On Aggregating Probabilistic Evidence

Sergei Artemov

CUNY Graduate Center

September 30, 2021

Objectives

Imagine a database – a set of propositions $\Gamma = \{F_1, F_2, \ldots, F_n\}$ with some kind of probability estimates and let X logically follow from Γ. What is the best lower bound of the probability of X?

The traditional approach computes a crude numeric lower bound for X from **probabilities** p_i's of F_i's corresponding to a worst-case configuration.

We offer a more flexible approach by first calculating **aggregated evidence** for X in Γ as a symbolic term of elementary evidence for the assumptions from Γ and then finding a tight lower bound for any, not only a worst-case, situation.

Outline

Probabilistic reading of logical entailment

Let proposition X logically follow from assumptions

$$\Gamma = \{F_1, F_2, \ldots, F_n\};$$

symbolically,

$$\Gamma \models X.$$

In standard set-theoretical semantics, this states that the truth set of X is the whole space if the truth set of each proposition from Γ is the whole space. A similar observation shows that if the probability of Γ is 1, then the probability of X is also 1.

But what happens in a general case when formulas from Γ have arbitrary probabilities?

Standard approach

The logical observation "X is true whenever Γ is true" yields the approach taken in inductive probability reasoning. The Suppes rule

$$\frac{P(A) \geq r \quad P(B \mid A) \geq p}{P(B) \geq rp}$$

is basically the special case of the logical rule *Modus Ponens*

$$A,\ A \rightarrow B \vdash B.$$

Further modifications were made for accumulating evidence for X throughout the entire database. In particular, the traditional Adams' Probability Logic introduces weights based on so-called *degrees of essentialness* of premises from Γ, etc.

Baby example

However, the very concept of drawing probability estimates for X **only from probability estimates for** Γ is limited: we are forced to consider worst-case scenarios leading to sub-optimal estimates.

Consider a simple derivation

$$A, B, C \models A \wedge B \wedge C.$$

For "high end" probabilities close to 1, probability-based estimates could make sense: in this case, if

$$P(A), P(B), P(C) \geq 0.99,$$

then it is easy to check that the tight low estimate is

$$P(A \wedge B \wedge C) \geq 0.97.$$

Baby example

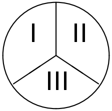
In social situations, however, "medium-range" probabilities are more typical, and for them the probability-based approach fails. For the same example

$$A, B, C \models A \wedge B \wedge C$$

and probabilities

$$P(A), P(B), P(C) \geq 2/3,$$

the corresponding **tight** lower bound for $P(A \wedge B \wedge C)$ is 0 (cf. the example of such A, B, and C below), which is meaningless.



$$A = \mathrm{I} \cup \mathrm{II}$$
$$B = \mathrm{II} \cup \mathrm{III}$$
$$C = \mathrm{I} \cup \mathrm{III}$$

Introducing evidence parameters

Instead of using a numerical lower bound p for the probability of F

$$p \leq P(F)$$

for some known p, we suggest using the evidence format:

$$u \subseteq F$$

for some event u. Given evidence $u_1, u_2, \ldots, u_n$ for $F_1, F_2, \ldots, F_n$ we build **aggregated evidence** $e(u_1, u_2, \ldots, u_n)$ for X,

$$e(u_1, u_2, \ldots, u_n) \subseteq X,$$

which provides a parameterized lower bound of probability of X:

$$P(e(u_1, u_2, \ldots, u_n)) \leq P(X).$$

Baby example with evidence parameters

For the same example

$$A, B, C \models A \wedge B \wedge C,$$

we introduce evidence variables u, v, w denoting events (subsets of a probability space) on which A, B, C respectively hold. The logical derivation suggests that $A \wedge B \wedge C$ is secured on event:

$$e(u, v, w) = uvw,$$

(we write st for $s \cap t$, for better readability) which we offer as **aggregated evidence** for $A \wedge B \wedge C$:

$$uvw \subseteq A \wedge B \wedge C.$$

If $P(u) = P(v) = P(w) = 2/3$, $P(uvw)$ ranges from 0 to 2/3.

Representative example

Suppose we are given events u, v, and w, each with probability $1/3$, which are supportive evidence for F, $F \to X$ and X respectively, i.e.,

1. $u{:}F$;
2. $v{:}(F \to X)$;
3. $w{:}X$.

Here $\Gamma = \{F, F \to X, X\}$ and Γ logically yields X

$$\Gamma \models X$$

in propositional logic.

Representative example

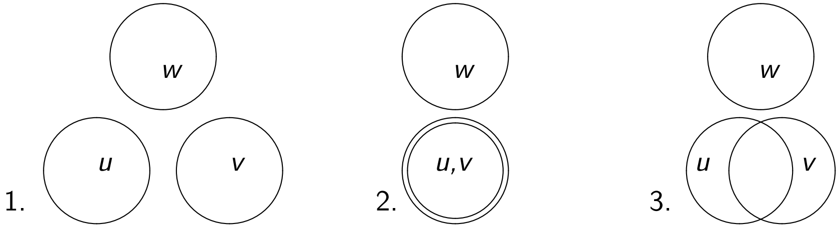
What is the aggregation of evidence for this X and what is its probability? The answer depends on set-theoretical configurations of u, v, and w.

In configuration 1, Figure 1, events u and v are incompatible and hence do not contribute to the evidence aggregation for X, which is therefore equal to w and has probability $1/3$.

In configuration 2, $u = v$ and hence the whole event $u \cup w$ supports X: the probability of X is $2/3$.

In configuration 3, the contributing sections to the evidence aggregation for X are $u \cap v$ (which we denote uv for better readability) and w, with the probability $P(uv \cup w)$ which, with an additional assumption that $P(uv) = 1/6$, is $1/2$.

Representative example



Some configurations of u, v, and w.

Representative example

These and all other configurations are covered by a uniform "evidence term"

$$uv \cup w^1 \tag{1}$$

which can be obtained by logical reasoning from Γ: there are two ways to justify X: either as a logical conclusion from F and $F \to X$, which is valid on uv, or directly from w, hence aggregated evidence (1).

We put this type of reasoning on solid logical and mathematical ground. In particular, we show that (1) is indeed the best logically justified evidence for X here.

[1] read these operations lattice-theoretically: "$\cdot$" as *meet* and "$\cup$" as *join*.

General setting

We are interested in estimating a probability $P(X)$ of an unspecified event X. Suppose X logically follows from a set of probabilistic assumptions

$$\Gamma = \{F_1, F_2, \ldots, F_n\}$$

over a given probability space.

Assume that for each F_i we have evidence u_i;

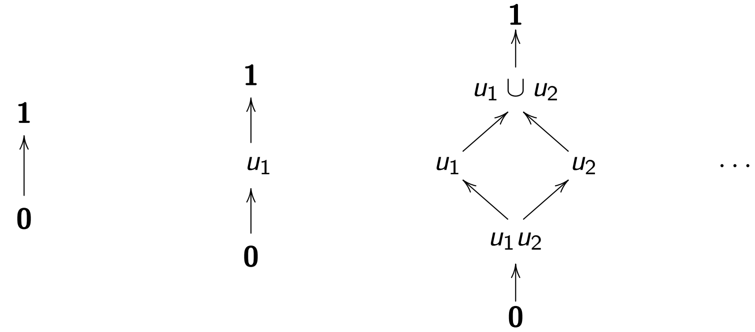
$$\vec{u}{:}\Gamma = \{u_1{:}F_1, u_2{:}F_2, \ldots, u_n{:}F_n\}.$$

Since each u_i is considered specified, we know each $P(u_i)$.

Why different vocabularies for evidence and events? They are "coefficients" and "variables": $u_1, u_2, \ldots, u_n$ are events which are presumed **specified** *and hence legitimate building material, or inputs, for aggregated evidence. Events $F_1, F_2, \ldots, F_n$ are not presumed specified.*

Logic of Evidence Aggregation LEA

Evidence lattice $\mathcal{L}_n$ is the free distributive lattice over $u_1, u_2, \ldots, u_n, \mathbf{0}, \mathbf{1}$ with operations "$\cap$," "$\cup$," and lattice order $\preceq$. Each term is equal to either $\mathbf{0}$, or $\mathbf{1}$, or a union of products $u_{i_1} \cap \ldots \cap u_{i_k}$ (or $u_{i_1} \ldots u_{i_k}$, for short). Example: $u_1 u_2 \cup u_2 u_3 \cup u_4$.



Logic of Evidence Aggregation LEA

Evidence terms are elements of the evidence lattice $\mathcal{L}_n$ (or $\mathcal{L}$, for short). Formulas are generated from propositional letters $p, q, r, \ldots$ by the usual logical connectives. We also allow formulas

$$t{:}F$$

where F is any formula and t any evidence term.

Notation:

- Fm is the set of all formulas.
- Tm is the set of all evidence terms.

Logic of Evidence Aggregation LEA

The logical postulates of LEA are

1. *axioms and rules of classical logic in the language of* LEA;
2. $s{:}(A \to B) \to (t{:}A \to [st]{:}B)$;
3. $(s{:}A \wedge t{:}A) \to [s \cup t]{:}A$;
4. $t{:}X \to s{:}X$, *for any evidence terms* s, t *such that* $s \preceq t$ *in* $\mathcal{L}$;
5. $\mathbf{1}{:}A$, *where* A *is an axiom* - a recursive postulate;
6. $\mathbf{0}{:}F$, *for each formula* F.

Necessitation rule:

$$\vdash F \Rightarrow \vdash \mathbf{1}{:}F$$

is derivable - by induction on derivation.

Some logical properties of LEA

1. LEA $\vdash (s{:}A \wedge t{:}A) \leftrightarrow [s \cup t]{:}A$.

Indeed, Axiom 4 secures $[s \cup t]{:}A \rightarrow s{:}A$ and $[s \cup t]{:}A \rightarrow t{:}A$.

2. Each evidence term t is a normal modality.

 a) LEA $\vdash t{:}(A \rightarrow B) \rightarrow (t{:}A \rightarrow t{:}B)$

(by Axiom 2 and identity $t = tt$ in $\mathcal{L}$);

 b) LEA $\vdash F$ *yields* LEA $\vdash t{:}F$. Indeed, given LEA $\vdash F$, by Necessitation, LEA $\vdash \mathbf{1}{:}F$, and, by Axiom 4, LEA $\vdash t{:}F$.

3. If $s = t$ in $\mathcal{L}$, then LEA $\vdash s{:}F \leftrightarrow t{:}F$ for each F.

(By Axiom 4).

4. Substitution: if LEA $\vdash X \leftrightarrow Y$, then LEA $\vdash F(X) \leftrightarrow F(Y)$

Probability interpretation - idea

The probabilistic reading of evidence terms t is measurable events from $\mathcal{F}$ in a given probability space $(\Omega, \mathcal{F}, P)$, formulas are subsets of Ω, constant $\mathbf{1}$ is Ω, $\mathbf{0}$ is $\emptyset$, and $t{:}F$ is understood as

t is a sufficient justification for F,

or

t is an event supporting F.

Probability interpretation

Consider a specific evaluation consisting of

- a probability space $(\Omega, \mathcal{F}, P)$;
- a mapping $*$ of propositions to 2^{Ω} and evidence terms to $\mathcal{F}$.

Assume $\mathbf{1}^* = \Omega$, $\mathbf{0}^* = \emptyset$ $(st)^* = s^* \cap t^*$, $(s \cup t)^* = s^* \cup t^*$. On propositions, interpretation $*$ is Boolean, i.e.,

$$(X \wedge Y)^* = X^* \cap Y^*, \quad (X \vee Y)^* = X^* \cup Y^*, \quad (\neg X)^* = \overline{X^*},$$

$$(t{:}X)^* = \overline{t^*} \cup X^* \text{ (a set theoretical reading of } t^* \subseteq X^*).$$

For a set of formulas Γ,

$$\Gamma^* = \bigcap \{F^* \mid F \in \Gamma\}.$$

Note that for each axiom A and each interpretation $*$, $A^* = \Omega$.

Theorem 1.
$\vdash F$ *yields that* F *holds under any probability interpretation.*

Note that LEA is not complete w.r.t. its probability semantics. For example, formula $\mathbf{1}{:}p \rightarrow p$ where p is a propositional letter is valid in probability semantics. Indeed,

$$(\mathbf{1}{:}p \rightarrow p)^* = \overline{(\mathbf{1}{:}p)^*} \cup p^* = \overline{\overline{\Omega} \cup p^*} \cup p^* = \overline{\emptyset \cup p^*} \cup p^* = \overline{p^*} \cup p^* = \Omega.$$

However, $\mathbf{1}{:}p \rightarrow p$ is not derivable in LEA. To check this, consider an artificial truth assignment under which formulas $t{:}F$ are true for all t and F, p is false, and Boolean connectives behave conventionally. Axioms of LEA are true under this assignment, hence all theorems of LEA are true. However, $\mathbf{1}{:}p \rightarrow p$ is false, hence this formula is not derivable in LEA.

We leave finding an axiomatization of probability tautologies in the language of LEA for future studies (**solved by Eoin Moore, 2021**).

Internalization

Theorem 2.

If

$$F_1, F_2, \ldots, F_n \vdash F,$$

then

$$u_1{:}F_1, u_2{:}F_2, \ldots, u_n{:}F_n \vdash [u_1 u_2 \ldots u_n]{:}F.$$

Proof. ($\Rightarrow$). Induction on the derivation of F from $F_1, F_2, \ldots, F_n$. The base cases are trivial. If F is obtained by *Modus Ponens* from $X \rightarrow F$ and X, then, by the IH, $u_1{:}F_1, u_2{:}F_2, \ldots, u_n{:}F_n \vdash p{:}(X \rightarrow F)$ and $u_1{:}F_1, u_2{:}F_2, \ldots, u_n{:}F_n \vdash q{:}X$ for some p and q. By Axiom 2, $u_1{:}F_1, u_2{:}F_2, \ldots, u_n{:}F_n \vdash [pq]{:}F$, and it suffices to set $t = pq$. So, we have found a non-zero evidence term t such that $u_1{:}F_1, u_2{:}F_2, \ldots, u_n{:}F_n \vdash t{:}F$. Since $u_1 u_2 \ldots u_n \preceq t$ for each such t, we also have

$$u_1{:}F_1, u_2{:}F_2, \ldots, u_n{:}F_n \vdash [u_1 u_2 \ldots u_n]{:}F.$$

Aggregated evidence for propositional formulas

Consider finite sets of propositional formulas $\Gamma = \{F_1, F_2, \ldots, F_n\}$ and evidence variables $V_n = \{u_1, u_2, \ldots, u_n\}$. Let LEA be built on evidence terms over V_n. We normally drop this n, for short.

Definition 3.

Evidence for a propositional formula X given Γ is a term $t(u_1, u_2, \ldots, u_n)$ such that in LEA,

$$u_1{:}F_1, u_2{:}F_2, \ldots, u_n{:}F_n \vdash t(u_1, u_2, \ldots, u_n){:}X.$$

Aggregated evidence $AE^\Gamma(X)$ for a proposition X given Γ is the evidence term

$$AE^\Gamma(X) = \bigcup\{t \mid t \textit{ is an evidence term for } X \textit{ given } \Gamma\}. \qquad (2)$$

Obviously, $AE^{\Gamma}(X)$ is the $\preceq$-largest evidence for X in Γ:

$$t \textit{ is evidence for } X \textit{ in } \Gamma \quad \textit{iff} \quad t \preceq AE^{\Gamma}(X).$$

Furthermore, $AE^{\Gamma}(X)$ can be found constructively given Γ and X.

Completeness: $AE^{\Gamma}(X)$ cannot be improved uniformly

Theorem 4.
If $\vec{u}{:}\Gamma \not\vdash t{:}X$, then $t{:}X$ fails in some probability model of $\vec{u}{:}\Gamma$.

Proof. Assume $\vec{u}{:}\Gamma \not\vdash t{:}X$, hence $t \not\preceq AE^{\Gamma}(X)$. Such t contains a meet $u_{i_1} u_{i_2} \ldots u_{i_k}$ which is not in $AE^{\Gamma}(X)$; w.l.g. we assume that $t = u_1 u_2 \ldots u_k$ for some $k \leq n$. By assumption,

$$u_1{:}F_1, u_2{:}F_2, \ldots, u_k{:}F_k \not\vdash [u_1 u_2 \ldots u_k]{:}X.$$

By Internalization

$$F_1, F_2, \ldots, F_k \not\vdash X.$$

Pick a Boolean assignment $\sharp$ which makes all F_i true and X false. Take an arbitrary probability space $(\Omega, \mathcal{F}, P)$ and evaluation $*$ that agrees with $\sharp$ on propositional letters. Apparently, all $F_i^* = \Omega$ and $X^* = \emptyset$.
Set $u_i^* = \Omega$ for $i = 1, 2, \ldots, k$ and $u_i^* = \emptyset$ for $i = k+1, k+2, \ldots, n$, which makes $(u_i{:}F_i)^* = \Omega$ for all $i = 1, 2, \ldots, n$. Furthermore, $t^* = \Omega$ as well, hence $(t{:}X)^* = \emptyset$, which means $\vec{u}{:}\Gamma \not\Vdash t{:}X$.

Model-theoretical view

Here is an algorithm of finding the aggregated evidence term $AE^{\Gamma}(X)$. Given Γ and X, find all set-theoretically minimal subsets Γ' of Γ such that

$$\Gamma' \models X$$

(this takes a finite number of satisfiability checkups) and form lattice meets

$$u_{i_1} u_{i_2} \ldots u_{i_k}$$

of evidence variables corresponding to all such Γ's. The aggregated evidence term $e = AE^{\Gamma}(X)$ is the join of these meets.

Proof-theoretical view

For each finite Γ and each X, the following is a description of $e = AE^{\Gamma}(X)$. Consider all possible derivations of X from Γ in tree-like form with axioms and assumptions at the leaf nodes and instances of *Modus Ponens* at all other nodes.

It is easy to see that in each of these derivations, aggregated evidence $s(\vec{v})$ of the root formula X is the meet of all variables from the leaf nodes. The desired aggregated evidence term $e(\vec{u})$ is the join of all these $s(\vec{v})$'s. Finiteness of the evidence lattice $\mathcal{L}$ guarantees that $AE^{\Gamma}(X)$ is a specific term in $\mathcal{L}$.

As a practical algorithm, try to build all reasonable proofs of X from Γ and have e be the join of the corresponding meets $u_{i_1} u_{i_2} \ldots u_{i_k}$. Check that any other Γ' does not yield X (a finite number of satisfiability tests).

Example of aggregated evidence

Let us return to the example with

$$\Gamma = \{F, F \to X, X\}$$

and the evidence variables assignment

$$\Delta = \{u{:}F, v{:}(F \to X), w{:}X\}.$$

Calculations show that

$$uv \cup w = AE^{\Gamma}(X).$$

Computational example

This will be a specific instance of the "motivating" example.

1. Probability space $\Omega = \{a, b, c, d, e, f\}$, each with probability $1/6$.
2. $\Gamma = \{F, F \to X, X\}$, evidence variables u, v, w.
3. $F^* = \{a, b, c\},\ X^* = \{c, d, e\}$, hence $(F \to X)^* = \{c, d, e, f\}$ (these are not known to the user).
4. Specified evidence $u^* = \{a, c\},\ v^* = \{c, d\},\ w^* = \{d, e\}$ (known to the user)

The goal is to recover the best possible evidence for X and estimate its probablity. We have already established that $AE^{\Gamma}(X) = uv \cup w$.
Evaluation: $[AE^{\Gamma}(X)]^* = (uv \cup w)^* = (u^* \cap v^*) \cup w^* = \{c, d, e\}$.
(In this case, we recover X entirely by its aggregated evidence.)

The justified lower bound p of probability of X:

$$p = P([AE^{\Gamma}(X)]^*) = P(\{c, d, e\}) = 1/2.$$

Example: comparisons with Adams' Probability Logic APL

Consider the evidence aggregation example from the SEP article "Logic and Probability":

$$\Gamma = \{p, q, r, s\}, \quad X = p \wedge (q \vee r).$$

Furthermore, $P(p) = 10/11$, $P(q) = P(r) = 9/11$, and $P(s) = 7/11$. APL yields the probability lower bound $8/11$.

The Probabilistic Evidence technique suggests introducing evidence variables x, y, z, u for propositions p, q, r, s, respectively, and calculating the aggregated evidence term, which in this case will be

$$xy \cup xz.$$

The lower bound for the probability of X, $P(xy \cup xz)$ ranges from $8/11$ to $10/11$ - up to 18 percentage points improvement.

Example: comparisons with Adams' Probability Logic APL

The contrast between APL and LEA becomes even more vivid with medium range probabilities. Consider the same example

$$\Gamma = \{p, q, r, s\}, \quad X = p \wedge (q \vee r)$$

but with $P(p), P(q), P(r), P(s) \leq 1/2$. In this case, APL yields trivial probability lower bound 0. LEA uses the same aggregated evidence term

$$xy \cup xz,$$

with parameterized probability lower bound $P(xy \cup xz)$ ranging from 0 to $1/2$ - an improvement of up to 50 percentage points.

Handling inconsistent data

In a general setting, one should not expect Γ to be logically consistent: in realistic situations, we have to deal with sets of assumptions which may contradict each other. Furthermore, we may want to gather evidence for X and for $\neg X$ from the same data. The framework of LEA naturally accommodates these needs: we can track both positive and negative evidence for X from Γ:

$$AE^{\Gamma}_{+}(X) = AE^{\Gamma}(X) = \bigcup\{t \mid t \textit{ is evidence for } X \textit{ in } \Gamma\},$$

$$AE^{\Gamma}_{-}(X) = AE^{\Gamma}(\neg X) = \bigcup\{t \mid t \textit{ is evidence for } \neg X \textit{ in } \Gamma\},$$

with positive and negative justified ratings of X in Γ for a given interpretation $*$ being the probabilities

$$P([AE^{\Gamma}_{+}(X)]^{*}) \text{ and } P([AE^{\Gamma}_{-}(X)]^{*}).$$

Handling inconsistent data

Once $*$ makes all formulas from $\vec{u}{:}\Gamma$ true, i.e., $[\vec{u}{:}\Gamma]^* = \Omega$,

$$[AE_+^\Gamma(X)]^* \subseteq X^*$$

and

$$[AE_-^\Gamma(X)]^* \subseteq \overline{X^*},$$

hence $[AE_+^\Gamma(X)]^*$ and $[AE_-^\Gamma(X)]^*$ are disjoint and

$$P([AE_+^\Gamma(X)]^*) + P([AE_-^\Gamma(X)]^*) \ \leq 1.$$

Positive and negative ratings do not necessarily sum to 1.

Handling inconsistent data: example

The same probability space and evaluation, with extended database

$$\Gamma = \{F, F \to X, X, \neg X\},$$

evidence variables $\{u, v, w, y\}$, and $y^* = \{a, b\} \subseteq (\neg X)^*$.
Γ is inconsistent, but admits meaningful evidence aggregation.

$AE_+^\Gamma(X) = uv \cup w$ and positive evidence for X is $[AE_+^\Gamma(X)]^* = \{c, d, e\}$ with probability $1/2$ (the positive rating of X in Γ).

$AE_-^\Gamma(X) = y$ and negative evidence for X is $[AE_-^\Gamma(X)]^* = \{a, b\}$ with probability $1/3$ (the negative rating of X in Γ).

Upper bounds

The event upper bound for a proposition X is a specified event s such that $X \subseteq s$ or, contrapositively, $\bar{s} \subseteq \neg X$. So, upper bounds for X corresponds to lower bounds for $\neg X$, hence the tight upper bound for X given Γ is determined by the negative evidence for X.

Specifically, for any given probability space $(\Omega, \mathcal{F}, P)$ and evaluation $*$, the tight upper evidence for X is given by the event

$$s = \Omega \setminus [AE^{\Gamma}(\neg X)]^{*}$$

and its probability

$$P(s) = 1 - \textit{negative rating of } X$$

is the tight justified upper bound for the likelihood of X. In the latter example, the upper bound of X is 1-(1/3)=2/3.

LEA from the logical perspective

Though LEA answers the question about evidence aggregation in its language, it is not complete w.r.t. probability semantics. The natural question of complete axiomatization of probability semantics was answered by Eoin Moore (the talk next week). Additional principles are

- $\mathbf{1}{:}A \rightarrow A$,
- $F \rightarrow t{:}F$ (this yields $\mathbf{1}{:}A$ for an axiom A),
- $[st]{:}F \leftrightarrow (s{:}A \vee t{:}A)$,
- $[s \cup t]{:}F \leftrightarrow (s{:}A \wedge t{:}A)$.

The corresponding logic LEA_{+} gives axiomatization of the "propositional" justification logic in which $t{:}F$ is read as the material implication $t \rightarrow F$. Moore's results also answer a long standing conceptual question about the scope of justification logics with propositional justifications.

Discussion on $\mathsf{LEA}_{+/-}$

Classical tautologies are sometimes counterintuitive. Consider the tautology which violates the causality of implication intuition:

$$[(A \wedge B) \rightarrow C] \rightarrow [(A \rightarrow C) \vee (B \rightarrow C)].$$

In LEA_{+} this reincarnates as

$$[st]{:}F \;\; \rightarrow \;\; (s{:}F \vee t{:}F). \qquad (3)$$

in which $u{:}F$ cannot be read as the proposition *'u is a subset of F.'*

This observation is a reminder that, in the probabilistic semantics, Boolean connectives are interpreted set-theoretically, whereas the evidence assertions $u{:}X$ are meta propositions with a natural *true/false* reading. Mixing Boolean connectives with evidence assertions can create tensions similar to (3).

In the evidence aggregation theory there is no such mixtures and the problem does not occur.