Jeffrey Conditioning and the Geometry of Reason

Stefan Lukits

Abstract

1 Introduction

Issues of epistemic utility in Bayesian epistemology have attracted some attention in the past few years. Patrick Maher has provided 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). 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). 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 Leitgeb and Pettigrew, 2010b). We will call these assumptions the 'geometry of reason,' a term coined by Leitgeb and Pettigrew themselves.

What makes the geometry of reason interesting is 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, by contrast, widely considered a commonsense extension of standard conditioning, fails to provide maximal epistemic utility. Another type of conditioning, which we will call L&P conditioning, takes the place of Jeffrey conditioning.

The claim of this paper is that 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 L&P conditioning we have good reasons to favour Jeffrey conditioning, especially because Jeffrey conditioning is continuous with standard conditioning and L&P conditioning is not.

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), an extension which is on the one hand commonsensical and on the other hand formally continuous with the standard conditioning which Leitgeb and Pettigrew have worked so hard to vindicate nonpragmatically.

There are four sections to come. 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 L&P conditioning, information theory supports Jeffrey conditioning. Section 4 provides reasons why Jeffrey conditioning is superior to L&P conditioning on independent grounds: Jeffrey conditioning maintains the ratio of partial beliefs that are unaffected by the evidence (see Wagner), but more importantly, Jeffrey conditioning is continuous with standard conditioning, a feature which should be important to Leitgeb and Pettigrew and which L&P conditioning does not have.

Section 4 also highlights two other weaknesses of Leitgeb and Pettigrew's approach: (a) zero-probability events change the updating results conditional on whether they are included in the update or not; and (b) Leitgeb and Pettigrew's notion of partial belief accuracy is a doubtful analogy to full belief accuracy, since it is evaluated by the lights of an earlier distribution which is in many cases known to be inaccurate. 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, normality, weak convexity, and symmetry cannot stand as Joyce assumes.

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 originally distinguished 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 the difficulties in the debate on verisimilitude in the philosophy of science. 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 positive 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 geometry itself might not be the best approach considering that we are concerned with epistemic virtue.

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 in order to avoid epistemic dilemmas trying to minimize inaccuracy on both measures, one must adopt a Brier score to measure inaccuracy. 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 $I : P(W) \times W \times \mathbb{R}_0^+ \to \mathbb{R}_0^+$ 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 $G : W \times \mathrm{Bel}(W) \to \mathbb{R}_0^+$ 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

$$I(A, w, x) = \lambda \left(\chi_A(w) - x\right)^2 \tag{1}$$

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

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

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

$$\sqrt{\sum_{i=1}^{n} (u_i - v_i)^2}. \tag{3}$$

We use (1) 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:

$$\mathrm{LExp}_b(I, A, E, x) = \sum_{w \in E} b(\{w\}) I(A, w, x) = \sum_{w \in E} b(\{w\}) \lambda \left(\chi_A(w) - x\right)^2. \tag{4}$$

We use (2) 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:

$$\mathrm{GExp}_b(G, E, b') = \sum_{w \in E} b(\{w\}) G(w, b') = \sum_{w \in E} b(\{w\}) \mu \|w - b\|^2. \tag{5}$$

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 justifies 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 to identify its shortcomings.

Continuous with Joyce's work, 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 L&P conditioning.

My claim is that L&P conditioning violates plausible expectations that (i) updating methods for Jeffrey-type scenarios should be continuous with standard conditioning; (ii) updating methods generally should be partition invariant; and (iii) updating methods should reflect the epistemic asymmetry of updating from b' to b'' versus updating from b'' to b'. (iii) may be interesting, but is not decisive; (i) and (ii), however, decisively undermine L&P conditioning so that 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 differ compared to the divergence of b' from b''. Updating methods based on information theory (standard conditioning, Jeffrey conditioning, the principle of maximum entropy) all fulfill expectations (i)–(iii).

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 $x \in \mathbb{R}_0^+$,
>
> $$I(A, w, x) = f\left(|\chi_A(w) - x|\right). \tag{6}$$
>
> **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 $b \in \mathrm{Bel}(W)$,
>
> $$G(w, b) = g\left(\|w - b_{\mathrm{glo}}\|\right). \tag{7}$$

Similarly to Joyce, these asioms 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 L&P conditioning fails all three plausible conditions (i)–(iii) outlined above. The conclusion is that a rational agent is better off with information theory than with the geometry of reason.

3 Geometry of Reason versus Information Theory

Consider the following three points in three-dimensional space:

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

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:

$$D_{\mathrm{KL}}(A,B)=\sum_{i=1}^{3}a_i\ln\frac{a_i}{b_i} \tag{9}$$

The Euclidean distance $\|A-B\|$ is defined as in equation (3). What is remarkable about the three points in (8) is that

$$\mathrm{GExp}_A(C)\approx 0.653<\mathrm{GExp}_A(B)\approx 0.656 \tag{10}$$

and

$$\|A-C\|\approx 0.057<\|A-B\|\approx 0.072 \tag{11}$$

assuming in (10) the global inaccuracy measure presented in (2) 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. [by the lights of b, L&P conditioning] 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.

4 Jeffrey Conditioning and L&P Conditioning

5 Asymmetry and Conclusion

6 References

References

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

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

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.

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.

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.

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