

The Value of Information, Under Ambiguity

Kevin Dorst

24.805, Fall 2025

I. Das 2023

Value of Information: When evidence is free and you know you'll respond to it rationally, you should prefer to gather evidence before making any decision.

Evidence Externalism: We can sometimes have conclusive evidence about (some bits of) the external world.

Conditionalization: When e is the strongest thing you learn between two times, you should update by conditioning your prior on e .

$$P^+(\cdot) = P(\cdot|e).$$

Das claims that if Evidence Externalism and Conditionalization are true, then the Value of Information is false.

→ It's false because sometimes rational agents can do **biased inquiries**—inquiries that they are *certain* will raise their credence in q .

Or are certain will lower their credence in q , i.e. raise credence in $\neg q$.

Why would biased inquiries, so defined, mean that VoI is false?

Suppose currently $P_a(q) = 0.5$.

Suppose you are *certain* that after you get some evidence, your credence in q will be at least 0.6: $P_a(P^+(q) \geq 0.6) = 1$.

Then you can face a decision problem where you'd prefer not to get the evidence:

$$B_0 = \$0 \text{ for sure} \quad B_1 = \begin{cases} +\$40 & \text{if } q \\ -\$60 & \text{if } \neg q \end{cases}$$

Current expected values:

$$\mathbb{E}_{P_a}(B_1) = 0.5(40) + 0.5(-50) = 20 - 25 = -5 < 0 = \mathbb{E}_{P_a}(B_0)$$

But you're sure that your *future* expected value for B_1 will be positive.

$$\mathbb{E}_{P^+}(B_1) \geq 0.6(40) + 0.4(-50) = 24 - 20 = +4$$

So you're sure that, if you get the evidence, your future self will take the bet.

⇒ So, as far as this decision problem goes, the choice to gather the evidence before deciding is equivalent to the choice to take B_1 . But that is worse than doing B_0 !

In fact, you should be willing to pay up to \$4 to avoid the evidence

This generalizes to any 'sure-win' investigation: if you are certain that your credence in a given proposition will increase, then the Value of Information fails.

(Notice: sure-win investigations imply Martingale failures. The con-

verse isn't quite true; but something near enough is: Martingale failures can be leveraged into failures of Vol)

Why would you be an evidence externalist?

Skepticism and good vs. bad cases.

→ Knowledge is factive. So unless skepticism is true, knowledge can violate **negative introspection**

If evidence is knowledge ($E=K$), then we can have failures of negative introspection on evidence.

- Don't we need to be able to know what our evidence is?
- Williamson and company will argue that there is *no* nontrivial state—including whether your evidence entails p —such that we are always in a position to know whether you're in that state.

Even if $E \neq K$, the same reasoning implies that either (1) our evidence can never *entail* (conclusively establish) anything about the external world, or (2) evidence can violate factivity¹, or (3) evidence can violate negative introspection.

Das assumes $\neg(1)$, so uses (2)-or-(3) as premise to argue against Vol.

How can violations of negative introspection lead to bias?

Single good/bad case, condition on your evidence.

→ Violates Martingale, but satisfies Vol.

But, Das says, if you can have a single case, you can have *two cases*, simultaneously: red, sandalwood wall.

→ Priors = 0.1, 0.1, 0.8. Posterior in rs either $\frac{8}{9} \approx 0.89$, or 1.

Possible aside: Analogy to Roger's case and Sleeping Beauty.

Reject conditionalization?

- Maybe, if your evidence is non-partitional (violating negative introspection) like this, you shouldn't condition on your evidence.
- Instead, should 'meta-conditionalize': condition on *which bit of evidence you received*.

If ro , evidence is $\{ro, rs\}$.

If ws , evidence is $\{ws, rs\}$.

If rs , evidence is $\{rs\}$.

→ So if update on what evidence you received, can figure out where you are!

- This is obviously right for introspective cases.
- Has mad consequences for non-introspective cases: handless BIV can be certain it doesn't have hands!

I.e. the principle that $\neg Kp \rightarrow K\neg Kp$.

Failures of $\neg Ep \rightarrow E\neg Ep$, where Ep is read as 'your evidence entails that p '

Eg what is the strongest n such that your evidence entails that the pole is at least n inches high? Your evidence doesn't settle that, he says.

¹ Violate $Ep \rightarrow p$

'Epistemic collider'

Schoenfield 2017; Hild 1998

If H , I'll tell you; if T , I'll say nothing

But Schoenfield-heads will say that's a consequence of mad assumptions about (non-introspective) evidence

II. Kevin

Even if we assume evidence is clear/partitional, if you're unsure what your priors are, similar reasoning will lead to violations of the Vol.

→ There can be a series of independent questions Q_1, \dots, Q_n where your priors and updates are (Ambiguous but) Valuable *with respect to each question*, yet your priors about their product $Q_1 \sqcap \dots \sqcap Q_n$ is not.

Take our toy model of the word search:

- Either you think you're good ($P = \pi_g$) or bad ($P = \pi_b$).
- Either way, c is 50%-likely.
- What you think is ambiguous:
If you think you're good, you're $2/3$ -confident of that.
If you think you're bad, you're $2/3$ -confident of that.
- Suppose you'll *find* iff *both* completable *and* you think you're good.
- What happens?

$$P \approx \begin{pmatrix} & g_c & g_{\bar{c}} & b_c & b_{\bar{c}} \\ g_c & .33 & .33 & .17 & .17 \\ g_{\bar{c}} & .33 & .33 & .17 & .17 \\ b_c & .17 & .17 & .33 & .33 \\ b_{\bar{c}} & .17 & .17 & .33 & .33 \end{pmatrix} \quad P^+ = \begin{pmatrix} & g_c & g_{\bar{c}} & b_c & b_{\bar{c}} \\ g_c & 1 & 0 & 0 & 0 \\ g_{\bar{c}} & 0 & .50 & .25 & .25 \\ b_c & 0 & .20 & .40 & .40 \\ b_{\bar{c}} & 0 & .20 & .40 & .40 \end{pmatrix}$$

This prior P and update (P, P^+) satisfies the Value of Information.

- The prior P trusts itself, in the sense that for any decision problem or accuracy metric, it expects using P to make it's decision leads to a better outcome than picking a constant distribution π .
- The prior P trusts P^+ , in the same sense—it's happy to outsource its decision to P^+ .

Generally: *partitional conditioning preserves Value.* If P is valuable and P^+ comes by conditioning P on a partition, then P values P^+ .

But recall that if $a = g_c$, then $\mathbb{E}_{P_a}(P^+(c)) = 0.55$, while $P_a(c) = 0.5$.

That's a Martingale failure, but not 'bias' in Das's sense—your credence in c isn't guaranteed to rise. But we can leverage it.

Suppose we have 1000 independent copies of the word search.

Let X be the number of searches that are completable, determined by a (fair) coin flip. So your initial estimate is $\mathbb{E}_{P_a}(X) = 500$.

Your initial estimate for your final probability in each c_i is 0.55: $\mathbb{E}_{P_a}(P^+(c_i)) = 0.55$.

Since $X = \sum_i \mathbb{1}_{c_i}$, it follows that your initial estimate for your final estimate of X is 550:

And, as a result, conditioning on facts about each Q_i can lead to (something very close to) bias in Das's sense

Suppose, in fact, $P = \pi_g$.

$$\pi_i(c|P = \pi_g) = 0.5 = \pi_i(c|P = \pi_b)$$

$$\pi_g(P = \pi_g) = 2/3, \text{ so } \pi_g(P = \pi_b) = 1/3$$

$$\pi_b(P = \pi_b) = 2/3, \text{ so } \pi_b(P = \pi_g) = 1/3$$

$$\text{So } f = \{g_c\}, \text{ and } \neg f = \{g_{\bar{c}}, b_c, b_{\bar{c}}\}$$

For example, the (global, singleton) Brier score of P is 0.1528 at all worlds, while of P is $(0, .094, .14, .14)$ —i.e. P^+ accuracy-dominates P on the Brier score.

In fact, it's not even *likely* to— $\frac{2}{3}$ of the time it'll drop.

Perhaps make there be 1000 different types of search—one is a math problem, one is a spot-the-difference, one is a logic puzzle, etc.—so it's plausible that how good you (think you) are is uncorrelated across searches

$$\mathbb{E}_{P_a}(\mathbb{E}_{P^+}(X)) = \mathbb{E}_{P_a}(\mathbb{E}_{P^+}(\sum_i \mathbb{1}_{c_i})) = \sum_i \mathbb{E}_{P_a}(\mathbb{E}_{P^+}(\mathbb{1}_{c_i})) = \sum_i \mathbb{E}_{P_a}(P^+(c_i)) = 1000(0.55)$$

More importantly, you're *very confident* that your estimate will go up.

You are confident that your posterior credence will likely be:

- 1 in about 333 of the c_i .
- 0.25 in about 333 of the c_i ; and
- 0.40 in about 333 of the c_i .

In order for your estimate to stay at 50, it would need to be the case that you are 1 in around 333 of the c_i and 0.25 in around 666 of them.

→ You're all-but-certain that won't happen.

So $\mathbb{E}_{P_a}(X) = 500$ but $P_a(\mathbb{E}_{P^+}(X) > 510) \approx 1$.

Suppose I offer the following bets:

$$B_0 = \$0 \text{ for sure} \quad B_1 = \$(X - 510)$$

Then B_1 is not worth the risk according to your prior: it's expected value is -10 .

But you are all-but-certain that your future self will take B_1 .

What is happening?

A form of the *lottery paradox* for Bayesians (under ambiguity).

Lottery paradox (for outright belief): rational beliefs about each of a series of questions, combined rationally, lead to irrational beliefs.

Paradox for Bayesians:

- Suppose, under ambiguity, your probabilistic beliefs about Q_i are rational iff they are Valuable.
- In the above case, your prior opinions about each Q_i are Valuable—hence rational.
- You are rationally certain that the Q_i are independent.
- But your prior opinions about $Q_1 \sqcap \dots \sqcap Q_n$ are *not* valuable

Reactions:

- 1) *Irrationality filters down:* Your priors about the product question are irrational, so your opinions (and updates) about the individual ones are too.
- 2) *Rationality filters up:* Your priors (and updates) about the individual questions are *rational*, so your opinions about the product question are too.

References

Hild, Matthias, 1998. 'Auto-epistemology and updating'. *Philosophical Studies*, 92(3):321–361.
Schoenfield, Miriam, 2017. 'Conditionalization Does Not (In General) Maximize Expected Accuracy'. *Mind*, To Appear.

It's like a fair die landing $1 \vee 2 \approx 333$ times and $3 \vee 4 \approx 666$ times—and never landing $5 \vee 6$ —in 1000 tosses.

So, if your goal is to make money on this bet, you'd pay up to \$10 to prevent yourself from learning how the searches go.

Believe t_i will lose for each i . But if conjoin beliefs, come to believe t_1, \dots, t_{100} will all lose

It's a graded analog of Das's double-bad case.

That's how conditioning can lead to biased updates and your prior P not valuing your posterior P^+ .

The book takes option (2). Option (1) seems unattainable for bounded agents