Searching for the Ideal:
The Fundamental Diversity Dilemma*

"Would you tell me, please, which way I ought to go from here?"
"That depends a good deal on where you want to get to," said the Cat.

Gerald Gaus and Keith Hankins

1. IDEAL THEORY AS ORIENTING OUR SEARCH FOR JUSTICE

1.1 Ideal Principles of Justice v. Ideal Social Worlds

There are numerous understandings of ideal theory and its attractions. On one understanding, it is about the ideal, optimal, or best theory of justice. This characterization may not make a great deal of sense on certain conceptions of political philosophy. If our understanding of the "best" principles of justice is that they are the "true" principles, it does not seem especially appropriate to call these principles "ideal." They simply are the truth about justice. However, on some views, especially those that employ an original position or contract to identify or justify principles of justice, the "ideal" modifier gets more traction. This, of course, is because in these theories we can idealize both the choosers and the conditions under which they choose. Suppose, that is, we entertain:

> *P* is an ideal principle of justice if and only if *P* would be selected by ideally rational and impartial choosers (or a single such chooser), *C*, assuming ideal background conditions, *B*.[1]

*We have greatly benefitted from the comments, encouragement (and strenuous objections) of a number of colleagues and friends. We are especially grateful to Fred D'Agostino for his unflagging support and insights. Thanks too to Sameer Bajaj, Alan Hamlin, Ryan Muldoon, John

[1] For a more detailed examination of chooser-based theories see Gerald Gaus and John Thasher, "Rational Choice in the Original Position" in *The Original Position*, edited by Timothy Hinton (Cambridge: Cambridge University Press, forthcoming).

Here it clearly makes sense to debate whether principles of justice should be "ideal principles," or whether a chooser-based theory should rely instead on more realistic assumptions about choosers and the background conditions of choice.

Our concern is not with this understanding of "ideal theory," but with a different sort of debate about the justice of social structures. As David Gauthier pointed out, sometimes when we are reflecting about ideal justice we are interested in the structure of the ideally just society. Our concern, in other words, is to identify the set of social structures and interpersonal relations that best conforms to our understanding of justice (e.g., that which most perfectly conforms to the relevant principles of justice).[2] To be somewhat more precise, let us say:

> *S* is the uniquely, ideally, just social structure (or, as we shall say, "social world") if and only if, among the set of social worlds *X* (where *X* is defined in terms of the possible, plausible, realistically achievable, etc.), *S* is most just in *X*.

1.2 The Ideal as Orienting

Although much of the contemporary debate about ideal theory has been focused on the first, ideal principle interpretation, in the history of political philosophy the second, ideal social world understanding, has often been a critical inspiration for ideal qua utopian thought. Right from the beginning political philosophy has sought to describe the ideal state, which, even if not itself fully achievable, tells us about the nature of justice. As Plato, the first of the ideal theorists, acknowledged, it is in "the nature of things that action should come less close to truth than thought," and so our ideal constructions will not be "reproduced in fact down to the least detail."[3] On this view, as one contemporary philosopher has put it, the ideal functions as "a mythical Paradise Island," which tells us where "*the endpoint of our journey lies*." Although the ideal does not "necessarily tell us *anything about the route to take to Paradise Island,"* it orients our journey.[4] It is only after identifying an ideal, in other words, that we can

---

[2] David Gauthier, *Morals by Agreement* (Oxford: Oxford University Press, 1986), pp. 233-4.

[3] Plato, *The Republic,* translated by Francis MacDonald Cornford (Oxford: Oxford University Press, 1941), p. 178 [v. 473]

[4] Ingrid Robeyns, "Ideal Theory in Theory and Practice," *Social Theory and Practice*, vol. 34 (July 2008): 314-362 at p. 345. Emphasis added.

take up the task of figuring out how to get there (or, if we cannot quite get to the ideal, to come as close to is as possible). Or, as Rawls, who spoke of ideal theory in both of the ways we have described, says of the ideal social world understanding of ideal theory: "By showing how the social world may realize the features of a realistic Utopia, political philosophy provides a long-term goal of political endeavor, and in working toward it gives meaning to what we can do today."[5] As Rawls also notes, though, "[the] idea of realistic Utopia is importantly institutional."[6] What we wish to identify are the institutional structures and patterns of interaction of a realistically achievable social world, for it this that ultimately provides the guidance we need to reform our own social world's institutions. Of course, we may never arrive at the ideal social world, but with an ideal guiding us the hope is that we can rest assured that our efforts to secure justice have at least moved us in the right direction.

1.3 Finding the Path Between Two Fixed Points or Finding the Ideal?

Underlying our discussion thus far has been what we take to be the predominant view of how the ideal can orient our search for justice. On this view we have two, relatively clear, fixed points: where we now are and the ideal. The practical problem of political philosophy is to find a route between them. This model of the task of political philosophy is amazingly close to using Google Maps to get directions. One has definite knowledge of where one is, and where one seeks to end up, but is uncertain of the distance or the terrain in-between. The wonder of Google is that it will inform you of the various routes you might take, what difficulties you are apt to meet on the way, and how long each route might take. And, more wondrous still, it is almost always correct.

But surely, despite the dreams and ambitions of Plato or Rawls, the path to a politically and morally better world cannot possibly be based on this model. If the perfectly just world is far away, how could we possibly think that we now have thorough knowledge of it such that it can meaningful orient our journey, let alone provide us with a firm grasp of our destination? As we (and Rawls) have said, on the social world interpretation of ideal theory the ideal is critically institutional. It is

[5] John Rawls, *The Law of Peoples* (Cambridge: Harvard University Press, 1999), p. 138.
[6] Ibid., p. 16.

about how a set of institutions will function given various background facts and conditions. Our understanding of institutional dynamics for sets of institutions very different from our own, though, must certainly be tentative. As we travel in the direction of the better, surely we must leave open the possibility (indeed, moral certainty) that we will actually learn something about the best. Surely it is the height of hubris to suppose that, despite our admission of the deep imperfections of our social world, the one thing that is perfect is our philosophical and social scientific knowledge of the best social world.

All this is rather broad-brushed. Later (§3.1) we shall provide some more systematic arguments for our conviction that part of our pursuit of the ideal is to learn just what it is, and — in a sense to be made more explicit — where it is. On our analysis, the ideal is not a predetermined destination, but something we come to discover as we search for it.

1.4 Plan of the Paper

Thus our question: how can we discover what the ideal is and in what direction it lies? And, with this information in hand, how can we use it to orient our pursuit of justice? Section 2 articulates the basic model of the search problem. Section 3 defends what we call the Neighborhood Constraint. Section 4 considers important, albeit limited, ways to mitigate the Neighborhood Constraint. Section 5 then turns to more radical solutions to the Neighborhood Constraint, focusing on what we call The Utopia Is At Hand Theorem. In section 6 we argue that, while these radical solutions hold out the promise of nullifying the Neighborhood Constraint, they also undermine the intelligibility of the very idea of a common pursuit of a shared ideal. We call this result the Fundamental Diversity Dilemma.

2. RUGGED LANDSCAPE OPTIMIZATION MODELS

2.1 Evaluative Perspectives

In his *Lectures on the History of Political Philosophy* Rawls tells us that "a normalization of interests attributed to the parties" is "common to social contract doctrines."[7] This

[7] John Rawls, *Lectures on the History of Political Philosophy*, edited by Samuel Freeman (Cambridge, MA: Harvard University Press, 2007), p. 226.

remark is made in the context of discussing Rousseau's idea of the general will, which is also said to required a shared "point of view."[8] On Rawls's reading of Rousseau, private individuals are characterized by a variety of different interests that are magnified by self-bias and selfishness. Such individuals can live together under freely endorsed common laws only if they "share a conception of the common good."[9] This shared conception, in turn, is generated by individuals' fundamental interests and capacities, which derive from their shared human nature. As Rawls sees it, these shared fundamental interests allow individuals to abstract from their differences and occupy a shared legislative point of view, based on a shared conception of the common good.[10] Furthermore, when occupying this shared view parties all have the same basis for their deliberations, and so everyone will the same laws, and this is what allows them to live together under freely endorsed common laws.

Rawls is informally articulating the conception of a perspective — or, as we shall say, an evaluative perspective — which has been more thoroughly and formally explored in the last decade.[11] As we model it, an *evaluative perspective,* Σ, includes three fundamental features (we shall later see more are required):

(A) A set of evaluative standards (principles of justice, values, etc.) by which alternative social worlds are to be evaluated.

(B) For all worlds *i* in *X*, a specification of the features of *i* that are relevant to (A). This yields a set of categorizations that constitute a description of *i* on perspective Σ. (We follow Rawls in holding that a scheme of basic institutions constitutes a "social world,"[12] and in supposing that a description of a social world includes an understanding of its dynamics and relevant background conditions, including economic and psychological facts.)

---

[8] Ibid., pp. 229ff.

[9] Ibid., p. 224.

[10] Ibid.

[11] See Scott E. Page, *The Difference* (Princeton: Princeton University Press, 2007), pp. 30ff. See also Ryan Muldoon, *Diversity and the Social Contract* (Ph.D. dissertation, University of Pennsylvania).

[12] Rawls, *Political Liberalism*, paperback edition (New York: Columbia University Press, 1996), p. 77. See also p. 41.

(C) A mapping function that applies (A) to (B), to determine the overall justice of the social world. Together (A) and (C) constitute a complete justice scoring function such that, for any social world in X and set of evaluative standards (A), a definite cardinal justice score can, in principle, be generated for every social world.

Rawls's description of the normalized interests supposed by contract models seems to only refer to feature (A), but it is clear that without (B) and (C), a shared set of evaluative standards will not suffice for a shared evaluation. If individuals are going to adopt a shared point of view they must also agree about precisely what they are evaluating and how to apply their shared standards to that which is being evaluated.

Two features of element B call for further clarification. (1) Note that the very idea of an ideal orienting our search for justice supposes a set of possible or feasible social worlds — for Rawls, ones that can be achieved under optimistic but plausible assumptions. An ideal theory performing this function must then be able to evaluate our social world, and alternative worlds, within the set. On the two-fixed points interpretation (§1.3), the current and ideal worlds are known, with those in-between a matter for discovery. On the other hand, on our discovery model we only know our own social world particularly well, and our task is to evaluate alternative worlds in the set, and come to better understand their justice and their relation to each other.

(2) By definition, no social world shares the exact same justice-relevant features as another (that is why they are not the same social world). Social world a may have justice-relevant features $\{f_1, f_2, f_3\}$, while b has some, but not all of these features (e.g. f_1 and f_3, but not f_2), and c a still different set of features (e.g. f_1 andf_2, but not f_3). Social worlds are thus individuated by the differences in their justice-relevant features.

To clarify the nature of the evaluation of each social world, consider three different procedures that might be employed. The first we might call "categorical," and is probably the most common type of evaluation in philosophy. Categorical judgments are concerned with whether or not a social world is just; employing this procedure would yield a series of yes/no judgments regarding the justice of various

social worlds. Whatever its attractions in other contexts, such judgments are of little use for an ideal theory that seeks to orient our quest for justice by guiding us to better worlds, short of the ideal. For that we need an evaluative function capable of generating a range of judgments. John Broome, for instance, contrasts the philosopher's use of "categorical" judgments with the economist's focus on "comparative" judgments. Whereas the philosopher asks "Is *a* just or unjust?" the economist asks "Is "*a* is more or less just than *b*?"[13]

The idea of a comparative judgment is ambiguous between the *strictly comparative* and the *scalar* interpretations, though.[14] On a strictly comparative reading one can only form a judgment about a given social world, *a*, by comparing it to another social world, *b*. On this reading, judgments of justice have the same logical structure as preferences: they are necessarily binary, and so "*a* is more just than *b*" becomes the basic primitive in our thinking about justice. Furthermore, because judgments of justice share this binary comparative structure with preference, just as we could have intransitive preferences (*a* is preferred to *b*, *b* is preferred to *c*, and *c* is preferred to *a*), we could also have intransitive judgments of justice (*a* is more just than *b*, *b* is more just than *c*, yet *c* is more just than *a*).[15] As is well-known, pairwise judgments made on the basis on *N*-dimensional underlying considerations can lead to a host of such pathologies.[16] Without modification, then, such a comparative approach is also inappropriate as a model of ideal theorizing. Unless we impose transitivity on the set of binary judgments in *X* (thereby creating an ordering of *X*) we cannot be assured that there will be a best element in the set. In other words, a set with cycles at the top would leave us without an ideal.

To avoid these problems, in this paper we model the evaluation of the justice of social worlds as based on a scalar function jointly defined by (A) and (C), which is not inherently comparative (though, importantly, neither is it categorical). On our model, an evaluative perspective Σ has a set of standards and a consistent way

[13] John Broome, *Ethics Out of Economics* (Cambridge: Cambridge University Press, 1999), p. 164.

[14] We have greatly benefitted from discussion with Christian Coons about these matters.

[15] For an argument that our moral thinking is characterized by such cycles, see Larry S. Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (Oxford: Oxford University Press, 2011).

[16] See Fred D'Agostino, *Incommensurability and Commensuration* (Aldershot, Hampshire: Ashgate, 2003).

applying them to social worlds, which generates for any social world a score of its justice.

2.2 Meaningful Structures and Perspectives: Similarity and Distance

In addition to identifying the relevant features of a social world, a perspective creates what Scott Page calls a "meaningful structure" or "meaningful relatedness" within the set X.[17] In other words, a perspective does not simply see X as a random collection of social worlds with diverse justice-relevant features, but rather as a set differentiated by systematic variations in their underlying properties such that some worlds appear closer to one another, and others farther away.[18] More formally, let us suppose our current social world a is composed of a set of features $\{f_1, f_2, f_3\}$, according to evaluative perspective Σ. Now consider two alternative social worlds, b and c, as in section 2.1. World b has features $\{f_1, f_3\}$ but not f_2; the closest feature it has to f_2 is f_2^*. World c has features $\{f_1, f_2\}$ but not f_3; the closest it has to f_3 is f_3^*. If, on evaluative perspective Σ, f_2 is more similar to f_2^* than f_3 is to f_3^*, we can say that the underlying structure of a is more similar to b than it is to c. Then, using a similar procedure, we can say whether social world c is closer to a or b, and from these sorts of comparative judgments we can generate a coherent set of similarity judgments for a members of X.[19] To simplify, we generally suppose that this sort of procedure induces a one-dimensional array, though later (§§4.2-4.3) we shall expand our analysis to include N-dimensional similarity.

Finally, we suppose that Σ enriches this one-dimensional array by applying a distance metric. Σ thus defines a metric space — an ordered pair (X, d), where X is a domain of social worlds and d is a function on X that defines the distance between all points in X. And, at least initially, we can assume that the distance metric is constrained by the prior judgments of similarity in underlying structures.

We thus have five elements of an evaluative perspective:

---

[17] Page, *The Difference*, pp. 48, 33.

[18] As Martin L. Weitzman noted in his classic analysis of measuring diversity, "Distance is … an absolutely fundamental concept in the measurement of dissimilarity." "On Diversity," *The Quarterly Journal of Economics*, vol. 107 (May, 1992): 363-405 at p. 365. See also Scott E. Page, *Diversity and Complexity* (Princeton: Princeton University Press, 2011), pp. 58ff.

[19] See the Appendix for a procedure for generating a single-dimensional structure.

(A) A set of evaluative standards or principles of justice;

(B) An identification of the relevant features of social worlds;

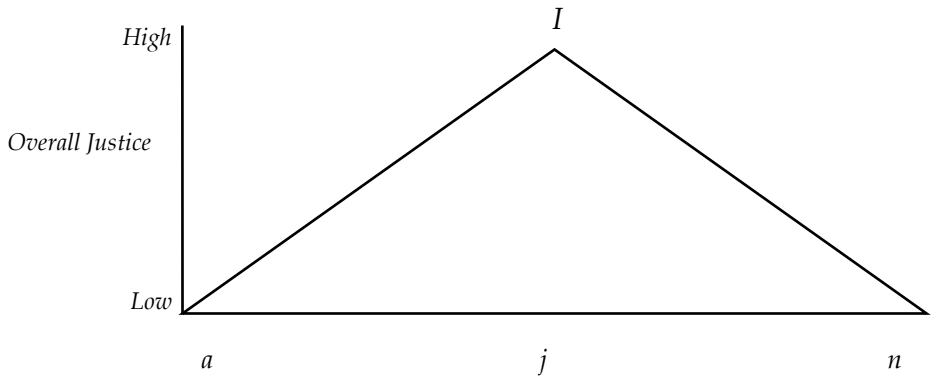
(C) A mapping relation from (A) to (B);

(D) An ordering of the underlying structures which meaningfully relates them in terms of similarity;

(E) A distance metric.

2.3 Smooth v. Rugged Optimization

Evaluative perspectives allow us to make judgments about the justice and structural similarity of a set of social worlds. Sometimes, as Display 1 indicates, a perspective can show us that our search for the ideal will be easy.



DISPLAY 1: A MOUNT FUJI LANDSCAPE

Here the *x*-axis represents Σ's understanding of the underlying structure of social worlds *a* through *n* in *X*. The *y*-axis, on the other hand, represents Σ's evaluation of the justice of these worlds. On this fortunate perspective, marginal changes in the underlying structure are always associated with marginal changes in their justice. Moreover, as we move from social world *a* towards *j* every small change in social structure leads to an increase in justice. Similarly, as we move from *j* towards *n* each small change yields a small loss in justice. Finding the ideal is simple. First move from where you are. If you get to a less just social world, stop, and move back. Then, move in the other direction, and keep on moving in that direction until a marginal

change yields a less just world. Finally, move one step back and you will have arrived at the ideal, the most just social world!

Does an ideal theorist usually face such a straightforward optimization problem? Recall that like Rawls we hold that institutional structure is fundamental to identifying just social worlds (§1.2). This is important because the justice of an institution (or more broadly, a policy) can be dependent on what other institutions or policies are in effect, as shown in Display 2.

Worlds	World with x	World with y	World with z	Total Justice
a	✔			10
b	✔		✔	12
c			✔	4
d		✔	✔	8
e	✔	✔	✔	2

DISPLAY 2: BLEEDING HEART LIBERTARIAN PERSPECTIVE

Here we consider a mild libertarian perspective, with "bleeding-heart" libertarian inclinations — that is, a perspective that places great weight on free markets and small states, but also values government aid to the less well off.[20] And:

Let x = Prohibition of deficits;

Let y = Prohibition of tax rises;

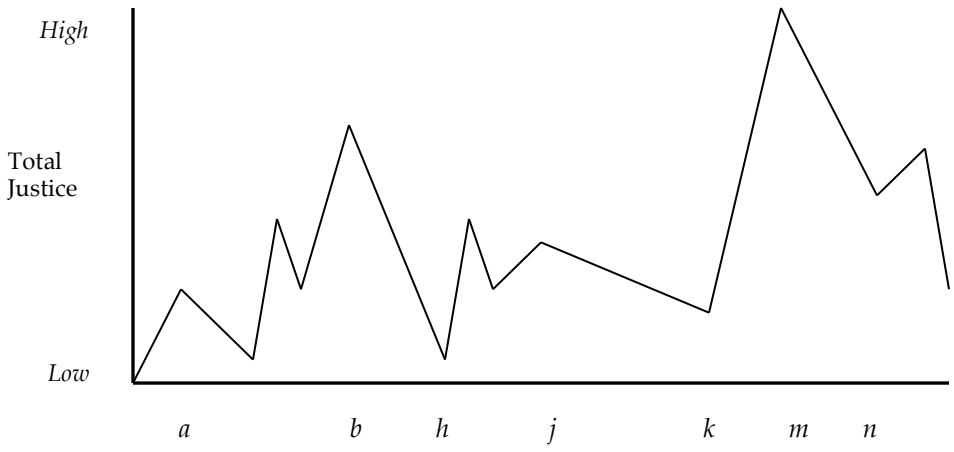
Let z = Prohibition on cutting vital services.

Suppose we start out in social world *a*, which has limits on deficits. Our libertarian may judge this world to be reasonably just because current generations cannot push the costs of their consumption on to the future, and so will be apt to be more cautious about governmental expenditure. But recall that our libertarian is concerned with the least well-off. She may therefore judge society *b* to be more just

[20] See, for example, John Tomasi's *Free Market Fairness* (Princeton: Princeton University Press, 2012).

than *a*, since *b* protects vital services upon which the least well off depend. However, now suppose a move to world *c* that keeps the prohibition on cutting vital services, but drops prohibition on deficits. World *c* may very well be judged a less just world than either *a* or *b* (with a score of 4 compared to 10 or 12), as the prohibition on cutting vital services is likely to inflate the size of the state whose costs will either be pushed onto future generations or funded through increases in taxation. Introducing a limit on taxation in world *d* at least mitigates some potential injustices, raising the justice of *d* compared to *c* (8 compared to 4), but still leaving *d* less just than *a* or *b*. Now suppose that in *e*, as in world *d*, there is a prohibition of increased taxation, but as in world *b*, prohibitions on both cutting vital services and deficits. One might think that the libertarian would judge this to be the best of all worlds. However, she may fear a California syndrome, in which expenses can neither be cut nor paid for, giving rise to the real possibility of default on the state's obligations which may pose the greatest threat to justice of all, leaving *e* the least just social world.[21]

Note that as we move from *a* to *e*, we have "peaks" at *b* and *d*, with gullies in-between. Accordingly, our model is more akin to the rugged landscape of Display 3 than to the Mount Fuji landscape of Display 1.



DISPLAY 3: A RUGGED LANDSCAPE

[21] Our thanks to Fred D'Agostino for suggesting this case.

More formally, if the various dimensions of social worlds interact as we've suggested they might, then we are confronted with an *NK* optimization problem — one in which we are optimizing over *N* dimension with *K* interdependencies among them.[22] When *K*=0, that is when there are no interdependencies between the justice of the dimensions, we have a simple optimization problem — the more of each element, the better — and each act of local optimization puts us on a path towards global optimization or the realization of an ideal. Not so when *K* begins to increase (as in evolutionary adaptation).[23] When multiple dimensions interact in complex ways to produce varying justice scores, as we saw in Display 2, then we are faced with a rugged landscape in which optimization is much more difficult..

Interestingly, although few political philosophers have analyzed their optimization problems in terms of rugged landscapes,[24] the idea underlies much of the recent discussion of ideal theory and non-ideal theory. For instance, in his criticism of the importance of identifying the ideal, Amartya Sen appeals to a topographic metaphor:

> The possibility of having an identifiably perfect alternative does not indicate that it is necessary, or indeed useful, to refer to it in judging the relative merits of two alternatives; for example, we may be willing to accept, with great certainty, that Mount Everest is the tallest mountain in the world, completely unbeatable in terms of stature by any other peak, but that understanding is neither needed, nor particularly helpful, in comparing the peak heights of, say, Mount Kilimanjaro and Mount McKinley. There would be something off in the general belief that a comparison of any two alternatives cannot be sensibly made without a prior identification of a supreme alternative.[25]

To which John Simmons, explicating Rawls's theory, replies:

> Which of two smaller "peaks" of justice is the higher (or more just) is a judgment that matters conclusively only if they are both on equally feasible paths to the highest peak of perfect

[22] The classical work exploring these problems is Stuart A. Kaufman, *The Origins of Order* (New York: Oxford University Press, 1993), especially chap. 2.

[23] See Gaus, "The Evolution, Evaluation and Reform of Social Morality: A Hayekean Analysis" in *Hayek and the Modern Economy*, edited by David Levy & Sandra Peart (New York: Palgrave Macmillan, 2013): 59-88.

[24] An important exception here is the work of Ryan Muldoon.

[25] Amartya Sen, *The Idea of Justice* (Cambridge, MA: Harvard University Press, 2009), p. 102.

> justice. And in order to endorse a route to that highest peak, we certainly do need to know which one that highest peak is. Perhaps for a while we can just aim ourselves in the general direction of the Himalayas, adjusting our paths more finely—between Everest and K2, say—only when we arrive in India. But we need to know a great deal about where to find the serious candidates for the highest peak before we can endorse any path to them from here.[26]

All this rather heavy reliance on topographical metaphors turns out to be more informative than at first glance. In more rigorous terms, Simmons can be understood to be making a point about justice as an *NK* optimization problem. *Pace* Sen, in rugged landscapes such as Display 3 a constant series of pairwise improvements can lead (*i*) to a local optimum (a small peak on Display 3) that is far inferior to the global optimum and (*ii*) lead us away the globally optimal social world. If we are at *h* we could move towards the nearby peak at *b*, but unfortunately this would take us farther away from the ideal social world, *m*. Whether theories of justice are tasked with solving optimization problems in rugged or smooth landscapes, then, matters quite a bit.

3. THE NEIGHBORHOOD CONSTRAINT

3.1 Why We Should Reject Claims to Know the Justice of Far-off Worlds

We began (§1.3) by contrasting two views of how we might search for the ideal: that which supposes two fixed points (the actual and the ideal), and that which supposes that we have firmest knowledge of the social worlds approximate to the actual. On the second view, as we proceed to evaluate more far-flung social worlds — that is, ones who basic structural features are very different from our own world — our understanding of their institutional dynamics, and so their justice, becomes far more speculative. Although, in principle, a perspective, Σ, can evaluate any social world, an actual evaluator needs to be able to model the institutional dynamics, and the social relations that characterize them, before a well-grounded judgment can be made.

Because social worlds are comprised of interdependent institutions and human actors interacting against background facts such as economics and history, social

[26] A. John Simmons, "Ideal and Nonideal Theory," *Philosophy and Public Affairs*, vol. 38 (Jan 2010): 1-13, at p. 35.

worlds are complex systems.[27] Complex systems are characterized by several features. (*i*) They have numerous heterogeneous variables; (*ii*) new patterns of interaction arise that could not have been predicted from the microelements (these are emergent properties); (*iii*) they are tightly coupled. The state of a complex system at any one time depends on a number of factors, and if even only a few are varied, there may be profound changes throughout the system. Thus we come to our critical point: the behavior of tightly coupled, complex, systems is difficult to predict beyond the very short term. This is primarily because such systems are characterized by *error inflation*. In complex systems small errors in predicting one variable at time *t* lead to drastic errors in predicting the overall system's state at *t+n* (even when *n* is small) as errors in the initial estimate of one variable both propagate to other variables and become magnified in subsequent periods. The combination of complexity and tight coupling is especially troublesome to successful manipulation of the system: "Complexity makes it impossible for anyone to understand how the system might act; tight coupling spreads problems once they begin."[28]

This is not to say that we cannot model or predict complex systems. Our ability to predict their behavior, however, radically decreases as we move into the future, or attempt to predict system's behavior in states very different from those we presently observe. The quintessential example of this is weather forecasting. Our ability to predict weather systems ten days out is drastically inferior to our ability to predict tomorrow's or the next day's weather. A crucial reason for this is that most of the variables influencing tomorrow's weather are set today and, because tomorrow's system is highly correlated with today's, we do not totally rely on abstract models in predicting it. As Stuart Kaufmann showed in his pioneering work on rugged landscapes (which model complexity), we can explain these results mathematically.[29] Consider again Display 3, which depicts an optimization problem with the sort of

---

[27] The analysis here draws on Gerald Gaus "Social Complexity and Evolved Moral Principles" in *Liberalism, Conservatism, and Hayek's Idea of Spontaneous Order*, edited by Peter McNamara. (London: Palgrave Macmillan, 2007), pp. 149-176.

[28] Edward Tanner, *Why Things Bite Back* (London: Fourth Estate, 1996), p. 16.

[29]. Kaufman, *The Origins of Order*, p. 70. This analysis applies only to landscapes that are not maximally complex (or rugged). A maximally rugged landscape is produced when, in an *NK* optimization problem, *K=N–*1. In this case we have a number of dimensions, and each dimension of dependent on all the others, leaving us with what Kaufmann calls a "complexity catastrophe" where the overall value of any point is uncorrelated with the points nearest to it (ibid., p. 52).

complexity we have been analyzing. Note that within some small distance from any point (say *b*), the value of the nearby social worlds are correlated with it. Even if one does not have a great predictive model, one knows that the justice of social worlds *b*+1 and *b*-1 will be correlated with *b*, because *b* contains most of the justice-relevant features of *b*+1 (and *b*-1). As we move far way from *b*, though, say to *j*, the two worlds share fewer features, and knowledge of the justice of *b* does not tell us anything about the justice of *j*. To evaluate *j* we must rely entirely on our models (which, we have seen, are subject to error inflation).

We believe there is overwhelming reason to reject the claim that we can be confident in our evaluations of the justice of far-off social worlds. Although this is hardly news to social scientists,[30] it runs counter to a great deal of philosophy which often seems to suppose that ideal states may be quite simple, and so easy to predict and evaluate. Consider, for example, G.A Cohen's famous story of a camping trip between friends that he uses to illustrate the basic dynamics of a socialist society.[31] Certainly Cohen's example looks like a simple model of a remarkably non-complex society, but that we can model socialism in this way does not show that actual socialism would be anything like Cohen's model predicts. Cohen suggests that his model captures the core dynamics of socialism, but, of course, that is precisely what is up for consideration. On what grounds could we have confidence that this toy model captures an actual social world of an advanced economy?

Although we have suggested that philosophers have long been too confident in the predictive power of their models, we are not the only philosophers to have voiced this concern. John Stuart Mill was deeply critical of those who take such models as guide to actions, because they are confident they have truly understood the dynamics of the ideal order. As he observed of the "revolutionary" socialists of his day:

> It must be acknowledged that those who would play this game on the strength of their own private opinion, unconfirmed as yet by any experimental verification—who would forcibly deprive all who have now a comfortable physical existence of their only present means of

[30] Though perhaps not to economists who, even after the last five years, appear to have inordinate faith in their models.

[31] G. A. Cohen, *Why Not Socialism?* (Princeton: Princeton University Press, 2009).

> preserving it, and would brave the frightful bloodshed and misery that would ensue if the attempt was resisted—must have a serene confidence in their own wisdom on the one hand and a recklessness of other people's sufferings on the other, which Robespierre and St. Just, hitherto the typical instances of those united attributes, scarcely came up to.[32]

Similarly, although we have approvingly cited Simmons's critical response to Sen (§2.3), it is worth noting that Sen's criticism of ideal theory is motivated in part by a concern that our descriptions of the ideal are often parochial, and that, accordingly, we should be less confident that they correspond to the truth about justice.[33] Of course, none of this is to say that, in proposing realistic reform, we can only plausibly move to adjacent social worlds. This would only be true if our knowledge of other social worlds was limited to just our closest neighbors. We do want to stress, however, that our knowledge of alternative social worlds is limited, and this limit is a function of the dissimilarity of other social worlds to the one which we currently occupy.

3.2 The Neighborhood Constraint

Having described some of the constraints on our knowledge of alternative social worlds that complexity introduces, we come to the critical concept of what Rawls called the "neighborhood" of our present set of social and political institutions. For our purposes, a neighborhood delimits a set of nearby social worlds characterized by relatively similar social structures. We assume that there exists a rough continuum of social worlds characterized by their social structures, some of which are in the neighborhood of our own social world (and some of which are not), and we propose that our understanding of the justice of alternative social worlds is far deeper within the neighborhood of our own social world. Furthermore, for simplicity, we assume that there is a clear boundary between the worlds that are in our neighborhood and those that are too dissimilar for us to make firm judgments about. Here, we follow Rawls who thought that judgments of the justice of basic structures were always

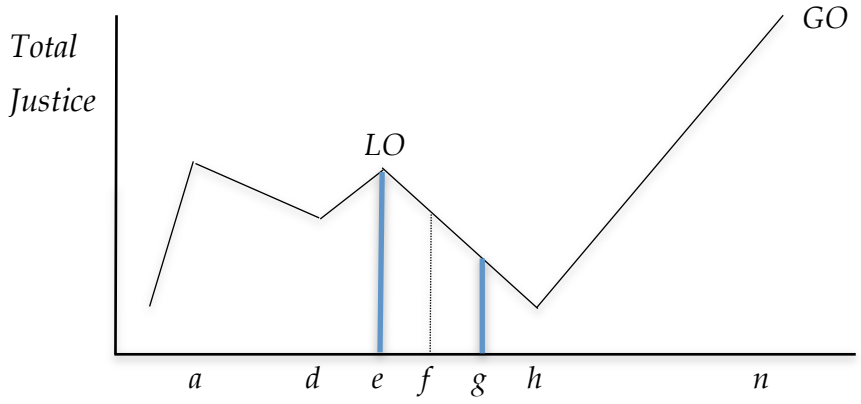
---

[32] John Stuart Mill, *Chapters on Socialism* in *The Collected Works of John Stuart Mill*, vol. 5, edited by John M. Robson (Toronto: University of Toronto Press, 1967), p. 737.

[33] See e.g., Amartya Sen, "What do we want from a theory of justice?" *Journal of Philosophy*, vol. 103 (2006): 215-238.

bound by a neighborhood. Indeed, Rawls says, the "difference principle itself [depends], on their being a rough continuum of basic structures, each very close (practically speaking) to some others in the aspects along which these structures are varied as available systems of social cooperation. (Those close to one another are in the same neighborhood)." [34]

Display 4 below illustrates how we can incorporate the idea of a neighborhood into a rugged landscape model.



DISPLAY 4: A NEIGHBORHOOD

Here, we are at social world *f*, our neighborhood runs from *e* to *g*, and *e* is the "local optimum" (LO) — the most just alternative in our neighborhood. We can immediately see the difficulty of optimization in rugged landscapes. While moving from *f* to *e* takes us to a more just social world, it also moves us further away from the global optimum (GO), *n*. So, we face a dilemma. On one hand, our understanding of the alternatives to our present world is limited. We can only confidently make judgments about social worlds in our neighborhood. On the other hand, ideal theory is intended to orient our quest for justice, but if the ideal (i.e., global optimum) lies outside our neighborhood it's hard to see how it can helpfully orient our search for justice. Indeed, as we try to optimize within our neighborhood we may move further away from the ideal. Our question, then, is whether we can mitigate the impact of the neighborhood constraint on our search for the ideal?

[34] Rawls, *Justice as Fairness: A Restatement*, edited by Erin Kelly (Cambridge, MA: Harvard University Press, 2001), pp. 69ff. Recall that we have explained how to make sense of the idea of a continuum of structures in section 2.2.

Of course, if we possess synoptic knowledge of the entire landscape such problems won't arise. As per usual, omniscience makes many problems easier to solve. It is important to note, however, that problems of the sort described above are liable to arise anytime our neighborhoods include something less than all possible worlds. For instance, we can imagine extending our knowledge to worlds that are even further away from our own so that our neighborhood now includes, *d-h*. In this case, though, *e* is still our local optimum, and so exploring a larger neighborhood fails to get us any closer to the global optimum.

3.3 Searching Wider for the Ideal: Diverse Heuristics

The problem confronting epistemically-bounded creatures (and, alas, even philosophers are) is to devise ways to explore different parts of a landscape, so that we might obtain, for example, information about the justice of as many different social worlds as possible. Interestingly, a number of convergent analyses have recently demonstrated that uniform adoption of the "best" search method is a surprisingly inefficient way of doing this.[35] These modeling and analytic results are impressive. Under plausible conditions, a community of inquirers employing a variety of "inferior" search methods will outperform communities where we all employ the best method. We can incorporate the idea of individuals employing diverse search methods into our model by introducing the notion of a *heuristic*. The idea which, like that of a perspective, we borrow from Hong and Page, is that individuals utilize rules (which may vary from simple rules of thumb to more complex algorithms) in order to navigate the neighborhoods in which they find themselves. More specifically, given the world we actually find ourselves in, a heuristic tells us which social worlds to explore (and in what order).

To begin to see the reasoning behind the results described above, suppose that we assume that Karl Popper's incremental, "piecemeal" approach to searching for

[35] For modeling of scientific optimization problems with this result, see Michael Weisberg and Ryan Muldoon, "Epistemic Landscapes and the Division of Cognitive Labor," *Philosophy of Science*, 76 (April 2009) pp. 225-252; Fred D'Agostino, *Naturalizing Epistemology* (New York: Palgrave Macmillan, 2010). For an analysis of economic optimization, with an analytic proof of the superiority of diverse search methods, see Lu Hong and Scott E. Page, "Problem Solving by Heterogeneous Agents," *Journal of Economic Theory* vol. 97 (2001): 123–163.

feasible improvements is best.[36] Rather than considering wholesale radical social innovation, we use our best social science to predict the new social world that will be produced by modest changes in current norms and institutions, and we then implement those changes and observe the results. If the results are positive, we might investigate further such changes; if not, we can undo the changes and revert back to the *status quo*.[37] In terms of our model, this is equivalent to a conservative "climb the gradient" search procedure, and, if one finds oneself on a slope, it will allow her to climb up to the peak. The downside to such a procedure, though, is that it will get stuck at the first local optima it comes to. In terms of Display 4, starting at *f* we can climb to *e*, but having arrived at it, the procedure stops there as all further experiments will result in a decrease in justice.

An alternative to the Popperian procedure would be to identify some justice score that one will not go below (this might be updated as we go along), and to keep experimenting until that bottom threshold is reached. This procedure will not get caught at local optima, and so has the virtue of allowing us to explore more terrain and potentially identify better optima. Conversely, though, it may also lead us down into a valley. If we think of the issue in terms of actual social experiments, this search technique is consistent with federalist style experimentation in which different communities implement different policies. As long as an experiment does not turn out too badly (so that its score is below the minimum acceptable level), communities will keep on experimenting. Other communities can then discard unsuccessful experiments and adopt the more successful experiments of their neighbors. While embracing federalism may lead some communities to do worse than they otherwise might, the important point is that it greatly increases our odds of identifying optimal policies.

Of course, not all investigations of alternative social worlds need to engage in actual social experimentation. A philosopher might, for instance, model far off social worlds and then report back to the rest of us on what they have discovered. Indeed,

---

[36] Karl R. Popper, *The Poverty of Historicism*, 2nd edn. (London: Routledge and Kegan Paul, 1961), esp. chap. 20.

[37] Of course, whether this is politically feasible in the sense that policy makers will admit their mistakes and politicians will be willing to reverse course presents an even more complex feasibility issue.

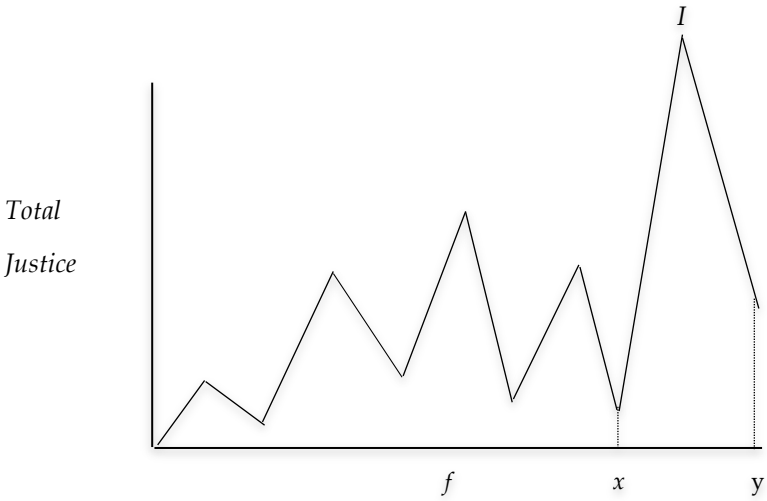
there is a sense in which this is precisely what political philosophy, starting at the beginning in Plato's *Republic,* has done.[38] In this case our heuristics tell the philosopher which worlds to model, and we can imagine the utopian theorist modeling distant social worlds and reporting back on the incredible heights of justice that can be reached *if only enough elements of our social world are simultaneously changed*. To flesh out our story further, we might analogize this approach to a simple "filling out the map" model of searching on which inquirers explore worlds both near and far, but where crucially, once an inquirer has explored a far-flung part of the landscape, the rest of us simply accept her reports as veridical and add that information to our maps. The upshot of this sort of model is that once we have explored enough of our landscape we will eventually confront a "simple" feasibility problem of how to get to the global optimum from wherever we happen to be. This, of course, is because the collective result of our searches will be that we now know the entire landscape, and the global optimum could again play its orienting role.

As Fred D'Agostino has stressed, though, this common story ignores a fundamental question, viz. whether I will be able to fit together explorers' reports with what I now believe, given the social world I inhabit and the tools that I think are reliable?[39] To put it mildly, some of us are skeptical of Plato's modeling, just as some of us are skeptical of claims about how a social world of "property-owning democracy" or "democratic socialism" might work.[40] Although modeling far-flung social worlds can be a useful and often enlightening enterprise, too little attention is paid to the accuracy of these models. We must be careful to ask whether what a model purports to be true of social world x is actually true about some other social world in its general neighborhood. Note, too, that accuracy can be a critical problem in rugged optimization landscapes. Consider Display 5:

[38] We put aside here whether the *Republic* is satire, irony, or something else entirely; we suppose it can be read as an ideal construction.

[39] D'Agostino, *Naturalizing Epistemology,* esp. chap. 1.

[40] These, of course, are the two regime types Rawls recommends in *Justice as Fairness*, Part IV.



DISPLAY 5: "BALLPARK" UTOPIANISM

Suppose that a utopian theorist provides us a model of the globally optimal social world that is, as Americans say, somewhere within the "ballpark" of *x* to y. If we are now at *f,* and the modeling of the social world at the ideal *I* is only accurate within the range *x*–*y*, then, even if we accept the normative evaluation of *I*, the attempt to achieve it may lead us to a far less just social world than our present state.

This is not just a peculiar result of a finicky model, either. Rather, it is fundamental to the problem of social reform. All-too-often reformers and revolutionaries have implicitly assumed that they are confronting Mount Fuji problems, where, even if the revolution fails to bring about the utopia that the reformers had in mind, it will bring us "closer" to the ideal and so leave us in a better state than we presently occupy. In rugged landscapes, though, *closer and better do not march hand-in-hand*. And, the appalling consequences of the great (distinctly anti-Popperian) social experiment of the twentieth century, communism, have shown us that we confront a decidedly rugged landscape.

4. MITIGATING THE NEIGHBORHOOD CONSTRAINT

4.1 Minimal Perspectival Diversity

The costs and benefits of employing different heuristics (or search strategies) dominates much of the literature on exploring rugged landscapes. However, once an

inquirer seeks to evaluate social worlds beyond our neighborhood, our rationale for adopting the neighborhood constraint will lead us to doubt her findings. *Ex hypothesi*, she is making claims about the justice of social worlds in which we cannot be confident. If we are to search more widely and accept the reasoning behind the neighborhood constraint, then we must explore ways to expand our ken consistent with the neighborhood constraint. Is there any way to do so?

Recall again the five elements of a perspective: (A) a set of evaluative standards or principles of justice; (B) an identification of the relevant features of social worlds; (C) a mapping relation from (A) to (B); (D) an ordering of the underlying structures which meaningfully relates them in terms of similarity; (E) a distance metric. Thus far we have been supposing that everyone shares all these features so that they have the same perspective and simply adopt different search procedures while exploring the same optimization landscape. In other words, though everyone agrees precisely about what the normative landscape looks like, by employing different search procedures they are able to explore more of it. Complete agreement on (A)-(E) is a rather extreme assumption, though. How likely is it that different individuals will all agree about exactly what a normative landscape looks like? Interestingly, however, analytic results indicate that if we relax the assumption of a common perspective, and consider searches among individuals who look at the world differently, results can be greatly improved.[41]

Let us, then, introduce a minimal degree of perspectival diversity into our model. Suppose, in other words, that the investigators in our model now all agree on every element of their perspectives except the *metric of distance* between social worlds (§3.1). That is, they agree on (A) – (D), but now disagree on (E).

To illustrate the significance of this sort of perspectival diversity we can begin by exploring the idea of a *distance contracting metric*. Let us define a distance contracting metric as any metric that increases the effective size of a neighborhood of social worlds relative to some other metric. Consider for instance the most minimal and straightforward way in which two distance metrics, d_1 and d_2, might differ from one

[41] See Lu Hong and Scott E. Page, "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, No. 46 (Nov. 16, 2004): 16385-16389; Hong and Page, "Problem Solving by Heterogeneous Agents."

another, namely if d_2 were to be a linear transformation of d_1. In this case if $d_2 = kd_1$ where $k \in (0,1)$, then d_2 would be a distance contracting metric relative to d_1, and a perspective Σ_2 utilizing the d_2 metric will view moves between certain social worlds (say, from our current socio-economic system to property-owning democracy) as moves within our neighborhood, while Σ_1, employing the d_1 metric, will see these moves as beyond our ken. The result of this sort of differences is likely to be debate about the real size and scope of our current neighborhood, and what constitutes a plausible experiment. An upshot of this debate sometimes will be the effective expansion of our neighborhood. Should those with distance contracting metrics like d_2 convince others that Σ_2 is a plausible perspective, we take a small but significant step towards mitigating the conflict between local experimentation and the search for global optima.

Once again, it is important to stress that the benefits of perspectival diversity are not simply an upshot of our formal representation of the problem. Consider again Mill's case for what he called "socialism." From Mill's perspective Victorian capitalism fell far below the moral optimum; a form of society centered around worker cooperatives was far better, and perhaps even the ideal. Mill did not simply analyze this ideal, though. Instead he sought to show how a society that might appear very far from the one he inhabited could be achieved via the institutions already in place. Mill insisted that the evolution of new forms of partnerships and corporations that render capitalism more efficient would also allow competitive market processes within capitalism to test the viability of socialist experiments, and by connecting the idea of worker cooperatives to a series of intermediate social worlds, he sought to bring socialism into the neighborhood of Victorian capitalism. Rather than a leap into the dark, Mill depicted socialism as a form of industrial organization within the current neighborhood.

4.3 Deeper Perspectival Diversity

The previous section focused on the number of social worlds that we can meaningfully model, using distance between worlds as the benchmark. Notice, however, that because the two perspectives we discussed are linear transforms of each other, they induce the same orderings of social worlds with respect to their

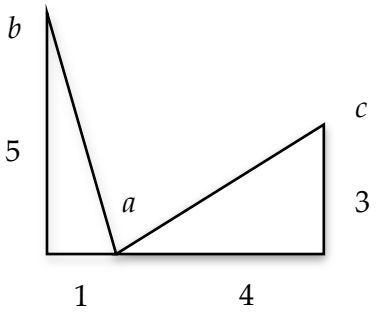
distance from the actual world (element D is the same in both perspectives). This is significant because it means that disagreements between theorists with perspectives Σ_1 and Σ_2 will not extend to which worlds are closer, but will instead be limited to how far away various worlds are. This is obviously a very limited type of diversity, and we may expect that it will have some, but not great, advantages in expanding the size of our neighborhoods. A somewhat deeper type of diversity is implicit in the notion of *equivalent metric spaces*. As we described in section 2.2, metric spaces are ordered pairs denoted (X, d) comprised of a domain X and a distance metric d (a function defining the distance between objects in the domain). Two metrics, d_1 and d_2, are said to be *strongly equivalent* when they are capable of inducing the same topology on a space, a sufficient condition for which is that there exist two constants, α and β, such that for every point in a domain (i.e. $\forall\ x,y \in M$): $\alpha d_1(x,y) \leq d_2(x,y) \leq \beta d_1(x,y)$.[42] Roughly speaking, equivalent metrics are capable of inducing the same sets over a given space. This is important because it means that sets described by one metric can be similarly described by an equivalent metric. Crucially, however, equivalent metrics will not always agree on which of two points in a space is closer to a third, and, as we are about to see, it is this fact that allows clever inquirers to more efficiently explore a landscape.

So far our discussion of distance metrics has all been rather abstract. To make things more concrete it will help to consider a toy example. Arguably the two most familiar metrics, which happen also to be equivalent, are the Euclidean metric and the taxi-cab metric (sometimes referred to as the Manhattan norm or box metric). The Euclidean metric defines distance according to the familiar formula for the shortest linear distance between two points: $d_{Euc}(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$ (which in two dimensions simplifies to the formula for the length of the hypotenuse of a right triangle).[43] The taxi-cab metric, on the other hand, defines distance in two-dimensional space according to the shortest distance between two points from the perspective of taxi navigating a rectangular street grid (hence the name), which in higher dimensional space can be represented as: $d_{Taxi}(x,y) = \sum_{i=1}^{n}|x_i|$. As Display 6

[42] Two metrics are *weakly equivalent* when for every pair of points in a domain there exist constants, α and β, such that $\alpha d_1(x,y) \leq d_2(x,y) \leq \beta d_1(x,y)$, where these constants need not be the same for different pairs of points.

[43] Notice that we are now supposing an N-dimensional similarity space.

indicates the Euclidean and taxi-cab metrics will sometimes disagree about which of two points is closer to a third:



$$d_{Euc}(a,c) = 5 \quad < \quad d_{Euc}(a,b) \approx 5.1$$
$$d_{Taxi}(a,c) = 7 \quad > \quad d_{Taxi}(a,b) = 6$$

DISPLAY 6: COMPARISON OF EUCLIDEAN AND TAXI-CAB METRICS

Because the Euclidean and taxi-cab metrics are equivalent, though, their disagreements will always be limited to the relative distance of points that lie within similar neighborhoods. In other words, equivalent metrics will never disagree about which of two points is relatively closer when, from at least one perspective, one of these is far away, while the other is close.[44]

To further illustrate the significance of metric equivalence consider two tourists, Alf and Betty, who are sightseeing in Manhattan. Suppose Alf and Betty agree about all the sights they want to see and how important it is to each of them to see each sight, so that they agree on the following rank-ordered list of destinations to see while they are in town: {Empire State Building, the Met, Statue of Liberty, Natural History Museum, Times Square, Central Park Zoo, Tom's Restaurant, Madison Square Garden}. Given the number of sights they want to see, however, Alf and Betty recognize that they will not be able to see all of the destinations on their list in the short time they have in the city. To see as many sights as possible, then, they agree to explore Manhattan in a way that maximizes the number of sights they can visit in a day and it is here that employing multiple distance metrics can help our sightseeing pair. Assume, for instance, that Alf who has always had trouble doing arithmetic in his head prefers to measure distances according to the taxi-cab metric so that he can avoid dealing with squares and roots, while Betty who is better with

[44] This is because for equivalent metric spaces disagreements about the distance between pairs of points are always bounded by the constants α and β.

numbers prefers to use the Euclidean metric. In this case, the pair might choose to skip the Statue of Liberty because they both agree that it is too far from all the other sights, but they might disagree about the relative distances of some of the other sights on their list. However, these disagreements can be useful. For example, when they are near the park Betty can identify the better routes for the pair to take because it is possible to take direct routes through the park (say between the Met and the Zoo). Once they leave the park, however, this is no longer possible, and Alf's metric becomes the better guide as the pair decides which route to take in order to minimize the number of blocks they walk. Provided that our Seinfeld loving pair can agree about who should take the lead when, then, their disagreements about the distance between destinations might allow them to squeeze in a stop by Tom's Restaurant on the upper west side that a less efficient route would have precluded.

As we have now seen, equivalent metrics characterize landscapes in subtly different ways. In particular, note that allowing more diversity about the metrics (element *E*), leads to some diversity in orderings (element *D*). Although equivalent metrics will agree on what is near and what is far away, they may order the various "nearby" options differently. We might call this consequence *diversity contagion*: allowing diversity at a lower level induces diversity at a more fundamental level. As we are about to see, this sort of contagion has important implications for our search for the ideal.

5. THE UTOPIA IS AT HAND THEOREM

According to what Page calls the Savant Existence Theorem: "For any problem there exist many perspectives that create Mount Fuji landscapes."[45] There are always arrangements of the elements in *X* (social worlds) that create Mount Fuji landscapes. Showing this is trivial in a one-dimensional similarity space: take the ordering of scores on the *y*-axis from high to low, and rearrange the *x*-axis to correspond with this ordering. This will yield a Mount Fuji landscape. There are many such landscapes. Take the highest justice score and place that social world at any point on the *x*-axis, and then rearrange the other social worlds so that on each side of this point the further one goes, the less just the social world. This too will yield a Mount

[45] Page, *The Difference*, p. 47.

Fuji landscape. If we can show that our problem is a smooth optimization landscape, then the conflict between local and global optimization is entirely obviated (§2.3).

A more modest version of the Savant Existence Theorem might be called:

> *The Utopia is at Hand Theorem:* There are always perspectives that show that the ideal is within our current neighborhood.

This is a more modest version of the Savant Existence Theorem as it does not require a reordering of the similarity dimension such that for all social worlds in X there is a smooth optimization landscape. First, it "only" requires that a subset of X, which includes the current world and the global optimum (and we assume some other near-by social words) are ordered such that they form a neighborhood. Neither does it require that within this neighborhood there is a smooth, Mount Fuji, landscape. "All" that is required is that the global optimum is within our current neighborhood.

It is easy to rearrange an x-axis. What is not easy is to show two things. (1) If a proposal (utopian or not) is to make sense to others they must be able to see this new arrangement of the set X as exemplifying a meaningful structure that relates the social worlds (§2.2). It is not enough simply to propose an ordering of the social worlds that brings utopia within our neighborhood. Rather, the underlying structure that an ordering manifests must make sense. Recall our initial discussion of the distance metric (§2.2). There we supposed that according to perspective Σ, our current social world a is defined by a set of features $\{f_1, f_2, f_3\}$. We then postulated alternative social worlds, b and c, with features $\{f_1, f_2^*, f_3\}$ and $\{f_1, f_2, f_3^*\}$, respectively, and given Σ's identification of these basic relevant features, the procedure for ordering nearby worlds was to ask whether feature f_2^* in b was more or less similar to f_2, than feature f_3^* in c was to f_3, and similarly for worlds d, e, f, etc. Now, if our utopian savant comes along and gives us a radically different ordering of the social worlds, it must either be the case that (*i*) her similarity judgments are simply wildly different; through she sees the precise same features as relevant, she entirely disagrees about what world is similar to which others. But this looks deeply puzzlingly. Take a prosaic example: suppose we are ranking in terms of similarity, a Ford F-150 pickup, a Chevy Silverado pickup, and an elephant, and we agree on precisely what they are. Suppose, then, that our savant proclaims that the elephant is

more similar to the F-150 than is the Chevy Silverado to the F-150. But how are we to make sense of this, if she agrees on precisely what they are? Alternatively, and more sensibly, (*ii*) she sees different features of the social worlds as relevant to justice — that is, she proposes that we should focus on different features of social worlds in making our similarity judgments. Notice that, in the latter case, our savant does not share element B of a perspective with the rest of us, and so we see the diversity contagion rearing its head.

(2) This first point is widely recognized. Unless a different perspective imposes a meaningful (and, it should be added, attractive) structure on the world, it will seem more like craziness to us than like insight. But what is not sufficiently appreciated in the literature is how odd the savant utopian's case is.[46] As we have seen, the utopian must be saying that our current perspective leads us to believe that the ideal lies outside our current neighborhood only because we are mistaken about what the fundamental features of the social worlds truly are. Display 7 makes this concrete with an example of a six-world set (X), with three-world neighborhoods.

World	*Justice*	Σ_1 *features*	Σ_1 *neighborhood*	Σ_2 *features*	Σ_2 *neighborhood*
a	10	$\{f_1, f_2, f_3\}$	✓	$\{f_1, f_4, f_5\}$	✓
b	15	$\{f_1, f_2, f_{3^*}\}$	✓	$\{f_1^*, f_4^*, f_5^*\}$	
c	5	$\{f_1, f_2^*, f_3\}$	✓	$\{f_1^*, f_4, ^* f_5\}$	
d	15	$\{f_1^*, f_2, f_3\}$		$\{f_1^*, f_4^*, f_5^*\}$	
e	20	$\{f_1^*, f_2, ^* f_3\}$		$\{f_1, f_4, ^* f_5\}$	✓
f	30	$\{f_{1^*}, f_2, ^* f_3^*\}$		$\{f_1, f_4, f_5^*\}$	✓

DISPLAY 7: RADICALLY DIFFERENT WORLDS WITH THE SAME JUSTICE SCORE

Notice first that Σ_1 and Σ_2 concur on the justice score of each world in the set (column 2). Σ_1 identifies certain fundamental features $\{f_1, f_2, f_3\}$, of *a*, our current social world, and then determines which other worlds, with slightly different features, are most similar (in this case *b* and c). Thus on Σ_1 the local optimum is *b*, with a justice of 15. Σ_2, our utopian savant perspective, identifies *a*'s fundamental features as $\{f_1, f_4, f_5\}$,

[46] Or, in the more general literature, the radically better perspective.

and given its similarity judgments, the most similar worlds are *e* and *f*. And, of course, *f* is both the local and the global optimum. So on Σ_2 the ideal is within the current neighborhood.

But Σ_1 must be perplexed. The utopian perspective radically disagrees about what the relevant features of the social world are, yet (somewhat miraculously) concurs on their justice. Why should a world of $\{f_1, f_2, f_3\}$ and of $\{f_1, f_4, f_5\}$ have the same justice score? If there are different features being evaluated, then we would expect variance in their justice. Of course there may be cases where the justice of a world as determined by different criteria converge, but it would be fantastical to expect (let alone assume) that this could be true for every social world.[47] Surely Σ_1's most reasonable conjectures are either that the utopian perspective is simply erroneous, or that it is using a different set of evaluative standards too, and thus, that Σ_2 does not in fact share element A of Σ_1's perspective (note the further diversity contagion). If, however, the utopian is confused, or has very different evaluative standards, it is not clear that Σ_1 has much to learn from Σ_2. At the very least, it's hard to see how Σ_1 could be expected to recognize that he can learn from Σ_2.

6 THE FUNDAMENTAL DIVERSITY DILEMMA

Embracing multiple perspectives can improve our ability to explore landscapes, especially when these perspectives utilize equivalent metrics (§§4.2-4.3). This is because perspectival diversity has the potential to increase the effective size of our neighborhood as well as alter our view of the terrain. Certain perspectives will view some social worlds as distant that other perspectives see as quite close by. Similarly, a social world that is a local optimum on one perspective may not be on a different perspective. Reflecting on these points allows us to see why equivalent metrics, in particular, are important. When we disagree about which worlds are in our neighborhood it may be hard for us to reach consensus about what to do or which social worlds it would be desirable for us to move to. Perspectives utilizing equivalent metrics, however, tend to agree about which worlds are in our

[47] The core Hong and Page theorem assumes problem solvers always agree on the evaluative function. As we think we've shown, this is a very strong assumption. Hong and Page, "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers." Note that Page considers evaluative diversity in *The Difference*, Part Three.

neighborhood (they simply disagree about which of these worlds are closest). In other words, equivalent metrics paint different pictures of the local landscape, but broadly similar portraits of the global landscape. This minimizes the chance of getting stuck at a poor local optimum within our neighborhood, without substantially inhibiting our ability to talk to one another.

As we have seen, if we extend perspectival diversity a bit deeper, our chances of identifying global optimum are much improved, while our chances of getting stuck at a local optimum are (at least potentially) reduced. Because non-equivalent metrics induce different topologies on a space, though, perspectives with non-equivalent metrics will have difficulty taking each other's results seriously. Not only will perspectives with non-equivalent metrics disagree about which worlds are in a neighborhood (which can sometimes be useful), they will radically disagree about where these worlds are, and what features characterize them. Indeed, given that the perspectives do not ascribe the same features to the social worlds, they might not even be able to communicate to other perspectives *which worlds they are describing*. Recall Display 7: if one perspectives is analyzing a world composed of $\{f_1, f_2, f_3\}$ while the other's sees the structure as based on $\{f_1, f_4, f_5\}$, in what sense are they talking about the same social world at all? Radical disagreement, in other words, may prevent us from characterizing certain social worlds in terms others are capable of understanding.

Nor is diversity only a problem when we disagree about the relevant features of a world. Even if we agree that the relevant features of our world are f_1, f_2, and f_3, if we radically disagree about how to measure similarity (or distance), a modification to some relevant feature of the world that I consider to be relatively minor might appear quite radical to you. For instance, I might ask you to imagine a world that is otherwise like ours, but in which people are slightly more equal, though at the cost of being slightly less free, and I might judge that world to be superior to our own. If you have a different conception of what counts as slightly less free, though, you might imagine an entirely different world – one which you, reasonably, might think is much less just than our own – and in this case, it is almost inevitable that we will find ourselves talking past one another.

The lesson is clear. A community of inquirers into justice drawing on diverse perspectives — in terms of metrics and similarity judgments — can, in principle, overcome the conflict between local experimentation and discovery of the ideal. As our perspectives become more deeply different, though, our ability to communicate with one another is hampered.[48] At some point – and certainly when we radically disagree about what constitutes the relevant neighborhood – the costs of perspectival diversity are likely to outweigh the benefits. Although embracing deeper diversity may improve our odds of identifying the global optimum it likely reduces our ability to meaningfully utilize this information. Indeed, empirical evidence tends to confirm this, indicating that those with very different outlooks have more trouble communicating and coordinating with one another.[49]

Furthermore, even when perspectival diversity does not hinder our ability to communicate it can present a different sort of problem. While we've shown that utilizing equivalent metrics can improve our ability to identify better social worlds, when we turn our attention away from the normative projects of identifying the more just worlds, and towards the more practical project of social reform, even equivalent metrics begin to pose problems. While equivalent metrics induce the same topology on a space they do not preserve certain other important characteristics of functions. Most notably, it is possible for a function to be a contraction-mapping in one metric space but not in another equivalent metric space, and when our concern is not just with whether there is a better social world in our neighborhood, but also with how far away various potentially better worlds are, this turns out to matter quite a bit.[50] More specifically, if we imagine a function telling us which world to move to that is a function of both how good a world is and how far away it is, then which world the function tells us to move to will depend crucially on

---

[48] Page is aware of this. See *The Difference*, p. 49.

[49] Roberto A. Weber and Colin F. Camerer, "Cultural Conflict and Merger Failure: An Experimental Approach, *Management Science*, Special Issue on Managing Knowledge in Organizations: Creating, Retaining, and Transferring Knowledge, vol. 49 (April 2003): 400-415. For a model showing that meaningful propositions emerge from shared overall perspectives, see Brian Hazelhurst and Edwin Hutchins, "The Emergence of Propositions from the Co-ordination of Talk and Action in a Shared World," *Language and Cognitive Processes*, vol. 13 (1998): 373-424.

[50] A contraction-mapping is a function for which $d(f(a),f(b)) \leq d(a,b)$. Several important results in mathematics regarding the existence of equilibria depend on certain functions being contraction-mappings.

which metric we use for evaluating the distance between worlds. Here, then, diversity which was once helpful threatens to cause new problems.

This brings us to the Fundamental Diversity Dilemma. Once we acknowledge that as perspectives become more deeply diverse their ability to effectively communicate is severely impaired (as is their ability to see each other as peers), we confront the sobering reality that optimization in rugged landscapes may present intractable problems. If we keep perspectives similar enough so that they will acknowledge that they are engaged in a similar problem-solving task with the same evaluative standards, they are almost sure to get caught in the same neighborhood. They may communicate in ways that encourage some expansion of the neighborhood or small changes in its similarity orderings, but by and large they will be pursuing whatever local optimum lies in their neighborhood, and, as a result, the global ideal cannot perform its orienting function (§1.1). Alternatively, we can allow deeper perspectival diversity, in which case it is more likely that some inquirers will find what they consider to be the ideal, but this can only come at the cost of introducing disagreement about exactly what constitutes the ideal. Thus, even having found what was once thought to be a mythical Paradise Island, an inquirer is unlikely to be able to convince others (at least not all others) to join her. Moreover, and this is why we have suggested that the problem is intractable, *no approach* to political philosophy (be it ideal or non-ideal) can avoid confronting this dilemma and the choice it poses.

Appendix

A QUICK GUIDE TO CONSTRUCTING A SINGLE-DIMENSIONAL ORDERING OF WORLDS

The basic idea we employ in section 2.2 is a similarity judgment of the form "*a* is more similar to *b* than *a* is to *c*." Let us use the following notation for this claim: $(a\sim b)>(a\sim c)$. We suppose that is this is a transitive relation: if $[(a\sim b)>(a\sim c)]$ & $[(b\sim c)>(b\sim d)] \rightarrow [(a\sim c)>(a\sim d)]$. In the text we claim that a series of these judgments can yield an ordering of social worlds. In might be queried how the basic similarity judgment can give rise to orderings.

To see how, we start with three social worlds, *a, b,* and *c*. Now, let us see how a specific ordered triplet can be derived from iterated applications of the basic similarity relation. We can begin by asking: what ordered triplets of the set {*a,b,c*} are consistent with the finding that $(a\sim b)>(a\sim c)$? Not all, but many: $(a\sim b)>(a\sim c)$ is consistent with {*a,b,c*}, {*c,b,a*} {*c,a,b*} OR {*b,a,c*}. We can thus exclude {*b,c,a*} and {*a,c,b*} as those orderings are not consistent with our initial similarity judgment.

Note that in an important sense {*a,b,c*} is equivalent to {*c,b,a*} insofar as, at least initially, we don't care whether our similarity orderings run (as it were) from the right to left or left to right. It would be the same dimension of similarity in either case. As long as once we impose a direction we stick with it, there is no worry about which way the dimension runs. (In a similar way {*c,a,b*} is equivalent in this sense to {*b,a,c*}.) Because it does not matter whether our dimensions go right to left or left to right, we can just pick a direction that constrains further orderings. Suppose we choose {*a,b,c*}. We can thus set aside the {*c,b,a*} ordering. That leaves us with the equivalent {*c,a,b*} and {*b,a,c*}, where, for the other alternative dimension, we choose {*c,a,b*}, dropping the opposite "direction" of {*b,a,c*}. Now, having done this for the first triplet, we have imposed a direction on our dimension, and we have that $(a\sim b)>(a\sim c)$ is consistent with {*a,b,c*} OR {*c,a,b*}.

Now suppose a second judgment, employing the method in section 2.2: $(c\sim b)>(c\sim a)$. This is not consistent with {*c,a,b*}. (Neither is it consistent with {*b,a,c*}, the "other direction" of the {*c,a,b*} ordering). So we are left with {*a,b,c*}. Notice that we have taken multiple similarity judgments and derived an ordered triplet that arrays the options along a dimension. We thus have generated "*in-betweeness*" (*b* is in-between *a* and *c*) from "more similar than" relations. We can build out from the

ordered triplet, to yield an entire dimension in terms of similarity, or in terms of our model, an ordered set of all social worlds in the set. Note that this is more enriched metric information than an ordered ranking. If we have {*a*,*b*,*c*} we can find out whether *b* is more similar to *a* or *c*, giving us more than in-betweeness, as we also get "in-between but closer to."

Of course a one-dimensional similarity dimension is often too simple. Consistent distant metrics can also be worked out for *N*-dimensional space, though, using a more complicated version of this procedure involving a sort of triangulation. The important point is that we do not simply know the relation of other social worlds from our own, but their relation to each other.