CHAPTER 24

SOCIAL EVALUATION UNDER RISK AND UNCERTAINTY

PHILIPPE MONGIN AND MARCUS PIVATO

24.1 Historical Background and Overview of the Chapter

Perhaps surprisingly, uncertainty plays no role whatsoever in the classical works on social welfare. The two fundamental theorems of welfare economics say that any general competitive equilibrium is an optimum, in Pareto's sense, and by a kind of converse, that any Pareto optimum can be decentralized as a general competitive equilibrium. These famous results presuppose that both consumers and producers are certain of, and agree on, what exactly they trade. Admittedly, Debreu (1959, ch. 7) redefined commodities so as to include uncertain states of nature into their description, for instance, counting an umbrella when it rains and an umbrella when it is sunny as if they were two distinct commodities, but this formal trick extends the validity of the original results only if there is a separate market for each commodity so redefined, and this is clearly not the case.

To take another area and another classic, Arrow ([1951] 1963) stated his impossibility theorem for an unstructured set of social alternatives, and the natural interpretation one can give of this formalism is that the individuals who rank the alternatives are certain of and agree on their description. This limitation of analysis comes as no surprise if one realizes that Arrow's work, innovative though it was, also depended on the "new welfare economics" of his time.

However, in these early years, an original thinker was running against the tide. Two papers by Harsanyi (1953, 1955) opened the way to a theory of social welfare under uncertainty. Their common idea was to construct a social welfare function (SWF) from the decision theory that von Neumann and Morgenstern ([1944] 1947) had newly developed. The first paper sketches in a few lines what has come to be known as the

Impartial Observer Theorem, and the second, more detailed paper reviews this result and introduces another, later to be called the Social Aggregation Theorem. The two pieces did not obviously belong to either welfare economics or social choice theory, and it took time to recognize that they had given a new turn to both fields, not to mention the neighboring area of formal ethics.

When it comes to uncertainty, economists distinguish between choice situations in which probability values already exist to represent the uncertain phenomenon (e.g., playing roulette), and choice situations in which this is not the case (e.g., betting on a horse race). Economists have made a habit of referring to the former as "risk" and to the latter as "uncertainty." This clashes with ordinary language, which takes "risk" to mean the possibility of a loss, and "uncertainty" to mean that various outcomes are possible, depending on the realized state of nature. However, the above two choice situations need distinguishing, and we will follow the received terminology for lack of a better one.[1]

Von Neumann and Morgenstern (vNM) deal only with risk, unlike Savage ([1954] 1972), who proposes a decision theory for uncertainty. In both cases, the problem is to derive the classic expected utility formula from a suitable set of preference axioms. Faithful to the economists' idea of risk, vNM take the probabilities as being part of the description of the objects of preference (referred to as "lotteries"). What they derive from their preference axioms are the utility function underlying the expected utility formula and that very formula itself. In contrast, faithful to the economists' idea of uncertainty, Savage does *not* include the probabilities in the objects of preference (there referred to as "prospects"). He rather takes these objects as being mappings from states of the world to consequences. His preference axioms are powerful enough to derive *both* components of the expected utility formula—that is, a probability function and a utility function—at the same time as this formula itself.

Economists attach canonical importance to Savage's demonstration. For them, uncertainty eventually reduces to risk, in the sense that even when probabilities are not usually available, as in horse betting versus roulette playing, they nonetheless exist implicitly. Agents are supposed to act *as if* they relied on them to compute an expected utility. "Subjective probabilities" and "subjective expected utility" are the received expressions for this scenario.[2] However, the economists' reductive move preserves a major difference between risk and uncertainty. In the former, all agents use the *same* probabilities, whereas in the latter, the implicit probabilities typically vary from one agent to the next.[3]

[1] When we revert to the ordinary use of "risk" or "uncertainty," we will make this clear.

[2] Symmetrically, economists say "objective probabilities" and "objective expected utility" in the case of risk, but we eschew this terminology to avoid confusion with "objective probabilities" in the frequentist sense.

[3] Decision theorists have recently considered more radical formalisms of uncertainty, where subjective beliefs are represented by entire *sets* of subjective probabilities (or by generalized probability measures called *capacities*). Since economists associate the word "uncertainty" with Savage, these developments go under the different rubric of "ambiguity."

Harsanyi's work is concerned with risk. Both in the Impartial Observer and Aggregation Theorems, he builds on vNM theory to obtain a SWF that is additive in terms of the individuals' utility functions. He then claims that this SWF is utilitarian, and that he has thus provided the old doctrine of Bentham, Mill, and Sidgwick with a modern preference foundation. Sen (1969, 1974) and followers have opposed this claim for technical reasons connected with vNM theory. Their claim should carefully be distinguished from the classical ethical objections directed at utilitarianism itself, which we will not review here. Although the chapter explains Sen's objections, it emphasizes another, which emerged later in the literature: *neither theorem extends well from the von Neumann–Morgenstern to the Savage framework.* Since economists take the latter to be the canonical one, this challenges the use of the theorems for any purpose, and not only to defend utilitarianism.

We will use another organizing distinction, this time from welfare theory. When uncertainty (in the ordinary sense) prevails, individual preferences can be considered either ex ante or ex post, that is, either before or after uncertainty is resolved. Accordingly, the Pareto principle—to the effect that society should respect unanimous individual preferences—can be defined in two ways, and similarly with social welfare criteria. The ex ante criterion ranks social prospects by applying the ex ante Pareto principle, and the ex post one ranks them by applying the ex post Pareto principle along with some decision-theoretic principle for society. As we will explain, *the two criteria can be endorsed together under risk, but not under uncertainty*. This limitation is essentially equivalent to the failure of Harsanyi's theorems to extend from risk to uncertainty. Having fully described the problem, we will examine the solutions, that is, adopting the ex ante criterion alone, adopting the ex post criterion alone, or devising a compromise between them.

Historically, Harsanyi was attacked on still another front. Supposing, *pace* Sen, that his additive SWF can be given a proper utilitarian sense, he would be open to egalitarian objections against utilitarianism. These objections first appeared in Rawls (1971) and were partly endorsed by Sen ([1970] 1979, 1976). They assume the certainty context, but Diamond (1967a) extended them to risk, and in this way challenged egalitarians themselves: how should their doctrine apply to risk and even uncertainty? The ensuing literature has shown that equality was amenable to both ex ante and ex post criteria that *clashed with each other* no less than the corresponding social welfare criteria do (this comes out by way of examples rather than theorems, as in the latter case). Having thus delineated another major problem, we will review its solutions, which can again be described as purely ex ante, purely ex post, or a compromise between them.

The two major debates of the chapter originate in the Social Aggregation Theorem. Meanwhile, the Impartial Observer Theorem had its own separate developments. Section 24.2 touches on the latter briefly, referring to more thorough coverage elsewhere (Pattanaik 1968; Mongin, 2001; Adler, 2014). Then, section 24.3 returns to the main thread by stating the Social Aggregation Theorem and discussing Sen's objections. Section 24.4 explains the collapse of this theorem under uncertainty, links this result to the conflicting ex ante and ex post social welfare criteria, and reviews the positive solutions. Sections 24.5 and 24.6 unfold similarly, the former explaining the conflict between ex ante and ex post egalitarian criteria, and the latter reviewing the positive solutions.

24.2 The Impartial Observer Theorem

The Impartial Observer Theorem is sometimes attributed to Lerner (1944–47) and Vickrey (1945, 1960), but Harsanyi (1953, 1955) is more explicit about it, although he does not yet give it the form of a theorem. It is only late (in Harsanyi 1977a, 48–55) that it reaches this stage, but the earlier versions also matter, not least because they aroused a famous controversy with Rawls (1971, III.27 and III.28).

Suppose that some poor person is asked about the desirable income distribution and propounds one heavily favoring the poor. From the outside, it is impossible to say whether or not this person is trying to abstract from his interests and inclinations, as would be required by the mental disposition of impartiality. But now suppose that the person knows neither his position on the income ladder, nor any other personal features that may influence his judgment. Then, the possible causes of bias having disappeared, his answer can be declared to be impartial. This is only a de facto sense, because it is still unclear whether the relevant mental disposition is present, but as it seems, it is a minimal ethical basis to determine what income distributions are just for society.[4]

According to many epistemologists, especially the Bayesians, situations of partial ignorance are to be analyzed in probabilistic terms. The classic form of this strategy, Laplace's Principle of Insufficient Reason, gives equal probabilities to the possible realizations of the unknowns—here, the personal features that may bias judgment. One may think of an observer facing an equal chance lottery, with the unknowns playing the role of outcomes. Now, if vNM theory applies to the evaluation of this lottery, one obtains a SWF that averages the individual utility values for income. It has the mathematical form of the average rule of utilitarianism, and by a further step, it may be taken to be utilitarian indeed.[5]

This is about the stage at which Harsanyi (1953, 1955) left the argument. Although vague, it may already be attacked from several angles. The objections against Laplace's Principle are nearly as old as the principle itself, and some of them were revived by Rawls (1971, 161–74), who also objected to the use of vNM theory, and more basically, of probability theory to represent ignorance.[6] The famous maximin principle, which the "original position" entails for any observer put under the "veil of ignorance,"

[4] The impartiality approach to distributive justice is compatible with various ethical standpoints. It can be analyzed in teleological terms, as in the utilitarian tradition claimed by Harsanyi, or in deontological terms, as in the Kantian tradition claimed by Rawls, and in still other ways.

[5] The sum rule and average rule of utilitarianism differ only if one considers variable populations, which is not the case here. Harsanyi (1977c, 1979) emphasizes the latter over the former only for broader philosophical purposes.

[6] Rawls hardly mentions Harsanyi, but it is transparent that he argues against him.

embodies simultaneous rejection of all three features. When they are disentangled from each other, theoretical possibilities emerge that have escaped consideration in the Rawls-Harsanyi debate.[7]

The above sketch does not say what personal features, over and beyond income distribution, must be ignored to secure impartiality. Harsanyi's successive formulations differ regarding precisely this basic question. His 1953 note is equivocal, but his 1955 paper suggests, and his 1977a book asserts, that *the observer must ignore his own preferences*. Meanwhile, Rawls had published his "veil of ignorance," which cancels even more of the observer's information, and this might have influenced Harsanyi despite their continuing disagreements. It seems commonsensical that to ignore one's preferences for luxuries (for example) will make one's distributive judgments more impartial, in the de facto sense explained. However, ignoring one's preferences altogether is like giving up one's personal identity, and it then becomes unclear *who* is now passing the judgments in question. Harsanyi's (1977a, 50) claim that individuals entertain both "personal" and "moral" preferences amounts to begging the question. Thus, while clarifying his notion of impartiality, he inadvertently opened the Pandora's box of personal identity metaphysics. Rawls (1971, 173–74) was probably the first to raise an objection based on personal identity against Harsanyi despite having to face one himself in the "veil of ignorance" construction.

Harsanyi's mature work combines his demarcation of ignorance with Laplace's Principle and vNM theory. The essential trick consists in introducing *extended alternatives* and *extended lotteries* as new objects of preference. The former are pairs (x, P_i), where x is any social state of affairs (an income distribution as a particular case), and P_i summarizes all information (numerical or otherwise) relative to individual i's preference. The latter are lotteries constructed from these alternatives (i.e., they attribute probabilities to extended alternatives regarded as possible outcomes). Prominent among extended lotteries are the *equiprobable* ones (here defined for a society of n individuals and in standard notation):

$$L_x = (1/n, (x, P_1); \ldots ; 1/n, (x, P_i); \ldots ; 1/n, (x, P_n)).$$

With these lotteries, the observer has an equal chance of experiencing a social state with the preferences of any member of the society. Importantly, they rise to prominence by an *epistemological* argument based on Laplace's Principle and not by an *ethical* argument based on the value of equality. The vNM axioms may now be applied to *extended preferences*, that is, preferences bearing on extended lotteries, and particularly on the main objects of concern, that is, the L_x.[8]

[7] Harsanyi's (1976) collection of essays is in part on this debate. A little known rejoinder by Rawls (1974) is discussed there.

[8] We have generalized the informal discussion slightly by assuming that extended alternatives take the form (x, P_i) rather than (x_i, P_i), where x_i is i's position in the income distribution x. This permits i being concerned about the social state of affairs as a whole. Another generalization is needed for the formal statement below. There x is replaced by a lottery p on all possible x, and an equiprobable (meta)lottery L_p is defined accordingly, i.e., $L_x = (1/n,(p,P_1); \ldots ;1/n,(p,P_i); \ldots;1/n,(p,P_n))$.

Harsanyi is indebted here to Sen ([1970] 1979, ch. 9*), who had introduced extended alternatives for the certainty context. There the notion already strikes one as being problematic. By a standard argument, to compare (x,P_i) and (y,P_j) for all x, y, i, and j, the observer draws upon the ability that humans share of being *empathetic* to each other.[9] However, some of the preference features of i or j may be inaccessible to the observer, meaning that there is no cognitive basis for his empathetic involvement. For instance, Alice can take up some relevant features of Cleopatra or Shakespeare, but presumably not all of them, and not even the same for both. The problem of personal identity surfaces again, and none of the later solutions has thus far won the day.[10]

Harsanyi takes the role of empathy for granted when he postulates that the observer's preferences between (x,P_i) and (y,P_i) coincide with i's own preference between x and y. This "principle of acceptance" was buried in the early sketches but can now be expressed formally. It has the same individualistic basis as the Pareto principle of welfare economics, that is, an individual's preferences should be respected when he is the only one concerned, but it is arguably weaker (see Mongin 2002; and Grant et al. 2010). Let us now sum up the construction into a formal statement (corresponding to Harsanyi 1997a, 65–66).

The Impartial Observer Theorem

Suppose that

(V) *The observer's preference satisfies the vNM axioms over the set of extended lotteries;*

(Eq) *Social welfare comparisons between any two lotteries p and p' over social states coincide with the observer's preference between the corresponding equiprobable lotteries L_p and $L_{p'}$; and*

(Ac) *The individuals have well-defined vNM preferences over lotteries p over social states, and these preferences and the social observer's jointly satisfy the acceptance principle.*

Then, there exists a SWF W of the following form: for all lotteries p,

$$W(p) = \frac{1}{n}\big(U_1(p) + \ldots + U_i(p) + \ldots + U_n(p)\big), \tag{24.1}$$

where for all i, U_i is a vNM utility function representing i's preferences.

This conclusion relative to lotteries p would be unnecessarily strong if one were only concerned with evaluating the social states of affairs x, but it is needed if one is to evaluate policies with risky outcomes. As a weaker variant, if vNM theory applies only

[9] Harsanyi (1977a and [1977] 1982, unlike 1977b) refers to Smith's "impartial but sympathetic observer." This unfortunately confuses sympathy with empathy, two concepts that today's scholars distinguish; see Fontaine 1997 and Adler 2014. Similarly, some economic literature says "extended sympathy" (after Arrow [1951] 1963, ch. 8) where empathy is the relevant concept.

[10] Mongin (2001) sketches an answer, but Adler (2014) disputes it. Pattanaik (1968) implicitly, Kaneko (1984) explicitly, and perhaps also Gauthier (1986, 239–45), had raised the personal identity problem.

to the social observer, one gets equation (24.1) holding for all social states x instead of all lotteries p.

A brief reminder may be in order to clarify the technical concepts used above. A *vNM utility function* is one having the expected utility form when the objects of preferences are lotteries. If a preference relation over them satisfies the vNM axioms, it can be represented by such a function. Denote a typical lottery by $p = (p_1, x_1; \ldots ; p_k, x_k)$, where $p_1, \ldots, p_k$ are the probabilities assigned the possible outcomes $x_1 \ldots, x_k$. According to the vNM theorem, there exists a function U such that

$$U(p) = p_1 U(x_1) + \ldots + p_k U(x_k),$$

and U *represents* the preference relation, meaning that p is preferred to p' if and only if $U(p) > U(p')$. In this equation, we can use the same U on both sides because any outcome x can be identified with a lottery (i.e., the degenerate one giving probability 1 to x). The vNM theorem adds that $U(p)$ can be replaced in the equation by *positive affine transforms*, that is, functions $U'(p)$ of the form

$$U'(p) = aU(p) + b,$$

where a is positive and b is of any sign, and that these replacements are the only ones that preserve the equation.

In equation (24.1), the vNM theorem accounts directly for the fact that individuals have vNM utility functions $U_i(p)$, and indirectly for the fact that W is additive in terms of these functions.

As we suggested critically about Rawls, one may disagree with the conclusion of the Impartial Observer Theorem without questioning all assumptions at once. The recent trend is precisely to make piecemeal revisions. Like Rawls, all contributors discard Laplace's Principle of Insufficient Reasons, and this opens up two distinctive possibilities. One may either keep (Eq), hence the equal weights in (24.1), but reinterpret it as expressing the observer's *ethical* preference for equality, or remain on the *epistemological* plane and reject (Eq) as an inadequate rendering of ignorance. In our view, the former line defeats the purpose of the impartiality construction, which is to minimize the number of directly ethical postulates and make the most of the supposed ignorance situation. The latter line is more challenging, and it can be pursued—Rawls notwithstanding—within the confines of probability theory.

This can be done if the vNM axioms are replaced by Savage's—or a suitable variant of them, like Anscombe and Aumann's (1963)—so as to generate subjective probabilities over the P_i instead of the Laplacean equal probabilities. Karni (1998) and Mongin (2001) have independently proposed such a scheme—the former with more elaboration.[11] However, as the latter explains, beside delivering generally nonequal

[11] Karni's (1998) final representation is a variant of formula (24.1) in which each U_i is normalized so as to vary between 0 and 1; this is the so-called *relative utilitarian* SWF. A similar formula has been obtained in different ways by Dhillon and Mertens (1999), Segal (2000), Gajdos and Kandil (2008), and Pivato (2009).

weights, this revision brings out the difficulty of the "many impartial observers." That is, there will be as many SWFs as there are subjective probability distributions conceivable for the observer. In the absence of an argument for selecting one, the conclusion of the Impartial Observer Theorem remains indeterminate. *This is the sense in which it does not carry through from risk to uncertainty*. Pattanaik (1968) had noticed a related difficulty with the initial theorem: by imposing the vNM axioms on the observer's preferences, one implicitly fixes a risk attitude for him, and since it is a purely subjective feature, the resulting SWF must be nonunique.[12]

Another option—also within probabilistic confines—is to replace the vNM assumptions put on the observer by one of the recent nonexpected utility versions of decision theory. This makes it possible to handle some of the fairness objections that additive SWFs classically raise. Most contributions along this line are made in the context of the Social Aggregation Theorem, but Gajdos and Kandil (2008) specifically address the Impartial Observer Theorem. They redefine extended alternatives as being *sets* of probability distributions, a move that is typical of the current nonexpected utility work (as section 24.6 will also exemplify). One of their SWFs averages the minimum and the sum of individual utility values, thus arguably reconciling Rawls and Harsanyi.[13]

All these works proceed from a denial of Laplace's Principle, with the last one also rejecting the expected utility axioms. Elsewhere, the main effort has been to weaken the completeness part of assumption (V) in the Impartial Observer Theorem, that is, the assumption that the observer entertains preferences over *all* extended lotteries. To require such a large domain is to stretch empathy beyond what seems reasonable. Karni and Weymark (1998) and Safra and Weissengrin (2003) show that slimmer domains are mathematically sufficient for the proof. Grant et al. (2010) motivate their own domain restriction by arguing that the two polar cases of extended lotteries, that is, "outcome lotteries" (just over the x) and "identity lotteries" (just over the P_i), call for distinctive preference treatment. They separate the two classes while preserving vNM features, and eventually generalize (24.1) into the following: for all lotteries p over the x (the "outcome lotteries"),

$$W'(p) = f_1(U_1(p)) + \ldots + f_i(U_i(p)) + \ldots + f_n(U_n(p)), \tag{24.2}$$

where for all i, U_i is a vNM utility function, and f_i is a (possibly nonaffine) increasing transformation of U_i. In technical parlance, an *additively separable representation* has replaced the additive one.[14] Many writers have already converged to additively

[12] In a different reconstruction, Karni and Safra (2002b) suppose that any individual simultaneously entertains both a personal preference and a moral preference ("fairness relation"). They connect the two axiomatically to derive an aggregative utility representation by which the individual gives some weight to each.

[13] In a more elementary way, an average could be obtained by applying Jaffray's (1988) axioms for risky preferences. Relevant arguments also exist for the certainty context.

[14] See Grant et al. 2012 for more results, and Karni and Safra 2000 and Safra and Weissengrin 2003 for related anticipations. Space prevents us from discussing hypothetical ignorance positions that include mechanism design constraints, as in d'Aspremont and Gérard-Varet 1991 and Nehring 2004. More broadly, Gauthier (1988) has argued for a game-theoretic, rather than decision-theoretic, modeling of these positions. Regrettably, we must also leave out Dworkin's (1981) hypothetical insurance market.

separable representations starting from the other theorem, and we will assess them while commenting this result.

24.3 The Social Aggregation Theorem

Harsanyi compares his two contributions as follows: "[The Social Aggregation Theorem] yields a lesser amount of philosophically interesting information about the nature of morality than [the Impartial Observer Theorem] does, but it has the advantage of being based on much weaker—almost trivial—philosophical assumptions" ([1977] 1982, 48). Admittedly, it relies on no more than the Pareto principle and the vNM axioms, two bread-and-butter assumptions of economics. But as will turn out, Harsanyi uses them controversially.

A building block of welfare economics and social choice theory, the *Pareto principle* says that social welfare judgments should coincide with individual preferences when they are unanimous. More precisely, it comes into two parts:

Pareto indifference. *If all individuals are indifferent between two options, then so is the social observer.*

Strict Pareto. *If each individual either prefers option 1 to option 2 or is indifferent between the two, and at least one of them prefers 1 to 2, then the observer prefers 1 to 2.*

(We still designate the maker of social welfare comparisons by "the social observer," regardless of whether it is an individual or an institution, and without any empathetic connotation being now implied.)

The Social Aggregation Theorem applies these Pareto conditions to specific objects—*social lotteries* over *social states of affairs*—which we already encountered in the Impartial Observer Theorem. It also needs vNM theory applied to *both* the social observer and the individuals.[15]

The Social Aggregation Theorem

Suppose that

(V) *Individual preferences satisfy the vNM axioms over the set of social lotteries,*

(V*) *The social observer's preferences are also vNM, and*

(P) *The two sets of preferences jointly satisfy the Pareto principle.*

Then, for any set of vNM utility representations $W, U_1,\ldots, U_i,\ldots,U_n$, for the observer and the n individuals respectively, there exist positive constants $a_1,\ldots, a_i,\ldots, a_n$ and a constant

[15] As mentioned in section 24.2, a version of the Impartial Observer Theorem still holds even though only the observer satisfies the vNM theory. This difference is rarely noticed.

b such that for all social lotteries p,

$$W(p) = a_1 U_1(p) + \ldots + a_i U_i(p) + \ldots + a_n U_n(p) + b. \tag{24.3}$$

In its core version, the Social Aggregation Theorem only assumes Pareto indifference and derives the same formula (24.3), but with weights $a_1,\ldots,a_n$ of any sign. Adding Strict Pareto, one can select them to be positive, as a utilitarian interpretation would require. Importantly, even for a fixed profile of vNM utility functions W, $U_1,\ldots,U_n$, the weights $a_1,\ldots,a_n$ may not be unique (think of identical U_i).[16]

Since the weights do not have to be equal, the conclusion is still at a distance from the average or sum rules of utilitarianism. Harsanyi (1955) acknowledged this defect, which may explain his above claim that the assumptions are weaker than in the other theorem. He also realized that the weights varied with the profile of vNM utility functions, even raising a "many observers" point in advance: each observer comes up with his own set of functions, which determines his personal weighting of the individuals. Later, Harsanyi (1977a and 1977 [1982]) proposed to bridge the gap with (24.1) by selecting vNM utility functions for the individuals such that the weights are equal.[17] However, no assumption of the theorem justifies this move, and one should be sorry that Harsanyi repudiated his better insight that each observer was associated with a particular weighting.

At any rate, Harsanyi has always claimed that the Social Aggregation Theorem was a significant step to utilitarianism.[18] The difficulties surrounding this interpretation were pointed out early on by Sen (1969, 1974, 1977, 1986) and by Pattanaik (1968), who relied on Sen's first paper in the series. Although there are other possible accounts, we will separate three major objections.[19] This analysis works mutatis mutandis for the Impartial Observer Theorem, and if we focus on the Social Aggregation Theorem, this is only because it has a clearer structure.

The first objection, to the effect that weights vary with the chosen profile of vNM functions, has just been explained. A possible solution, which Harsanyi (1955, 1977a, and 1977b) also floated, is to complement the theorem with a *scheme of interpersonal comparisons*, so as to fix the utility ratios U_i/U_j for distinct individuals i,j in equation

[16] The weights are unique if and only if the U_i are algebraically independent. There is a preference rendering of this condition. For this and further technical properties, see Coulhon and Mongin 1989; Weymark 1993; and De Meyer and Mongin 1995. Danan, Gajdos, and Tallon (2014) investigate the Social Aggregation Theorem under Pareto indifference when the individuals and the observer can have incomplete vNM preferences. They use *sets* of vNM utilities to represent this incompleteness and characterize the observer's set in terms of utility by sums that generalize (24.3). See also the related analysis by Danan, Gajdos, and Tallon (2013).

[17] Since each individual vNM utility function is defined only up to positive affine transformations, it is trivial to change (24.3) and have equal weights.

[18] Harsanyi's own utilitarian interpretation of (24.3) relates to a society of coexisting individuals, but there is another one in which individuals represent successive generations (with the weights playing the role of discount factors). For some development of the Social Aggregation Theorem along this line, see Bommier and Zuber 2008, and for the intertemporal form of utilitarianism, see Blackorby, Bossert, and Donaldson 2005. Note in this connection that the theorem can be proved for infinitely many individuals (Zhou 1997).

[19] Compare with Weymark 1991; Mongin and d'Aspremont 1998; Mongin 2002; Weymark 2005; Fleurbaey and Mongin 2012.

(24.3). However, the scheme should be expressed in the same theoretical language as the other assumptions. This can be done by shifting to a richer framework of analysis. We eschew the technical details, just stressing that they have the effect of making the weights profile-independent.[20]

The Impartial Observer Theorem is afflicted with a related problem: the individual weights are meaningless if they are not accompanied with any uniqueness restriction on the utilities. In conclusion (24.1), the typical term $1/nU_i$ can be rewritten as $a_i U'_i$, where a_i is any positive number and U'_i is defined to be $(1/na_i)U_i$. With this rewriting, the same formula holds with both new weights and new vNM individual utility functions. With their more powerful axioms, the subjective probability variants permit fixing both the weights and utilities uniquely, and thus avoids the problem, though generally not putting the weights equal to $1/n$. If one insists on equal weighting, it is best to employ the techniques suggested for the Social Aggregation Theorem in the last paragraph.

The other objections amount to questioning the choice of vNM utility functions to represent the individual preference relations. Emphatically, when the vNM theorem concludes that the preference relation P submitted to the vNM axioms can be represented by a vNM function $U(p)$, it automatically permits representing P by *all increasing transforms* of $U(p)$, including the nonaffine ones, such as $(U(p))^3$, $\exp[U(p)]$, and so on. Indeed, if U is positive-valued, then the equivalence

$$p\ P\ q \text{ if and only if } U(p) > U(q)$$

is trivially equivalent to the equivalences

$$p\ P\ q \text{ if and only if } (U(p))^3 > (U(q))^3 \text{ if and only if } \exp[U(p)] > \exp[U(q)],$$

and so on. The theorem is sometimes misunderstood as saying that the permissible representations of P reduce to the positive affine transforms, such as $7U(p)$, $2U(p) + 3.14$, and so on. However, it only says that the positive affine transforms are *those alternative representations that preserve the expected utility form*—a much more limited statement.

The previous reminder is essential to understanding why Harsanyi's utilitarian interpretation of (24.3) is objectionable. Suppose that an agent's preferences over lotteries satisfy the vNM axioms and that this agent ranks lottery (1/2, \$1000; 1/2, \$0) above (resp. below) the sure outcome of \$500. Representing this comparison by one of his vNM function U, one gets

$$1/2U(1000) + 1/2U(0) > (\text{resp.} <) U(500), \textit{or}$$
$$U(1000) - U(500) > (\text{resp.} <) U(500) - U(0).$$

[20] The simplest solution is to reformulate the Social Aggregation Theorem within the so-called *multiprofile* framework of social choice theory (Mongin 1994; Blackorby, Bossert, and Donaldson 2006; Blackorby, Donaldson, and Weymark 2008). Danan, Gajdos, and Tallon (2013) apply the multiprofile framework to the aggregation of *sets* of vNM utility functions. Pivato's (2013) more complex framework permits deriving profile-independent weights, while allowing for incomplete preferences on the social observer's part.

The second line suggests that the agent can make comparisons of *preference intensities*—he seems to be able to decide if replacing \$500 by \$1,000 satisfies him more, or less, than replacing \$0 by \$500. However, since U is just one representation of P beside many others, the second line does not have the desired meaning unless it holds for *all* admissible utility representations of P, and it is typically violated if $U(p)$ is replaced by its nonaffine transforms. Hence, under the vNM axioms, the second line is a blank formal restatement of the first. Both carry information about risk attitudes, but neither says anything about comparisons of preference intensities. However, utilitarianism is precisely concerned with such comparisons. Only if they can be expressed, both at the individual and the interindividual level, does the average or sum rule make any sense.[21]

A related objection emphasizes the *welfare* concept. For the theorem to be ethically relevant, one would need to take vNM utility as a measure of individual welfare, but the vNM axioms do not warrant this interpretation (see especially Sen 1976, 248–51). Thus formulated, the objection strikes the *welfarist* content of the theorem more broadly than its *utilitarian* content. Indeed, utilitarianism adds up welfare numbers, but welfarism, as defined by Sen (1979) and others, only requires the social preference to be increasing with these numbers.

Important as these arguments are, they only make logical claims and only deliver expressions of doubt. They leave open the possibility that *a reinforced set of assumptions will invest vNM utilities with the desired measurement properties*, be they in terms of preference intensities or welfare. To go beyond the mere expression of doubt, one must dismiss this possibility, and this would be making a factual claim on top of the logical one. By its nature, such a claim would require a separate justification.

It is unclear whether Sen's critique stopped at the logical claim or went as far as the factual one, and it is also unclear how much of either claim Harsanyi was able to foreshadow. However understood, the critique has met with approval among welfare economists and social choice theorists, especially after Weymark's (1991) survey of the "Sen-Harsanyi debate" circulated. Our own position, which follows Mongin and d'Aspremont (1998), is more balanced. First, there now exist an array of axiomatic possibilities to exclude the nonlinear transforms of $U(p)$, and thus formally endow this function with cardinal significance.[22] Each of these systems must be evaluated on its own merits. Second, the most vehement critics unwittingly confuse the logical and factual objections—they conclude that $U(p)$ *cannot be* cardinal from the point that it *might not* be so. At best, they make a rhetorical appeal to plausibility. More seriously, they could invoke whatever negative experimental data decision theory has

[21] In one choice of language, utilitarianism needs a *cardinal* form of utility, whereas vNM theory is only *ordinal* in nature.

[22] This is done with increasing elaboration in Mongin and d'Aspremont 1998; Mongin 2002; and Fleurbaey and Mongin 2012. Weymark (1991, 2005) also has a positive solution, which he intends mostly for the Impartial Observer Theorem. In a more radical move, Harvey (1999) entirely dispenses with vNM theory and reformulates the Social Aggregation Theorem for cardinal preferences in the certainty case. Pivato (2014) extends Harvey's approach to a setting with incomplete interpersonal comparisons of utility.

made available.[23] However, it is questionable whether the present debate should take this empirical angle rather than be resolved at the theoretical level of comparing the different axiomatic additions.

This inconclusive discussion leaves untouched a significant remainder, which is *the additively separable SWF* that we already encountered with the Impartial Observer Theorem. Indeed, if one fixes completely general representations U_i^* of the individuals' vNM preferences, and a similarly general representation W^* of the observer's vNM preferences, conclusion (3) in the Social Aggregation Theorem becomes

$$F(W^*(p)) = f_1(U_1^*(p)) + \ldots + f_i(U_i^*(p)) + \ldots + f_n(U_n^*(p)), \tag{24.4}$$

where the f_i and F are increasing (typically nonaffine) transformations of the given representations. By stopping at (24.4), one avoids the questionable measurability assumptions needed to sum up utilities, but retains a feature of the sum, that is, trade-offs between utilities in any subgroup are independent of utilities in the complementary group. If utilities could be given a welfare interpretation, this would be a major insight into the structure of social welfare and a computationally helpful property for policy analysis. Section 24.6 further investigates additive separable SWFs in connection with egalitarianism.

24.4 Subjective Probability and the Problem of Ex Ante versus Ex Post Welfare

Both the Impartial Observer and the Social Aggregation Theorems run into difficulties when they are restated for uncertainty instead of risk—the former, as mentioned, because of the "many observers" problem, and the latter, as we will now see, because it collapses into a logical contradiction. Mongin (1995) axiomatizes this impossibility within the canonical framework of preference under uncertainty, that is, Savage's, and we follow this approach here (alternatives will be discussed below), starting with a further decision-theoretic reminder.

In vNM theory, the objects of preference are *lotteries*, which assign probabilities to outcomes. In Savage's theory, the primitives are states of the world and consequences, and the objects of preference are *uncertain prospects*, which assign consequences (just a variant terminology for "outcomes") to the states without specifying probabilities for the latter. An intellectual tour de force, Savage's representation theorem shows that probabilities are available nonetheless. It says that, if an agent's preference relation over uncertain prospects satisfies seven axioms, then it admits an expected utility

[23] See the early data summarized in Bouyssou and Vansnick 1991. These authors also discuss the problem of "cardinalizing" vNM theory in the context of Allais's (1953) decision-theoretic work.

representation in terms of a specific pair of utility function and probability distribution. Formally,[24] let us denote the possible states of the world by $s_1,\ldots,s_k$ and a typical uncertain prospect by $\mathbf{g} = (x_1; \ldots ;x_k)$, where x_j is the consequence that $\mathbf{g}$ assigns to state s_j. Then, there exists a utility function V (on the consequences) and a probability distribution q (on the states) such that the function

$$V(\mathbf{g}) = q(s_1)V(x_1) + \ldots + q(s_k)V(x_k)$$

represents the preference relation; that is, $\mathbf{g}$ is preferred to $\mathbf{g}'$ if and only if $V(\mathbf{g}) > V(\mathbf{g}')$.[25] Moreover, in this formula, q is uniquely determined, while V can be replaced by its positive affine transforms, and in no other way. The existence and relative uniqueness of V reproduce those obtained for U in the vNM theorem. The existence and absolute uniqueness of q are the specific additions of Savage's theorem. They justify treating q as representing the agent's beliefs—his subjective probability. Different agents will typically entertain different subjective probabilities, as they reflect typically different preferences over prospects—this is a crucial consequence for the impossibility result to come. We give it for only two individuals A and B, although a good deal of the work was to establish it in the proper algebraic form for a general population.

The Impossibility of Social Aggregation Theorem

Suppose that

(S) *A's and B's preferences over the set of prospects satisfy Savage's axioms;*

(S*) *The social observer's preferences also satisfy them;*

(P) *The two sets of preferences jointly satisfy the Pareto principle; and*

(A) *A and B rank at least one pair of prospects in the same way.*

Then, if A and B have different utility functions V_A and V_B over the consequences,[26] they must have identical subjective probability distributions—that is, $q_A = q_B$.

This counts as an impossibility theorem because the conclusion entails a restriction of individual diversity that is neither empirically plausible nor normatively warranted. Our wording displays the impossibility in terms of *probability identity*,[27], assuming that subjective probability differ, is equally apt.[28]

Mongin (1995) obtains the theorem by first making a step similar to the Social Aggregation Theorem using only Pareto indifference in (P). Then, the Savage utility

[24] Though still heuristically, because we take the states to be finite in number, whereas Savage excludes this.

[25] The same V applies both to $\mathbf{g}$ and x, since x can be identified with the degenerate uncertain prospect that gives x in all states of the world.

[26] $V_A(x)$ is understood to be the same as $V_B(x)$ if they are positive affine transforms of each other.

[27] Some commentators (e.g., Broome 1991, 160) say "probability agreement," but this is misleading, since it suggests that the individuals know each other's probabilities and have somehow come to "agree" on them. The theorem only says that they are the same.

[28] If there are more than two individuals, then the two forms of *reductio* are no longer equivalent by contraposition.

functions Z, V_A, V_B for the observer and the individuals are related by a weighted sum formula: for all uncertain prospects $\mathbf{g}$,

$$Z(\mathbf{g}) = a_A V_A(\mathbf{g}) + a_B V_B(\mathbf{g}) + b, \tag{24.5}$$

with the weights a_A, a_B having any sign. The next step is to develop this formula using the subjective probabilities q, q_A, and q_B, and show that it collapses into an impossibility when the Pareto principle is taken into full account. We sketch the argument, supposing for simplicity that there are two states s_1 and s_2.

The left-hand side of (24.5) has a sum over the states of the world:

$$q(s_1)Z(x_1) + q(s_2)Z(x_2).$$

The right-hand side of (24.5) has a sum over *both* the individuals *and* the states (we disregard the constant term):

$$a_A\, q_A(s_1) V_A(x_1) + a_A\, q_A(s_2) V_A(x_2) + a_B\, q_B(s_1) V_B(x_1) + a_B\, q_B(s_2) V_B(x_2).$$

How can the four-term right-hand side be made equal to the two-term left-hand side for every possible prospect $\mathbf{g}$ simultaneously? Intuitively, either by setting two of the four terms equal to zero, or by factorizing the four terms into two products. Reinforcing Pareto indifference with Strict Pareto in (P) gives positive coefficients a_A and a_B, so the first solution is barred. (This solution would anyhow be unwelcome since it would amount to making either A or B a dictator.) Assuming that the individuals differ in both their probabilities and utilities bars the second solution. This heuristically proves the impossibility theorem.

The Social Aggregation Theorem of last section corresponds to the case where $q = q_A = q_B$, so that probabilities can be factored out, reducing (24.5) to

$$\begin{aligned} &q(s_1)Z(x_1) + q(s_2)Z(x_2) \\ &\quad = q(s_1)\big(a_A V_A(x_1) + a_B V_B(x_1)\big) + q(s_2)\big(a_A V_A(x_2) + a_B V_B(x_2)\big).^{29} \end{aligned}$$

Dually, if $Z = V_A = V_B$, utilities can be factored out, and (24.5) becomes:

$$\begin{aligned} &q(s_1)Z(x_1) + q(s_2)Z(x_2) \\ &\quad = \big(a_A q_A(s_1) + a_B q_B(s_1)\big)Z(x_1) + \big(a_A q_A(s_2) + a_B q_B(s_2)\big)Z(x_2). \end{aligned}$$

[29] The present derivation of the Social Aggregation Theorem is only for social consequences x but a stronger one for social prospects $\mathbf{g}$ can be proved from Savage's axioms. One can identify the vNM lotteries of the original framework with Savage uncertain prospects when the individuals and the observer are given identical subjective probabilities. Blackorby, Donaldson, and Weymark (1999) have another derivation of the Social Aggregation Theorem in this adapted Savage framework.

This time, one obtains an averaging theorem for probabilities—this is the *linear opinion pool* of decision theory.[30] Thus, two classic theorems have been recovered as limiting cases of the same analysis. Richer solutions—with probabilities *and* utilities possibly differing—can emerge if there are more than two individuals.

The Impossibility of Social Aggregation Theorem has been given other forms, both axiomatic and nonaxiomatic, but none involves the pivotal step (24.5), from which impossibilities and possibilities derive at the same time. Hylland and Zeckhauser (1979) use social choice theory axioms, including a version of Arrow's ([1951] 1963) "independence of irrelevant alternatives," and they ignore the Social Aggregation Theorem altogether. Broome (1990) draws a connection with it using Jeffrey's ([1965] 1983) axiomatic decision theory, however without reaching (24.5).[31] At a nonaxiomatic level, welfare economists have long foreshadowed the Impossibility of Social Aggregation Theorem, taken in its probability identity form (see Starr 1973 and Hammond 1981). Their strong microeconomic assumptions prove unnecessary.[32] What in retrospect matters most with their treatment is the illuminating connection with the problem of ex ante versus ex post social welfare criteria. We now move to this major topic of the chapter.

When uncertainty prevails both on the individuals' and observer's side, the Pareto principle admits two versions. The ex ante one says that unanimous ex ante preferences should be respected, and the ex post one says that unanimous ex post preferences should be respected. The former applies unanimity once to uncertain prospects, whereas the latter applies it to consequences in each state of the world. Welfare economists have carried out a comparison between two *social welfare criteria* that can be associated with each version. The ex ante criterion combines the ex ante Pareto principle with subjective expected utility being imposed on the individuals; the ex post criterion combines the ex post Pareto principle with subjective expected utility being imposed on the observer.

Let us reconsider the assumptions of the Impossibility of Social Aggregation Theorem with these new concepts. (S) and (P) exactly capture the ex ante criterion, while (S*) is part of the ex post criterion. The ex post Pareto principle seems to be missing, but, given (S*), the ex ante principle (P) implies it, so that the assumptions encompass the full ex post criterion.[33] Thus, the conclusion can be restated as follows: *the ex ante and ex post social welfare criteria are incompatible*. For the welfare economists, this conflict sounds like a disaster, since they take each of the hypotheses (S), (S*) and (P) to be prima facie compelling.

[30] See the reviews in Genest and Zidek 1986 and Clemen and Winkler 1999.

[31] Bradley (2005) explores this and other differences with Mongin's treatment. Some of the technical assumptions Broome needs are specific to Jeffrey's framework and difficult to assess.

[32] They assume that outcomes are commodity bundles, utility functions are increasing with one's consumption of commodities, and these functions are differentiable. Under similarly strong microeconomic assumptions, the probability identity conclusion may already underlie Diamond's (1967b) model of the stock market.

[33] The implication follows because consequences (to which the ex post principle applies) are identified with constant prospects (to which the ex ante principle applies as a particular case).

But perhaps the welfare economists are wrong to insist on subjective expected utility, as if it was compelled by individual rationality. Among the alternative representations explored by decision theory, some have acquired normative appeal through axiomatic characterizations.[34] The first rescue move is thus to explore the consequences of weakening (S) and (S*) in the Impossibility of Social Aggregation Theorem. This can be done in a number of ways, given that there are seven independent axioms to consider in Savage, and that they apply here both to the individuals and to the observer.

By common consent, Savage's crucial axiom is the so-called *sure-thing principle*. However, weakening it, even in *both* (S) and (S*), does not yet deliver a solution. After the sketch in Blackorby, Donaldson, and Mongin (2004), this has been properly demonstrated in various frameworks by Gajdos, Tallon, and Vergnaud (2008), Fleurbaey (2009), Chambers and Hayashi (2014),[35] and Mongin and Pivato (2014). Admittedly, none of these frameworks exactly reproduces Savage's axiom or the lack thereof, but translation devices exist that make the negative conclusion inescapable. In essence, the impossibility still obtains with the considerably milder property of *statewise dominance* replacing the sure-thing principle. It says that if $\mathbf{g}$ and $\mathbf{g}'$ are uncertain prospects, and if the consequence of $\mathbf{g}$ is preferred to that of $\mathbf{g}'$ for every state, then $\mathbf{g}$ should be preferred to $\mathbf{g}'$. Decision theorists regard statewise dominance as normatively compelling, and many nonexpected utility systems include it, thus falling prey to the Impossibility of Social Aggregation Theorem. Unsurprisingly, these systems exhibit well-behaved aggregative properties when the individuals are taken to have the same utility function.[36]

The property of *state independence* that Savage conveys by two of his axioms also plays a role in the impossibility.[37] It requires that the preferences over the consequences defined conditionally on a state be the same, whatever this conditioning state. To illustrate, suppose that the states and the consequences describe the environmental situation and income distributions, respectively. Then, by state independence, ex post distributional judgments hold regardless of the environmental situation. But extreme values of the greenhouse effect that entail the near destruction of life on earth would presumably make the observer indifferent between income distributions that he would have strictly ranked under normal values. This and other examples suggest that the state-independence contained in (S*) is restrictive. Since it also plays a role in the proof, a prima facie attractive solution is to drop it out.

However, this may be wrong on second thought, because Savage's state independence is necessary for his derivation of a uniquely defined probability in the expected utility representation, and losing uniqueness, one is not entitled to consider it as being a

[34] The reader may consult surveys by Machina (2008), Gilboa (2009), and Karni (2014).

[35] Chambers and Hayashi (2014) generalize a result by Nehring (2004).

[36] Under this assumption, Crès, Gilboa, and Vieille (2011) and Nascimento (2012) show how to aggregate the *sets of probabilities* that the recent "ambiguity" constructions deliver to represent beliefs.

[37] Specifically, the state-independence axioms are (P3) and (P4). The sure-thing principle is stated in (P2), and (P1) is the standard ordering axiom. (P5), (P6), and (P7) play a mostly technical role. See Savage [1954] 1972, chs. 2–6.

subjective probability anymore. So the solution throws the baby out with the bathwater. After investigating it in detail, Mongin (1998) and Chambers and Hayashi (2006) eventually argue against it.[38]

The decision-theoretic avenue being closed, one is led to call the ex ante Pareto principle (P) into question. Here is a classic objection: it amounts to extending the individuals' sovereignty from evaluative judgments to factual judgments, and this is normatively dubious. The argument is usually presented against the background of hypothesis (S), which permits mapping factual and evaluative judgments onto probability and utility assignments, respectively. By the ex ante principle, the observer's ranking of social prospects will be sensitive not only to the values of individual utilities, which are neither correct nor incorrect, and which the observer can only take for granted, but also to the values of individual probabilities, whose correctness the observer can dispute (most typically, by referring to suitable evidence). When the Impossibility of Social Aggregation Theorem is phrased as the clash between the ex ante and ex post criteria, this argument automatically endorses the latter as being the only correct part of the former.[39]

This is a stringent objection and a stringent conclusion, but here is a more balanced view of what goes astray with (P) and how to remedy it. Suppose the observer should compare prospects **g** and **g**′ on the basis of the following individual data (we again assume two individuals and two states):

	s_1	s_2
A's probability	0.6	0.4
B's probability	0.4	0.6

	s_1	s_2
A's utility for **g**	20	−10
B's utility for **g**	−10	20

	s_1	s_2
A's utility for **g**′	0	0
B's utility for **g**′	0	0

Computing expected utility values, we find that both A and B give 8 to **g** against 0 to **g**′, so both prefer **g** to **g**′; by principle (P), the observer should endorse this preference. However, applying (P) here is normatively dubious, because A and B effectively disagree

[38] There is a classic move that consists of redefining social consequences to get rid of state dependence; however, it is blocked here by the need to have a rich-enough domain to prove the relevant theorems (including our Savage-based Impossibility of Social Aggregation Theorem).

[39] The argument often surfaces in the welfare economists' defense of the ex post criterion, as Broome (1991, 161) notices. However, Hammond (1982, 1983) also buttresses the ex post criterion, and more specifically ex post utilitarianism, by drawing upon the observer's dynamic consistency and the proof that this property connects with the expected utility axioms. On this author's utilitarian position, see also Hammond (1987).

on the reasons for their preferences: A supports $\mathbf{g}$ in the expectation of a positive consequence on s_1, while B supports it in the expectation of a positive consequence on s_2. They disagree on the ranking of *both* utilities and probabilities, and the ex ante Pareto principle has traction only because these two disagreements *cancel* each other. Mongin (1997) labels this class of situations *spurious unanimity* and emphasizes their critical role in bringing about impossibilities.[40] His proof of the Impossibility of Social Aggregation Theorem already emphasizes this role, and it would be possible to check that any other proof must also involve a spurious unanimity step.

Gilboa, Schmeidler, and Samet (2004) have proposed a solution based on this analysis. They begin by dividing the events—sets of states of the world—into those that get the same probability values from all individuals and those that do not. They now divide uncertain prospects into those that can be defined in terms of events of the former class—call them *admissible*—and the others. Finally, they propose restricting (P) to comparisons of admissible prospects. This restriction is precisely meant to exclude any prospects, like $\mathbf{g}$ and $\mathbf{g}'$ above, that could foster spurious unanimity. The authors take this problem, rather than the extension of unanimity endorsement to factual matters, to be the real objection to the ex ante Pareto principle.

More formally, in the statement above of the Impossibility of Social Aggregation Theorem, the authors replace (P) by

(P′) *The two sets of preferences jointly satisfy the Pareto principle over admissible prospects only.*

With the other assumptions unchanged, two conclusions follow. First, for all consequences x,

$$Z(x) = a_A V_A(x) + a_B V_B(x) + b,$$

with positive weights a_A, a_B. Second, for all events E of the state space,

$$q(E) = c_A q_A(E) + c_B q_B(E),$$

with positive weights c_A, c_B summing to 1. In words, the outputs are a utility sum that is restricted to *consequences*, instead of holding over *prospects* as in (24.5), and the already encountered *linear pooling rule* for probabilities.

Both results are consistent with the ex post social welfare criterion, while being more precise in two ways. First, the ex post criterion only says that the observer's utility for consequences is functionally related to the individuals' corresponding utilities; it takes ex post utilitarianism to claim additivity. Second, the ex post criterion only endows the observer with *some* subjective probability, without necessarily relating it to individual probabilities by the linear pooling rule or any other way. The extra content follows from adopting a version of the Pareto principle that is halfway between ex post and ex ante.

[40] Early on, Raiffa (1968) had made a similar suggestion, thus prompting Hylland and Zeckhauser's (1979) work.

It seems to be assertive just in the right way by restricting (24.5) from prospects to consequences. However, difficulties remain because the linear pooling rule can get the wrong answer in situations where the individual probabilities arise from conditioning on private information. In these cases, the observer's probability should depend on some aggregate of this private information, and not simply on the probability data that individually result from conditioning.[41]

To illustrate the critical point, suppose there are three states of nature and A and B have a common prior probability:[42]

	s_1	s_2	s_3
A's and B's probability	0.49	0.02	0.49

Now suppose that each has access to some private information before uncertainty is completely resolved, for example, that A observes the event $\{s_1, s_2\}$, and B observes the event $\{s_2, s_3\}$. If each revises probabilities by conditioning, then the outcome of the linear pooling is as follows:

	s_1	s_2	s_3
A's revised probability	0.96	0.04	0
B's revised probability	0	0.04	0.96
Average	0.48	0.04	0.48

However, assuming that both A and B have reliable pieces of information, the social observer should infer *by comparing them* that the true state is s_2. Linear pooling misses this answer because it erases all information that is not already contained in A's and B's revised probabilities.

This example can be turned into an objection against the restricted ex ante Pareto principle (P′) proposed by Gilboa, Samet, and Schmeidler. Suppose A and B have the following common utility function:

[41] Another problem of linear pooling, which we do not pursue here, is that the observer's probability may also have to depend on the individuals' risk attitudes (Gollier 2007).

[42] Mongin (1997) has a related example and argument.

	s_1	s_2	s_3
A & B's utility for **g**	100	0	100
A & B's utility for **g**′	0	100	0

The initial expected utility values are 98 and 2 for **g** and **g**′, and A and B should revise them to 96 and 4 after revising their probabilities. Now, A and B agree on the probabilities of the events "**g** = 0", "**g** = 100", "**g**′ = 0" and "**g**′ = 100"; therefore, **g** and **g**′ are admissible prospects, so by (P′), the observer should prefer **g** to **g**′. However, we have seen that he should recognize s_2 as the true state, and this entails the opposite preference. As this contradiction illustrates, spurious unanimity is not so easily vanquished as Gilboa, Samet, and Schmeidler had hoped.

At this point, readers will naturally be tempted to return to the ex post social welfare criterion. However, a critical point due to Hild, Jeffrey, and Risse (2003, 2008) awaits them. These authors argue that what counts as a consequence in one description counts as an uncertain prospect in another, more thorough one. They model a "flipping" situation in which an ex post social observer constantly refines consequences, and as a result, constantly oscillates between two preferred options. By contrast, the ex ante principle is well defined in such a situation. If it should eventually prevail, despite its shortcomings, this is really because there is no genuine ex post alternative to it (see also Risse 2001, 2003).

This is about the stage where we should leave the problem of ex ante versus ex post social welfare.[43] We now move to another topic, changing our focus from utilitarianism to egalitarianism.

24.5 Egalitarianism and the Problem of Ex Ante versus Ex Post Welfare

The next developments originate in a modest theoretical point by Diamond (1967a). He supposes a two-individual social uncertainty situation, with two equiprobable states h and t, and two prospects

$$\mathbf{f} = (x_h; x_t) \text{ and } \mathbf{u} = (x'_h; x'_t),$$

where $x_h = x'_h = x'_t = (1,0)$ and $x_t = (0,1)$ are the two individuals' utility amounts. In table form:

[43] Some recent working papers follow up the issues of this section and especially the spurious unanimity objection to the ex ante Pareto principle: Gayer et al. 2014; Alon and Gayer 2014; Billot and Vergopoulos 2014; Danan et al. 2014; Gilboa, Samuelson, and Schmeidler 2014; Qu 2014. Jackson and Yariv (2014) prove an original form of the Impossibility of Social Aggregation Theorem in which individual discount rates play the role given to individual probabilities.

	h	t
A's utility for **f**	1	0
B's utility for **f**	0	1

	h	t
A's utility for **u**	1	1
B's utility for **u**	0	0

Strotz (1958, 1961) had already offered this example, or closely related ones, claiming that they raised a "paradox" for the notion of a just income distribution.[44] Diamond turned the example specifically against Harsanyi's additive SWF. With equal weighting of A and B, the sum-of-utility formula gives

$$W(\mathbf{f}) = \frac{1}{2}\Big(V_A(x_h) + V_B(x_h)\Big) + \frac{1}{2}\Big(V_A(x_t) + V_B(x_t)\Big) = 1,$$
$$\text{and } W(\mathbf{u}) = \frac{1}{2}\Big(V_A(x'_h) + V_B(x'_h)\Big) + \frac{1}{2}\Big(V_A(x'_t) + V_B(x'_t)\Big) = 1,$$

so that social indifference prevails. Diamond objects that **f** should be chosen because it gives B a "fair shake," unlike **u**—that is, **f** gives A and B equal chances to score a high utility level, whereas **u** puts A at his high level, and B at his low level, for sure. Diamond concludes that the social observer should not be subjected to expected utility theory—our hypothesis (S*) in section 24.3.[45]

The tiny example acquired great popularity after Sen expanded on it, though at first critically.[46] In article suggestively entitled: "Do Welfare Economists Have a Special Exemption from Bayesian Rationality?" Harsanyi (1975) defended (S*) on the ground that the social observer should obey ordinary rationality standards, but as Sen (1976 and 1977) readily noted, this of course begs the question of what they are. Indeed, Diamond's or equivalent examples can *precisely* be used to test one's intuition on the rational force of the expected utility axioms.[47] Harsanyi (1975, 1977b) has another reply of deeper relevance. He argues that **f** and **u** are not really ex ante prospects, since they come after the resolution of a prior uncertainty on A's and B's identities. A full analysis should take this "great lottery of life" (1975, 317) into account, as in the Impartial Observer Theorem, and once this is done, there is no intuitive reason left for ex ante preferring **f** to **u**, so that expected utility theory can safely be reaffirmed for the observer. This is a sophisticated dismissal, but it can be countered by Grant et al.'s (2010)

[44] See also Fisher and Rothenberg's (1961, 1962) comments on Strotz.

[45] Diamond takes states of the word to be explicit as the same time as probabilities. This is a mixed framework of uncertainty, as in note 29, so we will noncommittally say "expected utility" instead of referring to Savage or vNM.

[46] As Sen ([1970] 1979, 143–46) initially observed, Diamond's critique is compelling only if A's and B's utility *levels* are comparable, as in the utility-based version of the Rawlsian maximin, whereas a sum of utility only needs utility *differences* to be comparable. Sen (1974) takes over this measurability problem, but neglects it later when polemicizing with Harsanyi.

[47] Machina (1989) imagines a mother who could give a present to either of her two children directly, but prefers to decide by a random drawing who is the beneficiary. Expected utility theory cannot account for this preference, which Machina believes nonetheless to be rationally defensible. This is in effect Diamond's example, as Grant (1995) has noted, while pursuing Machina's intuition against expected utility theory in all axiomatic detail.

argument that "identity lotteries" and "outcome lotteries" should not be submitted to the same preference axioms (see section 24.2).

The example is actually richer than Diamond himself suggested. Let us suppose that the observer satisfies *statewise dominance* instead of expected utility. Section 24.4 already investigated this weakening in connection with social welfare. Let us also assume that the observer treats A and B symmetrically and is thus indifferent between the ex post utility vectors (0, 1) and (1, 0). Then, the observer will again be indifferent between **f** and **u**.[48] Seen from this broader perspective, the example raises a problem for egalitarianism itself. The equality concept, like the social welfare concept, generates two criteria under uncertainty. The ex ante egalitarian criterion first applies decision theory to the individuals, and then assesses prospects by considering the distributions of ex ante utility values that this application has generated. The ex post egalitarian criterion first assesses the distributions of individual utility values in each state, and then assesses prospects by applying decision theory to these data. Thus, even if expected utility theory is weakened to statewise dominance, there is a clash between the two egalitarian criteria. This has become the received interpretation of Diamond's example.

Intriguingly, no canonical statement like the Impossibility of Social Aggregation Theorem has formalized this conflict within egalitarianism. Thus, more examples are worth giving to show how tenacious it is. Adler and Sanchirico (2006) consider the prospects **e** and **e**′:

	h	t
A's utility for **e**	3	3
B's utility for **e**	1	1

	h	t
A's utility for **e**′	4	0
B's utility for **e**′	0	4

Under equiprobability and expected utility, A gives 3 and B gives 1 to **e**, and they give a common value 2 to **e**′, which ex ante equality would thus favor. However, in either state, the utility distribution is more egalitarian in **e** than **e**′, which ex post equality would thus disfavor. The conflict is not between ex post indifference and ex ante preference, as in Diamond, but between opposite preferences.

Consider now the prospects **m** and **m**′:

	h	t
A's utility for **m**	9	81
B's utility for **m**	81	9

	h	t
A's utility for **m**′	25	64
B's utility for **m**′	64	25

Under equiprobability and expected utility, A and B both get 45 in **m** and 44.5 in **m**′. So ex ante equality, in accord with ex ante Paretianism, will strictly prefer **m** to **m**′. But ex

[48] This reasoning uses statewise dominance in terms of indifference, which can be obtained from the standard version in terms of preference by a continuity argument.

post equality may rank $\mathbf{m}'$ over $\mathbf{m}$. To see the conflict, apply the inequality-averse SWF $W_{1/2}(a,b) = a^{1/2} + b^{1/2}$ to $\mathbf{m}$ and $\mathbf{m}'$ to either case. The chosen function is *concave* in the utilities, which entails valuing marginal increases of the higher one *less* than marginal increases of the lower one, so that the evaluation is sensitive to equality.[49] Now, in both states s and t, $\mathbf{m}'$ (having $W_{1/2}$ value 13) is ex post preferred to $\mathbf{m}$ (having $W_{1/2}$ value 12), while $\mathbf{m}$ is still ex ante ranked above $\mathbf{m}'$ (as generally $W_{1/2}(a,a) > W_{1/2}(a',a')$ if $a>a'$). The conflict between the two egalitarian criteria comes out vividly again, and it also turns out that ex post equality, unlike ex ante equality, can clash with the ex ante Pareto principle (a violation noted, e.g., in Myerson 1981).

As Adler and Sanchirico (2006) observe, the ex post criterion typically devotes more resources than the ex ante one to prevent public risks that threaten everyone with a small probability, and materialize by striking only small numbers, though very badly so. Examples involve pandemics, terrorist attacks, and natural disasters. Adler and Sanchirico express a considered preference for the ex post criterion, which they summarize as follows: "Policymakers should focus on the prospects for equity, not the equity of prospects" (2006, 350). The applied prescriptive work appears to be divided between the two criteria.

24.6 Solutions to the Problem of Ex Ante versus Ex Post Equality

We will now cover various solutions to the conflict. Given the broad analogy with the earlier problem of social welfare, one may expect two polar solutions, which simply endorse one criterion and discard the other, while others attempt a compromise (these last solutions actually prove easier to find than in the other problem). All proposals take the form of defining axioms for ranking social prospects, with the uncertain consequences stating either material allocations, or as in Diamond-like examples, already given utility vectors.

Let us first revisit the additively separable SWF that many have offered as being the true conclusion of Harsanyi's two theorems. We restate it for social prospects $\mathbf{g}$:

$$W(\mathbf{g}) = f_1(U_1(\mathbf{g})) + \ldots + f_i(U_i(\mathbf{g})) + \ldots + f_n(U_n(\mathbf{g})), \tag{24.6}$$

supposing that the U_i are subjective expected utility functions computed with the same given probabilities and the f_i are increasing concave transformations. As explained above, concavity introduces a concern for equality, so that (24.6) immediately provides

[49] It is a perennial idea of the utilitarian school that if utilities are concave, egalitarian consequences will follow from the sum or average rules. The standard argument takes utility functions to be alike for all individuals, though Lerner (1944) and Sen (1969) have generalized it somewhat. The use of concavity here is more refined since it concerns transformations of typically diverse utility functions.

an ex ante egalitarian solution. As to the corresponding ex post egalitarian solution, it results from restricting W to the consequences: for all allocations x,

$$W(x) = f_1(U_1\ (x)) + \ldots + f_i(U_i\ (x)) + \ldots + f_n(U_n\ (x)).$$

Then, using the $W(x)$ as ex post social utilities, one gets an ex post solution by taking expected values of W according to the given probabilities.

Section 24.5 illustrated the changes in utility comparisons brought about by taking f_i to be the square root, but applied welfare economists have used many other concave functions.[50] At the theoretical level, only the generic choice of concavity matters, and it must be justified by axioms put on the social observer's preferences. In the context of the Impartial Observer Theorem, Grant et al. (2010) axiomatize concave f_i in the ex ante SWF $W(\mathbf{g})$ by the condition that the observer's extended preferences favor "outcome lotteries" over "identity lotteries." Regrettably, there is no comparable foundation in the context of the Social Aggregation Theorem.[51]

Among the SWFs that pay attention to inequality, the additively separable ones are only one step removed from the additive ones. The next generalizing step is to replace the f_i by *rank-dependent transformations*, meaning that the transformation of *i*'s utility does not depend anymore on *i*'s identity, but on the rank that *i*'s component occupies in the utility vector. A typical example is the *Gini* SWF, which derives from the well-known inequality index with the same label (see, e.g., Sen 1973). It has the property that the weights of successively ranked quantities differ by a constant; for example, for two or three individuals, with the quantities stated in increasing order, the Gini SWF is, respectively, $(3x_1 + x_2)\ /\ 4$ and $(5x_1 + 3x_2 + x_3)\ /\ 9$. This and other rank-dependent transformations can be used to deliver both ex ante and ex post inequality-averse SWF.

There are even more radical departures from additivity, such as taking a higher-degree polynomial. Epstein and Segal (1992) recommend the first example in this class—a quadratic—as a suitable ex ante inequality-averse SWF. They interpret Diamond's example as reflecting "preference for randomization" and propose the following axiom for this property: if the observer is indifferent between two social prospects, yet some individuals prefer one, while others prefer the other, then the observer should rank both prospects *below any prospect randomizing between them*. Compare Diamond's initial prospects with the following one:

	h	t
A's utility for $\mathbf{u}'$	0	0
B's utility for $\mathbf{u}'$	1	1

[50] A common example is the function $f_i(r) = (1/n)\ r^{1-a}/(1-a)$, which connects with a convenient inequality measure.

[51] Note that beside being concave, the f_i can be used to make the range of the different U_i comparable in the "relative utilitarian" sense (see note 11).

Prospect **f**—Diamond's favorite—can be seen as a (1/2 , 1/2) randomization of **u** and $\mathbf{u}'$, which evoke opposite individual preferences. Hence if the observer is indifferent between **u** and $\mathbf{u}'$, as ex ante symmetry between the individuals would recommend, then by the proposed axiom, he prefers **f** to both **u** and $\mathbf{u}'$.[52]

Epstein and Segal also introduce an axiom of "mixture symmetry." In words, if the observer is indifferent between two prospects, then he is also indifferent between two randomizations of these prospects that interchange the weights; for example, if he is indifferent between $\mathbf{g}'$ and $\mathbf{g}''$, he is between their randomizations (¾, ¼) and (¼, ¾). Adding the ex ante Pareto principle and an expected utility condition on individual preferences, Epstein and Segal derive a *quadratic* SWF representation of social preference.[53] Let us put this conclusion formally in the two-individual case. For all expected utility functions $U_A(\mathbf{g})$ and $U_B(\mathbf{g})$, there exist a utility representation $\Phi(\mathbf{g})$ of social preference, and constants α, β, γ , δ and ε such that for all **g**,

$$\Phi(\mathbf{g}) = \alpha U_A(\mathbf{g}) + \beta U_B(\mathbf{g}) + \gamma\, U_A(\mathbf{g})^2 + \delta U_B(\mathbf{g})^2 + \varepsilon U_A(\mathbf{g}) U_B(\mathbf{g}).$$

Moreover, $\Phi(\mathbf{g})$ is increasing and quasi-concave in the two variables $U_A(\mathbf{g})$ and $U_B(\mathbf{g})$.[54] The former property reflects the Pareto principle, while the latter—a technical weakening of concavity—reflects "preference for randomization." Roughly speaking, $\Phi(\mathbf{g})$ balances the additive conclusion of the Social Aggregation Theorem against other terms that can express a concern for equality. Recall, however, that if Sen's critique of the welfare sense of expected utility measurements holds good, it does not spare Epstein and Segal's result any more than Harsanyi's.

The next solutions belong to the class of egalitarian compromises between ex ante and ex post. This class has emerged only recently and is still expanding with current research. We have selected three representative and interconnected axiomatic pieces.[55]

Ben-Porath, Gilboa, and Schmeidler (1997) propose ranking social prospects by the class of *min-of-means* SWF. To explain this, we first define a *weighting* w to be an assignment to individuals i and states s of nonnegative weights $w(i,s)$ that sum to 1. Using w, one computes the w-weighted average value of any social prospect. For example, take w as follows:

	h	*t*
A's weight	1/8	5/8
B's weight	1/8	1/8

[52] Preference for randomization is also a property of Karni and Safra's (2002a) moral preference relation.

[53] This solution follows the stride toward nonlinearity that has taken place in decision theory; see, e.g., Machina's (2008) review.

[54] The constants α, β, γ, δ, and ϵ must be restricted if Φ is to satisfy this property, but this is not done in Epstein and Segal.

[55] Earlier on, Keeney and Winkler (1985) proposed a compromise representation that adds up three terms, i.e., one for catastrophe avoidance and two for ex ante and ex post equality respectively.

With these data, Diamond's prospects **f** and **u** score $1/8 + 1/8 = 1/4$ and $1/8 + 5/8 = 3/4$. Now let **W** be a collection of such weightings. The **W**-min-of-means SWF evaluates each lottery by the *smallest weighted average* it obtains from any element in **W**. For example, take the four weightings:

0.3	0.4
0.2	0.1

0.4	0.3
0.1	0.2

0.1	0.2
0.4	0.3

0.2	0.1
0.3	0.4

Then, the new prospects $\mathbf{g}^*$ and $\mathbf{g}^{**}$, which are defined below, both score 0.5, while **f** and **u** get 0.4 and 0.3 respectively.

	h	*t*
A's utility for $\mathbf{g}^*$	1/2	1/2
B's utility for $\mathbf{g}^*$	1/2	1/2

	h	*t*
A's utility for $\mathbf{g}^{**}$	1	0
B's utility for $\mathbf{g}^{**}$	1	0

Accordingly, the social ranking is: $\mathbf{g}^*$ *indifferent to* $\mathbf{g}^{**}$ *preferred to* **f** *preferred to* **u**.

This reconciles Diamond's ex ante egalitarian intuition that **f** is better than **u** with the ex post egalitarian intuition that $\mathbf{g}^*$ and $\mathbf{g}^{**}$ are both better than **f** (since they equalize utility in each state).

Ben-Porath, Gilboa, and Schmeidler characterize min-of-means SWFs over prospects for income vectors (although a utility interpretation seems also applicable) by relevant axioms put on the observer's preferences.[56] When min-of-means SWFs are defined under certainty, they include average income, minimal income, and Gini SWF as particular cases.[57] Under uncertainty, they include SWFs that reconcile ex ante and ex post intuitions about equality, as our numerical example showed. However, it is not entirely clear how to select functions in this wide class, and one may wish to have a conceptual interpretation for the weights. In a different criticism, Gajdos and Maurin (2004) argue that Ben-Porath, Gilboa, and Schmeidler's solution is too specific, min-of-means being only one way of finding SWFs that are both ex ante and ex post egalitarian.

Gajdos and Maurin's more general approach stems from the following observation. Instead of applying the Gini SWF to the individuals' expected incomes, as in a typical ex ante solution, or taking the expectation of the Gini SWF applied to the individuals' incomes in each state, as in a typical ex post solution, one may attempt a compromise

[56] Once again, this axiomatic work parallels an earlier one in decision theory; see Gilboa 2009, ch. 17, on how min-of-means (or "maxmin expected utility") obtains in this context. A previous application to inequality did not yet involve uncertainty (Ben-Porath and Gilboa 1994).

[57] For the Gini SWF, **W** is the set of all weightings with the property that weights for successive ranks differ by a constant.

by *weighting one against the other*.[58] However, in the authors' view, this generalization would not yet be sufficient. The specific operators W (for the Gini SWF) and E (for mathematical expectation) should give way to abstract ones I_a and I_p according to some axiomatization of social preferences. The axioms should be so chosen that $I_a(I_p(\mathbf{g}))$ and $I_p(I_a(\mathbf{g}))$ would represent the pure ex ante and pure ex post egalitarian solutions, and that

$$I(\mathbf{g}) = \lambda I_a(I_p(\mathbf{g})) + (1-\lambda) I_p(I_a(\mathbf{g}))$$

would represent the final compromise between them. Gajdos and Maurin make significant additions to Ben-Porath, Gilboa and Schmeidler's axiom system. We only stress the Pareto condition that requires unanimity to hold between the individuals *both ex ante and ex post* for the social observer to endorse it. This directly connects with the observation in section 24.5 that equality criteria sometimes upset the usual forms of the Pareto principle.[59]

Commenting on these earlier works, Fleurbaey (2010) observes that weighted averages such as $I(\mathbf{g})$ may possess the drawbacks of both ex ante and ex post egalitarian criteria, namely violation of statewise dominance for the former and violation of the ex ante Pareto principle for the latter. He rather recommends a SWF behaving like an ex ante one in situations where this is appropriate, like an ex post one in situations where this is appropriate, and otherwise smoothly combining these two. More precisely, Fleurbaey identifies two classes of prospects: the *riskless* ones (where, for each agent, the outcome is the same in all states of nature), and the *egalitarian* ones (where all agents face the same individual lottery, hence getting the same consequence in each realized state). Here is an example with three agents and four states of nature:

Riskless prospect	s_1	s_2	s_3	s_4
A's utility	1	1	1	1
B's utility	3	3	3	3
C's utility	2	2	2	2

Egalitarian prospect	s_1	s_2	s_3	s_4
A's utility	1	3	5	2
B's utility	1	3	5	2
C's utility	1	3	5	2

Now come two Paretian axioms. "Weak Pareto for equal risk" is the ex ante principle as restricted to egalitarian lotteries, and "Weak Pareto for no risk" is the same principle as restricted to riskless prospects. Together with statewise dominance, "Weak Pareto for no risk" gives rise to an ex post SWF W such that, for all ex post utility

[58] Here is a symbolic rendering of the recommended SWF: for all prospects $\mathbf{g}$, $\lambda W(\mathrm{E}(\mathbf{U}(\mathbf{g}))) + (1-\lambda)$ $\mathrm{E}(\mathbf{W}(\mathbf{g}))$, where $\mathbf{U}(\mathbf{g})$ is the vector of individual expected utilities associated with $\mathbf{g}$, $\mathbf{W}(\mathbf{g})$ is the vector of values taken by W when applied to $\mathbf{g}$ state by state, and λ is a number between 0 and 1.

[59] Strengthening Gajdos and Maurin's system by including the ex ante and ex post Pareto principles unrestrictedly would lead back to the Impossibility of Social Aggregation Theorem. The comparison is most easily done with Mongin and Pivato's (2014) version.

vectors $(u_1, \ldots, u_n)$,

$$W(u_1, \ldots, u_n) = e,$$

where e is the utility level making $(u_1, \ldots, u_n)$ ex post socially indifferent to $(e, e, \ldots, e)$.[60] As usual, the more concave the W function is, the more inequality-averse the social ex post preferences. Then Fleurbaey shows that the only SWFs satisfying both Pareto axioms and statewise dominance have the *expected equally distributed equivalent* form; that is, for all

$$\mathbf{g}, \mathbf{g}', \mathbf{g} \text{ is socially preferred to } \mathbf{g}' \text{ if and only if } E[W(\mathbf{g}(s))] > E[W(\mathbf{g}'(s))].$$

This SWF satisfies the ex post equality criterion by construction. Although it violates the ex ante Pareto principle in general, it satisfies it when comparing either two egalitarian lotteries or two riskless lotteries, and in these cases, also satisfies the ex ante equality criterion. The intuitive force of this solution can be tested on the above set of four prospects. With a concave W, the resulting social ranking is

$\mathbf{g}^*$ *preferred to* $\mathbf{g}^{**}$ *preferred to* $\mathbf{f}$ *indifferent to* $\mathbf{u}$.

This ranking may be completely satisfactory to an ex post egalitarian, but not to an ex ante one, as it still suffers from Diamond's original criticism.[61]

24.7 Conclusion

We have reviewed the existing theories of social preference and social welfare under risk and uncertainty, borrowing the economists' division between these two cases. Another clue to the chapter was the ex ante and ex post division of welfare theory, which we also applied to egalitarian criteria. These abstract categories made it possible to locate the various theories in a more or less unified framework. All theories turn out to have shortcomings, which either the axiomatic decomposition or mere counterexamples reveal. But some solutions have emerged as representing interesting compromises between opposite criteria. Although the present debate over the ex ante and ex post social welfare seems inconclusive, further work could combine both more appropriately, and egalitarian thinkers have already devised relevant trade-offs between ex ante and ex post equality, thus suggesting that apparent impossibilities can sometimes be overcome.

[60] Defining *egalitarian equivalences* for utility or commodity vectors has become a standard tool of recent welfare economics; see Fleurbaey 2008.

[61] By varying the form of W, this SWF can be made to coincide with existing ones, such as the utility sum, the min of utilities, or variants of the Gini SWF.

Acknowledgments

Most of this chapter was written when M. Pivato was at the Department of Mathematics at Trent University, Canada. This research was supported by NSERC grant #262620-2008, and by Labex MME-DII (ANR11-LBX-0023- 01).

References

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

Adler, M. D. 2014. "Extended Preferences and Interpersonal Comparisons: A New Account." *Economics and Philosophy* 30: 123–62.

Adler, M. D., and C. W. Sanchirico. 2006. "Inequality and Uncertainty: Theory and Legal Applications." *University of Pennsylvania Law Review* 155: 279–377.

Allais, M. 1953. "Le comportement de l'homme rationnel devant le risque." *Econometrica* 21: 503–46.

Alon, S., and G. Gayer. 2014. "Utilitarian Preferences and Multiple Priors." Manuscript, Bar-Ilan University.

Anscombe, F. J., and R. J. Aumann. 1963. "A Definition of Subjective Probability." *Annals of Mathematical Statistics* 34: 199–205.

Arrow, K. J. [1951] 1963. *Social Choice and Individual Values*. 2nd ed. New Haven: Yale University Press.

Billot, A., and V. Vergopoulos. 2014. "Utilitarianism with Prior Heterogeneity." Manuscript, LEMMA, Université Paris II.

Ben-Porath, E., and I. Gilboa. 1994. "Linear Measures, the Gini Index, and the Income-Equality Trade-Off." *Journal of Economic Theory* 64: 443–67.

Blackorby, C., W. Bossert, and D. Donaldson. 2005. *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*. Cambridge: Cambridge University Press.

Blackorby, C., W. Bossert, and D. Donaldson. 2006. "Utilitarianism and the Theory of Justice." K. J. Arrow, A. K. Sen, and K. Suzumura, eds., *Handbook of Social Choice and Welfare*, vol. 1, 543–96. Amsterdam: North-Holland.

Blackorby, C., D. Donaldson, and P. Mongin. 2004. "Social Aggregation without the Expected Utility Hypothesis." Cahier no. 2004-020, Laboratoire d'économétrie, Ecole Polytechnique.

Blackorby, C., D. Donaldson, and J. A. Weymark. 1999. "Harsanyi's Social Aggregation Theorem for State-Contingent Alternatives." *Journal of Mathematical Economics* 32: 365–87.

Blackorby, C., D. Donaldson, and J. A. Weymark. 2008. "Social Aggregation and the Expected Utility Hypothesis." M. Fleurbaey, M. Salles, and J. A. Weymark, eds., *Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls*, 136–83. Cambridge: Cambridge University Press.

Bommier, A., and S. Zuber. 2008. "Can Preferences for Catastrophe Avoidance Reconcile Social Discounting with Intergenerational Equity?" *Social Choice and Welfare* 31: 415–34.

Bouyssou, D., and J. C. Vansnick. 1990. "'Utilité cardinale' dans le certain et choix dans le risque." *Revue économique* 41: 979–1000.

Bradley, R. 2005. "Bayesian Utilitarianism and Probability Homogeneity." *Social Choice and Welfare* 24: 221–51.

Broome, J. 1990. "Bolker-Jeffrey Expected Utility Theory and Axiomatic Utilitarianism." *Review of Economic* Studies 57: 477–503.

Broome, J. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Oxford: Blackwell.

Chambers, C., and T. Hayashi. 2006. "Preference Aggregation under Uncertainty: Savage vs. Pareto." *Games and Economic Behavior* 54: 430–40.

Chambers, C., and T. Hayashi. 2014. "Preference Aggregation with Incomplete Information." *Econometrica* 82: 589–99.

Clemen, R. T., and R. L. Winkler. 1999. "Combining Probability Distributions from Experts in Risk Analysis." *Risk Analysis* 19: 187–203.

Coulhon, T., and P. Mongin. 1989. "Social Choice Theory in the Case of von Neumann–Morgenstern Utilities." *Social Choice and Welfare* 6: 175–87.

Crès, H., I. Gilboa, and N. Vieille. 2011. "Aggregation of Multiple Prior Opinions." *Journal of Economic Theory* 146: 2563–82.

Danan, E., T. Gajdos, and J. M. Tallon. 2013. "Aggregating Sets of von Neumann–Morgenstern Utilities." *Journal of Economic Theory* 148: 663–88.

Danan, E., T. Gajdos, and J. M. Tallon. 2015. "Harsanyi's Aggregation Theorem with Incomplete Preferences." *American Economic Journal: Microeconomics* 7 (1): 61–9.

Danan, E., T. Gajdos, B. Hill, and J. M. Tallon. 2014. "Aggregating Tastes, Beliefs, and Attitudes Under Uncertainty." THEMA Working Paper No. 2014-13, Université de Cergy-Pontoise.

d'Aspremont, C. and L. A. Gérard-Varet. 1991. "Utilitarian Fundamentalism and Limited Information." J. Elster and J. Roemer, eds., *Interpersonal Comparisons of Well-Being*, 371–85. Cambridge: Cambridge University Press.

Debreu, G. 1959. *Theory of Value*. New Haven: Yale University Press.

De Meyer, B., and P. Mongin. 1995. "A Note on Affine Aggregation." *Economics Letters* 47: 177–83.

Dhillon, A., and J. F. Mertens. 1999. "Relative Utilitarianism." *Econometrica* 67: 471–98.

Diamond, P. A. 1967a. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment." *Journal of Political Economy* 75: 765–66.

Diamond, P. A. 1967b. "The Role of Stock Market in a General Equilibrium Model with Technological Uncertainty." *American Economic Review* 57: 759–76.

Dworkin, R. 1981. "What Is Equality? Part 2: Equality of Resources." *Philosophy and Public Affairs* 10: 283–345.

Epstein, L. G., and U. Segal. 1992. "Quadratic Social Welfare Functions." *Journal of Political Economy* 100: 691–712.

Fisher, F. M., and J. Rothenberg. 1961. "How Income Ought to Be Distributed: Paradox Lost." *Journal of Political Economy* 61: 162–80.

Fisher, F. M., and J. Rothenberg. 1962. "How Income Ought to Be Distributed: Paradox Enow." *Journal of Political Economy* 70: 88–93.

Fleurbaey, M. 2008. *Fairness, Responsibility and Welfare*. Oxford: Oxford University Press.

Fleurbaey, M. 2009. "Two Variants of Harsanyi's Aggregation Theorem." *Economics Letters* 105: 300–302.

Fleurbaey, M. 2010. "Assessing Risky Social Situations." *Journal of Political Economy* 118: 649–80.

Fleurbaey, M., and P. Mongin. 2011. "The Utilitarian Relevance of the Social Aggregation Theorem." WP No. 955, GREGHEC, HEC School of Management.

Fontaine, P. 1997. "Identification and Economic Behavior: Sympathy and Empathy in Historical Perspective." *Economics and Philosophy* 13: 261–80.

Gajdos, T., and F. Kandil. 2008. "The Ignorant Social Observer." *Social Choice and Welfare* 31: 193–232.

Gajdos, T., and E. Maurin. 2004. "Unequal Uncertainties and Uncertain Inequalities: An Axiomatic Approach." *Journal of Economic Theory* 116: 93–118.

Gajdos, T., J. M. Tallon, and J. C. Vergnaud. 2008." Representation and Aggregation of Preferences under Uncertainty." *Journal of Economic Theory* 141: 68–99.

Gauthier, D. 1986. *Morals by Agreement*. Oxford: Oxford University Press.

Gayer, G., I. Gilboa, L. Samuelson, and D. Schmeidler. 2014. "Pareto Efficiency with Different Beliefs." *The Journal of Legal Studies* 43 (S2): S151–S171, Conference proceedings of "Benefit-Cost Analysis of Financial Regulation."

Genest, C., and J. V. Zidek. 1986. "Combining Probability Distributions: A Critique and Annotated Bibliography." *Statistical Science* 1: 114–48.

Gilboa, I. 2009. *Theory of Decision under Uncertainty*. Cambridge: Cambridge University Press.

Gilboa, I., L. Samuelson and D. Schmeidler. 2014. "No-Betting Pareto Dominance." Manuscript, Tel Aviv University.

Gilboa, I., D. Schmeidler, and D. Samet. 2004. "Utilitarian Aggregation of Beliefs and Tastes." *Journal of Political Economy* 112: 932–38.

Gollier, C. 2007. "Whom Should We Believe? Aggregation of Heterogeneous Beliefs." *Journal of Risk and Uncertainty* 35: 107–27.

Grant, S. 1995. "Subjective Probability without Monotonicity: or How Machina's Mom May Also Be Probabilistically Sophisticated." *Econometrica* 63: 159–89.

Grant, S., A. Kajii, B. Polak, and Z. Safra. 2010. "Generalized Utilitarianism and Harsanyi's Impartial Observer Theorem." *Econometrica* 78: 1939–71.

Grant, S., A. Kajii, B. Polak, and Z. Safra. 2011. "Equally Distributed Equivalent Utility, *Ex Post* Egalitarianism and Utilitarianism." *Journal of Economic Theory* 147 (4): 1545–71.

Grant, S., A. Kajii, B. Polak, and Z. Safra. 2012. "A Generalized Representation Theorem for Harsanyi's 'Impartial' Observer." *Social Choice and Welfare* 39: 833–46.

Hammond, P. J. 1981. "*Ex-ante* and *Ex-post* Welfare Optimality under Uncertainty." *Economica* 48: 235–50.

Hammond, P. J. 1982. "Utilitarianism, Uncertainty and Information." A. K. Sen and B. Williams, eds., *Utilitarianism and Beyond*, 85–102. Cambridge: Cambridge University Press.

Hammond, P. J. 1983. "*Ex-post* Optimality as a Dynamically Consistent Objective for Collective Choice under Uncertainty." P. K. Pattanaik and M. Salles, eds., *Social Choice and Welfare*, 175–205. Amsterdam: North-Holland.

Hammond, P. J. 1987. "On Reconciling Arrow's Theory of Social Choice with Harsanyi's Fundamental Utilitarianism." G. R. Feiwel, ed., *Arrow and the Foundations of the Theory of Economic Policy*, 179–221. London: Macmillan; New York: New York University Press.

Harsanyi, J. C. 1953. "Cardinal Utility in Welfare Economics and in the Theory of Risk Taking." *Journal of Political Economy* 61: 434–35. Reprinted in *Essays on Ethics, Social Behavior, and Scientific Explanation.* Dordrecht: P. Reidel.

Harsanyi, J. C. 1955. "Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility." *Journal of Political Economy* 63: 309–21. Reprinted in *Essays on Ethics, Social Behavior, and Scientific Explanation.* Dordrecht: P. Reidel.

Harsanyi, J. C. 1975. "Nonlinear Social Welfare Functions: Do Welfare Economists Have a Special Exemption from Bayesian Rationality?" *Theory and Decision* 6: 311–32. Reprinted in *Essays on Ethics, Social Behavior, and Scientific Explanation.* Dordrecht: P. Reidel.

Harsanyi, J. C. 1976. *Essays on Ethics, Social Behavior, and Scientific Explanation*, Dordrecht, P. Reidel.

Harsanyi, J. C. 1977a. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.

Harsanyi, J. C. 1977b. "Nonlinear Social Welfare Functions: A Rejoinder to Professor Sen." R. E. Butts and J. Hintikka, eds., *Foundational Problems in the Special Sciences*, 293–96. Dordrecht: D. Reidel.

Harsanyi, J. C. 1977c. "Rule Utilitarianism and Decision Theory." *Erkenntnis* 11: 25–53.

Harsanyi, J. C. [1977] 1982. "Morality and the Theory of Rational Behavior." A. K. Sen and B. Williams, eds., *Utilitarianism and Beyond*, 39–62. Cambridge: Cambridge University Press. First published in *Social Research* 44: 623–56.

Harsanyi, J. C. 1979. "Bayesian Decision Theory, Rule Utilitarianism and Arrow's Impossibility Theorem." *Theory and Decision* 11: 289–317.

Harvey, C. M. 1999. "Aggregation of Individuals' Preference Intensities into Social Preference Intensity." *Social Choice and Welfare* 16: 65–79.

Hild, M., R. Jeffrey, and M. Risse. 2003. "Flipping and *Ex Post* Aggregation." *Social Choice and Welfare* 20: 267–75.

Hild, M., R. Jeffrey, and M. Risse. 2008. "Preference Aggregation after Harsanyi." M. Fleurbaey, M. Salles, and J. A. Weymark, *Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls*, 198–217. Cambridge: Cambridge University Press.

Hylland, A., and R. Zeckhauser. 1979. "The Impossibility of Bayesian Group Decision Making with Separate Aggregation of Beliefs and Values." *Econometrica* 47: 1321–36.

Jackson, M. O., and L. Yariv. 2014. "Collective Dynamic Choice: The Necessity of Time Inconsistency." Manuscript, California Institute of Technology.

Jaffray, J. Y. 1988. "Choice under Risk and the Security Factor." *Theory and Decision* 24: 169–200.

Jeffrey, R. [1965] 1983. *The Logic of Decision.* Chicago: University of Chicago Press. 2nd ed.

Kaneko, M. 1984. "On Interpersonal Utility Comparisons." *Social Choice and Welfare* 1: 165–75.

Karni, E. 1998. "Impartiality: Definition and Representation." *Econometrica* 66: 1405–15.

Karni, E. 2003. "Impartiality and Interpersonal Comparisons of Variations in Well-Being." *Social Choice and Welfare* 21: 95–111.

Karni, E. 2009. "A Reformulation of the Maxmin Expected Utility Model with Application to Agency Theory." *Journal of Mathematical Economics* 45: 97–112.

Karni, E. 2014. "Axiomatic Foundations of Expected Utility and Subjective Probability." M. J. Machina and W. K. Viscusi, eds., *Handbook of the Economics of Risk and Uncertainty*, 1–40. Amsterdam: North-Holland.

Karni, E., and Z. Safra. 2000. "An Extension of a Theorem of von Neumann and Morgenstern with an Application to Social Choice." *Journal of Mathematical Economics* 34: 315–27.

Karni, E., and Z. Safra. 2002a. "Individual Sense of Justice: A Utility Representation." *Econometrica* 70: 263–84.

Karni, E., and Z. Safra. 2002b. "Intensity of the Sense of Fairness." *Journal of Economic Theory* 105: 318–37.

Karni, E., and J. Weymark. 1998. "An Informationally Parsimonious Impartial Observer Theorem." *Social Choice and Welfare* 15: 321–32.

Keeney, R. L., and R. L. Winkler. 1985. "Evaluating Decision Strategies for Equity of Public Risks." *Operations Research* 33: 955–70.

Lerner, A. P. [1944] 1947. *The Economics of Control*. 2nd ed. London: Macmillan.

Machina, M. J. 1989. "Dynamic Consistency and Non-expected Utility Models of Choice under Uncertainty." *Journal of Economic Literature* 27: 1622–68.

Machina, M. J. 2008. "Non-expected Utility Theory." S. N. Durlauf and L. E. Blume, eds., *The New Palgrave Dictionary of Economics*, 74–84. 2nd ed. New York: Palgrave Macmillan.

Mongin, P. 1994. "Harsanyi's Aggregation Theorem: Multiprofile Version and Unsettled Questions." *Social Choice and Welfare* 11: 331–54.

Mongin, P. 1995. "Consistent Bayesian Aggregation." *Journal of Economic Theory* 66: 313–51.

Mongin, P. 1997. "Spurious Unanimity and the Pareto Principle." THEMA Working Paper, University of Cergy-Pontoise.

Mongin, P. 1998. "The Paradox of the Bayesian Experts and State-Dependent Utility Theory." *Journal of Mathematical Economics* 29: 331–61.

Mongin, P. 2001. "The Impartial Observer Theorem of Social Ethics." *Economics and Philosophy* 17: 147–80.

Mongin, P. 2002. "Impartiality, Utilitarian Ethics, and Collective Bayesianism." Cahier No. 2002-030, Laboratoire d'économétrie, Ecole Polytechnique

Mongin, P., and C. d'Aspremont. 1998. "Utility Theory and Ethics." S. Barberà, P. Hammond, and C. Seidl, eds., *Handbook of Utility Theory*, vol. 1: *Principles*, 371–481. Dordrecht: Kluwer.

Mongin, P., and M. Pivato. 2015. "Ranking Multidimensional Alternatives and Uncertain Prospects." *Journal of Economic Theory* 157: 146–71.

Myerson, R. B. 1981. "Utilitarianism, Egalitarianism, and the Timing Effect in Social Choice Problems." *Econometrica* 49: 883–97.

Nascimento, L. 2012. "The *Ex Ante* Aggregation of Opinions under Uncertainty." *Theoretical Economics* 7: 535–70.

Nehring, K. 2004. "The Veil of Public Ignorance." *Journal of Economic Theory* 119: 247–70.

Pattanaik, P. K. 1968. "Risk, Impersonality, and the Social Welfare Function." *Journal of Political Economy* 76: 1152–69.

Pivato, M. 2009. "Twofold Optimality of the Relative Utilitarian Bargaining Solution." *Social Choice and Welfare* 32: 79–92.

Pivato, M. 2013. "Risky Social Choice with Incomplete or Noisy Interpersonal Comparisons of Well-Being." *Social Choice and Welfare* 40: 123–39.

Qu, X. 2014. "Separate Aggregation of Beliefs and Values under Ambiguity." Manuscript, Université de Cergy-Pontoise.

Raiffa, H. 1968. "Decision Analysis." *Introductory Lectures on Choices under Uncertainty*, Reading, MA: Addison-Wesley.

Rawls, J. 1971. *A Theory of Justice.* Cambridge, MA: Harvard University Press.

Rawls, J. 1974. "Some Reasons for the Maximin Criterion." *American Economic Review* 64, *Papers and Proceedings*, 141–46.

Risse, M. 2001. "Instability of *Ex Post* Aggregation in the Bolker-Jeffrey Framework and Related Instability Phenomena." *Erkenntnis* 55: 239–69.

Risse, M. 2003. "Bayesian Group Agents and Two Modes of Aggregation." *Synthese* 135: 347–77.

Safra, Z., and E. Weissengrin. 2003. "Harsanyi's Impartial Observer Theorem with a Restricted Domain." *Social Choice and Welfare* 20: 177–87.

Savage, L. J. [1954] 1972. *The Foundations of Statistics*. 2nd ed. New York: Dover.

Segal, U. 2000. "Let's Agree That All Dictatorships Are Equally Bad." *Journal of Political Economy* 108: 569–89.

Sen, A. K. 1969. "Planner's Preferences, Optimality, Distribution and Social Welfare." J. Margolis and H. Guitton, eds., *Public Economics*, 201–21. London: Macmillan.

Sen, A. K. [1970] 1979. *Collective Choice and Social Welfare*. San Francisco: Holden Day. Reprint, Amsterdam: North Holland.

Sen, A. K. 1973. *On Economic Inequality*. Oxford: Clarendon Press.

Sen, A. K. 1974. "Rawls versus Bentham: An Axiomatic Examination of the Pure Distribution Problem." *Theory and Decision* 4: 301–9.

Sen, A. K. 1976. "Welfare Inequalities and Rawlsian Axiomatics." *Theory and Decision* 7: 243–62. Also in R. E. Butts and J. Hintikka, eds., *Foundational Problems in the Special Sciences*, 271–92. Dordrecht: D. Reidel, 1977.

Sen, A. K. 1977. "Non-linear Social Welfare Functions: A Reply to Professor Harsanyi." R. Butts and J. Hintikka, eds., *Logic, Methodology and Philosophy of Science*, 297–302. Dordrecht: Reidel.

Sen, A. K. 1979. "Utilitarianism and Welfarism." *Journal of Philosophy* 76: 463–89.

Sen, A. K. 1986. "Social Choice Theory." K. J. Arrow and M. Intriligator, eds., *Handbook of Mathematical Economics*, vol. 3, 1073–81. Amsterdam: North-Holland.

Starr, R. M. 1973. "Optimal Production and Allocation under Uncertainty." *Quarterly Journal of Economics* 87: 81–95.

Strotz, R. H. 1958. "How Income Ought to Be Distributed: A Paradox in Distributive Ethics." *Journal of Political Economy* 66: 189–205.

Strotz, R. H. 1961. "How Income Ought to Be Distributed: Paradox Regained." *Journal of Political Economy* 69: 271–78.

Vickrey, W. S. 1945. "Measuring Marginal Utility by Reaction to Risk." *Econometrica* 13: 319–33.

Vickrey, W. S. 1960. "Utility, Strategy, and Social Decision Rules." *Quarterly Journal of Economics* 74: 507–35.

von Neumann, J., and O. Morgenstern. [1944] 1947. *The Theory of Games and Economic Behavior*. 2nd ed. Princeton, NJ: Princeton University Press.

Weymark, J. A. 1991. "A Reconsideration of the Harsanyi-Sen Debate on Utilitarianism." J. Elster and J. E. Roemer, eds., *Interpersonal Comparisons of Well-Being*, 255–320. Cambridge: Cambridge University Press.

Weymark, J. A. 1993. "Harsanyi's Social Aggregation Theorem and the Weak Pareto Principle." *Social Choice and Welfare* 10: 209–21.

Weymark, J. A. 2005. "Measurement Theory and the Foundations of Utilitarianism." *Social Choice and Welfare* 25: 527–55.

Zhou, L. 1997. "Harsanyi's Utilitarianism Theorems: General Societies." *Journal of Economic Theory* 72: 198–207.