An Objective Justification of Bayesianism I: Measuring Inaccuracy

Hannes Leitgeb and Richard Pettigrew

December 19, 2009

Abstract

One of the fundamental problems of epistemology is to say when the evidence in an agent's possession justifies the beliefs she holds. In this paper and its sequel, we defend the Bayesian solution to this problem by appealing to the following fundamental norm:

> Accuracy An epistemic agent ought to minimize the inaccuracy of her partial beliefs.

In this paper, we make this norm mathematically precise in various ways. We describe three epistemic dilemmas that an agent might face if she attempts to follow Accuracy, and we show that the only inaccuracy measures that do not give rise to such dilemmas are the quadratic inaccuracy measures. In the sequel, we derive the main tenets of Bayesianism from the relevant mathematical versions of Accuracy to which this characterization of the legitimate inaccuracy measures gives rise, but we also show that unless the requirement of Rigidity is imposed from the start, Jeffrey conditionalization has to be replaced by a different method of update in order for Accuracy to be satisfied.

1 Introduction

One of the fundamental problems of epistemology is to say when the evidence in an agent's possession justifies the beliefs she holds; and, when it does, how it does this, and to what extent. In this paper and its sequel, we defend the Bayesian solution to this problem for those cases in which the set of possible worlds about which the agent holds an opinion is finite. If W is such a set of possible worlds, let $\mathcal{P}(W)$ be the power set of W, and let $\mathrm{Bel}(W)$ be the set of functions $b : \mathcal{P}(W) \to \mathbb{R}^+_0$. We regard each function in $\mathrm{Bel}(W)$ as a (potential) *belief function* on the power set of W. Indeed, one of the distinctive presuppositions of Bayesianism is that, if W is the set of possible worlds about which an agent holds an opinion, then that agent's epistemic state at a given time t may be represented by a belief function $b_t \in \mathrm{Bel}(W)$ that takes each proposition A, represented as a subset of W, to a real number $b_t(A)$ that measures the degree of credence the agent assigns to A. Thus, to solve the fundamental problem of epistemology, the Bayesian must say, for a given body of evidence, which belief functions it would be rational for an agent to have at a time when she is in possession of that evidence.

At the core of Bayesianism lie two claims that go some way to solving the fundamental problem of epistemology as set out above: they are **Probabilism** and **Conditionalization**. **Probabilism** is a *synchronic* norm: that is, it concerns the intrinsic properties of an agent's belief function at particular moments in her epistemic life. In particular, it demands that an

agent's belief function should be a probability function at any such moment. **Conditionalization**, on the other hand, is a *diachronic* norm: that is, it concerns the relation between an agent's belief functions at different times. In particular, it demands that an agent who learns the truth of a proposition E between times t and t' (and nothing stronger) ought to update her belief function by conditionalizing on E.[1]

These central claims have been extended in various ways by various philosophers. Two such extensions will be of particular interest to us. First, the synchronic claim that characterizes (a version of) Objectivist Bayesianism: if an agent has opinions only about finitely many possible worlds, and if E is the strongest proposition given to her by her evidence so far, then her belief function ought to be the uniform probability distribution over the worlds in E: we will call this **Uniform Distribution**. And, second, Richard Jeffrey proposed a diachronic norm, which we will call **Jeffrey Conditionalization**, and which is meant to cover those instances of updating in which the evidence learned does not come in the form of a proposition learned with certainty, as in **Conditionalization**, but rather in the form of some weaker side constraints on the agent's belief function.

In this paper and in its sequel ("An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy"), we defend the two core tenets of Bayesianism (and, to a much lesser extent, the additional Objectivist Bayesian tenet) by appealing to the following fundamental norm:

> Accuracy An epistemic agent ought to approximate the truth. In other words: she ought to minimize her inaccuracy.[2]

Also, we use this norm to criticize Jeffrey's updating rule, and we defend an alternative to Jeffrey conditionalization that applies to the same type of situations. We consider all of this to be an *objective* manner of justifying Bayesianism, which is based on theoretical considerations on how to get to the truth rather than on practical considerations on how to make prudent decisions. Indeed, for us an agent's degree of belief in a proposition A is such that the agent ought to minimize its distance from the truth value of A; for all epistemological purposes, this feature is in fact constitutive of the notion of degree of belief. Although the chosen type of justification is objective in this sense, it should be kept in mind that what gets justified in this way is still mainly just standard *subjective* Bayesianism.

We agree with Jim Joyce ([6], [7]) that the relevant notion of accuracy here is what he calls *gradational accuracy*, where gradational accuracy depends only on the truth values of propositions at worlds and on the agent's belief function.[3] We will make this feature of accuracy more precise in the Local Normality and Dominance postulate in section 5 below, when we will give the notion a mathematical analysis, as promised. The quantitative notion of accuracy that interests us differs from Popper's [13] comparative concept of verisimilitude, according

[1] As Jon Williamson reminded us, **Conditionalization** must be qualified: it holds only when plain factual evidence about the world is involved. Cf. [17]

[2] Why not *maximize your accuracy* instead? Because we like to think of inaccuracy as being given by a distance: the lesser the distance from the truth, the lesser the inaccuracy; the greater the distance from the truth, the greater the inaccuracy. Since distances from the truth are bounded from below, i.e., by the zero distance, they can be minimized, and that is what we are asking for in Accuracy. Using some means of transformation, these properties might perhaps be captured just as well by accuracy or 'inverse distance', but employing the notion of inaccuracy directly seems to give us a much more appealing way of stating our central epistemic goal.

[3] Indeed, Joyce's original paper has been a major inspiration for the project in this paper and its sequel. We discuss Joyce's own account in detail in the sequel of this article.

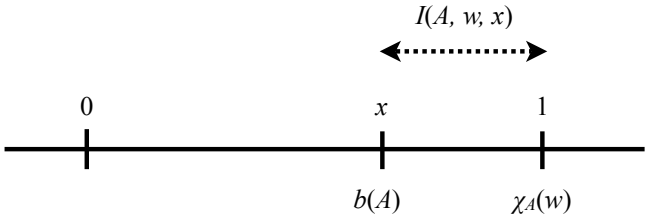

Figure 1:

to which some sets of statements are closer to the truth than other sets of statements. This concept was proven inadequate by Miller's [9] and Tichý's [16] triviality results.[4]

Given this understanding of the notion of accuracy, it is the purpose of this paper to make the Accuracy norm precise. In the sequel, we will then investigate the consequences of this norm.

2 The basic concepts and the argument in brief

Our argument is long and it involves a number of distinctions. Thus, for the sake of clarity, we present an overview of its underlying concepts and its structure before we give it in a fully detailed form.[5]

We begin by drawing a distinction between *local* and *global* measures of inaccuracy. A *local inaccuracy measure* is a mathematical function that takes a proposition $A \subseteq W$, a world $w \in W$, and a non-negative real number x and gives a measure $I(A, w, x)$ of the *local* inaccuracy of having degree of credence x in proposition A at world w. So $I(A, w, x)$ measures the distance of x from the truth value $\chi_A(w)$ of A at w, where the truth values are represented by the real numbers 0 and 1, as shown in Figure 1. Intuitively, $I(A, w, x)$ will be greater the more x differs from the truth value of A in w.

On the other hand, a *global inaccuracy measure* is a mathematical function that takes a belief function b and a world w and gives a measure $G(w, b)$ of the *global* inaccuracy of having belief function b at world w. So $G(w, b)$ measures the distance of b from the world w, where both belief function and world will be represented geometrically in terms of vectors, as we will explain in detail in section 3.2 and as shown in Figure 2. Again intuitively, $G(w, b)$ will be greater the more the degree of belief assignment that is determined by b differs from the

[4] In the meantime, refined theories of verisimilitude have been introduced which do not suffer from triviality results; some of them do resemble our theory in important respects. In particular, Niiniluoto's theory of estimated truthlikeness (cf. [10]) defines a relational notion of truthlikeness in terms of an expected value of quantitative truthlikeness which bears some similarity to the expected inaccuracy of propositions that we will be interested in. We leave it to another paper to work out the details of this correspondence which was pointed out to us by Theo Kuipers.

[5] Bas van Fraassen's comments were instrumental in making this section clearer.

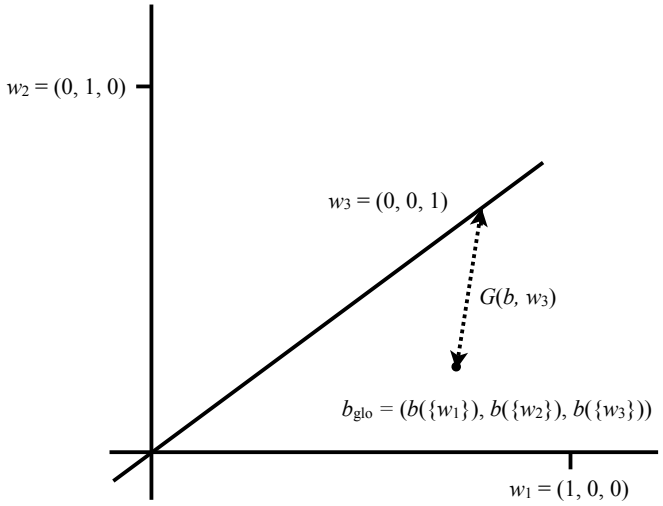

Figure 2:

truth value assignment that is determined by w. Obviously, not *any* such function I or G will do: rather we will have to restrict ourselves to sensible or *legitimate* choices of such functions, which will be achieved later by formulating postulates on what such legitimate I or G are like.

Thus, our first attempt to make Accuracy precise results in the norm bifurcating:

> Accuracy (Local) An agent ought to minimize the *local* inaccuracy of her degrees of credence in all propositions $A \subseteq W$ relative to a legitimate measure of local inaccuracy.
>
> Accuracy (Global) An agent ought to minimize the *global* inaccuracy of her belief function relative to a legitimate measure of global inaccuracy.

Note that, while we will be interested throughout in versions of both of these norms, Joyce is concerned solely with a version of Accuracy (Global) ([6], [7]).

Thus, we have two norms to which Accuracy gives rise. However, as formulated above, the norms are still incomplete: both local and global inaccuracy is only defined *relative to a world*, but we have not specified yet relative to which world or set of worlds the inaccuracies in questions are to be calculated. The obvious answer at this point would seem to be: relative to the actual world. But from an internalist point of view on justification—which we are going to adopt, as we will explain at greater length below (section 3.3)—this will not do, since we should not presuppose that the agent knows which world w in W is the actual world. Instead, the agent should take into account inaccuracies with respect to all and only the worlds that are epistemically possible for her; if this set is taken to be epistemically accessible

to her, then we do not violate internalism about justification by demanding that she assesses her overall inaccuracy in terms of it. Hence, we focus on the measures of *expected* local and global inaccuracy to which any pair of legitimate local and global inaccuracy measures will give rise, and we evaluate these expected inaccuracy measures over the set E of epistemically possible worlds.

The expected local inaccuracy of a degree of credence is defined, as one would expect, as the sum of its inaccuracies at various worlds weighted by the degree of belief assigned to each of those worlds (or rather to their singleton sets). Thus, to determine the expected local inaccuracy of a degree of credence x in a proposition A, we must specify three parameters:

(i) The belief function that gives the degree of belief assigned to each of the worlds over which the sum is taken. That is, the belief function we use to weight the inaccuracies that we sum to give the expected local inaccuracy.

(ii) The set of worlds over which the sum is taken. This is the set of worlds that are epistemically possible for the agent.

(iii) The local inaccuracy measure that gives the inaccuracies of the degree of credence at the various worlds over which the sum is taken.

Thus, we have the following definition:

Definition 1 (Expected local inaccuracy) *Given a local inaccuracy measure I, a belief function b, a degree of credence x, and propositions $A, E \subseteq W$, we define the* expected local inaccuracy of x in proposition A by the lights of b, with respect to I, and over the set E of epistemically possible worlds *as follows:*

$$\mathrm{LExp}_b(I, A, E, x) = \sum_{w \in E} b(\{w\}) I(A, w, x)$$

Expected *global* inaccuracy requires us to fix the same parameters. It is defined as follows:

Definition 2 (Expected global inaccuracy) *Given a global inaccuracy measure G, belief functions b and b', and a proposition $E \subseteq W$, we define the* expected global inaccuracy of b' by the lights of b, with respect to G, and over the set E of epistemically possible worlds *as follows:*

$$\mathrm{GExp}_b(G, E, b') = \sum_{w \in E} b(\{w\}) G(w, b')$$

Thus, our second attempt to make Accuracy precise gives:

> Accuracy (Expected local) An agent ought to minimize the *expected* local inaccuracy of her degrees of credence in all propositions $A \subseteq W$ relative to a legitimate measure of local inaccuracy.
>
> Accuracy (Expected global) An agent ought to minimize the *expected* global inaccuracy of her belief function relative to a legitimate measure of global inaccuracy.

However, neither of these proposals is fully specified in its current form either, for we have not said by the lights of which belief function an agent ought to assess her expected local or global inaccuracy, nor over which set of epistemically possible worlds. Specifying the belief function by the lights of which we assess expected local or global inaccuracy leads us to

a further distinction: the distinction between *synchronic* and *diachronic* versions of both Accuracy (Expected local) and Accuracy (Expected global). Here are the synchronic versions of Accuracy (Expected local) and Accuracy (Expected global):

> Accuracy (Synchronic expected local) An agent ought to minimize the expected local inaccuracy of her degrees of credence in all propositions $A \subseteq W$ by the lights of her *current* belief function, relative to a legitimate local inaccuracy measure, and over the set of worlds that are *currently* epistemically possible for her.
>
> Accuracy (Synchronic expected global) An agent ought to minimize the expected global inaccuracy of her current belief function by the lights of her *current* belief function, relative to a legitimate global inaccuracy measure, and over the set of worlds that are *currently* epistemically possible for her.

And here are the diachronic versions of Accuracy (Expected local) and Accuracy (Expected global), where an agent has learned evidence between time t and time t' that imposes constraints C on her belief function $b_{t'}$ at time t', or on the set E of worlds that are epistemically possible for her at t', or both.

> Accuracy (Diachronic expected local) *At time t'*, such an agent ought to have a belief function that satisfies constraints C and is minimal amongst belief functions thus constrained with respect to the expected local inaccuracy of the degrees of credence it assigns to each proposition $A \subseteq W$ by the lights of her belief function *at time t*, relative to a legitimate local inaccuracy measure, and over the set of worlds that are epistemically possible for her *at time t' given the constraints C*.
>
> Accuracy (Diachronic expected global) *At time t'*, such an agent ought to have a belief function that satisfies constraints C and is minimal amongst belief functions thus constrained with respect to expected global inaccuracy by the lights of her belief function *at time t*, relative to a legitimate global inaccuracy measure, and over the set of worlds that are epistemically possible for her *at time t' given the constraints C*.

These are very nearly our final versions of Accuracy. All that remains is to specify the legitimate local and global inaccuracy measures. This is the work of section 5, where we argue for the main thesis of this paper: the only legitimate local and global inaccuracy measures are *quadratic inaccuracy measures*, which are known as *Brier scores* in the literature on scoring rules ([2]). We call these characterizations Local Inaccuracy Measures and Global Inaccuracy Measures respectively. By substituting them into the appropriate norms above, we obtain the four mathematically precise versions of Accuracy that are the aim of this paper. In the sequel, we investigate the consequences of these norms.

Of course, as with any set of norms, it may turn out that some or all of these norms are simply unsatisfiable. Prior to investigation, there is no reason to think that there are such minimally inaccurate belief functions in the senses required by these norms. However, as we will show in the sequel to this paper, such a situation arises only for certain instances of Accuracy (Diachronic expected local); for each of the other norms and indeed for many instances of this norm, there are belief functions that satisfy them and, moreover, they are the belief functions that the Bayesian demands. Furthermore, in the situations in which Accuracy (Diachronic expected local) cannot be satisfied, its global analogue Accuracy (Diachronic expected

global) can be. Thus, our approach does make a demand in these situations. We will discuss this further in the second paper.

Note that $\mathrm{LExp}_b(I, A, E, x)$ and $\mathrm{GExp}_b(G, E, b')$ will always be zero if the global belief function b is identical to the constant zero function on singletons of members of E. So if an agent has ruled out all worlds in E by means of b, any degree of belief whatsoever may be assigned to A in order to minimize expected local and global inaccuracy. In order to do better, we would have to generalize our framework and aim at a justification of Popper functions (see [12], for one reference among many) rather than standard absolute probability measures. We leave this topic to a different paper.

In section 3, we discuss in greater detail the transition from Accuracy to the four norms just listed, making explicit all the formal and philosophical presuppositions of our theory. We are not going to justify these presuppositions in any substantial manner, but at least we will make sure we put all of our cards on the table, and we will formulate explicitly the main questions that will remain open. In section 4, we give a brief sketch of the argument in favour of quadratic inaccuracy measures, both local and global, and in section 5, we give this argument in full detail.

3 The presuppositions of our argument

3.1 The Ought-Can principle

We wish to make the following norm precise:

> Accuracy An epistemic agent ought to approximate the truth. In other words: she ought to minimize her inaccuracy.

In doing so, we will be guided at a number of points by the following version of a well-known normative or metanormative principle, the function of which is to constrain our choice of normative systems:

> Ought-Can A norm should not demand anything of an agent that is beyond her epistemic reach.

This is just a variant of the classical Ought-Can principle applied in the present context. We leave open the exact character of the possibility modality that is implicit in 'Can' and 'reach(ability)'. However, as will become clear from our applications of Ought-Can, it is certainly stronger than mere logical possibility, and not too far from a notion of realistic achievability in the epistemic domain.

One important consequence of Ought-Can is that one should not demand of an agent that she draw distinctions that she is conceptually unable to draw. Here is one way in which this is relevant to our argument: In the following sections, we presuppose that our agents hold opinions about a finite and non-empty set W of possible worlds of which they assume the actual world to be one. There is nothing particularly philosophical about our decision to stick to the case of finitely many worlds in this paper; we simply assume this to be so and postpone the discussion of the infinite case to another time. However, already this seemingly harmless supposition may appear to have drastic—and even drastically wrong—consequences: Since there are, presumably, infinitely many possible worlds, every choice of a finite set of 'possible worlds' must consist in carving the actual set of worlds into finitely

many chunks, which are then presupposed to figure as the chosen '(pseudo-)possible worlds' of the framework. Worse, if a uniform probability measure has been defined over the resulting finite set W, as demanded by Objectivist Bayesianism (**Uniform Distribution**), it is unclear whether this measure is also uniform over the actual set of worlds, and if so, in what sense. So are we buying into a presupposition that makes our approach appear highly questionable from the start? Not necessarily, in light of the Ought-Can principle. Assume that we are solely concerned with agents whose conceptual resources cut the space of logical possibilities into finitely many pieces—the members of W. Indeed, we may assume these agents to be epistemically unable to distinguish between any two of the possible worlds that belong to one and the same partition set, and to be incapable also of altering their conceptual framework on any rational grounds for the time of our investigation. From the viewpoint of any of these agents, at the time of our investigation, there is thus no way of having any sort of epistemic access to the 'actual' infinite set of possible worlds. Demanding of such agents that they transcend their conceptual boundaries would mean demanding that they go beyond their epistemic reach, in conflict with Ought-Can. Therefore, whenever we refer to agents in this article, let them be as just described, and Ought-Can will allow us to proceed without any further worries. For many practical purposes, maybe even *we* can be taken to be such finitely constrained agents. In all other contexts, i.e., whenever we are dealing with fully-fledged human agents and their (maybe) infinitary conceptual capacities, the arguments of this paper do not apply. But hopefully, even in the latter case, our arguments will be interesting in themselves.

Ought-Can will prove even more relevant in section 3.3 where we derive from it an internalist view of justification that forces us to move from Accuracy (Local) and Accuracy (Global) to Accuracy (Expected local) and Accuracy (Expected global), as we mentioned above.

And finally the principle will play a crucial role in section 5 as well, where we argue for Local and Global Inaccuracy Measures, which specify the legitimate local and global measures of inaccuracy: It follows from Ought-Can that a normative system should not be such that it may lead an agent who obeys the norms of this system into an epistemic dilemma, i.e., into a situation in which she ought to change her epistemic state in two or more ways that are jointly impossible. We show that an agent who employs an inaccuracy measure that is not permitted by Local and Global Inaccuracy Measures may face an epistemic dilemma akin to the so-called discursive dilemma in the theory of judgement aggregation. From this conclusion and Ought-Can, our characterizations of the legitimate inaccuracy measures as being the quadratic ones will follow.

3.2 The geometrical framework

So much for Ought-Can; we will return to it below. In the meantime, we turn to the other background assumptions of our approach. Broadly speaking, we are taking a geometrical approach; to coin a slogan, we are interested in the 'geometry of reason'. Of course, it is common procedure to go back and forth between probability measures as characterized by the Kolmogorov axioms and probability measures viewed as points in geometrical space, so it might seem that putting forward a geometrical account of belief dynamics is completely unproblematic. However, our task is not so much to presuppose probability theory and to exploit its mathematical features in applications of probabilistic reasoning but rather to justify probabilistic reasoning in a context in which probability theory is not yet in place. So we will have to take things more slowly.

Since W is finite, there is an n such that $n = |W|$. We start by positioning the members of W in the n-dimensional Euclidean space $\mathbb{R}^n$ by 'identifying' each world $w_i \in W = \{w_1, \ldots, w_n\}$ with the i-th unit vector in $\mathbb{R}^n$, i.e., with the vector $(\delta_{i,1}, \ldots, \delta_{i,n})$ where $\delta_{i,j} = 1$ if $j = i$ and $\delta_{i,j} = 0$ if $j \neq i$. If the real number 1 is taken to represent truth and the real number 0 is taken to stand for falsity, then each such vector $(\delta_{i,1}, \ldots, \delta_{i,n})$ can be considered as an assignment of the geometrical counterparts of truth and falsity to the singleton propositions $\{w_1\}, \ldots, \{w_n\}$, such that the j^{th} coordinate of the vector $(\delta_{i,1}, \ldots, \delta_{i,n})$ is identical to the real number 1 if, and only if, the proposition $\{w_j\}$ is true in the world that is represented by this vector. Given this mapping from worlds to geometrical points or vectors, we can measure distances or 'closeness' between worlds with respect to each other in terms of the Euclidean distance between their geometric counterparts. Accordingly, regarding truth values 1, 0 as points 1, 0, respectively, in the one-dimensional Euclidean space, i.e., $\mathbb{R}$, allows us to measure distances between truth values geometrically. Now, there is certainly not just one geometrically 'natural' notion of distance in Euclidean space. In fact, there are lots of geometrically plausible metrics on $\mathbb{R}$ and $\mathbb{R}^n$ that are distinct from the Euclidean metric, and indeed we are going to argue later that inaccuracy has to be measured in terms of one of them. However, we will demand that every such notion of the geometrical distance between two points supervenes on—is functionally dependent on—the Euclidean distance between these points. In this sense, measuring closeness will always amount to a geometrical, and indeed Euclidean, procedure in the context of this paper.

The next step is to locate degrees of belief within $\mathbb{R}$ and to place belief functions into $\mathbb{R}^n$. As pointed out at the beginning of this paper, we consider belief functions to be mappings of the form $b : \mathcal{P}(W) \to \mathbb{R}_0^+$, so degrees of belief are assumed to be quantitative objects from the start.[6] Following Accuracy from above, we regard rational agents to be aiming at distributing their degrees of belief in such a way that every such degree $b(A)$ approximates the truth value of the proposition A (though it's not yet clear exactly in what sense). Hence, a rational agent's degree of belief for a proposition is nothing but the agent's best possible estimate or 'simulation' of the truth value of that proposition, given her present epistemic situation. Since truth and falsity have been represented by real numbers, too, degrees of belief and truth values are comparable—they occupy the same quantitative or geometrical scale. So, e.g., assigning a degree of belief 1 to a proposition A would mean that the agent believes that A is true rather than false, since the degree of belief 1 is closer—in fact, identical—to the real number 1 that represents truth than it is to the real number 0 representing falsity. In this way, closeness of a degree of belief to the truth can be measured again according to a metric on the one-dimensional Euclidean space.

In order to see how belief functions determine points in the n-dimensional Euclidean space, by which it then becomes possible to measure their distances from the point vectors that represent possible worlds, it is useful to introduce the following terminology: call any function $b_{\text{glo}} : W \to \mathbb{R}_0^+$ a *global belief function* and let $\text{Bel}_{\text{glo}}(W)$ be the set of all global belief functions. Obviously, every such global belief function may be regarded as a vector $(a_1, \ldots, a_n) \in \mathbb{R}^n$ with $a_j = b_{\text{glo}}(w_j)$ for $j = 1, \ldots, n$. Since every belief function b induces a global belief function b_{glo} by means of $b_{\text{glo}}(w_j) = b(\{w_j\})$, we can position b in $\mathbb{R}^n$ through its corresponding global belief function. Apart from supplying belief functions with a geometrical

[6] We could have regarded belief functions as mappings into $\mathbb{R}$ rather than $\mathbb{R}_0^+$; some parts of our argumentation are in fact not going to hang on this. Our reason for not doing so from the start is quite simply that some readers might find negative degrees of belief odd.

interpretation, the latter equation also yields a way of interpreting global belief functions: we call them *global* belief functions as they may be taken to summarize an agent's attitude towards all the worlds w in W, i.e., towards all of the singleton propositions $\{w\}$ for $w \in W$. E.g., a global belief function with the vector $(1, 0, \ldots, 0)$ represents an agent who is—in the *degrees of belief being best possible estimates or 'simulations' of truth values* sense mentioned above—certain of the truth of the proposition $\{w_1\}$ and certain that every other singleton proposition is false. Geometrically, the distance between the vector that belongs to this global belief function and the vector that corresponds to the world w_1 is 0, which is exactly what we want to be the case in such a situation. So, (global) belief functions and worlds have become comparable as well, again by means of their geometrical representations.[7]

We have seen how belief functions determine corresponding global belief functions. Is there also a way of inverting this procedure? That is: Given a global belief function b_{glo}, is there a similarly salient way of determining a belief function b that assigns degrees of belief not just to singleton propositions (or worlds) but to all propositions whatsoever? If Bayesianism were taken for granted, the answer would of course be 'Yes', by iterated application of finite additivity. But Bayesianism is exactly what is at issue here, so without further argument there does not seem to be any obvious way of determining a unique belief function from a given global belief function *justifiedly*, nor, for that matter, of determining justifiedly any belief function from a given global belief function at all. For this reason, presupposing that an agent's epistemic state at a time involves the acceptance of a global belief function is at least *prima facie* a weaker presupposition than assuming that the agent's epistemic state at that time involves the acceptance of a belief function.

3.3 Internalism, expected inaccuracy, and Ought-Can again

In this section, we return to Ought-Can. Above, we noted that this principle forces the shift from Accuracy (Local) and Accuracy (Global) to Accuracy (Expected local) and Accuracy (Expected global). Here we explain how.

In section 5, we will argue that the only legitimate inaccuracy measures are quadratic inaccuracy measures. However, while this conclusion gives us an important auxiliary notion of inaccuracy, it does not yet by itself yield a notion of inaccuracy that an agent can actually make use of in order to determine the local inaccuracy of her degree of credence in a proposition or the global inaccuracy of her belief function. The reason, as we observed above, is simply that the agent cannot be assumed to know at which world she is evaluating inaccuracies. So, by Ought-Can, we require a conception of the minimization of inaccuracy towards which the agent is epistemically capable of aiming. On such a conception, an agent's assessment of the *local* inaccuracy of a degree of belief in a proposition will be bound to take into

[7] We are not suggesting that this is the only 'natural' way of identifying belief functions with vectors nor that this type of identification is free of presuppositions. In particular, representing belief functions geometrically in such a manner corresponds in some sense to a 'bias' towards worlds. It might well be that if a different form of representation were chosen, then we would not end up justifying Bayesianism but some different account of belief dynamics (say, the Dempster-Shafer approach or something else). At the same time, it is not as if this 'bias' towards worlds gives us anything like the Bayesian tenets in any obvious and immediate manner: it will need elaborate arguments, substantial postulates, and mathematical proofs until we will have finally derived these tenets (as we will in the sequel to this paper). As we will point out later, there are various parameters in our theory which we set in a particular way and for which we investigate the consequences of setting them as such. But we would be equally interested in studying alternative ways of setting these parameters and in determining the consequences of these alternative settings.

account the *local* inaccuracies of that degree of belief *at all possible worlds* that are not excluded by the evidence available to her, i.e., all worlds that are epistemically possible for her. Similarly, on this conception, an agent's assessment of the *global* inaccuracy of a belief function will be bound to take into account the *global* inaccuracies of that belief function *at all epistemically possible worlds*. Thus, the Ought-Can principle gives rise to a form of internalism about justification; and this internalism leads in turn to the notions of expected local and global inaccuracy defined above (Definitions 1 and 2) and to versions of the Accuracy norm that demand that an agent should minimize not her *actual* local or global inaccuracy, but her *expected* local or global inaccuracy instead. And, as we described above, this move results in a further bifurcation of the norms: Accuracy (Expected local) splits to give Accuracy (Synchronic expected local) and Accuracy (Diachronic expected local), while Accuracy (Expected global) splits to give Accuracy (Synchronic expected global) and Accuracy (Diachronic expected global). The synchronic norms constrain the intrinsic nature of an agent's belief function at particular times in her epistemic life: at any such time t, the agent's belief function b_t must have minimal expected local (respectively, global) inaccuracy by the lights of b_t itself. And the diachronic norms constrain the relation between an agent's belief function at times t and t' between which the agent obtains evidence that places constraints C on the legitimate belief functions at t' or on the epistemically possible worlds at t', or on both: $b_{t'}$ must satisfy C and must be minimal, amongst those belief functions that satisfy C, with respect to expected local (respectively, global) inaccuracy by the lights of b_t and over the possible worlds that are epistemically possible at t' and which satisfy the constraint C.

We should say a little more about the mathematical notion of expectation in this context. Usually, expectations are defined only for probability measures. However, in the absence of **Probabilism**—a claim that we wish to establish, not presuppose—there is no reason for thinking that belief functions are probability measures. On the other hand, while probability theory is the usual context in which expectations are defined, there is no objection in principle to extending the definition to cover the case of belief functions that may not be probability measures. Of course, if the belief function is not additive, we will not be able to prove the equivalence of the definitions we have given (see Definitions 1 and 2 from before) with two alternative definitions that are standard in probability theory: namely,

$$\mathrm{LExp}_b(I, A, E, x) = \sum_{r \in \mathrm{Ran}(I)} r \cdot b(\{w \in W : I(A, w, x) = r\})$$

$$\mathrm{GExp}_b(G, E, b') = \sum_{r \in \mathrm{Ran}(G)} r \cdot b(\{w \in W : G(w, b) = r\})$$

But we will never use the alternative definitions, and we take the definitions that we have given before to be the conceptually basic ones, so we do not regard this as a problem. It would be great to support this view by having to hand a general and abstract theory of the concept of *expected value* for which one could actually *prove* that given the presuppositions of our approach the resulting notion of expected value must be the one that is employed in definitions 1 and 2.[8] We will have to leave any investigations into that topic for a different occasion.

By using what we take to be the basic definitions of expectation, we avoid the objection that Joyce (p. 589, [6]) raises against Rosenkrantz's appeal to expected inaccuracy in [14].

[8] We thank Franz Dietrich for highlighting this in personal communication.

Like us, Rosenkrantz favours quadratic global inaccuracy measures. He lists constraints on global inaccuracy measures, and conjectures that they are satisfied uniquely by the quadratic functions. However, he gives no proof of this claim, and, as Joyce notes, he gives no non-circular arguments in favour of his constraints. In particular, Rosenkrantz demands of an agent that she minimize her expected inaccuracy calculated over every possible partition of the space simultaneously. As Joyce points out, unless her belief function is already assumed to be a probability function, this will not be possible. Our precise versions of the Accuracy norms are not vulnerable to this objection since we only ever appeal to the most fine-grained partition of W that is conceptually available to the agent, i.e., the set of singletons of worlds in W.

3.4 The status of our presuppositions

Now that the underlying formal framework of our theory has been made more explicit, how are we going to argue for it? Short answer: we don't. Not that there is nothing at all to say in favour of it: Amongst other things, one could point out that some aspects of our geometrical framework are purely conventional and thus do not need any further justification at all. E.g., we could have represented truth by, say, the real number 2, and accordingly worlds by vectors that are like unit vectors but where the coordinate 1 is replaced by 2, and so forth.[9] In other words, the absolute position of truth values, worlds, and belief functions in the Euclidean plane is arbitrary. However, fortunately, none of our results depends on it.

Furthermore, if that Kantian move helped at all, the choice of Euclidean geometry rather than a non-Euclidean one could perhaps be grounded in our intuitions about space and distance in general—although the 'geometry of reason' we are after can hardly be called Kantian, and in a context in which *information* is represented geometrically, often metrics other than the Euclidean one are used. In fact, one would be completely justified in wondering why degrees of belief should "live" in a Euclidean space at all—after all, physical space turned out to be non-Euclidean, and even from our internalist perspective it is perfectly reasonable to assume that there might be agents whose "cognitive spaces" are non-Euclidean.[10]

Other aspects of the framework do not even have a conventional or intuitive character at all: E.g., why is the notion of expected local inaccuracy given by taking the *sum* of weighted local inaccuracies rather than their product or their maximum or whatever other function comes to mind? Again, one could surely do better than just leaving the discussion at that point: indeed, one should be able to defend some necessary conditions on expected inaccuracy on the basis of assuming Ought-Can and the rest of the geometrical framework; it might also be possible to prove representation theorems by which our expected local inaccuracy functions would turn out to have an equivalent, and intrinsically plausible, qualitative or comparative formulation; and so forth. But it is hard to see how any such justification would be ultimate. Similarly: Why does each world have the same distance from each other world according to their geometric representations? Is this because we have renormalized the scales of our given coordinate system in a way that leads to this result trivially, or do we commit ourselves to a substantial assumption that is at least implicitly pointing towards Objectivist Bayesianism? We will have to say just slightly more about this last point in the sequel to this paper, but otherwise we simply leave the status of the framework untouched, i.e., we take this geometrical

[9] Franz Huber cites a corresponding worry raised by Colin Howson in an unpublished manuscript (p. 2, [5]).

[10] We are grateful to Alan Hájek and Kenny Easwaran for highlighting this as an important open problem.

framework as a presupposition of our justification of (Objectivist) Bayesianism without giving it any further defence.

One final remark, though: we do regard the question of just how much work in our argumentation is done by its geometrical Euclidean background framework as a very important one; eventually, this paper ought to be complemented by one that abstracts from the epistemic-geometrical models that we presuppose all and only the essential axioms that are needed in order for our arguments to go through, and only then will it be possible to see how much weight is carried by our Euclidean presuppositions. We will return to this as an open question to be formulated in the final section of this paper. When we determine the quadratic inaccuracy measures as the legitimate ones in section 5, this ought to be understood in the way that they are the legitimate ones *relative to the chosen Euclidean framework*; other inaccuracy measures might be legitimate if given a different framework.[11]

4 The argument for Local and Global Inaccuracy Measures: An overview

Section 5 is devoted to justifying two claims: Local Inaccuracy Measures and Global Inaccuracy Measures. The former says that the legitimate local inaccuracy measures are the quadratic scoring rules. The latter says that the legitimate global inaccuracy measures are the quadratic functions of the Euclidean metric on $\mathbb{R}^n$. In this section, we give an overview of the argument.

In fact, we will state three arguments for identifying these characterizations of the legitimate local and global inaccuracy measures. Each begins in the same way by imposing four conditions that restrict the class of legitimate inaccuracy measures (both local and global) on the basis of the principles highlighted in section 3. And each continues by noting that these restrictions fail to exclude inaccuracy measures that give rise to a dilemma for the agent: i.e., a situation in which the epistemic norms of the previous section, combined with these inaccuracy measures, entail two mutually exclusive prescriptions for the agent. In each case, when we restrict the class of legitimate inaccuracy measures to exclude those that lead to the dilemma in question, we are left only with the quadratic inaccuracy measures and Local and Global Inaccuracy Measures follow immediately.

Thus, in section 5, we give three separate arguments, each of which shares its first four

[11] We thank Branden Fitelson for making this point in discussion of our paper.

premises with the others. We set them out here for ease of reference:

First argument	Second argument	Third argument
Shared premises (§5.1):	*Shared premises* (§5.1):	*Shared premises* (§5.1):
Local Norm & Dom	Local Norm & Dom	Local Norm & Dom
Global Norm & Dom	Global Norm & Dom	Global Norm & Dom
Local & Global Comp.	Local & Global Comp.	Local & Global Comp.
Local & Global Min. Inacc.	Local & Global Min. Inacc.	Local & Global Min. Inacc.
Dilemma 1 premise (§5.2.1):	*Dilemma 2 premise* (§5.2.2):	*Dilemma 3 premises* (§5.2.3):
Agreement on Inaccuracy	Separability	Cont. Differentiability
		Agreement on Urgency
Conclusions	*Conclusions*	*Conclusions*
Local Inaccuracy Measures	Local Inaccuracy Measures	Local Inaccuracy Measures
Global Inaccuracy Measures	Global Inaccuracy Measures	Global Inaccuracy Measures

Each of the premises will be supported by our presuppositions—some strictly, some defeasibly. Furthermore, each of the arguments mentioned in this section will be seen to be strictly valid, i.e., deductively valid given mathematics, as shown by the proofs given in the appendix.

The conclusions of these three arguments will be our promised two characterizations of legitimate inaccuracy measures which we want to establish in section 5:

> Local Inaccuracy Measures If $I : \mathcal{P}(W) \times W \times \mathbb{R}_0^+ \to \mathbb{R}_0^+$ is a legitimate measure of the local inaccuracy of a degree of credence x in a proposition A at a possible world w, then there is $\lambda \in \mathbb{R}_{>0}$ such that
>
> $$I(A, w, x) = \lambda(\chi_A(w) - x)^2$$
>
> where $\chi_A : W \to \{0, 1\}$ is the characteristic function of the proposition A.
>
> Global Inaccuracy Measures If $G : W \times \mathrm{Bel}(W) \to \mathbb{R}_0^+$ is a legitimate measure of the global inaccuracy of a belief function b at a possible world w, then there is $\lambda \in \mathbb{R}_{>0}$ such that
>
> $$G(w, b) = \lambda||w - b_{\mathrm{glo}}||^2$$
>
> where w and b_{glo} are represented by vectors as in section 3.2 and $||u-v||$ is the Euclidean distance between vectors u and v: i.e. $||u-v|| = \sqrt{(u_1 - v_1)^2 + ... + (u_n - v_n)^2}$.

The claim that quadratic inaccuracy measures yield the only legitimate scoring rules is similar to Selten's central claim in [15]. Thus, it might seem that we could easily adapt Selten's ingenious argument to establish Local Inaccuracy Measures. However, at a number of points in his proof, Selten relies on the assumption that belief functions are probability functions. For his avowed purpose, this is perfectly legitimate. On the other hand, for the purposes of this paper, we could not avail ourselves of a result premised upon this assumption to establish Local Inaccuracy Measures, which we then in turn wish to use to derive **Probabilism**, amongst other things (and, to the best of our knowledge, similar points can be made about much of the excellent and highly evolved literature on scoring rules and decision theory). Thus, we must beat our own path to our conclusion.

As pointed out, we beat three paths, which begin together and diverge only at the final premises of the arguments. Thus, we begin with the shared premises in section 5.1; then, in section 5.2, we consider the three dilemmas that motivate the three different final premises. All three arguments turn on mathematical theorems; their proofs are annexed in the appendix at the end of the paper (section 7). Amongst the three arguments, we consider the final one (section 5.2.3) to be the strongest and most convincing; but the first two arguments (sections 5.2.1 and 5.2.2) are easier to state, which is why we will turn to them before we give the third, and philosophically central, argument.

5 Measuring inaccuracy

5.1 The shared premises

The first premise of each of our arguments combines local analogues of Joyce's Normality and Dominance conditions:[12] the local version of Joyce's Normality condition says that the inaccuracy of degree of credence x in proposition A at world w ought to depend only on the difference between x and the value of the characteristic function of A at w (that is, the truth value of A at w); the local analogue of Dominance merely states that local inaccuracy increases as this difference increases.

> 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 $A \subseteq W$, $w \in W$, and $x \in \mathbb{R}_0^+$,
>
> $$I(A, w, x) = f(|\chi_A(w) - x|)$$

Note that this also implies that distances from the truth—$\chi_A(w) = 1$—and distances from falsity—$\chi_A(w) = 0$—are measured in the same way, which is entailed by our geometrical take on truth and falsity as points in a space. It is clear that once a Euclidean framework such as ours is in place, a condition analogous to Local Normality and Dominance ought to hold for global inaccuracy measures as well. Local Normality and Dominance asserts that the local inaccuracy of a degree of credence x in proposition A at world w ought to be a strictly increasing function only of the *difference* (i.e., the Euclidean distance) between x and $\chi_A(w)$. Its analogue, Global Normality and Dominance asserts that the global inaccuracy of a global belief function b at a world w ought to be a strictly increasing function only of the *Euclidean distance* between the vector representation of b and the vector representation of w. That is,

> 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 $b \in \mathrm{Bel}(W)$,
>
> $$G(w, b) = g(||w - b_{\mathrm{glo}}||)$$

Global Normality and Dominance is a consequence of taking seriously the talk of inaccuracy as 'distance' from the truth and it endorses 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

[12] Respectively, p. 596 and p. 593 in [6].

to offer in its defence, except for stressing that we would be equally interested in studying the consequences of minimizing expected inaccuracy in a *non-Euclidean* framework. But without doubt starting with the Euclidean case is a natural thing to do.

The third premise that is shared by each of our arguments for Local and Global Inaccuracy Measures says that any function of the reals that gives rise to a legitimate local inaccuracy measure, also gives rise to a legitimate global inaccuracy measure and vice versa:

Local and Global Comparability

(i) If $I(A, w, x) = f(|\chi_A(w) - x|)$ is a legitimate local inaccuracy measure, then $G(w, b) = f(||w - b_{\text{glo}}||)$ is a legitimate global inaccuracy measure.

(ii) If $G(w, b) = g(||w - b_{\text{glo}}||)$ is a legitimate global inaccuracy measure, then $I(A, w, x) = g(|\chi_A(w) - x|)$ is a legitimate local inaccuracy measure.

Again, this is a consequence of our geometrical interpretation of accuracy: we interpret inaccuracy as distance from the truth, and we interpret distance as being given by a strictly increasing function of the Euclidean metric. Since distances are independent of dimension, it should always be possible to use legitimate local inaccuracy measures in order to determine their global counterparts, and also the other way round; it is simply not relevant on which dimensions Euclidean distances are measured.

The final premise shared by each of our three arguments for Local and Global Inaccuracy Measures does nothing more than to lay down a convention: we will permit only inaccuracy functions that take value zero when the distance between truth value and degree of belief or between world and global belief function is zero.

Minimum Inaccuracy

(i) If $I(A, w, x) = f(|\chi_A(w) - x|)$ is a legitimate local inaccuracy measure, then $f(0) = 0$.

(ii) If $G(w, b) = g(||w - b_{\text{glo}}||)$ is a legitimate global inaccuracy measure, then $g(0) = 0$.

In the presence of Local and Global Comparability, we can derive (i) from (ii) and (ii) from (i). Thus, we need only impose one of these conditions. However, we state both lest the reader be given the mistaken impression that one or the other is more fundamental.

Together, Local Normality and Dominance, Global Normality and Dominance, Local and Global Comparability, and Minimum Inaccuracy restrict the class of legitimate inaccuracy measures. However, as we shall see in the next section, they do not restrict them enough. There are functions that satisfy these restrictions but which have undesirable properties. In the next section, we call attention to one particular sort of undesirable property: that is, the property of giving rise to a dilemma for an epistemic agent. In each case, the dilemmas in question concern possible discrepancies between measuring inaccuracy in a local and in a global fashion. We show that, when we exclude the inaccuracy measures that give rise to these dilemmas, we are left with only the quadratic inaccuracy measures. This will complete our argument for Local and Global Inaccuracy Measures. Quadratic inaccuracy measures will be the ones that allow the local and the global perspective on belief functions to be compatible with each other.

5.2 Excluding dilemmas: Completing the three arguments

An inaccuracy measure gives rise to a dilemma for an agent if the prescription to be as accurate as possible with respect to that inaccuracy measure entails two prescriptions for the agent such that she cannot satisfy both together. In sections 5.2.1, 5.2.2, and 5.2.3, we consider three dilemmas to which an inaccuracy measure may give rise. In each case, we introduce a principle to exclude such inaccuracy measures and show that Local and Global Inaccuracy Measures follow from this stipulation along with the four conditions enumerated in section 5.1. The respective dilemmas are *serious* in the following sense: (i) They are about minimizing inaccuracy, the central goal of our epistemic agents; (ii) they involve an agent's having to choose to follow either of two options or norms; (iii) there does not seem to be any principled way of ranking the two options or norms, such that one would become epistemically prior or superior to the other. Since the dilemmas below are serious in this sense, they have to be avoided, and the only manner in which this can be done is by making sure that the two options or norms never lead to different epistemic recommendations to the agent.

This—defeasible—argumentation in favour of principles by which the dilemmas may be avoided can only be defeated in either of two ways: First of all, by attacking (iii), i.e., by showing that for each of the three dilemmas there is in fact a way of ranking one option or norm over the other. E.g., if someone were to put forward a sufficiently strong argument in favour of a form of epistemic holism according to which considerations of global inaccuracy always overrule considerations of local inaccuracy, then this would defeat the seriousness of each of our three dilemmas. Secondly, by presenting yet another serious dilemma in the sense of (i)–(iii) into which an agent is led by opting for our quadratic inaccuracy measures. In that case, no inaccuracy measure *whatsoever* could protect an agent from being confronted with some serious dilemma, and the best one could hope for would be a multitude of mutually exclusive and partially defective choices of inaccuracy measures, such that each one of them would avoid some epistemic dilemmas, but none of them would avoid all. We hope that at least as things stand, our arguments are undefeated as yet.

5.2.1 Agreement on Inaccuracy

By Local and Global Comparability from section 5.1, if the function f gives rise to a legitimate local inaccuracy measure, then it gives rise to a legitimate global inaccuracy measure as well, and vice versa. However, the four conditions enumerated in section 5.1 can be shown not to exclude functions f such that: (1) the global inaccuracy measure $f(||w - b_{\text{glo}}||)$ determines the inaccuracy of a belief function at a world; (2) its counterpart local inaccuracy measure $f(|\chi_A(w) - b(A)|)$ yields an indirect way of also determining the inaccuracy of a belief function at a world by summing up the local inaccuracies of degrees of belief assigned to world-propositions in the expected manner; and yet the outcomes of the two determination procedures *differ*. Such a disagreement would give rise to a dilemma: the agent who uses both the global inaccuracy measure and its local counterpart will come to two conflicting conclusions concerning the inaccuracy of her beliefs. Of course, it might be the case that despite of the numerical disagreement some formal properties are still shared by the globally and by the indirectly locally determined inaccuracies of belief functions at a world—for instance the ordering of belief functions according to their inaccuracies. And searching for the belief function that minimizes expected inaccuracy first in a globally and then in a locally induced way might still yield one and the same output even when the globally and the locally

determined inaccuracies diverge in value for some or even all arguments. But the only way for the agent to have a guarantee that the global and the local procedure will never lead to any conflict whatsoever—and thus that the global and the local procedure always lead to the same epistemic recommendations, independently of how sensitive the agent is to the exact numerical inaccuracy values—is to postulate a convergence between the global and local way of determining the inaccuracy of any belief function at any world.

Otherwise the agent's situation would be analogous to that of a group of individuals faced with making a collective judgement that is vulnerable to the paradoxes of judgement aggregation. For instance, consider the stock example of the so-called *discursive dilemma*, the most vivid of these paradoxes.[13]

> *Discursive Dilemma* Three judges must decide whether or not to convict a defendant. By law, the defendant may be convicted if, and only if, propositions P and Q hold. The judges' judgements on P and Q and the consequence for conviction are recorded in the following table, along with the majority judgement in each case.
>
> | | P | Q | Conviction |
> |---|---|---|---|
> | Judge 1 | True | True | Yes |
> | Judge 2 | True | False | No |
> | Judge 3 | False | True | No |
> | Majority | True | True | Yes/No |
>
> Thus, while the majority of the individual judgements concerning P and Q leads to a conviction, the majority of the consequences of those judgements for conviction leads to acquittal. Which conviction consequence reflects the aggregate of the judges' judgements? This is the *discursive dilemma*.

In the discursive dilemma, there is a tension between two algorithms by which to aggregate individual judgements on two propositions by three individuals into a single judgement. The first algorithm begins by deriving the conviction consequences from the individual judgements on P and Q and then takes the majority verdict; the other begins by taking the majority verdict on each of P and Q and then derives the conviction consequence. They lead to conflicting results and it is not clear at all how to tell between them.

The agent who uses global and local inaccuracy measures that give rise to different values for the inaccuracy of her belief function faces a similar problem. She faces the problem of aggregating her various degrees of belief in various propositions into a value for the inaccuracy of this belief function as a whole. She has at her disposal two obvious measures by which to obtain this value: one is just given by *applying the global inaccuracy function itself*, the other *by summing up the relevant local inaccuracies*. If they disagree, the agent faces an irresolvable dilemma, analogous to that faced by the judges in the discursive dilemma.[14]

We exclude the possibility that gives rise to this dilemma by imposing the following condition on legitimate inaccuracy measures which yields a principled way of relating global and local inaccuracy judgements:

[13] See, for instance, [11].

[14] It would be great if we had some results which would show that even if the agent did not aggregagate local inaccuracies by summing them up but rather by applying some other numerical operation to them that would satisfy certain natural constraints, then a theorem similar to the one stated below could be derived. Unfortunately, we do not have to offer anything like that at this point; so we have to leave this for future work.

Agreement on Inaccuracy Suppose I is a legitimate local inaccuracy measure. Then, by Local Normality and Dominance, there is a strictly increasing function, $f : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ such that $I(A, w, x) = f(|\chi_A(w) - x|)$. Further, by Local and Global Comparability, $G(w, b) = f(||w - b_{\text{glo}}||)$ is a legitimate global inaccuracy measure. Then, the following must hold: If b is a belief function and w_i is a world,

$$G(w_i, b) = \sum_{j=1}^{n} I(\{w_j\}, w_i, b(\{w_j\}))$$

That is,

$$f(||w_i - b_{\text{glo}}||) = \sum_{j=1}^{n} f(|\chi_{\{w_j\}}(w_i) - b(\{w_j\})|)$$

From this, along with the four conditions stated in section 5.1, Local and Global Inaccuracy Measures follow by the following theorem:

Theorem 3 *The following two propositions are equivalent:*

(i) *f is strictly increasing and, for all belief functions, b, and worlds w_i,*

$$f(||w_i - b_{\text{glo}}||) = \sum_{j=1}^{n} f(|\chi_{\{w_j\}}(w_i) - b(\{w_j\})|)$$

(ii) *There is $\lambda \in \mathbb{R}_{>0}$ such that, for all $x \in \mathbb{R}_0^+$:*

$$f(x) = \lambda x^2$$

This theorem is proved in section 7.

5.2.2 Separability of Global Inaccuracy

In the previous section, we drew attention to a possible dilemma that results from using the legitimate local and global inaccuracy measures given by the same function f. And we ruled out the possibility by introducing Agreement on Inaccuracy. In this section, we describe another way in which an inaccuracy measure could give rise to conflicting values for the inaccuracy of a belief function at a world.

To state the problem, we introduce the following terminology: if $1 \leq j \leq n$ and $(a_1, \ldots, a_n) \in R_0^n$, then $\text{proj}_j((a_1, \ldots, a_n))$ is the projection of $(a_1, \ldots, a_n)$ onto the linear subspace that is spanned by the unit vectors that represent the worlds in $W \setminus \{w_j\}$: that is,

$$\text{proj}_j((a_1, \ldots, a_n)) = (a_1, \ldots, a_{j-1}, 0, a_{j+1}, \ldots a_n).$$

Hence, for $i \neq j$, $\text{proj}_j(w_i) = w_i$, whereas $\text{proj}_j(w_j) = (0, \ldots, 0)$.

Now, as in the previous section, suppose that f is a function that gives rise to a local inaccuracy measure I and a global inaccuracy measure G. And suppose further that our epistemic agent's global belief function is represented by the vector $(a_1, \ldots, a_n)$. Then, given a world w_i, there seem to be two ways to measure the inaccuracy of the agent's belief function at w_i that arise from combining I and G:

(1) One might simply use G: that is, the inaccuracy of $(a_1, \ldots, a_n)$ at w_i is

$$G(w_i, (a_1, \ldots, a_n))$$

(2) Or, for any world w_j with $i \neq j$, one might take the inaccuracy of $(a_1, \ldots, a_n)$ at w_i to be

$$I(\{w_j\}, w_i, a_j) + G(\text{proj}_j(w_i), \text{proj}_j((a_1, \ldots, a_n)))$$

That is, one might take the local inaccuracy of the degree of credence in proposition $\{w_j\}$ at world w_i, and add it to the global inaccuracy at w_i of the 'remainder' of $(a_1, \ldots, a_n)$ when world w_j is not considered: that is, geometrically speaking, one adds the global inaccuracy at world w_i of the belief function represented by the projection of $(a_1, \ldots, a_n)$ onto the subspace spanned by $W \backslash \{w_j\}$.

As in the previous section, the conditions listed in section 5.1 do not rule out the possibility that these two ways of measuring the inaccuracy of the agent's belief function at w_i disagree; and, if they do, a dilemma might arise for the agent. As before, we rule out the functions f that give rise to this dilemma by laying down a further principle:

Separability of Global Inaccuracy Suppose I is a legitimate local inaccuracy measure. Then, by Local Normality and Dominance, there is a strictly increasing function, $f : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ such that $I(A, w, x) = f(|\chi_A(w) - x|)$. Further, by Local and Global Comparability, $G(w, b) = f(||w - b_{\text{glo}}||)$ is a legitimate global inaccuracy measure. Then, the following must hold: for all $w_i, w_j \in W$ with $i \neq j$,

$$G(w_i, (a_1, \ldots, a_n)) = f(||w_i - (a_1, \ldots, a_n)||) =$$
$$f(|\chi_{\{w_j\}}(w_i) - a_j|) + f(||\text{proj}_j(w_i) - \text{proj}_j((a_1, \ldots, a_n))||)$$

As in the case of Agreement on Inaccuracy, this condition may be justified by noting that, if it were to fail, two legitimate ways by which an agent may determine her inaccuracy would lead to different results in at least one situation; such disagreement would lead to a situation analogous to that described in the discursive dilemma.

From Separability of Global Inaccuracy along with the four conditions stated in section 5.1, Local and Global Inaccuracy Measures follow by the following theorem:

Theorem 4 Separability of Global Inaccuracy *and* Minimum Inaccuracy *entail* Agreement on Inaccuracy (*which, in combination with the assumptions of section 5.1, yields* Local *and* Global Inaccuracy Measures*, by Theorem 3*).

The theorem is proved in the appendix again (section 7).

5.2.3 Agreement on Directed Urgency

Before we can state our final dilemma—saving the best for last—we must restrict the class of legitimate accuracy measures a little more than is done by the four conditions of section 5.1. In particular, we demand:

Continuous Differentiability

(i) If $I(A, w, x) = f(|\chi_A(w) - x|)$ is a legitimate local inaccuracy measure, then f is continuously differentiable on $\mathbb{R}_0^+$.

(ii) If $G(w, b) = g(||w - b_{\text{glo}}||)$ is a legitimate global inaccuracy measure, then g is continuously differentiable on $\mathbb{R}_0^+$.

Again, in the presence of Local and Global Comparability, we can derive (i) from (ii) and (ii) from (i). However, again, we state both conditions to avoid the mistaken impression that one is more fundamental than the other. Having said this, we will give our argument for Continuous Differentiability only in terms of (i). This is not because we derive (ii) by inferring it from (i), rather, it is because the argument for (ii) is exactly analogous and may be easily reconstructed from the argument for (i).

Suppose that f is a function that gives rise to a local inaccuracy measure. By Local Normality and Dominance, f is an increasing function. But it is clear that $f(|\chi_A(w) - x|)$ should be also a continuous function of x, and thus f should be a continuous function. After all, if $f(|\chi_A(w) - x|)$ were discontinuous as a function of x at some particular $x_0 \in \mathbb{R}_0^+$, an agent's accuracy could improve or deteriorate dramatically by an arbitrarily small change to her degree of credence in the neighbourhood of x_0. Thus, f must be continuous.

However, Continuous Differentiability demands something further. It demands that f be continuously differentiable on $\mathbb{R}_0^+$. To justify this claim, consider again the notion of expected local inaccuracy, which we introduced in section 2. Given a belief function b, propositions $A, E \subseteq W$, a degree of credence x in A, and a local inaccuracy measure I, we have interpreted $\text{LExp}_b(I, A, E, x)$ as the expected value of the inaccuracy of x by the lights of b, with respect to I, and over the epistemically possible worlds $w \in E$.

Now, if I were a legitimate local inaccuracy function, the function $\text{LExp}_b(I, A, E, x)$ would provide not only a measure of the expected inaccuracy of x in A by the lights of b, with respect to I, and over E. It would provide also a means by which to measure the *urgency* and *direction*—in short, the *directed urgency*—with which an agent for whom E is the set of epistemically possible worlds ought to change her degree of credence in A by the lights of some belief function b. Clearly, this measure would be provided by the derivative of $\text{LExp}_b(I, A, E, x)$ with respect to x, were this derivative to exist. Wherever it is defined, the absolute value of the function $\frac{d}{dx}\text{LExp}_b(I, A, E, x)$ measures the rate at which the expected inaccuracy of x by the lights of b is changing (the slope of the tangent): thus, by the lights of b, it would be more urgent to change degree of credence r in proposition A than to change degree of credence s in the same proposition just in case the absolute value of $\frac{d}{dx}\text{LExp}_b(I, A, E, x)$ evaluated at r were greater than that the absolute value of the same derivative evaluated at s.[15] Furthermore, if the sign of $\frac{d}{dx}\text{LExp}_b(I, A, E, x)$ evaluated at r differed from the sign of the same derivative evaluated at s, then the degree of credence r in proposition A ought to be increased when the degree of credence s in the same proposition ought to be decreased, or vice versa. Indeed, *only* the derivative of $\text{LExp}_b(I, A, E, x)$ could supply the agent with this sort of information. Thus, if I is to be a legitimate local inaccuracy measure, then $\frac{d}{dx}\text{LExp}_b(I, A, E, x)$ should be defined on $\mathbb{R}_0^+$, since there *ought* to be a measure of directed urgency that an agent can use to determine a local recommendation of where to go epistemically and, as it were, how quickly she should move. If Ought-Can by itself does not support this claim sufficiently, then hopefully it does so in conjunction with the

[15] Gibbard, too, interprets the derivative of the inaccuracy measure as a measure of the urgency of updating ([3]).

geometrical framework that we presuppose. In any case, it is straightforward to show that, if $\frac{d}{dx}\text{LExp}_b(I, A, E, x)$ is defined on $\mathbb{R}_0^+$, for every belief function b, propositions $A, E \subseteq W$, and degree of credence x, then $I(A, w, x)$ must be differentiable on this domain as well. Thus, if $I(A, w, x) = f(|\chi_A(w) - x|)$, then f must be differentiable on $\mathbb{R}_0^+$. Of course, this argument requires some amount of idealization, since for all "real world" cases in which an agent ought to determine the directed urgency of a belief change, computing small real-valued differences rather than infinitesimal ones should be sufficient.[16] But if the agent wants to be certain about this, whatever the level of precision required, then f ought to be differentiable.

What's more, just as we wish our local inaccuracy measure to be a continuous function of the degree of credence whose inaccuracy it is measuring, we would like our directed urgency measure to be a continuous function of the degree of credence the directed urgency of the change to which it is measuring: we would not wish the urgency and direction by which an agent should change her degree of credence to change drastically after an arbitrarily small shift in her degree of credence. From which the local part of Continuous Differentiability—namely, (i) from above—follows.[17]

As we mentioned above, our argument in favour of the global part of Continuous Differentiability—namely, (ii)—is analogous to the argument just given in favour of the local part. Focusing just on propositions A which are singleton propositions of the form $\{w_j\}$ for $j = 1, ..., n$, we say that if $G(w, b) = g(||w - b_{\text{glo}}||)$ is a legitimate global inaccuracy measure, then it also *ought* to give rise to a measure of the urgency with which an agent must change her degree of belief such a singleton proposition. Given a belief function b, then the urgency to change the degree of belief x in proposition $\{w_j\}$ for an agent with belief function b' by the lights of b would be given by

$$\frac{d}{dx}\text{GExp}_b(G, E, (b'(\{w_1\}), \ldots, b'(\{w_{j-1}\}), x, b'(\{w_{j+1}\}), \ldots, b'(\{w_n\}))$$

if this derivative were to exist (and which, given Continuous Differentiability, indeed exists). And, as in the local case, its sign would indicate the direction in which the change must occur. Thus, we interpret the derivatives both of expected global inaccuracy and expected local inaccuracy as measures of directed urgency. Granted this, we claim that these two measures ought to agree on singleton propositions whenever the global and local inaccuracy measures from which they arise are based on the same strictly increasing and continuously differentiable function, f. This is the content of Agreement on Directed Urgency.

> **Agreement on Directed Urgency** If $I(A, w, x) = f(|\chi_A(w) - x|)$ and $G(w, b) = f(||w - b_{\text{glo}}||)$ are legitimate local and global inaccuracy measures respectively, and if f is differentiable, then, for all belief functions b and b' and all worlds $w_j \in W$,
>
> $\frac{d}{dx}\text{LExp}_b(I, \{w_j\}, E, x) =$
> $\frac{d}{dx}\text{GExp}_b(G, E, (b'(\{w_1\}), \ldots, b'(\{w_{j-1}\}), x, b'(\{w_{j+1}\}), \ldots, b'(\{w_n\}))$

Suppose this condition were not to hold. Then the agent who employed these measures of inaccuracy, and the measures of directed urgency to which they give rise, would be left with

[16] Alan Hájek and Kenny Easwaran pointed this out to us.

[17] If this little transcendental argument is not convincing enough, here is a much more mundane thought: let us restrict ourselves just to 'geometrically nice' local and global inaccuracy measures. But in order to be 'geometrically nice', these measures will have to be given by continuously differentiable functions.

a dilemma. Where (1) the local measure of directed urgency for $\{w_j\}$ differed from (2) the global measure of directed urgency for the coordinate w_j, she would be unable to determine the urgency with which she must update her belief in order to minimize her expected inaccuracy, and perhaps even the direction in which her update should proceed; the two measures would give conflicting values between which she could not choose in a principled way.

Together with Continuous Differentiability and the four conditions stated in section 5.1, Agreement on Directed Urgency entails Local and Global Inaccuracy Measures by means of the following theorem:

Theorem 5 *The following two propositions are equivalent:*

(i) $f : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ *is strictly increasing and continuously differentiable,* $f(0) = 0$*, and, for all belief functions* $b \in \mathrm{Bel}(W)$*, all* $w_j \in W$*, and all* $a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_n \in \mathbb{R}_0^+$*:*

$$\begin{aligned} & \frac{d}{dx} \sum_{w_i \in W} b(\{w_i\}) f(|\chi_{\{w_j\}}(w_i) - x|) \\ = \; & \frac{d}{dx} \sum_{w_i \in W} b(\{w_i\}) f(||w_i - (a_1, \ldots, a_{j-1}, x, a_{j+1}, \ldots, a_n)||) \end{aligned}$$

(ii) *There is* $\lambda \in \mathbb{R}_{>0}$*, such that, for all* $x \in \mathbb{R}_0^+$*:*

$$f(x) = \lambda x^2$$

The proof is given in section 7.

6 A look ahead to the sequel and to future work

This concludes our argument for Local and Global Inaccuracy Measures and, with it, our defence of the four mathematically precise versions of the Accuracy norm introduced in section 2, that is, the synchronic local and global versions and the diachronic local and global versions, now supplied with the right inaccuracy measures.

In the sequel to this paper, we investigate the consequences of these norms. Before considering some open questions about our approach, we report the results of that investigation:

(1) From the synchronic local version of Accuracy, we derive **Probabilism**.

(2) From the diachronic local version of Accuracy, we derive **Conditionalization**.

(3) From a related, but much stronger norm, we derive **Uniform Distribution**.

(4) We show that, in the situations normally assumed to be covered by Jeffrey's updating rule, there is no updating rule that satisfies the diachronic *local* version of Accuracy. However, the diachronic *global* version can be satisfied. We show that Jeffrey's updating rule sometimes violates this diachronic version of the norm, unless the so-called postulate of rigidity is not required by fiat, and we describe the alternative updating rule that satisfies it.

We finish with five open questions that point towards future research:

- How can the approach taken in this paper be extended to the case of an *infinite* set of worlds; in particular, how can it be extended to the case of a non-denumerable set of worlds? What role does countable additivity play in such extensions?
- Which of our conclusions depend essentially on our geometrical background machinery being *Euclidean*? Which conclusions can be drawn in a *non-Euclidean* setting?
- How can the theory be translated into a more abstract system of *axiomatic* constraints on both belief update and the geometrical background system (along the lines of Joyce, but also Greaves & Wallace [4])? How robust are our results if, at the relevant places, *summing up* of inaccuracies gets replaced by any numerical operation that satisfies some set of plausible constraints?
- How does the theory relate to theories of *verisimilitude* in which relational truthlikeness is analysed in terms of the expected degree of truthlikeness (as in Niiniluoto's [10] theory of estimated truthlikeness)?
- How does our theory of expected inaccuracy work in a framework that permits partial beliefs in self-locating propositions? Such an application would cover the well-known Sleeping Beauty problem. See [8] for a related attempt, where it is assumed that quadratic inaccuracy measures provide the only legitimate scoring rule.
- Could a variant of our theory of expected inaccuracy justify probabilistic methods of *judgement aggregation* or *amalgamation*?

(To be continued by the sequel, 'An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy'.)

7 Appendix: Proofs of Theorems 3, 4, and 5

In this section, we prove the three theorems used to argue for Local and Global Inaccuracy Measures in section 5.

7.1 Proof of Theorem 3

It will suffice to show that the following two propositions are equivalent:

(i') g is strictly increasing and, for all belief functions, b, and worlds $w_i \in W$,

$$g(||w_i - b_{\text{glo}}||^2) = \sum_{j=1}^{n} g(|\chi_{\{w_j\}}(w_i) - b(\{w_j\})|^2)$$

(ii') There is $\lambda \in \mathbb{R}_{>0}$, such that, for all $x \in \mathbb{R}_0^+$:

$$g(x) = \lambda x$$

Suppose (i') and (ii') are equivalent. Then, if (i), then $g(x) = f(\sqrt{x})$ satisfies (i') and thus (ii'), so $f(\sqrt{x}) = \lambda x$, which gives (ii). Similarly, if (ii), then $g(x) = f(\sqrt{x})$ satisfies (ii') and thus (i'), so f satisfies (i). Thus, we will prove the equivalence of (i') and (ii').

First, we show that (ii') implies (i'). Thus, suppose $g(x) = \lambda x$. Clearly, g is strictly increasing. Now suppose that b is a belief function and w_i a world; then

$$\begin{aligned}\lambda||w_i - b_{\text{glo}}||^2 &= \lambda\left[b(\{w_1\})^2 + \cdots + (1 - b(\{w_i\}))^2 + \cdots + b(\{w_n\})\right] \\ &= \lambda\sum_{j=1}^{n}|\chi_{\{w_j\}}(w_i) - b(\{w_j\})|^2\end{aligned}$$

Thus, (ii') implies (i').

Now, we show that (i') implies (ii'). Our strategy is to show that, from (i'), it follows that, for any $x, y \in \mathbb{R}_0^+$, $g(x+y) = g(x) + g(y)$. Since, by (i'), g is also strictly increasing, (ii') follows by Cauchy's classical result that all monotone additive functions on $\mathbb{R}$ are linear on $\mathbb{R}$.[18]

Thus, suppose $x, y \in \mathbb{R}_0^+$. Then let $b(\{w_1\}) = \cdots = b(\{w_{n-2}\}) = 0$, $b(\{w_{n-1}\}) = \sqrt{x}$, and $b(\{w_n\}) = 1 - \sqrt{y}$. Then, by (i'),

$$g(||w_n - (0, \ldots, 0, \sqrt{x}, 1 - \sqrt{y})||^2) = g(|0 - \sqrt{x}|^2) + g(|1 - (1 - \sqrt{y})|^2)$$

From this, we have $g(x+y) = g(x) + g(y)$, as required. □

7.2 Proof of Theorem 4

Consider the following statements:

(i) f is strictly increasing, $f(0) = 0$, and, for all belief functions b and all worlds w_i and w_j such that $i \neq j$,

$$f(||w_i - b_{\text{glo}}||) = f(|\chi_{\{w_j\}}(w_i) - b(\{w_j\})|) + f(||\text{proj}_j(w_i) - \text{proj}_j((a_1, \ldots, a_n))||)$$

(ii) f is strictly increasing and, for all belief functions, b, and worlds w_i,

$$f(||w_i - b_{\text{glo}}||) = \sum_{j=1}^{n} f(|\chi_{\{w_j\}}(w_i) - b(\{w_j\})|)$$

We must show that (i) entails (ii). Suppose (i) holds. Then, by repeatedly separating the local inaccuracy measure from the global one, we obtain:

$$f(||w_i - b_{\text{glo}}||) = \textstyle\sum_{j=1}^{n} f(|\chi_{\{w_j\}}(w_i) - b(\{w_j\})|) \; + \\ f(||\text{proj}_i(w_i) - \text{proj}_n(\text{proj}_{n-1}(\ldots \text{proj}_2(\text{proj}_1((a_1, \ldots, a_n))) \ldots))||)$$

But

$$f(||\text{proj}_i(w_i) - \text{proj}_n(\text{proj}_{n-1}(\ldots \text{proj}_2(\text{proj}_1((a_1, \ldots, a_n))) \ldots))||) = f(0) = 0$$

by Minimum Inaccuracy. Thus, (ii), as required. □

[18] In fact, we need a slightly different version, which states that all monotone additive functions on $\mathbb{R}_0^+$ are linear on $\mathbb{R}_0^+$. Suppose $g : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ is additive on $\mathbb{R}_0^+$. Then define $g' : \mathbb{R} \to \mathbb{R}$ as follows:

$$g'(x) = \begin{cases} g(x) & \text{if } 0 \leq x \\ -g(-x) & \text{if } x < 0 \end{cases}$$

Then it is easy to show that g' is additive and monotone on $\mathbb{R}$ if g is additive and monotone on $\mathbb{R}_0^+$. Thus, g' satisfies the hypotheses of Cauchy's result. For the proof of Cauchy's result, see [1].

7.3 Proof of Theorem 5

For reasons analogous to those given in the proof of Theorem 3, it will suffice to show that the following two statements are equivalent:

(i') $g : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ is strictly increasing and differentiable, $g(0) = 0$, and, for all belief functions $b \in \mathrm{Bel}(W)$ and $w_j \in W$:

$$\begin{aligned} & \frac{d}{dx} \sum_{w_i \in W} b(\{w_i\}) g(|\chi_{\{w_j\}}(w_i) - x|^2) \\ = \; & \frac{d}{dx} \sum_{w_i \in W} b(\{w_i\}) g(||w_i - (a_1, \ldots, a_{j-1}, x, a_{j+1}, \ldots, a_n)||^2) \end{aligned}$$

(ii') There is $\lambda \in \mathbb{R}_{>0}$, such that, for all $x \in \mathbb{R}_0^+$:

$$g(x) = \lambda x$$

First, we prove that (ii') implies (i'). If $\lambda \in \mathbb{R}_{>0}$ and $g(x) = \lambda x$, then g is certainly strictly increasing and differentiable: indeed, $g'(x) = \lambda$. Thus, by straightforward differentiation and direct calculation, if $a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_n \in \mathbb{R}_0^+$, then

$$\begin{aligned} & \frac{d}{dx} \sum_{w_i \in W} b(\{w_i\}) g(|\chi_{\{w_j\}}(w_i) - x|^2) \\ = \; & \sum_{w_i \in W} b(\{w_i\}) [-2\lambda \chi_{\{w_j\}}(w_i) + 2\lambda x] \\ = \; & \frac{d}{dx} \sum_{w_i \in W} b(\{w_i\}) g(||w_i - (a_1, \ldots, a_{j-1}, x, a_{j+1}, \ldots, a_n)||^2) \end{aligned}$$

as required.

Next, we prove that (i') implies (ii'). Our strategy will be to show that (i') implies that g' is constant. This will suffice, since, by (i'), g is increasing, so $\lambda > 0$. We do this in two stages. First, we prove that, on the assumption of (i'), $g'(1 + x) = g'(1)$ on $\mathbb{R}_0^+$ and then we prove that, on the assumption of (i'), $g'(x) = g'(x + 1)$ on $\mathbb{R}_0^+$.

Thus, suppose (i') holds. We wish to show that $g'(x + 1) = g'(1)$ on $\mathbb{R}_0^+$. So, suppose $a \in \mathbb{R}_0^+$. Then let $j = 1$, $a_2 = \sqrt{a}$, and $a_3 = \ldots = a_n = 0$. Then, by (i'),

$$\begin{aligned} & \frac{d}{dx} \sum_{w_i \in W} b(\{w_i\}) g(|\chi_{\{w_1\}}(w_i) - x|^2) \\ = \; & \frac{d}{dx} \sum_{w_i \in W} b(\{w_i\}) g(||w_i - (x, \sqrt{a}, 0, \ldots, 0)||^2) \end{aligned}$$

But

$$\begin{aligned} & \frac{d}{dx} \sum_{w_i \in W} b(\{w_i\}) g(|\chi_{\{w_1\}}(w_i) - x|^2) \\ = \; & \frac{d}{dx} \left[b(\{w_1\}) g((1 - x)^2) + b(\{w_2\}) g(x^2) + \ldots + b(\{w_n\}) g(x^2) \right] \\ = \; & -2(1 - x) b(\{w_1\}) g'((1 - x)^2) + 2x b(\{w_2\}) g'(x^2) + \ldots + 2x b(\{w_n\}) g'(x^2) \end{aligned}$$

and

$$\begin{aligned}
& \frac{d}{dx} \sum_{w_i \in W} b(\{w_i\}) g(||w_i - (x, \sqrt{a}, 0, \ldots, 0)||^2) \\
= \; & \frac{d}{dx}[b(\{w_1\}) g((1-x)^2 + a) + b(\{w_2\}) g(x^2 + (1-\sqrt{a})^2) + \\
& b(\{w_3\}) g(x^2 + a + 1) + \ldots + b(\{w_n\}) g(x^2 + a + 1)] \\
= \; & -2(1-x) b(\{w_1\}) g'((1-x)^2 + a) + 2x b(\{w_2\}) g'(x^2 + (1-\sqrt{a})^2) + \\
& 2x b(\{w_3\}) g'(x^2 + a + 1) + \ldots + 2x b(\{w_n\}) g'(x^2 + a + 1)
\end{aligned}$$

Thus, taking $x = 0$, we have

$$-2b(\{w_1\}) g'(1) = -2b(\{w_1\}) g'(1 + a)$$

which gives $g'(1) = g'(1 + a)$, as required.

Now we wish to show that $g'(x) = g'(x+1)$ on $\mathbb{R}_0^+$. Thus, suppose $a \in \mathbb{R}_0^+$. Then let b be a belief function in $\mathrm{Bel}(W)$ such that $b(\{w_2\}) = 1$ and $b(\{w_1\}) = b(\{w_3\}) = \ldots = b(\{w_n\}) = 0$; and let $a_2 = \ldots = a_n = 0$. Then, by (i'),

$$\frac{d}{dx}\, g(|\chi_{\{w_1\}}(w_2) - x|^2) = \frac{d}{dx}\, g(||w_2 - (x, 0, 0, \ldots, 0)||^2),$$

which gives $g'(x^2) = g'(x^2 + 1)$ for $x \in \mathbb{R}_0^+$ such that $x \neq 0$ and thus $g'(x) = g'(x + 1)$ for $x \in \mathbb{R}_0^+$ such that $x \neq 0$, since x^2 is bijective between $\mathbb{R}_0^+$ and $\mathbb{R}_0^+$. Thus, since g' is continuous on $\mathbb{R}_0^+$, $g'(x) = g'(x + 1)$ for $x \in \mathbb{R}_0^+$. Thus, for all $x \in \mathbb{R}_0^+$, $g'(x) = g'(x + 1) = g'(1)$, so g' is constant on $\mathbb{R}_0^+$. This completes our proof. □

Acknowledgements: We would like to thank F. Arntzenius, F. Dietrich, K. Easwaran, B. Fitelson (and his Berkeley reading group), A. Hájek, L. Horsten, F. Huber, J. Joyce, T. Kuipers, W. Myrvold, S. Okasha, G. Schurz, T. Seidenfeld, R. Williams, J. Williamson, and B. van Fraassen for their comments on earlier versions of this paper. Hannes Leitgeb would like to thank the Leverhulme Trust and the Alexander von Humboldt Foundation for their generous support of this work. Richard Pettigrew would like to thank the British Academy with whom he was a postdoctoral fellow during work on this paper.

References

[1] J. Aczél and J. Dhombres. *Functional Equations in Several Variables*. Cambridge University Press, Cambridge, 1989.

[2] G. W. Brier. Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review*, 78:1–3, 1950.

[3] A. Gibbard. Rational Credence and the Value of Truth. In T. Gendler and J. Hawthorne, editors, *Oxford Studies in Epistemology*, volume 2, pages 143–164. Oxford University Press, Oxford, 2008.

[4] H. Greaves and D. Wallace. Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind*, 115(459):607–632, 2006.

[5] F. Huber. The Consistency Argument for Ranking Functions. *Studia Logica*, 82:1–28, 2006.

[6] J. M. Joyce. A Nonpragmatic Vindication of Probabilism. *Philosophy of Science*, 65(4):575–603, 1998.

[7] J. M. Joyce. Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. In F. Huber and C. Schmidt-Petri, editors, *Degrees of Belief*, pages 263–297. Synthese Library, volume 342, Springer.

[8] B. Kierland and B. Monton. Minimizing Inaccuracy for Self-Locating Beliefs. *Philosophy and Phenomenological Research*, 70(2):384–395, 1999.

[9] D. W. Miller. Popper's Qualitative Theory of Verisimilitude. *The British Journal for the Philosophy of Science*, 25:166–177, 1974.

[10] I. Niiniluoto. *Truthlikeness*. Reidel, Dordrecht, 1987.

[11] P. Pettit. Deliberative Democracy and the Discursive Dilemma. *Philosophical Issues (Supp. Noûs)*, 11:268–299, 2001.

[12] K. R. Popper. *The Logic of Scientific Discovery, Revised Edition*. Hutchinson, London, 1968.

[13] K. R. Popper. *Objective Knowledge. An Evolutionary Approach*. Clarendon Press, Oxford, 1972.

[14] R. D. Rosenkrantz. *Foundations and Applications of Inductive Probability*. Ridgeview Press, Atascadero, CA, 1981.

[15] R. Selten. Axiomatic Characterization of the Quadratic Scoring Rule. *Experimental Economics*, 1(1):43–61, 1998.

[16] P. Tichý. On Popper's Definition of Verisimilitude. *The British Journal for the Philosophy of Science*, 25:155–160, 1974.

[17] J. Williamson. Objective Bayesianism, Bayesian Conditionalization, and Voluntarism. *Synthese*, forthcoming.