisfies neither the multiplicative identity nor the log-space consistency.

The practical difference between the two forms is significant only in the moderate-confidence regime. At $C = 0$ and $C = 1$ they agree exactly; at $C = 0.5$ and $\mathrm{BF}_{\mathrm{max}} = 10$, the linear form gives 5.5 while the geometric form gives 3.16 — a ratio of 1.74. This is the regime in which the model's honesty most matters, and the choice of interpolation is therefore not cosmetic.

8.2 The conditional-independence assumption

The multiplicative aggregation of Equation (10) assumes that a node's children are conditionally independent given the truth of the parent. This assumption is standard in Bayesian network modelling, but it is nonetheless a substantive claim, and its violation is the single largest source of over-confident Truth Seeker outputs.

Two structures particularly violate the assumption. *Common-cause structures* — where multiple children draw on the same underlying evidence — cause the multiplicative aggregation to double-count the shared evidence. *Chain structures* — where the truth of one argument

depends on the truth of another — create direct dependencies between siblings that the tree model does not represent. The Dependency system exists precisely to allow the user to declare such structures and to apply corrections that discount the redundant portion.

The corrections themselves — information-theoretic overlap discounts for common-cause dependencies, causal dampening for chain dependencies, and geometric-mean aggregation for correlated groups — are engineering heuristics rather than derivations from a specified joint distribution. Their design principles and formulas are documented in the companion *Dependency Mathematics* reference. What is important to state here is that the baseline Truth Seeker model, in the absence of any user-declared dependencies, is asking the user to assert conditional independence by default. Users who leave the Dependency system unused are, by implication, claiming that their argument tree respects this assumption. That claim is the user's, not the application's.

8.3 Parameterised versus data-derived Bayes factors

The Truth Seeker model is subjective-Bayesian in the sense of Savage (1954) and de Finetti (1974): its Bayes factors are declared by the user via the Importance dial and interpreted by the system through the calibration of Equation (6). This is a legitimate mode of Bayesian reasoning, but it is not the same operation as computing a Bayes factor from data via specified likelihood functions.

The practical consequence is that the model's outputs inherit their calibration from the user's Importance declarations. If a user consistently over-rates the importance of their arguments, their Truth Seeker outputs will be systematically over-confident; if consistently under-rates, systematically under-confident. The Argument Style parameter provides a global control on this calibration, and the Jeffreys' scale correspondence documented in Section 4.2 gives users an external reference for what different Importance values mean in the traditional evidence-strength vocabulary. Beyond this, the model does not attempt to correct for user miscalibration — it treats the user's declarations as authoritative.

Users who wish to align their Importance declarations to conventional evidence-strength categories can refer to the Jeffreys' scale in Table 1: an argument they consider "strong" evidence should receive an Importance approaching 100 at Conservative Style; "very strong" at Balanced Style; "decisive" at Aggressive Style. The scale is a suggestion, not a requirement.

8.4 Boundary handling

The odds transformation is undefined at $p = 1$ (infinite odds) and vanishes at $p = 0$. Both models clamp probabilities to $[\varepsilon, 1 - \varepsilon]$ with $\varepsilon = 10^{-9}$ before conversion to odds. The clamp is defensive; it exists to prevent numerical overflow rather than to modify the model's outputs. Users cannot in practice reach these bounds through normal parameter settings, but they should nonetheless understand that Truth Seeker cannot output exactly 0.0 or exactly 1.0. The application's expression of categorical belief is "essentially certain" rather than "certain," which is itself a defensible feature of a Bayesian system: no finite quantity of evidence can produce mathematical certainty.

8.5 What the model does not do

For completeness, we list several things the assessment engine does not attempt, so that the reader is clear about the scope of its claims:

- **The model does not detect cycles.** A tree with cyclic dependencies declared via the Dependency system is handled by an iterative-convergence procedure with a bounded iteration

count; convergence is not guaranteed, and the fixed-point interpretation of the resulting numbers is heuristic.

- **The model does not learn user calibration.** If a particular user's Importance declarations are systematically too high or too low, the model produces systematically biased outputs. No calibration correction is applied. This is a deliberate choice: automatic recalibration would obscure the transparency that makes user-declared Bayes factors defensible in the first place.
- **The model does not enforce coherence across trees.** If a user builds two argument trees on the same topic with different Importance values on shared claims, both trees produce their own internally-consistent results, without cross-tree consistency checks.
- **The model does not aggregate uncertainty over user parameters.** The user's declared Importance and Confidence values are treated as point estimates. A fully-hierarchical Bayesian treatment would place distributions over these parameters and marginalise; the Truth Seeker model does not.
- **The model does not represent negation as a distinct operation.** An oppose argument at strength $\mathrm{BF}_{\mathrm{act}}$ is treated as evidence for the parent claim at strength $1/\mathrm{BF}_{\mathrm{act}}$. This is correct under a strict binary partition of the parent's hypothesis space (H versus $\neg H$), but is a simplification when the parent claim admits richer alternatives (three or more mutually exclusive hypotheses). The model does not attempt to represent such richer hypothesis spaces.

None of these limitations is a bug; each is a modelling choice that trades expressiveness for tractability, transparency, or user comprehension. The reader who understands them is in a position to use the tool honestly.

9 Tooltip and UI Text Reference

The following are the authoritative tooltip strings adopted by the application for the numerical-assessment parameters. Each tooltip has been written to be honest about the mathematical role of the parameter it describes, in the sense established in the preceding sections. The intent is that a user who reads only the tooltip receives a correct if abbreviated understanding of the parameter, and a user who reads this document and the tooltip together finds no disagreement between them.

9.1 Central Argument parameters

Prior Probability. *Your starting belief in the Central Argument, before any of the arguments in the tree have been considered. Expressed as a percentage from 0 to 100.*

Argument Style. *Controls how aggressively evidence updates belief across the tree, in Truth Seeker mode. Conservative styles produce cautious updates; Aggressive styles produce larger belief shifts for the same evidence. Has no effect in Scorekeeper or Gatekeeper modes.*

Assessment Model. *Selects the mathematical model used to compute results: Truth Seeker (Bayesian belief updating), Scorekeeper (weighted-average scoring), or Gatekeeper (weighted-average scoring with essential-criterion vetoes).*

9.2 Sub-argument parameters

Importance. *How strong this argument would be as evidence for the parent claim, if fully established. A higher Importance means this argument, when supported, will shift belief in the parent claim more.*

Confidence. *How well this argument is itself established. For arguments with no sub-arguments (leaves), this is the final assessed truth level of the argument. For arguments with sub-arguments, this acts as a prior belief that the sub-arguments will then update.*

Essential. *When enabled, this argument becomes a hard requirement for the parent claim in Gatekeeper mode. If this argument's score is zero or negative, the parent claim's result will be FAIL regardless of the other arguments.*

9.3 Terminology reference

UI term	Formal meaning
Central Argument (CA)	Root hypothesis of the tree
Main Argument (MA)	Immediate child of the CA
Sub-argument	Any non-root node
Leaf	Node with no children
Prior Probability	$P_{\text{CA}} \in [0, 1]$, prior belief in the CA
Importance	$I \in [0, 100]$, evidential strength if fully established
Confidence	$C \in [0, 1]$, degree to which the argument itself is established
Argument Style	$\sigma \in [0, 1]$, updating aggressiveness
Support/Oppose	Type $\tau \in \{\text{support}, \text{oppose}\}$
Essential	Boolean $E \in \{\text{false}, \text{true}\}$
Truth Seeker	Hierarchical Bayesian assessment mode
Scorekeeper	Weighted-average assessment mode
Gatekeeper	Weighted-average mode with essential-criterion vetoes
Dependency	User-declared correlation or causal link between arguments

10 References

The mathematical framework of this document rests on standard results in probability theory, Bayesian statistics, and decision theory. Readers wishing to consult the primary sources for the specific results invoked will find the following useful.

10.1 Bayes' theorem and Bayesian statistics

- Bayes, T. (1763). "An Essay towards solving a Problem in the Doctrine of Chances." *Philosophical Transactions of the Royal Society of London*, 53, 370–418. The original statement of what is now called Bayes' theorem.
- Jeffreys, H. (1961). *Theory of Probability*, 3rd ed. Oxford University Press. Chapter 5 introduces the Bayes factor evidence-strength scale used to calibrate the Argument Style parameter in Section 4.2 of this document.
- Good, I. J. (1950). *Probability and the Weighing of Evidence*. Charles Griffin. Introduces the "weight of evidence" as log BF; the base-2 logarithm gives the quantity in bits, on which our Importance parameter is calibrated.

- Jeffrey, R. C. (1965). *The Logic of Decision*. McGraw-Hill. Chapter 11 develops the classical theory of updating on uncertain (soft) evidence; the approximation $\mathrm{BF}_{\mathrm{act}} = \mathrm{BF}^{C}_{\mathrm{max}}$ used by the Truth Seeker model is a standard reduction of Jeffrey conditioning.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann. The reference work on Bayesian networks and their conditional-independence assumptions.

10.2 Subjective Bayes and decision theory

- Savage, L. J. (1954). *The Foundations of Statistics*. Wiley. Foundational work on subjective probability.
- de Finetti, B. (1974). *Theory of Probability*. Wiley. The classical treatment of Bayesian probability as personal degree of belief; provides the theoretical grounding for our parameterised (rather than data-derived) Bayes factors.
- Keeney, R. L., and Raiffa, H. (1976). *Decisions with Multiple Objectives*. Wiley. The reference for multi-attribute utility theory underlying the Scorekeeper model.
- Edwards, W., and Barron, F. H. (1994). "SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement." *Organizational Behavior and Human Decision Processes*, 60(3), 306–325. A practical treatment of weighted-average decision models.
- Tversky, A. (1972). "Elimination by aspects: A theory of choice." *Psychological Review*, 79(4), 281–299. The classical treatment of lexicographic and conjunctive decision rules underlying the Gatekeeper model.

10.3 General references and further reading

- MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press. A modern textbook covering both the Bayesian framework and its information-theoretic foundations. Chapter 3 covers Bayesian inference; Chapter 37 covers Bayesian model comparison and Bayes factors.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*, 3rd ed. Chapman & Hall/CRC. The reference textbook on modern applied Bayesian analysis; useful for the reader who wishes to see how the framework generalises beyond the binary-hypothesis setting of this document.
- Kass, R. E., and Raftery, A. E. (1995). "Bayes factors." *Journal of the American Statistical Association*, 90(430), 773–795. A comprehensive review of Bayes factors in practical statistical inference, including a discussion of the Jeffreys' scale we adopt for calibration.

The companion Dependency Mathematics *document extends this material to cover the corrections applied when users declare violations of the conditional-independence assumption of Section 4.6.*