Jeffrey Conditioning and the Geometry of Reason

(Author)

Abstract

Defenders of the epistemic utility approach to Bayesian epistemology use the geometry of reason to justify the foundational Bayesian tenets of probabilism and standard conditioning. The geometry of reason is the view that the underlying topology for credence functions is a metric space, on the basis of which axioms and theorems of epistemic utility for partial beliefs are formulated. It implies that Jeffrey conditioning must cede to an alternative form of conditioning. The latter fails five plausible expectations, which Jeffrey conditioning fulfills, and brings with it unacceptable results in certain cases. The solution to this problem is to reject the geometry of reason and accept information theory in its stead. Information theory comes fully equipped with an axiomatic approach which covers probabilism, standard conditioning, and Jeffrey conditioning. It is not based on an underlying topology of a metric space, but uses asymmetric divergences instead of a symmetric distance measure.

1 Introduction

The 'geometry of reason,' a term coined by Richard Pettigrew and Hannes Leitgeb, refers to a view of epistemic utility in which the underlying topology for credence functions (which may be subjective probability distributions) on a finite number of events is a metric space. Consider a 6-sided die. Probabilism assumed (we may drop this assumption later on and try to show that probabilism is justified on the basis of maximizing epistemic utility), the possible credence functions of an agent are isomorphic to the 6-dimensional simplex ($\mathbb{S}^6 \subset \mathbb{R}^6$) for which

$$p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1 \tag{1}$$

Since the isomorphism is to a metric space, there is now a distance relation between credence functions which can be used to formulate axioms relating credences to epistemic utility and justify or criticize contentious positions such as Bayesian conditionalization, the principle of indifference, other forms of conditioning, or probabilism itself (for probabilism the isomorphism is usually to $\mathbb{R}^6_{\geq 0}$, since the simplex isomorphism assumes probabilism). Before I introduce the notion of epistemic utility and some of the substantial claims in the literature that epistemic utility together with the geometry of reason give us, I want to spell out my claim that (a) given an epistemic utility approach and some intuitive axioms, the geometry of reason leads itself ad absurdum; and (b) there is a viable alternative to the geometry of reason which avoids the problematic implications: information theory. For information theory, as opposed to the geometry of reason, the underlying topology for credence functions is not a metric space. Before I argue for my claim, we need to achieve clarity on the epistemic utility approach (contrasted most usefully with a pragmatic utility approach) and look at the impressive results of applying the geometry of reason.

Epistemic utility in Bayesian epistemology has attracted some attention in the past few years. Patrick Maher provides a compelling acceptance-based account of epistemic utility (see Maher, 1993, 182–207). James Joyce, in "A Nonpragmatic Vindication of Probabilism," defends probabilism supported by partial-belief-based epistemic utility rather than the pragmatic utility we are used to in Dutch-book style arguments (see Joyce, 1998). For Joyce, norms of gradational accuracy characterize the epistemic utility approach to partial beliefs, analogous to norms of truth for full beliefs.

David Wallace and Hilary Greaves investigate epistemic utility functions along 'stability' lines and conclude that for everywhere stable utility functions standard conditioning is optimal, while only somewhere stable utility functions create problems for maximizing expected epistemic utility norms (see Greaves and Wallace, 2006; and Pettigrew, 2013). Richard Pettigrew and Hannes Leitgeb have published arguments that under certain assumptions probabilism and standard conditioning (which together give epistemology a distinct Bayesian flavour) minimize inaccuracy, thereby providing maximal epistemic utility (see Leitgeb and Pettigrew, 2010a and 2010b).

Leitgeb and Pettigrew show, given the geometry of reason and other axioms inspired by Joyce (normality, dominance), that in order to avoid epistemic dilemmas we must commit ourselves to a Brier score measure of inaccuracy and subsequently to probabilism and standard conditioning. Jeffrey conditioning (also called probability kinematics) is widely considered to be a common sense extension of standard conditioning. On Leitgeb and Pettigrew's account, it fails to provide maximal epistemic utility. Another type of conditioning, which we will call LP conditioning, takes the place of Jeffrey conditioning.

The failure of Jeffrey conditioning to minimize inaccuracy on the basis of the geometry of reason casts, by reductio, doubt on the geometry of reason. To relate probability distributions to each other geometrically, using the isomorphism between the set of probability distributions on a finite event space W with $|W| = n$ and the n-dimensional simplex $\mathbb{S}^n \subset \mathbb{R}^n$, is initially an arbitrary move. Leitgeb and Pettigrew do little to substantiate a link between the geometry of reason and epistemic utility. I will show that between Jeffrey conditioning and LP conditioning we have good reasons to favour Jeffrey conditioning.

The question then remains whether we have a plausible candidate to supplant the geometry of reason. The answer is yes: information theory provides us with a measure of closeness between probability distributions on a finite event space that has more intuitive appeal than the geometry of reason, especially with respect to epistemic utility—it is intuitively correct to relate coming-to-knowledge to exchange of information. More persuasive than intuitions, however, is the fact that information theory supports both standard conditioning (see Williams, 1980) and the extension of standard conditioning to Jeffrey conditioning (see Caticha and Giffin, 2006; and Lukits, 2015), an extension which is on the one hand commonsensical (see Wagner, 2002) and on the other hand formally continuous with the standard conditioning which Leitgeb and Pettigrew have worked so hard to vindicate nonpragmatically (see Levinstein, 2012).

There are four sections to come [note that this needs to be updated]. Section 2 articulates the geometry of reason and provides a brief overview of Leitgeb and Pettigrew's strategy to give probabilism and standard conditioning a foundation in epistemic utility. Section 3 gives a simple example where the geometry of reason and information theory give different results about the closeness of probability distributions. The geometry of reason supports LP conditioning, information theory supports Jeffrey conditioning. Section 4 provides reasons why Jeffrey conditioning is superior to LP conditioning on independent grounds, since LP conditioning violates five commonsense expectations which Jeffrey conditioning fulfills. Section 5 draws the conclusion that information theory, not the geometry of reason, reflects in formal terms what epistemic utility expresses in informal terms. Information theory notably is not a geometry of reason because its measure of closeness between probability distributions is not symmetrical. This asymmetry speaks in favour of information theory because it reflects epistemic asymmetries for which a non-geometrical approach can provide the better account.

A further conclusion of my argument is that Joyce's axioms of gradational accuracy, based as they are on the geometry of reason, need to be reformulated. Fortunately, Joyce's result still stands, vindicating probabilism on epistemic merits rather than prudential ones: partial beliefs which violate probabilism are dominated by partial beliefs which obey it, no matter what the facts are. Without the geometry of reason, however, it is necessary to modify Joyce's axioms of normality, weak convexity, and symmetry.

2 Epistemic Utility and the Geometry of Reason

There is more epistemic virtue for an agent in believing a truth rather than not believing it and in not believing a falsehood rather than believing it. Accuracy in full belief epistemology can be measured by counting four sets, believed truths and falsehoods as well as unbelieved truths and falsehoods, and somehow relating them to each other such that epistemic virtue is rewarded and epistemic vice penalized. Accuracy in partial belief epistemology must take a different shape since as a 'guess' all partial non-full beliefs are off the mark so that they need to be appreciated as 'estimates' instead. Richard Jeffrey distinguishes between guesses and estimates: a guess fails unless it is on target, whereas an estimate succeeds depending on how close it is to the target.

The gradational accuracy needed for partial belief epistemology is reminiscent of verisimilitude and its associated difficulties in the philosophy of science [references]. Both Joyce and Leitgeb/Pettigrew propose axioms for a measure of gradational accuracy for partial beliefs relying on the geometry of reason, i.e. the idea of geometrical distance between distributions of partial belief expressed in non-negative real numbers. In Joyce, the geometry of reason is adopted without much reflection. Terms such as 'midpoint' between two distributions and $\lambda b' + (1-\lambda)b''$ for distributions 'between' two distributions b' and b'' are used freely. Leitgeb and Pettigrew muse about alternative geometries, especially non-Euclidean ones. They suspect that these would be based on and in the end reducible to Euclidean geometry but they do not entertain the idea that they could drop the requirement of a metric topology.

Once a geometry of reason is in place, interesting and substantial results follow. Leitgeb and Pettigrew define two notions, local and global inaccuracy, and show that one must adopt a Brier score to measure inaccuracy in order to avoid epistemic dilemmas trying to minimize inaccuracy on both measures. To give the reader an idea what this looks like in detail and for purposes of later exposition, I want to provide some of the formal apparatus. Let W be a set of worlds and $A \subseteq W$ a proposition. Then

$$I : P(W) \times W \times \mathbb{R}_0^+ \to \mathbb{R}_0^+ \tag{2}$$

is a measure of local inaccuracy such that $I(A, w, x)$ measures the inaccuracy of the degree of credence x with respect to A at world w. Let $\mathrm{Bel}(W)$ be the set of all belief functions (what we have been calling distributions of partial belief). Then

$$G : W \times \mathrm{Bel}(W) \to \mathbb{R}_0^+ \tag{3}$$

is a measure of global inaccuracy of a belief function b at a possible world w such that $G(w, b)$ measures the inaccuracy of a belief function b at world w.

Axioms such as normality and dominance guarantee that the only legitimate measure of inaccuracy are Brier scores if one wants to avoid epistemic dilemmas where one receives conflicting advice from the local and the global measures. For local inaccuracy measures, this means that there is $\lambda \in \mathbb{R}^+$ such that

$$I(A, w, x) = \lambda \left(\chi_A(w) - x\right)^2 \tag{4}$$

where χ_A is the characteristic function of A. For global inaccuracy measures, this means that there is $\mu \in \mathbb{R}^+$ such that

$$G(w, b) = \mu \|w - b\|^2 \tag{5}$$

where w and b are represented by vectors and $\|u - v\|$ is the Euclidean distance

$$\sqrt{\sum_{i=1}^{n} (u_i - v_i)^2}. \tag{6}$$

We use (4) to define expected local inaccuracy of degree of belief x in proposition A by the lights of belief function b, with respect to local inaccuracy measure I, and over the set E of epistemically possible worlds as follows:

$$\mathrm{LExp}_b(I, A, E, x) = \sum_{w \in E} b(\{w\}) I(A, w, x) = \sum_{w \in E} b(\{w\}) \lambda \left(\chi_A(w) - x\right)^2. \tag{7}$$

We use (5) to define expected global inaccuracy of belief function b' by the lights of belief function b, with respect to global inaccuracy measure G, and over the set E of epistemically possible worlds as follows:

$$\mathrm{GExp}_b(G, E, b') = \sum_{w \in E} b(\{w\}) G(w, b') = \sum_{w \in E} b(\{w\}) \mu \|w - b\|^2. \tag{8}$$

To give a flavour of how attached the axioms are to the geometry of reason, here are Joyce's axioms called Weak Convexity and Symmetry, which he uses to justify probabilism:

> **Weak Convexity**: Let $m = (0.5b' + 0.5b'')$ be the midpoint of the line segment between b' and b''. If $I(b', \omega) = I(b'', \omega)$, then it will always be the case that $I(b', \omega) \geq I(m, \omega)$ with identity only if $b' = b''$.

> **Symmetry**: If $I(b', \omega) = I(b'', \omega)$, then for any $\lambda \in [0, 1]$ one has $I(\lambda b' + (1 - \lambda) b'', \omega) = I((1 - \lambda) b' + \lambda b''), \omega)$.

Joyce advocates for these axioms in geometrical terms, using justifications such as "the change in belief involved in going from b' to b'' has the same direction but a doubly greater magnitude than change involved in going from b' to [the midpoint] m" (see Joyce, 1998, 596). Once I have established my alternative account, I will give counterexamples where these axioms are violated. The final task will be to show how these axioms and the geometry of reason justifying them saddle us with counterintuitive results on their own terms. This will establish the alternative (information theory) as a superior alternative. Before I do this, however, I will show how the geometry of reason works in Leitgeb and Pettigrew's account, since their account more so than Joyce's will give us leverage in identifying its shortcomings.

Leitgeb and Pettigrew's work is continuous with Joyce's work, but significantly goes beyond it. Joyce wants much weaker assumptions and would be leery of expected inaccuracies (7) and (8), as they might presuppose the probabilism that Joyce wants to justify axiomatically without begging the question. Leitgeb and Pettigrew investigate not only whether probabilism and standard conditioning follow from gradational accuracy based on the geometry of reason, but also uniform distribution (their term for the claim of objective Bayesians that there is some principle of indifference for ignorance priors) and Jeffrey conditioning. They show that uniform distribution requires additional axioms which are much less plausible than the ones on the basis of which they derive probabilism and standard conditioning (see Leitgeb and Pettigrew, 2010b, 250f); and that Jeffrey conditioning does not fulfill Joyce's Norm of Gradational Accuracy (see Joyce, 1998, 579), in short that it violates the pursuit of epistemic virtue. Leitgeb and Pettigrew provide us with an alternative method of updating for Jeffrey-type updating scenarios, which I will call LP conditioning.

Here is a brief example of a Jeffrey-type updating scenario. Sherlock Holmes attributes the following probabilities to the propositions E_i that k_i is the culprit in a crime: $P(E_1) = 1/3, P(E_2) = 1/2, P(E_3) = 1/6$, where k_1 is Mr. R., k_2 is Ms. S., and k_3 is Ms. T. Then Holmes finds some evidence which convinces him that $P'(Y) = 1/2$, where Y is the proposition that the culprit is male and P is relatively prior to P'. What should be Holmes' updated probability that Ms. S. is the culprit? We will look at the recommendations of Jeffrey conditioning and LP conditioning for this case in the next section. For now, we note that LP conditioning violates all of the following plausible expectations for an amujus, an 'alternative method of updating for Jeffrey-type updating scenarios.'

- CONTINUITY An amujus ought to be continuous with standard conditioning as a limiting case.

- INVARIANCE An amujus ought to be partition invariant.
- LEVINSTEIN An amujus ought not to give "extremely unattractive" results in a Levinstein scenario (see Levinstein, 2012).
- REGULARITY An amujus ought not to assign a posterior probability of 0 to an event which has a positive prior probability and about which the intervening evidence says nothing except that a strictly weaker event has a positive posterior probability.
- ASYMMETRY An amujus ought to reflect epistemic asymmetries.

We are faced with the choice of rejecting the geometry of reason or accepting these unpleasant consequences. Fortunately, there is a live alternative to the geometry of reason: information theory. Information theory has its own axiomatic approach to justifying probabilism and standard conditioning (see Shore and Johnson, 1980). Furthermore, information theory provides a justification for Jeffrey conditioning and generalizes it (see Lukits, 2015). Information theory is not a geometry of reason in the sense that it measures divergences, not distances, between distributions of partial belief. In other words, the divergence of b'' from b' may not be equal to the divergence of b' from b''. Updating methods based on information theory (standard conditioning, Jeffrey conditioning, the principle of maximum entropy) fulfill expectations CONTINUITY, INVARIANCE, LEVINSTEIN, REGULARITY, and ASYMMETRY.

Returning from general argument to mathematical detail, salient axioms in Leitgeb and Pettigrew are both local and global Normality and Dominance (see Leitgeb and Pettigrew, 2010a, 219):

> **Local Normality and Dominance**: If I is a legitimate inaccuracy measure, then there is a strictly increasing function $f : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ such that, for any $A \in W$, $w \in W$, and $x \in \mathbb{R}_0^+$,
>
> $$I(A, w, x) = f\left(|\chi_A(w) - x|\right). \tag{9}$$
>
> **Global Normality and Dominance**: If G is a legitimate global inaccuracy measure, there is a strictly increasing function $g : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ such that, for all worlds w and belief functions $b \in \mathrm{Bel}(W)$,
>
> $$G(w, b) = g\left(\|w - b_{\mathrm{glo}}\|\right). \tag{10}$$

Similarly to Joyce, these axioms are justified on the basis of geometry, but this time more explicitly so:

> Normality and Dominance [are] a consequence of taking seriously the talk of inaccuracy as 'distance' from the truth, and [they endorse] the geometrical picture provided by Euclidean n-space as the correct clarification of this notion. As explained in section 3.2, the assumption of this geometrical picture is one of the presuppositions of our account, and we do not have much to offer in its defense, except for stressing that we would be equally interested in studying the consequences of minimizing expected inaccuracy in a non-Euclidean framework. But without a doubt, starting with the Euclidean case is a natural thing to do.

The next section provides a simple example where the distance of geometry and the divergence of information theory differ. With this difference in mind, I will show how LP conditioning fails the five expectations outlined above. The conclusion is that a rational agent uses information theory, not the geometry of reason.

3 Geometry of Reason versus Information Theory

Consider the following three points in three-dimensional space:

$$A=\left(\frac{1}{3},\frac{1}{2},\frac{1}{6}\right)\qquad B=\left(\frac{1}{2},\frac{3}{8},\frac{1}{8}\right)\qquad C=\left(\frac{1}{2},\frac{5}{12},\frac{1}{12}\right)\tag{11}$$

All three are elements of the three-dimensional simplex $\mathbb{S}^3$: their coordinates add up to 1. Thus they represent probability distributions over a partition of the event space into three events. Now call $D_{\mathrm{KL}}(A,B)$ the Kullback-Leibler divergence of B from A defined as follows, where a_i are the Cartesian coordinates of A:

$$D_{\mathrm{KL}}(A,B)=\sum_{i=1}^{3}a_i\ln\frac{a_i}{b_i}\tag{12}$$

The Euclidean distance $\|A-B\|$ is defined as in equation (6). What is remarkable about the three points in (11) is that [what is remarkable here?? shouldn't we be comapring the differences in global expectation, metric distance AND Kullback-Leibler?]

$$\mathrm{GExp}_A(C) \approx 0.653 < \mathrm{GExp}_A(B) \approx 0.656 \tag{13}$$

and

$$\|A - C\| \approx 0.057 < \|A - B\| \approx 0.072 \tag{14}$$

assuming in (13) the global inaccuracy measure presented in (5) and $E = W$ (all possible worlds are epistemically accessible). The Kullback-Leibler divergence and Euclidean distance give different recommendations with respect to closeness. If A corresponds to my prior and my evidence is such that I must change the first coordinate to $1/2$ and nothing stronger, then information theory via the Kullback-Leibler divergence recommends the posterior corresponding to B; and the geometry of reason as expounded in Leitgeb and Pettigrew recommends the posterior corresponding to C.

There are several things going on here that need some explanation. First, we note that for Leitgeb and Pettigrew, expected global inaccuracy of b' is always evaluated by the lights of another partial belief distribution b. This may sound counterintuitive. Should we not evaluate b' by its own lights? It is part of a larger Bayesian commitment that partial belief distributions are not created ex nihilo. They can also not be evaluated for inaccuracy ex nihilo. Leitgeb and Pettigrew say very little about this, but it appears that there is a deeper problem here with the flow of diachronic updating. The classic Bayesian picture is one of moving from a relatively prior probability distribution to a posterior distribution. This is nicely captured by standard conditioning, Bayes' formula, and updating on the basis of information theory (the Kullback-Leibler divergence reflects this flow by asymmetries which we will bring up again below).

The geometry of reason and notions of accuracy based on it sit uncomfortably with this idea of flow, as the suggestion is that partial belief distributions are evaluated on their accuracy without reference to a prior probability distributions—why should the accuracy or epistemic virtue of a posterior probability distribution depend on a prior probability distribution which has already been debunked by the evidence? I agree with Leitgeb and Pettigrew that there is no alternative here but to evaluate the posterior by the lights of the prior. Not doing so would saddle us with Carnap's Straight Rule, where priors are dismissed as irrelevant (see Carnap, 1952, 40ff). Yet we shall note that a justification of evaluating a belief function's accuracy by the lights of another belief function is a lot less persuasive than the way Bayesians and information theory integrate prior distributions into forming posterior distributions by virtue of an asymmetric flow of information.

Second, I want to outline how Leitgeb and Pettigrew arrive at posterior probability distributions in Jeffrey-type updating scenarios. I will call their method LP conditioning. Consider a possibility space $W = E_1 \cup E_2 \cup E_3$ (the E_i are sets of states which are pairwise disjoint and whose union is W) and a partition $\mathcal{F}$ of W such that $\mathcal{F} = \{F^*, F^{**}\} = \{E_1, E_2 \cup E_3\}$. Let P be the prior probability function on W and P' the posterior. I will keep the notation informal to make this simple, not mathematically precise. Jeffrey-type updating scenarios give us new information on the posterior probabilities of partitions such as $\mathcal{F}$. In our example, let

$$\begin{array}{rcl} P(E_1) & = & 1/3 \\ P(E_2) & = & 1/2 \\ P(E_3) & = & 1/6 \end{array} \tag{15}$$

and the new evidence constrain P' such that $P'(F^*) = 1/2 = P'(F^{**})$.

Jeffrey conditioning works on the following intuition, which elsewhere I have called Jeffrey's updating principle JUP (see also Wagner, 2002) and where the posterior probabilities conditional on the partition elements equal the prior probabilities conditional on the partition elements (since we have no information in the evidence that they should have changed):

$$\begin{aligned} P'_{\text{JC}}(E_i) \quad & = P'(E_i|F^*)P'(F^*) + P'(E_i|F^{**})P'(F^{**}) \\ & = P(E_i|F^*)P'(F^*) + P(E_i|F^{**})P'(F^{**}) \end{aligned} \tag{16}$$

Jeffrey conditioning is controversial (for an introduction to Jeffrey conditioning see Jeffrey, 1965; for its statistical and formal properties see Diaconis and Zabell, 1982; for a pragmatic vindication of Jeffrey conditioning see Armendt, 1980, and Skyrms, 1986; for criticism see Howson and Franklin, 1994). Information theory, however, supports Jeffrey conditioning. Leitgeb and Pettigrew show that Jeffrey conditioning does not in general pick out the minimally inaccurate posterior probability distribution. If the geometry of reason as presented in Leitgeb and Pettigrew is sound, this would constitute a powerful criticism of Jeffrey conditioning. Leitgeb and Pettigrew introduce an alternative to Jeffrey conditioning, which we have called LP conditioning. It proceeds as follows for our example and in general provides the minimally inaccurate posterior probability distribution in Jeffrey-type updating scenarios.

Solve the following two equations for x and y:

$$\begin{array}{rcl} P(E_1)+x & = & P'(F^*) \\ P(E_2)+y+P(E_3)+y & = & P'(F^{**}) \end{array} \tag{17}$$

and then set

$$\begin{array}{rcl} P'_{\mathrm{LP}}(E_1) & = & P(E_1)+x \\ P'_{\mathrm{LP}}(E_2) & = & P(E_2)+y \\ P'_{\mathrm{LP}}(E_3) & = & P(E_3)+y \end{array} \tag{18}$$

For the more formal and more general account see Leitgeb and Pettigrew, 2010b, 254. The results for our toy example are:

$$\begin{array}{rcl} P'_{\mathrm{LP}}(E_1) & = & 1/2 \\ P'_{\mathrm{LP}}(E_2) & = & 5/12 \\ P'_{\mathrm{LP}}(E_3) & = & 1/12 \end{array} \tag{19}$$

Compare these results to the results of Jeffrey conditioning:

$$\begin{array}{rcl} P'_{\mathrm{JC}}(E_1) & = & 1/2 \\ P'_{\mathrm{JC}}(E_2) & = & 3/8 \\ P'_{\mathrm{JC}}(E_3) & = & 1/8 \end{array} \tag{20}$$

Note that (15), (20), and (19) correspond to A, B, C in (11). Now we will show how LP conditioning violates the five expectations. Because Leitgeb and Pettigrew's reasoning is valid, it cannot be sound. The premise to reject is the geometry of reason. Fortunately, information theory replaces it and yields results that fulfill the five expectations.

4 Five Expectations

It remains to provide more detail for the five expectations and to show how LP conditioning violates them. The full-length paper does so with formal rigour and by giving simple hands-on examples. In this abstract, I can only provide a short synopsis. LP conditioning violates CONTINUITY because standard conditioning gives a different recommendation than a series of Jeffrey-type updating scenarios which get arbitrarily close to standard event observation. This is especially troubling considering how important the case for standard conditioning is to Leitgeb and Pettigrew.

LP conditioning violates INVARIANCE because two agents who have identical credences with respect to a partition of the event space may disagree about this partition after LP conditioning, even when the Jeffrey-type updating scenario provides no new information about the more finely grained partitions on which the two agents disagree.

LP conditioning violates LEVINSTEIN because of "the potentially dramatic effect [LP conditioning] can have on the likelihood ratios between different propositions" (Levinstein, 2012, 419). Consider Benjamin Levinstein's example: There is a car behind an opaque door, which you are almost sure is blue but which you know might be red. You are almost certain of materialism, but you admit that there's some minute possibility that ghosts exist. Now the opaque door is opened, and the lighting is fairly good. You are quite surprised at your sensory input: your new credence that the car is red is very high. Jeffrey conditioning leads to no change in opinion about ghosts. Under LP conditioning, however, seeing the car raises the probability that there are ghosts to an astonishing 47%, given Levinstein's reasonable priors. Levinstein proposes a logarithmic inaccuracy measure as a foundation to avoid violation of LEVINSTEIN (vaguely related to the Kullback-Leibler divergence), but his account falls far short of the formal scope, substance, and integrity of information theory.

LP conditioning violates REGULARITY because formerly positive probabilities can be reduced to 0 even though the new information in the Jeffrey-type updating scenario makes no such requirements (as is usually the case for standard conditioning). Ironically, Jeffrey-type updating scenarios are meant to be a better reflection of real-life updating because they avoid extreme probabilities. The violation becomes especially egregious if we are already somewhat sympathetic to an information-based account: the amount of information required to turn a non-extreme probability into one that is extreme (0 or 1) is infinite. Whereas the geometry of reason considers extreme probabilities to be easily accessible by non-extreme probabilities under new information (much like a marble rolling off a table or a bowling ball heading for the gutter), information theory envisions extreme probabilities more like an event horizon. The nearer you are to the extreme probabilities, the more information you need to move on. For an observer, the horizon is never reached.

LP conditioning violates ASYMMETRY for reasons related to REGULARITY. Even the scrupulous about partial beliefs (such as Isaac Levi in Jeffrey's "Dracula Meets Wolfman: Acceptance vs. Partial Belief") concede that extreme probabilities are special and induce asymmetries in updating: moving in direction from certainty to uncertainty is asymmetrical to moving in direction from uncertainty to certainty. Geometry of reason's metric topology, however, allows for no asymmetries. Henri Poincaré once suggested that it could never be experimentally demonstrated that physical space was best modeled by a Euclidean topology, but a Euclidean topology was the simplest and therefore the preferable model. Impressed by Albert Einstein's relativity theory, Ernst Cassirer reinterpreted Poincaré's argument and suggested that once the universe is populated (by objects creating gravitational fields) a non-Euclidean topology is a simpler model for physical space. I would use this analogy here to suggest that once epistemology is populated with certainties, the geometry of reason (whether Euclidean or non-Euclidean) is no longer the simplest and most effective explanatory model.

5 Conclusion

In conclusion, Leitgeb and Pettigrew's reasoning to establish LP conditioning on the basis of the geometry of reason is valid. Given the failure of LP conditioning with respect to the five expectations, it cannot be sound. The premise to reject is the geometry of reason. Fortunately, information theory replaces it and yields results that fulfill the five expectations.

References

Armendt, Brad. "Is There a Dutch Book Argument for Probability Kinematics?" *Philosophy of Science* 583–588.

Carnap, Rudolf. *The Continuum of Inductive Methods.* University of Chicago, 1952.

Caticha, Ariel, and Adom Giffin. "Updating Probabilities." In *MaxEnt 2006, the 26th International Workshop on Bayesian Inference and Maximum Entropy Methods.* 2006.

Diaconis, Persi, and Sandy Zabell. "Updating Subjective Probability." *Journal of the American Statistical Association* 77, 380: (1982) 822–830.

Greaves, Hilary, and David Wallace. "Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility." *Mind* 115, 459: (2006) 607–632.

Howson, Colin, and Allan Franklin. "Bayesian Conditionalization and Probability Kinematics." *The British Journal for the Philosophy of Science* 45, 2: (1994) 451–466.

Jeffrey, Richard. *The Logic of Decision.* New York, NY: McGraw-Hill, 1965.

Joyce, James. "A Nonpragmatic Vindication of Probabilism." *Philosophy of Science* 65, 4: (1998) 575–603.

Leitgeb, Hannes, and Richard Pettigrew. "An Objective Justification of Bayesianism I: Measuring Inaccuracy." *Philosophy of Science* 77, 2: (2010a) 201–235.

———. "An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy." *Philosophy of Science* 77, 2: (2010b) 236–272.

Levinstein, Benjamin Anders. "Leitgeb and Pettigrew on Accuracy and Updating." *Philosophy of Science* 79, 3: (2012) 413–424.

Lukits, Stefan. "Maximum Entropy and Probability Kinematics Constrained by Conditionals." *Entropy* Special Issue "Maximum Entropy Applied to Inductive Logic and Reasoning": (2015) forthcoming.

Maher, Patrick. *Betting on Theories.* Cambridge University, 1993.

Pettigrew, Richard. "Epistemic Utility and Norms for Credences." *Philosophy Compass* 8, 10: (2013) 897–908.

Shore, J., and R.W. Johnson. "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy." *IEEE Transactions on Information Theory* 26, 1: (1980) 26–37.

Skyrms, Brian. "Dynamic Coherence." In *Advances in the Statistical Sciences: Foundations of Statistical Inference*, Springer, 1986, 233–243.

Wagner, Carl. "Probability Kinematics and Commutativity." *Philosophy of Science* 69, 2: (2002) 266–278.

Williams, Paul. "Bayesian Conditionalisation and the Principle of Minimum Information." *British Journal for the Philosophy of Science* 131–144.