Revue des livres / *Book Review*
Comptes rendus / *Reviews*

Marc Fleurbaey and François Maniquet, *A Theory of Fairness and Social Welfare*

Cambridge, MA: Cambridge University Press, 2011, 293 pages, ISBN: 978-0521887427 (hardback); 978-0521715348 (paper)

Matthew D. Adler*

A Theory of Fairness and Social Welfare ("TFSW") is a brilliant book. The authors, Marc Fleurbaey and François Maniquet, are world-class social choice theorists, and this book will burnish their reputations as leading figures in the field.

The key concept of TFSW is that of the "social ordering function" (SOF). This represents a major intellectual extension of the literature on so-called "fair allocation." (Thomson, 2011). That literature seeks to identify allocations of goods that are not only *optimal*, in the sense of being Pareto efficient, but satisfy fairness properties—more specifically properties that do *not* presuppose utility numbers which are cardinal or interpersonally comparable (for short, "ONC" fairness properties). One well-known such ONC property of fair allocation theory is "no envy": the selected optimal allocation should be such that no person prefers anyone else's bundle to his own. Another is "equal split selection": if an equal division of the total feasible stock of goods is optimal, then it should be selected.

TFSW seeks to identify a rule for ordering all possible allocations, not merely optimal ones: a rule that takes the form of an ordering (a transitive, reflexive and complete binary relation on allocations); that is Paretian in the sense of satisfying the Pareto indifference axiom and the strong or at least weak Pareto axiom; and that satisfies fairness properties that are both ONC properties, and capture some intuitive sense of what makes one allocation of goods more fair or equal than another.

*Duke University. adler@law.duke.edu

The book makes a large contribution to at least three literatures: not only fair allocation theory, but Arrovian social choice theory and, finally, egalitarian welfare economics. Arrow's theorem and extensions (e.g., to economic domains; see Bordes and Le Breton 1989) show that a Paretian nondictatorial social ordering of outcomes, defined just as a function of individuals' ordinal preferences, is impossible (given sufficient diversity of preferences), if the ordering must also conform to a strong independence requirement ("independence of irrelevant alternatives"). TFSW demonstrates the possibility of a Paretian nondictatorial—indeed, fairness-regarding—ordering of allocations, defined on ordinal preferences, and notwithstanding the diversity thereof, if a slightly weaker (but still demanding) independence requirement is substituted for Arrow's version. And, as I will explore further in a moment, TFSW substantially enriches our understanding both of the "Pigou-Dalton" principle (surely a key component of any adequate account of equality), and the place in egalitarianism of maximin or leximin approaches that give absolute priority to the worst- or worse-off.

TFSW is rigorous and axiomatic, in the best tradition of social choice theory. It delivers a stunning array of impossibility results or logical implications. But the presentation is also very accessible. Results are expressed both formally and in English. What drives them is made clear with many helpful graphs. Long proofs are relegated to an appendix.

Parts I and II develop the SOF concept in the "pure distribution" context: society has some stock of goods, already produced, and the comparative fairness of different possible allocations of this total stock now needs to be determined. Part III shows how the ideas developed in the first two parts can be extended to scenarios where production decisions remain to be made. For example, individuals can increase the stock of goods through labor, or private resources can be taxed by government to fund a public good. Parts I and II (chapters 1-7) are the intellectual core of TFSW, and shall be the focus of my comments. Part III (chapters 8-11) is icing on a very rich cake.

In the "pure distribution" context, TFSW defines the SOF as follows. (My presentation, here and below, generally follows the TFSW notation, making a few slight changes where a simpler symbolism suffices for my purpose.) There are $l \geq 2$ kinds of goods, and N individuals. The social endowment of goods is Ω, an l-dimensional vector. An allocation z of goods specifies a bundle z_i for each individual i, with such bundle having l components. Each good is most naturally thought of as some kind of transferrable physical item (although TFSW in chapter 7 discusses how to extend the model to internal, nontransferable "functionings"), and is represented by a nonnegative number. An allocation is *feasible* if the total amount of each of the l

goods (summed across the N individuals) is less than or equal to the amount in Ω.

Each individual i has preferences R_i, namely a complete ordering of all possible bundles. These are assumed to be continuous, monotonic and (to begin with) convex. R^N is the N-fold profile of these preferences. An "economy" E is a pair (R^N, Ω), and a SOF (denoted in bold as **R**) is a mapping from economies to orderings of allocations. That is, an SOF **R** takes the form **R**(E) = **R**(R^N, Ω). z **R**(E) z^* is shorthand for: according to the ordering of allocations generated *by* SOF **R** from economy E, allocation z is at least as good as allocation z^*.

It is critical to note that SOFs have *two* arguments: not just profiles of preferences (a familiar idea from multiprofile social choice theory going back to Arrow), but also social endowments.

Chapter 1 presents the SOF concept, and situates it relative to fair allocation theory and Arrovian social choice, as well as other components of welfare economics (such as social welfare functions, cost-benefit analysis, and multidimensional methods). Chapter 2 explores different versions of the Pigou-Dalton (PD) principle. This principle, generically, concerns a transfer of some "currency," from a transferor who starts out with more, to a transferee who starts out with less, which satisfies two properties: it is "non-leaky" in the sense that the transferee gains by precisely as much as the transferor loses; and it is small enough that even after the transfer the transferee has no more of the currency than the transferor. Chapter 2, specifically, discusses versions of the PD principle framed in terms of *goods*. The simplest version presented in that chapter says: If allocation z^* is reached from z by simultaneous non-leaky transfers in all l goods, from a transferor with more of *each of the l* goods, and which leaves the transferee still with strictly less of each than the transferor (and if no one else's holdings change), *then* the SOF must (weakly) prefer z^* to z.

Strikingly, this simple version of the PD principle is inconsistent with the weak Pareto principle. Chapter 2 is thus led to explore Pareto-consistent restrictions thereof, such as: "equal-split transfer," "transfer among equals," and "nested-contour transfer." The first requires an SOF to satisfy the PD principle in the sense stated in the previous paragraph *if* the transferee ends up with less than $1/N$ of Ω in each good, and the transferor with more. The second does so *if* the transferee and transferor have the same preferences. The third, a strengthening of the second, does so *if* the transferee's post-transfer indifference curve is strictly below the transferor's.

Note that all four versions of the PD principle just discussed are ONC principles: the simple principle is framed in terms of transfers of physical quantities of goods, *not* utilities; and the restrictions in the latter three versions are either good-based (in the case of equal-split transfer), or based on ordinal properties of each individual's preferences (what her indifference curves look like).

Chapter 3 concerns the relation between the PD principle and a much stronger *absolute* priority for the worse off. It contains two landmark theorems. "Priority among equals" is an absolute-priority principle that says: if in allocation z one individual, the transferee, has less of all l goods than a second, the transferor; and if a transfer (*however leaky*) occurs such that in z^* the transferee's holdings of each of the goods have increased by some amount, and the transferor's have decreased by some amount, and the transferee remains with less than the transferor in all l dimensions, with no one else's holding changed; and if the two individuals have the same preferences, *then*: z^* is weakly socially preferred to z. Theorem 3.1 shows that if an SOF satisfies Pareto indifference, transfer among equals, and "unchanged-contour independence", then it satisfies "priority among equals."

Unchanged-contour independence, a weakened version of the Arrow "independence of irrelevant alternatives" requirement, forms one of the linchpins of TFSW. Deriving originally from work by Hansson (1973), it here says: if two economies E and E' contain the same Ω and contain profiles of preferences, R^N and R^N+, such that for all N individuals each individual i's indifference curves in R^N and R^N+ are the same at his bundle z_i and the same at his bundle z_i^*, then the SOF must not differentiate between these economies at these bundles: z $\mathbf{R}(E)$ z^* iff z $\mathbf{R}(E')$ z^*. Unchanged-contour independence not only figures centrally in Theorem 3.1, but is satisfied by the two main SOFs highlighted by TFSW: $\mathbf{R}^{\Omega\text{lex}}$ and $\mathbf{R}^{\text{EW}}$, to be described momentarily.

Theorem 3.1 shows, astonishingly, that Paretianism plus a minimal ONC egalitarian requirement (transfer among equals) *plus* a robustness requirement (independence) yields a kind of absolute priority. To be sure, this robustness requirement is contestable, as I'll discuss. But Theorem 3.2 traces a different path from Paretian ONC egalitarianism to absolute priority. "Nested-contour priority" is the absolute-priority analogue of nested-contour transfer, favoring a transfer (however leaky) as long as the transferee's post-transfer indifference curve lies strictly below the transferor's. Theorem 3.2 shows that if an SOF satisfies Pareto indifference and nested-contour transfer, and is separable with respect to unaffected individuals (in the sense of being invariant to both their holdings and preferences), then it satisfies nested-contour priority. Robustness has been swapped out for separability (harder to contest), and yet we still get absolute priority.

Chapter 4 addresses the social-endowment sensitivity of SOFs; considers variations on "unchanged contour independence"; and discusses in depth the difference between SOFs and fair allocation rules. Chapter 5 describes two favored SOFs, $\mathbf{R}^{\Omega\text{lex}}$ and $\mathbf{R}^{\text{EW}}$. Each converts an allocation z into an N-fold vector of individual indices. In the case of $\mathbf{R}^{\Omega\text{lex}}$, individual i's index is the fraction λ_i such that i is indifferent between that fractional share of Ω and z_i. In the case of $\mathbf{R}^{\text{EW}}$, individuals i's index is z_i converted into money at certain Ω-relative (and allocation-relative) prices. $\mathbf{R}^{\Omega\text{lex}}$ applies the leximin rule to its indices, while $\mathbf{R}^{\text{EW}}$ the maximin rule.

In both cases, the indices might be termed "utilities," since they are numerical representations of preferences, but if so are ordinal and interpersonally noncomparable "utilities." They are generated from the profile R^N of individual ordinal rankings of bundles, plus the social endowment Ω. Thus, if U is an N-fold vector of individual utility functions $(u_1(.), ..., u_N(.))$, and U^* a profile of individual-specific ordinal transformations thereof $(f_1(u_1(.),...., f_N(u_N(.))$—with $f_i(.)$ any strictly increasing function— $\mathbf{R}^{\text{EW}}$ and $\mathbf{R}^{\Omega\text{lex}}$ using utility functions rather than preferences as initial inputs (and then inferring the profile of preferences R^N from those utility functions) are invariant to the substitution of U^* for U.

$\mathbf{R}^{\text{EW}}$ satisfies Pareto indifference, weak Pareto, and nested contour transfer and priority in their weak forms. All three "weaks" can be strengthened to "strong" with a leximin extension of $\mathbf{R}^{\text{EW}}$, but doing so has problematic implications in terms of separability. Moreover, chapter 6 shows that $\mathbf{R}^{\text{EW}}$ is problematic with nonconvex preferences. $\mathbf{R}^{\Omega\text{lex}}$ satisfies Pareto indifference, strong Pareto, and nested contour transfer and priority in their strong forms; is fully separable with respect to unaffected individuals; and generalizes nicely to nonconvex economies. All in all, $\mathbf{R}^{\Omega\text{lex}}$—the SOF analogue of the concept of "egalitarian equivalence" in fair allocation theory—seems to emerge the winner from TFSW's tournament of axioms.

$\mathbf{R}^{\Omega\text{lex}}$ is social-endowment sensitive. Ω, a feature of the economy—the total stock of distributable goods—is not merely a formal argument of $\mathbf{R}^{\Omega\text{lex}}$ as for any SOF. $\mathbf{R}^{\Omega\text{lex}}$ is such that its ranking of allocations for a given profile of preferences can *vary* as Ω does. However, it should be noted that $\mathbf{R}^{\Omega\text{lex}}$ as well as $\mathbf{R}^{\text{EW}}$ are invariant to multiplication of Ω by a positive constant. They thus take account of the relative proportions of goods in Ω, as well as preferences, in ranking allocations.

The Arrow problem, *sensu stricto*, is to arrive at a single ranking of a set of social states for each profile of preferences. $\mathbf{R}^{\Omega\text{lex}}$ and $\mathbf{R}^{\text{EW}}$, because they are social-endowment sensitive, do not actually demonstrate how this problem can be solved by weakening the "independence of irrelevant alternatives" requirement to unchanged-contour independence. But a related SOF which is discussed (although not endorsed) by TFSW does illustrate this. This SOF, $\mathbf{R}^{\Omega 0\text{lex}}$, follows the same approach as $\mathbf{R}^{\Omega\text{lex}}$, but using an arbitrary, fixed stock of goods Ω_0 rather than the economy's stock Ω.

Chapters 6 and 7 extend the analysis of the first chapters to various specific economics domains (nonconvexity, indivisible goods, homothetic preferences), and in other directions.

An "internal" critique of TFSW—internal to the project of describing social orderings that build upon the intellectual tradition of the fair-allocation literature—appears impossible. To the eyes of this reviewer (admittedly not someone who writes in the fair-allocation tradition), TFSW seems to have fully succeeded at this project. The more important criticism is "external," namely that the SOF framework has built-in flaws which are avoidable by using a social welfare function to rank vectors of interpersonally comparable utilities corresponding to allocations, *if* a normatively plausible construction of such utilities is available.

Assume, for the moment, that a plausible utility function $u(.)$ with the following features can indeed be constructed: (1) $u(.)$ maps pairs of bundles and preferences onto utility numbers. $u(.) = u(b, R)$, with b a bundle (a vector of the l goods) and R a preference. (2) For different bundles with the same preferences, $u(.)$ conforms to the preference. $u(b, R) \geq u(b^*, R)$ iff $b\ R\ b^*$. (3) $u(.)$ is unique or at least unique up to a positive ratio transformation.

If so, $u(.)$ can be used to define a "fair social welfare function" **W**, which is a mapping from a profile of preferences onto a complete ordering of allocations. Define **W** as follows. $z\ \mathbf{W}(R^N)\ z^*$—to be read as "z at least as good as z* according to **W**, given preferences R^N"—iff

$$\sum_{i=1}^{N} g(u(z_i, R_i)) \geq \sum_{i=1}^{N} g(u(z_i^*, R_i))$$

with $g(.)$ a strictly increasing and strictly concave real-valued function. (For a general discussion, see Adler 2012, ch. 5).

If $u(.)$ is unique up to a ratio transformation, $g(.)$ can be the power function $g(x) = (1-\gamma)^{-1}x^{1-\gamma}$, with γ positive. **W** with this particular $g(.)$ function will order allocations the very same way, for a given profile of preferences, if $u(.)$ is replaced with $u^*(.) = ku(.)$, k positive. If $u(.)$ is fully unique, any strictly increasing and strictly concave $g(.)$ can be used. Note that the utilities assigned by $u(.)$ in either case are cardinal and interpersonally comparable, since the only admissible transformation is either a ratio rescaling with a common positive k, or none at all.

Because $u(.)$ by hypothesis has been constructed to conform to R, **W** shares with $\mathbf{R}^{\Omega\text{lex}}$ the virtue of satisfying both the Pareto indifference and the strong Pareto axioms. Like $\mathbf{R}^{\Omega\text{lex}}$, it is separable with respect to unaffected individuals (being invariant to changes in either their holdings or preferences). However, it lacks two features of $\mathbf{R}^{\Omega\text{lex}}$ that some egalitarians, at least, will find problematic. First, many find the PD principle compelling, but are troubled by absolute priority for the worse- or even worst-off. **W** satisfies the PD principle in *utility*. If i, j, z, and z^* are such that $u(z_i, R_i) > u(z_i^*, R_i) \geq u(z_j^*, R_j) > u(z_j, R_j)$, and $u(z_i, R_i) - u(z_i^*, R_i) = u(z_j^*, R_j) - u(z_j, R_j)$, with everyone else indifferent between z and z^*, then **W** strictly prefers z^* to z. However, **W** does not give absolute priority to those at lower utility levels, and the extent of priority it gives (how leaky a utility transfer can be to be acceptable) can be adjusted by changing the concavity of *g*(.). Whether **W** gives absolute priority with respect to transfers of *goods* (not utility) is not so straightforward. But it is easy to see that if *u*(.) as well as the transformation function *g*(.) are unbounded above, **W** will not give absolute priority in terms of goods. A loss to one with fewer goods will be a finite loss in transformed utility, which (given the unboundedness assumptions) can be counterbalanced by a sufficiently large increase in the holdings of someone with more goods.

Second, the social-endowment sensitivity of $\mathbf{R}^{\Omega\text{lex}}$ is puzzling, indeed doubly so. Let Z(Ω) be the set of allocations whose sums of each of the *l* goods are less than or equal to the amounts in Ω. TFSW characterizes each allocation within Z(Ω) as "feasible." Even if this is true, the classical approach to rational choice is to maximize the attainment of the decision maker's objective given feasibility constraints, rather than to build constraints into the description of the objective. **W** assumes this classical form, since it makes the goal (the identification of better or worse allocation) depend upon individuals' preferences, but not on Ω itself.

But, in fact, it cannot be true that *every* allocation within Z(Ω) is feasible. If this were the case, the Paretian social planner would be irrational to choose *any* Pareto inefficient allocation in Z(Ω), one that is Pareto inferior to some other allocation in Z(Ω). The very justification for an SOF approach that ranks all allocations in Z(Ω), according to TFSW, is that the social planner might find herself forced to choose between two allocations *z* and *z** both of which are Pareto-inefficient in Z(Ω), and might have good reason to choose one or the other. But this can only be the case if there are hidden feasibility constraints, not explicitly included in the "pure distribution" model of TFSW. Ω identifies the total stock of goods that exist within the society in question, but administrative costs, political economy considerations, the threat of violence by those who currently possess goods, etc., prevents their redistribution.

Wouldn't it be a better approach for the social planner to rank allocations as a function of individual well-being (determined by allocations and preferences), and then choose among them with a view to what's actually feasible—rather than to make the ranking a function of an Ω that is "feasibly" distributable only in a hypothetical world *different* from the one the planner actually finds herself in, a hypothetical world lacking the actual feasibility constraints that make it rational for her (in *her* world) to choose an allocation Pareto inefficient in Z(Ω)?

In chapter 4, TFSW gives a partial riposte to this skeptical line of questioning. Assume that Ω is genuinely feasible, and that the social planner is therefore rationally constrained (if she is Paretian) to choose a Pareto-efficient allocation in Z(Ω). If her social ordering is just a Ω-independent function of individual preferences and allocations, and satisfies weak Pareto, then Theorem 4.1 shows that the ordering may fail to choose an equal split of Ω even if it is Pareto-efficient. But the proponent of **W** has a ready response: individuals with equal bundles at a Pareto-efficient allocation may have (globally if not locally) different preferences, and thus different *utilities* as assigned by the hypothesized *u*(.) that takes preferences as well as bundles as its arguments. An equal split in resources, even if Pareto efficient, does not necessarily leave individuals with diverse preferences equal in terms of interpersonally comparable well-being.

Finally, it is worth comparing **W** and $\mathbf{R}^{\Omega\text{lex}}$ with respect to the robustness property that is central to TFSW: unchanged-contour independence. There is no particular reason to think that **W** will have this property, but this does not seem especially troubling. Consider the case in which *u*(.) is unique up to a positive ratio transformation. *u*(.) assigns determinate utility ratios between any two pairings of preferences and bundles. In particular, *u*(.) will assign determinate utility ratios between any two indifference curves of the same particular preference *R*. That ratio, intuitively, may depend upon the overall pattern of indifference curves for that preference. But "unchanged contour independence" requires, in this context, that if two given indifference curves (specified by the bundles they connect) belong to some preference, the utility ratios between the two curves must be the same *regardless of the preference to which they belong*. If utility is indeed interpersonally comparable on a ratio scale, that seems like a very strong and unwarranted demand.

So we come back to the hypothesized *u*(.). The "external critique" of TFSW that I have set forth hinges upon the availability of such a utility function. In other work, building upon John Harsanyi's concept of "extended preferences," I have outlined an approach to constructing utilities that are interpersonally comparable, determine utility ratios, and take account of individuals' preferences. I cannot discuss the details here. (Adler 2012, ch. 3; Harsanyi 1986, ch. 4). Whatever its promise, the approach is hardly well-established. Nor is any other. It must be conceded that the proponents of social welfare functions currently lack an established, consensus account for arriving at an interpersonally comparable measure of individual well-being that is both sensitive to individual preferences, and allows for the diversity thereof.

So the "external critique" is perhaps not a "critique" at all. It shows how the SOF lacks certain arguable virtues of a methodology for social choice, the social welfare function, which is itself still a "work in progress." And, again, the "critique" (if it is one) is external, not internal. In writing TFSW, Fleurbaey and Maniquet aspired to fully develop the concept of an SOF. They have succeeded masterfully in that aim. For those (such as this author) who remain inclined toward alternative approaches, TFSW will function as both spur and exemplar—challenging us to draw blueprints for *our* favored methodologies that are as precise and elegant as the blueprint Fleurbaey and Maniquet have set down for theirs.

References

Adler, Matthew D. 2012. *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*. Oxford: Oxford University Press.

Bordes, Georges and Michel Le Breton. 1989. Arrovian Theorems with Private Alternatives Domains and Selfish Individuals. *Journal of Economic Theory*, *47*(2): 257-281.

Hansson, Bengt. 1973. The Independence Condition in the Theory of Social Choice. *Theory and Decision*, *4*(1): 25-49.

Harsanyi, John C. [1977] 1986. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.

Thomson, W. 2011. Fair Allocation Rules. In K. Arrow, A. Sen, and K. Suzumura, eds, *Handbook of Social Choice and Welfare*, vol. 2, 393-506. Amsterdam: Elsevier.