# Bayes and Brentano

Stephen Yablo MIT

*TMS & Related Approaches*
July 2025, Munich

# They meet

You remember Bayes. He told us how to adjust our confidence in *H* when evidence *E* is acquired. *H* had beforehand a probability *conditional* on *E*. *That*'s how confident to be in *H* afterwards.

$$(\text{B-CON})\ \mathrm{p}_n(H) = \mathrm{p}_o(H|E) = \mathrm{p}_o(H \wedge E)\ /\ \mathrm{p}_o(E). \qquad (1)$$

You remember Brentano. Mental states are *directed*, he said; they have an *intentional object* or *subject matter*. Exceptions have been alleged (undirected anxiety, nameless dread). But for propositional attitudes – $\varphi$ing that *E* — the idea seems right.

$$(\text{B-ABT})\ \varphi\text{ing that } E \text{ involves attending to } \mathbf{e}\text{, its subject matter.} \qquad (2)$$

These may seem unrelated, but let $\varphi$ be belief. If believers attend to **e**, so also presumably do those with a certain degree of belief.

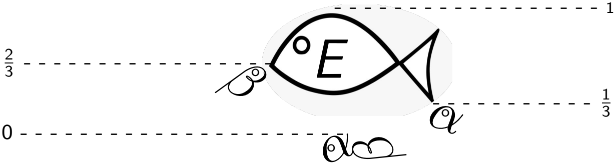
B-CON is strangely uncurious about this. It asks only for *E*'s intension **E**, the set of *E*-worlds; p(*E*) is the measure of that set. *E*'s aboutness properties are ignored.

# "Strangely" uncurious?

Curiosity would be strange, if **e** was irrelevant. Let's check. **e** for our purposes is made up of *E*'s (exact) truthmakers $\epsilon$. p($E$) is a function of their probabilities, alone and together – of, that is, the probabilities of its *t-makers* $\varepsilon$.[1] Say $E = A \vee B$,

$$\mathrm{p}(E) = \mathrm{p}(\alpha \vee \beta) = \mathrm{p}(\alpha) + \mathrm{p}(\beta) - \mathrm{p}(\alpha\beta) \quad (3)$$

$\alpha$, $\beta$, $\alpha\beta$ are like a school of helper-fish nudging $E$ somehow or other up to 1. Any number of ways of doing this, as many as reals in [0,1] such that $r+s\text{-}t = 1$.




Are the helper-fish supposed to be irrelevant? *E*'s truth consists in the holding of certain $\varepsilon$s. The ones it *probably* consists in get more of a say, surely. If $\beta$ favors $H$, $\alpha$ would have to *strongly* disfavor it, for $H$ not to come out ahead.

[1] *E*'s t-makers are its truthmakers closed under $\wedge$. Its *dt*-makers = its t-makers closed under $\vee$.

# Rigidity

T-maker neglect .... just an oversight? No, it's written into the rule's DNA. B-CON $=_{Df}$ the one and only way of incorporating $p(E)=1$ that is *rigid* in the sense that probabilities given $E$ don't change: $p_n(X|E)=p_o(X|E)$.

From rigidity it follows that $F$s that imply $E$ have got to appreciate at the same rate.

$$p_n(F) = p_o(F|E) = p_o(E\wedge F)/p_o(E) = p_o(E)\times \underline{1/p_o(E)}. \qquad (4)$$

$E$'s t-makers $\varepsilon$ certainly imply $E$. So they too appreciate at the underlined rate.

No wonder B-CON doesn't feel the need to ask about prob underpinnings $p(\varepsilon)$. There was only ever one option: $p_n(\varepsilon)=p_o(\varepsilon)/p_o(E)$. Underpinnings are in effect stipulated.

*****

But ... what do you mean they are stipulated? Who gave the classical updater that power? It's not theorists who decide these things, it's the world. T-makers $\varepsilon$ can have various probabilities, depending how $E$ is learned,

# Underpinnings

Al, Betty, and Cleo toss coins. Your money is on all three throwing heads ($ABC$), so you are pleased to learn that $A \vee B$ ($E$). p($E$) was originally $\frac{3}{4}$ ($p_o(\overline{\alpha}\overline{\beta}) = \frac{1}{4}$) and now becomes 1. p($AB$) and p($ABC$) were 1/4 and 1/8, and now become ... what?

That depends. $\alpha\beta$ can't stay put (at $\frac{1}{2}, \frac{1}{2}, \frac{1}{4}$) else p($\alpha$)+p($\beta$)-p($\alpha\beta$) is stuck at $\frac{3}{4}$. p($E$) needs new underpinnings. Continuum many can be imagined, e.g.,

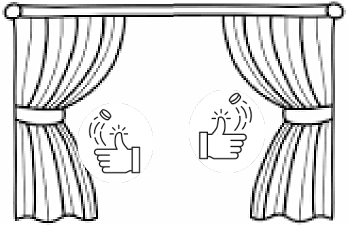
| | $A \vee B$ | $\alpha$ | $\beta$ | $\alpha\beta$ | $ABC$ |
|---|---|---|---|---|---|
| before | 3/4 | 1/2 | 1/2 | 1/4 | 1/8 |
| after | 1 | ↓ | ↓ | ↓ | ↓ |
| $\text{UP}_0$ | 1 | 2/3 | 2/3 | 1/3 | 1/6 |
| $\text{UP}_1$ | 1 | 3/4 | 3/4 | 1/2 | 1/4 |
| $\text{UP}_2$ | 1 | 7/8 | 7/8 | 3/4 | 3/8 |
| $\text{UP}_3$ | 1 | 9/10 | 9/10 | 4/5 | 2/5 |

Table: Alternative underpinnings for p($A \vee B$)=1

p($ABC$) is up in the air until you decide which $\text{UP}_i$ — which p($\varepsilon$)s — to go with. The p($\varepsilon$)s derive in turn from other probabilities, of this or that t-maker sending $E$ your way. Accounts of which $\varepsilon$s were probably implicated are *backstories* $\text{BK}_i$.

# Backstories

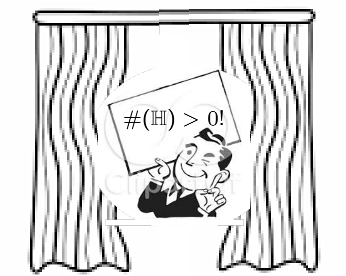
Lefty and Righty are flipped side by side. One is Al's silver dollar and the other Betty's loonie. *Something* is seen that makes clear that $A \vee B$ but not whether $A$ ($B$).



What? Three scenarios will be sketched. In the first and Bayes-friendliest, the curtains draw shut before either coin lands. A sign assures us that not both landed tails. Lefty is *seen* landing heads in the second; Righty's still hidden. The last has two sightings, of $\text{Lefty}_1$ and then $\text{Lefty}_2$. They may or may not be the same coin. Both are heads-up.

# UNSEEN (= $BK_0$), the Bayesian pick

Lefty and Righty land behind their respective curtains.
$A \vee B$ is conveyed by testimony; we read the sign.



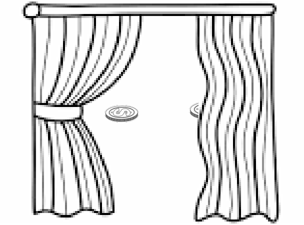

This supports $UP_0$: 2/3, 2/3, 1/3. Why? $p(\alpha\beta) = p(\alpha\overline{\beta}) = p(\overline{\alpha}\beta)$ and they sum to 1, so each is 1/3. $p(\alpha) = p(\alpha\beta) + p(\alpha\overline{\beta}) = 1/3+1/3 = 2/3$; similarly for $p(\beta)$. All as B-CON predicted. E.g., $p_n(\alpha) = p_o(\alpha)/p_o(A \vee B) = 1/2 \div 3/4 = 2/3$.

# ONESEEN (=$\text{BK}_1$), and a famous fallacy

$A \vee B$: a coin landed heads. $A \wedge B$ iff the other one did too. The probability of the other coin landing heads is $\frac{1}{2}$. So $p(\alpha\beta) = \frac{1}{2}$, not $\frac{1}{3}$.[2]

Gardner originally thought this fallacious. But it needn't be, he eventually decided. $A \vee B$ could for example have been learned like so.

Lefty landed heads. Lefty is as likely Al's dollar as Betty's loonie. Righty we don't know. The curtain was pulled before it landed.

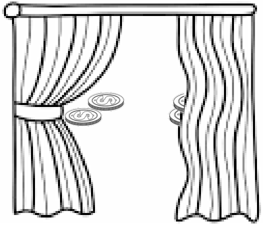


Lefty-heads=$\lambda$, Righty-heads=$\rho$. $p(\lambda\rho)$=$p(\lambda)\times p(\rho)$=$1\times\frac{1}{2}$. $p(\alpha\beta\equiv\lambda\rho)$=1, $\therefore$ $p(\alpha\beta)$=$\frac{1}{2}$. $p(\alpha)$= $p(\textit{saw-}\alpha)$ + $p(\textit{saw-}\beta)\times p(\alpha|\textit{saw-}\beta)$ = $\frac{1}{2} + \frac{1}{2}\times\frac{1}{2} = \frac{3}{4}$. $p(\beta)$ too.
Backstory $\text{BK}_1$ thus generates $\text{UP}_1$: $\frac{3}{4}, \frac{3}{4}, \frac{1}{2}$.

[2] Gardner [1959]

# TWOSEEN ($=$BK$_2$). and higher probabilities

Two sightings are made on the left, both of heads.
Was that (i) $\alpha$ twice, (ii) $\beta$ twice, or (iii) $\alpha$ and $\beta$, in either order?



p(*see*-$\alpha\beta$) = p((iii)); p(*see*-$\alpha$) = p($\neg$(ii)); p(*see*-$\beta$) = p($\neg$(i)). These fix the probabilities of $\alpha$, $\beta$, and $\alpha\beta$. Writing $\mathcal{S}[\theta]$ for seeing $\theta$,

1. $p(\alpha)=p(\mathcal{S}[\alpha])+p(\alpha|\bar{\mathcal{S}}[\alpha])\times p(\bar{\mathcal{S}}[\alpha])= (p(\mathcal{S}[\alpha])+1)/2$
2. $p(\beta) = (p(\mathcal{S}[\beta])+1)/2$
3. $p(\alpha\beta) = p(\alpha)+p(\beta)\text{-}p(\alpha\vee\beta) = (p(\mathcal{S}[\alpha])+p(\mathcal{S}[\beta]))/2$

This yields UP$_2$ $(\frac{7}{8},\frac{7}{8},\frac{3}{4})$ if $p(\mathcal{S}[\alpha])=p(\mathcal{S}[\beta])=\frac{3}{4}$, $p(\mathcal{S}[\alpha\beta])=\frac{1}{2}$, and UP$_3$ $(\frac{9}{10},\frac{9}{10},\frac{4}{5})$ if they're $\frac{4}{5}$ and $\frac{3}{5}$. Almost any underpinnings can be obtained by such means.

# Bayes at bat

| | $\mathcal{S}[\alpha]$ | $\mathcal{S}[\beta]$ | $\mathcal{S}[\alpha\beta]$ | | $\alpha\vee\beta$ | $\alpha$ | $\beta$ | $\alpha\beta$ | |
|---|---|---|---|---|---|---|---|---|---|
| before | N/A | N/A | N/A | | 3/4 | 1/2 | 1/2 | 1/4 | |
| after | $\downarrow$ | $\downarrow$ | $\downarrow$ | | 1 | $\downarrow$ | $\downarrow$ | $\downarrow$ | |
| $\text{BK}_0$ | 0 | 0 | 0 | $\Rightarrow$ | 1 | 2/3 | 2/3 | 1/3 | $\text{UP}_0$ |
| $\text{BK}_1$ | 1/2 | 1/2 | 0 | $\Rightarrow$ | 1 | 3/4 | 3/4 | 1/2 | $\text{UP}_1$ |
| $\text{BK}_2$ | 3/4 | 3/4 | 1/2 | $\Rightarrow$ | 1 | 7/8 | 7/8 | 3/4 | $\text{UP}_2$ |
| $\text{BK}_3$ | 4/5 | 4/5 | 3/5 | $\Rightarrow$ | 1 | 9/10 | 9/10 | 4/5 | $\text{UP}_3$ |

Bayes' rule to go by the chart is batting .250. Adding further scenarios would push the average lower, as low as you like. How is this not a reductio?

Reply: Rules are tools; they come with a manual telling you how to use them. FOR INERT *E*s ONLY, it says. B-CON is *on the cases for which it's intended* batting 1.000.[3]

| | |
|---|---|
| Trivial | "Inert" just *means* B-CON-satisfying. (I bat 1.000 on my hits.) |
| Circular | One needs $p_n(\varepsilon)$ to test for inertness,..to use B-CON,..to find $p_n(\varepsilon)$. |
| Stultifying | Inert *E*s are *special*. You learn *E* by engaging with its subject matter. |

"Doctor, it hurts when I go like this."
"So, don't go like that!"

"Reverend, your rule leads me sometimes astray."
"So, use it the other times!"

[3] Not really, as *other E*-impliers might appreciate too quickly/slowly. No harm in being concessive.

# Active learning (1): Uncertainty

A rule for active evidence would be nice. Three hurdles to clear. (1) We're used to updating on certainties. but $p(\epsilon^i)<1$. If we think of $E$ as *ambiguous* between the $\epsilon^i$s, then Jeffrey has this covered (Jeffrey [1990], 165):

> *Observation by candlelight*: The agent inspects a piece of cloth by candlelight, and gets the impression that it is green, although... it might be blue or even violet...originally his degrees of belief in $G$, $B$, and $V$ were [thus and such], his credences afterward are .70, .25, and .05.

If $X$ is the disjunction of incompatible alternatives $X^i$, $H$'s new probabiliity should be a weighted average of its old ones conditional on the $X^i$s:

$$\text{J-CON} \qquad p_n(H) = \Sigma_i \; p_o(H|X^i)\times p_n(X^i) \tag{5}$$

This might look familiar. It's the *expectation* of $p_o(H|X^i)$ considered as a random variable. Let's quickly review how that works.

A random variable V is a function from partition-cells $Z^i$ to reals: `the total mass of the stars`, for instance. $\text{EXP}[\mathtt{V}] = \Sigma_i \; \mathtt{V}(Z^i)\times p(Z^i)$, where V has a fixed value in each cell. $H$'s new probability for Jeffrey is precisely this is the case where $Z^i = p_o(H|X^i)$.

$X$ can hold various ways: $X^1$-ly, $X^2$-ly, etc. If we knew which obtained, we'd condition on it: $p_o(H|X^{\checkmark})$. As it is we have to guesstimate, taking the expectation of $H$`'s-probability-given-whichever-`$X^i$`-is-correct`.

# Active learning (2): Partitionality

That solves one problem, two to go. $E$'s t-makers $\varepsilon$ rarely form a partition. Take $A \vee B$. Its t-makers are $\alpha$, $\beta$, and $\alpha\beta$. Any two of these overlap.

The issue with non-partitions is that overlap worlds will be double counted, pushing $p_n(A \vee B)$ over 1. Jeffrey hints at the possibility of generalizing J-CON to allow for overlap. There are definitely things one could try, e.g., "normalizing."

Another option is to partition on *which* truthmakers obtain. Worlds are equivalent where the **the truth in E** is concerned iff $E$ is true in the same way(s) in them.[4]

$$p_n(H) = \Sigma_k \; p_o(H|E^k) \times p_n(E^k) \tag{6}$$

$E^k$ ranging over $E$-consistent state-descriptions $\pm\epsilon^1 \wedge \pm\epsilon^2 \wedge \ldots$. No need to pursue this, since the $\varepsilon^i$s are unsuitable for *another* reason, and our response to that other reason will get us back a partition.

[4] The cells when $E{=}A \vee B$ would be $\alpha\beta$, $\alpha\bar{\beta}$, and $\bar{\alpha}\beta$.

# Active learning (3): Availability

Cell probs have to be available pre-update. T-maker probabilities often aren't. $\alpha$ isn't *seen* to be 3/4 probable. The seeing $\mathcal{S}$ has prob 1/2 of being of $\alpha$, and so on.

If p($\alpha$) remains to be determined, condition on its determinants: p($\mathcal{S}[\alpha]$), etc. Or, since many $\varepsilon^i$s may be presented, on the prob p($\mathcal{O}[\![\theta]\!]$) that $\theta$ is *all* that's presented.

Total meanings $\theta$ have the advantage over inter alia meanings $\varepsilon$ that $\mathcal{O}[\![\theta^1]\!]$, $\mathcal{O}[\![\theta^2]\!]$,... form a partition. Inter alia probabilities are recoverable so it's a win/win:

$$\text{p}(\mathcal{O}[\varepsilon]) = \Sigma\ \{\text{p}(\mathcal{O}[\![\theta^k]\!]):\ \theta^k \vDash \varepsilon\} \quad (7)$$

The prob $\mathcal{O}$ means $\alpha$ = the summed probs of "$\alpha$+.." being all it means.

What is *E*'s evidential import?

↓

What are the probabilistic underpinnings p($\varepsilon^i$) of *E*'s rise to 1?

↓

How was *E* learned? What are the probabilities p($\mathcal{O}[\varepsilon^i]$) of $\mathcal{O}$ meaning $\epsilon^i$ inter alia?

↓

How shall probability be distributed over hypotheses $\mathcal{O}[\![\theta^k]\!]$ about $\mathcal{O}$'s total meaning?

The distribution over total disambiguations is what we bring to the update table.

# Amending the rule

**The truth in E** was a function from $E$-worlds $w$ to (the conjunction $\theta$ of) $E$'s obtaining t-makers in $w$ ($w \mapsto \theta$). $H$'s expected probability conditional on **the truth in E** would be good to know. But it depends alas on $p_n(\theta^1)$ etc which remain to be determined.

**The truth in $\mathcal{O}$** substitutes *presented* t-makers ($w \mapsto \mathcal{O}[\![\theta]\!]$).[5] $p(\mathcal{O}[\![\theta]\!])$ is available as it concerns what's in view. ($p(\alpha) = p(\mathcal{O}[\alpha])+$the prob of $\alpha$ obtaining offstage.)

Available probabilities are ones we can update on. What credence in $H$ is licensed by an event of $E$-learning? Bayes: $H$'s probability conditional on $E$. Jeffrey: $H$'s prob conditional on **the truth in E**. Us: $H$'s prob conditional on **the truth in $\mathcal{O}$**:

$$\text{A-CON} \quad p_n(H) = \Sigma_\theta \; p(H|\mathcal{O}[\![\theta]\!]) \times p_n(\mathcal{O}[\![\theta]\!]) \qquad (8)$$

What if $\mathcal{O}$ brings word of *no* particular t-makers ($p_n(\mathcal{O}[\![\theta]\!])=0$)? Then A-CON goes trivial. $\theta$ *is extended therefore to dt-makers* = t-makers closed under disjunction. $E$'s truth-condition $\vee_i \, \epsilon^i$ is certainly presented, or $\mathcal{O}$ would not be a vehicle for $E$-learning.

B-CON sees $\vee_i \, \epsilon^i$ as $\mathcal{O}$'s total meaning, and sometimes it is. The sign "says" $\alpha \vee \beta$, full stop; $p(\mathcal{O}[\alpha]$ etc are 0.) A-CON agrees *in that case* with B-CON: $p(H|\mathcal{O}[\![\vee_i \epsilon^i]\!]) \approx p(H|E)$. Often though $p(\mathcal{O}[\![\theta]\!]) > 0$ for stronger $\theta$s. And then things get interesting.

---

[5] Presented $\theta$s do obtain, for presentation is factive; $\mathcal{O}[\varepsilon] \Rightarrow \varepsilon$.

# E1: Implicature

$A \vee AB$ is good news, intuitively, for $B$. Why, when $A$ is not good news and **A=A∨AB**?

Usual answer: Context pragmatically strengthens $A \vee AB$ to include implicatures.
Our answer: It always supports $B$, *unless* context forces weakest possible reading.

The issue is $B$'s expected probability conditional on $\mathcal{O}$**'s-true-meaning**, aka **the truth in** $\mathcal{O}$. $A$ has just the one dt-maker $\alpha$; that's the only thing $\mathcal{O}_A$ can mean; so p($B \| A$) =

$$\underbrace{p(B|\mathcal{O}[\![\alpha]\!])}_{1/2} \times \underbrace{p(\mathcal{O}[\![\alpha]\!])}_{1} = 1/2$$

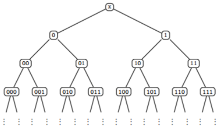
$\mathcal{O}_{A \vee AB}$ can mean either $\alpha$ (with probability $k$) or $\alpha\beta$ (1-$k$). p($B \| A \vee AB$) =

$$\underbrace{p(B|\mathcal{O}[\![\alpha]\!]) \times p(\mathcal{O}[\![\alpha]\!])}_{k/2} + \underbrace{p(B|\mathcal{O}[\![\alpha\beta]\!]) \times p(\mathcal{O}[\![\alpha\beta]\!])}_{1-k} = 1 \text{ - } \frac{k}{2}$$

This exceeds 1/2 provided that $k$<1, i.e., p($\mathcal{O}[\alpha\beta]$> 0. $A \vee AB$ fails to support $B$ only if sent by an $\mathcal{O}$ that definitely does *not* present $\alpha\beta$. That can happen (we'll see).

But it takes work to arrange. $A \vee AB$'s just by its content raises for A-CON the question of what p($\mathcal{O}[\alpha\beta]$) should be. $B$-favoring answers hugely outnumber $B$-neutral ($2^{\aleph_0}$:1).

# E2: A paradox




$H$ = *My coin ($c_0$, aka c) landed heads*

$D$ = *All but finitely many of $c_1$, $c_2$, $c_3$,... landed heads*

$E$ = *All but finitely many of c, $c_1$, $c_2$,... landed heads*

*E seems* like better news for $H$ than $D$.[6] The arguments pointing that way must be fallacious, it's thought, since p($H|D$)=p($H|E$).[7] p($H|D$) is $\frac{1}{2}$, so p($H|E$) must despite appearances be the same.

Claim: p($H\|E$) $>$ 1/2.

Let $S^j$ range over cofinite subsets of $\{c_0, c_1, ...\}$. $E$'s truthmakers (t-makers, dt-makers) are of the form: $\hbar(S^j)$ (all $S^j$s are heads). $p(H\|E) = \text{EXP}[p(H|\mathcal{O}[\![\hbar(S^j)]\!]) = \text{EXP}[c \epsilon S^j] + \text{EXP}[c \notin S^j]/2$. So $p(H\|E) = \frac{m+1}{2}$, where $m = \text{EXP}[c \epsilon S^j]$,

That's at least 1/2, more if $m$>0, that is, c is *not certainly* among the finitely many exceptions. So p($H\|E$)>1/2. How much greater? p($H\|E$) $\geqslant$ 3/4, if c is as likely to be among the cofinite multitudes as the finite few.

6 This is Builes' paradox (Builes [2020]).

7 Given that p($E|D$)=1 and vice versa (Dorr et al. [2021]).

# E3: Induction

The straight rule STR: $\mathrm{p}(Fb|\wedge_i Fa_i) > \mathrm{p}(Fb)$.[8] It commits us supposedly, when the predicate is $Rx =_{Df} Gx \& Ox \vee \bar{G}x \& \bar{O}x$, to expecting $b$ to be non-$G$ on learning $\wedge_i Ra_i$.

That can't be, for STR is silent on expectations; it contains no $\mathrm{p}_n(\ldots)$. STR+B-CON may commit us. But, B-CON is for $E$s whose t-makers get $1/\mathrm{p}_o(E)$ times likelier. And gruesome $E$s' don't; $\mathrm{p}(\gamma_i\omega_i)$ rises while $\mathrm{p}(\bar{\gamma}_i\bar{\omega}_i)$ falls. Apply A-CON then.

$$\mathrm{p}(Rb||Ra) = \mathrm{p}(Rb|\mathcal{O}[\![\gamma\omega]\!]) \times \underbrace{p(\mathcal{O}[\![\omega]\!])}_{1} + \mathrm{p}(Rb|\mathcal{O}[\![\bar{\gamma}\bar{\omega}]\!]) \times \underbrace{p(\mathcal{O}[\![\bar{\gamma}\bar{\omega}]\!])}_{0}$$

$$= \mathrm{p}(Rb|\mathcal{O}[\![\gamma\omega]\!]) = \mathrm{p}(\neg Gb||\mathcal{O}[\![\gamma\omega]\!]) \approx \mathrm{p}(\neg Gb||\gamma\omega) \approx \mathrm{p}(\neg Gb|Ga).$$

STR puts $\mathrm{p}(\neg Gb|Ga)$ *below* $\mathrm{p}(\neg Gb)$ (by putting $\mathrm{p}(Gb|Ga)$ above $\mathrm{p}(Gb)$). A straight-rulers' credence in $\neg Gb$ will *drop* when she learns $Ra$.

Goodman: STR generates gruesome expectations; $\mathrm{p}_n(\neg Gb) > \mathrm{p}_o(\neg Gb)$.
From our perspective: STR generates

1. no expectations until *a is grue* is fed into an update rule,
2. *b isn't green*-expectations when fed (as it shouldn't be) into B-CON,[9]
3. *b is green*-expectations when plugged (as it should be) into A-CON

[8] $b$ is understood to have been set aside for later observation ($\mathrm{p}(Ob)=0$).

[9] Proof that it shouldn't be: known falsehoods like *a was neither observed nor green* gain probability too.

# O1: Inference

"*Almost all heads* is said to be better evidence than *Almost all heads, c aside* for *c landed heads*. But if I know the one, I can infer the other, giving myself thereby the evidence I formerly lacked."

You're forgetting that evidential powers depend on how a thing is learned.

You see Al's coin land heads; that's $\mathcal{O}_A$. You infer $A \vee AB$; that's $\mathcal{O}_{A \vee AB}$. If $\mathcal{O}_A$ did not put $\beta$ before you — $p(\mathcal{O}[\beta])=0$ — then $\mathcal{O}_{A \vee AB}$, doesn't either; it gets its meaning from $\mathcal{O}_A$. That's an extreme case. $A \vee AB$ come by honestly does support $B$.

$\mathcal{O}_D$ is assumed to present only dt-makers for $D$: $\delta$s. These concern later coins and are are neutral on whether c lands heads. $\mathcal{O}_E$ = inferring-$E$-from-$D$ gets its meaning from $\mathcal{O}_D$. $E$ though normally good news for *c landed heads* isn't when thus obtained.[10]

Learning $E$ has by default a chance of presenting any $\varepsilon$....a presumption defeated if it piggybacks on $\mathcal{O}_D$. $p(\mathcal{O}_E[\hbar(\text{all})])=p(\mathcal{O}_D[\hbar(\text{all})])= 0$ since $\hbar(\text{all})$ is not a $\delta$.

[10] $A$ vs $A \vee AB$ is like $D$ vs $E$ insofar as $E$ is exactly equivalent to $D \vee DE$.

# O2: Remote learning

What is it for $\mathcal{O}_E$ to "present" something? This has a stipulative aspect; only $\varepsilon$s are presented. Not all of them, though, presumably, indeed some may be incompatible. Which $\varepsilon$s are presented by a particular $\mathcal{O}$? I have only a picture/model to offer.

Learning *E* is like glimpsing its dt-makers through a glass darkly (Kratzer [1989]). $p(\mathcal{O}[\varepsilon]) > 0$ if one could say later: "Ah, so that was you the whole time." $p(\mathcal{O}[\varepsilon])$ is higher if one could add, "As I suspected," lower if "Who'd have thunk?"

The model may seem to depend on our seeing, or sensing, that *E*.[11] We might instead read it or be told that *E*. Language has its own ways of conveying source information.

There may be t-maker differences, as with $A$ vs $A \vee AB$. $\neg\mathcal{O}_A[\beta]$ just because $\beta$ is not an $\alpha$. Even where t-makers agree, *S* can be instructive about which were on display:

- E Not: no coins came up heads .... $p(\mathcal{O}[\![\alpha\vee\beta]\!])=1$, $p(\mathcal{O}[\![\alpha]\!])=p(\mathcal{O}[\![\beta]\!])=p(\mathcal{O}[\![\alpha\beta]\!])=0$
- F A certain coin came up heads.... $p(\mathcal{O}[\![\alpha\vee\beta]\!])=0$, $p(\mathcal{O}[\![\alpha]\!])=1/2$, $p(\mathcal{O}[\![\beta]\!])=1/2$, $p(\mathcal{O}[\![\alpha\beta]\!])=0$
- G A certain # of coins came up heads .... $p(\mathcal{O}[\![\alpha\vee\beta]\!])=0$, $p(\mathcal{O}[\![\alpha]\!])=p(\mathcal{O}[\![\beta]\!])=1/4$, $p(\mathcal{O}[\![\alpha\beta]\!])=1/2$

One needs to look here at the semantics literature on evidential meaning and specificity. *Durian must be smelly* suggests indirect access. *Durian smells yuck* suggests direct (the "acquaintance inference"). *A certain F is G* suggests $\overline{\exists x}$ ($x$ is $F$ and this message is sourced in *its* being G).[12] "Ah, so *you* are the rumored someone."

---

[11] Dretske [1969], Dretske [1979], Barwise [1981]; I thought I saw Al drowning, but it was Betty swimming.

[12] Hintikka [1986], Farkas [2002], Goodman and Lassiter [2015].

# O3: Overgeneration

"A certain coin landed heads" presents itself as sent our way by a *particular* coin. The speaker may not know which, or their informant either, back through the centuries. Still *one* landed heads, and that's enough to push p($AB$) to 1/2.

Can't this *always* be done? Headsy $=_{Df}$ *the* heads, if it's unique, otherwise *a* heads picked at random. p(Othersy-heads)=1/2, right? So p($AB$)=1/2 regardless!

That's the fallacy Gardner was warning against. p(*Othersy-heads*) is NOT 1/2 given how it was chosen. It's $0 \times$p(forced-choice) + $1 \times$p(unforced) = 1/3. So p($AB$)=1/3.

Now, Othersy may be Righty. "They" differ probabilistically because of how they came by those names. *Righty*-ness is independent of how it landed. *Other*-ness is achieved in most cases by landing tails.

"A certain $F$" seems at first to be one we're acquainted with, and have in mind. But (i) we may be clueless about it, & (ii) why would in-mindness be prob relevant?

Have we got things backwards perhaps? Rather than acquaintance driving the probabilities, $n$ counts as an object of our acquaintance if its status as $n$ is independent of the question under discussion, e.g., how it landed.[13]

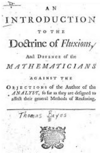
[13] "What is $357 \times 9792$?" "Easy, 10. (I answered in base $357 \times 9792$.)"

# Outro

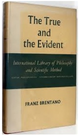
Evidence acquired by engaging with the state of things **e**-wise is *organically* conceived. Bayes' rule seems better adapted to immaculately conceived $E$s, ones that appear fully formed in your belief box. Immaculate conception is not an everyday thing.

Jeffrey's rule is organic. But he partitions on external-world hypotheses $E^i$. You can't always observe your way to a distribution over $E^i$s opinionated enough to form a partition. p($\alpha\beta$) isn't seen but inferred from the prob $\alpha$, $\beta$, $\alpha\beta$ are seen ($\frac{1}{2}$,$\frac{1}{2}$,0).

Observation does enable assessments of what our eyes are probably telling us, or $\mathcal{O}$ is probably tracking. Update on those assessments, then. That's what A-CON does.




*So Bayes-style formal epistemology*
*is still basically OK, provided this small adjustment*
*is made, that seems (sometimes) called for anyway?*




*Absolutely! ......Though,*
*"What is at first small is often large in the end."*

Jon Barwise. Scenes and other situations. *The journal of Philosophy*, 78(7):369–397, 1981.

David Builes. A paradox of evidential equivalence. *Mind*, 129(513):113–127, 2020.

Donald Davidson. Emeroses by other names. *The Journal of Philosophy*, 63(24):778–780, 1966.

Cian Dorr, John Hawthorne, and Yoaav Isaacs. Solving a paradox of evidential equivalence. *Mind*, 130(520): 1159–1182, 2021.

Fred I. Dretske. *Seeing and Knowing*. University Of Chicago Press, Chicago, 1969.

Fred I Dretske. Simple seeing. In *Body, mind, and method: Essays in honor of Virgil C. Aldrich*, pages 1–15. Springer, 1979.

Donka F Farkas. Specificity distinctions. *Journal of semantics*, 19(3):213–243, 2002.

Martin Gardner. The two children problem, "Mathematical Games". *Scientific American*, 1959.

Noah D Goodman and Daniel Lassiter. Probabilistic semantics and pragmatics: Uncertainty in language and thought. *The handbook of contemporary semantic theory*, pages 655–686, 2015.

Jaakko Hintikka. The semantics of a certain. *Linguistic Inquiry*, 17(2):331–336, 1986.

Richard C Jeffrey. *The logic of decision*. University of Chicago Press, 1990.

Angelika Kratzer. An investigation of the lumps of thought. *Linguistics and Philosophy*, 12:607–653, October 1989.