On justifying an account of moral goodness to each individual: contractualism, utilitarianism, and prioritarianism

Richard Pettigrew
Richard.Pettigrew@bristol.ac.uk

February 2, 2024

Many welfarists—be they utilitarians, prioritarians, or something else—wish to assign to each possible state of the world a numerical value that measures something like its moral goodness. Using this number, they can compare the moral goodness of any two states of the world; and perhaps they can use also this quantity to determine the morally right actions to choose in a situation of uncertainty—perhaps the morally right action is the one that maximizes moral goodness in expectation from the point of view of the chooser's subjective probabilities.

How, then, are we to determine this quantity? How should we measure the moral goodness of a state of the world? For the welfarist, it is a function of the levels of well-being of each morally relevant individual who inhabits that world. These individuals might include non-humans as well as humans; they might individuals extended over their whole lifetime, or instead temporal parts of individuals that span some shorter part of their life. I take no position on either of these questions here. Formally, we need only assume that there is a population of size $n > 0$; we assume this population is fixed, so that it contains exactly the same individuals at each state of the world; and for each individual i within that population, there is a numerical measure u_i of their well-being at that state of the world. (Later, we will return to the scale on which their well-being is measured.)

Given this, we can state the utilitarian measure of moral goodness:

Utilitarianism.

$$U(u_1, \ldots, u_n) = \frac{u_1 + \ldots + u_n}{n}$$

(Since we are working with a fixed population, the difference between average and total utilitarianism is not relevant.)

For the utilitarian, increasing an individual's well-being by a given amount, while holding everyone's fixed, increases the moral goodness of the state by the same amount regardless of how much well-being that individual has already. For the prioritarian, on the other hand, increasing someone's well-being by a set amount adds less to the moral goodness of the world the greater their prior level of well-being (Parfit, 2000, 2012; Adler & Holtug, 2019). So, we can state the prioritarian measure of moral goodness as follows:

Prioritarianism$_g$.

$$P_g(u_1, \ldots, u_n) = \frac{g(u_1) + \ldots + g(u_n)}{n}$$

where g is a strictly increasing and strictly concave function, so that $g(a + \varepsilon) - g(a) > g(b + \varepsilon) - g(b)$ whenever $a < b$. Different choices of g give different versions of prioritarianism.

Finally, we state a slight variant on standard prioritarianism, which we call expected equally distributed equivalent (or EEDE) prioritarianism (Fleurbaey, 2010). This takes the prioritarian's measure and applies to it the inverse of the concave function that was used to generate it. An important consequence of this is that the moral goodness of a state is the level of well-being such that, if everyone were to have it, that would be the moral goodness of the state.

EEDE$_g$.

$$E_g(u_1, \ldots, u_n) = g^{-1}\left(\frac{g(u_1) + \ldots + g(u_n)}{n}\right)$$

(Note: $E_g(u, \ldots, u) = g^{-1}(g(u)) = u$. Also note: E_g is well-defined, since g is strictly increasing.)

How are we to motivate the utilitarian's measure of moral goodness, or the prioritarian's, or the EEDEist's? On the standard approach, we axiomatise them. That is, we lay down a set of axioms that govern the ordering of states of the world by their moral goodness and we prove that only the orderings that agree with our measure of moral goodness satisfy them all; or we lay down a set of axioms that govern the preferences of the social chooser over uncertain prospects, and show that they must be combining their subjective probabilities with our favoured measure of moral goodness to set those preferences. So Harsanyi (1955) axiomatizes the utilitarianism in the second way, McCarthy (2008) axiomatizes prioritarianism in the first way, and Fleurbaey (2010) introduces and axiomatizes the EEDE approach in the second way. In this paper, I want to explore an alternative approach.

1 Justifying the measure of moral goodness to each

The central idea behind this alternative approach is contractualist: it says that a legitimate measure of moral goodness is one that could be justified to each member of the population in question. Contractualists tend to focus on the legitimacy of decisions made on behalf of the population, and they say that such a decision is legitimate if it can be justified to each member of the population. Here is Scanlon:

> An act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced, general agreement. (Scanlon, 2000, 153)

Our approach here is only slightly different: here, it is the measure of moral goodness that must be justified to each, not the decisions made using it, though of course those might be justified to each indirectly by appealing to this measure of moral goodness and noting that it can be justified to each. So we seek account of moral goodness that none of the individuals in question could reasonably reject as a basis for measuring the moral value of a state of the world.

How, then, do we justify a measure of moral goodness to each individual? Each individual i would prefer the social chooser to use i's own measure of well-being to make choices that will affect her, just as i would prefer them to use i's own preference ordering for the choice. So, i would prefer the social chooser to use u_i. But they recognise that such a choice is not justifiable to others in the population; some of the others could reasonably reject that. They recognise that they must reach some sort of compromise with those others. But of course some compromises are more reasonable than others; and some are better justifiable to a given member of the population than others. Each member recognises that the social chooser's measure of moral goodness is going to have to deviate from the well-being function of at least some of the members of the population, providing some of them have different levels of well-being. But we can nonetheless justify it to each of them if it doesn't deviate more than is necessary, and if the deviations from each member are given equal weighting in whatever process we use to determine it.

Here's a proposal for how to satisfy those conditions. First, we need a measure of the distance $\mathfrak{d}$ from a proposed compromise to an individual's level of well-being. If u is the proposed compromise and u_i is the individual's well-being, $\mathfrak{d}(u \| u_i)$ is the distance from u to u_i. In fact, as we'll see below, we don't want to restrict to functions that have all the standard mathematical properties of a distance, such as symmetry and the triangle inequality, so instead we assume only that $\mathfrak{d}$ is a divergence: that is, $\mathfrak{d}(u \| u) = 0$ for all u in the domain of $\mathfrak{d}$, and $\mathfrak{d}(u \| v) > 0$ whenever $u \neq v$. (We'll return to the domain of these functions below.) Given such a function d, a natural way to ensure that, for each individual i, the divergence $\mathfrak{d}(u \| u_i)$ from a candidate measure of moral goodness u to their level of well-being u_i is given due weight is to say that the moral goodness is the candidate compromise that minimizes the sum of divergences from it to the individuals' levels of well-being. That is, the moral goodness of a state in which individual 1 has well-being u_1, individual 2 has well-being u_2, and so on, is

$$\arg\min_u \sum_{i=1}^{n} \mathfrak{d}(u \| u_i)$$

at least where such a unique minimum exists.

That's the proposal. It is inspired by a suggestion from the judgment aggregation literature. It appears first in work by Konieczny & Pino Pérez (1998, 1999) on merging databases; it is then extended by Osherson & Vardi (2006), Predd et al. (2008), and Pettigrew (2019, 2022) in their treatment of aggregating sets of probabilistic judgments. The idea is that the set of probability judgments that aggregates the probability judgments of some set of individuals is the one whose total distance to the probability judgments of the individuals is minimal.

In what follows, I will list some well-known divergences and families of divergences; I'll show the measures of moral goodness to which they give rise when combined with this proposal; and I'll end by adapting two existing axiomatic characterizations of such divergences to pin down divergences that give us utilitarianism and some natural versions of EEDE, respectively.

2 Measuring divergence between individual and compromise

In this section, I describe some divergences. These are all popular divergences in the wide variety of disciplines that appeal to such measures. I will define them on $S \times S$, where S is either the set of all real numbers $\mathbb{R}$ or the set of positive real numbers $\mathbb{R}_+$.

Quadratic.[1] $\mathfrak{q}(u \| v) = (u - v)^2 \quad (u, v \in \mathbb{R})$

I-divergence.[2] $\mathfrak{k}(u \| v) = u \log \frac{u}{v} - u + v \quad (u, v \in \mathbb{R}_+)$

[1] This is sometimes known as the *squared Euclidean distance* and it is used in the least squares method in statistics.

[2] This is sometimes known as the *generalized Kullback-Leibler divergence*, where the Kullback-Leibler divergence measures distance from one probability distribution to another, and is sometimes known as the relative entropy (Kullback & Leibler, 1951). The generalization is due to Imre Csiszár (1991).

Itakura-Saito.[3] $\mathfrak{i}(u\|v) = \log \frac{v}{u} + \frac{u}{v} - 1 \quad (u, v \in \mathbb{R}_+)$

Power$_\alpha$. $\mathfrak{p}_\alpha(u\|v) = \frac{1}{\alpha}(v^\alpha - u^\alpha) + v^{\alpha-1}(u - v) \quad (u, v \in \mathbb{R}_+, \alpha < 1, \alpha \neq 0)$

Each of these four divergences belongs to the family of Bregman divergences, which is defined as follows. Suppose $S = \mathbb{R}$ or $S = \mathbb{R}_+$. Then, if φ is a strictly convex, differentiable function on S, then

Bregman$_\varphi$.[4] $\mathfrak{d}_\varphi(u\|v) = \varphi(u) - \varphi(v) - \varphi'(v)(u - v) \quad (u, v \in S)$

So, to calculate $\mathfrak{d}_\varphi(u\|v)$, we take the linear function that gives the tangent to φ at v, extend it to u and take its value there, and subtract that from the value of φ at u. This is illustrated in Figure 1.

The following table gives the strictly convex, differentiable functions that generate the four divergences given above:

Divergence	$\varphi(x)$
Quadratic	x^2
I-divergence	$x \log x$
Itakura-Saito	$-\log x$
Power$_\alpha$	$-\frac{1}{\alpha}x^\alpha$

Next, I define the α-divergences:

α-divergence. $\mathfrak{a}_\alpha(u\|v) = \frac{v}{\alpha} - \frac{u}{\alpha-1} + \frac{u^\alpha}{v^{\alpha-1}\alpha(\alpha-1)} \quad (u, v \in \mathbb{R}_+, \alpha \neq 0, 1)$

Each divergence in this set is a member of the family of f-divergences introduced by Imre Csiszár (1963). These divergences have the form $\mathfrak{d}(u\|v) = vf\left(\frac{u}{v}\right)$, for a continuously differentiable, strictly convex function f on $\mathbb{R}_+$. The α-divergence is the f-divergence generated by the function $f_\alpha(x) = \frac{x^\alpha - \alpha x + \alpha - 1}{\alpha(\alpha-1)}$. The so-called Hellinger distance is the α-divergence for $\alpha = \frac{1}{2}$ (Hellinger, 1909).

3 Minimizing the total divergence to the individuals

In this section, I work through the divergences we met in the previous section and say what happens when we minimize total divergence from compromise to the individuals. In each case, the fact is an easy consequence of calculus, since the divergences are all differentiable in their first argument.

[3] This is due to Itakura & Saito (1968).

[4] This definition is due to Bregman (1967).

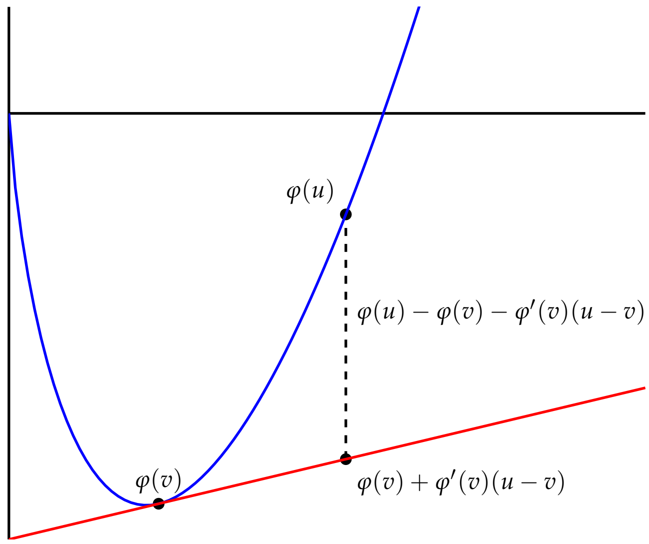


Figure 1: The blue line graphs the function φ, the red line graphs the linear function that gives the tangent to φ at v, and the length of the dashed line gives the difference between the value at u of φ, and the value at u of that linear function; that difference is the Bregman divergence from u to v.

Proposition 1.

$$
\begin{array}{c|rcl}
\mathfrak{q} & \arg\min_u \sum_i \mathfrak{q}(u \| u_i) & = & \frac{u_1 + \ldots + u_n}{n} \\
\mathfrak{k} & \arg\min_u \sum_i \mathfrak{k}(u \| u_i) & = & \sqrt[n]{u_1 \ldots u_n} \\
\mathfrak{i} & \arg\min_u \sum_i \mathfrak{i}(u \| u_i) & = & \frac{n}{\frac{1}{u_1} + \ldots + \frac{1}{u_n}} \\
\mathfrak{p}_\alpha & \arg\min_u \sum_i \mathfrak{p}_\alpha(u \| u_i) & = & \left(\frac{u_1^{\alpha-1} + \ldots + u_n^{\alpha-1}}{n}\right)^{\frac{1}{\alpha-1}} \quad \textit{where } \alpha < 1, \alpha \neq 0. \\
\mathfrak{d}_\varphi & \arg\min_u \sum_i \mathfrak{q}(u \| u_i) & = & g^{-1}\left(\frac{g(u_1) + \ldots + g(u_n)}{n}\right) \quad \textit{where } g = \varphi'. \\
\mathfrak{a}_\alpha & \arg\min_u \sum_i \mathfrak{a}_\alpha(u \| u_i) & = & \left(\frac{n}{\frac{1}{u_1^{\alpha-1}} + \ldots + \frac{1}{u_n^{\alpha-1}}}\right)^{\frac{1}{\alpha-1}} = \left(\frac{u_1^{1-\alpha} + \ldots + u_n^{1-\alpha}}{n}\right)^{\frac{1}{1-\alpha}}
\end{array}
$$

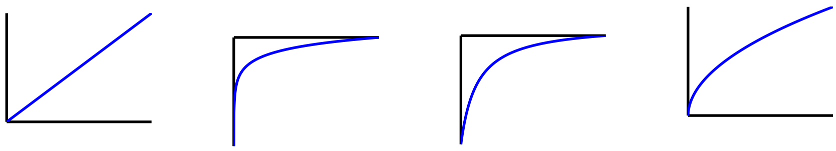
The first thing to note is that:

(i) the quadratic divergence gives utilitarianism;[5]

(ii) the I-divergence gives EEDE prioritarianism with $g_{\mathfrak{k}}(x) = \log x$;[6]

[5]Of course, this has the form of EEDE definitions of the moral goodness of a state, with $g(x) = x$; but since $g(x) = x$ isn't strictly concave, it is not a version of EEDE.

[6]This is because $\sqrt[n]{u_1 \ldots u_n} = e^{\frac{\log u_1 + \ldots + \log u_n}{n}}$.



(a) $g(x) = x$ (b) $g(x) = \log x$ (c) $g(x) = -\frac{1}{x}$ (d) $g(x) = x^{\frac{1}{2}}$

Figure 2: The linear function that generates utilitarianism (a), and concave functions that generate different versions of EEDE prioritarianism (b)-(d).

(iii) the Itakura-Saito divergence gives EEDE prioritarianism with $g_{\mathfrak{i}}(x) = -\frac{1}{x}$;

(iv) the power$_\alpha$ divergence gives EEDE prioritarianism with $g_{\mathfrak{p}_\alpha}(x) = -x^{\alpha-1}$, for $\alpha < 1, \alpha \neq 0$;

(v) the Bregman$_\varphi$ divergence gives EEDE prioritarianism with $g_{\mathfrak{d}_\varphi}(x) = \varphi'(x)$, whenever φ' is strictly concave (φ' is guaranteed to be strictly increasing because φ is strictly convex, and so $\varphi'' > 0$).

(vi) the α-divergence gives EEDE prioritarianism with $g_{\mathfrak{a}_\alpha}(x) = x^{1-\alpha}$, for $0 < \alpha < 1$.

These functions are plotted in Figure 2. To illustrate the differences between the versions of EEDE prioritarianism to which $g_{\mathfrak{k}}$, $g_{\mathfrak{i}}$, and $g_{\mathfrak{p}_\alpha}$ give rise, consider populations with two individuals; write (u_1, u_2) for the welfare distribution in which the first has well-being level u_1 and the second has u_2; and write $<_U$ for the utilitarian ordering of worlds and $<_{E_g}$ for the EEDE ordering with concave function g. Then:

(i) $(28, 12) >_U (18, 19)$ but $(28, 12) <_{E_{g_{\mathfrak{k}}}} (18, 19)$.

(ii) $(28, 12) >_{E_{g_{\mathfrak{k}}}} (18, 16)$ but $(28, 12) <_{E_{g_{\mathfrak{i}}}} (18, 16)$.

(iii) $(28, 12) >_{E_{g_{\mathfrak{i}}}} (3, 114)$ but $(28, 12) <_{E_{g_{\mathfrak{p}_{1.5}}}} (3, 114)$.

(iv) $(28, 12) >_{E_{g_{\mathfrak{k}}}} (6, 40)$ but $(28, 12) <_{E_{g_{\mathfrak{p}_{1.5}}}} (6, 40)$.

Aside: Another interesting fact to note is that each of the quadratic divergence, I-divergence, and Itakura-Saito divergence takes as its minimum a different Pythagorean mean. The total quadratic divergence to a set of numbers is minimized at the arithmetic mean; the total I-divergence is minimized at the geometric mean; and the total Itakura-Saito divergence is minimized at the harmonic mean.

4 Minimizing the total divergence from the individuals

Above, I mentioned I would use divergences rather than distances because I didn't want to assume symmetry: that is, I want to allow that the divergence from u to u_i is different from the divergence from u_i to u. Indeed, of the divergences I introduced in Section 2, only the quadratic divergence is symmetric. And it is well-known that it is the only symmetric divergence in the whole family of Bregman divergences.

In Section 3, I described the definitions of moral goodness that minimize total divergence *to* the individual levels of well-being. In this section, I explain which definitions minimize total divergence *from* the individuals. Since the divergences can be asymmetric, the answer might be different in the two cases. And indeed, for all Bregman divergences except quadratic divergence, it is. Indeed, for all Bregman divergences it is the utilitarian account of the moral goodness of a state that minimizes divergence from the individual levels of well-being.

Proposition 2. *Suppose φ is a strictly convex and differentiable function on $S = \mathbb{R}$ or $S = \mathbb{R}_+$, and $\mathfrak{d}_\varphi$ is the Bregman divergence it generates. Then*

$$\arg\min_u \sum_{i=1}^n \mathfrak{d}_\varphi(u_i \| u) = \frac{u_1 + \ldots + u_n}{n}$$

Interestingly, for the α-divergences, it is not the utilitarian account that minimizes, but a very slightly different EEDE prioritarian account:

Proposition 3.

$$\arg\min_u \sum_{i=1}^n \mathfrak{a}_\alpha(u_i \| u) = \left(\frac{u_1^\alpha + \ldots + u_n^\alpha}{n}\right)^{\frac{1}{\alpha}}$$

5 Characterizing the divergences

The specific divergences I introduced in Section 2 are popular in the areas of science that use such measures, such as statistics, image reconstruction in tomography, and content search in databases (Csiszár & Shields, 2004; Herman & Lent, 1976; Zhang et al., 2009). But for the normative purposes of telling between different varieties of welfarism, we need principled reasons in favour of one divergence or another. In this final section, I briefly survey two of the most interesting approaches to such characterizations.

5.1 The Order-Sensitivity Approach

The first is due to Marcello d'Agostino and Valentino Dardanoni (2009). Let's assume that $\mathfrak{d} : S \times S \to [0, \infty]$ for $S = \mathbb{R}$ or $S = \mathbb{R}_+$. Writing $\mathbf{u}$ for the vector $(u_1, \ldots, u_n)$ in S^n and $\mathbf{v}$ for the vector $(v_1, \ldots, v_n)$ in S^n, they make two assumptions.

> **Difference Sensitivity.** There is continuous and strictly increasing $G : \mathbb{R}_+ \to \mathbb{R}_+$ such that
>
> $$\mathfrak{d}(u \| v) = G(|u - v|)$$
>
> **Monotone-Order Sensitivity.** For $n > 1$, define
>
> $$D_{\mathfrak{d}}(\mathbf{u} \| \mathbf{v}) = \sum_{i=1}^{n} \mathfrak{d}(u_i \| v_i)$$
>
> Now suppose we have $\mathbf{u}, \mathbf{v}$ and $1 \leq i, j, k, l \leq n$ with the following properties:
>
> (i) u_i, u_j are ordered the same way as v_i, v_j;
> (ii) u_k, u_l are ordered the same way as v_k, v_l;
> (iii) $\mathfrak{d}(u_i \| u_j) = \mathfrak{d}(u_k \| u_l)$;
> (iv) $\mathfrak{d}(v_i \| v_j) \leq \mathfrak{d}(v_k \| v_l)$.
>
> Then, if
>
> - $\mathbf{v}_{ij}$ is the vector that swaps the i^{th} and j^{th} entries in $\mathbf{v}$,
> - $\mathbf{v}_{kl}$ is the vector that swaps the k^{th} and l^{th} entries in $\mathbf{v}$,
>
> then
>
> $$D_{\mathfrak{d}}(\mathbf{u} \| \mathbf{v}_{ij}) \leq D_{\mathfrak{d}}(\mathbf{u} \| \mathbf{v}_{kl}))$$

And they prove:

Theorem 4. *If $\mathfrak{d} : S \times S \to [0, \infty]$ satisfies Difference Sensitivity and Monotone-Order Sensitivity, then $\mathfrak{d}$ is the quadratic divergence, that is, $\mathfrak{d}(u \| v) = \mathfrak{q}(u \| v) = (u - v)^2$.*

Now, recall from Proposition 1 that, if we measure divergence from one level of well-being to another using the quadratic divergence, then the compromise between that minimizes the sum of divergence from it to the individuals in a population is the utilitarian measure of moral goodness. So, if we can motivate the two axioms that D'Agostino and Dardanoni use to characterize that divergence, we have an argument for utilitarianism. So let's turn now to these.

The idea behind Difference Sensitivity is straightforward: the divergence from u to v should depend only on the difference between u and v. One thing to note about this axiom is that it entails Symmetry:

Symmetry. $\mathfrak{d}(u\|v) = \mathfrak{d}(v\|u)$, for all u, v in S.

As we noted above, the only Bregman divergence that satisfies Symmetry is the quadratic divergence $\mathfrak{q}$. We state that result formally here now:

Proposition 5. *Suppose φ is a strictly convex, differentiable function on S. Then, if $\mathfrak{d}_\varphi$ satisfies Symmetry, then it is the quadratic divergence, that is, $\mathfrak{d}_\varphi(u\|v) = \mathfrak{q}(u\|v) = (u-v)^2$.*

So, once we meet Csiszár's characterization of the Bregman divergences below, we can add this to obtain a characterization of the quadratic divergence.

The idea behing Monotone-Order Sensitivity is this. Sometimes, you need to measure the divergence from one sequence of levels of well-being to another. This is common practice, for instance, in the study of intergenerational economic mobility, where you wish to measure how far the well-being of one generation lies from the well-being of their offspring (Cowell, 1985; Fields & Ok, 1996). In such cases, if two individuals in the first generation are ordered in one way—let's say that $u_i > u_j$—and the corresponding individuals in the second generation are ordered in the same way—so that $v_i > v_j$—then if we look at the sequence of second generation well-being levels, but we swap v_i and v_j, so that individual i in the second generation now gets v_j and individual j gets v_i, then we call that an *order-reversing swap*. And one assumption that is often made about divergences from one sequence of well-being levels to another is that order-reversing swaps should increase divergence (Atkinson, 1983; Dardanoni, 1993). The idea is that, as well as measuring the sum of the divergences from one well-being level to another, it would be beneficial if our divergence could reflect similarities and differences between more global properties of welfare distributions, like the order of two individuals' levels of well-being. Keeping everything else equal, simply reversing the order of two individuals in the second generation who were also ordered in that way in the first generation increases the distance from the first generation.

Now that isn't quite what Monotone-Order Sensitivity demands, and indeed there is a whole family of divergences measures for which order-reversing swaps increase divergence, namely, the Minowski distances (when the power $p > 1$). Instead, Monotone-Order Sensitivity tells us something about how two order-reversing swaps relate to one another. It says that, if one order-reversing swap switches v_i and v_j, while another switches v_k and v_l, and the first-generation well-being levels u_i and u_j are exactly as far apart as u_k and u_l, and the second-generation well-being levels v_i and v_j are no further apart than those of v_k and v_l, then the first generation and the second generation with v_i and v_j switched are no further apart than the first generation and the second generation with v_k and v_l switched. That is, roughly, larger divergence between the swapped well-being levels leads to larger increases in the divergence from first generation to second.

5.2 The Projection Approach

The second approach I'll describe is due to Imre Csiszár (1991). He is interested in the following sort of situation, where, as before, S is either $\mathbb{R}$ or $\mathbb{R}_+$. You have an n-dimensional vector of real numbers $\mathbf{v}$ from S^n, and you have a set of such vectors $L \subseteq S^n$ and you want to identify the vector in L whose divergence from $\mathbf{v}$ is minimal. This problem comes up in a number of different contexts. For instance, $\mathbf{v}$ might give your prior probabilities over a range of possibilities and L might contain the probabilities that satisfy certain constraints that your new evidence places on your posterior probabilities, and you might wish to find the candidate posterior probabilities that satisfy the constraints of your new evidence but deviate as little as possible from your priors (Diaconis & Zabell, 1982). Another example, closer to our purpose here: the vectors in S^n might give distributions of well-being; $\mathbf{v}$ might be the ideal distribution of well-being within a society with n members; and yet your society, in its non-ideal way, might be able only to achieve a distribution of well-being that lies in the set L, and again you might wish to identify the distribution from that feasible set that lies the shortest distance from the ideal, for you think of that as the best you can do within the limits of your constraints.

Throughout, Csiszár is interested only in sets $L \subseteq S^n$ that are defined by linear constraints. That is, he restricts to those $L \subseteq S^n$ such that there is a matrix $\mathbf{A}$ in $\mathbb{R}^{m \times n}$ and a vector $\mathbf{b}$ in $\mathbb{R}^m$ for which

$$L = \{\mathbf{u} \in S^n \mid \mathbf{A}\mathbf{u} = \mathbf{b}\}$$

We write $\mathcal{L}^n_S$ for the set of such spaces. We write $\mathcal{M}^n_S$ for the set of $(n-1)$-dimensional spaces, so that L is in $\mathcal{M}^n_S$ iff there is a vector $\mathbf{a}$ in $\mathbb{R}^n$ and a number b in $\mathbb{R}$ such that

$$L = \{\mathbf{u} \in S^n \mid \mathbf{a}^T\mathbf{u} = b\}$$

Given a divergence $\mathfrak{d}$, let

$$\Pi_{\mathfrak{d}}(L \mid \mathbf{v}) = \underset{\mathbf{u} \in L}{\arg\min} \sum_{i=1}^{n} \mathfrak{d}(u_i \| v_i) = \underset{\mathbf{u} \in L}{\arg\min}\, D_{\mathfrak{d}}(\mathbf{u} \| \mathbf{v})$$

when this minimum exists and is unique. We can think of this as the *projection of* $\mathbf{v}$ *into* L. Initially, Csiszár makes four assumptions:

Existence. For any $\mathbf{v}$ in S^n and any L in $\mathcal{L}^n_S$, $\Pi_{\mathfrak{d}}(L \mid v)$ exists.

Distinctness. Given any two $L \neq L'$ in $\mathcal{M}^n_S$,

$$\Pi(L \mid \mathbf{v}) \neq \Pi(L' \mid \mathbf{v})$$

Continuity. $\mathfrak{d}(u \mid v)$ is a continuous function of u.

Transitivity. For any $\mathbf{v}$ in S^n and any $L' \subseteq L$ in $\mathcal{L}^n_S$,

$$\Pi_{\mathfrak{d}}(L' \mid \Pi_{\mathfrak{d}}(L \mid \mathbf{v})) = \Pi(L' \mid \mathbf{v})$$

And he proves the following theorem (Csiszár, 1991, Theorem 4):

Theorem 6. *If $\mathfrak{d} : S \times S \to [0, \infty]$ satisfies Existence, Distinctness, Continuity, and Transitivity, then there is strictly convex and differentiable φ such that $\mathfrak{d}$ is the Bregman divergence generated by φ, that is, $\mathfrak{d}(u \| v) = \mathfrak{d}_\varphi(u \| v) = \varphi(u) - \varphi(v) - \varphi'(v)(u - v)$.*

Note that, combining Proposition 5 and Theorem 6, we obtain:

Proposition 7. *If $\mathfrak{d} : S \times S \to \mathbb{R}_+$ satisfies Existence, Distinctness, Continuity, Transitivity, and Symmetry then $\mathfrak{d}(u \| v) = \mathfrak{q}(u \| v) = (u - v)^2$.*

He then states two further conditions. To state these, we introduce a little standard notation:

- $\lambda \mathbf{u} = (\lambda u_1, \ldots, \lambda u_n)$;
- $\lambda L = \{\lambda \mathbf{u} \mid \mathbf{u} \in L\}$;
- $\mathbf{u} + \mu \mathbf{1} = (u_1 + \mu, \ldots, u_n + \mu)$;
- $L + \mu \mathbf{1} = \{\mathbf{u} + \mu \mathbf{1} \mid \mathbf{u} \in L\}$.

Scale invariance. For any $\lambda > 0$, L in $\mathcal{L}^n_S$, and $\mathbf{v}$ in S^n,

$$\Pi(\lambda L \mid \lambda \mathbf{v}) = \lambda \Pi(L \mid \mathbf{v})$$

Translation invariance. For any μ, L in $\mathcal{L}^n_S$, and $\mathbf{v}$ in S^n,

$$\Pi(L + \mu \mathbf{1} \mid \mathbf{v} + \mu \mathbf{1}) = \Pi(L \mid \mathbf{v}) + \mu \mathbf{1}$$

And then he proves the following theorem (Csiszár, 1991, Theorem 4):

Theorem 8.

(i) *If $\mathfrak{d} : \mathbb{R} \times \mathbb{R} \to [0, \infty]$ satisfies Existence, Distinctness, Continuity, Transivity, Scale Invariance, and Translation Invariance, then $\mathfrak{d}(u \| v) = \mathfrak{q}(u \| v) = (u - v)^2$.*

(ii) *If $\mathfrak{d} : \mathbb{R}_+ \times \mathbb{R}_+ \to [0, \infty]$ satisfies Existence, Distinctness, Continuity, Transivity, and Scale Invariance, then one of the following holds:*

- $\mathfrak{d}(u \| v) = \mathfrak{k}(u \| v) = u \log \frac{u}{v} - u + v$
- $\mathfrak{d}(u \| v) = \mathfrak{i}(u \| v) = \log \frac{v}{u} + \frac{u}{v} - 1$
- $\mathfrak{d}(u \| v) = \mathfrak{p}_\alpha(u \| v) = \frac{1}{\alpha}(v^\alpha - u^\alpha) + v^{\alpha - 1}(u - v) \quad (\alpha < 1, \alpha \neq 0)$

So, if we can motivate Existence, Distinctness, Continuity, Transitivity, and Scale and Translation Invariance for divergences defined on all the real numbers, then we can pin down the quadratic divergence and thereby gain another argument for utilitarianism, since it is their account of the moral goodness of a welfare distribution that minimizes the total quadratic divergence to the individual levels of well-being.

On the other hand, if we can motivate Existence, Distinctness, Continuity, Transitivity, and Scale Invariance for divergences defined on all the positive real numbers, then we can narrow down the legitimate divergences to: the I-divergence, the Itakura-Saito divergence, and the power$_\alpha$ divergences (for $\alpha < 1, \alpha \neq 0$). And if we use any one of these divergences to measure how far an individual's level of well-being lies from a putative compromise, then the compromise that minimizes the total divergence to the individual levels is a version of EEDE prioritarianism, as we saw in Proposition 1.

So let's work through these conditions to see what they have going for them. Existence is a surprisingly powerful axiom. It requires that, for any $\mathbf{v}$ and L, there is in $\mathbf{u}$ *in L* that uniquely minimizes divergence to $\mathbf{v}$. Depending on whether we take S to be $\mathbb{R}$ or $\mathbb{R}_+$, this can rule out some standard divergences:

- First, suppose $S = \mathbb{R}_+$. Then quadratic divergence doesn't satisfy Existence. For instance, if $\mathbf{v} = (2, 20)$ and $L = \{(u_1, u_2) \in \mathbb{R}^2_+ \mid u_2 = 10 - u_1\}$, there is no $\mathbf{u} = (u_1, u_2)$ in L that minimizes the quadratic divergence from $\mathbf{u}$ to L. There is $\mathbf{u}$ in the line $L^* = \{(u_1, u_2) \in \mathbb{R}^2 \mid u_2 = 10 - u_1\}$ of which L is a segment that minimises this uniquely, namely, $(-4, 14)$. But that minimizer, is not in L.

- Second, suppose $S = \mathbb{R}$. Then I-divergence doesn't satisfy Existence. Indeed, $\mathfrak{k}(u\|v)$ is simply not defined for all u, v in $\mathbb{R}$, specifically, if $u > 0$ and $v = 0$.

This is why we see that, using similar conditions in the two parts of Theorem 8, Csiszár characterizes quite different sets of divergences: in the first case, S is $\mathbb{R}$; in the second, S is $\mathbb{R}_+$. In our context, then, it matters the scale on which well-being is measured.

Csiszár's other powerful axiom is Transitivity. It is really this that does much of the work in narrowing down to the Bregman divergences. And yet it is a very natural assumption. Let us suppose that our task is to find the welfare distribution, among those that are available to our imperfect society, that minimizes distance to the ideal distribution. Suppose $\mathbf{v}$ represents the ideal distribution, L represents certain constraints, while $L' \subseteq L$ represents more stringent constraints. Let $\mathbf{u}$ be the distribution in L that minimizes divergence to the ideal $\mathbf{v}$, that is, $\mathbf{u} = \Pi(L \mid \mathbf{v})$. That is, $\mathbf{u}$ is the best distribution we can manage within the confines of L. So let's assume we achieve that. But now say that further constraints come to be imposed on our society, so that $\mathbf{u}$ is no longer feasible—these further constraints are represented by L'. Then we might naturally say that we should move to the distribution $\mathbf{u}'$ within L' that minimizes the divergence to $\mathbf{u}$, that is, $\mathbf{u}' = \Pi(L' \mid \mathbf{u})$. But we might also naturally say that we should simply move to the distribution in L' that minimizes divergence to the ideal $\mathbf{v}$. Transitivity says that these two strategies always agree. If we project $\mathbf{v}$ into L, and then project the resulting vector into L', then we obtain the same vector as if we were to project $\mathbf{v}$ directly into L'.

The axioms that distinguish quadratic divergence from I-divergence, Itakura-Saito divergence, and the power$_\alpha$ divergences, and that help distinguish all of them from the other Bregman divergences are Scale and Translation Invariance. (Symmetry also distinguishes quadratic divergence from the other Bregmans.) Scale Invariance says that, if you scale up the constraints by a given factor and scale up the ideal vector by the same factor, then the projection of the scaled up ideal onto the scaled up constraints is the vector you get from scaling up the projection of the original ideal onto the original constraints. One way to see why that might be desirable is again to think of the projection of $\mathbf{v}$ onto L as the best possible welfare distribution among the feasible ones in L, that is, the one for which divergence to the ideal is minimal. Scale Invariance then says that, if we multiply the amount of welfare that can be distributed by a given amount, and thereby scale up the ideal distribution by that amount as well as the feasible distributions, then the best in this new situation should be the best in the old situation scaled up by that amount.

Translation Invariance says something similar, but instead of scaling up the amount of wealth available and the distributions that are feasible, we add a certain amount to each individual in the ideal distribution and the same to the distributions that are feasible. Here we might imagine that, exogenously, everyone is automatically given the same fixed amount of well-being as a sort of starter pack, and that is outside the control of the social chooser. The social chooser must then find the best way of distributing further well-being over this fixed base amount so that it satisfies the constraints and is closest to the ideal. Translation Invariance says that the best distribution of further well-being that the social chooser can disburse is not sensitive to the base amount given exogenously.

6 Conclusion

The considerations in the previous section in favour of the various divergences are not decisive, but they do provide a framework in which we might tell between utilitarianism and various forms of prioritarianism. The central idea is contractualist: the measure of moral goodness of a state of the world is a compromise between the levels of well-being enjoyed by individuals in that state of the world, just as the choices of a state are compromises between the choices that the various individuals within it might make; in both cases, the compromises must be justifiable to each individual involved; they must be such that no member of the population can reasonably reject them. We can ensure this by showing that the compromise minimizes the sum of the divergences to the individual levels of well-being. As we saw, different divergences give different results: some utilitarian, some prioritarian. Future work should seek further characterizations of the divergences that would allow us to better distinguish between utilitarianism and prioritarianism. The present note is more concerned with laying out the approach.

References

Adler, Matthew D., & Holtug, Nils (2019). Prioritarianism: A response to critics. *Politics, Philosophy & Economics*, *18*(2), 101–144.

Atkinson, Anthony B. (1983). The measurement of economic mobility. In A. B. Atkinson (Ed.) *Social justice and public policy*. London: Wheatsheaf Books Ltd.

Bregman, Lev M. (1967). The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, *78*(384), 200–217.

Cowell, Frank A. (1985). Measures of Distributional Change: An Axiomatic Approach. *The Review of Economic Studies*, *52*(1), 135–51.

Csiszár, Imre (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Erdodizität von Markoffschen Ketten. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, *8*, 85–108.

Csiszár, Imre (1991). Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Annals of Statistics*, *19*, 2032–2066.

Csiszár, Imre, & Shields, Paul C. (2004). Information Theory and Statistics: a tutorial. *Communications and Information Theory*, *1*(4), 417–528.

D'Agostino, Marcello, & Dardanoni, Valentino (2009). What's so special about Euclidean distance? A characterization with applications to mobility and spatial voting. *Social Choice and Welfare*, *33*(2), 211–233.

Dardanoni, Valentino (1993). Measuring social mobility. *Journal of Economic Theory*, *61*, 372–94.

Diaconis, P., & Zabell, S. L. (1982). Updating Subjective Probability. *Journal of the American Statistical Association*, *77*(380), 822–830.

Fields, Gary S., & Ok, Efe A. (1996). The Meaning and Measurement of Income Mobility. *Journal of Economic Theory*, *71*(2), 349–77.

Fleurbaey, Marc (2010). Assessing risky social situations. *Journal of Political Economy*, *118*, 649–680.

Harsanyi, John C. (1955). Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. *Journal of Political Economy*, *63*, 309–321.

Hellinger, Ernst (1909). Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *Journal für die reine und angewandte Mathematik*, *136*, 210–271.

Herman, Gabor T., & Lent, Arnold (1976). Iterative reconstruction algorithms. *Computers in Biology and Medicine*, *6*, 273–294.

Itakura, Fumitada, & Saito, Shuzo (1968). Analysis synthesis telephony based on the maximum likelihood method. In Y. Kohasi (Ed.) *Reports of the 6th of the International Congress on Acoustics*, (pp. 17–20). IEEE.

Konieczny, S., & Pino Pérez, R. (1998). On the logic of merging. In *Proceedings of KR'98*, (pp. 488–498).

Konieczny, S., & Pino Pérez, R. (1999). Merging with integrity constraints. In *Fifth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'99)*, (pp. 233–244).

Kullback, Solomon, & Leibler, Richard A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*(1), 79–86.

McCarthy, David (2008). Utilitarianism and Prioritarianism II. *Economics and Philosophy*, *24*, 1–33.

Osherson, Daniel, & Vardi, Moshe Y. (2006). Aggregating disparate estimates of chance. *Games and Economical Behavior*, *56*(1), 148–173.

Parfit, Derek (2000). Equality or Priority. In M. Clayton, & A. Williams (Eds.) *The Ideal of Equality*, (pp. 81–125). Houndmills: Palgrave.

Parfit, Derek (2012). Another defence of the priority view. *Utilitas*, *24*, 399–440.

Pettigrew, Richard (2019). Aggregating incoherent agents who disagree. *Synthese*, *196*, 2737–76.

Pettigrew, Richard (2022). Aggregating agents with opinions about different propositions. *Synthese*, *200*(5), 1–25.

Predd, Joel B., Osherson, Daniel, Kulkarni, Sanjeev, & Poor, H. Vincent (2008). Aggregating Probabilistic Forecasts from Incoherent and Abstaining Experts. *Decision Analysis*, *5*(4), 177–189.

Scanlon, T. M. (2000). *What We Owe to Each Other*. Cambridge, Mass.: Harvard University Press.

Zhang, Zhenjie, Ooi, Beng Chin, Parthasarathy, Srinivasan, & Tung, Anthony K. H. (2009). Similarity search on Bregman divergence: towards non-metric indexing. *Proceedings of the VLDB Endowment*, *2*(1), 13–24.