

Cornell Law Faculty Publications

Cornell Law Library

Year 2008

Pareto versus Welfare

Robert C. Hockett
Cornell Law School, rch37@cornell.edu

CORNELL LAW SCHOOL

LEGAL STUDIES RESEARCH PAPER SERIES



Pareto versus Welfare

Robert Hockett

Cornell Law School
Myron Taylor Hall
Ithaca, NY 14853-4901

Cornell Law School research paper No. 08-031

This paper can be downloaded without charge from:
The Social Science Research Network Electronic Paper Collection:
<http://ssrn.com/abstract=1309699>

PARETO VERSUS WELFARE

BY ROBERT HOCKETT*

Many normatively oriented economists, legal academics and other policy analysts appear to be “welfarist” and Paretian to at least moderate degree: They deem positive responsiveness to individual preferences, and satisfaction of one or more of the familiar Pareto criteria, to be reasonably undemanding and desirable attributes of any social welfare function (SWF) employed to formulate social evaluations. Some theorists and analysts go further than moderate welfarism or Paretianism, however: They argue that not simply liberal values, but “the Pareto principle” itself requires the SWF be responsive to individual preferences alone – a position I label “strict” welfarism – and conclude that all social evaluation should in consequence be formulated along strictly welfarist lines. I show that no strictly welfarist social welfare function can give complete expression to a normative social evaluation, as distinguished from simply describing a social allocation. I show also that SWFs employed to formulate complete social evaluations must accordingly, in virtue of the entailment relations obtaining among strict welfarism and the several Pareto criteria when these latter are interpreted along preference-regarding lines, be constructed without regard to Weak, Strong or Indifferentist Pareto. The results derived here generalize Arrow’s (1951), Sen’s (1970), and cognate impossibility results.

Keywords: Pareto, Paretianism, welfare, welfarism, welfare economics, social welfare function, social choice, impossibility result.

1. INTRODUCTION

PARETIAN CRITERIA FOR SOCIAL EVALUATION have been with us for about a century now. (See Pareto (1909).) The principal reason for their persistence of course is that they are widely perceived, at first blush at any rate, to be normatively “attractive.” (See, e.g., Basu (1984), Sen (2003), Tungodden (2003).)

Pareto’s attractions for their part are not altogether mysterious. If a shift from one law or policy set to another stands to render one or more people better off and nobody worse off, why *wouldn’t* “we” undertake it? Abstention would seem to forgo a costless increase in something that, as the term “better” suggests, is “good.” That would in turn be a thing we can plausibly think of, as “good’s” adverb “well” certainly invites, as “well-being” or “welfare.”¹ Needless surrender of welfare for its part would seem to be wasteful, a gratuitously occasioned opportunity-cost. It is scarcely surprising in this light that so-called “Weak” and “Strong” Paretian criteria, which trade on the picture just sketched, are widely described as “efficiency” criteria.²

Weak and Strong Pareto, of course, do not exhaust the menu of Paretian attractions. Distinctly but near as compellingly comes this thought: If no individual in a

* Copyright © 2008. Thanks to [Matthew Adler, Jon Bakija, Kaushik Basu, Brian Bix, Robin Boadway, Mike Dorf, Matti Eklund, Dan Farber, Marc Fleurbaey, Rick Geddes, George Hay, Douglas Kysar, Jerry Mashaw, Alistair McCleod, Cherie Metcalf, John Roemer, Josh Teitelbaum, Bertil Tungodden, participants at Cornell, Penn, and Queens workshops, the editors and a referee] for helpful comments, questions and suggestions. Thanks of course do not imply agreement.

¹ “Welfare” derives from “well faring,” the “fare” here in turn stemming from French “faire,” meaning to do or to make. One is “faring” well when one is “doing” well, which is when “things are well” with one.

² Strong and Weak Pareto are duly formalized *infra*, Section V.

particular society fares better under one given law or policy set than some other, shouldn't "society as a whole" be considered indifferent as between the two sets? Surely a society's faring well, ill, or indifferent is derivative in some sense of its individual members' thus faring. Permitting a measure of social welfare to vary in response to matters apart from such person-centered groundings, moreover, would seem to invite fetishism or "perfectionism" of some sort. (See, e.g., Arneson (1990), Hurka (2002), Rawls (1971), Sen (1979). Also Hockett (2009a, b, c).) We abandoned that sort of thing long ago, along with state-sponsored religion and cognate public *teloi*. We did that at least partly on grounds that respect for our fellows as agents demands we respect their autonomy in judging their own good. So-called "Pareto Indifference" is thus often taken to give fair expression to widely accepted, modern "liberal" or "individualist" values.³ (See, e.g., Samuelson (1947). But see also Sen (1970a). See generally Hockett (*idem.*).

Considerations of the sorts just articulated generally prompt social welfare theorists, as well as some economically trained legal and policy analysts, to take one or more of the familiar Pareto criteria legitimately to constrain the construction of those social welfare functions, functionals, orderings, quasi-orderings, partial orderings, and cognate metrics in which they couch much of their research. (See, e.g., Adler (2007), Kaplow & Shavell (2002). See generally Arrow (1951), d'Aspremont (2002), Samuelson (1947), Sen (1970b), Sen (1986). Also Hockett (*idem.*).

Satisfaction of one or more flavors of Pareto is considered a reasonably minimal condition to impose upon any plausible method employed in the comparative normative evaluation of possible states of the world, including states of the world effected or affected by law and policy. (*Idem.*)

It might not be immediately apparent from the foregoing considerations as thus worded, however, precisely what normative *work* Pareto is doing in such social evaluations as are made to comport with it. More precisely, it might not be apparent what work Pareto is *not* doing.

The notions of "better" and "worse off" employed in the second and third paragraphs above, upon which Pareto is transparently parasitic, are as yet still unspecified. I have not thus far characterized any determinate mode of evaluation such as might lend any *content* to the evaluative terms "good," "well," "better," or "worse." I have not said on what grounds – pursuant to what normative standard or standards – ascriptions of "better" and "worse," of the sort upon which application of Pareto still pends, might be made. I have said only that Pareto has something secondary to say about what "we" ought to do, once those primary standards have been antecedently specified. Pareto's normative clutch in the meanwhile has nothing with which to engage; it awaits an anterior driveshaft and engine.

But now I have also suggested, in effect, that at least one consideration seemingly underwriting one of the Pareto conditions – Pareto Indifference – might well militate *against* "our" antecedently specifying any such primary standard at all. For surely our specifying the still requisite primary standard – the "driveshaft and engine" – might violate individuals' autonomy in judging their own good. It might do so in precisely the illiberal way that I said modern polities tend to eschew. It isn't only Pareto Indifference that might raise this prospect, moreover. Inasmuch as the "veto" (per Sen (1970a, b)) latent in Strong Pareto's "no one worse off" constraint mentioned earlier is *itself* understood in autonomy-vindicating terms (see Basu (1984) on this prospect), it too

³ Pareto Indifference likewise is duly formalized *infra*, Section V.

would seem hostile to “our” framing a primary standard. And if it does, then one is left wondering on what grounds “we” might ever be able to recommend or do *anything* qua “we,” since Pareto’s proscriptive, “no one worse off” operation seems to *block* recognition of social *teloi* even as its prescriptive, “someone better off” operation awaits them.

These observations raise an interesting question for legal, policy, and social welfare analysis: If “we” cannot recommend or do anything consistently with Pareto, what coherent conception of “social welfare” can there *be* that is consistent with Pareto? Can any antecedently given primary standard of the sort that can underwrite a conception of social welfare and guide social evaluation or collective action, if consistently applied in the manner that one expects of a normative principle, comport with Paretian criteria?

I have come to believe this is possible only on two complementary conditions, conditions that are not generally met in the voluminous Paretian literature. The first condition is that the normative standard which does the primary work – the standard that Pareto’s prescriptive face transparently endorses – must indeed be a *standard*: It must be incumbent upon preferences rather than merely an object of preferences,⁴ and in this sense must be at least partly definitive of “welfare” itself.⁵ The second condition, which is really just a corollary of the first, is that the antecedently given normative standard must be consistent with *both* of Pareto’s two edges, the prescriptive and the proscriptive. We must not equivocate by first employing the primary standard to determine when one is “better” off, and then some contrary standard – or pseudo-standard, such as normatively unbounded preference-satisfaction – to determine when someone is “worse” off. The “veto,” in short, must not trump the standard, lest there be no standard at all.

These conditions, if sound, are interesting, I think, because most who employ Paretian criteria do not seem to meet them. Most seem, in particular, to read “better” and “worse” as endorsed by Paretian criteria to mean none but “individually preferred” and

⁴ One often hears claims to the effect that the apparatus of SWFs can accommodate “preferences for justice,” “tastes for fairness,” and so forth. These locutions signal, I think, a misapprehension of the logic of normative standards qua standards. Standards can of course be preferred, and one can prefer to act in conformity to standards. But standards are standards by dint of their incumbency even on those who prefer contrariwise. You may prefer, for example, not to put all of us at risk by driving at 120mph through our neighborhood. But the speed limit’s being a “limit” is its applicability even in the event of your preferring to speed.

⁵ It is worth noting once more here the inherently evaluative nature of words such as “welfare” and its cognates, which the root “well” – adverbial form of the word “good” – indeed immediately suggests. “Social welfare” must surely import some conception of “wellness” that “we” – the “society” in question – can agree upon. As emerges below, the minimally intrusive such conception is one to which a conception of fairness sufficient to prompt an SWF’s argument domain and aggregation rule is anterior.

In this connection it bears noting that there is a venerable doctrine, associated with Rawls (1971), to the effect that “the right” – justice – enjoys a form of “priority” relative to “the good” – welfare. I do not think this articulation does full justice to justice. In light of what emerges below, the right is not so much “prior” to the good, in the sense of capacity to “trump” it, so much as it is what I shall call “internal” to it.

That is to say there just *is* no coherent social *conception* of “good” in relation to which a conception of distributive propriety – fairness, or justice – incumbent on individual preferences is not constitutive. Fairness, which finds expression in a social welfare function or cognate metric’s argument domain and aggregation rule, is accordingly as it were the genetic or logical structure of welfare in the same sense that the metric itself is the measure of welfare.

“—dispreferred,” on a person by person basis.⁶ That means they have no primary normative standard with which prescriptive Pareto can engage. And it means that proscriptive Pareto would veto, in any event, any such standard as they might wish to proffer. And this means in turn, I’ve effectively claimed, that their “preferential Paretianism,” as I shall call it, prevents their would-be normative social evaluation itself – which in turn is to say that it rules out any coherent conception of social “welfare” itself.

I can best explicate and substantiate these claims, as well as elaborate their fuller significance, by reference to a two-sided puzzle that is raised when ever a particular social welfare metric – in this case, the familiar Bergson-Samuelson social welfare function (SWF), whose continuity features render it intuitively most readily tractable – is made to comport with Pareto on the one hand, but not the standard-based conditions I have stated on the other. The puzzle takes the form of an indeterminacy in one case, a contradiction in the other.

The indeterminacy arises from the want of any clear means of deciding upon an SWF’s argument domain or aggregation rule in the absence of some preference-incumbent normative standard.⁷ That should be disturbing to would-be “welfarist” – i.e., strictly preference-regarding⁸ – normative legal theorists and policy analysts because the argument domain and aggregation rule are “where the action is” so far as any SWF meant to bear normative legal or policy significance is concerned.⁹

The contradiction, for its part, arises from cashing out Pareto in terms of preference-satisfaction even after having settled upon some such preference-incumbent standard as can determine an SWF’s argument domain and aggregation rule. This will be disturbing to anyone wishing to continue honoring Paretian constraints in constructing their SWFs, while interpreting these – that is, while reading “better” and “worse off” – in terms of individual preference-satisfaction alone. But as we’ll see, there is no reason to wish to do this. Indeed, for any who believe in right and wrong, or good and ill, there is every reason to wish against it.

The remainder of the analysis proceeds as follows. The next Section, Section 2, briefly reprises the standard Bergson-Samuelson SWF model. As noted a moment ago,

⁶ Note that, if we do not read Pareto in this, what I shall call “preference-based” or “preferential” way, there seems little point in employing Pareto at all. The “veto” appears to have been one of the principal reasons for Strong Pareto’s introduction. See Pareto (1909).

⁷ In case the meaning of “preference-incumbent” has been forgotten, see *supra* note 4 and associated text. Preference-incumbent normative standards are normative standards that are incumbent on preferences – which is to say, they are *standards*.

⁸ One can deem preference-satisfaction a critical “input” to or determinant of social welfare without thereby committing to the untenable view that it is the sole such determinant. Many who call themselves “welfarists” today do commit to that view, however. In order to disambiguate as between tenable and untenable preference-regarding views, I shall below introduce the term “preference-regard” for the judgment that preference-satisfactions are determinants of social welfare, and the term “strict preference-regard” for the untenable claim that they are the sole such determinant.

⁹ This might not be as immediately apparent in the case of the argument domain as it is in the case of the aggregation rule, but it is true in both cases. The value judgment implicit in electing to quantify, in the SWF, over all or some people’s preferences is not itself simply a preference. It is a judgment, indeed an inherently distributive judgment like that implicit in the choice of an aggregation rule. Animals’ preferences are not counted in the social welfare aggregate by most social welfare theorists, for example. One suspects that this is not merely because the animals have not asked.

this is the social welfare metric most readily accessible to intuition, which is why I'll employ it. The results I'll derive, however, generalize straightforwardly to more abstract models.

Section 3 then focuses on a question raised by the standard model as currently configured: How are the SWF's argument domain and aggregation rule determined, and how might such considerations as guide that determination find explicit expression in the SWF? The upshot of this inquiry is that social welfare theorists both must and in general do decide upon argument domains and aggregation rules, whether self-cognizantly or otherwise, by reference to normative distributional principles that they take for incumbent on preferences. Such principles are, in effect, the minimal "primary standards" mentioned above upon which Pareto must ride to say anything determinate at all.

It is easy to lose sight of these principles' preference-incumbency and welfare-constitutive role, however, in the standard apparatus as presently constituted. For this is manifest principally in the "shape" of the SWF, and in the cut between what is and is not included in its argument domain, all while attention tends to be drawn only to those items that *are* named in the argument domain.¹⁰

In other words, though theorists must decide upon social-welfare-shaping argument domains and aggregation rules "by reference" to preference-incumbent normative distributional principles, they do not actually "refer" to those principles in the idiom of SWFs themselves.¹¹ They refer to them only in the vernacular while constructing or justifying their SWFs, then tend to forget them, along with their role in defining the theorist's working conception of "welfare," once using and thus "speaking in," rather than "about," the language of their chosen SWFs. This is problematic for purposes of "high theory" if the whole point of the SWF apparatus remains, as originally envisaged, to give comprehensive expression to all values that enter into a social evaluation. (On these purposes, see Bergson (1938). Also Samuelson (1947). And Sections 2 and 6.)

Section 4 next queries what it would be to *give* full and explicit expression, in the comparatively simple formal "language" of SWFs itself, to that preference-incumbent normative standard which guides the choice or design of an SWF's argument domain and aggregation rule. It finds that the one means available, given the formal constraints of standard functional notation, is not presently exploited in the standard model. The most straightforward available means is explicitly to open a variable-space for the normative standards in question – a space in which these values can be explicitly named – then to deem states of the world better, *ceteris paribus*, in virtue of the correct such standards' having been sought, found, and employed in the world.¹²

Section 5.1 then shows that, when we thus supplement the standard model in a manner enabling it to give complete and explicit expression to the value judgments that

¹⁰ Items in the argument domain constitute, as one might say in the current "behavioralist" idiom, the primary "focal points."

¹¹ The sense in which argument domains and aggregation rules are "welfare-shaping" is explained *supra*, note 5, second paragraph. The thing to remember is that the SWF itself affords the measure, and in that sense the working conception, of "welfare" employed by the theorist. Its shape is welfare's shape. And that shape is determined by the argument domain and aggregation rule, hence by what ever considerations of fairness determine the latter.

¹² This imparts to the SWF a recursive structure isomorphic to that of normative inquiry itself, in which we query what we ought to do. More on this fact below.

determine its own construction, the resulting normatively transparent SWF cannot be exclusively “welfarist,” per the received jargon: It cannot be, that is to say, what I shall more precisely call “strictly preference-regarding.”¹³ It must also, in short, be what I shall call “*standard-regarding*.”¹⁴

Section 5.2 then shows that, this being the case, the normatively transparent SWF also cannot be Paretian, if Pareto’s “better” and “worse” be interpreted, as most appear unsurprisingly to interpret them,¹⁵ in terms not of that standard of value which has determined the SWF’s argument domain and aggregation rule, but of individual preferences alone. Since Pareto quite typically *does* seem to be interpreted in terms of individual preferences, this result is apt to be striking. It means among other things that most of the best-known and dreaded impossibility results in the welfare economic and social choice literatures, from Arrow’s (1951) result on down, were all but transparently predictable. Preferential Pareto’s exclusion of preference-incumbent norms results in normative indeterminacy quite as a matter of course.

Section 6 briefly addressed objections that I anticipate to the fine-tuned model constructed in Section 4, hence the results derived in Section 5. Section 7 then concludes with a few thoughts on the future employment of SWFs in explicitly normative law, policy, and social welfare analysis.

In anticipation of what Section 7 will ultimately conclude, I wish to emphasize up front what I am *not* saying here. I am not saying that there are no good uses whatever for preferential Paretian criteria in SWFs of any sort. I am saying, rather, that there seem to be good grounds for purging Pareto from SWFs in at least one of what look to be three dominant usages: Pareto should not figure into SWFs as used in their original Bergsonian “high theory,” normative employments, as would-be “complete” formulations of social evaluative judgments. For preferential Pareto is inimical to preference-incumbent normativity itself.

Where, by contrast, SWFs are pragmatically employed merely as compact means of instructing government functionaries as to what they are to do,¹⁶ or of describing correlations between particular social allocations and associated social welfare measures,

¹³ See supra note 8. I use the term “strict preference-regard” where some theorists use “welfarist.” The reason is that my expression is more explicit about what is meant.

¹⁴ It bears repeating that those who construct SWFs, if they do so for normative purposes, are themselves standard-regarding in the sense I intend. The upshot here is thus that normatively complete SWFs which explicitly reflect all of the theorist’s value judgments per the original Bergson-Samuelson program (Bergson (1938), Samuelson (1947)) will be transparently standard-regarding as well, and will accordingly not be consistently Paretian if Pareto is cashed out in terms of preferences alone.

¹⁵ The preferentialist interpretation is unsurprising because Pareto’s whole point in introducing his criteria seems to have been to avoid difficult distributional questions. See again supra, note 6, and Pareto (1909). That desideratum might be pragmatically justifiable in some circumstances, but we shall see there is no reason for it in welfare theory.

¹⁶ Suppose we have reason to be confident that, so long as functionaries act as to maximize social welfare as defined by a particular SWF, distributions of welfare-relevant resources over citizens will tend to converge toward what those values which have prompted our SWF’s aggregation rule would deem best. In such case we would not find occasion to worry, looking forward, about the Paretian veto’s preventing our doing what we think normatively best. And so in that case Pareto would operate, pragmatically speaking, as no more than an unobjectionable efficiency criterion. We would be able to view the “pie,” that’s to say, as properly divided already, and so Pareto would say no more than to maximize the pie’s size. Pragmatic prospect such as these, I suspect, underwrite Pareto’s intuitive appeal. Here, however, as in other “impossibility” inquiries, our concerns will be matters more of “pure” than “applied” theory.

there might be some room for Pareto – provided that we remain clear and explicit that these are not normative-theoretical uses.

2. THE STANDARD MODEL

As indicated above, the analysis below proceeds by reference to the familiar apparatus of Bergson-Samuelson social welfare functions, which I shall accordingly briefly rehearse in this Section.

The aim of the SWF apparatus as employed in “pure” welfare theory, since its first introduction some seventy years ago, is to afford clarity in respect of the logical structures and consequences of our value commitments. (See Bergson (1938), Samuelson (1947).) It aims to do so by rendering these commitments, along with their analytic relations, transparent. Doing this in turn is thought to enable us straightforwardly to trace all of these commitments’ logical consequences, and incidentally thereby to ascertain their compatibilities and incompatibilities inter se on various possible environments.¹⁷

Familiarly, then, we shall let x (or x' , x'' , etc.) designate a possible state of the world as viewed under a normatively complete description. A normatively complete description includes all details that might be found salient pursuant to some standard-guided method of evaluation or form of prescriptive judgment – that is, judgment concerning a state of the world’s being “better” or “worse” than another such state. These details can, familiarly, include facts about any number of individuals i belonging to a defined population N of n members. It will sometimes be convenient to call x (or x' , x'' , etc.) simply “a world,” or “possible world.”

We shall also assume that there exists a subclass N_t of those persons i among the population N , whom we’ll call *theorists* (elliptical for “welfare, legal, or policy theorists”).¹⁸ Theorists are all persons who engage in normative inquiry or investigation with a view to determining what law and social policy ideally “should” be.

Use of the word “should” here of course suggests that theorists will typically be bringing normative *standards*, or *principles* to bear in their theorizing, principles that they take to be both generally applicable and ethically or morally – in a word, normatively – apt or compelling. In many if not most cases, moreover, it is likely that commitment to such standards, or at minimum an aim to ascertain appropriate such standards, is what prompts and informs the theorists’ decision to engage in normative legal or policy assessment in the first place.

It also bears noting, if only in passing, that the class of theorists N_t might constitute the full population N (such that $N_t = N$), or might instead constitute a proper subclass of N , indeed even a singleton such subclass. A parent might be the sole normative “theorist” in a single-parent family, for example, at least for a while. An academic or other institutionally affiliated “policy elite,” in turn, might constitute the full

¹⁷ All of the most notorious impossibility results are of course proofs of incompatibility of this sort. See *infra*, Section 7.

¹⁸ “Theorists” might also be called, in keeping with a more venerable idiom, “planners,” or “the planner.” See, e.g., Lerner (1946).

class of theorists in some elitist or otherwise moral-labor-divided societies, while the full citizenry might do so in some ideal ethical republic or participatory democracy.

The identity of the theorist(s), and the method of evaluative assessment employed by – i.e., the standards or principles guiding the judgments made by – the same, are among the facts figuring into the aforementioned normatively complete descriptions of possible worlds. Worlds can be deemed better or worse, that is to say, by dint of who the theorists are in those worlds, as well as by dint of there being theorists seeking, and possibly finding, the right normative standards at all.

Now let X be the set of possible worlds invariant in N . We are interested only in possible worlds inhabited by the n individuals i who jointly constitute N .

A *social welfare function*, W , is, familiarly, a (“direct,” i.e., simple, or “indirect,” i.e., composite) mapping from the set of possible states of the world that concern us, X , to the real line, \mathbb{R} . Formally, $W: X \rightarrow \mathbb{R}$. As use of the term “welfare,” which, again, derives etymologically from the notion of “faring well,” suggests, one possible, and indeed the usual, interpretation is that W ’s assigning a greater \mathbb{R} value to world x than to world x' corresponds to x ’s being found to be in some sense “faring better than,” or simply “better than” world x' .

An individual i ’s *preference function*, U_i , is also a (simple or composite) mapping from X onto \mathbb{R} , again one possible, and indeed the usual, interpretation being that greater \mathbb{R} values correspond to some conception of “preferredness.” Formally, $U_i: X \rightarrow \mathbb{R}$.

As noted parenthetically in the previous two paragraphs, a social welfare function W or an individual preference function U_i can in principle be simple (“direct”) or composite (“indirect”). Here is what that means.

Where individuals assign some worlds higher or lower real-numeric valuations than other worlds in virtue of one or more variable *properties* of those worlds, they effectively prefer worlds “indirectly.” They prefer them in virtue of their separate attributes, in functionally composite fashion. The individual i ’s preference function in such case takes the form $U_i(f_1(x), \dots, f_m(x))$, where each f represents a feature of x in virtue of which it may score “more” or “less” highly in i ’s preference schedule. Where $m > 1$ in such case, i must of course commensurate – comparatively weight – and then aggregate f scores in order to arrive at a scalar \mathbb{R} value to assign x .

Likewise, when a social welfare function W maps possible worlds x onto \mathbb{R} in a manner responsive at least in part to individuals’ mappings of those worlds onto \mathbb{R} , it too is effectively composite. It is a function, at least in part, of at least one other function – viz. one or more individual’s or individuals’ preference function(s). It thus takes the form, in whole or in part, $W(U_1(x), \dots, U_m(x))$.¹⁹ Where $m > 1$ in such case, theorists must of course commensurate – comparatively weight – and then aggregate U_i ’s in order to arrive at a scalar \mathbb{R} value to assign x .

¹⁹ Combining the aforementioned individual preference and social welfare cases, one should say technically, using the standard function-compositional notation, that social welfare $SW = W \circ U_i \circ f_j$ for all i and j . Because the present analysis will not hinge on the compositions of individuals’ first-order preference functions, however, we shall take interest only in functional compositions of the form $W \circ U_i$, or, rendered more simply, $W(U_i(x))$.

Decisions about how so to weight individual preference functions effectively amount to decisions concerning distributive fairness, or justice. Such decisions, like those concerning what indeed even to count in W 's domain of arguments at all, are decisions concerning what makes a world better not simply per somebody's individual preferences, but better full stop: Such decisions are considered, if the theorist is normatively clear about what s/he is doing, to be the "right" decisions per the theorist's principled, standard-governed estimation. There will be more to say of this immediately below in Section 3.

As noted in passing through the previous several paragraphs, the internal structure of either a social welfare function W or an individual preference function U_i can in principle be not only "direct" (simple) or "indirect" (composite). It can also – and quite distinctly – be multivariate, mapping vector values (either directly or indirectly) onto \mathbb{R} , or univariate, mapping scalar values (either directly or indirectly) onto \mathbb{R} . The multivariate cases simply are those in which $m > 1$ per the foregoing paragraphs: Multiple f s in the U_i case, or multiple U_i s – possibly with other arguments as well – in the W case.

In such multivariate cases, the mappings that are the preference or social welfare functions in question amount to mappings of pairs of vectors (n -tuples) and worlds, on the one hand, onto single numeric values R along \mathbb{R} on the other. That is to say, they are of the form $((f_1, f_2 \dots f_n), x), R$. Equivalently, one can describe these as mappings of vectors (n -tuples) onto pairs of worlds x and numeric values R , of the form $((f_1, f_2 \dots f_n), (x, R))$. These matters of commutativity are of course matters of formal convenience, which I shall exploit below.

It will not have escaped notice that individual preference functions and social welfare functions as just described are in a certain sense isomorphic. In effect, the social welfare theorist models social welfare determinations much as the price theorist models individual consumption decisions. In this sense the theorist assimilates "society," "the planner," herself, or some variant of the Harsanyian (1953, 1955) "impartial observer" to an individual choosing from a feasible set of consumption bundles.

Assimilations of this sort will not be intelligibly objectionable, of course, unless relative to some theoretical purpose that they might frustrate. Might some such purpose be frustrated here? I think possibly one: The standard model of an individual's consumption behavior is meant to be no more than descriptive. It affords formal expression to a positive fact – the fact of the individual's consumption decisions. It is thus perfectly in order in this case to read-off exchange rates among vector components as implied or implicit. Consumers need not decide beforehand, for example, that they shall trade off but one apple for three bananas. "By their fruits ye shall know them": We learn the price ratios from described behavior after the fact.

The normative social welfare theorist's purposes, I take it, are different from those of the positive price theorist. They are not merely descriptive, but globally prescriptive in character. The theorist purports to guide action on the part of all possible functionaries before the fact, rather than modeling actions taken by particular consumers after the fact. In such case the SWF's counterpart to a preference function's implied price ratio between things like apples and bananas – namely, the weighting rule applied to those individual preference functions that figure into the SWF's argument domain – must be specified in

advance. There is no other way to tell functionaries what to do – what precisely to maximize and how.

This is all of it to say that the temporal “direction,” so to speak, between aggregation and action is reversed in the prescriptive case – as was of course only to be expected in view of that prefix – “pre-” – occurring in “to prescribe.” For specific action determinately to be prescribed *ex ante*, the SWF’s aggregation rule must be rendered explicit *ex ante*, rather than read off as implicit *ex post*. If we do no more than read it off as implicit *ex post*, we are effectively doing no more than describing how the society whose decisions we model effectively treats its members.²⁰ We say nothing as yet in respect of how we believe that it *ought* to treat its members.

Obscurity about this “directional” fact in the standard social welfare model, it is tempting to conjecture, tends to obscure what will now emerge in Section 3, hence the need of that minimal fine-tuning I shall make to the standard model in Section 4, then the results that the fine-tuned model straightforwardly enables us to derive in Section 5. There will be more to say of this in Sections 6 and 7.

3. A PUZZLE RAISED BY THE STANDARD MODEL

The model as just elaborated raises an interesting normative question: How is the SWF’s aggregation rule, which functions as a normative principle of distribution for the society whose welfare is both measured and to be optimized per the SWF in question, to be determined? How, for that matter, is the domain of arguments that the multivariate SWF takes, which also bears distributive consequences, for its part determined?²¹ And how, finally, might the considerations by reference to which one addresses those questions be given explicit expression in the welfare theorist’s would-be “complete” SWF?

As to the first two questions, it might be tempting, particularly to those who have not happened to notice the normatively critical distinction between preferences on the one hand, and normative standards that are incumbent upon preferences on the other hand,²² to suppose that the argument domain and aggregation rule might *themselves* be selected by reference to preferences. “Take a poll,” you might say; “put it to a vote.”

A moment’s reflection, however, suffices to satisfy ourselves that recourse to preferences, in selecting an argument domain and aggregation rule that effectively distribute the satisfactions of preferences, simply defers the question that it purports to answer. For we shall of course now have to choose from among and aggregate the preferences expressed in respect of the normatively appropriate argument domain and aggregation rule themselves.

What “second-order” choice of argument domain and aggregation rule are we then to employ in counting and aggregating the preferences by reference to which we would select the first, “first-order” argument domain and aggregation rule, whose

²⁰ Once again, “by their fruits ye shall know them”: For example, “These people value the blondes among them more highly than the brunettes, as evidenced by the fact that blondes’ preference functions are weighted more heavily than those of brunettes in determining that normative measure they call ‘social welfare.’”

²¹ See the last paragraph but two in the previous Section. Also *supra*, notes 5, 11.

²² Please see again note 4 in this connection.

selections set us off on this course in the first place? Shall we take another vote? If so, won't that just replicate the original problem? Clearly we're off on a regress. And as was once sagely observed, the trouble with a regress is not that it has no end; it is that it has no beginning. It is ungrounded, we don't understand *any* of its terms. (See Wittgenstein (1984: 693).) Hence we are faced with a clear case of normative indeterminacy – thus far at least.

Now it is natural enough in such cases to propose that the selection of an argument domain and aggregation rule, fraught with distributive import as they are, be conducted by vote, with the votes in turn counted and aggregated pro rata. The suggestion seems only fair. But the crucial words here, I think, are those words “natural” and “fair.” We are drawn to such focal points, one suspects, in virtue of a natural resonance they have with a standard of fairness that we find antecedently compelling, one that operates independently of – indeed, as “compelling” suggests, one that “compels” or at any rate guides and in that sense determines – our decisive or operative preferences themselves.

This is surely the case with most, if not all SWFs. Certainly it is the case in respect of all normatively constructed, selected, or employed SWFs. Social welfare theorists, when they are careful and normatively prompted, choose argument domains and aggregation rules in keeping with conceptions of distributive propriety.

Utilitarians, for example, are commonly heard to defend utilitarian SWFs by reference to the putative egalitarian fairness imminent in their counting each individual's utility function “only once.” (See, e.g., Smart (1973).) It is far from clear, of course, why counting *utility* functions equally is any more normatively compelling than would be counting thigh-length or forehead-height-functions equally, at least if the former are genetically accidental as are the latter. (See Hockett (2009a).) But that's neither here nor there for the moment.

The critical point for present purposes is that the SWF's argument domain and, especially, its aggregation rule register an antecedently operative normative distributional standard which, in determining how preferences are aggregated in reaching a measure of social welfare, operates both as incumbent upon individual preferences and as constitutive of the theorist's working conception of “welfare” itself. Theorists work with a guiding idea of (a) who is entitled, as a matter of normative priors, to be counted, in virtue of what normatively interesting attributes, and (b) how one counted party's interests are to be comparatively weighed as against others', and why.

Judgments of fairness accordingly determine these theorists' measures, and in this sense determine their very conceptions, of social welfare. The values that determine their choices of argument domain and aggregation rule, that is to say, are constitutive of – they are “internal to” and definitive of – their views of what counts as welfare itself. Welfare is fairly aggregated satisfaction.²³

²³ See also *supra*, notes 4, 5, 11, 21, and associated text. In this connection, locutions like “fairness versus welfare,” which figures into the titles of a surprising number of publications (see Kaplow & Shavell (1999, 2000, 2002, 2003)), looks transparently to register a category confusion, a confusion that one suspects stems from the same source as that noted above in note 4. An immediate hint that something has gone wrong is the fact that “fairness” names a pattern of distribution, while “welfare” on the dominant preference-satisfaction understanding names an object of distribution. “Fairness versus welfare” accordingly rings a bit as would “velocity versus projectile.” But recognition that a conception of fairness is internal to a conception of welfare itself tends to sharpen the sense in which the locution can be seen to

The trouble, it turns out, is that the standard SWF model as presently rendered offers no means of explicitly registering this fact. This gap in the standard model accounts, in turn, not only for much confused talk about welfare to fairness “tradeoffs,”²⁴ but also for obscurity, to date, about what the results next derived in Section V show. It also amounts to a sense in which the standard model as presently configured falls short, for the time being, of those normative-theoretic completeness ambitions that originally prompted its introduction and continue to prompt its use. (See again Bergson (1938), Samuelson (1947). Evidence of this motivation’s continued force is found, e.g., in Kaplow & Shavell (2001, 2002).

4. THE MODEL SHARPENED

I turn now to fine-tuning the standard model with a view to enabling it explicitly to register the preference-incumbency just considered – that of those normative standards which guide the theorist in determining any SWF’s argument domain and aggregation rule, what one might call its “distributive structure.” Once these minimal adjustments to the standard model are in place, the formal results derived in Section 5 are straightforwardly derived, and their import transparently manifest.

DEFINITION 1: A *preference-regarding* social welfare function, W_{pr} , is a composite social welfare function of the form $W(\bullet, \dots, U_1(x), \dots, U_m(x))$ the immediate domain of (non-“dummy” arguments of) which includes, but need not be limited to, m individuals’ preference functions U_i , and that is monotonically increasing in the value of each of its individual U_i arguments. It is of course possible in such case that there exist $x, x' \in X$ such that $U_i(x) = U_i(x')$ for all i and yet $W_s(x) \neq W_s(x')$. It is also the case here that n might exceed m . In such case W_{pr} could be characterized as “elitist.” (The preferences of some members of N “would not count.”) If m were precisely 1, W_{pr} could be characterized as “despotic,” “monocratic,” or, approximately though not fully following Arrow (1951), “dictatorial.”

COMMENT 1: “Preference-regard” corresponds to what some since Samuelson (1947) have called “individualism,” and what some since Sen (1979) have called “welfarism.” Because regard for persons’ preferences is what is in fact distinctive here, while “individualism” is quite imprecise and “welfarism” encourages conflation of the mere “is” of positive individual demand with the traditionally more ethically resonant “ought” of normative social or ethical command, I believe we’ll do well to regard “preference-regard” as the preferable term.

be a category error. For fairness now can be seen to constitute what one might call the “form” of welfare, while preference-satisfaction constitutes the “substance” thereof. Welfare, in short, just is this: Fair satisfaction. It is appropriately distributed satisfaction. If this is correct, then to say “fairness versus welfare” in the domain of law and policy design is a counterpart to what it would be to say “three-leggedness versus wood” in the domain of furniture design.

²⁴ See supra notes 4, 5, 11, and 23. “Fairness versus welfare” is simply the *reductio ad absurdum* of this impossible line of thought.

DEFINITION 2: A *strictly preference-regarding* social welfare function, W_{spr} , is a preference-regarding social welfare function of the form $W(U_1(x), \dots, U_m(x))$ the domain of (non-“dummy” arguments of) which is restricted to individuals’ preference functions U_i alone. Hence it is a social welfare function for which there exist no $x, x' \in X$ such that $U_i(x) = U_i(x')$ for all i and yet $W(x) \neq W(x')$. Thus the following hold true: $\forall i \in N; x, x' \in X: U_i(x) > U_i(x') \supset W_{spr}(x) > W_{spr}(x')$. And: $\forall i \in N; x, x' \in X: U_i(x) < U_i(x') \supset W_{spr}(x) < W_{spr}(x')$.

COMMENT 2: “Strict” preference-regard corresponds to what most others (than those mentioned in Comment 1) since Sen (1979) have called “welfarism.” Because it is possible for an SWF to respond positively either to preferences alone, or to such preferences along with one or more other factors that a theorist might find to render a world better when present therein, we require an adjective that can distinguish between the two cases. “Strict” seems a fitting such adjective given its analogous employment in other formal contexts.

DEFINITION 3: An *explicitly standard-regarding*, or equivalently, *normatively complete* social welfare function, W_{sr} , is a social welfare function that gives complete and explicit expression to the normative standards or value judgments that prompt its design and construction – in particular, those that determine its argument domain and its aggregation rule. A *non-explicitly standard-regarding*, or equivalently, *normatively incomplete*, or merely *descriptive* social welfare function is simply a social welfare function that is not complete and explicit in this sense. Note that an explicitly standard-regarding social welfare function can also be preference-regarding. But as reflection on Definition 2 will immediately suggest, it is difficult to see how a standard-regarding social welfare function could be strictly preference-regarding.²⁵

COMMENT 3: The distinction between an explicitly standard-regarding, normatively complete SWF on the one hand, and a non-explicitly standard-regarding,

²⁵ A perspicuous symbolism would render this incompatibility immediately apparent. The simplest way would be for it to take the form displayed in Definition 1, with at least one “●.” This would register the normative standard or standards to which the SWF had explicit regard. It is essentially the want of transparency of this sort that accounts for the failure of Section 5’s results thus far to have been noticed. An adequate formalism would wear what the proofs reveal on its face.

“But doesn’t the prefix, ‘Max,’ typically found before SWFs, itself suffice fully to register the normative character of the theorist’s SWF?,” someone might ask. “After all, it is an order.” The answer is no. For people can issue demands in the form of commands, without the commands’ thereby acquiring normative status – i.e., without their becoming incumbent upon the preferences of those to whom the demands are issued. Such commands as those issued by slaveholders, for example, or by brigands demanding you “stand and deliver are attempted conscriptions rather than normative prescriptions. (See Hockett (2009c) for fuller discussion of this matter.)

What distinguishes normative social prescription from mere attempted conscription is that the former can be justified by explicit reference to generally applicable normative principles – precisely those items that guide the choice of an SWF’s argument domain and aggregation rule. The minimal counterpart of this form of explicit reference to normative standards in an SWF is the “●.” The dot is a stand-in for those normative standards to which the theorist might explicitly appeal in reply to an individual who, in response to the theorist’s order to maximize the SWF, rejoins that like Bartleby the Scrivener, she would “prefer not to.” (See Melville (1856).)

normatively incomplete, or merely descriptive SWF on the other, bears crucially upon the *uses* to which social welfare functions might be put.²⁶

As noted in the final paragraphs to Section 1, a social welfare function ordinarily employed (incompletely) to formulate normative theoretical judgments might be used instead, advertently or inadvertently, simply to formulate a descriptive report. Such a report might be, say, a description of resource allocations over individuals with which corresponding levels or amounts of some positive (not necessarily normative) aggregate output (GDP, for example, or “hedonic pleasure” as measured in endorphins) are empirically associated. Such outputs would be, strictly speaking, matters in respect of which the theorist effectively reserved normative judgment, not committing herself to any particular normative theoretical attitude.

Another circumstance, likewise envisaged at the end of Section 1, in which theoretic incompleteness might do little or no mischief would be that in which one wishes to frame an instruction meant to guide functionaries in doing their jobs. In such cases the SWF would not be meant to give complete or transparent expression to the full complement of a theorist’s nameable value commitments with a view to tracing all logical consequences thereof. It would be meant simply to guide actions on the part of others who take orders from above, hence employed to formulate an instruction rather than a complete value judgment.

Social welfare functions often seem to figure in these less normative-theoretic ways in the social choice literature, as well as in governmental budget offices, suggesting that it would be helpful to have to hand some formal means by which SWFs might be made transparently to indicate whether they are giving full expression to complete and fully articulated normative value judgments on the one hand, or merely descriptive reports or instructions on the other hand, on particular occasions of use.

The want of a standard, straightforward, unequivocal formal indicator marking off social welfare functions in their normative theoretical as distinguished from their positive or pragmatic employments accounts, I conjecture, for frequent confluences of positive preference with normative principle, hence ultimately of “is” with “ought,” in much of the welfare economic and social choice literature.²⁷

The next concept is introduced in order to give expression, in the apparatus of SWFs, to one particular evaluative judgment to which anyone who normatively inquires, judges, or prescribes with full articulation appears to commit herself, at a bare minimum, in any act of normative inquiry, judgment or prescription. It is an implicit judgment that comes to light – a “second-order preference” that is “revealed,” so to speak – the moment we take care expressly to distinguish normative theoretic from descriptive or pragmatic inquiry or activity.

DEFINITION 4: A social welfare function W is *incumbent*, or equivalently, *decisive*, insofar as a theorist actually conducts or communicates normative-theoretic assessments of possible worlds either with it, or, derivatively, pursuant to the principles

²⁶ The idea that the uses of expressions bear upon their meanings is of course not unfamiliar. See, e.g., Wittgenstein (1953). The link is of course also manifest in regular treatment of the word “usage” as a rough synonym for “meaning” itself.

²⁷ There are other sources of this conflation as well, one suspects. A remark of Pareto’s ((1909): 318) is apposite: “In general, men are prone to give their private demands the form of general demands.”

to which it gives formal expression. Note that W 's being incumbent or not is a feature of the world, hence is partly constitutive of any x – i.e., any state of the world as given complete normative description per the second paragraph of Section 2.

COMMENT 4: The fuller significance of incumbency emerges in Axiom 1, below. See also Sections 1 and 3 above, in which this concept and its prompting considerations are first broached. I'll address anticipated objections to incumbency, and to the claim made immediately below concerning normatively complete social welfare functions' valuing their own incumbency, further below in Section 6.

I now request the reader's assistance in assessing the self-evidence or otherwise of the following axiom, which I for my part see no way of sincerely denying:

AXIOM 1—Minimal Normative Incumbency: All else being equal, an explicitly standard-regarding social welfare function W_{sr} deems a world better, i.e., assigns to it a higher value along \mathbb{R} , if W_{sr} is itself incumbent in that world. Somewhat more formally, $\forall x, x' \in X$, if x is a world in which W_{sr} is incumbent and x' is an otherwise identical world in which W_{sr} is not incumbent, then $W_{sr}(x) > W_{sr}(x')$.

REMARK 1: Here is the sense in which the Axiom seems to me undeniable: If I engage in normative inquiry, then I judge normative inquiry to be a good thing to do. The judgment appears to be latent in my deliberate engagement in this form of inquiry itself, rather as first-order preferences are said to be “revealed” by trading or consumption behavior. (Assuming that I am not merely “acting out” an addiction of some sort.)

But if I judge normative inquiry a good thing, then, since the way in which I give expression to my judgments of good and ill, better and worse, and so forth in the idiom of SWFs is by assigning numeric values to “worlds” in response to their featuring such good or bad things, I must in consequence judge the world better for my engaging in normative inquiry in it. At least that is so inasmuch as I am indeed *in* the world. A good thing happens in the world by dint of my doing the good thing that is engaging in normative investigation in that world. Hence the world is a better world, *ceteris paribus*, for my doing so. And my normatively complete, *explicitly* standard-regarding SWF registers this fact, like any other world-bettering fact, by assigning the world a higher numerical value in virtue of the fact's obtaining.

Note also that the incumbency registered in Axiom 1 as being valued by the standard-regarding SWF is, as the name of the Axiom suggests, indeed *minimal*. It involves no more than the function's registering the value of the theorist's engagement in her normative analysis. In the actual world that we inhabit, many if not most theorists – particularly those in the academy – appear effectively to value a significantly *stronger* form of incumbency: Theorists generally endeavor, through reasoned argument, *to convince others* to “see things their way,” normatively speaking. That is to say, they seek to convince others to favor the same policies, for the same normative reasons – on the basis of the same normative standards or principles – as do they.²⁸ Their actions

²⁸ See Hockett (2009b) for further development of this observation.

accordingly bespeak their valuing the incumbency of their particular principles upon the analyses of others as well as upon themselves. That's more than minimal incumbency.

Here, though, again, we restrict ourselves to the minimal, least contestable claim in respect of the standard-regarding SWF's valuation of its own incumbency. That is, per Axiom 1, its simply registering the good, in the theorist's judgment, of the theorist's seeking to ascertain and employ the normatively most appropriate social welfare function itself. The form of recursion that this imports into the SWF is nothing other than the normatively complete SWF's full reflection of the theorist's activity itself, which is *itself* reflective inasmuch as the theorist does what she does *in* the world to whose states her SWF assigns values.

The next concept is implicit in any multivariate function. But it is implicated with particular salience by the distinction between normatively complete and incomplete SWFs, as we'll presently see:

DEFINITION 7: Call a subset (\circ) or (\bullet) of a function's full set of arguments (\bullet, \circ) a "subdomain" of the function. Where (\circ) is the set of individual *preference-functional* components of a preference-regarding social welfare function W_{pr} 's *full set* of arguments, call it W_{pr} 's "preference subdomain." Finally, call W_{pr} as applied to its preference subdomain alone the "preference-regarding subfunction" $W_{pr/\circ}$ of W_{pr} .²⁹ We shall then say that a preference-regarding social welfare function W_{pr} *preference-wisely replicates* (symbol " $\approx \circ$ ") another preference-regarding social welfare function W_{pr} in world x when and only when W_{pr} and W_{pr} share a common preference-regarding subfunction with respect to x . W_{pr} preference-wisely replicates W_{pr} , that is to say, when and only when W_{pr} and W_{pr} share a common preference subdomain with which is associated the same world and numeric value pair (x, R) . Somewhat more formally, $W_{pr}(x) \approx \circ W_{pr}(x) \equiv .$
 $W_{pr/\circ}(x) = W_{pr'/\circ}(x)$.³⁰

COMMENT 7: The concept of one SWF's preference-wisely replicating another is introduced to afford ready means of registering an obvious but seemingly oft-ignored prospect: namely, that (a) two such functions' argument domains might be coextensive with respect to the individual preference schedules that are among (or that exhaust) their elements, and (b) the persons whose individual preference schedules are thus shared by the functional domains might receive identical allocations of resources in two worlds x and x' all while (c) the two functions nevertheless do not assign x and x' the same values along \mathbb{R} precisely because some feature of worlds to which some *other* element of a function's full argument domain might be sensitive *varies* between worlds. Precisely this

²⁹ The computational term "subfunction" seems intuitively apt here in light of its resonance with the familiar "subdomain." Mathematical purists will recognize "subfunction's" synonymy with the more traditional set- and function-theoretic "restriction." If D is the domain of $W(\bullet, \circ)$, D_{\bullet} the subset of D 's " \bullet " factors, and D_{\circ} the subset of " \circ " factors, then since W is $W: D \rightarrow \mathbb{R}$, the subfunction W_{\circ} is simply the restriction of W to D_{\circ} , i.e., $W|D_{\circ}$, or $W_{\circ}: D_{\circ} \rightarrow \mathbb{R}$.

³⁰ By way of paradigmatic example, if $W(\bullet, \circ)$ is a weighted sum of its arguments, then $W(\circ)$ is the subfunction resulting when we "drop" – weight at zero – all arguments (\bullet). Where (\circ) is the set of individual preference functional components of a *preference-regarding* social welfare function W_{pr} 's full set of arguments, $W(\circ)$ is what I am calling " $W_{pr/\circ}$."

prospect is raised by the distinction between normatively complete and incomplete SWFs. Its not having been generally remarked in the literature is a consequence, I believe, of a tendency found in much of the literature to elide the distinction between complete normative prescription and mere positive description, a tendency reinforced by, if not rooted in, the tendency to collapse normative standard or principle into positive preference.

The sense in which the just-described prospect to which preference-wise replication gives expression emerges from the distinction between complete normative prescription and mere positive description is this: As noted above in connection with Definition 3, SWFs *might*, but *need not*, be employed to formulate complete and explicit normative judgments or prescriptions. That is to say that they might, but need not, be rendered normatively incumbent by registering or figuring into actual normative judgments or prescriptions concerning described possible worlds. For again, as noted above in Comment 3, a function can be employed to describe or locally prescribe allocations of individually preferred benefits, dispreferred burdens, and so forth over individuals, and can correlate “levels” or “amounts” of some conceived social aggregate that varies in response to varying such allocations, without thereby committing the theorist to any complete *normative judgment or global prescription* concerning those allocations or associated aggregates.

Insofar as the theorist remains indifferent or agnostic about whether it be generally “good” or “ill” that others’ preferences be satisfied or no, for example, she need not value or, therefore, express valuation of the normative incumbency of the SWF with which she tracks and registers such patterns of preference-satisfaction or -dissatisfaction. Her SWF will in such case remain, per the terms of Definition 3, either normatively incomplete or merely descriptive.³¹ By contrast, the moment the theorist (a) adopts a normative stance or formulates a normative judgment as to the goodness or otherwise, generally speaking, of the pattern of satisfactions and dissatisfactions registered by her SWF, and (b) elects to employ the latter to give what theorists call “complete” expression to her value judgment, the SWF will come to value its own incumbency in addition to valuing the preference-satisfactions experienced by those whose preference schedules figure into its argument domain. (See, again, Axiom 1.) It will accordingly preference-wise replicate (and of course be replicated by) its merely descriptive counterpart, even while adding at least one non-preference-grounded argument – viz., that associated with its own incumbency – to those features of possible worlds in virtue of which it assigns variable numerical values to worlds.

I now request the reader’s assistance once more, in reflecting upon and assessing the self-evidence or otherwise of the following axiom, which, like Axiom 1, I can see no way of sincerely denying:

³¹ In view of the normative coloration of such words as “welfare,” it would of course be preferable in such case to call her SWF something like an “aggregate preference-satisfaction function” (an “APSF”?) rather than a “social *welfare* function.” But I’ll leave that terminological gripe to one side for the present, contenting myself with the “explicitly standard-regarding,” “normatively complete” and “globally prescriptive” versus “non-explicitly standard-regarding,” “normatively incomplete” and “descriptive” SWF terminology introduced above. More on these matters in Hockett (2009a, b, c).

AXIOM 2—Replicability: *For any normatively complete preference-regarding social welfare function $W_{sr/pr}$, and any possible world x , there is a strictly preference-regarding social welfare function W_{spr} which preference-wise replicates $W_{sr/pr}$ in x .*

REMARK 2: Simply take $W_{sr/pr}$'s preference-regarding subfunction $W_{pr/\circ}$, by way of immediate example.³²

5. TWO IMPOSSIBILITY RESULTS

I turn now to the results mentioned above in Section 1, which are readily derived in the augmented model. The first result concerns strict preference-regard (a.k.a. strict “welfarism”). The second, derived in the next Section, concerns Paretianism.

5.1 The Impossibility of Strictly Preference-Regarding Social Evaluation

LEMMA: *If all persons i whose preference functions jointly constitute the preference-regarding subfunction $W_{pr/\circ}$ of a preference-regarding social welfare function W_{pr} assign the same value to two worlds x and x' , then so does the subfunction.* Formally, $\forall i: U_i(x) = U_i(x') \equiv W_{pr/\circ}(x) = W_{pr/\circ}(x')$.

PROOF: This is immediate from Definition 7 of a preference-regarding subfunction of a social welfare function and Definition 2 of a strictly preference-regarding social welfare function. ■

We're now positioned to derive strict preference-regard's formal inconsistency with normatively complete social evaluation.

PROPOSITION 1: *No normatively complete social welfare function is strictly preference-regarding.*³³

PROOF: Assume a world x in which the standard-regarding social welfare function W_{sr} is incumbent. Assume, per Axiom 2 (Replicability), another world x' in which W_{sr} is

³² It is worth noting in passing that replicability is straightforwardly connected to a familiar entailment of the “fundamental theorems” of welfare economics – viz., that for any efficient allocation x , there exists some SWF whose domain comprises a weighted sum of concave, continuous and monotonically increasing preference functions that x maximizes. (See, e.g., Varian (1992): 333-35.) Such SWFs would be, in an infinite number of cases, quite arbitrary from the point of view of any known normative standard or principle that might actually prompt and guide normative assessment. (See Hockett (2009b, c) for more on this point.) Insofar as it is normatively or prescriptively arbitrary, in turn, such an SWF will be merely positive or descriptive; it will do no more than mathematically characterize an allocation, irrespective of any recognized or indeed recognizable sense in which that allocation is socially “rightful” or “good.” It will at best serve as a post hoc “rationalization,” or rubber stamp, arbitrarily deeming normatively “optimal” any randomly selected positive allocation. That's what one might call, alluding to the *Essay on Man* (“what ever is, is right”), “Pope with a vengeance.” (See Pope (1732).)

³³ Note that, by Definition 3, if this Proposition holds true, then a strictly preference-regarding SWF can at most be descriptive.

preference-wise replicated but not incumbent. We then know the following to hold true: By Observation 1, $W_{sr}(x) > W_{sr}(x')$. Hence $W_{sr}(x) \neq W_{sr}(x')$. But by Replicability and the Lemma, $U_i(x) = U_i(x')$ for all i . Hence by Definition 2, W_{sr} is not strictly preference-regarding. ■

5.2 The Impossibility of Preferential Paretian Social Evaluation

Any form of Paretianism that interprets “better” and “worse” in terms of preference-satisfaction as described in Section 1 is readily shown to inherit strict preference-regard’s inconsistency with normatively complete social evaluation, as we now demonstrate.

DEFINITION 9: A social welfare function W is “Paretian” if and only if it satisfies one or more of the Pareto indifference (PI), weak Pareto (WP) or strong Pareto (SP) conditions. Somewhat more formally: Either PI holds of W , i.e. $\forall i \in N; x, x' \in X: U_i(x) = U_i(x') \Rightarrow W_p(x) = W_p(x')$; or WP holds of W , i.e., $\forall i \in N; x, x' \in X: U_i(x) > U_i(x') \Rightarrow W_p(x) > W_p(x')$; or SP holds of W , i.e., $\forall i \in N; x, x' \in X: U_i(x) \geq U_i(x') \ \& \ \exists i \in N; x, x' \in X: U_i(x) > U_i(x') \Rightarrow W_p(x) > W_p(x')$.

COMMENT 9: So-called “full Pareto” (FP), defined in the literature as the conjunction of PI and SP, is of course trivially embraced by Definition 9. Definition 10 and Proposition 2, below, will afford warrant to our lumping all versions of Pareto together as here.

SUPPOSITION—Normal Good Profile: There might exist at least one indefinitely divisible resource, or profile of such resources, r , which, were every individual i to possess an increment $\varepsilon > 0$ more of it, would leave each individual i ’s preferences more satisfied. Somewhat more formally, taking $r_i/r(x)$ for each person i ’s allotment of r in world x , and $r_i/r(x')$ for i ’s allotment of r in world x' , we have $\forall i \in N: r_i/r(x) - r_i/r(x') = \varepsilon \Rightarrow \forall i \in N: U_i(x) > U_i(x')$.³⁴

³⁴ One could word this somewhat more technically by saying that each U_i is responsive only to i ’s own allotment, not to others’ allotments, of the resource r and is at least weakly continuous and monotonically increasing in that allotment. See, e.g., Suzumura (2001). By “weakly” continuous here I mean simply that if $r_i/r(x) > r_i/r(x')$, then there is a neighborhood $V(r_i/r(x'))$ of $r_i/r(x')$ all of whose members $r_i/r(x)$ exceeds. See, e.g., Campbell & Kelly (2002).

I proceed for the most part in what follows as if the Supposition is to be understood in terms of the single resource r , but do so only for the sake of simplicity in exposition. The same idea could, with a bit more complexity, be put forward in terms of each individual i ’s preferring x' to x by virtue of the former’s yielding an added increment of *some* resource or other feature of the world represented by “ r ,” the specific r ’s potentially differing from individual to individual, each being anything at all by virtue of which somebody might prefer a possible world. (The r ’s might, that is to say, be indexed by individual, such that r for person $i1$ is $r1$, for person $i2$ is $r2$, etc. In such case “ r ” will designate, not a single resource, but a profile of resources corresponding to a set of individuals.) Thanks to Matt Adler for emphasizing this prospect in conversation.

REMARK 3: This supposition serves as a counterpart to the “Assumptions” hypothesized in several proofs cognate to that offered below, including, e.g., Suzumura (2001), and Kaplow & Shavell (2001), which restrict themselves to a single good. A profile of goods, such that different goods play the same role for different persons, will do just as well. I trust that the reason for the name is clear: The resource (or profile) here specified affects preference-satisfaction precisely as do the normal goods familiar to price theory.

It should be noted that the Supposition serves indeed as a conditional supposition, rather than as a committal assumption or axiom. It is set forth in order both to set the stage for and to make plain the full significance of a definition, namely that of a “preferentially Paretian” social welfare function, which is defined, next, as a social welfare function that behaves in particular ways when and if ever the Supposition might hold, but need not so behave when the Supposition does not hold.

DEFINITION 10: A social welfare function W is *preferentially Paretian* if and only if for each individual i it is at least weakly continuous and monotonically increasing in $r_i / r(x)$ when ever the Normal Good Profile Supposition holds. (That is to say, W is itself weakly continuous and monotonically increasing in each individual’s allotment of the resource r .)

COMMENT 10: Specifying a few implications of Definition 10 in prose might facilitate more immediate intuitive grasp of the next proposition and proof as well as of their import. Here then is the intended interpretation:

A preferentially PI social welfare function allows no value apart from preference-satisfactions – including, for example, incumbency per Definition 6 above, which any standard-regarding, hence normatively complete or prescriptive social welfare function per Axiom 1 will at a bare minimum count for something – to count for anything in determining an optimum. So a preferentially PI SWF will represent the world as unambiguously improving or worsening precisely insofar as it gains or loses – even if but infinitesimally – in the quantum of r that it contains, quite irrespective of any accompanying loss or gain in respect of the incumbency of some normative standard or principle such as might find expression in the social welfare function itself. If everybody likes ice cream, for example (and surely everybody likes ice cream), the addition of (even an arbitrarily small fraction of) one scoop of ice cream to the world will be deemed by the preferentially PI social welfare function to render the world infinitely better than, for example, has the theorist’s having elected to ascertain and employ the most normatively appropriate social welfare function in her normative analyses of completely described possible worlds, even if it is this social welfare function itself that she has found to be most normatively appropriate in this sense.

Preferentially WP or SP social welfare functions, for their parts, will not “trade off” so much as an infinitesimally small increment of gain in r , whether the traded increment of forgone gain is itself divided over everyone (WP) or is a traded-off increment of gain that would be realized by but one person (SP), for any gain, no matter how large, in respect of some other value – again including incumbency per Definition 6 above, which any standard-regarding, hence normatively complete or prescriptive social welfare function at a bare minimum will count for something. A scoop (or arbitrarily

small fraction of that scoop) of ice cream's addition to the world, for example, whether it be divided over everyone or simply added to Dick Cheney's next Halliburton dividend, will be deemed by the social welfare function to constitute a greater improvement to the world than has the theorist's having elected to look for the normatively most appropriate social welfare function and come to employ the result – this social welfare function itself – in normatively evaluating completely described possible worlds.

PROPOSITION 2: *No normatively complete social welfare function is preferentially Paretian.*

PROOF: There are three cases to consider:

For PI: We know from Sen (1986) that where that portion of the domain of what I have been calling a preference-regarding social welfare function which comprises individual preferences is restricted to a unique profile of individual preference functions, weak transitivity and Pareto indifference jointly entail what I have been calling strict preference-regard. (See Fleurbaey, Tungodden & Chang (2003) on this point.) Here we are indeed considering but a single such profile (recall, that of the members of N) and at least weakly transitive social welfare orderings (both per our opening remarks in Section 2 and per Definition 10), so our preferentially PI social welfare function is strictly preference-regarding. But by Proposition 1, no standard-regarding, hence normatively complete or prescriptive social welfare function is strictly preference-regarding. So the preferentially PI social welfare function cannot be standard-regarding, normatively complete or prescriptive.³⁵

For WP: Modify the proofs offered in Suzumura (2001), and Kaplow & Shavell (2001), with benefit of the foregoing observations and additional observations made by others. Do so, then, (a) partly in keeping with the sharpened, normatively more transparent model laid out in Section 4 and employed in the present analysis, and (b) partly in light of observations made by Campbell & Kelly (2002) concerning the needlessly strong assumptions employed in some cognate proofs, notably that of Kaplow & Shavell (2001): By Definition 2, we know that if a social welfare function W is not strictly preference-regarding, then $\exists x, x' \in X: \forall i \in N: U_i(x) = U_i(x')$ and yet $W(x) \neq W(x')$. Now suppose a particular such function and two worlds x and x' for which (a) $\forall i \in N: U_i(x) = U_i(x')$ and (b) $W(x) > W(x')$. Next suppose a world x'' identical to x' in all respects save that each $r_i/r(x')$ per the Normal Good Supposition has been increased by ε

³⁵ Two additional observations are germane in the present connection: First, assuming independence of irrelevant alternatives and unrestricted domain per Arrow (1951) – which many social choice theorists and welfare economists, including Kaplow & Shavell themselves (2000, 2004), appear to find attractive – Pareto indifference proves equivalent to what I have been calling strict preference-regard even in respect of *all possible* profiles of individual satisfaction functions, rather than simply in respect of a unique such profile, meaning that we could dispense with the latter assumption. See again Sen (1986). Second, under a weak set of four assumptions, three of which – viz. (a) one indefinitely divisible resource, in which individual U_i 's and W are (b) continuous and (c) monotonically increasing – are largely incorporated into the analysis here, and one of which – compensability in increments of the divisible resource – is commonly invoked in the literature of fair allocations and presumably acceptable to such as Kaplow & Shavell as well, all three Pareto criteria have been proved to be jointly equivalent. See again Suzumura (2001). Were we to assume compensability, then, which one suspects most Paretians would happily do, the proof of Proposition 2 would now be complete.

> 0 , so that $\forall i \in N: U_i(x'') > U_i(x')$. Now observe that, granted (even weak, per the Supposition and Definition 10, not necessarily Kaplow & Shavell's full) continuity, as well as monotonicity, in $r_i/r(x')$, there is some ε sufficiently small that our W , which, recall, is assumed not to be strictly preference-regarding, will rank worlds thus: $W(x) > W(x'')$. In that case since, per (a) just above we have supposed $\forall i \in N: U_i(x) = U_i(x')$, it must also be the case that $\forall i \in N: U_i(x'') > U_i(x)$. Hence if weak Pareto is satisfied, $W(x'') > W(x)$. But that contradicts what we found supposing weak continuity and monotonicity per the Normal Good Supposition. Hence the preferentially WP social welfare function per Definition 10 relinquishes non-strict preference-regard. (It has to be *strictly* preference-regarding.) But by Proposition 1, with that it relinquishes its standard-regard, its normative completeness, and hence its capacity to give full expression to a prescription.³⁶

For SP: Strong Pareto trivially entails weak Pareto. What is true of the WP case here is accordingly true *a fortiori* of the SP case. ($SP \supset WP$; $WP \supset$ no prescription; $\vdash SP \supset$ no prescription.) Hence the latter too is incapable of prescribing.³⁷ ■

6. ANTICIPATED OBJECTIONS

Here I shall briefly address a few questions that I anticipate might be raised in connection with the fine-tuning of the standard model offered in Section 4, upon which the results in Section 5 rely. The principal objections that I anticipate are directed to the concepts of explicit standard-regard and normative incumbency.

It might be objected that there is no need of what I call explicit "standard-regard," because regard for the standard or standards that prompt the theorist's design of her SWF is already implicit in the argument domain and aggregation rule themselves. In effect echoing the "early" Wittgenstein, this objector objects that standard-regard is already "shown" in the SWF, and thus need not – or perhaps even in some obscure sense cannot – be "said." (The allusion is to Wittgenstein (1921): 4.121 *passim*.³⁸)

³⁶ Note that this portion of the proof goes through even when W is replaced by a weaker, nontransitive ordering relation falling short of functionality. See, e.g., Campbell & Kelly ((2002): 80-81).

³⁷ I'll note here that Chang (2000a) constructs a number of interesting *mixed* "liberal consequentialist," or equivalently "Paretian liberal" social welfare functions that *reject* continuity. These in effect give formal expression to the venerable mixed teleological/deontological systems of ethics pursuant to which maximization of one or more goods is enjoined, subject to a side constraint sounding in rights to fairness or some basic minimum. See, e.g., Fuller (1964) (jointly operative moralities of "aspiration" and "duty"), Calabresi (1970) (lexicographically ordered "just[ice] or fair[ness]" and cost-minimization as goals of accident law), Rawls (1971) ("right" ordered as prior to "good"), Dworkin (1977) (rights as "trumps") for canonical articulations. Chang's social welfare functions also of course escape the consequences of the proofs I here offer, particularly in such circumstances as those in which their non-Paretian components kick in. The reason is clear: They are not *strictly* preference-regarding or preferentially Paretian.

³⁸ I say "obscure" sense because the distinction in Wittgenstein's case is notoriously controversial. As Frank Ramsey once said, "if you can't say it you can't say it – and you can't whistle it either." The point is that, if there is anything there to be "shown," it can be articulated too. The *Tractatus*, incidentally, is notoriously and indeed deliberately self-contradictory on this point, concluding with the observation that it has itself been all along saying what can only be shown, and is accordingly nonsensical.

It might relatedly be objected that Axiom 1's "Minimal Normative Incumbency" is contestable as a matter of metaethics. Might there not, after all, be what Parfit (1984) used to call "self-effacing" moral principles? And in all events, is any incumbency that normative standards might indeed have upon preferences really so far reaching as to require not only that we act in conformity to such standards, but that we seek to ascertain them and act deliberately upon them as well?³⁹

I cannot help but feel some sympathy with these objections.⁴⁰ But I also cannot, on reflection, seem to find myself persuaded by them. The whole point of the SWF apparatus as employed in "pure" welfare theory, recall, is to give complete and explicit – not implicit – expression to the full array of value judgments that enter into the social welfare theorist's comparative evaluation of possible states of the world. The reason for this form of completeness, for its part, is not simply gratuitous; it isn't a matter of "obsessive-compulsive disorder" on the part of the theorist. It is to enable the theorist to be fully precise and explicit about what all of her normative commitments are.

We seek full transparency of this sort, in turn, in order more readily to trace all logical entailments of, including all logical incompatibilities possibly lurking among, the full complement of those commitments. We want, in other words, to know what our normative priors are getting us into, logical consequence-wise. We seek such knowledge both in order that we might better refine these priors in response to any unintended or unpalatable such logical entailments, and that we might better design law and policy in keeping with them insofar as we do ultimately find them robust and compelling.

Now it seems to me that all of this effectively implies that we want, as Bergson (1938) himself seems to have wanted, to be able to say *in* our SWFs all that we say by way of justifying those value-fraught choices we make in *designing* our SWFs. The aim of the SWF apparatus, that is to say, is not to enable people explicitly to "say" what in truth they need or must only implicitly "show," per the first objection countenanced above. It is rather to enable one to say *formally* what one *already* says *informally*. We wish to translate our vernacular – all of it that is relevant – into a more precise, logically tractable idiom. If the SWF does not permit this, it will both fail to discharge its appointed task, and, accordingly, fall short of effectively warranting such all-things-considered normative judgments as we aim to justify by reference to it. (The "all-things-considered" locution is from Davidson (1978).)

Now in the idiom of SWFs, the only way that there is to register value judgments is by assigning possible worlds higher rankings in response to their instantiating the values in question. And the only way to "show" this, in turn, is by opening a space for it in the SWF's argument domain. In simple formal languages like functional calculi, "to be is to be the value of a variable," quoth the oft-quotable Quine (1951: 15). It is just so with those "values" which are the normative standards that prompt the design of our SWFs.⁴¹ If we do not possess names for them in the language in which we conduct and express our analyses, they cease to exist in those possible worlds to which our language

³⁹ See *infra*, note 42, and associated text for additional material bearing upon this question.

⁴⁰ I could scarcely fail to be moved by objections sounding in Wittgenstein's Tractarian sensibility, nor could I fail to be moved by Parfitian objections raised as forcefully as Matt Adler has done in conversation and correspondence.

⁴¹ What is true of first-order formal languages with relational predicates and quantification quite generally per Quine carries over to the special cases of real-valued functions and, accordingly, SWFs.

assigns rankings. And that is so even though we inhabit, qua theorists, an infinite number of these possible worlds, and act pursuant to normative standards in conducting our analyses in them.

Without names for our standards in our SWFs, then, in short, we're vulnerable to the prospect of forgetting that these standards exist and are guiding our analyses. And if we justify our SWFs to others by reference to such standards, and our SWFs are themselves then to give full formal expression to our standards in turn per the original Bergsonian ambition, then the SWFs themselves must offer means of referring to those standards "by reference to which" we argue on behalf of them.

Now all of this admittedly means, of course, that the normatively perspicuous SWF will bear a recursive structure – a capacity for self-reference. For it values possible worlds partly in virtue of its own normative incumbency therein. (See generally Hockett (2008b). Also Hockett (2009b, c).) But recursion is hardly alien to high theory as practiced by economists. (See, e.g., Barberà & Jackson (2004); Bergstrom (1999); Farmer (1999); Koray (2000); Ljungqvist & Sargent (2004); Stokey, Lucas & Prescott (1989).)

Nor is the prospect of a recursive SWF surprising, for that matter, when we notice that normativity and normative theory themselves always import self-reference, inasmuch as they invite and involve self-assessment.⁴² Legal, policy, and social welfare theorists are inquiring after what "we" ought to be doing. And that carries with it a question wherein "we" consider "ourselves" – namely, the question of how "we" measure up to the standards we employ in addressing that "ought" question. In thinking about social welfare, then, we think among other things about ourselves – not only about how we are faring and how we might fare, but about what we should do with a view to that faring.

It seems to me that these same considerations suffice to address any Parfitian "self-effacement" objection that might be raised to my Minimal Incumbency axiom – which, recall, was indeed minimal. (See *supra*, Remark 1.) Inasmuch as the social welfare theorist is both (a) in the world and yet (b) prompted by normative considerations to seek means of evaluating the world, there seems no way for that theorist not, on pain of a species of incompleteness or opacity running contrary to the whole aforementioned *purpose* of the SWF apparatus, both: (a) to assign numeric values or rankings to possible worlds at least partly by reference to what she and/or other theorists do in those worlds; and (b) in particular, to deem or rank worlds better, *ceteris paribus*, in virtue of there being such people – social welfare theorists – in these worlds doing the sort of thing she is doing: And that is, again, seeking the right normative principles upon which to frame law and policy.

⁴² Indeed all forms of intentional action, individual as well as collective, appear to involve forms of self-reference inasmuch as an intention is always an intention to act pursuant to the intention itself. On this point see, e.g., Anscombe (1961); Harman (1976); Hockett (2009b, c); Velleman (2007 [1989]). It is also intriguing, I think, that at least one prominent theory of schizophrenia and the interpersonal ineptitude that tends to accompany it – the "Double Bind" account associated with Gregory Bateson – attributes both to a blindspot in respect of self-reference. (See Bateson et al. (1956).)

7. CONCLUSION

Assuming that the foregoing is sound, where does it leave us? Well, the first thing to note of course is that all of the most notorious impossibility results familiar to the welfare-theoretic literature, from Arrow (1951) through Koopmans (1960) and Diamond (1965) on down through Sen (1970), Hart (1975), Basu & Mitra (2003) and others, involve Pareto. But if the foregoing is sound and Pareto has been none but a veto of normative standards themselves rather than something that's normatively "attractive," then none of these theorems ought to have surprised us, let alone left us paralyzed or pessimistic in the way some of these results sometimes have done. At most they should simply confirm us in our determination to seek normatively intelligible determinants of true welfare – i.e., welfare as shaped by plausible SWF argument domains and aggregation rules.

I think that the analysis above also means, in this latter connection, that all normative legal, policy, and social welfare theorists are in effect justice – i.e., "fair welfare" – theorists.⁴³ The only real question in future, I think, will be whether we're good, careful justice theorists or less conscious and deliberate, default justice theorists. For we have seen here that every conception of social welfare incorporates, *eo ipso*, a conception of fairness as well. And that is so just as surely as every SWF or cognate ordering which serves as the theorist's measure of social welfare is internally constituted by an argument domain and an aggregation rule that jointly add up to a conception of distributive propriety.

It is difficult not to anticipate that the careful new "fair welfare" theorists, moreover, for reasons of both perspicuity and direct operational applicability, will either dispense with or supplement maximization formulae in formulating satisfactory social optimization programs. (They'll presumably retain maximization formulae in modeling individuals' optimization decisions.) For of course any multivariate maximization formula can be translated into an equivalent equalization formula. (See Hockett (2009a) for fuller discussion of the implications of this.⁴⁴) And working with formula of the latter

⁴³ See Hockett (2009a, b) for fuller elaboration of the "fair welfare" ideal. Strictly speaking, the name involves a redundancy, as we have just seen that true welfare is necessarily fair welfare. Welfare is fairly distributed satisfaction – satisfaction derived in conformity to those normative standards that determine the SWF's argument domain and aggregation rule. But "fair welfare" sounds nice, and serves to counteract the unfortunate and untenable claim advanced by some that there is some manner of "conflict" between fairness and welfare. See also *supra*, notes 4, 5, and 11 on this.

⁴⁴ Here, in a bit more detail, is the idea in a nutshell: The ethical considerations that go into formulating an aggregation rule take individuals to be equals in one or another respect adjudged to be ethically relevant. The classical utilitarian rule, for example, treats them as equals in respect of their utility functions and thus sums over the latter in unweighted form. (Because I can't see the ethical relevance of utility functions for which persons are not responsible, utilitarianism seems to me fetishist for precisely this reason; it's like summing over randomly distributed endorphin counts, or even thigh-lengths or forehead-heights.) Now any such summand implies a complement equalisand embracing all attributes of individuals considered *not* to be ethically relevant. An opportunity-egalitarian distribution rule, for example, deems factors beyond an individual's control to be ethically arbitrary, hence judges that such factors' effects should be equalized across individuals. Call the latter equalisand "ethically exogenous opportunity and risk," or if you prefer, simply "opportunity" for short. (The longer form's better because we "make" some of "our own" opportunities and risks; they're not all of them accidents, some are what we can call "ethically endogenous.")

sort directs attention both (a) to that which must be attended to in satisfactorily addressing distributive questions, as in effect noted at the end of Section 2, and (b) to that which can actually be *done*, via implementary mechanism, to ensure that the normatively interesting social welfare aggregate per our discussion above ends up being the aggregate that *is* indeed “maximized.”

The employment of social maximization formulae is, I think, no more than an artifact of the tendency, when SWFs were first introduced, simply to project the model of individual rational action as a species of maximizing behavior upon “society as a whole.” But the species of distributive decision implicated by social welfare concerns is as a normative matter critically distinct from that implicated by individual consumption decisions.

In the latter, individual case, rates of exchange among vector components of commodity bundles can without causing mischief simply be read off of actual consumption behavior, which so far as price theory is concerned can be prompted by tastes or preferences alone. In the social case, by contrast, the order is reversed: It is the normative distributional decision that must be made first – in the choice of argument domain and aggregation rule – in order for there to be so much as a normatively intelligible measure of social “welfare” at all. Only thereafter is the maximandum implied. But this means we ought to be acting, in our collective capacity, pursuant to some such maxim as this: Distribute in the way that you ought, and the appropriate form of maximization will follow.⁴⁵

Recasting our formal social welfare inquiries and programs will of course leave all the room in the world for carefully parsed formal treatment, including statistical-empirical treatment. (See, e.g., Baumol (1986); Fleurbaey (2008); Kolm (1971); Moulin (1995, 2003); Roemer (1994, 1996, 1998); Sen (1992, 1997).) Indeed it demands it.⁴⁶ But it leaves less room for social maximization formulae, and no room at all for either strict preference-regard (strict “welfarism”) or the preferential Paretianism that we’ve seen here to entail it.

Now index such opportunity individual by individual and symbolize it thus: “ O_i ” (Please think of this, not as spelling out a familiar East End London ejaculation, but as standing for person i ’s opportunity.) Now call that which the individual produces out of such opportunity – something which I think our commitments to autonomy commit us to allowing her to maximize untrammelled – “ethically cognizable welfare,” and symbolize it thus: “ WE_i .” (Please think of this as standing for person i ’s ethically (E) cognizable welfare (W .) Then an opportunity-egalitarian social welfare function can be formulated either thus: $\forall i: \text{Max } \sum WE_i$, or equivalently and operationally more helpfully, $\forall i: \text{Eq } O_i$.

The appropriately characterized – i.e., ethically contoured or “weighted” – summand in the first formula corresponds as full complement to the equalisand of the second formula. And because the most direct way, operationally speaking, to ensure that the first formula is satisfied is by acting on the second formula (maximization of the first will “take care of itself,” i.e. will follow directly, upon our satisfying the second), formulae of the second type will prove ultimately more helpful to those who implement policy. Our attachment to maximization formulae even in the social and not just the individual context, I think, is simply a relic. It is a relic of the confluences I’ve been at pains here to disentangle, a relic partly also of the mathematical methods that happened to be prevalent in the most admired natural sciences during a formative period of welfare economics, and a relic that’s now easily and best-advisedly dispensed with.

⁴⁵ See supra, note 44. In short, if you are equalizing the right thing – e.g., material opportunity – and distributing resources accordingly, the recipients of the resources will quite automatically maximize the normatively appropriate social aggregate – e.g., equal opportunity grounded satisfaction, i.e., welfare.

⁴⁶ It demands it, for example, in parsing the boundary between ethically exogenous and ethically endogenous opportunity per note 44.

Indeed, if there is but one lesson to draw from the foregoing Sections, it is that while “fairness versus welfare” is a locution that is quite literally nonsensical in virtue of fairness’s constitutivity of welfare itself, “Pareto versus welfare” would be quite apt. For it is none but the routine deployment of a normativity-inimical, preferential Pareto that has blocked theorists from developing a coherent conception of social welfare that normative legal and policy theory might seek to advance.

It is now long forgotten that Pareto the man was the dubious beneficiary of a Fascist state funeral presided over by none other than Il Duce himself, a man whom Pareto would surely have loathed. Certainly a more fitting tribute on this, the *Manual*’s centenary, would be to lay both our misdeployed Paretianism and misnamed strict “welfarism” to rest, and move on to advancing real *welfare* – which has always been no more and no less than fair satisfaction.⁴⁷

School of Law, Cornell University, 316 Myron Taylor Hall, 524 University Avenue, Ithaca, NY 14853, U.S.A.; robert-hockett@lawschool.cornell.edu

REFERENCES

- ADLER, MATTHEW D. (2007): “Well-being, Inequality, and Time: The Time-Slice Problem and Its Policy Implications” (working paper, on file with the author).
- ANSCOMBE, G. E. M. (1963): *Intention*. Cambridge, MA: Harvard Univ. Press.
- ARNESON, RICHARD. (1990): “Primary Goods Reconsidered,” *Nous*, 24, 429-54.
- ARROW, KENNETH J. (1951): *Social Choice and Individual Values*. New Haven, CT: Yale Univ. Press.
- BARBERÀ, SALVADOR, AND MATTHEW O. JACKSON (2004): “Choosing How to Choose: Self-Stable Majority Rules,” *Quarterly J. Econ*, 119, 1011-48.
- BASU, KAUSHIK (1984): “The Right to Give Up Rights.” *Economica*, 51, 413-22.
- BASU, KAUSHIK, AND TAPAN MITRA (2003): “Aggregating Infinite Utility Streams with Intergenerational Equity: The Impossibility of Being Paretian,” *Econometrica*, 71, 1557-63.
- BATESON, GREGORY; ET AL. (1956): “Toward a Theory of Schizophrenia,” *Behav. Sci.* 1, 27-52.
- BAUMOL, WILLIAM (1986): *Superfairness*. Cambridge, MA: MIT Press.
- BERGSON, ABRAM (1938): “A Reformulation of Certain Aspects of Welfare Economics,” *Quarterly J. Econ*, 52, 310-34.
- BERGSTROM, THEODORE (1999): “Systems of Benevolent Utility Functions,” *J. Pub. Econ. Theory*, 1, 71-90.
- CALABRESI, GUIDO (1972): *The Costs of Accidents: A Legal and Economic Analysis*. New Haven, CT: Yale Univ. Press.
- CAMPBELL, DONALD E., AND JERRY S. KELLY (2002): “Impossibility Theorems in the Arrovian Framework.” In *Handbook of Social Choice and Welfare*, vol. 1, edited by Kenneth J. Arrow, Amartya K. Sen, and Kotaro Suzumura. Amsterdam: Elsevier Sci.

⁴⁷ “Put the ‘fair’ back in ‘welfare,’” one might say, should she be seeking a slogan and willing to pardon a pun.

- CHANG, HOWARD F. (2000a): "A Liberal Theory of Social Welfare: Fairness, Utility, and the Pareto Principle," *Yale Law J.*, 110, 173-235.
- _____ 2000b. "The Possibility of a Fair Paretian," *Yale Law J.*, 110, 251-58.
- DAVIDSON, DONALD (1980 [1978]): "Intending." Reprinted in *Essays on Actions and Events*. Oxford: Oxford Univ. Press.
- DIAMOND, PETER (1965): "The Evaluation of Infinite Utility Streams," *Econometrica*, 33, 170-76.
- DWORKIN, RONALD M. (1977): *Taking Rights Seriously*. Cambridge, MA: Harvard Univ. Press.
- FARMER, ROGER E. A. (1999): *Macroeconomics of Self-Fulfilling Prophecies*. Cambridge, MA: MIT Univ. Press.
- FLEURBAEY, MARC (2008): *Fairness, Responsibility, and Welfare*. Oxford: Oxford Univ. Press.
- FLEURBAEY, MARC, BERTIL TUNGODDEN, AND HOWARD F. CHANG (2004): "Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle: A Comment," *J.P.E.*, 111, 1382-85.
- FULLER, LON L. (1964): *The Morality of Law*. New Haven, CT: Yale Univ. Press.
- HARMAN, GILBERT (1976): "Practical Reasoning," *Rev. Metaphysics*, 29, 431-63.
- HARSANYI, JOHN (1953): "Cardinal Welfare in Welfare Economics and in the Theory of Risk-Taking," *J. Pol. Econ.*, 61, 115.
- _____ (1955): "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *J. Pol. Econ.*, 63, 309.
- HART, OLIVER D. (1975): "On the Optimality of Equilibrium When the Market Structure is Incomplete," *J. Econ. Theory* 11, 418.
- HICKS, JOHN (1940): "The Valuation of Social Income," *Economica*, 7, 105-12.
- HOCKETT, ROBERT (2008a): "Noncomparabilities and Nonstandard Logics." Manuscript under revision for *Econ. & Phil.* Ithaca, NY: Cornell University Law School.
- _____ 2008b. "Recursive Social Welfare Functions." Working Paper. Ithaca, NY: Cornell University Law School.
- _____ 2009a. "Taking Distribution Seriously." Working Paper. Ithaca, NY: Cornell University Law School.
- _____ 2009b. "Why Paretians Can't Prescribe: Preference, Principles, and Imperatives in Law and Policy." Working Paper. Ithaca, NY: Cornell University Law School.
- _____ 2009c. "Reflective Intensions: Two Foundational Decision-Points in Mathematics, Law, and Economics." Working Paper. Ithaca, NY: Cornell University Law School.
- _____ 2009d. "Justice in Time." *George Washington Law Review*, 76, ___ (forthcoming symposium issue on intergenerational equity).
- HURKA, THOMAS (2002): *Perfectionism*. Cambridge, MA: Harvard Univ. Pr.
- KALDOR, NICHOLAS (1939): "Welfare Propositions of Economics and Interpersonal Comparisons of Utility," *Econ. J.*, 49, 549-52.
- KAPLOW, LOUIS, AND STEVEN SHAVELL (1999): "The Conflict between Notions of Fairness and the Pareto Principle," *Am. Law & Econ. Rev.*, 1, 63-77.
- _____ (2000): "Notions of Fairness versus the Pareto Principle: On the Role of Logical Consistency," *Yale Law J.*, 110, 237-49.

- _____. 2001. "Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle," *J.P.E.*, 109, 281-86.
- _____. 2002. *Fairness versus Welfare*. Cambridge, MA: Harvard Univ. Press.
- _____. 2004. "Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle: Reply," *J.P.E.*, 112, 249-51.
- KOLM, SERGE-CHRISTOPHE (1971): *Justice et Équité*. Paris: Centre d'Études Prospectives et d'Économie Mathématique Appliquées à la Planification.
- KOOPMANS, T. (1960): "Stationary Ordinal Utility and Impatience," *Econometrica* 28, 287-93.
- KORAY, SEMIH (2000): "Self-Selective Social Choice Functions Verify Arrow and Gibbard-Satterthwaite Theorems," *Econometrica*, 68, 981-95.
- LERNER, ABBA (1975 [1946]): *The Economics of Control*. London: A. M. Kelley.
- LJUNGQVIST, LARS; AND THOMAS J. SARGENT (2004): *Recursive Macroeconomic Theory* (2nd ed.). Cambridge, MA: MIT Press.
- MELVILLE, HERMAN (1856): "Bartleby, the Scrivener: A Story of Wall Street," in *The Piazza Tales*. New York, NY: Dix & Edwards.
- MOULIN, HERVÉ (1995): *Cooperative Microeconomic Theory*. Princeton: Princeton Univ. Press.
- _____. (2003): *Fair Division and Collective Welfare*. Cambridge, MA: MIT Press.
- PARETO, VILFREDO (1909): *Manuale di Economia Politica*. Milan: Societa Editrice Libreria. English translation, 1927. *Manual of Political Economy*. New York, NY: A. M. Kelly.
- PARFIT, DEREK (1984): *Reasons and Persons*. Oxford: Oxford Univ. Press.
- POPE, ALEXANDER (1732): *Essay on Man*.
- QUINE, W. V. O. (1951): *From a Logical Point of View*. Cambridge, MA: Harvard Univ. Press.
- RAWLS, JOHN (1971): *A Theory of Justice*. Cambridge, MA: Harvard Univ. Press.
- ROEMER, JOHN E. (1993): "A Pragmatic Theory of Responsibility for the Egalitarian Planner," *Philosophy & Public Affairs*, 22, 46-66.
- _____. (1996): *Theories of Distributive Justice*. Cambridge, MA: Harvard Univ. Press.
- _____. (1998): *Equality of Opportunity*. Cambridge, MA: Harvard Univ. Press.
- SAMUELSON, PAUL A. (1947): *Foundations of Economic Analysis*. Cambridge, MA: Harvard Univ. Press.
- SEN, AMARTYA K. (1970a): "The Impossibility of a Paretian Liberal," *J.P.E.*, 78, 152-57.
- _____. (1970b): *Collective Choice and Individual Values*. Amsterdam: North-Holland.
- _____. (1979): "Utilitarianism and Welfarism." *J. Phil.*, 76, 463-89.
- _____. (1986): "Social Choice Theory." In *Handbook of Mathematical Economics*, vol. 3 edited by K. Arrow and M. Intriligator. Amsterdam: North-Holland.
- _____. (1992): *Inequality Reexamined*. Cambridge, MA: Harvard Univ. Press.
- _____. (1997): *On Economic Inequality*. Oxford: Oxford Univ. Press.
- _____. (2003): *Rationality and Freedom*. Cambridge, MA: Belknap Press.
- SMART, J. J. C. (1973): "An Outline of Utilitarianism," in Smart & Williams, *Utilitarianism: For and Against*. Cambridge: Cambridge Univ. Press.
- STOKEY, NANCY L.; ET AL. (1989): *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard Univ. Press.

- SUZUMURA, KOTARO (2001): "Pareto Principles from Inch to Ell," *Econ. Letters*, 70, 95-98.
- TUNGODDEN, BERTIL (2003): "The Value of Equality." *Econ. & Phil.* 19: 1-23.
- VARIAN, HAL (1992): *Microeconomic Analysis* (3rd ed.). New York: Norton.
- VELLEMAN, J. DAVID (2007 [1989]): *Practical Reflection*. Stanford: CSLI Publications.
- WITTGENSTEIN, L. (1921): *Tractatus Logico-Philosophicus*. London: Routledge & Kegan-Paul.
- _____ (1953): *Philosophical Investigations*. Oxford: Basil Blackwell.
- _____ (1984): *Zettel*. Oxford: Basil Blackwell.