Spurious unanimity and the Pareto principle

Working paper

SPURIOUS UNANIMITY AND THE PARETO PRINCIPLE

Paper presented at the Conference on Utilitarianism, New Orleans,
March 1997

Philippe MONGIN,
Centre National de la Recherche Scientifique

1. Introduction

The present paper is a critique of the Pareto principle, one of the building blocks of traditional welfare economics as well as an influential principle in recent normative economics and moral philosophy. The principle states that if the members of society express the same preference judgment between two options, this judgment is compelling for the social observer (who may be either a policy-maker or a moral observer, depending on the interpretation). More technically, the *weak* version of the principle stipulates only that unanimous strict preference judgments should be respected, and the *strong* version has two parts: first, unanimous indifference should be respected (this part is called Pareto-indifference); second, when weak preference is unanimous and at least one individual's preference is strict, then the observer should entertain that individual's strict preference (this is sometimes called strict Pareto). We are concerned here with both the weak and strong versions.

To debunk the Pareto principle, we will adopt the following roundabout strategy. First, we need to distinguish it from other unanimity principles with which it is often - in our opinion incorrectly - identified. This preliminary groundwork capitalizes on some of the recent work in normative economics (section 2). Second, we argue that for the chosen interpretation of the principle, preference unanimity alone is not sufficient to compel the social observer. Unanimity without unanimous reasons - what we call spurious unanimity - has no normative standing (section 3). Next, we investigate various cases of spurious and non-spurious unanimity in detail. The major cases involve the Bayesian model of rational choice under uncertainty which we show leads to a proliferation of spuriously unanimous preferences (section 4). There is an impossibility theorem in social choice theory to the effect that Bayesian preferences not do aggregate consistently. Because this impossibility theorem leads to a powerful argument against the Pareto principle, we compare it with the dismissive argument based on spurious unanimity (section 5). There are examples, again borrowed from the uncertainty context, where even non-spurious preference unanimity is not compelling. We argue this point while discussing the (normatively spurious) probabilistic analogue of the Pareto principle (section 6). Finally, we discuss spurious preference unanimity in the remaining contexts of risk and certainty, with a view of showing that it can also occur in these contexts (section 7).

At this point, a hopefully complete argument against the Pareto principle will have been produced, since the following two conclusions will then be established: spurious preference unanimity, for one, deprives the principle from its alleged universal force, and for another, can occur in any informational context in which the principle is used. The extension of the critique beyond the case where the individuals entertain different

subjective probabilities on the set of states of the world is a delicate, and probably novel, step in the discussion of the Pareto principle. One already existing position in normative economics is that the principle does not apply to preferences over uncertain prospects when subjective probabilities differ. The point of the present paper is in part to extend the critique of the Pareto principle beyond this existing position. For another part, it is to comfort the negative conclusion already obtained for the uncertainty context by using different, possibly more direct arguments.

We would like to emphasize at the outset that nowhere in the paper does our critique depend on assuming that all or some of the members of society are irrational. It would be too easy to deny the principle that preference unanimity is compelling by pointing out that individuals sometimes have spurious reasons, or have no reasons at all, for their preference judgments. We are concerned here with ideally reflective individuals who can defend their preference judgments in terms of statable reasons. It is not the lack or the inadequacy of individual reasons, but their *distribution* across society, which leads to what we call spurious preference unanimity. To formalize rationality we will employ the standard models of contemporary decision theory. Our concept of rationality under uncertainty is the Bayesian one, and more particularly, Savage's (1972). To take the Bayesian framework for granted is a self-consented limitation of the present analysis.

2. The Pareto principle versus other unanimity principles

We first need to distinguish between several common interpretations of the Pareto principle. This preliminary analysis will lead to separate the Pareto principle proper, which is the topic of this paper, from other related unanimity principles. Following one common view the former is a particular rendering of the *democratic sovereignty principle* : political decisions are legitimate to the extent only that they originate in the people. The members of society are the source of the law which applies to them; if they do not themselves legislate, they should appoint their legislators, who will be accountable to them. When political scientists and political philosophers discuss the Pareto principle, this is often done in connection with democratic sovereignty. Viewed in this way, the principle seems to be basic and unexceptionable. Disagreements among the citizens raise difficulties for democratic politics, but agreement does not; what the citizens decide unanimously should automatically be implemented.

The political reading of the Pareto principle equivocates on the meaning of individual preferences. Plainly, what one decides for oneself is not necessarily what one prefers. This simple point had to be made repeatedly by philosophers against those economists who defend "revealed preference theory" and thus confuse one kind of decision - i.e., a choice between alternatives - with preference.[1] The dismissal of "revealed preference theory" appears to be relevant here, even if the context of discussion is a non-economic one. In the political realm there is the further complication that the citizens do not decide for themselves but really *propose* that something should be decided. This further enlarges the gulf between preference and decision. The connection between the two does not become tighter since a proposal is made unanimously. A candidate X can conceivably be unanimously elected to be President not only by those who prefer X to competitors W and Z, but - typically, for some strategic reason - also by those who

[1]See in particular Sen's critique (1982, ch.2).

prefer Z to X and those who prefer W to X; among X's electors there might even be people for which X was the worst of all three candidates. Voting theory has made this sort of cases familiar. The choice of X as a president will be legitimized by the democratic sovereignty principle, but *not* by the Pareto principle. The latter simply does not apply to the case, since the electors were not unanimous in their preferences.

Thus, the political interpretation of the Pareto principle relies on an equivocation. Arrow (1951) uses a formalism of binary relations which he explicitly wants to be interpretable not only in terms of preferences, but also in terms of votes for candidates or political motions. Following the latter interpretation his famous impossibility theorem uncovers a potential dysfunctioning of democracy. This way of looking at Arrow's work is respectable, and indeed standard in political science. It is just unfortunate that this political interpretation is stated in terms of the Pareto principle, which, again, typically does not apply to the context of democratic decision-making. Notice that Arrow's work antidates, and in some sense was superseded by, the work of those theorists who emphasized the strategic side of the vote. If we ask whether spurious unanimity of preferences, as in the example of X above, creates a problem for the democratic sovereignty principle, the answer is probably that it does, but not to the point of calling democratic sovereignty into question. The issue should really be phrased in terms of spurious *majority* - a familiar occurrence in the history of western democracies. However, these political issues are beyond the scope of a paper concerned with the Pareto principle proper.

The Pareto principle is often discussed in an interpretation in which preferences refer to *tastes*. For instance, see again Arrow (1951-1963, p.71-72). At least, this is a defensible interpretation for the word "preference": many of my preference judgments just express my tastes (when I say that I prefer wine to beer, or vacationing in the mountains to vacationing on the seaside). The taste interpretation of the Pareto principle underlies the ill-defined but influential "consumer sovereignty" doctrine, and can probably be traced back to Pareto himself (although his exposition of Pareto optimality in the *Manuel d'économie politique* (1906) is notoriously obscure). Despite the air of respectability that the "consumer sovereignty" doctrine enjoys, one wonders whether one can go very far in normative economics with just a taste interpretation of preference. By themselves individual tastes do not carry much normative force. That I prefer reading a book to doing gymnastics of course does not mean that I should indulge in bookreading at the expense of doing physical exercises. Clearly also, this taste-based preference does not warrant a claim on social resources. I cannot just mention my *taste* for books to claim more resources for libraries or the publishing business; that would not be the proper argument to make. The normative force of individual tastes does not change all the sudden when they happen to be shared unanimously. Thus, the Pareto principle does not seem to have much to recommend itself in a taste interpretation of preferences. Contrary to the political interpretation, this one does not distort the meaning of preference, but it does not make unanimous preference compelling for the social observer.

What the writers on consumer sovereignty were groping for is probably that the satisfaction of one's tastes contributes towards one's welfare even if it is - of course - not the whole of it. It is one consideration to be balanced against others in determining the individual (and derivately, the social) welfare or well-being. Contrary to the satisfaction of tastes, the realization of individual welfare carries with it considerable normative force. Accordingly, the principle recommending to maximize all the members of society's welfare whenever this does not raise any mutual incompatibility deserves to be

taken very seriously. This principle can be stated as follows: whenever option x conduces to more welfare for all members of society than does option y, then the social observer take option x rather than y. A weak and a strong version of this *principle of compatible welfare* can be devised, exactly as for the standard Pareto principle. It is far from being ethically indisputable, as the lively critique of "welfarism" in current normative economics demonstrates, but it is a principle worth serious consideration.[2]

It is important to realize that the principle of compatible welfare is not simply a variant of the Pareto principle. The latter is formulated in terms of unanimous preference for x over y, not in terms of x enhancing everybody's welfare more than y. As Samuelson (1947-1983, p.223) puts it concisely, the Pareto principle stipulates that "individuals' preferences are to count".[3] There is a world of difference between one formulation and the other. This point would perhaps not deserve emphasis if welfare economists had not obscured it for a long time. A once respected welfare economist, de Graaff wrote bluntly that
"a person's welfare map is defined to be identical with his preference map" (1957, p.5),
an extraordinary pronouncement if one thinks of it. It seems painfully obvious that what I prefer is not necessarily what everything considered, will increase my well-being. I prefer continuing to smoke, even if stopping would on the whole be better for me. ("On the whole" indicates that allowance should be made for the fact that my taste for smoking will be frustrated.) Today's welfare economists are more aware of the conceptual gap between preference and welfare than was de Graaff in his time. However, they are still busy drawing conclusions about the society's welfare from reasonings constructed only in terms of preferences. For instance, the claim is often made that under a wide range of circumstances, the opening of a new market increases the overall welfare. It would be more accurate to rephrase the conclusion like this: if the individuals were given the opportunity to compare, all would rather have the economy with the new market than without. A Pareto improvement does not necessarily amount to a welfare improvement.

Thus, once again, it seems as if the Pareto principle draws its persuasive force from another principle to which it is somehow related, but which is distinct from it. Is there a genuine version of the principle that could be normatively defended? To be genuine this version should be stated in terms of preferences; to be defensible, it should involve something besides preferences.[4] An exclusively preference-based version of the Pareto principle will fail to be compelling, exactly as in the particular case in which preferences were equated with tastes. Suppose that all the members of society *prefer* a

---

[2]Recent reviews of the welfarism debate include Hausman and MacPherson (1996), Roemer (1996), and Mongin and d'Aspremont (1997).

[3]It is worth quoting the whole passage from the *Foundations*: "A more extreme assumption, which stems from the individualist philosophy of modern Western Civilization, states that individual preferences are to "count". If any movement leaves an individual on the same indifference curve, then the social function is unchanged, and similarly for an increase or decrease. Actually, an examination of the principles of jurisprudence, the folkway and mores, shows that in its extreme form this assumption is rarely proposed. Even "sane" adults are not permitted to eat and drink what they think best, individuals cannot sell themselves in order to consume more in the present, ...But economists in the orthodox tradition have tended to consider the above cases as exceptions" (1947-1983, p.223).
We interpret this passage by a prominent economic theorist (who in 1983 revisited his work largely approvingly) as confirming two things: (i) for what it is worth, the Pareto principle is stated in terms of preferences; (ii) it is undefended within economics.

[4]Compare this assessment with Broome's (1991, p. 155). He also contrasts a political and a (roughly speaking) welfarist interpretation of the Pareto principle. But contrary to him, we give a chance - at least for the sake of the argument - to a unanimity principle stated in terms of preferences.

certain bridge to be built rather than not built; or they *prefer* itinerary x to be chosen rather than y for the new motorway around the capital. Why should this unanimous preference be compelling? One reason why in the absence of further information, it cannot be compelling has already been mentioned: the common preference judgment might very well be based only on the members of society's tastes, and tastes by themselves do not normatively compel to anything; they are just one consideration to be balanced against others. Once built, the bridge will redirect commercial relations in ways that the individuals might not have considered when forming their preferences. The alternative itineraries x and y have distinctive environmental consequences which, again, the individuals might not have considered. All this suggests that preference data are not by themselves sufficient information to incline the social observer one way or another. Hence, the Pareto principle, in the preference interpretation, cannot be defended unless it is modified to take *also* some non-preference information into account. This critical point is closely related to Sen's at one stage of his critique of "welfarism", notably in "Utilitarianism and Welfarism" (1979).[5]

What kind of non-preference information must be added to qualify the Pareto principle appropriately is not at all easy to say. The "post-welfarist" literature does provide interesting (though non-unanimous and non-compelling) guidance. However, given the critical purpose of this paper, we need perhaps not investigate this literature in detail. We can follow a short-cut here, and just reason on what appears to be a necessary condition for *any* defensible version of the Pareto principle proper (i.e., preference-based): it should imply that not only individual preference data but the*reasons for these preferences* are both known to the social observer and used by him to make his preference judgment. Our notion of "reasons" here include conformity to the formal principles of decision theory, such as transitivity, the "sure-thing principle" of Bayesian decision theory, and the like, but it goes far beyond that. We mean *substantial* reasons as well - i.e., reasons which matter because of their content. How reasons are to be delineated depends on the particular context of evaluation. In the next section we specialize in a context where the reasons for a preference judgment are the utility values for the consequences of each action, and the probabilities of states of nature. To bring home the point about substantial reasons, we will not be concerned solely with the formal notion that the individuals base their preferences on *some* pair of probability and utility functions, but also with the *values* taken by these probability and utility functions.

3. Spurious unanimity defined and excluded

Any defensible version of the Pareto principle will have to satisfy the following normative claims:
(i) the observer should have reasons for ranking the unanimously preferred alternative above the other;
(ii) taken in themselves, individual preference comparisons between social alternatives do not constitute a reason for the observer to rank one alternative above another;
(iii) the observer's reasons are derived from considering no more than the individuals' preference comparisons between social alternatives and their respective reasons for these comparisons.

---

[5]Sen's (1985) later critique extends beyond this point by suggesting that welfare information might be not so much insufficient as altogether irrelevant.

Claims (i) and (ii) essentially formalize the discussion of the previous section, but claim (iii) also involves a novel consideration. In whatever version, the Pareto principle should include an *individualistic assumption*. (Sometimes in the past, it was even "defended" on the *sole* basis of individualism; this was a defective argument, of course.[6]) In essence, individualism means that the social observer is non-creative. His reasons for preferring one option to another should be based only on the knowledge of relevant individual "items". That is all that individualism implies. What is specific to claim (iii) is that in accordance with the previous discussion, it includes the individuals' reasons among the relevant "items". The conventional (but inadequate) version of the Pareto principle, in the preference-based version, presumably hinges on (i), the negate of (ii), and the following strengthening of (iii):
(iii') the observer's reasons are derived from considering no more than individual preference comparisons.

Unanimity of preference comparisons can prevail either with or without the individuals' reasons for these comparisons being themselves unanimous. We contrast these two cases as *genuine* and *spurious* preference unanimity, respectively, and proceed to argue that the Pareto principle is not compelling in the case of spurious preference unanimity. This appears to follow straightforwardly from points (i), (ii) and (iii) above. Since the observer can take into account no more than the individuals' preference comparisons and reasons for these, and preference comparisons by themselves are not reasons, only the individuals' reasons can provide the observer with reasons for his own preference ranking. But when spurious unanimity prevails, the individuals' reasons conflict with each other; they simultaneously point in different directions. In the absence of an arbitration principle, which would exceed the content of (iii), there is no way to decide between these conflicting arguments. Hence, spurious unanimity does not compel the social observer to assent to the commonly expressed preference. Of course, it does not commit him to the opposite preference either; the observer's deliberative process has no determinate outcome.

Spurious unanimity can be illustrated as follows. The social alternatives to be evaluated are to construct a bridge (x) across a river which is also the border of the country or not to construct it (y). The country - call it L - happens to be poor, though not miserable. It can claim century-old traditions of handicraft and unsophisticated but relatively environment-friendly agriculture, as well as peaceful social interaction. The neighbouring country - call it T - happens to be one of these fast-growing, quickly westernized countries which have conquered access to the amenities of modern life at the expense of jeopardizing their traditions and damaging the environment. One group of individuals in L believe that the more likely consequence of bridging the two countries with each other is that theirs will be swamped with an influx of products, people and techniques from T, and the old, backward way of life which still plagues L will hopefully disappear before the next generation. The remaining group in L believe that this country is too small to be a significant stake for T's businessmen, so that commercial and technological invasion is excluded; the limited extra flows across the bridge will just bring about a little more convenience to the poor people of beautiful L without jeopardizing their valuable way of life. We assume that there is a well-intentioned international organization which is eager to build the bridge as soon as the people in L agree, so that costs are not an issue between the two camps. We claim that the L people's unanimous preference for having the bridge built carries no normative

[6]Hence the label "individualism" in some classic papers of early welfare economics.

weight; the international organization should not base a case for funding the bridge works on these people's spurious unanimity.[7]

4. How spurious unanimity proliferates in the uncertainty context

More formally, let us introduce some basic notions of subjective expected utility (SEU) theory. (The latter is also commonly referred to as Bayesianism; indeed, Bayes' rule for revision will also be used at some point in the argument below.) In SEU theory the objects of preference are *acts*. The latter are defined to be functions from states of the world $s \in S$ to consequences $c \in C$; let us denote them by $a \in A$. This formalization neatly captures the notion that the individual is uncertain about the consequences of his actions. Which consequence c occurs is determined by the joint data of his act a and the state s, which he does not know beforehand. Preferences over consequences are a particular case of preferences $\leq$ over acts, since a consequence c can be identified with the act having constant value c. Axiomatic versions of SEU theory - Savage's (1954-1972) being the most elaborate among them - imply the theorem that there exists a representation of $\leq$ in terms of an expected utility functional $\sum U(a(s))P(s)$ or $\int U(a(s))dP(s)$, where U(.) is a function on C and P(.) a probability measure on S. The function U(.) is to be referred to as the individual's *utility function* (implicitly: over consequences), and P(.) as the individual's *subjective probability*. Both items are well-identified because - this also follows as a theorem from any standard axiomatization of SEU- U(.) is unique up to positive affine transformations and P(.) is absolutely unique.[8]

Take a two-person society in which the preference of each member i=1, 2 over acts can be represented by a SEU functional with utility function U^i and probability P^i. We make the technical assumption - let us denote it by (A) - that $P^1 \neq P^2$ and that there are two consequences c, c' such that $U^1(c)>U^1(c')$ and $U^2(c)<U^2(c')$. We make the further technical assumption - call it (B) - that for some event E, the two individuals strongly disagree in their relative probability assignments of E and E^c (this denotes the complementary event of E), that is:
$P^1(E)>P^1(E^c)$ and $P^2(E)<P^2(E^c)$.
Now, consider acts a and b defined as follows:
$a(s)=c$ if $s \in E$ and $a(s)=c'$ if $s \in E^c$;
$b(s)=c'$ if $s \in E$ and $b(s)=c$ if $s \in E^c$.
A straightforward expected utility calculation shows that the two individuals will agree in preferring a to b:
$P^i(E)\, U^i(c) + P^i(E^c)\, U^i(c') > P^i(E)\, U^i(c') + P^i(E^c)\, U^i(c)$, for i=1, 2.
Remarkably, they have *opposite reasons* to conclude that a is preferable to b. Individual 1 thinks that E is strictly more likely than E^c, and c strictly preferable to c'; individual 2 thinks that E^c is strictly more likely than E, and c' strictly preferable to c. By the magic of the expected utility calculation, the final evaluations turn out to be concordant. This is the paradigmatic case of spurious unanimity: unanimity prevails in terms of

[7]The acute reader will find out which Asian countries these initials refer to. Kadaré's (1981) beautiful novel contains a different, but not totally unrelated illustration of the ambiguous effects of bridging two countries with each other.

[8]We exclude the famous complication of state-dependent utilities (i.e., when the utility evaluation of consequences depends on the state in which these consequences occur, and when each consequence is perhaps not available in each state).

preferences over acts, but not in terms of the reasons justifying this common preference. This is the bridge example in algebraic language.

The diagnostic that unanimity is spurious depends on how one delineates the individuals' reasons for preferring an act over another. Some alternative rationales for both 1 and 2 preferring a to b would perhaps imply that the individuals also agree at the level of their underlying reasons. We do not deny this possibility. But since we assume throughout a SEU model of rationality, this sets the standard for what counts as a reason and what does not. Following the SEU logic there are only two considerations that can be called upon to justify or dispute one individual's ranking of acts - i.e., his probability assignments to events and his utility assignments to consequences. The distinction between probability and utility assignments is a particular rendering of the commonsensical distinction between beliefs and desires. Nonbayesian theories either imply that beliefs and desires are mapped onto technical concepts other than probability and utility, or involve agregation rules more complex than the additive procedure followed in SEU theory, or both. We suspect that spurious unanimity cases will occur in *every* theory which takes desires and beliefs to be the reasons for acting, treats them as being independent variables, and is intended to cover a sufficiently rich domain of acts. (The last two conditions essentially mean that every conceivable configuration of desires and beliefs is made available). In other words, we suspect that spurious unanimity examples could be constructed for each of the existing nonbayesian modellings of preference under uncertainty, but we do not investigate this claim here.

Since SEU theory covers, among others, the configurations of utilities and probabilities implied by spurious preference unanimity, the diagnostic is well-established. We can strengthen our case by saying something about the "frequency" of the paradoxical configurations. As it turns out, spurious unanimity cases literally proliferate in Savage's framework of Bayesianism. This can be checked at the same time as one shows that Savage's axioms make assumption (B) redundant with (A). Savage's (P6) (1954-1972, p.39) imposes constraints on the state space and the individual's preference relation which are eventually reflected in the property that the derived subjective probability is *nonatomic* . Nonatomicity says that any event with positive probability can be partitioned into two complementary subevents having themselves positive probability. Formally, if $P(E)>0$, there exist two disjoint events E' and E" such that $E' \cup E" = E$ and $P(E')>0$ and $P(E")>0$. It is actually permissible to take $P(E') = \alpha P(E)$ for any α strictly between 0 and 1. Applying the nonatomicity property, it is not difficult to show that (A) alone will lead to cases of spurious unanimity. This is worth emphasizing since the probabilistic assumption in (A) is much weaker than in (B). The former condition simply requires that P^1 and P^2 be distinct. The beauty of Savage's axioms is that they automatically imply that *distinct* subjective probability measures lead to strong probabilistic disagreements as in (B). Also, the construction shows that for no individual does spurious unanimity have zero measure; in particular, there are infinitely many cases of spurious unanimity.

5. Spurious unanimity and the impossibility of consistent Bayesian aggregation

We have gone far enough to be able to compare the present argument against the Pareto principle and the much discussed impossibility theorem that Bayesian preferences cannot be aggregated consistently. Here is an informal statement of the theorem relative

to a multi-agent Savage framework; for technical details the reader is referred to Mongin (1995).[9] Suppose that a group of individuals i= 1,...,n obey Savage's axioms of preference under uncertainty, and accordingly, that their preferences can be represented in terms of a SEU functional with a utility function U^i and a nonatomic probability P^i. Suppose also that there is an observer, to be referred to by index i=0, who registers the individuals' preference judgments and attempts to aggregate them. By assumption, this observer's preferences are *consistent* with the individuals', which means two things: (C1) they satisfy the same decision-theoretic axioms (here: Savage's) as those satisfied by individual preferences; (C2) they satisfy the Pareto principle in either the weak or strong version. The theorem states that if some "technical" conditions - akin to (A) above - are also satisfied, then the weak Pareto version leads to some form of dictatorship, and the strong Pareto version to sheer logical impossibility. (Two forms of dictatorship are considered: probability dictatorship, when the observer's probability P^0 coincides with one of the individuals' probability, and utility dictatorship, when the observer utility U^0 is identical to one of the individuals' utility up to a relevant transformation.) The theorem can be interpreted as saying that the two consistency conditions overdetermine the observer's aggregation rule. It is typically discussed in terms of which of the conditions should be dropped. The majority's view seems to be that (C1) has a higher normative standing than (C2); these writers recommend to resolve the conflict by keeping the former and dropping the latter.

The usual argument for reaching this conclusion is that (C1) is a rationality requirement on the observer's preferences, and thus indispensable in any normative context of application. The problem with this way of arguing is that it makes the conclusion *against* (C2) contingent on an argument *for* (C1). Strong as it seems, condition (C1) has to be defended. Even if Bayesian rationality is taken to be the standard for the individuals' rational preference, it remains to show that the social observer too should be subjected to it. This is why the consistency principle (C1) is introduced - a uniformity principle which is independent of which rationality standard is selected for the individuals.

One advantage of the present critique is perhaps that it is quicker in reaching the same negative conclusion against the Pareto principle. It does not depend on arguing for the observer's Bayesianism, but only on making the points that the observer should have reasons for his preference judgment and that these reasons must be inferred from considering the individuals' own reasons. On the other hand, the present argument is not as sharp as the existing one, which is structured by a mathematical statement and has the simple form of a dilemma. That (C1) and (C2) logically conflict with each other under relevant circumstances is an important result of collective choice theory. Some readers will be inclined to see the present argument as comforting rather than replacing the existing one. It is good to have further reasons for dropping (C2) if they turn out to be independent from those adduced for adhering to (C1). This is the case with the spurious unanimity argument.

There is a further lesson to be drawn from the discussion of spurious unanimity: it tentatively suggests a partial remedy to the failure of the Pareto principle in the uncertainty context. Instead of dropping (C2) altogether, why not first try to restrict it to

[9]There are alternative formulations of this impossibility theorem. Seidenfeld, Kadane, and Schervish (1989) provide a two-person illustration within a more accessible axiomatic framework than Savage's. See the further references, as well as the extensive discussion, in Mongin and d'Aspremont (1997, section 5).

cases of non-spurious unanimity? Suppose that this can be done meaningfully; suppose also that somehow, the restricted principle eschews the impossibility-of-consistent-Bayesian conclusion; then, all objections would vanish, and at least some form of Paretianism could be salvaged. Accordingly, we should explore the following tentative principle:

(*) In the uncertainty context, unanimous preference comparisons are compelling for the observer if and only if the individuals agree on both the probability and utility rankings underlying these preference comparisons.

6. The uncertainty context continued, and the difficulty aggravated

Principle (*) is tailor-made to address the spurious unanimity objection. Remarkably, it is also appropriate to circumvent the impossibility of consistent Bayesian aggregation. The latter theorem says that (C1) and (C2) clash with each other*under relevant technical assumptions*. These assumptions might go unnoticed but turn out to be crucial to the proof. Roughly speaking, they amount to saying that the Pareto principle will apply to at least one case of genuine and at least one case of spurious unanimity. That the aggregative rule is degenerate, or cannot exist, then follows from applying the principle to the genuine and the spurious cases in succession, and drawing the consequences. This proof is blocked once the spurious unanimity case is excluded. Two-person examples neatly illustrate this logic.[10]

More than that is true. When spurious unanimity is excluded by assumption from the collective Savage framework, consistency in the double sense does not overdetermine the observer's preference anymore. It can be checked that an aggregative rule exists, and that it is non-dictatorial. This sketch of resolution will be of interest to those writers[11] who have suggested that in the uncertainty context, the Pareto principle should be suitably restricted rather than abandonned altogether. Principle (*) provides a restriction that is both precise and technically effective. However, it will soon turn out that a resolution in terms of (*) is unconvincing. Spurious unanimity and the impossibility of consistent Bayesian aggregation are not the only objections that can be raised against the Pareto principle in the uncertainty context. We are now to raise an altogether different problem. It will be shown to hit principle (*).

To introduce this further objection we need to make a detour. We will first discuss the following, purely probabilistic unanimity principle:

(**) If the individuals all agree that event A has greater probabiility than event B, so should the social observer.

This unanimity principle can be made precise in a weak and a strong versions, exactly like the Pareto principle. It has an interest of its own, as witnesses the abundant literature on probability aggregation in statistics and management science. For instance, a widely circulated survey by Genest and Zidek (1986) discusses axiomatizations of the

[10]The technical argument for this paragraph is explained in Mongin (1995, Example 3, p.331) who use the weak version of the Pareto principle. Compare with Broome's (1991, p.152) two-person example, which relates to the strong version.

[11]Among them is Levi (1990) in his comment of Seidenfeld, Kadane and Schervish (1989).

"linear pooling rule": for any set of individual probability measures P^1, ..., P^n, the observer's probability measure is some convex combination of P^1,..., P^n. The "linear pooling rule" can be axiomatized variously, but it should be clear that it satisfies at least the weak version of principle (**). When every coefficient in the convex combination is positive, the strong version is satisfied too. The "linear pooling rule" is but the probabilistic counterpart of generalized utilitarianism (i.e., utilitarianism with possibly different nonnegative weights for the individuals).

We claim that principle (**) cannot be normatively compelling. This is shown by means of the following urn example.[12] Suppose there is an urn which has many black balls and only a few white balls. There are two individuals and an observer; they all know the proportions of white and black balls. Each individual will be asked to draw a ball (with replacement). He will just observe the result of his own drawing; the observer does not observe the result of either drawing. After each individual has drawn a ball, they are asked by the observer to compare the probabilities of a given event E and its complementary E^c. Having obtained the two individuals' answers, the observer makes his own probability assessment of E. The design of this experiment is common knowledge to the three participants. Each individual's probabilistic comparison is just reported to the observer, not to the other individual.

We take event A_1 (A_2) to mean that individual 1 (resp. 2) has drawn a white ball. Remember that A_1 and A_2 have very low prior probability. The event E of interest to the observer is that the total number of white balls drawn by the two individuals is even (hence, either 0 or 2). Now, suppose the two individuals are unanimous to say that E is less probable than E^c:

$P^i(E) < P^i(E^c)$, i=1,2.

Should the observer endorse this unanimous judgment? On the contrary, he should conclude that:

$P^0(E) > P^0(E^c)$.

The argument is as follows: given the informational assumptions, the observer should conclude that each individual will state that E is less probable than E^c *if and only if this individual has drawn a white ball*. (If the individual draws a white ball, he reasons that his ball is likely to be the only white ball drawn, which makes E less probable than E^c. If he does not draw a white ball, he thinks it likely that no white ball at all has been drawn, which makes E more probable than E^c.) In view of the unanimous judgment that E is less probable than E^c, the observer should conclude that each individual has observed a white ball, and thus that E is *more* probable than E^c. This counterexample defeats the probabilistic variant (**) of the Pareto principle. As a by-product, it leads to questioning the rationality of the popular "linear pooling rule".

The counterexample relies on a Bayesian updating argument which is potentially very general. Difficulties for principle (**) are likely to arise any time that: (a) the individuals have different private information and update their prior probabilities by conditioning them on this information; (b) the observer has no direct access to the individuals' private information but has some indirect access to it by observing their updated probabilities; and (c) the individuals do not communicate their updated

[12] This example, or a closely related variant, was suggested to us by Ed Green. We are grateful to him, as well as David Schmeidler, for illuminating conversations on the principle of probabilistic unanimity.

probabilities to each other. To see the role of condition (c), suppose that the updated comparisons:
$P^i(E) < P^i(E^c)$, i= 1,2,
are public knowledge between the individuals. Then, each should reason in the way the observer did in the previous paragraph, and conclude that the other has drawn a white ball. Each should conclude that two white balls have been drawn. Accordingly, the individuals will make a second updated comparison opposite to the first:
$P^i(E) > P^i(E^c)$, i= 1,2.
A moment's thought shows that if these inequalities are in turn made public between the individuals, they will remain unchanged, and that the social observer should now endorse them. Hence, the probabilistic Pareto principle is not defeated anymore. This illustrates the role of (c), and more generally, of a careful phrasing of the informational assumptions.

More complicated examples would perhaps involve a sequence of changes in the two individuals' posterior probabilities, the first updated comparisons bringing about a second updating once they are made public, the second updated comparisons a third updating, and so on. In similar examples, to make the first updated comparisons public between the individuals is probably not enough to remove the objection against principle (**). We conjecture that for any given n, if successive updatings are cross-communicated only to the order n, a relevant probabilistic space can be devised to defeat (**). Only when one assumes the *infinite* sequence of updated probabilities to be cross-communicated can principle (**) be in safety. This conjecture is closely related to Aumann's (1976) famous proof that if two agents share the same prior probability distribution and their posterior probabilities are common knowledge, then these probabilities must be equal.

We needed the negative argument just made about probabilistic unanimity in order to discuss *preference* unanimity. The conclusion that in the uncertainty context, the Pareto principle is inadequate will now follow from the lemma that under certain conditions, the Pareto principle implies (**), which has just been said to be inadequate. Actually, since we are particularly interested in assessing the restricted version (*) of the Pareto principle, we will prove the lemma in a slightly stronger form: we will show that the restricted version (*) implies (**). The conditions for this implication to hold are, for one, that principle (*) is non-vacuous and, for another, and crucially, that not only the individuals *but also the observer* satisfy SEU theory. The lemma is very easy to check. It revolves around the familiar fact that given a well-chosen normalization of the utility function, the value of the subjective probability is equal to the expected utility value of a suitably selected act. This fact was put to use in Ramsey's early work on SEU, where acts are interpreted as bets; it does not depend on adopting a particular axiom system, Savage's or any other. For the sake of completeness we footnote a little proof for a two-person society and the weak version of the Pareto principle.[13]

---

[13]Principle (*) implies that *if* there are a partition E, E^c and a pair of consequences c, c' such that: $P^i(E)\,U^i(c) + P^i(E^c)\,U^i(c') > P^i(E)\,U^i(c') + P^i(E^c)\,U^i(c)$, for i=1, 2,
with $P^i(E) > P^i(E^c)$ and $U^i(c) > U^i(c')$ for i=1, 2,
so that the two individuals agree on their reasons as well as on their preference,
then the social observer endorses the common preference, i.e.: $P^0(E)\,U^0(c) + P^0(E^c)\,U^0(c') > P^0(E)\,U^0(c') + P^0(E^c)\,U^0(c)$.
The last inequality depends on assuming that the observer himself satisfies subjective expected utility theory. Now, principle (*) holds nonvacuously. This implies in particular that there exist consequences d,

7. Risk and certainty

The difficulties surrounding the Pareto principle in the uncertainty context have long been recognized by some writers in welfare economics. The so-called ex post school (e.g., Hammond, 1982) rejects the Pareto principle at the stage where uncertainty prevails (the "ex ante" stage) but accepts it at the stage where uncertainty is resolved (the "ex post" stage). The school maps this temporal distinction onto two versions of the Pareto principle; it rejects the ex ante principle while upholding the ex post principle. (Notice in passing that the ex post school authors are explicitly concerned with the Pareto principle proper, not with the principles of democratic sovereignty or compatible welfare. If needed, their writings would further testify to the relevance of analyzing preference-based interpretations of the principle, like those selected in this paper.) Although they are not always very precise about this, the crucial dividing line for them appears to be between differing and identical subjective probabilities rather than between two temporal stages of analysis. The case in which probability values are *given* by a commonly perceived chance mechanism (as in the economic theory of risk) would be absolutely unproblematic to ex post writers. More generally, the Pareto principle at the ex ante stage becomes acceptable to them if subjective probabilities have somehow been already agreed upon between all agents in the model, including the observer.[14]

Despite its label, the ex post Pareto principle is intended to serve at the ex ante stage. The school adheres to Bayesianism firmly. It claims that the social observer's ex ante preferences should, like the individuals', satisfy the Bayesian axioms. The ex post principle is intended to determine the observer's utility function U^0 in terms of the individuals' U^1, ..., U^n.[15] The derived function U^0 will have to be combined with P^0, the observer's subjective probability, in the way prescribed by SEU theory. This is all that the school has to contribute. Importantly, it does not say how P^0 is to be determined. At best, it mentions the reasonable but rather vague principle that the observer should take all relevant information into account in forming his subjective probability; this may or may not include taking into account the individuals' subjective probabilities as manifested by their own ex ante preferences. That ex post school writers do not determine the observer's probability in any way is the major difference between their position and the tentative reformulation (*) of the ex ante Pareto principle. Remember that (*) implies the principle of probabilistic unanimity (**), a stringent constraint on

---

d' such that $U^i(d) > U^i(d')$ for i=1, 2. Applying principle (*) to the acts having constant values d and d' respectively, we also have that $U^0(d) > U^0(d')$, and the following joint normalization of utility functions becomes available:

$U^i(d) = 1$ and $U^i(d') = 0$ for i= 0, 1, 2.

To see that the weak version of (**) holds, assume that for some event E:

$P^i(E) > P^i(E^c)$, for i=1, 2.

By the chosen normalization, this is a particular case of a unanimous expected utility statement with unanimous reasons, and the observer's expected utility statement implied by (*) can be reexpressed as:

$P^0(A) > P^0(A^c)$.

[14]Evidence for this claim is provided by the ex post writers' occasional use of Harsanyi's framework of expected utility, in which probabilities are always taken to be identical between agents, including the social observer. See Mongin and d'Aspremont (1997, section 5) for details and references.

[15]The ex post principle will typically imply that U^0 can be expressed as an increasing function of U^1, ..., U^n. This requires mild regularity assumptions on the U^i.

P^0.[16] The ex post school avoids the pitfalls of probabilistic unanimity but at the cost of being very little informative. It can also foster a misunderstanding that we would like to dispel. To adopt the ex post Pareto principle does *not* imply that all ex ante preference information becomes irrelevant. Think again of the urn example: in this case, it is useful for the observer to know the individuals' preferences between some acts, since he can infer from these preferences what the individuals' subjective probabilities of the event of interest E are, and the example demonstrates that he can know for certain whether E occurs or not just from knowing these two probabilities! This shatters a point sometimes made in favour of the ex post school resolution - that it is informationally economical.

Recently, the ex post resolution has been criticized as being shallow.[17] A common theme to these criticisms is that it would be inconsistent to stop the case against the Pareto principle just when the ex post stage is reached. In the Bayesian decision-theoretic model consequences are taken to be *ultimate* consequences but this is just a definition or a conventional device. Consequences are not final in any real sense. They should rather be seen as aggregates of further uncertain consequences, each of which could be analyzed in turn, and so on ad infinitum. In effect, consequences partake in the nature of uncertain prospects, or acts, to which the ex post school does not want to apply the Pareto principle; hence an inconsistency. This critique strikes us as powerful. It would be instructive to formalize it within the context of Savage's theory in order to connect it with some of the formal raised above. However, we do not pursue it because the emphasis of this paper is put differently. The critical argument of section 3 was intended to be perfectly general. Nothing in the way in which it is stated suggests that uncertainty should be its only case of application. Spurious unanimity can also occur in the remaining contexts of risk and certainty. Since the ex post school preserves the Pareto principle for these two contexts, it should fall prey to the spurious unanimity argument.

The following variant of the bridge example illustrates how the argument may work in the certainty context. There are two consequences of bridging countries L and T with each other, and there is no uncertainty whatever about these consequences. One is that the traditional way of life will disappear from L; the other is that more convenience will be brought to the poor people in L. If it were possible, group 1 ("the cynicals") would like to bring about consequence 1 without consequence 2 (say, because the poor will make heavier demands on political participation once better off and this is considered to be disruptive). If it were possible, group 2 ("the sentimentalists") would like to bring about consequence 2 without consequence 1. So the two groups have conflicting preferences in a very precise sense. However, they will both support the bridge project if group 1 weighs the benefit of consequence 1 more heavily than the cost of consequence 2, and group 2 weighs the cost of consequence 1 less heavily than the benefit of consequence 2. Should the social observer endorse the common preference for the bridge project? We think not. The abstract argument of section 3 applies here: since the observer should base his own reasons on the individuals' reasons, and the latter are opposite, he would need an arbitration principle to decide, but that is excluded by Paretian logic. The observer should decide whether consequences 1 and 2 are mostly beneficial or detrimental, and nothing in the Pareto principle determines how this decision should be made. Exactly as in the uncertainty case, we cannot exclude that the

[16] In this connection we note the following consequence of Savage's framework: (**) is not only a necessary, but also a sufficient condition for the "linear pooling rule" (Mongin, 1995, Propositions 1 and 2).

[17] See in particular Broome (1990, 1991) and Hilde, Jeffrey and Risse (1996).

observer will after all agree with the spuriously unanimous preference judgment. If this, rather than the opposite issue, turns out to be the case, it must be justifiably because the observer has endorsed one of the two systems of reasons - i.e., either the cynics' or the sentimentalists'. The two systems of reasons cannot be both endorsed by the observer; although he outwardly agrees with the common preference, his own judgment not depend on applying any unanimity rule at all.

It is an interesting problem why leaving aside the delicate issue of preference externalities[18], economists have generally not been willing to admit that the Pareto principle met objections in other contexts than uncertainty. One possible explanation is that uncertainty is the only context where they are prepared to analyze the agent's preferences in terms of more primitive factors. But elsewhere, preferences are just what they are provided that they satisfy certain formal requirements. Economics, it is said, does not have to inquire about the individuals' tastes or values or whatever underlies the people's preferences. Not only is this methodological principle dubious in itself but there are hints that it has been violated even in conventional economics. For instance, the micro-theorist Lancaster has suggested that the consumer's actual objects of preferences are not commodities as empirically described, but rather vectors of underlying "characteristics". The analysis of the bridge example can be rephrased in Lancasterian language. Neither the cynicals nor the sentimentalists are interested in the bridge as such; rather, they are concerned with a vector of relevant features (we initially call these features consequences) which they weigh very differently. To analyze preferences in terms of characteristics is to go one step towards what we think would be the most appropriate analysis - the analysis in terms of the *reasons* underlying individual preferences.

8. Summing-up

We have mentioned three arguments against the Pareto principle in the preference-based interpretation which we argued is the relevant one to investigate the economists' claims:
(1) Bayesian preferences do not aggregate consistently;
(2) Probabilistic unanimity is not compelling for the social observer; still, the principle that it is follows from the Pareto principle;
(3) Spurious preference unanimity (i.e., when the common preference results from opposite reasons), is not compelling for the social observer; but the Pareto principle does not distinguish between spurious and non-spurious preference unanimity.
Argument (1) has been reviewed extensively elsewhere, and was here mostly for the sake of comparison. At least explicitly, (2) and (3) appear to be novel arguments in the discussion of the Pareto principle. Argument (2) leads to a clear-cut dismissal, but exactly like (1), it is limited to the uncertainty context, and it needs the assumption that not only the individuals but also the social observer are Bayesian. As to (3), it is not as sharp as the other two arguments, which rely on mathematical facts, but it is applicable outside the uncertainty context and it does not make heavily specific demands on the observer's rationality. This argument is closely related to one of the critiques of "welfarism" in the recent literature.

[18]We refer here to Sen's "liberal paradox" and the wide literature stemming from it.

REFERENCES

Arrow, K.J. (1951), *Social Choice and Individual Values,* New Haven, Yale University Press; 2nd revised edition, 1963.

Aumann, R. J. (1976), "Agreeing to Disagree", *The Annals of Statistics*, 4, 1236-1239.

Broome, J. (1990) "Bolker-Jeffrey Expected Utility and Axiomatic Utilitarianism", *Review of Economic Studies*, 57, 477-502.

Broome, J. (199l), *Weighing Goods,* Oxford, Blackwell.

Genest, C. and J. V. Zidek, "Combining Probability Distributions: A Critique and an Annotated Bibliography", *Statistical Science*, 1, 1986, 114-148.

Graaff, J. de Van (1957), *Theoretical Welfare Economics*, Cambridge, Cambridge University Press.

Hammond, P.J. (1982), "Utilitarianism, Uncertainty and Information," in A. Sen and B. Williams (eds.) *Utilitarianism and Beyond*, Cambridge, C.U.P., ch. 4, 85-102.

Hausman, D. and M. McPherson (1996), *Economic Analysis and Moral Philosophy*, Cambridge, Cambridge University Press.

Hild, M., R. Jeffrey, and M. Risse (1996), "What it Takes to Be a Group: Pareto vs. Diversity", mimeo, Department of Philosophy, Princeton University.

Kadaré, I. (1981), *Le pont aux trois arches*, Paris, Fayard (translated from the Albanian).

Levi, I. (1990), "Pareto Unanimity and Consensus", *Journal of Philosophy*, 87, 481-492.

Mongin, P. (1995a), "Consistent Bayesian Aggregation," *Journal of Economic Theory,* 66, 131-351.

Mongin, P. and C. d'Aspremont (1997), "Utility and Ethics", in *Handbook of Utility Theory*, S. Barbera, P. Hammond and C. Seidl (eds.), Dordrecht, Kluwer, forthcoming.

Pareto, V. (1905), *Manuel d'économie politique*, in *Oeuvres Complètes*, 7, Genève, Droz, 1966.

Roemer, J. (1996), *Theories of Distributive Justice*, Cambridge, Mass., Haravard University Press.

Samuelson, P. A. (1947), *The Foundations of Economic Analysis*, Cambridge, Mass., MIT Press; new edition, 1983.

Savage, L.J. (1954), *The Foundations of Statistics*, New York, Dover; 2nd revised edition, 1972.

Seidenfeld, T., J.B. Kadane and M.J. Schervish (1989), "On the Shared Preferences of Two Bayesian Decision Makers," *Journal of Philosophy,* 86, 225-244.

Sen, A. (1973a), "Behaviour and the Concept of Preference," *Economica,* 40, 241-259.

Sen, A. (1979) "Utilitarianism and Welfarism", *Journal of Philosophy*, 76, 463-489.

Sen, A. (1985), *Commodities and Capabilities,* Amsterdam, North Holland.