Soc Choice and Welfare 25:381–418 (2005)
DOI 10.1007/s00355-005-0010-1

ORIGINAL PAPER

Marc Fleurbaey · Philippe Mongin

The news of the death of welfare economics is greatly exaggerated

Published online: 3 December 2005

Abstract The paper reexamines the controversy about Bergson–Samuelson social welfare functions (BSFs) that took place between welfare economists and social choice theorists as a consequence of Arrow's (1951) impossibility theorem. The 1970's witnessed a new version of the theorem that was meant to establish that BSFs "make interpersonal comparisons of utility or are dictatorial." Against this, Samuelson reasserted the existence of well-behaved "ordinalist" BSFs and generally denied the relevance of Arrovian impossibilities to welfare economics. The paper formalizes and reassesses each camp's arguments. While being also critical of Samuelson's, it eventually endorses his conclusion that welfare economics was left untouched by the controversy. It draws some connections of BSFs with contemporary normative economics.

> "Many readers can be forgiven for thinking that Arrow has proved the impossibility of a Bergson Social Welfare Function, thereby dealing a death blow to the magnificent edifice of modern economics" (Samuelson PA (1967) Arrow's mathematical politics. In Hook S (ed) Human values and economic policy. New York University Press, New York, p. 418)

1 Introduction and preview

Welfare economics does not enjoy a flattering reputation among today's economists. Nearly all of them believe that it is a theory of the past, and if a few writers exceptionally lament its "strange disappearance" (Atkinson 2001), the larger number believe that it was conceptually flawed and deserved its fate anyway. Our paper aims at critically re-examining—and as will be seen, rebutting—this received view

M. Fleurbaey (✉)
CERSES, CNRS and Université Paris 5, 59-61 rue Pouchet, 75017 Paris, France
E-mail: marc.fleurbaey@cerses.cnrs.fr

P. Mongin
Laboratoire d'économétrie, C.N.R.S. et Ecole Polytechnique, 1 rue Descartes, 75005 Paris, France
E-mail: mongin@poly.polytechnique.fr

of welfare economics. A complete account of the fate of welfare economics would go beyond the scope of a single paper, and we have limited attention to those controversies which attended to the "Bergson–Samuelson social welfare function", leaving aside what should also be said about the "Compensation Principle".[1]

A culmination in the new welfare economists' work, Bergson's (1938) "social welfare function" was intended to put Paretian evaluation into a logically complete and coherent framework of thought. Much of its fame is due to the *Foundations* (1947, ch. VIII), where Samuelson both clarified and vindicated Bergson's statement. The success of the "social welfare function" was to be short-lived, since as early as 1951, Arrow's *Social Choice and Individual Values* argued that no such function existed, unless it was dictatorial. The book is categorical about the failure, but not about the remedy; however, it was usually taken also to provide a positive recommendation, i.e., that interpersonal comparisons of utility should be made. Arrow's *reductio to absurdum* depended on his impossibility theorem, and the welfare economists—i.e., Little (1952), Bergson (1954), and Samuelson (1967)-responded to it by denying that the theorem applied to their field. This was an experts' dispute, which the profession could only follow from a distance, but Arrow's reasoning generally made a strong impression, and thus already contributed to shatter the welfare economists' respectability.

The final blow to the "Bergson–Samuelson social welfare function" (as it was relabelled in recognition of Samuelson's role) was given later, when the social choice theorists Kemp and Ng (1976) and Parks (1976) devised a version of the impossibility theorem that was meant to be better suited to the polemical target than was the original one. More simply and explicitly than in *Social Choice and Individual Values*, the chosen strategy was to use the theorem in order to conclude that "it is impossible to find a "reasonable" Bergson–Samuelson [function] based on individual orderings" (Kemp and Ng 1976, p. 59), or equivalently, that "the Bergson function must make interpersonal comparisons [of utility] or be dictatorial" (Parks 1976, p. 450). Claiming that the new variant was no more relevant than Arrow's, Samuelson (1977) rejected the alleged dilemma of dictatorship versus interpersonal comparisons. He argued that it was possible to remain within the confines of "ordinalism", i.e., to rely only on information provided by individual indifference curves, and nonetheless escape the dictatorship conclusion. In brief, there existed well-behaved "Bergson–Samuelson social welfare functions". Despite the arbitration attempted by Pollak (1979), and later rejoinders by Samuelson (1981, 1987), the welfare economists' defence was generally considered weak. Sen's (1970, 1972) pathbreaking work on interpersonal comparisons of utility had already set the pace among social choice theorists, and this quickly growing community accepted Kemp and Ng's and Parks' conclusion almost unreservedly. The message got across to the non-specialists, and it became part of the official history of economics that a major refutation had taken place. If the official death of welfare economics were to be dated with some precision, the years 1976–79 would suggest themselves.

This apparently decisive controversy, with some of its antecedents and sequels, make the object of the present essay. About the earlier discussions involving Arrow, we will say little. They could not be decisive, if only because social choice

[1] On the Compensation Principle, see Chipman and Moore (1978). More recent investigations include Gravel (2001) and Suzumura (1999).

theory had not yet tightened the links between Bergson's "social welfare function" and Arrow's identically named, but distinct concept that drove his version of the impossibility theorem (see Mongin 2002). The controversy of the 1970's is more to the point: it took place at a stage when both camps were in a position to develop their polemics more analytically and less semantically, as it were. However, being struck by remaining unclarities in the arguments, we set out to question the social choice theorists' famous dilemma. Our doubts were reinforced by some recent work on criteria of distributive justice that are in effect well-behaved Bergson–Samuelson functions (Fleurbaey and Maniquet 1996, 2001). These comments summarize our motivations for embarking in the present reconstruction.

Having reformulated the major arguments made by each side, we can bring the news that the social choice theorists' arguments are mostly beside the point. Accordingly, the earlier news of the death of welfare economics is greatly exaggerated—like the news of Mark Twain's death. While drawing this logical and historical conclusion, we do not intend to rehabilitate welfare economics as it was practiced up to the 1970's. The "Bergson–Samuelson social welfare function" was useless for making policy decisions, because over and beyond its commitment to Paretianism, its mathematical shape was left unspecified (see Samuelson's claim that "any possible opinion is admissible", 1947, p. 221, and similarly in Bergson 1966, p. 70). We believe that the theories of fair allocation can give content to the hollow notion. To illustrate this surprising connection, the paper shows that Pazner and Schmeidler's (1978) egalitarian-equivalent criterion comes close to the non-dictatorial function that Samuelson sketched during the controversy. Owing to this unexpected posterity, the ill-famed theory of welfare economics might have a future in the profession after all.

The paper uses a social-choice-theoretic formalism that is presented in two stages. Section 2 recasts the impossibility theorem in both its Arrovian form and in the different form invented in the 1970's. This part of the formalism already permits relevant clarifications, in particular about the role of "single-profile" versus "multi-profile" approaches. Section 3 attempts to clarify what Bergson and Samuelson said of the "social welfare function", as against what later theorists made of it. While Section 2 is limited to preference relations, Section 4 considers utility functions (in specialized language, we move from "social welfare functions" to "social welfare functionals"). Using this more expressive formalism, Section 5 warrants the basic claim that the impossibility theorem, in either its initial or later version, is irrelevant to Bergson's and Samuelson's function. Section 6 reviews additional discussions pertaining to the controversy; in particular, it identifies two theorists whose (unfortunately little-noticed) work antedates our reconstruction (Mayston 1974, 1979, 1982; Pazner 1979). Section 7 briefly concludes.

2 The impossibility theorem and the single vs multi-profile controversy

This section recapitulates the impossibility theorems that were the basis of both the early and late controversies. It also clears the ground from an objection that welfare economists considered essential—Arrow's theorem was said to be irrelevant to welfare economics because it depended on considering many preference configurations ("profiles") at a time. By looking for single-profile counterparts to Arrow's theorem, social choice theorists gave some credence to this famous objection. We will show instead that it was a red herring.

In the usual style of Arrovian social choice theory we assume that there is a finite set of individuals $N=\{1,\dots,n\}$ and a set of alternatives X containing at least three elements; stronger cardinality restrictions will be added when necessary. Individual i's weak ordering is denoted by R_i, and a profile of individual orderings by $\overrightarrow{R} = (R_1, \dots, R_n)$ or $\overrightarrow{R'} = \left(R'_1, \dots, R'_n\right)$. A *social ordering function* F is defined to be a mapping from some domain $\mathcal{D}^F \subseteq (\mathcal{P}_X)^n$ of profiles of individual ordering to $\mathcal{P}_X$ where $\mathcal{P}_X$ refers to the set of all orderings on X.

As usual, R will stand for $F\left(\overrightarrow{R}\right)$, and R' for $F\left(\overrightarrow{R'}\right)$, and any time that P, I replace R in the notation of some ordering, these letters indicate its asymmetric and symmetric parts respectively. Given $\overrightarrow{R}$ we denote by $\overrightarrow{R}\Big|_{\{x,y,z,\dots\}}$ the profile of individual orderings restricted to the subset $\{x,y,z,\dots\} \subseteq X$.

Arrow's (1963, ch. III) abstract definition of a "social welfare function" does not specify its domain, but is usually understood in terms of a *maximum domain*, i.e., F is defined on $\mathcal{D}^F = (\mathcal{P}_X)^n$. This is but one case of a "social welfare function" in Arrow's sense, because his chapter VI introduces what is now called an *economic domain*, i.e., it imposes standard microeconomic restrictions on both X and the R_i. In whichever case, Arrow's analysis always depends on considering several profiles at a time. In contradistinction, Kemp and Ng's (1976) and Pollak's (1979) "social orderings" have a singleton domain, i.e., their F is defined on $\mathcal{D}^F = \left\{\overrightarrow{R}\right\}$ for some given profile $\overrightarrow{R}$. Parks' (1976) "social welfare function" is formulated directly in terms of utility amounts, but substantially reduces to this case. The terminology of "multi-profile" versus "single-profile" settings seems to come from Pollak (1979, p. 73), but without the words, Parks (1976, p. 447) and Kemp and Ng (1976, p. 59) draw the same distinction. Fishburn's (1973, ch. 13–14) distinction between "interprofile" and "intraprofile" conditions has a different conceptual basis. It makes it possible to classify axiomatic conditions *in the multi-profile context*, depending on whether they involve considering one profile or several profiles at a time. We will illustrate it in terms of the conditions on F and eventually argue that it is the more relevant distinction of the two.

The familiar conditions of *dictatorship*, *weak Pareto*, *independence*, and *neutrality*, are restated as follows:

$$(\mathbf{D})\exists i : \forall \overrightarrow{R} \in D^F,\ \forall x,\ y \in X,\ xP_i y \Rightarrow xPy;$$
$$(\mathbf{WP})\ \forall \overrightarrow{R} \in D^F,\ \forall x,\ y \in X,\ (xP_i y,\ \forall i \in N) \Rightarrow xPy;$$
$$(\mathbf{I})\ \forall \overrightarrow{R}, \overrightarrow{R'} \in D^F,\ \forall x,\ y \in X,\ \left[\left(xP_i y \Leftrightarrow xP'_i y\right)\&\left(yI_i x \Leftrightarrow yI'_i x\right),\ \forall i \in N\right] \Rightarrow (xRy \Leftrightarrow xR'y);$$
$$(\mathbf{N})\ \forall \overrightarrow{R}, \overrightarrow{R'} \in D^F, \forall x,\ y, z, w \in X,\ \left[\left(xP_i y \Leftrightarrow zP'_i w\right)\&\left(yI_i x \Leftrightarrow wI'_i z\right), \forall i \in N\right] \Rightarrow (xRy \Leftrightarrow zR'w).$$

In terms of Fishburn's distinction, (D), (I), and (N) are interprofile, while (WP) is intraprofile. When the domain reduces to a single profile $\overrightarrow{R}$, interprofile conditions lose their force but do not necessarily collapse. Here, only (I) becomes vacuous, and (D) and (N) lose force but remain binding. The former becomes *single-profile dictatorship* and the latter *single-profile neutrality*:

$(\mathbf{D}')\ \exists i : \forall x, y \in X, xP_i y \Rightarrow xPy$
$(\mathbf{N}')\ \forall x, y, z, w \in X, [(xP_i y \Leftrightarrow zP_i w)\&(yI_i x \Leftrightarrow wI_i z), \forall i \in N] \Rightarrow (xRy \Leftrightarrow zRw).$

The following, less familiar "triple" condition must also be introduced. It is reserved for the case in which $\mathcal{D}^F = \left\{\overrightarrow{R}\right\}$. In this statement, {*a, b, c*} is any set of three distinct elements.

(T) For any logically possible preference profile over {*a, b, c*}, there are distinct alternatives $x, y, z \in X$ such that $\overrightarrow{R}\Big|_{\{x,y,z,\ldots\}}$ reproduces this profile with x instead of a, y instead of b, z instead of c.[2]

The controversy of the 1970's involved Arrow's theorem as well as – more actively – its newly discovered single-profile analogue. We restate them both, adopting a doctored version of the latter, which is Pollak's but implicitly covers the others.

Theorem 1 (Arrow 1963) If F is defined on $\mathcal{D}^F = (\mathcal{P}_X)^n$, *then* (WP) *and* (I) *imply* (D).

Theorem 2 (Kemp and Ng 1976; Parks 1976; Pollak 1979). If F is defined on $\mathcal{D}^F = \left\{\overrightarrow{R}\right\}$ *and* (T) *is satisfied, then* (WP) *and* (N′) *imply* (D′).[3]

As such, these abstract theorems were not applicable to welfare economics, since the economic domains needed for the latter involve making special assumptions on the structure of X and the orderings allowed to enter $\mathcal{D}^F$. However, the theorems were used as convenient proxies for the relevant results —an expository device that we retain in this paper. It is sufficient to check once and for all that Theorems 1 and 2 carry through to economic domains.[4]

An important consequence of the single-profile theorem was that *the neutrality condition rose to prominence in social choice theory*. The theorists of the 1970's were led to dig into Arrow's proof in order to find an assumption to start with, and what they found was neutrality in the form of (N'). The second edition of

[2] This statement is borrowed from Pollak (1979, p. 76). Parks (1976) has a different formulation in terms of utility values.

[3] As a matter of attribution, it appears that Kemp and Ng on the one hand, Parks on the other worked independently and simultaneously, while Pollak was stimulated by the former's work and did not initially know of the latter's. The argument in Kemp and Ng involved an irrelevant topological framework, and had many loose ends. Pollak extracted the neat Theorem 2 from this jumble. Parks had already reached Theorem 2, but his paper is terse, and does not draw the connections with Theorem 1 in the same useful way as Pollak does.

[4] Arrow (1963, ch. VI) initiated the study of economic domains in social choice theory, but full proofs had to await the 1980's. For a survey, see Le Breton and Weymark (1996).

Arrow's book (1963, p. 100–102) had discussed neutrality in connection with May's (1952) classic axiomatization of the majority rule, but perhaps surprisingly, had not mentioned that this condition underlay the impossibility theorem. Sen (1970, ch. 5–5*) discussed it again primarily in connection with the majority rule, but contrary to Arrow, provided a clear glimpse of its use to derive impossibilities.[5] At any rate, it was really Theorem 2 which directed the specialists' attention towards neutrality.

Samuelson was quick to point out the major defect of this theorem. In his reply to Kemp and Ng, he complained that the neutrality assumption was much too close to the dictatorship conclusion:

> "I believe that Plato, Aristotle and Hobbes would be interested in Arrow's "Impossibility Theorem" on ideal democracy. I doubt that such ethical philosophers as Bentham, Kant, Sidgwick or Rawls would be jarred by the Kemp–Ng "impossibility theorem", once they perceive how it transparently follows from the gratuitous Axiom 3". (1977, p. 81, our emphasis).

Here is a two-person geometric argument, based on Samuelson (1977, p. 87), which makes it clear that (D′) readily follows from (WP) once (N′) is granted.[6] Suppose that the alternatives can be represented in the two-dimensional utility space $\mathbb{R}^2_+$ and take an interior point x, as in Fig. 1. By (WP) the social ordering ranks any point north-east of x, such as a, above x and any point south-west of x, such as b, below x. Now, (N′) entails that all points south-east of x, such as c, d, are ranked in the same way relative to x, which means that they are all above x or all below x. (If c, d were indifferent to x, by transitivity they would be indifferent to each other, and this would contradict (WP).) Similarly, by another application of (N′), all points north-west of x, such as e, are either above x or below x, depending on whether x is above or below c, d. In conclusion, the social ordering follows either 1's or 2's strict preference. To turn this diagrammatic argument into a rigorous proof that dictatorship follows from neutrality, it would be enough to justify in terms of the basic conditions the initial claim that the alternatives can be represented in the utility space.

The mathematical simplicity of the single-profile variant reflects negatively on the merit of the multi-profile theorem. The standard proof, as is still tought today, involves two lemmas, one to show that an individual who is "almost decisive" on a pair of alternatives is "decisive" on any pair, and the other to show that there is indeed an individual who is "almost decisive" on some pair (see, e.g., Sen 1970, ch. 3*, for formal definitions and proofs). This sequence loses some of its mystery when it is realized that the first lemma amounts to *making a neutrality statement*. As Sen puts it aptly:

> "If we object to the characteristic of neutrality and the axioms that lead to it, then we question the applicability not merely of single-profile impossibility results, but also of multi-profile impossibility theorems. There isn't much of a

[5] Turning May's characterization into an Arrovian theorem, Sen (1970, p. 73) proves that no F can satisfy the universal domain condition, neutrality, *positive responsiveness* (an interprofile condition which is related to (WP)), and *anonymity* (which is a strong denial of (D)). This may be the first impossibility theorem involving neutrality in its premisses.

[6] Blackorby et al. (1984) diagrammatic analysis of social choice uses a similar argument. See also Blackorby et al. (1990).

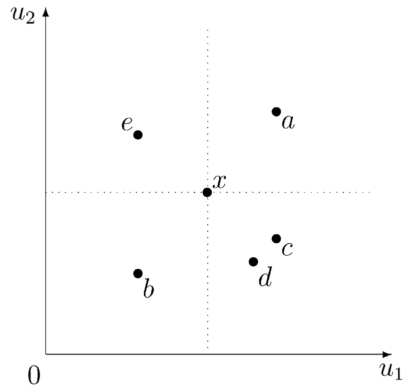

Fig. 1 Diagram of alternatives represented in the two-dimensional utility R^2_+

> line to draw between them from this particular point of view" (1977, in 1982 p. 256).

Of course, the greater elegance of Theorem 1, compared with Theorem 2, is that it involves an element of neutrality merely as a logical step; it does not bluntly assume it. This is why Samuelson never downgraded Arrow's theorem to the status of a triviality, which he did crushingly for Kemp and Ng's.

Whether it is substantially distinct or not from Theorem 1, Theorem 2 had a lasting effect on the relations between social choice theory and welfare economics. It helped deflate a seemingly persuasive point that had been raised earlier against applying Arrow's impossibility theorem to the latter field. The "Bergson–Samuelson welfare function", it was said, is a single-profile construal, and this automatically disqualifies the theorem from being applicable. This line of defence originates in Little (1952), but it was Samuelson who pursued it most actively:

> "My exposition is well designed to bring out the difference between a Bergson Social Welfare Function and an Arrow Constitutional Function (or so-called "social welfare function"). For Bergson, one and only one of the 2,197 possible patterns of individuals' orderings is needed. It could be any one, but it is only one. From it (not from each of them all) comes a social ordering". (1967, p. 48–49).

In other words, Arrow's attempted *reductio* can only be a non-starter since it needs a universal, or at least a large domain of profiles, among its assumptions. Samuelson's heavy-handed argument is the likely reason why single-profile impossibility theorems flourished in the 1970's. Kemp and Ng, Parks, and Pollak, all motivate their work by citing Samuelson's (1967) paper.

However, a rebuttal was possible without going to the pain of proving Theorem 2. The Bergson function is defined for any admissible profile; accordingly, it is a multi-profile notion. Turning Samuelson's comment on its head, Sen argued as follows:

> "Given the fact that Samuelson is admitting only one preference n-tuple, but not restricting that one in any way ("it could be any one, but it is only one",

p. 49), Samuelson is in fact combining unrestricted domain with the absence of any inter-profile condition" (1977, in 1982 p. 252).

Similarly, Arrow:

"After all, he wants any society to have a social welfare function. Hence, treating the function as itself defined by some characteristics of the society, including its individual preference orderings but possibly also using other information, is not by itself a very large deviation from the Bergson–Samuelson program…. What gives the discussion its bite is the assumption that there are or should be some consistency conditions between the social orderings associated with different societies" (1983, p. 26).

In brief, the disagreement with welfare economists has to do with *inter-profile* conditions in a *multi-profile* framework. They should not have suggested that $\mathcal{D}^F = \left\{ \overrightarrow{R} \right\}$, but rather that whatever $\mathcal{D}^F$ may be, the "Bergson–Samuelson social welfare function" for a given profile $\overrightarrow{R}$ is not constrained by its values on other profiles $\overrightarrow{R'}$. Once it is realized that the framework is not a significant issue, it is simpler, as Sen suggests, to take the maximum domain $\mathcal{D}^F = (\mathcal{P}_X)^n$.[7]

Neither Little nor Samuelson ever thought through the profile issue. The former had the excuse of writing before Fishburn's (1973) distinction and Sen's (1977) related clarifications had become available.[8] The latter maintained his views in print even beyond this late stage. Not only was he persistently confused about what mattered and what did not in the profile issue, but at some point he appeared to be almost inconsistent. While rejecting an intra-profile condition—Kemp and Ng's neutrality—he praised the underlying inter-profile condition—Arrow's independence—as if they were significantly different (see 1967, p. 47; 1981, p. 262). We will show in Section 5 how this inconsistency contributed to blur his message during the controversy of the 1970's. The single-profile theorists are not beyond reproach either. By making a big deal of their single-profile impossibility, as if it were a result truly different from Arrow's, they deflected the discussion from its proper course: it would have developed more fruitfully if they had made (N), and eventually (I), the focus of analysis. Their own condition (N′) was confusing. It had the virtue of revealing the role of (N) behind Arrow's proof and argument against Bergson, but being stated just for one profile, it reinforced the welfare economists' prejudiced view that the number of profiles was an issue between them and their opponents.

3 Bergson–Samuelson social welfare functions

This section attempts to clear up the misunderstandings about the "Bergson–Samuelson social welfare function" that plagued the controversy, most of which are

[7] It is not always the case that the maximal domain should be selected when there is no reason to select any particular domain. For a discussion in the context of choice functions, see Sen (1971, 1982, p. 48–49) and Mongin (2000, p. 92–93).

[8] Sadly, Little's (1999, p. 17–18) late retraction does not show that he has taken advantage of the intervening clarifications.

still with us today. Returning to the founders' work, we will show that it is *ordinalist*, in the classic sense of requiring only non-comparable ordinal utility functions, and nonetheless *comparative*, in the sense of requiring comparisons that are *not* of utility functions or values. A novelty of the present paper, this interpretation clashes with the textbook interpretation, which contemplates only two possibilities: either Bergson–Samuelson functions are ordinalist and do not involve any interpersonal comparisons, or they make interpersonal comparisons of utility.

Take the set of alternatives to be a set of economic allocations X^e, and for any allocation $x \in X^e$, denote by x_i the subvector made out of the i-relative components. Allocation x can be rewritten as a grand vector $x=(x_1,\ldots,x_n)$. A *Bergson–Samuelson function* (henceforth BSF) is any function E associating numerical ("welfare") values with allocations:

$$E : X^e \to \mathbb{R}, x \mapsto E(x).$$

Samuelson stresses that the function "need only be ordinally defined" (1947, p. 221), i.e., that E can be replaced by any ordinal transform of it. This condition is implicit in Bergson (1938), whose purpose in introducing E ("the Economic Welfare Function") was to facilitate the derivation of the classic marginal equalities for a general optimum. The first-order conditions are of course the same whether E or any ordinal transform is maximized; so Bergson's function is ordinal even if he does not stop at mentioning it.

By the middle of his paper, Bergson (1938, p. 318–319) specializes E to the subclass of those functions which satisfy the "Fundamental Value Propositions of Individual Preference". In Samuelson, the same set of conditions is labelled "individualism", and stated as follows:

> "If any movement leaves an individual on the same indifference curve, then the social welfare function is unchanged, and similarly for an increase or a decrease" (1947, p. 223; see also 1981, p. 230).

This is simply the Pareto principle, defined in such a way as to include the Pareto-indifference condition. The latter is responsible for the convenient expression of a BSF in terms of individual utility values. Suppose that the individuals' utility functions are $U=(U_1,\ldots,U_n)$. The function E satisfies *Pareto-indifference*, i.e.,

$$\forall x, x' \in X, (U_i(x) = U_i(x'), \forall i \in N) \Rightarrow E(x) = E(x'),$$

if and only if there exists a function $W : \mathbb{R}^n \to \mathbb{R}$ such that E factors out in terms of U and W, i.e.,

$$\forall x \in X, E(x) = W \circ U(x).$$

Once they have assumed the Pareto principle, Bergson and Samuelson usually work on $W \circ U$ representations. However, they have made it clear that W and U are used for convenience, these notions being derivative to the initial E.

Samuelson spells out what pairs of W and U are admissible, given E:

> "There are an infinity of equally good indicators... which can be used. Thus, if one of these is written as $E= W(U_1,\ldots,U_n)$, and if we were to change from one

set of cardinal indexes of individual utility to another set $(U_1',\ldots,U_n')$, we should simply change the form of the function W so as to leave all social decisions invariant". (1947, p. 228; adapted notation).

This explanation encapsulates the ordinalist claim that E depends only on the individuals' preferences R_i.[9] This time more explicit than Samuelson, Bergson says that U_i, labelled by him "indifference function",

> "expresses the loci of combinations of commodities consumed and work performed which are indifferent to the ith individual" (1938, p. 319) and that the U_i
>
> "are understood to be only ordinal indicators... Hence, for any given [vector] of ordinal indicators, there is a corresponding W" (1966, p. 81; adapted notation).

Formally, any $(U_1',\ldots,U_n')$ that also represents[10] $(R_1,\ldots,R_n)$ can replace $(U_1,\ldots, U_n)$, and when this change is made, W must be changed into W' in order to represent the same E. The new representation W is defined implicitly by $E= W\circ U=W'\circ U'$, and explicitly by

$$W'(\xi_1,\ldots,\xi_n) = W\left(\psi_1^{-1}(\xi_1),\ldots,\psi_n^{-1}(\xi_n)\right),$$

where ψ_i is the strictly increasing transformation connecting U_i' and U_i.[11] Actually, since E is ordinal, this set of (U',W') may be enlarged by requiring $W'\circ U'$ to be not equal, but ordinally equivalent, to $W\circ U$.

Throughout their work, Bergson and Samuelson make the standard economic assumptions that U_i only depends on, and strictly increases with, the i-relative components. Given the *strict Pareto condition*, i.e.,

$$\forall x, x' \in X,\ \left(U_i(x) \geq U_i(x'), \forall i \ \&\ \exists j, U_j(x) > U_j(x')\right) \Rightarrow E(x) > E(x')$$

the final form of a BSF is

$$E(x) = W(U_1(x_1),\ldots,U_n(x_n)),$$

where W is strictly increasing in each $U_i(x_i)$. This formula is often assumed from the beginning in textbooks.

Disappointingly, the two authoritative sources, i.e., Bergson's 1938 article and Samuelson's *Foundations*, provide no proof that the above definition of a BSF eschews unacceptable mathematical restrictions. It is no less than thirty years after publishing his book, when he was pushed to the wall by the social choice

[9] See also Samuelson's rebuttal of Lange and Fisher: "Lange, Fisher, and others have contended that measurable utility, while superfluous from the standpoint of positivistic behavioristic description, is necessary for the purpose of a normative science of welfare economics... It is well to point out that this is not at all necessary" (1947, p. 173).

[10] A function $U_i \in \mathbb{R}^X$ *represents* R_i on X if, for all $x, y \in X$, xR_iy iff $U_i(x)\geq U_i(y)$. The vector-valued function $U = (U_i)_{i\in N}$ is said to represent $(R_i)_{i\in N}$ when U_i represents R_i for each i.

[11] This formula can be found in Arrow's comment of Samuelson (1983, p. 22).

theorists, that Samuelson finally delivered an example of a well-behaved BSF (1977, p. 84–86; the same example is taken up in 1981, p. 234). Even then, he did not go beyond a coarse diagrammatic sketch; so it will help if we push the formalization one step further. Consider the problem of sharing a bundle of goods $\Omega \in \mathbb{R}^{\ell}_{++}$ among n individuals. This defines a subset of feasible allocations

$$X = \left\{(x_1, \ldots, x_n) \in \mathbb{R}^{n\ell}_{+} \middle| x_1 + \ldots + x_n \leq \Omega\right\}.$$

Each individual i has an ordering R_i of the x_i that satisfies the standard economic assumptions. For any x, define:

$$\lambda(x) = \min_i \quad \max\left\{\lambda_i \in \mathbb{R} | x_i R_i \lambda_i \Omega\right\}$$

and

$$E^*(x) = \Omega_1 \lambda(x),$$

where Ω_1 is the quantity of good 1 in the reference bundle (in order to facilitate graphical illustration below, we take E^* to be measured in units of this good). Informally, E^* evaluates an allocation x by paying attention only to the individual(s) whose bundle(s) in this allocation is (are) the least satisfying of all. Each individual's bundle is assessed by comparing it with an equivalent bundle along the common reference direction Ω. As the formula makes clear, E^* depends only on the R_i. It satisfies the weak Pareto and Pareto-indifference conditions; hence, one can find U and a weakly increasing W such that E^* and $W \circ U$ are equal or, as it is only relevant, ordinally equivalent. There are infinitely many such possibilities, among which for instance:

$$E^*(x) = W^*\left(U_1^*(x_1), \ldots, U_n^*(x_n)\right) = \min_i U_i^*(x_i),$$
$$\text{where} \quad U_i^*(x_i) = \Omega_1 \max\{\lambda_i \in \mathbb{R} | x_i R_i \lambda_i \Omega\},$$

or

$$E^*(x) = W'\left(U_1'(x_1), \ldots, U_n'(x_n)\right) = \Omega_1 \min_i \sqrt{U_i'(x_i)},$$
$$U_i'(x_i) = \max\left\{\lambda_i \in \mathbb{R} | x_i R_i \sqrt{\lambda_i} \Omega\right\}.$$

The function E^* is non-dictatorial and even *anonymous*, that is to say, if the names of the individuals were permuted, the social ranking of allocations would not be altered. Figure 2 illustrates the computation of U_i^* for the case of two goods, and the social indifference curves for W^* for the case of two individuals.

This figure is similar to Samuelson's Fig. 1 in his 1977 paper, which also describes a two-individual, two-good economy in terms of indifference curves drawn in the commodity space (for the individuals) and a social indifference curve drawn in the utility space (for the "ethical observer"). The shape of the latter indifference curve suggests that Samuelson would like to have continuous trade-offs between individual utilities, and this is confirmed in the text. By contrast, our E^* is represented by angular social indifference curves. We chose this function partly because it leads to an explicit analytical definition, partly because it connects nicely with the theory of fair allocation. This example is indeed

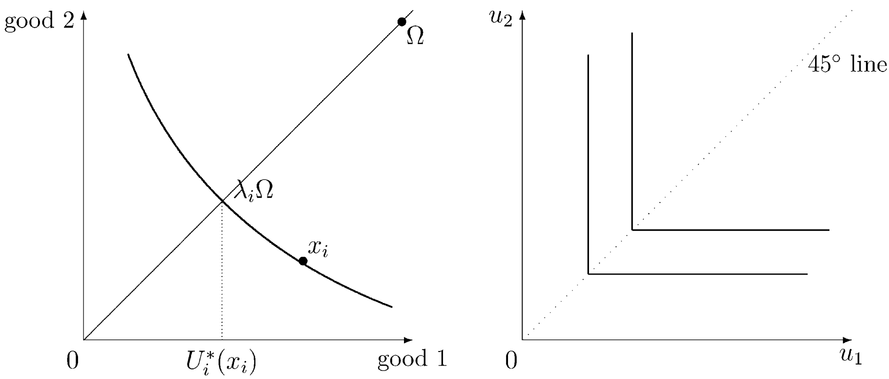

Fig. 2 Illustration of the computation of U_i* for the case of two goods and the social indifference curves of W* for the case of two individuals

suggested by Pazner (1979, p. 169), conveying the influence of his earlier work with Schmeidler on egalitarian-equivalent allocations. According to Pazner and Schmeidler (1978), an allocation x is said to be *egalitarian-equivalent* if there is a commodity bundle x_0, identical for each individual i, that would leave each i with exactly the same satisfaction as he gets from his bundle x_i in x. Now, if E* is maximized over the set of feasible x, the solution x* will be both egalitarian-equivalent and Pareto-optimal.[12]

It has just been shown that BSF *could*, as a matter of logic, eschew interpersonal comparisons of utility. Obviously, this does not imply that they *should*, as a normative or "ethical" matter. Bergson and Samuelson were categorical on the former issue. By contrast, what they have to say about the latter is either vague or disconcerting. This step in our reconstruction deserves textual evidence that we will now provide.

Bergson (1938) carefully distinguishes his own construction of a BSF based on "indifference functions" (which he denotes by S_i instead of U_i), from "the Cambridge analysis" of Marshall and Pigou, in which "the welfare of the community, stated symbolically, is an aggregate of the form

$$\overline{E}(x) = \sum_{i \in N} U_i(x_i).$$

In this expression U_i is some function of the indifference function, S_i, and measures the satisfactions derived by the i^{th} individual" (p. 324; notation adapted). Bergson does not accept "the Cambridge analysis", but the only argument he makes is that the notion of social welfare cannot be purely factual. An emphasis on value judgments would be compatible with the possibility that $\overline{E}$ may be the appropriate "ethical" BSF after all. While Bergson does list the kinds of value judgments that are needed for the elaboration of BSF (1938, p. 327), he says very little on the substance of such value judgments. His 1966 paper shows a continuing hostility towards the "Cambridge conception" and is more explicit about his belief

[12] Compare this statement with Pazner and Schmeidler (1978, p. 679).

that empirical interpersonal comparisons of utility are impossible—he still does not feel the need for a critical *ethical* argument. The novelty in this retrospective paper is Bergson's notion of a "rule of equity", which he meant to be an alternative to utility comparisons:

> "The optimum income distribution… is not determined by an empirical comparison of marginal social welfare per dollar among different households. Rather it is determined by the rule of equity, which itself defines social welfare in the sphere of income distribution" (p. 66).

Despite this interesting hint, Bergson ended up with the exceedingly cautious claim that ordinality "apparently is quite sufficient" for his criterion (p. 67).[13] For related evidence of Samuelson's vacillations, see the following passage from the *Foundations*:

> "Our ethical observer need only decide then what his preferences are as between the given levels of satisfactions of different individuals" (1947, p. 228)

Even more strikingly, under the ambiguous heading "Interpersonal Optimal Conditions", Samuelson discusses the continuum of Pareto optima along the contract curve in the following terms:

> "An infinity of such positions exist ranging from a situation in which all of the advantage is enjoyed by one individual, through some sort of compromise position, to one in which another individual has all the advantage. Without a well-defined W function, i.e., without interpersonal comparisons *of utility*, it is impossible to decide which of these points is best". (1947, p. 244, our emphasis).

The safe conclusion to draw is that Bergson and Samuelson have worked with two definitions of a BSF at a time. Their "social welfare function" is sometimes the purely ordinalist E that was exemplified above by E*, sometimes a nondescript

$$\widetilde{E}(x) = \widetilde{W}\left(\widetilde{U}_1(x_1), \ldots, \widetilde{U}_n(x_n)\right),$$

which covers all possible cases of cardinality and ordinality, comparability and non-comparability, depending on what is being assumed on $\widetilde{U}$ and $\widetilde{W}$. In the secondary literature, $\widetilde{E}$ quickly ousted E. This seems already to happen in Graaff's classic text Theoretical Welfare Economics (1957, p. 37).[14] Here is a more recent and absolutely unequivocal example:

> "We are assuming full comparability of utility, so that these levels are themselves expressed in units that can be meaningfully compared across individuals. Although it is clearly of interest to compare the different social choices that result from adopting different cardinalizations..., the social choice involved is not, in general, invariant with respect to such changes. *It is thus not*

[13] We read no further advance in Bergson's (1976) latest retrospective paper.

[14] This text was regarded by Samuelson as an authoritative restatement of the new welfare economics; see his preface.

> *very useful to talk about Bergson–Samuelson welfare functions being defined over ordinal utilities*." (Deaton and Muellbauer 1980, p. 222, our emphasis)

To compound the difficulty in interpretation, Bergson and Samuelson repeatedly emphasize that interpersonal comparisons are necessary to define social welfare.

> "In welfare economics objection is made not to interpersonal comparisons but to the contention that these comparisons can be made without the introduction of ethical premisses" (Bergson 1954, p. 245). "There is no avoiding such interpersonal judgments if we are to be provided with a complete ethical ordering of all the states of the world". (Samuelson 1981, p. 234).

This seems in apparent contradiction with Bergon and Samuelson's claim that ordinal non-comparable preferences are a sufficient informational basis. But the contradiction vanishes when one understands that they might have in mind comparisons of wealth, economic positions, or indifference curves, not comparisons of utility figures. In other words, when mentioning the necessity of interpersonal comparisons, Bergson and Samuelson simply refer to the obvious point that without distributional judgments a complete ordering of social states is impossible. There is an interesting evolution, in their writings, with respect to the kind of comparisons involved in such distributional judgments. In 1938, Bergson introduces two "value propositions" which are directly based on the comparison of individual wealth at fixed prices (p. 321, 332). In 1947 Samuelson has a cryptic formulation about comparing consumption bundles:

> "Assuming that Welfare Economics involves comparisons between individuals, it is sufficient that explicit welfare judgments be made such that we are able to relate *ordinally* all possible combinations of goods and services consumed by each and every individual". (p. 173).

The idea of comparing indifference curves appears in 1954 when Bergson mentions the possibility

> "to pair by separate ethical premises all the indifference curves of each household with all those of every other one" (p. 245).

In 1977, Samuelson shows that it is sufficient to select one point on each indifference curve, such as point $\lambda_i \Omega$ on Fig. 2, and to compare these selections. This is what our E^* does indeed.

4 A social ordering functional framework for the controversy of the 1970's

This section extends the mathematical framework from social ordering *functions* to social ordering *functionals*, a concept which will provide a natural bridge between welfare economics and social choice theory, and will enable us to further clarify the matter of interpersonal comparisons. In particular, we reformulate and extend arguments that appeared in Blackorby et al. (1990), Fleurbaey (2003) and Fleurbaey and Hammond (2004).

As in Section 2, there are n individuals and at least three distinct elements in the set of alternatives X. A *social ordering functional* f is defined to be a mapping from a set $\mathcal{D}^f$ of individual utility function profiles $U=(U_1,\ldots,U_n)$ to the set $\mathcal{P}_X$ of orderings on X. The case of a maximal domain $\mathcal{D}^f = \left(\mathbb{R}^X\right)^n$ corresponds to Sen's (1970, 1986) and many other theorists' "social welfare functionals". The letters R, P, and I will be used to denote the ordering $f(U)$, and its asymmetric and symmetric parts, respectively. The conditions of Section 2 may be reformulated in terms of utilities:

$$(\mathbf{D})_f \exists i : \forall U \in \mathcal{D}^f, \forall x, y \in X, U_i(x) > U_i(y) \Rightarrow xPy;$$

$$(\mathbf{WP})_f \forall U \in \mathcal{D}^f, \forall x, y \in X, (U_i(x) > U_i(y), \forall i \in N) \Rightarrow xPy;$$

$$(\mathbf{I})_f \forall U, U' \in \mathcal{D}^f, \forall x, y \in X,$$

$$\left[\begin{array}{c} \left(U_i(x) > U_i(y) \Leftrightarrow U_i'(x) > U_i'(y)\right) \& \\ \left(U_i(x) = U_i(y) \Leftrightarrow U_i'(x) = U_i'(y)\right), \forall i \in N \end{array}\right] \Rightarrow (xRy \Leftrightarrow xR'y);$$

$$(\mathbf{N})_f \forall U, U' \in \mathcal{D}^f, \forall x, y, z, w \in X,$$

$$\left[\begin{array}{c} \left(U_i(x) > U_i(y) \Leftrightarrow U_i'(z) > U_i'(w)\right) \& \\ \left(U_i(x) = U_i(y) \Leftrightarrow U_i'(z) = U_i'(w)\right), \forall i \in N \end{array}\right] \Rightarrow (xRy \Leftrightarrow zR'w).$$

We briefly introduce some standard material of invariance. For a vector of increasing transformations $\varphi=(\varphi_1,\ldots,\varphi_n)$, the abusive notation $\varphi \circ U$ abridges $(\varphi_1 \circ U_1,\ldots,\varphi_n \circ U_n)$. A domain $\mathcal{D}^f$, and by extension the social ordering functional f, will be said to be *ordinally complete* if $\varphi \circ U \in D^f$ for all $U \in \mathcal{D}^f$ and all vectors of increasing transformations φ.

An important special case occurs when $\mathcal{D}^f$ is the set of all increasing transformations of a single profile $\overline{U} = \left(\overline{U}_1, \ldots, \overline{U}_n\right)$; denote it by $D^f_{\overline{U}}$. Now, define an ordinally complete f to be *ordinally non-comparable* if:

$$(\mathbf{ON})_f \forall U \in \mathcal{D}^f, \forall \varphi, f(\varphi \circ U) = f(U).$$

For the small domains $\mathcal{D}^f_{\overline{U}}$, we will need a "triple" condition:

$(\mathrm{T})_f$ For any logically possible preference profile over $\{a, b, c\}$, there are distinct alternatives $x, y, z \in X$ such that $\overline{U}\big|_{\{x,y,z\}}$ represents this profile with x instead of a, y instead of b, z instead of c.

The f and F formalisms are easily connected with each other. For every F on $\mathcal{D}^F$, define its *canonical associate* f by the conditions that:

(i) $U \in \mathcal{D}^f$ iff U represents some $\overrightarrow{R} \in D^F$, and
(ii) for all $\overrightarrow{R} \in \mathcal{D}^F$, if $U \in \mathcal{D}^f$ represents $\overrightarrow{R}$, then for all $x, y \in X$, $xF\left(\overrightarrow{R}\right)y \Leftrightarrow xf(U)y$.

Clearly, if F satisfies (D), (WP), (I) or (N), its canonical associate f satisfies the corresponding f-condition, and conversely. When this translation device is applied, the following variants of Theorems 1 and 2 arise:

Theorem 1$_f$ *If* f *is defined on* $\mathcal{D}^f = \left(\mathbb{R}^X\right)^n$, *then* (I)$_f$ *and* (WP)$_f$ *imply* (D)$_f$.

Theorem 2$_f$ *If* f *is defined on* $\mathcal{D}^f_{\overline{U}}$ *for some* $\overline{U}$ *and* (T)$_f$ *is satisfied, then* (N)$_f$ *and* (WP)$_f$ *imply* (D)$_f$.[15]

It is customary to restate the impossibility theorems in a slightly different form. Define the *utility-independence* and *utility-neutrality* conditions.

$$(\mathbf{UI})_f\ \forall U, U' \in \mathcal{D}^f, \forall x, y \in X, \left[U(x) = U'(x)\ \&\ U(y) = U'(y)\right] \Rightarrow (xRy \Leftrightarrow xR'y);$$

$$(\mathbf{UN})_f \forall U, U' \in \mathcal{D}^f, \forall x, y, z, w \in X, \left[U(x) = U'(z)\ \&\ U(y) = U'(w)\right] \Rightarrow (xRy \Leftrightarrow zR'w).$$

These are the "independence of irrelevant alternatives" and "extended neutrality" conditions of the theory of "social welfare functionals".[16] This theory employs them in the following variants of Theorems 1 and 2:

Theorem 3 *If* f *is defined on* $\mathcal{D}^f = \left(\mathbb{R}^X\right)^n$, *then* (ON)$_f$, (UI)$_f$ *and* (WP)$_f$ *imply* (D)$_f$.

Theorem 4 *If* f *is defined on* $\mathcal{D}^f_{\overline{U}}$ *for some* $\overline{U}$ *and* (T)$_f$ *is satisfied, then* (ON)$_f$, (UN)$_f$ *and* (WP)$_f$ *imply* (D)$_f$.

The classic advantage of the f framework is that it makes it possible to formalize interpersonal comparisons of utility, or the lack thereof. It has the further, less obvious advantage of permitting a refined analysis of independence and neutrality. Some of the clarifications brought by this paper depend on logically decomposing the ordinalist conditions (I)$_f$ and (N)$_f$, as follows.

[15] The proof of Theorems 1$_f$ and 2$_f$ from Theorems 1 and 2 uses the fact that the latter remain true when D^F is restricted to the set of those orderings on X which can be represented by a utility function. Fishburn (1970, p. 27–28) states that there exists U_i representing R_i if and only if P_i is order-dense on the quotient of X by the indifference relation I_i.

[16] Interestingly, Hammond (1976) derives (UN)$_f$ from (UI)$_f$ and a weaker form of neutrality. He proves a variant of Theorem 4 stated below.

Proposition 1 If f is ordinally complete, it satisfies $(\mathrm{I})_f$ *iff it satisfies* $(\mathrm{ON})_f$ *and* $(\mathrm{UI})_f$ *and it satisfies* $(\mathrm{N})_f$ *iff it satisfies* $(\mathrm{ON})_f$ *and* $(\mathrm{UN})_f$*. If* $D^f = D^{f}_{\overline{U}}$ *then* $(\mathrm{I})_f$ *and* $(\mathrm{ON})_f$ *are equivalent.*

Proof. That $(\mathrm{I})_f$ implies $(\mathrm{UI})_f$ is clear. That $(\mathrm{I})_f$ implies $(\mathrm{ON})_f$ for an ordinally complete f can be seen by applying $(\mathrm{I})_f$ to $U'=\varphi\circ U$, which by assumption belongs to the domain.

Conversely, fix U, U', x, y and assume the antecedent of $(\mathrm{I})_f$. For each i, take φ_i to be an affine positive transformation such that $U_i'(x)=\varphi_i(U_i(x))$ and $U_i'(y)=\varphi_i(U_i(y))$. (There is always one such transformation, and it is unique iff $(U_i(x), U_i(y))$ and $(U_i'(x), U_i'(y))$ are affinely independent vectors.) Apply $(\mathrm{ON})_f$. Then,

$$\left(xR_U y \Leftrightarrow xR_{\varphi\circ U} y\right).$$

Applying $(\mathrm{UI})_f$, we can replace $\varphi\circ U$ by U', and thus get the conclusion of $(\mathrm{I})_f$.

For the second part, $(\mathrm{N})_f$ obviously implies $(\mathrm{UN})_f$. It also implies $(\mathrm{I})_f$, and therefore $(\mathrm{ON})_f$. The proof of the converse is the same as in the first part, except that $(\mathrm{UN})_f$ is applied instead of $(\mathrm{UI})_f$.

If $\mathcal{D}^f = \mathcal{D}^{f}_{\overline{U}}$, then $(\mathrm{ON})_f$ is satisfied if and only if f is constant. When f is constant, $(\mathrm{I})_f$ is satisfied. □

For further reference, we note that the equivalence between $(\mathrm{I})_f$ and $(\mathrm{ON})_f$ does not generally hold for domains larger than $D^{f}_{\overline{U}}$. A counter-example is given at the end of the section.

The next proposition involves a special domain restriction. A domain $\mathcal{D}^f$, and by extension the social ordering functional f, will be said to be *interpolating* if for all $x, y, x', y' \in X$, and $U, U' \in \mathcal{D}^f$ such that $U(x)=U'(x')$, $U(y)=U'(y')$, there exist $U^1, U^2 \in \mathcal{D}^f$ and $x'', y'' \in X$ such that

$$\begin{aligned} U(x) &= U^1(x) = U^1(x'') = U^2(x'') = U^2(x'), \\ U(y) &= U^1(y) = U^1(y'') = U^2(y'') = U^2(y'). \end{aligned}$$

We also need the Pareto-Indifference condition:

$$(\mathbf{PI})_f \ \forall U \in D^f, \forall x, y \in X, (U_i(x) = U_i(y), \forall i \in N) \Rightarrow xIy.$$

Proposition 2 If f is interpolating, it satisfies $(\mathrm{UN})_f$ *iff it satisfies* $(\mathrm{PI})_f$ *and* $(\mathrm{UI})_f$.

The statement is usually made for a maximal domain, but inspection of its proof (e.g., d'Aspremont 1985, p. 34) shows that this is unnecessary. The interpolation condition is all that is needed.[17]

Except for the trivial case of constant utility functions U_i, the interpolation condition is not satisfied when $\mathcal{D}^f = \mathcal{D}^{f}_{\overline{U}}$. For such a domain, the implication from $(\mathrm{PI})_f$ and $(\mathrm{UI})_f$ to $(\mathrm{UN})_f$ does not generally hold. This may be checked by selecting a

[17] We choose this particular condition because it is simple and can be satisfied on economic domains. Bordes et al. (1996) investigate more basic conditions, and so does Weymark (1998).

suitable constant f satisfying $(\mathrm{PI})_f$ and violating $(\mathrm{UN})_f$ (e.g., the function f^* derived from E^*, as defined below, and restricted to this small domain). Since any constant f satisfies $(\mathrm{UI})_f$, the implication just mentioned fails. What about richer domains than $\mathcal{D}^f_U$? The following counter-example shows that when f is ordinally complete but not interpolating, the implication does not generally hold either.

Example 1 Take $X=\{a, b, c, d\}$, $N=\{1, 2\}$, $\mathcal{D}^f = \{U, U'\}$, *with* U, U' *defined as follows*:

$$U_1(a) = 1 = U_1(d), U_1(b) = 0, U_1(c) = -1,$$
$$U_2(a) = -1 = U_2(d), U_2(b) = 0, U_2(c) = 1,$$
$$U'_1 = U_2; U'_2 = U_1.$$

Let $f(U)=U_1$, $f(U')=U'_2$. *The Pareto conditions, including* $(\mathrm{PI})_f$, *trivially hold. So does* $(\mathrm{UI})_f$ *since* f *is constant. But* $(\mathrm{UN})_f$ *is violated*: $U_1(a)=1=U'_1(c)$, $U_2(a)=-1=U'_2(c)$, $U_1(b)=0=U'_1(b)$, $U_2(b)=0=U'_2(b)$, $af(U)b$, *but not* $cf(U')b$. *The conclusion is unchanged if* $\mathcal{D}^f$ *contains all ordinal transforms of* U, U'.

Corollary 1 If f *is both ordinally complete and interpolating, it satisfies* $(\mathrm{N})_f$ *iff it satisfies* $(\mathrm{UN})_f$ *and* $(\mathrm{ON})_f$ *iff it satisfies* $(\mathrm{PI})_f$, $(\mathrm{UI})_f$ *and* $(\mathrm{ON})_f$ *iff it satisfies* $(\mathrm{PI})_f$ *and* $(\mathrm{I})_f$.

Proof. From Propositions 1 and 2. □

The following figure summarizes the results, first for the general case of a domain $\mathcal{D}^f$, and then for the special case of the domain $\mathcal{D}^f_U$, which is ordinally complete by definition, but is never interpolating.

This brief analysis clarifies the informal comments made in Section 2 about the impossibility theorems. First, it formalizes Sen's point that the multi-profile and single-profile variants are closely related. The corollary is the key result here: it provides a way of proving Theorems 3 and 4 at once, under the unsubstantial addition of $(\mathrm{PI})_f$ to (WP).[18] Once $(\mathrm{N})_f$ is reached, the dictatorship conclusion follows easily—remember Samuelson's graphical argument. The only difference between Theorem 3 and Theorem 4 is that one assumes $(\mathrm{UN})_f$ and $(\mathrm{ON})_f$, and the other starts one step remote by assuming $(\mathrm{UI})_f$ and $(\mathrm{ON})_f$, and then proving $(\mathrm{UN})_f$. All in all, this is not much of a difference.[19]

Second, and more importantly, it becomes easy to question the received view that permitting interpersonal comparisons of utility—i.e., replacing $(\mathrm{ON})_f$ by any weaker invariance condition such as ordinal or cardinal comparability—is the natural way out of dictatorship. $(\mathrm{N})_f$ is the proximate cause of dictatorship in the theorems. As Fig. 3 shows, relaxing $(\mathrm{ON})_f$ makes it possible to keep $(\mathrm{UI})_f$ and $(\mathrm{UN})_f$, as well as $(\mathrm{PI})_f$, without reaching $(\mathrm{N})_f$. But there is an alternative solution—

[18] It is well-known that on relevant economic domains, $(\mathrm{PI})_f$ follows from $(\mathrm{WP})_f$ and a continuity condition (see, e.g., Suzumura, 2001).

[19] Compare this technical analysis with Roberts's (1980) and d'Aspremont's (1985). In ways slightly different from ours, they tighten the links between Arrow's multi-profile theorem and his followers' single-profile version.

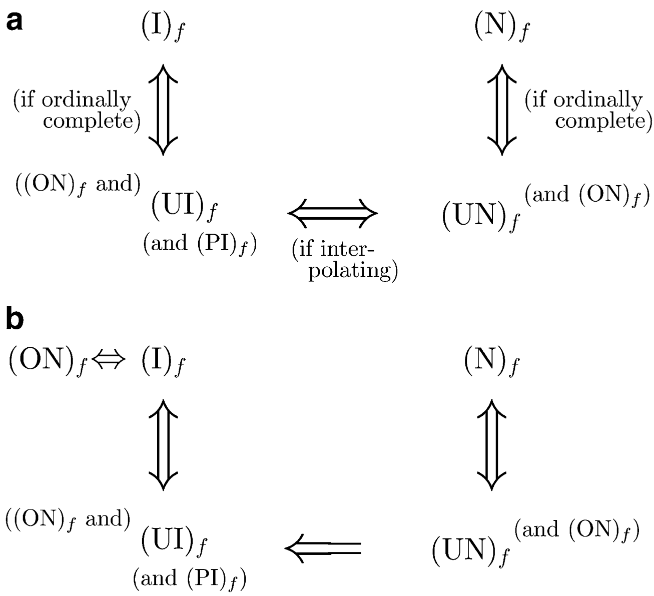

Fig. 3 **a**. The general case $\mathcal{D}^f$. **b**. The $\mathcal{D}^f_{\overline{U}}$ domain

keep (ON)$_f$ and (PI)$_f$ but relax (UN)$_f$. On the domain $\mathcal{D}^f_{\overline{U}}$, (UI)$_f$ and (I)$_f$, which are then equivalent to (ON)$_f$, are innocuous, but on a general interpolating domain $\mathcal{D}^f$, this solution requires dropping (UI)$_f$ and (I)$_f$, too. Therefore, on the domain $\mathcal{D}^f$, both the standard and alternative solutions require dropping (I)$_f$. As we see, (I)$_f$ is a *compound* condition of (UI)$_f$ and (ON)$_f$. Accordingly, the blame of the dictatorship conclusion in Arrow's theorem can be put either on the lack of interpersonal utility comparisons captured by (ON)$_f$, or on the binariness property captured by (UI)$_f$. *Prima facie*, there is nothing to recommend rejecting one more than the other.

By expressly adopting (ON)$_f$ and (PI)$_f$, the welfare economists did not leave much choice between these two solutions. They had to reject (UN)$_f$ and, considering the possible application of their analysis to the multi-profile context, (UI)$_f$ and (I)$_f$ as well.[20]

Examine the E^* example again. It gives rise to a social ordering functional f^* defined by: $xf^*(U)y$ if and only if

$$\min_i \ \max\{\lambda_i \in \mathbb{R} | U_i(x_i) \geq U_i(\lambda_i \Omega)\} \geq \min_i \ \max\{\lambda_i \in \mathbb{R} | U_i(y_i) \geq U_i(\lambda_i \Omega)\}.$$

This f^* satisfies (ON)$_f$ and (PI)$_f$, but none of the axioms (I)$_f$, (UI)$_f$ (except when $\mathcal{D}^f = \mathcal{D}^f_{\overline{U}}$), (UN)$_f$, (N)$_f$. More can be learned from f^*: it satisfies (ON)$_f$, hence we have produced the counter-example to show that (I)$_f$ and (ON)$_f$ are generally not equivalent for non-singleton domains $\mathcal{D}^f$.

For a two-individual, two-good economy, Fig. 4 shows why f^* does not satisfy (UN)$_f$ on the domain $\mathcal{D}^f_{\overline{U}}$ or on any larger domain $\mathcal{D}^f$. From this and the above results one deduces that it does not satisfy the other axioms on the relevant

[20] Blackorby et al. (1990, p. 283) reach a similar conclusion.

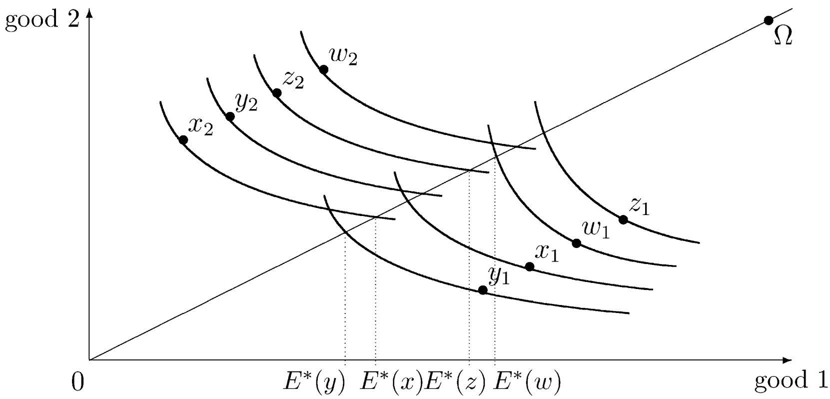

Fig. 4 Illustration of why f^* does not satisfy $(\mathrm{UN})_f$ on the domain $\mathcal{D}^f_{\overline{U}}$ or on any larger domain $\mathcal{D}^f$

domains. In Fig. 4, allocation x is better for f^* than allocation y, and allocation w is better than allocation z. But it is easy to find two profiles U, U' such that $U(x)=U'(z)$ and $U(y)=U'(w)$.[21]

Nonetheless, as will be detailed in the next section, many social choice theorists made the astounding claim that the BSF had to satisfy neutrality. Consider a social ordering functional f defined in terms of a BSF, so that $xf(U)y$ if and only if $W\circ U(x)\geq W\circ U(y)$. The satisfaction of $(\mathrm{ON})_f$ requires W to change with U, as explained in the previous section. Now suppose, on the contrary, that W does not change with U. In this case, it is true indeed that f satisfies $(\mathrm{UN})_f$ and also $(\mathrm{UI})_f$. The fact that W changes with U in order to satisfy $(\mathrm{ON})_f$ is therefore not foreign to the fact that the BSF generally does not satisfy $(\mathrm{UN})_f$. But forgetting this property of W and assuming that it is a fixed function is an easy mistake which may lead one to believe that neutrality is a typical property of the BSF. With these observations made, it becomes possible to reconstruct the controversy formally.

5 The controversy reconstructed

This section re-examines and rejects the social choice theorists' critique that the impossibility theorems apply to the BSF. But we will not endorse all and everything that the other camp's main respondent, Samuelson, said in his defence. Some of his arguments are strong, though just sketched; others are weak or even incorrect; hence the point of a detailed reconstruction, which also purports to explain why the social choice theorists eventually missed what was important in Samuelson's apparent confusion.

[21] For clarity of Fig. 4, the two pairs of allocations (x, y) and (z, w) use different amounts of resources, but it is easy to adapt the example to a case without free disposal.

5.1 An abstract summary

How could the social choice theorists attempt to refute "Bergson–Samuelson welfare functions" by way of the impossibility theorem? If it were feasible, the strongest move of all would be to demonstrate that Bergson and Samuelson included $(\mathrm{N})_f$ within their notion of a BSF. In view of the previous section, there is one, and only one way in which this move could be performed—the social choice theorists should try to deduce $(\mathrm{UI})_f$ from $(\mathrm{ON})_f$ and $(\mathrm{PI})_f$, two conditions that Bergson and Samuelson explicitly adopt, and from there, try to reach $(\mathrm{N})_f$. We call this the *logically-based strategy*. In words, it is an attempt at deriving neutrality from ordinalism and individualism taken as the only assumptions. To analyze it, we must make a decision on the domain $\mathcal{D}^f$ of the social ordering functional f associated with E.

On one interpretation, the welfare economists are concerned with a unique preference profile, and accordingly, their f is defined on some $\mathcal{D}^f_{\overline{U}}$. The trouble with this domain is that it is too small to support the logically-based strategy *up to the end*. Figure 3b shows that for f on $\mathcal{D}^f_{\overline{U}}$, $(\mathrm{I})_f$ and $(\mathrm{ON})_f$ are equivalent, so that $(\mathrm{ON})_f$ implies $(\mathrm{UI})_f$. However, f is never interpolating on $\mathcal{D}^f_{\overline{U}}$, and the next step from $(\mathrm{UI})_f$ and $(\mathrm{PI})_f$ to $(\mathrm{UN})_f$ cannot generally be achieved on this domain—see Fig. 3b and the comments after Proposition 2. This dashes the hope of reaching $(\mathrm{N})_f$ from $(\mathrm{ON})_f$ and $(\mathrm{PI})_f$.

On the other interpretation, a larger domain can be selected. In the absence of specific restrictions, it would be standard to take $\mathcal{D}^f = \left(\mathbb{R}^X\right)^n$, although any non-singleton domain could do as well. What goes wrong this time is explained in the comments after Proposition 1 and at the end of Section 4—the *initial* step from $(\mathrm{ON})_f$ to $(\mathrm{I})_f$ cannot be performed anymore. Again, the derivation fails.

The upshot is that welfare economists have a reply to make whatever the chosen domain. Once this reply is made, there remains only one possible use for the impossibility theorem. If any of the axioms $(\mathrm{N})_f$, $(\mathrm{UN})_f$, $(\mathrm{I})_f$ or $(\mathrm{UI})f$ could be shown to be compelling, or at least strongly commendable, the welfare economists would have to accept the social choice theorists' conclusions. We call this the *normative strategy*; one should be very clear about what distinguishes it from the previous, logically-based strategy. Let us now compare this abstract summary with the actors' own understanding and respective moves.

5.2 Kemp and Ng

Kemp and Ng (1976, 1977, 1982) developed a formalism of BSF for a $\mathcal{D}^f_{\overline{U}}$ domain exclusively, a choice that was dictated by their reading of Samuelson (1967). Beside being mathematically loose, their work constantly equivocates between the normative and the logically-based strategy. In the 1976 paper, the former line prevailed over the latter. Their spectacular conclusion that it was impossible to find a "reasonable" BSF hinged on the prior claim that the conditions of their theorem (here Theorem 4) were all "reasonable":

> "Any social ordering derived by a rule that satisfies certain "reasonable" conditions must be lexicographical...." (1976, p. 59).

Kemp and Ng provide no argument in favour of the neutrality condition, except for an appeal to Arrow's authority: it is allegedly "similar to Arrow's Independence of Irrelevant Alternatives" (1976, p. 61). Their verbal statement of neutrality is that "the social ordering of any two alternatives depends only on the individual orderings of alternatives" (p. 60), an evidently incomplete description, which reduces it to independence. The slip was carefully noted by one of their respondents, Mayston (1979, p. 184). Even if the "similarity" of the two conditions could be accepted, Kemp and Ng would have to argue that independence is normatively attractive, and there is not even the beginning of an argument to this effect.

After Samuelson had shown by means of his chocolate example (see below) that neutrality had no normative standing for distributional issues, Kemp and Ng reinterpreted their contribution. Their second paper—a brief reply to Samuelson—makes the distinction between the two possible lines of attack against BSF, and bends their contribution towards the logically-based strategy:

> "We do not claim that our [neutrality] axiom is reasonable or unreasonable. *We claim only that [it] gives expression to the requirement that only individual orderings count*" (1977, p. 90; our emphasis).

This major claim was left unsupported. In their third paper, Kemp and Ng acknowledged the failure, and announced that they had remedied it:

> "We have already stated, in our reply to Paul Samuelson, that [our neutrality axiom] is implied by Individualism and Ordinalism... In the present note we supply the missing proof" (1982, p. 33).

It is very instructive to compare the last two quotes. If Ordinalism is identified with "the requirement that only individual orderings count", as the comparison suggests, there must be a shift from one paper to another. Their 1977 "claim" was that neutrality followed from Ordinalism alone, whereas the 1982 "proof" is now meant to establish that this condition follows from Ordinalism *and* Individualism. In the f formalism, neutrality is $(\mathrm{N})_f$ and Ordinalism is $(\mathrm{ON})_f$. What is then Individualism? Kemp and Ng (1982, p. 34) analyze it into $(\mathrm{PI})_f$, oddly relabelled "Weak Individualism", *and* $(\mathrm{UI})_f$, conventionally labelled "Independence". Granting once again $(\mathrm{PI})_f$, how are they to account for adding $(\mathrm{UI})_f$? Since they cannot prove that it is implied by the welfare economists' assumptions, they simply argue that it is "extremely reasonable" (p. 34). In other words, Kemp and Ng revert to the normative strategy in the middle of a failed attempt at carrying out the logically-based strategy.

Beside making its point incoherently, the 1982 paper is technically inaccurate. The claimed result is that $(\mathrm{PI})_f$, $(\mathrm{UI})_f$ and $(\mathrm{ON})_f$ together imply $(\mathrm{N})_f$, which is the Corollary of Section 4, except that Kemp and Ng skip the required domain restrictions (1982, p. 35). Evidently, they have missed the dilemma induced by the choice of a domain $\mathcal{D}^f$. Samuelson did not react to Kemp and Ng's "proof", but Mayston (1982) did, and some of his objections overlap with those made here.

5.3 A step back to Arrow's independence

In their 1987 paper on Arrow, Kemp and Ng tried to patch up their normative strategy by explaining why they believed independence to be such a compelling condition. Their argument is an elaboration of Arrow's (1963, p. 26). The latter had claimed that the social ordering between two candidates A and B should not depend on how the electors ranked a third candidate C vis-à-vis these two. In order to defend his claim, he had envisaged the possibility that C died before the ballot, and submitted that the date of C's death should not be allowed to interfere with the electoral decision between A and B. Accordingly, it should be possible to decide between A and B by "blotting out" C's name in the individual orderings. Kemp and Ng try to reinforce Arrow's point by adding this:

> "If whether C is around is a relevant consideration, then a complete specification of alternatives will not just involve A versus B but rather A (as chairman) with C around, B with C around, A without C, etc. With a complete specification of the alternatives, (I) is not open to the objection.... It is not an overstatement to say that "to understand independence is to accept it" " (1987, p. 225–226).

This semantic manoeuvre of redefining the options is popular in individual or collective decision theory, where, for example, it has served to protect von Neumann–Morgenstern independence from criticism. The usual objection is that it salvages the condition under attack only by making it either inapplicable or vacuous. Under Kemp and Ng's reformulation, (I) could play no role in the analysis of concrete electoral systems, since a ballot takes the form of a straightforward ranking of either A and B, or A, B and C, and not of A and B knowing that C is or is not part of the contest. In the case of the independence condition, there is a further objection to be made. Suppose that the individuals express preferences between redescribed alternatives A′=A given that B and C are also available, and B′=B given that A and C are also available. In order to rank A′ and B′ socially, there is *prima facie* more information about individual preferences to be considered than just these preferences between A′ and B′. *Prima facie*, the social ranking may be affected by how A′ and B′ are ranked with respect to C′=C given that A and B are also available. It remains to establish that C′ is irrelevant, which means that the initial argument has not fulfilled its aim.[22] In other words, the assumption that *alternatives* include a description of everything ethically relevant about social states does not entail that *binary preferences* over these alternatives exhaust what is ethically relevant for the ranking of two social states.

[22] Similarly, consider again the economic example of the distribution of Ω among n individuals. Suppose alternatives are redescribed in this way:

$$\widehat{x} = (x_1, \ldots, x_n, \Omega, U_1^*(x_1), \ldots, U_n^*(x_n)).$$

If one believes that E^* is the proper ethical criterion for this example, then one must admit that $\widehat{x}$ contains all the relevant information. But it still holds that individual i weakly prefers $\widehat{x}$ to $\widehat{y}$ if and only if $x_i R_i y_i$. Therefore (I) applied to individual preferences over $\widehat{x}$, $\widehat{y}$ is just the same as (I) applied to the initial x, y, and is just as restrictive and questionable.

As to Arrow's argument itself, it is in effect just a claim. Some well-established electoral procedures, like Borda's rule of scoring, are sensitive to how the relevant set of candidates is defined, while others, like the majority rule, do satisfy the binariness property encapsulated in (I). Arrow lucidly describes the consequences of adopting either type of rule, but gives no reason to warrant his strong suggestion of the "reasonableness" of the latter type as against the former (1963, p. 27). It may be that either type is reasonable from a particular perspective—e.g., the former because it takes account of interactions that are an integral part of the decision problem, the latter because it is computationally economical. The counter-argument we are sketching has become common in the post-Arrovian literature of mathematical politics, which does not revere independence anymore.

This leaves welfare economics aside, for which Arrow had a specific example and tentative justification.

> "Suppose that there are just two commodities, bread and wine. A distribution, deemed equitable by all, is arranged, with the wine-lovers getting more wine and less bread than the abstainers. Suppose now that all the wine is destroyed. Are the wine-lovers entitled, because of that fact, to more than an equal share of bread?…My own feeling is that tastes for unattainable alternatives should have nothing to do with the decision among the attainable ones; desires in conflict with reality are not entitled to consideration, so that [(I)], reinterpreted in terms of tastes…, is a valid value judgment, to me at least' (1963, p. 73).

This wine-and-bread example is hardly more helpful than the dead candidate example. Arrow only suggests that in absence of wine the selected allocation should ignore individual preferences over wine, even if this hurts those who would had previously received a lot of wine. This fails to imply that the ranking of two allocations of bread and wine should depend only on individual preferences over these two allocations.

5.4 Parks

A consistent representative of the logically-based strategy, Parks (1976) carefully refrained from any suggestion that neutrality might be a "reasonable" condition. After stating his variant of $(\mathrm{N})_f$, he stressed that his purpose was "not to argue in favor or against the axiom but rather to present its logical consequences" (1976, p. 448). So Parks' argument entirely depends on showing that Bergson and Samuelson have implicitly accepted $(\mathrm{N})_f$. He should be noted for making a claim that will become popular among social choice theorists, i.e., that *neutrality formalizes the lack of interpersonal comparisons of utility.*

Having recast a BSF as a composed function:

$$W(U(x)) = W(U_1(x), \ldots, U_n(x)),$$

Parks takes the following crucial step:

> "Exactly what this means in terms of interpersonal comparisons is not totally clear. We take it to mean that the individual preferences can be used in determining W(.) and nothing more. We formalize this by… the following

axiom. Let φ_i be any increasing affine function. Then, for all $u, u' \in \mathbb{R}^n$, $W(u) > W(u')$ if and only if $W(\varphi(u)) > W(\varphi(u'))$, where $\varphi(u) = (\varphi_1(u_1), \ldots, \varphi_n(u_n))$" (1976, p. 450; notation adapted).

From his axiom, Parks correctly concludes that BSF must satisfy neutrality, and after invoking the analogue of Theorem 4, that they are dictatorial. Returning to the interpretation, he points to the lack of interpersonal comparisons as to the source of the trouble.

The mistake in this reasoning is straightforward. Parks' informal comment on interpersonal comparisons would be captured by an f satisfying only $(\mathrm{PI})_f$ and $(\mathrm{ON})_f$, but his axiom makes it satisfy $(\mathrm{UN})_f$. This is essentially the same problem as in Kemp and Ng, but it is easier to locate because of Parks' clearer exposition. Admittedly, he expresses a reservation at the beginning of the passage, and says "interpersonal comparisons" rather than "interpersonal comparisons *of utility*". However, it is unlikely that Parks would draw a distinction between two kinds of comparisons without elaborating on it, since it was not—and it is still not—well understood among economists.

5.5 Samuelson

Samuelson (1977, 1981, 1987) rebutted neutrality by making two distinctive moves. When examined separately, each of them turns out to be relevant and even convincing, but his hurried 1977 reply jumbles them together, and his 1981 and 1987 papers, on Bergsonian economics and Arrow's social choice theory, respectively, create more difficulties in interpretation than they resolve. In terms of our organizing taxonomy, the first move counters the social choice theorists' *normatively-based* strategy, while the second blocks their *logically-based* strategy.

In order to dispel the impression that neutrality might be a "reasonable" condition, Samuelson devised the amusing chocolate example. Suppose that the "ethical observer" is to divide a stock of 100 chocolates between two individuals.

> "What is the meaning of [Kemp and Ng's neutrality condition] in this context? It says, "If it is ethically better to take something (say 1 chocolate or, alternatively, say 50 chocolates) from Person 1 who had all the chocolates in order to give to Person 2 who had none, then it must be ethically preferable to give all the cholates to Person 2". One need not be a doctrinaire egalitarian to be speechless at this requirement. Is it "reasonable" to put on an ethical system such a straightjacket? Few will agree that it is" (1977, p. 83).

This ingenious parable is virtually all that is needed to deprive (N) or $(\mathrm{N})_f$ from their normative appeal as far as *distributive* applications are concerned, which is what the BSF is intended for. Incidentally, the parable also illustrates Samuelson's mathematical point that neutrality implies dictatorship "transparently"; the connection between the two conditions is made tangible by the advantage given to Person 2 over Person 1. With some exaggeration, it could be said that Samuelson argues from dictatorship to the inadmissibility of the condition that implies it, hence turning the social choice theorists' favourite weapon against them.

Samuelson also tried to dispel the impression that neutrality somehow underlay Bergon's and his conception of ordinalism. Later in his reply, he puts the matter as follows:

> "Is there some semantic or philological reason for failure to agree to satisfy [Kemp and Ng's neutrality condition] to be a judged by the jury to be a lapse from ordinalism in ethical judging? "No" is I believe the proper answer to such a question" (1977, p. 83–84).

This is a uselessly complicated and hardly grammatical way of making the point, and the ensuing discussion is obscure, in part because of its *ad hominem* style, in part because it returns ineffectively to the normative point already made. Contrary to the first, Samuelson's second line of attack left many readers unshaken, and this is definitely his fault. He should have gone straight to his example of a well-behaved BSF and explained that it satisfied $(\mathrm{ON})_f$ but not $(\mathrm{UN})_f$ or $(\mathrm{N})_f$, which the diagram by itself does not establish (1977, p. 85–87). Had Samuelson got to these details, the misunderstandings about the BSF would not have spread beyond the initial papers of the controversy.

We have already mentioned that Samuelson was reluctant to extend his critique of neutrality to independence. The hindrance is visible in his 1977 reply, where he praises Arrow with exaggeration compared with Kemp and Ng, and even more clearly in his next paper, where he makes a detailed comparison between a BSF and an Arrovian social ordering function (1981, p. 258–259). There, he lists all conceivable profiles of two individuals' strict preferences on a set X of three elements, and for each such a profile, all Paretian social orderings that are compatible with it. After this painstaking exercise, Samuelson spells out what the possible Bergson–Samuelson and Arrow functions are. For any profile, any Paretian ordering defines a BSF. Many of these orderings do not coincide with one individual's ordering in the profile. If a BSF were redefined as a social ordering function defined on the whole set of profiles, it would be seen immediately that it is dictatorial if and only if it satisfies (I)—Arrow's case. Carried by his own analysis, Samuelson should have scrutinized this condition. But he preempted the discussion with the clichè he was accustomed to make since 1967, i.e., that Arrow's work belongs not to welfare economics, but to mathematical politics, and that it is unobjectionable in this other field.[23] For Samuelson, as far as the "Constitutional Function" interpretation of F is concerned, Arrow made "four plausible axioms" (1981, p. 262; see also 1967, p. 47:"all three conditions seem reasonable"). When it comes to the BSF, Arrow's contribution simply disappears from the scene, and the logical analysis comes to a halt.

It is only in Samuelson (1987, p. 169–171) that independence is discussed in connection with welfare economics. He considers two individuals, 1 and 2, and two allocations, i.e., x=(15, 5), y=(5, 15), where the numbers denote apples. Samuelson assumes that the E function is weighted additive:

$$E(x) = \sum_i w_i U_i(x_i),$$

[23] See, e.g.:"I shall argue that the Arrow result is much more of a contribution to the infant discipline of mathematical politics than to the traditional mathematical theory of welfare economics" (Samuelson 1967, p. 168).

with U_1=U_2, and $E(z)$=$E(w)$, with z=(10, 5) and w=(5, 10). It immediately follows that w_1=w_2, and hence that $E(x)$=$E(y)$. The point of this curiously trivial example is to suggest that a social decision about two states of the world (x and y) may sensibly depend on a social decision on two other states of the world (z and w). Samuelson turns the example into a denial of independence:

> "The axiom… seems to rule [this] out, saying 'In deciding between the binary states x and y, don't let the reactions to the non-relevant states z and w affect your judgement rendering'." (p. 170).

And he eventually concludes that he and Bergson

> "are explicitly (and reasonably) deciding to violate [the independence axiom]. Third states of the world do seem to force themselves legitimately into our binary choices…. Most ethical systems purport to define who is the deserving one by how the contemplated individuals react to a vast panoply of possible situations. So each and every binary $x{\sim}y$ decision depends on people's $\sim_i$ decisions in many, many other (z, w,…) situations" (p. 170).[24]

The forceful conclusion is exactly that which one would have hoped Samuelson to reach earlier. Unfortunately, the supporting example is only loosely related to $(\mathrm{UI})_f$, and it is flawed anyway, because the assumption that U_1=U_2 clashes with ordinalism. (The U_i can be replaced by arbitrary increasing transforms, so that no significance can be attached to their equality.) The source of Samuelson's persisting difficulties with independence can perhaps be located in the fact that this apple example, exactly like the chocolate parable, *involves only one commodity*. The assumption entails that there is only one profile of economic (self-centered and strictly monotonic) preferences, so that $\mathcal{D}^f = \mathcal{D}^f_{\overline{U}}$, and $(\mathrm{I})_f$ and $(\mathrm{UI})_f$ are automatically satisfied if $(\mathrm{ON})_f$ is. Amazingly, Samuelson focused on the only special case in which the BSF is bound to satisfy independence!

To conclude about the major protagonist's contribution, he had a winning point to make against the social choice theorists, but he wasted it both during the controversy and (with less excuse) when revisiting the issues. His disorganized 1977 plea for the BSF did not help social choice theorists to realize where they had gone astray. This rhetorical fault is partly explained by Samuelson's dislike for what he thought was an uninteresting discussion. When he returned to it, he could not refrain from expressing his aloofness:

> "If it were not that a number of writers in the social choice vineyard (Kemp and Ng, Parks, Hammond, Pollak,…) attach disproportionate importance to this issue, I would be even briefer in disposing of it here" (1981, p. 235).

Samuelson might be right that this was not a high-brow controversy after all, but the derogatory tone is out of place, given the serious work done by his op-

[24] Samuelson is justified in attributing to Bergson a rejection of independence. However, Bergson's (1954, p. 244–245; see also 1966, p. 75–76) criticism relied on the concept of *need*, and virtually ignored the point that Samuelson is making here, i.e., that individual *preferences* about third allocations are relevant in the evaluation of two allocations.

ponents, and the fact that he had himself tried so hard to win the stake. Given Pollak's (1979) intervening clarifications, Samuelson could have done better.

5.6 Pollak

When he stepped in the controversy, Pollak (1979) knew Kemp and Ng's, Parks' and Samuelson's papers. The benefit of hindsight may explain why he has the most refined formalism of all. Our statement of Theorem 2 is borrowed from his paper, and he also offers a lexical dictatorship variant based on (*SP*) instead of (*WP*). His technical comments clarify the comparison between Theorem 1 and Theorem 2, in preparation for the general meta-theoretical claim that "it is likely that there are single profile analogues of virtually all results in the theory of social choice" (p. 86). This perspective implies a rejection of Samuelson's enduring view that the framework is by itself an issue.

Not surprisingly either, Pollak's paper has also the most balanced conclusions of all. It appears that he never contemplated the logically-based strategy against the BSF, and concerning the normative strategy, he is explicit that it would not work, because neutrality is "unacceptable" (p. 86) in the case of the BSF.[25] This sets him aside from other social choice theorists, except for Mayston and Pazner, who also found themselves on the welfare economists' line. Concerning the political interpretations of neutrality, Pollak is not as black-and-white as Samuelson, who is prepared to grant all of Arrow's assumptions as soon as they do not bear on welfare economics. When reviewing Arrow's defence of independence, we argued that independence is not compelling for electoral systems. Pollak makes a related point about neutrality in the political realm:

> "[Neutrality] is clearly a vulnerable axiom: on its face, it implies that if a group of jurors can elect Smith over Jones as foreman over the opposition of the rest, then this same group of jurors can convict the defendant over the opposition of the rest" (1979, p. 80).

In sum, it is no less naive to separate the "political" from the "economic" interpretation in terms of the independence or neutrality condition than it is to do so in terms of the number of profiles. Pollak's courteous exposition does not draw this conclusion as sharply as we do, which may be the reason why it was lost to Samuelson.

6 Further perceptions of the controversy

This section singles out for discussion some significant comments that were not part of the controversy, but accompanied or followed it. The first group illustrates the received view of the BSF and its genesis. The second group focuses on two—unfortunately little-known—contributions by social choice theorists who dissented from this received view and whose interpretation of the BSF inspired ours. We have

[25] Pollak's conclusion coincides with Samuelson's, but the main argument he provides is oddly the ineffective Bergsonian one "that the Bergson–Samuelson social welfare function reflects the judgments of an ethical observer" (1979, p. 86).

added a brief section on Arrow, which puts the complex story in historical perspective and is also meant to do him justice; as will be seen, he came to disapprove of the social choice theorists' position.

6.1 Sen and the mainstream of social choice theory

In their rejoinder, Kemp and Ng (1977) had argued that Samuelson's well-behaved BSF relied on *more* information than could be made available by ordinal non-comparable preferences. Careful scrutiny of their argument reveals that they mistook the numerical measurement of physical goods with a cardinal measurement of utility. However, because Samuelson's exposition was incomplete, there was something to be said for their peculiar reading of his diagram, and Sen went to their rescue in the following comment. Kemp and Ng are correct, he writes, if the BSF

> "is interpreted to include...welfarism. Welfarism may appear to be implied by the form in which the Bergson–Samuelson function is sometimes written: $E=W(U_1,\ldots,U_n)$. *There is an ambiguity here*: if E and U_i are taken not to be welfare numbers but functions defined over X, then welfarism is not, in fact, implied. (Indeed then $W(.)$ will be very like a social welfare functional f... with E being a real-valued representation of the social ordering determined by f.) However, it appears that this 'functional' interpretation of $W(.)$ was not intended in the formulations in question (see...Samuelson 1947, p. 246...). And if $(U_1,\ldots,U_n)$ is simply a vector of individual utilities, then welfarism will follow, and impossibilities will be round the corner given unrestricted domain and the absence of interpersonal comparisons" (1977, p. 255; our notation and emphasis).

To paraphrase, Sen is considering two possible interpretations for the formula $E=W(U_1,\ldots,U_n)$, one (i) involving utility numbers and leading to impossibilities (i.e., Theorem 4), the other (ii) involving utility functions and immune to impossibilities. Kemp and Ng implicitly adopt the former interpretation, and they are justified in doing so to the extent that Samuelson's *Foundations* can be read as excluding (ii). This sophisticated comment sounds like a relief for Kemp and Ng. However, it runs into a problem of its own—it collapses two different distinctions into a single one. There is a first question, which is addressed here: is W defined on utility numbers or functions? If the answer is "W is defined on utility numbers", a second question arises, which is *not* being addressed: does W vary with the underlying profile U, or does it remain the same? It is true that Samuelson does not consider (ii), but this does not establish that he accepts (i), because he could also answer (iii) "W is defined on utility numbers and varies with U". This option, like (ii), is immune to impossibilities.

We have encountered the critical point about W more than once, but since the issue of "welfarism" might now obscure it, we will restate our response to Sen formally. By definition, "welfarism" obtains when the social preference ordering $R=f(U)$ can be expressed as an ordering R^* of utility vectors that does not depend on the utility profile U. That is to say, the definition requires that for some ordering R^*, and for all $U \in \mathcal{D}^f$ and $x, y \in X$,

$$x\mathcal{R}y \Leftrightarrow U(x)R^*U(y).$$

Assuming that R^* can be represented, it is equivalent to require that for some V, and for all $U \in \mathcal{D}^f$ and $x, y \in X$,

$$xRy \Leftrightarrow V(U(x)) \geq V(U(y)).$$

That a Paretian BSF can be written as $E = W \circ U$ does not by itself ensure that it satisfies "welfarism". This will be the case if and only if W does not change with U—see the order of quantifiers in the last requirement. The non-"functional", non-"welfarist" case in which W in $W \circ U$ varies with U is the option (iii), and Sen does not appear to take it into consideration.

The following comment by Roberts was made shortly after the controversy:

> "Most Bergson–Samuelson social welfare functions take the following basic form: there exists a real valued function W, with welfares achieved in a state as arguments, with the property that $W(U(x)) \geq W(U(y)) \Leftrightarrow x f(U) y$. Thus $(\mathrm{UN})_f$ is satisfied for some measurability/comparability assumption and $f(U)$ is complete, reflexive and transitive. Further, $(\mathrm{WP})_f$ is generally invoked. By the analogue of Arrow's impossibility theorem, when [a suitable domain assumption] is satisfied, f will be dictatorial if $(\mathrm{ON})_f$ holds" (1980, p. 449; our notation).

Like Parks, whose reasoning he closely follows here, Roberts does not envisage that W may vary with U. Among various comments in the same vein, we have selected Sen's and Roberts's because their early dates testify to the fact that the received view of the controversy was quickly established (no doubt, they have also been influential in establishing it).

In what may be his latest review of the controversy, Sen (1986, p. 1149) briefly repeats his defence of Kemp and Ng, and adds an interesting footnote that again links the individualistic form of the BSF to neutrality:

> "From the exchange between Kemp and Ng (1976, 1977) and Samuelson (1977), it would appear that it is not—indeed never was—Samuelson's intention to insist on neutrality. It is certainly the case that Samuelson (1947) made critical comments on this "extreme assumption"..., and while this did not stop him from dealing extensively with cases in which this condition is fulfilled..., the traditions of economic theory do not, of course, permit one to deduce belief from extensive use" (p. 1149–1150).

As a matter of fact, what Samuelson dubbed an "extreme assumption" (1947, p. 223) and indeed used extensively is the "individualism" (Pareto) condition. It does not appear that *The Foundations of Economic Analysis* deals with cases in which neutrality is fulfilled.

6.2 Two exceptions: Mayston and Pazner

Mayston (1974, 1979, 1982) offers an array of arguments about independence, neutrality, ordinalism, and the BSF, which inexplicably failed to attract his fellow theorists's attention. The exceptions are Kemp and Ng (1982, 1983, 1987), whose work, conversely, is discussed at great length by Mayston. For him, "the problem of social choice is not an impossibility of consistency, but rather an *embarras de richesses*" (1974, p. 79). How could he reach such a dissident conclusion? To debunk the neutrality and independence conditions is a major objective of Mayston's book, nonetheless entitled *The Idea of Social Choice* (1974, p. 69–81). Several years in advance, Mayston analyzed neutrality exactly as Samuelson would do in his chocolate example:

> "[one] may well find condition (N) a very unattractive one in formulating 'welfare' judgements about different social states. Consider the following situation: x_1=individual I is starving; x_3=individual II is starving; x_2=both individuals are well fed but individual I is £1 lower in income than in x_3 and individual II is £1 lower than in x_1, together with $x_3P_1x_2P_1x_1$; $x_1P_2x_2P_2x_3$. A formulation by the outside observer that socially x_2Px_3 implies under condition (N) that he must also state x_1Px_2" (1974, p. 78).

Concerning independence, he developed the following argument:

> "We are concerned to achieve consistency over the whole of X. Alternatives outside any proper subset S of X, but still in X, do therefore become relevant to the achievement of this basic goal. More especially in seeking a SWF over X, one causes to be admissible as a means of deciding between any pair $\{x_1, x_2\}$ in X not simply a direct comparison between x_1 and x_2, but also any path of the form $x_1Rz_2Rz_3 \ldots Rx_2 \ldots$ Indeed in the process of constructing a social ordering, a degree of dependence can be shown to be necessary in the formulation of the social choice between a pair $\{x_1, x_2\}$, once we have independently formulated the social preference for another pair $\{y, z\}$" (p. 70).

Mayston illustrates his argument in terms of an example that also has a bearing on neutrality, and for this reason, was actively debated between Kemp and Ng and him (see 1979, p. 195–198, and again 1982, p. 116–118). Suppose that in a two-individual population,

$$xP_1zP_1wP_1y,$$
$$wP_2yP_2xP_2z.$$

Suppose further that xIy. For Kemp and Ng, this example contains the same ordinal information about x and y as that about z and w; hence, one should have zIw. Against this, it may be said that the information available about (x, y) differs from that available (z, w) in a subtle way. It is true that individual 1 prefers z to w and x to y, while individual 2 has the opposite preferences for both pairs, but it is also the case that z and w are located between x and y in 1's ordering, and outside them in 2's ordering. Mayston claims that the information just described explains why x and y should not be treated by the social ordering in the same way as z and w. Accordingly, he finds nothing objectionable in the conclusion that follows from the standard assumptions: because x Pareto-dominates z and w Pareto-dominates y, by transitivity, xIy implies wPz. This conclusion clashes with Kemp and Ng's claim above that zIw should prevail.

The clash comes as no surprise, since Kemp and Ng's claim is an application of neutrality, and the example as a whole is a toy model of how to prove Theorem 2. The added value of this example lies elsewhere—it illustrates conflicting intuitions about *ordinalism*. The conclusion that wPz means that 2's difference in preference between w and z weighs larger in the social ordering than 1's difference in preference between z and w. For Mayston, this conclusion illustrates that ordinalism has more logical power than is usually thought—*it can take into account a notion of preference difference*. By contrast, for Kemp and Ng, ordinalism is *definitionally* not about preference differences, hence it must include either the independence or even the neutrality condition, in order to cancel any information on preference differences that orderings may inadvertently deliver. It is Mayston who is the more faithful to the meaning of ordinality in economics. The information he proposes to take into account is simply contained in the data of individual rankings; it is invariant to the choice of utility representations, and there is nothing cardinal about it. There may normative reasons to ignore preference differences, so that (I) or (N) may be justified at the end of the day, but these reasons cannot be found within ordinalism itself. Kemp and Ng are *semantically* inconsistent. This analysis complements our rebuttal in 5.2.

Having rejected (I) and (N), Mayston (1974) sets out to find a weaker condition than independence, and in particular proposes a "weak independence of irrelevant preferences", which stipulates that the social ranking of any pair of alternatives x, y depend on individual indifference sets *at x and y and in between*. Formally, if

$$B_i(x,y) = \{z \in X | xR_i zR_i y \quad or \quad yR_i zR_i x\}.$$

the condition reads:

$$(\textbf{WIIP}) \ \forall \overrightarrow{R}, \overrightarrow{R'} \in \mathcal{D}^F, \forall x,y \in X,$$
$$\left[\left(B_i(x,y) = B_i'(x,y)\right) \& \left(R_i|_{B_i(x,y)} = R_i'|_{B_i'(x,y)}\right), \forall i \in N\right] \Rightarrow (xRy \Leftrightarrow xR'y)^{26}.$$

For instance, the social ordering function F^* derived from E^* (and having f^* as its canonical associate) does satisfy this axiom. But Mayston's general formulation of a social ordering function is more complex. It involves defining individual weights $w_i(z)$ for every element of a "social weighting path", such that the variation in social welfare between x and y is computed by a weighted integration of the distance covered on this reference path by the indifference curves between x and y. This amounts to a weighted utilitarianism, in which social welfare at a given alternative x is a sum of individual integrals and each such integral, for i, adds up the weights $w_i(z)$ over the path from the origin up to the alternative $z_i(x)$ from the path such that $xI_iz_i(x)$. Every possible choice of $w_i(.)$ will produce "aggregate

[26] In a stronger version called "independence of irrelevant preferences", Mayston requires only the intersection of $B_i(x, y)$ with a reference path Z crossing all indifference curves to be the same in $\overrightarrow{R}$ and $\overrightarrow{R}'$.

consistency" (1974, p. 79). At this point, Mayston's work is no more helpful than Bergson's and Samuelson's. He leaves the reader with a weighted sum to represent the social ordering, but does not spell out what principles could guide society in the choice of weights.

The other exception among social choice theorists is Pazner (1979), whose innovative contribution was no more noticed than Mayston's. No doubt because of the author's premature disappearance, there is only one paper to report, and this unique sample is itself rather terse. Without even mentioning the single-profile impossibility theorems, Pazner makes it clear that he sides with Samuelson and considers the existence of well-behaved BSF as non-problematic. He suggests something like the E^* function as evidence for this claim. He relates it to Pazner and Schmeidler's (1978) egalitarian equivalence concept, and may be credited for bridging the gap between traditional welfare economics and the recently born theory of fair allocation. Another contribution of his 1979 paper is the weakening of independence called "Independence of non-indifferent alternatives" (p. 172). Differing from Mayston's, this condition says that the social ranking between two alternatives should only depend on the individual indifference sets these alternatives belong to. Formally, using the notation $B_i(x, x)$ for the indifference set of x,

$$(\mathbf{INIA})\ \forall \overrightarrow{R}, \overrightarrow{R'} \in \mathcal{D}^F, \forall x, y \in X,$$
$$\left[\left(B_i(x,x) = B'_i(x,x)\right)\&\left(B_i(y,y) = B'_i(y,y)\right), \forall i \in N\right] \Rightarrow (xRy \Leftrightarrow xR'y).$$

In the economic context this condition is stronger than Mayston's (WIIP), but it is nonetheless satisfied by E^* and is very natural. In particular, in the theory of fair allocation it is very common to assess the efficiency and equity properties of an allocation by examining the individuals' indifference curves at this allocation.

There may be a subtle difference between Pazner and Schmeidler's (1978) and Pazner's (1979) papers with respect to the semantics of ordinalism. The statement that egalitarian-equivalent allocations are generated by the maximin criterion "under a particular method of interpersonal comparisons" (1978, p. 680) suggests that Pazner and Schmeidler initially saw E^* as relying on ordinal *comparable* utilities. If this was the case, the misunderstanding was eventually dispelled by Pazner:

> "The apparent coordinality of [E^*] is misleading since, rather than using any particular utility representation, we can look at the underlying ordering of allocation space induced by this maximin function as being the BSF in question... The analysis is purely ordinal since independent positive monotonic transformations... will not alter the results" (1979, p. 169).

Even before Mayston and Pazner, Hansson (1973) had critically discussed (I) and proposed to weaken it into a condition similar to Pazner's (INIA). Unfortunately, Hansson's condition is stated in an abstract framework of social choice theory, and he did not examine its possible applications to welfare economics. Being primarily concerned with the BSF, neither the participants to the controversy, nor even Pazner and Mayston, had a chance to pay justice to Hansson's innovation.[27]

6.3 Arrow

Arrow's early analysis of the BSF does not belong to the topic of this paper but we may cite the following two passages from *Social Choice and Individual Values* as signposts for the work to follow in social choice theory:

> "If we exclude the possibility of interpersonal comparisons of utility, then the only methods of passing from individual tastes to social preferences which will be satisfactory and which will be defined for a wide range of sets of individual orderings are either imposed or dictatorial" (1963, p. 59),

And:

> "The Bergson social welfare function is mathematically isomorphic to the social welfare function under individualistic assumptions. Hence, the Possibility Theorem under Individualistic Assumptions... is applicable here; we cannot construct a Bergson social welfare function" (1963, p. 72).

Without the second statement, the controversy we have scrutinized would not have taken place, and without the first, the standard conclusion that BSFs "make interpersonal comparisons or are dictatorial" would not have been reached.

A puzzling feature of the first quotation is that it does not mention (I). But in the same passage he explains that (ON) and (I)

> "taken together serve to exclude interpersonal comparison of *social* utility either by some form of direct measurement or by comparison with other alternative social states" (p. 59; our emphasis).

In other words, (I) serves to exclude a particular kind of interpersonal comparisons, namely, those that make use of indifference loci and refer to third alternatives. This rhetorical connection between ordinalism and independence is reminiscent of Kemp and Ng's line of argument, so that Arrow may have been a source of the logically-based strategy against the BSF. But in light of Subsection 5.3, he was also, obviously, the ancestor of the normative strategy and one may argue that the latter is more faithful to Arrow's general approach.[28]

At any rate, Arrow's (1983) late assessment of the BSF is in stark contrast with his original work. As he now explains, a BSF is in effect a social ordering function *F* defined on wide domains of profiles and satisfying the Paretian conditions, and nothing else. Hence, it is non-problematic that non-dictatorial BSFs exist:

> "Is there any problem in existence? As is clear from the discussion, the answer is no. If there are "rumors that Kenneth Arrow's Impossibility Theorem rendered Bergson's "social welfare function" somehow non-existent or self-

[27] Similarly, Fleurbaey and Maniquet (1996) were not aware of either Hansson, Mayston or Pazner when they proposed in their turn to weaken the independence condition.
[28] See Mongin (1999, 2002).

contradictory", they are indeed "quite confused" (p. 21; quoted expressions echo Samuelson).

Consistently with this claim, Arrow recognizes that W and U vary together, and even provides the formula of Section 3 to compute W' from W, U, U'. As another contribution to the welfare economists' case, he brings Samuelson's 1977 incomplete analysis to its end. Whereas our formula for Samuelson's example involves the *min* of the U_i, he uses the sum (p. 23). Actually, this construction was already discussed, long before Samuelson, in Arrow (1963, p. 31), where it was noted that a violation of independence would typically follow.

In contrast, the weakness of this 1983 paper is that it does not clarify further the connection between the BSF and (I). Is it just the case that the BSF may or may not satisfy the condition, or is it rather the case that it negates it in a determinate way that would then justify introducing specific weakenings such as Mayston's and Pazner's? This is all the more disappointing as not much later than the 1963 edition of his monograph, Arrow considered relaxing (I) in a positive light, and even went some way in the direction of accepting non-utility comparisons mediated by "irrelevant alternatives".[29]

7 Conclusion

We have reached the end of this re-examination of a seemingly obscure controversy. Although eventually siding with the losers' side against the winners', we have tried to emphasize both camps' successes and failures. The welfare economists made some very poor defensive moves—e.g. Bergson's habit of answering technical points with the hollow claim that "judgments of value should be made", or Samuelson's intellectual resistance to the multi-profile framework and persistent inconsistency in his assessment of independence. The social choice theorists were generally correct when they made mathematical points, and their work actually contains most of the pieces of the puzzle that we have rearranged in a different way. But they were poor exegetes of those semi-formal concepts which originated in the pre-Arrovian period, like the "Bergson–Samuelson welfare function" and "ordinalism", offering for them strong implausible readings that made their *reductio* a simplistic winning game. If there is something to salvage from the old concepts, as we believe, this can be done only by applying the formal methods of social choice theory to the actual content of welfare economics, which for different reasons, neither the welfare economists nor their mainstream opponents managed to do.

To exploit Bergson's hint that a "rule of equity" must be invoked to compare indifference curves, together with Mayston's or Pazner's weak independence conditions, is a promising alley for welfare economics. There is no reason why the alternative direction of exploiting interpersonal comparisons of utility should be denigrated; it is rather a question of correcting the current imbalance in research. The current theories of fairness or equity have thus far had a limited impact on

[29] Witness this remarkable statement: "I now feel... that the austerity imposed by this condition is stricter than desirable; in many situations we do have information on preferences for nonfeasible alternatives. It can certainly be argued that when available this information should be used in social choice... *The potential usefulness of irrelevant alternatives is that they may permit empirically meaningful interpersonal comparisons*" (Arrow 1967, p. 19, our emphasis).

public economics; this is because they have concentrated on Pareto-efficient allocations, and public economics must of course compare non-efficient allocations. By taking the shape of an explicitly defined BSF, these theories may be turned into proper tools for second-best evaluations. This heuristics may revitalize not only welfare economics, but public economics itself, which has often restricted attention to unspecified "social welfare functions" of the generalized utilitarian type.[30]

Acknowledgements We thank J. Bone, P. Hammond, S. Kolm, A. Lapidus, F. Maniquet, J. Roemer, D. Schmeidler, K. Suzumura, K. Tadenuma for helpful comments or discussions, and especially M. Salles. We also benefited from presenting the paper at conferences held in Strasbourg, 2003 and Boulogne-sur-Mer, 2004.

References

Arrow KJ (1963) Social choice and individual values. Wiley, New York (2nd revised ed.; 1st ed. 1951)
Arrow KJ (1967) Public and private values. In: Hook S (ed) Human values and economic policy. New York University Press, New York
Arrow KJ (1983) Contributions to welfare economics. In: Brown EC, Solow RM (eds) Paul Samuelson and modern economic theory. McGraw-Hill, New York
Atkinson A (2001) The strange disappearance of welfare economics. Kyklos 54:193–206
Bergson (Burk) A (1938) A reformulation of certain aspects of welfare economics. Q J Econ 52:310–334
Bergson A (1954) On the concept of social welfare. Q J Econ 68:233–252
Bergson A (1966) Essays in normative economics. Harvard University Press, Cambridge, MA
Bergson A (1976) Social choice and welfare economics under representative government. J Public Econ 6:171–190
Blackorby C, Donaldson D, Weymark J (1984) Social choice theory with interpersonal utility comparisons: a diagrammatic introduction. Int Econ Rev 25:327–356
Blackorby C, Donaldson D, Weymark J (1990) A welfarist proof of Arrow' theorem. Rech Econ Louvain 56:259–286
Bordes G, Hammond PJ, Le Breton M (1996) Social welfare functionals on restricted domains and in economic environments, mimeo, Stanford University
Chipman JS, Moore JC (1978) The new welfare economics, 1939–1974. Int Econ Rev 19:547–584
d'Aspremont C (1985) Axioms for social welfare orderings. In: Hurwicz L, Schmeidler D, Sonnenschein H (eds) Social goals and social organization. Cambridge University Press, Cambridge
Deaton AS, Muellbauer J (1980) Economics and consumer behaviour. Cambridge University Press, Cambridge
de Van Graaf J (1957) Theoretical welfare economics. Cambridge University Press, Cambridge
Fishburn PC (1970) Utility theory for decision making. Wiley, New York
Fishburn PC (1973) The theory of social choice. Princeton University Press, Princeton
Fleurbaey M (2003) On the informational basis of social choice. Soc Choice Welf 21:347–384
Fleurbaey M (2005) Health, wealth and fairness. J Public Econ Theory 7:253–284
Fleurbaey M, Hammond PJ (2004) Interpersonally comparable utility. In: Barbera S, Hammond PJ, Seidl C (eds) Handbook of utility theory, vol. 2: extensions. Kluwer, Boston
Fleurbaey M, Maniquet F (1996) Utilitarianism versus fairness in welfare economics, forthcoming In: Salles M, Weymark JA (eds) Justice, political liberalism and utilitarianism: themes from Harsanyi and Rawls. Cambridge University Press, Cambridge
Fleurbaey M, Maniquet F (2001) Fair social orderings, mimeo, U. of Pau and U. of Namur
Fleurbaey M, Maniquet F (2002a) Help the low-skilled or let the hard-working thrive? A study of fairness in income taxation, mimeo, U. of Pau and U. of Namur

[30] Studies of tax issues in Fleurbaey and Maniquet (2002a,b) and Fleurbaey (2005) exemplify both this kind of heuristics and some of its preliminary results.

Fleurbaey M, Maniquet F (2002b) Fair income tax, forthcoming in Rev Econ Studies
Gravel N (2001) On the difficulty of combining actual and potential criteria for an increase in social welfare. Econ Theory 17:163–180
Hammond PJ (1976) Why ethical measures of inequality need interpersonal comparisons. Theory Dec 7:263–274
Hansson B (1973) The independence condition in the theory of social choice. Theory Decis 4:25–49
Kemp MC, Ng YK (1976) On the existence of social welfare functions, social orderings and social decision functions. Economica 43:59–66
Kemp MC, Ng YK (1977) More on social welfare functions: the incompatibility of individualism and ordinalism. Economica 44:89–90
Kemp MC, Ng YK (1982) The incompatibility of individualism and ordinalism. Math Soc Sci 3:33–37
Kemp MC, Ng YK (1983) Individualistic social welfare functions under ordinalism: a reply to Mayston. Math Soc Sci 4:305–307
Kemp MC, Ng YK (1987) Arrow's independence condition and the Bergson–Samuelson tradition. In: Feiwel GE (ed) Arrow and the foundations of the theory of economic policy. New York University Press, New York
Le Breton M, Weymark JA (1996) An introduction to Arrovian social welfare functions on economic and political domains. In: Schofield NJ (ed) Collective decision-making: social choice and political economy. Kluwer, Dordrecht
Little IMD (1952) Social choice and individual values. J Polit Econ 60:422–432
Little IMD (1999) Collections and recollections: economic papers and their provenance. Clarendon Press, Oxford
May KO (1952) A set of independent necessary and sufficient conditions for simple majority decisions. Econometrica 20:680–684
Mayston DJ (1974) The idea of social choice. Macmillan, London
Mayston DJ (1979) Where did prescriptive welfare economics go wrong? In: Currie DA, Peters W (eds) Contemporary economic analysis, Proc. AUTE Conference, vol. 2. Croom Helm, London
Mayston DJ (1982) The generation of a social welfare function under ordinal preferences. Math Soc Sci 3:109–129
Mongin P (1999) Normes et jugements de valeur en économie normative. Soc Sci Inf 34:521–553
Mongin P (2000) Does optimization imply rationality? Synthese 124:73–111
Mongin P (2002) Is there progress in normative economics? In: Boehm S, Gehrke C, Kurz HD (eds) Is there progress in economics? Knowledge, truth and the history of economic thought. Edward Elgar, Cheltenham
Parks R (1976) An impossibility theorem for fixed preferences: a dictatorial Bergson–Samuelson welfare function. Rev Econ Stud 43:447–450
Pazner E (1979) Equity, nonfeasible alternatives and social choice: a reconsideration of the concept of social welfare. In: Laffont JJ (ed) Aggregation and revelation of preferences. North-Holland, Amsterdam
Pazner E, Schmeidler D (1978) Egalitarian equivalent allocations: a new concept of economic equity. Q J Econ 92:671–687
Pollak RW (1979) Bergson–Samuelson socialWelfare functions and the theory of social choice. Q J Econ 93:73–90
Roberts KWS (1980) Social choice theory: the single and multi-profile approaches. Rev Econ Stud 47:441–450
Samuelson PA (1947) Foundations of economic analysis. Harvard University Press, Cambridge, MA
Samuelson PA (1967) Arrow's mathematical politics. In Hook S (ed) Human values and economic policy. New York University Press, New York
Samuelson PA (1977) Reaffirming the existence of 'Reasonable' Bergson–Samuelson social welfare functions. Economica 44:81–88
Samuelson PA (1981) Bergsonian welfare economics. In: Rosefielde S (ed) Economic welfare and the economics of Soviet Socialism: essays in honor of Abram Bergson. Cambridge University Press, Cambridge
Samuelson PA (1987) Sparks from Arrow's Anvil. In: Feiwel GE (ed) Arrow and the foundations of the theory of economic policy. New York University Press, New York
Sen AK (1970) Collective choice and social welfare. Holden Day, San Francisco

Sen AK (1971) Choice functions and revealed preference. Rev Econ Stud 38:307–317 (reprinted in Sen (1982))
Sen AK (1972) Interpersonal aggregation and partial comparability. Econometrica 38:393–409 (reprinted in Sen (1982))
Sen AK (1977) On weights and measures: informational constraints in social welfare analysis. Econometrica 45:1539–1572 (reprinted in Sen (1982))
Sen AK (1982) Choice, welfare and measurement. Blackwell, Oxford
Sen AK (1986) Social choice theory. In: Arrow KJ, Intriligator MD (eds) Handbook of mathematical economics, vol. 3. North-Holland, Amsterdam
Suzumura K (1999) Paretian welfare judgments and Bergsonian social choice. Econ J 109:204–220
Suzumura K (2001) Pareto principles from Inch to Ell. Econ Lett 70:95–98
Weymark J (1998) Welfarism on economic domains. Math Soc Sci 36:251–268