Summary

This revised manuscript is improved and may possibly be accepted after carrying out *major revisions*. Given the number of comments and the severity of some of them, I can see an argument for *rejecting* this manuscript; see below for details.

Major Comments

The major problem I have with this problem is that there are a large number of things in it which are not quite right – to my eyes.

Minor Comments

Page 2: "the updating scenario *permits* it"?
"if there is no further information, mutually disjoint and jointly exhaustive events are equiprobable" – that's too simplistic. Surely, for $\Omega = \{\omega_1, \omega_2, \omega_3\}$ the PoI does not recommend that $P(\omega_1) = P(\omega_2 \cup \omega_3)$.
"report" ?
Some of the c_i may be zero, right? It's worth stating this explicitly.
"she has no other concerns" – why not say that L captures all the agent cares for?
Please, do use the word strict/strictly everywhere. It is too confusing for readers otherwise.
Page 4, can we not visualise non-Euclidean space?
Page 5, there seems to be a word missing in the last sentence (after the first appearance of 'Bayesian').
Page 6, cross entropy updates can cope with more complicated constraints, not just affine constraints. Convexity of the feasible region is normally required to guarantee uniqueness of the solution.
"The Log score is asymmetric, but unlike the Brier score not unique among its asymmetric peers" – these (non-)uniqueness claims require more hedging.
List A and List B: It is preferable to only have the explanation of the feature in the list. All evaluations and examples should appear outside the list.
List A, no substantiation is made for the "other scores" column. Given that his column plays no role in this manuscript, I suggest to delete it.
Page 8, the set up of the probability space is unsatisfactory. Why not simply say that "the finite set of elementary events Ω is generated by n binary variables"?

The negation symbol "¬" is normally not a superscript. When logical notation (¬) is used, then one normally uses $\wedge$ instead of $\cap$.

Page 9/10: This just means that you accept the additivity axiom. At this point [at the very latest], you need to give a precise definition of the credence functions you are working with:

$$\mathcal{C} := \{c : \mathcal{P}\Omega \to [0,\infty) \mid \sum_{\omega\in\Omega} c(\omega) > 0 \ \& \ F\cap E = \emptyset \Rightarrow c(F)+c(E) = c(F\cup E)\}??$$

" All arguments defending probabilism in this paper do not only justify probabilistic credence functions over non-probabilistic ones that obey the logical entailment relationships, but ad fortiorem also over non-probabilistic ones that disobey the logical entailment relationships." No. See my later comments on logarithmic scores.

The first paragraph of Section 2.2 is a mess. Since the set of probability functions is convex, its convex hull is equal to this set. This set is a proper subset of the set of credence functions. For a credence function c which is also a probability function, one can of course chose $p = c$. This result holds much more widely for continuous strictly proper scoring rules and credence functions which do not satisfy additivity, see [2].

[3] only considers probabilistic credences.

delete bracket on evidence.

Classical information theory only considers probabilistic credences. Their logarithmic loss function is an expectation, $\sum P(\omega)\log(Q(\omega))$. According to this expected loss, $Q(\omega) = 1$ for all $\omega \in \Omega$ is the best possible choice. Since standard information rules out such credences by stipulation, this is no internal problem for them. The relation between their logarithmic scoring rule and the logarithmic scoring rule popular in formal epistemology circles [which is strictly proper also for non-probabilistic credences] is discussed in [1]. In your equation 12, this logarithmic scoring rule is not local for non-probabilistic credences.

Add 'p.' to the Hendrickson bracket.

Page 14: add that close relatives have the same extrema.

Lemma 3.2: You should define $\leq$ for functions. Presumably, it's some point-wise dominance notion.

Lemma 3.4: Do you mean $(\nabla H)^{-1} = \nabla(H^*)$ or $(\nabla H)^{-1} = (\nabla H)^*$?

"Briar" – Brier (twice)

"Lemmas 3.4, 3.5, and 3.6 establish theorem 3.1" – how, why?

Page 22: A scoring rule cannot embrace an epistemic norm.

The paragraph on confirmation measures is under-developed and best omitted.

Last paragraph of Section 4.1: Agreed; but so what? You are really comparing apples to oranges here.
Page 25 on Joyce: No, Joyce shows no such thing. The result you refer to is proved in [2] for continuous and strictly proper scoring rules. The result of Joyce is properly stated in 5.1.
Page 27: I found the last part of the second to last paragraph opaque.
The next paragraph should be connected to the rest of the text.
Theorem 4.4: Again, this only holds for probabilistic credences. How to save some kind of locality notion for logarithmic scoring rules which are strictly proper for all credence functions is discussed in [1].
How can a scoring rule vindicate conditioning?
Page 32: Sure, but we can also embed into a different space (or even manifold) with a different [e.g., Riemannian or symplectic] metric.
INVARIANCE: that's very vague.
[4] does not justify probabilism.
Page 40: these are different (non-equivalent) metrics which – of course – give different distances and hence make different recommendations.
(49): $GExp_A$ not defined.
5.4.1: LP updating differs from conditionalisation. No continuity argument is required to make this point.
(58): that's not a probability function.
Page 48/49: This only applies within the standard Bayesian framework. LP are silent about how to use their probabilities for decision making. Also, in their approach a finite amount of information can change posterior probabilities from or to zero.
5.4.4: this needs to be better motivated. It does look to me like these agents use different partitions and hence it is not immediately clear why Invariance should apply here.
Page 52: Why would one use standard conditioning in a previous step when one later uses LP updating?
Surely, the proper way is to specify the total available evidence, which either pronounces on X_0 or not. Either way, it's clear how the update should be carried out.
Page 55: The first paragraph conflates all sorts of intuitions and metrics.
Figure 5: How can it be unclear, whether a point is a midpoint or not? A calculation should make things clear, no?
(70): What if $\beta > \alpha$?
C is not right: there are some $\min\{..., 0\}$ missing.
Please, do show the algebra.

Page 59: Do not use the “simplification”. There’s no point in introducing a symbol which is only used a handful of times.
(78): The reasonable thing to require might be to point out that $||b - a||$ is symmetric in a and b, while information geometry is not. Hence, collinear horizon should then only require that $D_{KL}(B, A) > D_{KL}(B, C)$.
I can’t make much sense out of the caption of Figure 6.

References

[1] Jürgen Landes. Probabilism, Entropies and Strictly Proper Scoring Rules. *International Journal of Approximate Reasoning*, 63:1–21, 2015.

[2] J.B. Predd, R. Seiringer, E.H. Lieb, D.N. Osherson, H.V. Poor, and S.R. Kulkarni. Probabilistic Coherence and Proper Scoring Rules. *IEEE Transactions on Information Theory*, 55(10):4786–4792, 2009.

[3] Leonard Jimmie Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971.

[4] John E. Shore and Rodney W. Johnson. Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy. *IEEE Transactions on Information Theory*, 26(1):26–37, Jan 1980.