Harsanyi’s theorem without the sure-thing principle

Stéphane Zuber

Abstract

We study a version of Harsanyi’s theorem (Harsanyi, 1955) in a framework involving uncertainty. Without assuming the sure-thing principle but sticking to consequentialism, we obtain that a Paretian social aggregation should be affine and that all decision makers must use additively separable preferences that resembles expected utilities. Whenever preferences are state independent, we find that decision makers must be expected utility maximizers and share the same beliefs. The sure-thing principles is hence necessary for the social aggregation to be possible. And consequentialist non-expected utility models cannot be obtained as an aggregation of preferences.

Keywords: Harsanyi’s Theorem, Pareto Principle, Sure-Thing Principle, Consequentialism.

JEL Classification numbers: D6, D8.

---

1Université Catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium. E-mail: Stephane.Zuber@uclouvain.be.
1 Introduction

Harsanyi (1955) proved the following theorem: if (i) all individuals and the social observer are expected utility maximizers on the set of simple lotteries over a set $K$; (ii) whenever all individuals prefer lottery $p$ to lottery $q$, the social observer must also prefer $p$ to $q$ (the Pareto principle); then the social observer’s utility function must be an affine combination of individuals’ von Neumann and Morgenstern utility functions. This result is known as Harsanyi’s aggregation theorem.

Harsanyi’s aggregation theorem has spawned a prolific literature. In particular, several versions of Harsanyi’s aggregation theorem have been established for different expected utility models in frameworks involving uncertainty, as opposed to risk (see Knight, 1921, on the distinction between risk and uncertainty). Noticeable references are De Meyer and Mongin (1995), Mongin (1995), Mongin (1998) and Blackorby, Donaldson and Weymark (1999).

Harsanyi’s result has however been criticized by Diamond (1967) on fairness considerations. Diamond provided an example suggesting that Harsanyi’s principles cancel out any possibility of redistributing utility ex ante. He hence suggested to abandon the expected utility model at the societal level. Diamond’s suggestion was followed by Epstein and Segal (1992) and, in the context of Harsanyi’s impartial observer theorem, by Grant, Polak, Kajii, Safra (2007).

The difficulty with the solutions proposed by Diamond (1967), Epstein and Segal (1992) and Grant, Polak, Kajii, Safra (2007) is that they yield the social observer to be non-consequentialist. Consequentialism requires choices in each state of the world to be independent of what occurs in unrealized states of the world. It enables the use of standard recursive technique in dynamic settings. On the contrary, non-consequentialist criteria are prone to induce time inconsistencies in the usual - consequentialist - sense (see however the discussion by Machina, 1989, and Epstein and Segal, 1992, who suggest a broader definition of time consistency). They also generically violate the following dominance property (Fleurbaey, 2007): if in each state of the world the distribution of utilities in an alternative $x$ dominates the distribution of utilities in an alternative $\hat{x}$, the social observer should prefer $x$ to $\hat{x}$.

The purpose of this paper is to take up the issue considered by Harsanyi, namely the aggregation of preferences under uncertainty. But we considerably widen the investigation by considering a class of preferences, regular consequentialist preferences, that encompasses many models of decision under uncertainty.
This widening is primarily motivated by Diamond’s objection: we do not want to impose subjective expected utility at the societal level. However, to avoid the difficulties of non-consequentialist preferences, we follow the vast majority of the decision theoretic literature in sticking to consequentialism. This permits to include many models of decision under uncertainty that have been proposed recently, in particular the Choquet expected utility model by Schmeidler (1989), the multiple-prior preferences by Gilboa and Schmeidler (1989), the variational preferences of Maccheroni, Marinacci and Rustichini (2006) and the second order expected utility model by Klibanoff, Marinacci and Mukerji (2005) and Ergin and Gul (2009).

Our general framework also encompasses most settings that have been proposed to model uncertainty, for instance the ones by Arrow (1953), Savage (1954), Anscombe and Aumann (1963), or Karni, Schmeidler and Vind (1983). In particular, preferences are allowed to be state dependent.

We obtain a result very close to Harsanyi’s. Under the strong Pareto principle, social preferences must be represented by an affine aggregation of individuals’ utilities. Besides, the aggregation is possible only when all decision makers’ preferences satisfy the sure-thing principle. Under consequentialist principles, we hence obtain a generalized version of the subjective expected utility model as a result. This suggests that Diamond’s proposal can only be implemented in a non-consequentialist framework. Besides, whenever preferences are also required to be state independent, the aggregation is possible only if all decision makers are expected utility maximizers and use the same probability weights.

These results are obtained in two frameworks. In our general framework, we have to make assumptions on the utility possibility set and the form of social indifference curves. The assumption on the utility set (convexity and existence of a worst element) is satisfied in many problems. The assumption on social indifference curves (concavity) represents inequality aversion with respect to the utility profile under consideration.

In the more specific Anscombe-Aumann framework, we are able to obtain sharper results. Under two common choice-theoretic assumptions (independence for preferences over roulette lotteries and independent prospects for individuals), we find that the social aggregation is possible only if decision makers are subjective expected utility maximizers. The result is more in line with Harsanyi’s approach which consists in deriving ethical consequences from choice-theoretic assumptions and the Pareto principle.

Our findings complement several previous results in the literature. In particular, four recent studies have tried to investigate the aggregation of preferences under un-
certainty when preferences are not expected utilities. Chambers and Hayashi (2006) considered non-expected utility preferences that may not satisfy the sure-thing principle (P2) of Savage. In a Savage framework, they showed that, when individuals have different beliefs, eventwise monotonicity (P3) is incompatible with the Pareto axiom. Our setting relaxes axiom (P3) to allow for state-dependence preferences. Yet, we are able to obtain an impossibility similar to the one they describe.

Gajdos, Tallon and Vergnaud (2008) study the aggregation of what they call ‘rank-dependent additive preferences.’ These possibly state dependent preferences are obtained by weakening the independence axiom in the Anscombe-Aumann framework. They prove that the aggregation is affine and that it can be obtained if and only if the society is uncertainty neutral. However, the aggregation result is obtained on part of the domain of acts only. They do not identify the expected utility model as a requirement. The present paper thus significantly enlarges the knowledge we have on the conditions that are necessary to aggregate preferences under uncertainty.

Fleurbaey (2009) relaxes the expected utility at the societal level for a more general model, close to our regular consequentialist preferences. But he sticks to expected utility for individuals. He finds that the social aggregation must be affine. Our paper broaden the investigation to non expected-utilities individual preferences. We also do not assume objective probabilities. In this more general context, we not only prove that the social aggregation must be affine but also that individual preferences must satisfy the sure-thing principle in order to be aggregated.

The paper the closest to ours is the one by Blackorby, Donaldson and Mongin (2004). They also studied the aggregation of preferences in a large class of non-expected utility models. However, the framework they adopt is different from the one of the present paper. They jointly study the aggregation of beliefs and utilities, while we do not take beliefs as given. They also make several assumptions (most notably a normalization condition close to state independence and a strong assumption on the structure of the utility possibility sets) that we relax. In the end, they do obtain that the social aggregation should be affine, but they have weaker conclusions concerning the structure of preferences required for the aggregation to be possible.

Beside these four papers, our results are related to several other results in the literature. The aggregation result we obtain in Theorem 3 is similar to the one by Mongin (1998). The identity of the probability weights required in Theorem 2 conforms with the ill-named ‘probability agreement theorem’, which actually require beliefs to be the same (rather than showing that they are). It relates to a line of inquiry investigated
by several authors (Hylland and Zeckhauser, 1979; Hammond, 1981; Broome, 1990; Mongin, 1995). All conclude that the aggregation of subjective expected utility agents’ preferences is not possible as soon as they have different beliefs. We obtain the result on much weaker grounds.

The remainder of the paper is divided into four sections and one appendix. In Section 2, we lay down the model, define regular consequentialist preferences and describe the social aggregation of preferences. In Section 3, we present the main theorem (Theorem 1) and a variant thereof (Theorem 1’). We also examine the additional restrictions imposed by state-independence: the aggregation of preferences is possible only when decision makers are expected utility maximizers and share the same beliefs (Theorem 2). In Section 4, we focus on the more convenient Anscombe-Aumann framework. Assuming that decision makers’ preferences satisfy a property of statewise independence, we show that the familiar condition of ‘independent prospects’ is sufficient to obtain our main result (Theorem 3). Section 5 contains some concluding remarks. An appendix presents and discusses a result by Chateauneuf and Wakker (1991) used in the proof of Theorems 1’ and 3.

2 The framework

2.1 Preferences over state contingent alternatives

Consider the following framework of choice over state contingent alternatives. Let $S$ be a finite state space. To save on notation, $S > 2$ also denotes the number of states of the world. For each state $s \in S$, the set of state contingent alternatives is $X_s$ with typical element $x_s \in X_s$. We assume that each $X_s$ is a connected subset of a metric space. The set of social alternatives or prospects is $X = \prod_{s \in S} X_s$. Because each $X_s$ is a connected subset of a metric space, so is $X$. A typical element of $X$ is $x = (x_1, \cdots, x_S)$, with $x_s$ the outcome if state $s$ occurs.

We consider decision makers with preferences $\succeq$ on $X$. Preferences satisfy the following three axioms.

Ordering (O). $\succeq$ is a complete and continuous weak order on $X$.

As usual, we let $\succ$ and $\sim$ respectively denote the asymmetric and symmetric parts of $\succeq$. 

4
Non Degeneracy (ND). For each \( s \in S \), there exist \( \bar{x}_s \) and \( \bar{x}_s' \) in \( X_s \) such that, whenever \( x \) and \( \hat{x} \) in \( X \) are such that \( x_s = \bar{x}_s \), \( \hat{x}_s = \bar{x}_s' \), and \( x_{s'} = \hat{x}_{s'} \) for all \( s' \in S \setminus \{s\} \) then \( x \succ \hat{x} \).

Monotonicity (MON). For each \( s \in S \), if the four social alternatives \( x, \hat{x}, x' \) and \( \hat{x}' \) are such that \( x_s = x'_s \), \( \hat{x}_s = \hat{x}'_s \), and, for all \( s' \neq s \), \( x_{s'} = \hat{x}_{s'} \) and \( x'_{s'} = \hat{x}'_{s'} \), then

\[ x \succeq \hat{x} \iff x' \succeq \hat{x}' \]

Axioms (O) and (ND) imply that there exists a non-constant and continuous utility function \( U : X \to \mathbb{R} \) representing \( \succeq \). Axiom (MON) implies that we can define continuous state contingent preference orderings \( \succeq_s \) representing the decision maker’s preferences over state contingent alternatives. There exist non-constant continuous state contingent utility functions \( u_s : X_s \to \mathbb{R} \) representing these preferences.\(^1\)

Together the three axioms yield the following preference structure:

**Lemma 1** If the decision maker’s preferences \( \succeq \) satisfy Axioms (O), (ND) and (MON), then there exist continuous state contingent utility functions \( u_s \) and a continuous aggregator \( F : \mathbb{R}^s \to \mathbb{R} \) increasing in its components such that the function \( U \) represents \( \succeq \) where \( U \) is defined as follows:

\[
U(x) = F\left(u_1(x_1), \ldots, u_S(x_S)\right)
\]

**Proof.** The proof involves standard separability arguments (see for instance Blackorby, Primont, Russel, 1999). It is therefore omitted. \( \blacksquare \)

We will say that a decision maker whose preferences have the structure exhibited in Lemma 1 has **regular consequentialist preferences**. Preferences are called consequentialist because the state contingent preferences are independent of unrealized states of the world: they only depend on the actual consequences. The term ‘regular’ refers to the fact that ex ante judgments depends on ex post judgments so that a social alternative which is always preferred ex post will also be preferred ex ante.

\(^1\)Alternatively, we could assume the existence of non-constant continuous state contingent preference orderings \( \succeq_s \) and impose a consistency axiom named dominance:

\[
\forall x, \hat{x} \in X,

x_s \succeq_s \hat{x}_s \forall s \in S \implies x \succeq \hat{x}
\]

If furthermore, \( \exists s \in S : x_s \succ \hat{x}_s \) then \( x \succ \hat{x} \).
Axiom (MON) weakens well-known axioms that are commonly used in the theory of decision under uncertainty. It is of course a weakening of the sure-thing principle (axiom P2 of Savage):

**Sure-Thing Principle (STP).** For any $\bar{S} \subset S$, if the four social alternatives $x$, $\hat{x}$, $x'$ and $\hat{x}'$ are such that $x_s = x'_s$ for all $s \in \bar{S}$, $\hat{x}_s = \hat{x}'_s$ for all $s \in S \setminus \bar{S}$ and $x'_{s'} = \hat{x}'_{s'}$ for all $s' \in S \setminus \bar{S}$, then

$$x \succeq \hat{x} \iff x' \succeq \hat{x}'$$

Combined with (O) and (ND), (STP) delivers the following additively separable representation of preferences that generalizes the state contingent expected utility model:

$$U(x) = \Phi\left(u_1(x_1), \cdots + u_S(x_S)\right)$$

These preferences are a special case of regular consequentialist preferences.

Monotonicity also weakens axiom P3 of Savage which is known as the eventwise monotonicity property. To introduce the property, assume that $X_s = X$ for all $s \in S$, so that $X = X^S$. For any $x \in X$, denote $x^c \in X$ the social alternative such that $x^c_s = x$ for all $s \in S$.

**Eventwise Monotonicity (EMON).** For any $s \in S$, if the social alternatives $x$ and $\hat{x}$ are such that $x_s \neq \hat{x}_s$ and $x_{s'} = \hat{x}_{s'}$ for all $s' \in S \setminus \{s\}$, then

$$x \succeq \hat{x} \iff x^c_s \succeq \hat{x}^c_s$$

Eventwise monotonicity (EMON) is similar to (MON) but it further entails that the state contingent utility functions are the state-independence. Indeed, it requires that

$$u_s(x_s) \geq u_s(\hat{x}_s) \iff U(x^c_s) \geq U(\hat{x}^c_s),$$

whatever the state of the world $s \in S$. Defining

$$\Phi\left(u_1(x_1), \cdots + u_S(x_S)\right)$$

Thus, it is also weaker the following monotonicity requirement commonly used in the literature (see for instance Ghirardato, Maccheroni, Marinacci, 2004; and Maccheroni, Marinacci, Rustichini, 2006):

**P3.** For any $\bar{S} \subset S$, if the social alternatives $x'$ and $\hat{x}'$ are such that $x'_s = x$ for all $s \in \bar{S}$, $\hat{x}'_s = \hat{x}$ for all $s \in S \setminus \bar{S}$, and $x'_{s'} = \hat{x}'_{s'}$ for all $s' \in S \setminus \bar{S}$, then

$$x' \succeq \hat{x}' \iff x^c \succeq \hat{x}^c$$

(MON). For any $x$ and $\hat{x}$ in $X$, if $x^c_s \geq \hat{x}^c_s$ for all $s \in S$, then $x \succeq \hat{x}$. 

---

\footnote{The formulation of eventwise monotonicity that we propose here is actually closer to Assumption 1 in Anscombe and Aumann (1963) and weaker than the strict formulation of P3 in Savage which is as follows:

**P3.** For any $\bar{S} \subset S$, if the social alternatives $x'$ and $\hat{x}'$ are such that $x'_s = x$ for all $s \in \bar{S}$, $\hat{x}'_s = \hat{x}$ for all $s \in S \setminus \bar{S}$, and $x'_{s'} = \hat{x}'_{s'}$ for all $s' \in S \setminus \bar{S}$, then

$$x' \succeq \hat{x}' \iff x^c \succeq \hat{x}^c$$}
$u(x) \equiv U(x^c)$, we obtain the following structure for preferences when (EMON) is combined with (O) and (ND):

$$U(x) = \tilde{F}(u(x_1), \ldots, u(x_S))$$

with $\tilde{F}$ such that $\tilde{F}(u, \ldots, u) = u$.

Using regular consequentialist preferences, we are therefore able to encompass many decision models that have been suggested in the literature on decision under uncertainty. We allow for state dependent models. We also allow for non-expected utility models.

### 2.2 The aggregation of preferences

We consider a society composed of $N$ members and denote $I = \{1, \ldots, N\}$ the set of individuals. There is another decision maker in the society, namely the social observer. She is indexed by 0. We let $I^* = \{0, 1, \ldots, N\}$.

All decision makers (both individuals and the social observer) have regular consequentialist preferences on $X$. These preferences are represented by utility functions $U^i$, $i \in I^*$. We also let $u^i_s$ denote a state contingent utility function of decision maker $i$ and $F^i$ denote the associated aggregator function that defines preferences like in Equation (1).

The aim of a social aggregation is to derive the social preference ordering from individual utility functions. Many ethical principles can be invoked to determine how this aggregation should be made. But the most widely accepted principle is that social preferences should be Paretian.

A weak version of the Pareto principle is as follows:

**Pareto Indifference (PI).** For any $x$ and $\hat{x}$ in $X$, if $U^i(x) = U^i(\hat{x})$ for all $i \in I$ then $U^0(x) = U^0(\hat{x})$.

A stronger version is Strict Pareto:

**Strict Pareto (Strict P).** For any $x$ and $\hat{x}$ in $X$, if $U^i(x) \geq U^i(\hat{x})$ for all $i \in I$ and there exists $j \in I$ such that $U^j(x) > U^j(\hat{x})$ then $U^0(x) > U^0(\hat{x})$.

In this paper, we shall consider a combination of these two principles:

**Strong Pareto (SP). (PI) and (Strict P).**
Strong Pareto implies that social preferences can be derived from an aggregation of individuals’ preferences in a sense specified by the next lemma:

**Lemma 2 (Blackorby, Donaldson, Weymark, 1999)** The utility functions $U^i$ on $X$, $i \in I^*$, satisfy axiom (SP) if and only if there exists a social aggregator function $W$ which is continuous and increasing in each of its components and such that:

$$U^0(x) = W\left(U^1(x), \cdots, U^N(x)\right)$$  \hspace{1cm} (2)

**Proof.** For any $x$ and $\hat{x}$ in $X$, if $U^i(x) = U^i(\hat{x})$ for all $i \in I$ then $U^0(x) = U^0(\hat{x})$. Under Pareto Indifference, Blackorby, Donaldson and Weymark (1999: Lemma 1, p. 369; Lemma 2, p.370) show that there must exist a continuous social aggregator function $W$ such that $U^0(x) = W\left(U^1(x), \cdots, U^N(x)\right)$. Axiom (Strict P) implies that $W$ must be increasing in each of its components. ■

Equation (2) indicates how preferences are aggregated ex ante. But, when all decision makers have regular consequentialist preferences, it is also the case that preferences can be aggregated ex post. This possibility of ex post aggregation, which was assumed by most of the existing literature (in particular Hammond, 1981, and Myerson, 1981), is proved by Lemma 3.

**Lemma 3** If the utility functions $U^i$ on $X$, $i \in I^*$ satisfy axiom (SP) and if all decision makers have regular consequentialist preferences, then, for any $s \in S$, there exists a social aggregator function $w_s$ which is continuous and increasing in each of its components and such that:

$$u^0_s(x_s) = w_s\left(u^1_s(x_s), \cdots, u^N_s(x_s)\right)$$  \hspace{1cm} (3)

**Proof.** We start by proving that a state contingent Strong Pareto property must hold. Assume that $u^i_s(x_s) \geq u^i_s(\hat{x}_s)$ for all $i \in I$ for some $x_s$ and $\hat{x}_s$ belonging to $X_s$. Then consider $x'$ and $\hat{x}'$ such that $x'_s = x_s$, $\hat{x}'_s = \hat{x}_s$ and $x'_{s'} = \hat{x}'_{s'}$ for all $s' \neq s$. By monotonicity, it must be the case that $U^i(x') \geq U^i(\hat{x}')$ for all $i \in I$. By (SP), this implies $U^0(x') \geq U^0(\hat{x}')$. But given the definitions of $x'$ and $\hat{x}'$ and given that the social planner’s preferences state monotonicity, this is possible only if $u^0_s(x_s) \geq u^0_s(\hat{x}_s)$.

As a consequence, we obtain that $\left(u^i_s(x_s) \geq u^i_s(\hat{x}_s)\right)$ $\forall i \in I \implies u^0_s(x_s) \geq u^0_s(\hat{x}_s)$. If furthermore there exists $j \in I$ such that $u^j_s(x_s) > u^j_s(\hat{x})$ then $u^0_s(x_s) > u^0_s(\hat{x}_s)$.

8
Because the utility functions $u^i_s$ on $X_s$, $i \in I^*$, satisfy this state contingent Strong Pareto axiom, the same line of arguments as the one used for Lemma 2 yields the result.

The congruence between ex ante and ex post aggregations of preferences (as described by Lemmata 1 and 2) has been studied by Hammond (1981), Myerson (1981), Blackorby, Donaldson and Weymark (1999) and Blackorby, Donaldson and Mongin (2004). The first three studies restrict attention to situations where all decision makers have state independent expected utility preferences. The fourth study jointly studies the aggregation of utilities and beliefs; it also makes a normalization assumption close to state independence.

The present paper is more general since it allows for state dependent non-expected utility preferences without assuming the existence of well-defined beliefs. Though we are able to reach a very similar conclusion.

3 A generalization of Harsanyi’s theorem

3.1 The basic result

In this section, we present a basic aggregation result that generalizes Harsanyi’s theorem. The result contains the essence of why consequentialism and the Pareto principle require additive representation of preferences.

Denote $U_s = \left\{ (u^1_s, \cdots, u^N_s) \in \mathbb{R}^N : \exists x_s \in X_s, u^i_s = u^i_s(x_s) \forall i \in I \right\}$ the utility possibility set in state of the world $s \in S$ for a specific collection of state dependent utility functions $u^i_s$, $i \in I$. Also denote $u^i_s(X_s) = \left\{ u^i_s \in \mathbb{R}^N : \exists x_s \in X_s, u^i_s = u^i_s(x_s) \right\}$ the range of the utility function $u^i_s$.

**Assumption 1** For all $s \in S$, there exist a collection of state dependent utility functions $u^i_s$, $i \in I$, representing individuals’ preferences and such that $U_s = \prod_{i \in I} u^i_s(X_s)$.

Assumption 1 bears on the form of the ex post utility possibility set. This is typical in this literature (see Blackorby, Donaldson and Weymark, 1999). Assumption 1 is rather substantial though. The Cartesian product structure for the utilities typically precludes the consideration of interdependencies between individuals’ utilities, whether they are psychological interdependencies or interdependencies mediated by a distributional constraint. The assumption is similar to Assumption 4 in Blackorby, Donaldson and Mongin (2004). It is motivated by the following example.
Example 1 Let $m$ be the number of commodities in the economy in each state of the world and $\mathbb{R}^N_+$ represent the set of possible allocations of baskets of commodities between the $N$ individuals. Suppose that individuals’ ex post preferences $\succeq^i_s$ are individualistic and satisfy the usual properties of monotonicity and convexity. For any collection of utility functions $u^i_s$, Assumption 1 is satisfied.

Under Assumption 1, we have the following generalization of Harsanyi’s theorem:

Theorem 1 Assume that the social observer and all individuals have regular consequentialist preferences. Assume furthermore that Assumption 1 is satisfied for a collection of state contingent utility functions $(u^i_s)_{i \in I, s \in S}$. Axiom (SP) holds if and only if there exist continuous increasing transformations $\psi_s$ and $\varphi^i_s$ such that, whatever $x \in X$:

1. $U^0(x) = U^1(x) + \cdots + U^N(x)$ and $u^0_s(x_s) = \psi^{-1}_s\left(\sum_{i \in I^*} \varphi^i_s\left(u^i_s(x_s)\right)\right)$.

2. (a) For each $i \in I$, $U^i(x) = \sum_{s \in S} \varphi^i_s\left(u^i_s(x_s)\right)$, represents $i$’s preferences.

(b) $U^0(x) = \sum_{s \in S} \psi_s\left(u^0_s(x_s)\right)$, represents social preferences.

Proof. Consider a utility profile satisfying the conditions in Assumption 1.

The social observer has regular consequentialist preferences. Hence Equation (1) yields: $U^0(x) = F^0\left(u^0_1(x_1), \cdots, u^0_S(x_S)\right)$. Because of the ex post aggregation of preferences expressed in Equation (3), this equality can be rewritten:

$$U^0(x) = F^0\left(w_1\left(u^1_1(x_1), \cdots, u^1_S(x_S)\right), \cdots, w_S\left(u^0_1(x_S), \cdots, u^0_S(x_S)\right)\right)$$

(4)

On the other hand, we know (Equation (2)) that $U^0(x) = W\left(U^1(x), \cdots, U^N(x)\right)$. Applying the characterization of regular consequentialist preferences to individuals’ preferences, we obtain that:

$$U^0(x) = W\left(F^1\left(u^1_1(x_1), \cdots, u^1_S(x_S)\right), \cdots, F^N\left(u^N_1(x_1), \cdots, u^N_S(x_S)\right)\right)$$

(5)

Combining Equations (4) and (5), we end up with:

$$F^0\left(w_1\left(u^1_1(x_1), \cdots, u^1_1(x_1)\right), \cdots, w_S\left(u^0_1(x_S), \cdots, u^0_S(x_S)\right)\right) =$$

$$W\left(F^1\left(u^1_1(x_1), \cdots, u^1_S(x_S)\right), \cdots, F^N\left(u^N_1(x_1), \cdots, u^N_S(x_S)\right)\right)$$

(6)
Using the notation $U_s$ introduced in Assumption 1, define $u = (u_1, \ldots, u_S) \in \prod_{s \in S} U_s$, where $u_s$ has the form $u_s = (u^s_1, \ldots, u^s_N)$. Whatever $u \in \prod_{s \in S} U_s$, there exists $x \in X$ such that, $\forall i \in I$ and $\forall s \in S$, $u^i_s = u^i_s(x_s)$. And, since Equation (6) is true for any $x \in X$, the following functional equation holds for any $u \in \prod_{s \in S} U_s$:

$$F^0(w_1(u^1_1, \ldots, u^1_N), \ldots, w_S(u^S_1, \ldots, u^S_N)) = W(F^1(u^1_1, \ldots, u^1_S), \ldots, F^N(u^N_1, \ldots, u^N_S))$$

(7)

Denote $U_s^o$ the relative interior of $U_s$. By Assumption 1 and Axiom (ND), it is rectangular and non-empty. Hence the set $\prod_{s \in S} U_s^o$ is also open, rectangular and non-empty. Given that the functions $F^0, W, w_s$ and $F^i$ are all continuous and increasing, the solution of Equation (7) on $\prod_{s \in S} U_s^o$ is (Aczel and Maksa, 1996, Theorem 5):

$$F^0(w_1, \ldots, w_S) = \Phi \left( \sum_{s \in S} \psi_s(w_s) \right)$$

(8)

$$W(f^1, \ldots, f^N) = \Phi \left( \sum_{i \in I} \phi_i(f^i) \right)$$

(9)

$$\forall s \in S, \quad w_s(u^1_s, \ldots, u^N_s) = \psi_s^{-1} \left( \sum_{i \in I} \phi_s^i(u^i_s) \right)$$

(10)

$$\forall i \in I, \quad F^i(u^1_i, \ldots, u^i_S) = \phi_i^{-1} \left( \sum_{s \in S} \phi_s^i(u^i_s) \right)$$

(11)

where the real functions $\Phi, \psi_s, \phi_i, \phi_s^i$ are continuous and increasing.

The additive representation on $U_s^o$ extends to $U_s$ by continuity. ■

The first part of Theorem 1 is really the extension of Harsanyi’s results: the social welfare function must be an affine aggregation of an individual utility profile. Several authors have discussed the ethical significance of the theorem.$^3$ But, in any case, Harsanyi’s theorem is remarkable. First, it is a possibility social choice theorem. Second, it shows that the Pareto principle implies a very specific form for the social aggregation whenever individuals’ preferences respect particular choice-theoretic restrictions. This form is best described as generalized utilitarian (Mongin and d’Aspremont, 1998).

The second part of Theorem 1 can be viewed as a characterization theorem. Within the class of regular consequentialist preferences, only those satisfying condition 2 can

---

$^3$See Sen (1986) for the original argument. See also Weymark (1991) and Mongin and d’Aspremont (1998) for clear and comprehensive reviews of the literature on Harsanyi’s theorem, with different views on the ethical significance of the theorem.
be aggregated under the Strong Pareto requirement. As explained in Section 2, the additively separable form of preferences in condition 2 is obtained when axioms (O) and (ND) are combined with the sure-thing principle (STP). This model corresponds to a generalization of the expected utility model to state dependent preferences.

The intuition behind the theorem is that the standard Pareto principle and consequentialist principles are separability conditions. Combined together, they have more power than is typically acknowledged. Indeed, consequentialism induces separability between states of the world, while Pareto implies separability between individuals. This provides a sufficient separability structure to obtain the complete separability to which additive representations of preferences correspond. A Paretian social observer must take into account what happens to individuals in the different states of the world. But ex post she cannot do so. This would break the consistency between ex ante and ex post judgements unless individual’s preferences are additively separable.

### 3.2 Extension to non product utility possibility sets

Assumption 1 rules out interdependencies between individuals’ utilities. In this section, we weaken the assumption to allow typical economic environments where resources must be shared between individuals. We are able to obtain the same result as in Theorem 1. This extension has a cost: we need to impose a restriction on the form of social aggregation functions.

**Assumption 1’** For all $s \in S$, there exist state dependent utility functions $u_i^s$, $i \in I$, representing individuals’ preferences and such that:

1. The set $U_s = \left\{ (u_1^s, \ldots, u_N^s) \in \mathbb{R}^N : \exists x_s \in X_s, u_i^s = u_i^s(x_s) \forall i \in I \right\}$ is convex and has a non-empty relative interior.

2. There exists $x_s \in X_s$ such that, for all $x_s \in X_s$, $u_i^s(x_s) > u_i^s(x_s)$ for all $i \in I$.

3. Each function $w_s$ such that $u_0^s(x_s) = w_s\left(u_1^s(x_s), \ldots, u_N^s(x_s)\right)$ is quasi concave on $U_s$.

Condition 1 resembles the strong regularity condition of Blackorby, Donaldson and Weymark (1999). The connectedness of the utility possibility set is however strengthened here to convexity, a very common assumption in the context of Harsanyi’s theorem (it is emphasized for instance by De Meyer and Mongin, 1995). The non-empty interior
assumption ensures that the utility possibility set is of full dimension. In particular, the assumption rules out that individuals all have the same preferences. Hence the condition can be viewed as a preference diversity condition, which is also typical in the literature (our Assumption 2 in Section 4 will play the same role).

Condition 2 of Assumption 1’ states that there exists a situation that is unanimously deemed worse than any other. Condition 1 and 2 are motivated by the following canonical allocation problem.

**Example 2** Let \( m \) be the number of commodities in the economy in each state of the world. Let \( x^i_{sm} \) be the consumption of good \( m \) by individual \( i \) in state of the world \( s \). An allocation is a list \( x_s = (x^1_{s1}, \ldots, x^N_{s1}, \ldots, x^1_{sm}, \ldots, x^N_{sm}) \in \mathbb{R}^{mN} \) such that \( \sum_{i \in I^s} x^i_{sm} \leq y_{sm} \). The quantity \( y_{sm} \) is the quantity of good \( m \) to be shared in state of the world \( s \). Suppose that individuals’ ex post preferences \( \succeq^i_s \) are individualistic and satisfy the usual properties of monotonicity and convexity. For any collection of utility functions \( u^i_s \), conditions 1 and 2 of Assumption 1’ are satisfied.

Condition 3 of Assumption 1’ represents the idea that the social observer is inequality averse in each state \( s \) for the utility profile under consideration. Combined with a symmetry condition, quasi concavity of \( w_s \) would indeed imply that \( w_s \) is Schur-concave. Define \( u^i_s(x_s) \) as the \( N \)-dimensional vector whose \( i \)th component is \( u^i_s(x_s) \). When \( w_s \) is Schur-concave, it is the case that, whatever \( x_s \in X_s \), if \( \hat{x}_s \in X_s \) is such that \( u^i_s(\hat{x}_s) = Q \cdot u^i_s(x_s) \), with \( Q \) a bistochastic matrix, then \( w_s\left(u^1_s(\hat{x}_s), \ldots, u^N_s(\hat{x}_s)\right) \geq w_s\left(u^1_s(x_s), \ldots, u^N_s(x_s)\right) \). The transformation \( u^i_s(\hat{x}_s) = Q \cdot u^i_s(x_s) \) means that the utility profile \( u^i_s(x_s) \) can be obtained from the utility profile \( u^i_s(x_s) \) by a sequence of Pigou-Dalton utility transfers. Hence \( \hat{x}_s \) is ‘less unequal’ while having the save average utility. Condition 3 ensures that the social observer prefers the ‘less unequal’ distribution of utility.

Under Assumption 1’, we obtain an alternative version of Theorem 1:

**Theorem 1’** Assume that the social observer and all individuals have regular consequentialist preferences. Assume furthermore that Assumption 1’ is satisfied for a collection of state contingent utility functions \( (u^i_s)_{i \in I, s \in S} \). Axiom (SP) holds if and only if there exist continuous increasing transformations \( \psi_s \) and quasi-concave continuous increasing transformations \( \phi^i_s \) such that, whatever \( x \in X \):

1. \( U^0(x) = U^1(x) + \cdots + U^N(x) \) and \( u^0_s(x_s) = \psi^{-1}_s\left(\sum_{i \in I^s} \phi^i_s\left(u^i_s(x_s)\right)\right) \)
2. (a) For each $i \in I$, $U^i(x) = \sum_{s \in S} \varphi^i_s(u^i_s(x_s))$, represents $i$’s preferences.

(b) $U^0(x) = \sum_{s \in S} \psi_s(u^0_s(x_s))$, represents social preferences.

Proof. With the same line of reasoning as in the proof of Theorem 1, we obtain that the functional equation (7) holds for any $u \in \prod_{s \in S} U_s^o$. By Assumption 1’, $U_s^o$ is convex and non-empty. Hence the set $\prod_{s \in S} U_s^o$ is also open, convex and non-empty. Consider any $u \in \prod_{s \in S} U_s^o$. There necessarily exists a rectangular subset of $\prod_{s \in S} U_s^o$ in the neighborhood of $u$ such that the functional equation (7) holds. On this rectangular set, Theorem 1 applies. We hence obtain an additively separable representation of preferences in any rectangular neighborhood in the interior of the utility possibility set. By quasi concavity of the ex post social aggregation function, it must also be the case that all functions $\varphi^i_s$ are quasi concave.

The additively separable representation holds locally. As discussed by Chateauneuf and Wakker (1993), it is not always possible to extend local additive representations to global additive representations, even if the local additive representation holds in any neighborhood (as it is the case here). But it is showed in the Appendix that this extension is possible whenever Assumption 1’ holds. ■

Theorem 1’ extends the basic result to many natural economic environments. This extension has a cost, namely condition (iii) in Assumption 1’. Although the condition may seem natural, it is non-typical in the literature on Harsanyi’s theorem. The originality of Harsanyi’s approach was to derive ethical conclusions from the sole Pareto principle combined with choice-theoretic premises. We have argued that condition (iii) is an additional ethical requirement. Theorem 1’ wanders off the pure Harsanyi’s approach, but we believe that it is in line with Harsanyi’s research program.

In the next section, we show that the expected utility model with well-defined beliefs would be obtained if (MON) is strengthened to (EMON). We also obtain the known result that the Paretian aggregation is possible only if all decision makers share the same beliefs.

---

4Harsanyi himself introduced an anonymity requirement to represent fairness concerns in Harsanyi (1955). Although the application of the axiom was not formally correct (see Mongin and d’Aspremont, 1998, for a correct use of the axiom), this indicates that equity concerns are not in contradiction with Harsanyi’s approach.
3.3 State independent preferences and beliefs

In order to consider a state independent model of choice, assume that $X_s = X$ for all $s \in S$, so that $X = X^S$. We say that a decision maker has state independent regular consequentialist preferences whenever her preferences satisfy axioms (O), (ND) and (EMON). We use this denomination because (EMON) implies (MON) in the framework of this section. And we know that it guarantees that state-contingent preferences are state independent.

We obtain the following theorem for state independent regular consequentialist preferences.

**Theorem 2** Assume that the social observer and all individuals have state independent regular consequentialist preferences. Assume furthermore that Assumption 1 (or 1’) is satisfied for a collection of state contingent utility functions $(u^i)_{i \in I}$. Axiom (SP) holds if and only if there exist scalars $p_s \in (0, 1)$, with $\sum_{s \in S} p_s = 1$, and (quasi-concave) continuous increasing transformations $\tilde{\varphi}^i$ such that, whatever $x \in X$:

1. $U^0(x) = U^1(x) + \cdots + U^N(x)$ and $u^0(x_s) = \sum_{i \in I} \tilde{\varphi}^i(u^i(x_s))$.

2. (a) For each $i \in I$, $U^i(x) = \sum_{s \in S} p_s \tilde{\varphi}^i(u^i(x_s))$, represents $i$’s preferences.
   (b) $U^0(x) = \sum_{s \in S} p_s u^0_s(x_s)$, represents social preferences.

**Proof.** Let $(u^i)_{i \in I^*}$ be representations of state contingent preferences satisfying Assumption 1 (or 1’). Since (EMON) implies (MON), Theorem 1 (1’) yields that the social observer’s preferences are represented by $u^0_s(x_s) = \psi_s^{-1}\left(\sum_{i \in I} \varphi^i_s(u^i_s(x_s))\right)$ in state $s \in S$. By (EMON), it must be the case that, whatever $i \in I$ and $s \in S$, $\varphi^i_s(u^i_s(x)) = \psi^i_s(u^i_s(x))$, for some continuous increasing function $f^i_s$. Since the social observer has state independent preferences, there must exist an increasing function $G_s$ such that, for all $s \neq 1$, $\sum_{i \in I} f^i_s(u^i(x)) = G_s\left(\sum_{i \in I} f^i_s(u^i(x))\right)$ whatever $x \in X$.

Denote $\varphi^i \equiv f^i_1, h^i_s \equiv f^i_s \circ (\varphi^i)^{-1}$ and $V = \{v^1, \cdots, v^N\} \in \mathbb{R}^N : \exists x \in X, v^i = \varphi^i(u^i(x)), \forall i \in I\}$. Because the functions $\varphi^i$ are continuous and increasing, and by Assumption 1 (1’), the set $V$ is convex and has a non-empty interior.

The following functional equation holds on the set $V$:

$$\sum_{i \in I} h^i_s(v^i) = G_s\left(\sum_{i \in I} v^i\right)$$

---

5 When Assumption 1’ holds.
This is a Pexider equation whose solution is (see Radò and Baker, 1987, Theorem 1 and Corollary 3):

\[ G_s(v) = a_s v + \sum_{i \in I} b_i \]

\[ h_i(v) = a_s v + b_i \]

for some scalars \( a_s > 0 \) and \( b_i \in \mathbb{R} \).

The equation \( h_i(v) = a_s v + b_i \) implies that \( f_i(u^i(x)) = a_s \varphi^i(u^i(x)) + b_i \) for any \( x \in X \). Denote \( \chi = 1 + \sum_{s \neq 1} a_s, \ p_1 = \frac{1}{\chi}, \ p_s = \frac{a_s}{\chi} \) for \( s \neq 1 \), \( \beta^i = \sum_{s \in S} b_s^i \), and \( \widetilde{\varphi}^i \equiv \chi \varphi^i + \beta^i \). By Theorem 1 (1') and the above results, the function \( U^i \) represents decision maker \( i \)'s decision where

\[ U^i(x) = \sum_{s \in S} \varphi^i u^i_s(x_s) = \sum_{s \in S} f_i^s(u^i(x_s)) = \chi \sum_{s \in S} p_s \varphi^i(u^i_s) + \beta^i = \sum_{s \in S} p_s \widetilde{\varphi}^i(u^i_s) \]

The representation holds for any \( x \in X \) and any \( i \in I \).

Remark that, by the solution of the Pexider equation, the probability weights \( p_s \) are the same for all individuals. By construction, they are such that \( \sum_{s \in S} p_s = 1 \).

Finally, by Theorem 1 (1'), we know that \( U^0(x) = U^1(x) + \cdots + U^N(x) \), whatever \( x \in X \). This yields \( U^0(x) = \sum_{s \in S} p_s u^0(x_s) \), with \( u^0 \equiv \sum_{i \in I} \widetilde{\varphi}(u^i) \).

Theorem 2 brings two additional insights with respect to Theorems 1 and (1'): for the Paretian aggregation of preferences to be possible, all decision makers 1/ must have well-defined beliefs and share the same beliefs; 2/ must be uncertainty neutral.

The impossibility of aggregating preferences when individuals’ beliefs differ and the social planner also has her own beliefs is a typical result known as the ‘probability agreement theorem’ (Broome, 1990). It arises whenever individuals have diverse preferences over consequences, which is the case here since the utility possibility sets have a non-empty interior.\(^6\) The specificity of Theorem 3 is that, contrary to most previous works, the expected utility model is obtained here as a result rather than an assumption. Individuals must have well-defined and identical beliefs for the social aggregation to be possible.

Most of the recent literature on choices under uncertainty has insisted that decision makers may not be uncertainty neutral - contrary to the expected utility model.

---

\(^6\)See Mongin (1995) on the duality between beliefs or identical preferences in the context of subjective uncertainty.
The literature has therefore proposed several (non-expected utility) models of choice exhibiting uncertainty aversion. Theorem 2 shows that it is impossible to find a satisfactory aggregation of the preferences of ambiguity averse individuals. More precisely, the subjective expected utility form of preferences implies that individuals must be uncertainty neutral for the aggregation to be possible. This result concurs with the one by Gajdos, Tallon and Vergnaud (2008). It is strengthened here: under consequentialist principles, if we assume state independence, it is impossible to abandon the subjective expected utility model.

4 The Anscombe-Aumann framework

Assumption 1 and (1') are based on utility representations of preferences rather than bearing directly on the preferences which are the primitives of the problem. In this section, we focus on the Anscombe-Aumann framework and we make widely admitted assumptions on preferences. The result of Theorem 1 is strengthened: decision makers must be subjective expected utility maximizers and the social aggregation affine. The approach, which derive ethical consequences from choice-theoretic axioms and the Pareto principle is conform to Harsanyi’s research program.

4.1 A state contingent condition

Let us first introduce the Anscombe-Aumann framework. The only difference with Section 3 is that the sets $X_s$ have a particular structure. Namely each $X_s$ is a set of simple lotteries over a set of consequences $Y_s$. This is commonly written $X_s = \Delta(Y_s)$. Therefore $X_s$ is a convex set which can be endowed with the mixture operation for lotteries, $\tilde{+}$: for any lotteries $x_s$ and $\hat{x}_s$ in $X_s$ and any scalar $\alpha \in (0, 1)$, $\alpha x_s \tilde{+} (1 - \alpha) \hat{x}_s$ denotes the lottery consisting in playing lottery $x_s$ with probability $\alpha$ and lottery $\hat{x}_s$ with probability $1 - \alpha$.

In this framework, the set $X$ can also be endowed with a mixture operation $\oplus$ defined as follows: for any $x$ and $\hat{x}$ in $X$ and any scalar $\alpha \in (0, 1)$, $x' = \alpha x \oplus (1 - \alpha) \hat{x}$ is such that, for any $s \in S$, $x_s = \alpha x_s \tilde{+} (1 - \alpha) \hat{x}_s$. Note that by definition $x' \in X$.

Assume that decision makers have regular consequentialist preferences so that state contingent preferences $\succeq_s$ are well-defined. We can introduce the property of Statewise Independence.

Statewise Independence (SI). For any $s \in S$, whatever $x_s$, $\hat{x}_s$ and $\tilde{x}_s$ in $X_s$ and
for any $\alpha \in (0,1)$:

$$x_s \succeq_s \hat{x}_s \iff \alpha x_s \oplus (1-\alpha)\tilde{x}_s \succeq_s \hat{x}_s \oplus (1-\alpha)\tilde{x}_s$$

To the best of our knowledge, Axiom (SI) has not been introduced before in the literature. However, it corresponds to the generalization of the Weak Certainty Independence (WCI) Axiom to a state dependent framework. As noted in the next section, the (WCI) Axiom is widely admitted in the literature. Indeed, (SI) like (WCI) basically states the preferences over roulette lotteries (lotteries with objective probabilities) are von Neumann and Morgenstern (vNM). There are well-placed rationality arguments in favor of the vNM model in the case of objective risk. Axiom (SI) requires preferences to satisfy these rationality requirements in each state of the world.

In order to obtain our result for state dependent regular consequentialist preferences satisfying (SI), we make use of a simple condition.

**Assumption 2** For any $i \in I$ and any $s \in S$, there exist $x_s$ and $\hat{x}_s$ in $X_s$ such that $x_s \succ_s^i \hat{x}_s$ and $x_s \sim_s^j \hat{x}_s$ for all $j \in I \setminus \{i\}$.

Assumption 2 is known in the literature on Harsanyi’s theorem as ‘Independent Prospect’ (Weymark 1991). It is a preference diversity condition: there exists a minimal disagreement between individuals concerning the evaluation of social alternatives. Under the Pareto principle, it also guarantees that each individual has the opportunity to express his own preferences between some alternatives.

We have the following classical results for preferences satisfying vNM axioms.

**Lemma 4** Assume that the social observer and all individuals have regular consequentialist preferences satisfying (SI). Assume furthermore that Assumption 2 is verified. If axiom (SP) holds then there exist non-constant state contingent utility functions $u^i_s$, $i \in I^*$, that represent decision makers’ preferences and such that:

1. $u^i_s(\alpha x_s + (1-\alpha)\hat{x}_s) = \alpha u^i_s(x_s) + (1-\alpha)u^i_s(\hat{x}_s)$ for all $i \in I^*$, $x_s, \hat{x}_s \in X_s$ and $\alpha \in (0,1)$.

2. For any $s \in S$, there exists a unique vector $(a^1_s, \cdots, a^N_s) \in \mathbb{R}^N_{++}$ and a unique scalar $b_s$ such that $u^0_s(x_s) = a^1_s u^1_s(x_s) + \cdots + a^N_s u^N_s(x_s) + b_s$, whatever $x_s \in X_s$.

3. The set $U_s = \left\{(u^1_s, \cdots, u^N_s) \in \mathbb{R}^N : \exists x_s \in X_s, u^i_s = u^i_s(x_s) \forall i \in I \right\}$ is convex and has a non-empty relative interior.
Proof. State contingent preferences are non-constant continuous order satisfying the independence condition (SI). By vNM theorem (more precisely its formulation in Fishburn, 1970, Theorem 8.4, pp. 112-113) we know that state contingent preferences can be represented by a continuous affine real-valued function. This is implication 1 in Lemma 4.

Implication 2 in Lemma 4 is Harsanyi’s theorem. The theorem holds because state contingent utility functions are vNM. The fact that social weights \( a_s^i \) are unique and strictly positive is due to the combination of Independent Prospects and Strong Pareto (Weymark 1991).

The convexity of the range of vNM utility functions is well-known in the literature on Harsanyi’s theorem (De Meyer and Mongin, 1995; Blackorby, Donaldson, Weymark, 1999). The non-emptiness of the interior is due to Assumption 2. □

Lemma 4 guarantees that adequate conditions are satisfied to extend Theorem 1. The lemma also provides a collection of state contingent preferences that will be used in the representation.

Theorem 3 Assume that the social observer and all individuals have regular consequentialist preferences satisfying (SI). Assume furthermore that Assumption 2 is verified. Axiom (SP) holds if and only if there exist a collection of non-constant affine state contingent utility functions \( (u_s^i)_{i \in I^*, s \in S} \) that represent decision makers’ preferences, scalars \( p_s^i \in (0,1) \), with \( \sum_{s \in S} p_s^i = 1 \) for all \( i \in I \), positive scalars \( (\alpha^i)_{i \in I} \) and scalars \( (\beta^i)_{i \in I^*} \) such that, whatever \( x \in X \):

1. \( U^0(x) = \alpha^1U^1(x) + \cdots + \alpha^NU^N(x) \) and \( u_s^0(x_s) = \sum_{i \in I} a_i^s u_s^i(x_s) \), with \( a_i^s = \alpha^i p_s^i \).

2. (a) For each \( i \in I \), \( U^i(x) = \sum_{s \in S} p_s^i u_s^i(x_s) + \beta^i \), represents \( i \)'s preferences.

(b) \( U^0(x) = \sum_{s \in S} u_s^0(x_s) + \beta^0 \), represents social preferences.

Proof. Consider the collection of state contingent preferences \( (u_s^i)_{i \in I^*, s \in S} \). With the same line of reasoning as in the proof of Theorem 1, we obtain that the functional equation (7) holds for any \( u \in \prod_{s \in S} U_s \). By point 3 of Lemma 4, \( U_s^0 \) is convex and non-empty. Hence the set \( \prod_{s \in S} U_s^0 \) is also open, convex and non-empty. Consider any \( u \in \prod_{s \in S} U_s^0 \). There necessarily exists a rectangular subset of \( \prod_{s \in S} U_s^0 \) in the neighborhood of \( u \) such that the functional equation (7) holds. On this rectangular set, Theorem 1 applies.
The local additively separable representation we have just obtained can be extended globally. Indeed, the three conditions of Chateauneuf and Wakker (1993) presented in the Appendix are satisfied: the utility possibility set is convex and hence satisfied their conditions 1 and 2; the social indifference curves are linear (by point 2 of Lemma 4) and hence connected on the utility possibility set. The global additive representation can be extended to boundaries by continuity.

We hence obtain the same representations as in Theorem 1 using the utility profile \((u^i_s)_{i \in I^r, s \in S}\). But, by point 2 of Lemma 4, the ex post aggregation must be affine. So it is necessary that functions \(\varphi^i_s\) in Theorem 1 are affine. There exists positive numbers \((a^i_s)_{i \in I^r, s \in S}\) and scalars \((b^i_s)_{i \in I^r, s \in S}\) such that \(\varphi^i_s(u^i_s(x_s)) = a^i_s u^i_s(x_s) + b^i_s\). Denote \(\alpha^i = \sum_{s \in S} a^i_s\), \(p^i_s = \frac{a^i_s}{\alpha^i}\) and \(\beta^i = \sum_{s \in S} b^i_s\), we obtain \(U^i(x) = \alpha^i \left(\sum_{s \in S} p^i_s u^i_s(x_s) + \beta^i\right)\).

By Theorem 1 we also know that \(U^0(x) = U^1(x) + \cdots + U^N(x)\), whatever \(x \in X\). In Theorem 3, we use the alternative representation of individual preferences \(\tilde{U}^i = U^i/\alpha^i\). We also have \(U^0(x) = \sum_{s \in S} u^0_s(x_s) + b^0\), with \(u^0 \equiv \sum_{i \in I} a^i_s u^i_s\) and \(b^0 = \sum_{i \in I} b^i\).

Theorem 3 brings two additional insights compared to Theorem 1. First, the state contingent utilities used must be affine. In other words, they must be NM utility functions. Second, we explicitly obtain that individuals’ preferences must be represented by state dependent subjective expected utilities for the social aggregation to be possible.

The aggregation of preferences described in Theorem 3 is actually the one of Proposition 6 in Mongin (1998). But it relies on much weaker grounds, because Mongin (1998) selected vNM representations at the start. Theorem 3 shows that we cannot go further than Mongin’s result if we stick to consequentialism.

A clear advantage of Theorem 3 over Theorem 1 is that it is obtained from choice-theoretic premisses only, namely (SI) and Assumption 2. It therefore fits better into Harsanyi’s research program. Axiom (SI) is a very natural requirement that many scholar endorse. Assumption 2 is more transparent than Assumptions 1 and 1’. Theorem 3 constitutes a clear proof that Harsanyi’s result cannot be extended outside the expected utility model.

4.2 The state independent case

Most of the literature on decision theory in the Anscombe-Auman framework consider state independent preferences. It is therefore natural to study the implications of our results in this context. The only difference with Section 4.1 is that the set \(X = \Delta(Y)\).
is the same in each state of the world.

In this context, the Axiom (SI) corresponds to the property of Weak Certainty Independence.

**Weak Certainty Independence (WCI).** For any $x$, $\hat{x}$ and $\tilde{x}$ in $X$ and for any $\alpha \in (0,1)$:

\[ x \succeq \hat{x} \iff \alpha x \oplus (1 - \alpha) \hat{x} \succeq \alpha \hat{x} \oplus (1 - \alpha) \tilde{x} \]

Axiom (WCI) is the weakest of the independence axioms that have been proposed in the literature, viz. independence, comonotonic independence and certainty independence.\(^7\) It is satisfied by the class of variational preferences which nests subjective expected utility preferences, Choquet expected utility preferences, multiple-prior preferences and multiplier preferences. (WCI) corresponds to the axiom of ‘Independence over Risk’ used by Neilson (2009) to characterize the second-order expected utility model that was proposed by Klibanoff, Marinacci and Mukerji (2005) and Ergin and Gul (2009).

We obtain the following corollary of Theorem 3.

**Corollary 1** Assume that the social observer and all individuals have state independent regular consequentialist preferences satisfying (WCI). Assume furthermore that Assumption 2 is verified. Axiom (SP) holds if and only if there exist a collection of non-constant affine utility functions $(u^i)_{i \in I}$ that represent decision makers’ preferences, scalars $p_s \in (0,1)$, with $\sum_{s \in S} p_s = 1$, positive scalars $(\alpha^i)_{i \in I}$ and scalars $(\beta^i)_{i \in I}$, such that:

1. $U^0(x) = \alpha^1U^1(x) + \cdots + \alpha^N U^N(x)$ and $u^0(x) = \sum_{i \in I} \alpha^i u^i(x)$.

2. (a) For all $i \in I$, $U^i(x) = \sum_{s \in S} p_s u^i(x_s) + \beta^i$, represents $i$’s preferences whatever $x \in X$.

   (b) $U^0(x) = \sum_{s \in S} p_s u^0(x_s) + \beta^0$, represents social preferences whatever $x \in X$.

The interest of Corollary 1 is that it applies directly to many models that have been recently proposed in the decision theoretic literature. In this respects, it appears that the only model of choice that can be properly aggregated is the subjective expected

---

\(^7\)For more details, see Maccheroni, Marinacci, Rustichini (2006). Independence corresponds to the case where $x$, $\hat{x}$ and $\tilde{x}$ can be any elements of $X$, rather than only certain alternatives. Comonotonic independence restricts $x$, $\hat{x}$ and $\tilde{x}$ to be comonotonic (certain alternatives are necessarily comonotonic). Certainty independence restricts $\tilde{x}$ only to be certain.
utility model. Corollary 1 excludes several decision models at the societal level, among others the Choquet expected utility model, multiple-prior preferences, variational preferences and the second-order expected utility model. As a consequence, the notion of a representative agent having such preferences may not seem appealing.

5 Concluding Remarks

We have explored the possibility of aggregating preferences under uncertainty in a general model of choice. We have established that, under rather weak consequentialist premises, an additively separable utility form close to the subjective expected utility is actually necessary to obtain a possibility result. If we restrict attention to state independent preferences we actually find that only subjective expected utilities can be aggregated under the strong Pareto principle. In both cases, the social aggregation should be affine.

Two natural responses to the negative conclusions contained in our theorems present themselves. One is to argue against the Pareto principle in an environment of subjective uncertainty. The other is to argue against the proposed decision model, in particular at the societal level.

The first direction, which entails calling into question the Pareto requirement, has been suggested by several authors in the context of choice uncertainty (for instance Gilboa, Samet and Schmeidler, 2004; or Fleurbaey, 2007). One possible (though radical) route could be for instance to apply the Pareto principle ex post. When all individuals and the social observer have regular consequentialist preferences, it is clear from Lemma 3 that ex post strong Pareto implies that the social welfare function must have the form:

$$U^0(x) = F^0\left(w_1(u_1^1(x_1), \ldots, u_1^N(x_1)), \ldots, w_S(u_S^1(x_S), \ldots, u_S^N(x_S))\right)$$

Doing so, we can be able to endorse any decision model at the societal level, leaving for instance some room for uncertainty aversion (which the additively separable form excludes). But we would also have to abandon welfarism.

The second direction consists in abandoning the monotonicity axiom. Indeed, like in Gajdos, Tallon and Vergnaud (2008), the monotonicity condition is at the heart of the impossibility results. Dropping it at the society’s level is likely to restore the possibility of an aggregation for general individuals’ preferences. This direction however
entails abandoning consequentialism. As we have said in the introduction, there are
several difficulties with non-consequentialism, among which the impossibility to have
an independent definition of contingent social preferences. We hence cannot use the
kind of dominance property suggested by Fleurbaey (2007) in his ‘omniscient principle’.
In addition, non-consequentialism requires to remember in each state of the world all
the unrealized branches of the decision tree in order to make consistent judgments.
Practically, this can render the process of choice extremely complicated.

In any case, the two directions should be explored further. Both the non-consequentialist
and the non-Paretian options are currently underdeveloped, with only few papers
seeking to derive attractive social criterion within these two classes. In view of our
results, these are attractive research avenues, if one is not willing to consider subjective
expected utility as the only relevant decision model under uncertainty.

Appendix

Chateauneuf and Wakker (1993) identify three conditions that are sufficient for the
local additive representability of an ordering \( \succeq \) on an open set \( E \subset \prod_{k=1}^{K} E_k \) to be
extended to global additive representability:

1. \( E \) must be connected

2. all sets of the form \( \{ e = (e_1, \cdots , e_K) \in E : e_k = z \} \), where \( k \in \{1, \cdots , K\} \) and \( z \in E_k \), must be connected;

3. the indifference curves on \( E \) for the ordering under consideration must be con-

nected.

Let \( \mathcal{U}^o = \prod_{s \in S} \mathcal{U}^o_s \). And denote \( G : \mathcal{U}^o \rightarrow \mathbb{R} \) the function such that, for any \( \mathbf{u} = (u_1, \cdots , u_1^N, \cdots , u_S^1, \cdots , u_S^N) \in \mathcal{U}^o \), \( G(\mathbf{u}) = F^0 \left( w_1(u_1^1, \cdots , u_1^N), \cdots , w_S(u_S^1, \cdots , u_S^N) \right) \).
We want to show that ordering \( \succeq \) defined on \( \mathcal{U}^o \) by \( \mathbf{u} \succeq \hat{\mathbf{u}} \iff G(\mathbf{u}) \geq G(\hat{\mathbf{u}}) \) satisfies the
three condition by Chateauneuf and Wakker (1993) under Assumption 1.

Condition 1 in Assumption 1 implies that each \( \mathcal{U}^o_s \) is an open, convex and non-
empty subset of the Euclidian \( N \) space, \( \mathbb{R}^N \). As a consequence, the set \( \mathcal{U}^o = \prod_{s \in S} \mathcal{U}^o_s \) is also an open, convex and non empty subset of the Euclidian \( N \) space. The first and
second conditions by Chateauneuf and Wakker are therefore satisfied.

To show that the indifference curves are connected, we proceed in two steps.
Step 1: We first show that the social indifference curves in the space $\mathcal{U}_s^o$ are connected.
These social indifference are defined by the function $w_s : \mathcal{U}_s^o \rightarrow \mathbb{R}$.

To show that the indifference curves are connected, we will show that they are path-connected. Recall that a set $Z$ is path connected if for any $z$ and $\hat{z}$ in $Z$ there exists a continuous function $\Psi^{z,\hat{z}} : [0,1] \rightarrow Z$ such that $\Psi^{z,\hat{z}}(0) = z$ and $\Psi^{z,\hat{z}}(1) = \hat{z}$.

The elements of $\mathcal{U}_s^o$, denoted $u_s = (u_s^1, \ldots, u_s^N)$, are vectors in $\mathbb{R}^N$. For any $w \in Im_{w_s} = \{ w \in \mathbb{R} : \exists u_s \in \mathcal{U}_s^o, w_s(u_s^1, \ldots, u_s^N) = w \}$, denote $Ind_{w_s}(w) = \{ u_s \in \mathcal{U}_s^o : w_s(u_s^1, \ldots, u_s^N) = w \}$ the indifference curve of level $w$.

For any $u_s$ and $\hat{u}_s$ in $Ind_{w_s}(w)$, we define the function $\Psi^{u_s,\hat{u}_s} : [0,1] \rightarrow Ind_{w_s}(w)$ in the following way. For any $\alpha \in [0,1]$, $u_s^\alpha$ is the point in $\mathbb{R}^N$ such that $u_s^\alpha = \alpha u_s + (1 - \alpha) \hat{u}_s$. We also denote $\underline{u}_s$ the point in $\mathbb{R}^N$ such that $u_s = (u_s^\underline{\alpha}(\underline{x}_s), \ldots, u_s^\underline{1}(\underline{x}_s))$, where $\underline{x}_s$ is defined like in Assumption 1. Then, for any $\alpha \in [0,1]$, we define $\Psi^{u_s,\hat{u}_s}(\alpha)$ as the point in the segment $[u_s, u_s^\alpha]$ such that $w_s(\Psi^{u_s,\hat{u}_s}(\alpha)) = w$. Remark that $\Psi^{u_s,\hat{u}_s}(0) = u_s$ and $\Psi^{u_s,\hat{u}_s}(1) = \hat{u}_s$.

We need to prove that the function $\Psi^{u_s,\hat{u}_s}$ is well-defined and that it is continuous. We first show that $\Psi^{u_s,\hat{u}_s}$ is well-defined. By convexity of $\mathcal{U}_s^o$, $u_s^\alpha$ belongs to $\mathcal{U}_s^o$. Convexity also implies that the whole segment $[u_s, u_s^\alpha]$ is a subset of $\mathcal{U}_s^o$, since $u_s \in \mathcal{U}_s^o$. The function $w_s$ is hence a continuous on the compact set $[u_s, u_s^\alpha]$. In addition, $w_s(u_s) < w \leq w_s(u_s^\alpha)$, by definition of $u_s$ and concavity of the function $w_s$. Thus there exists $\Psi^{u_s,\hat{u}_s}(\alpha) \in [u_s, u_s^\alpha]$ such that $w_s(\Psi^{u_s,\hat{u}_s}(\alpha)) = w$. It is unique, due to the monotonicity of $w_s$. Hence, for any $\alpha \in [0,1]$, $\Psi^{u_s,\hat{u}_s}(\alpha)$ exists and is unique: the function is well-defined.

We now need to show that $\Psi^{u_s,\hat{u}_s}$ is continuous with respect to the usual Euclidian distance, $d$. This means that, for any $\alpha \in [0,1]$ and whatever $\varepsilon > 0$, there exists $\delta > 0$ such that, $|\alpha - \alpha| < \delta$ implies $d(\Psi^{u_s,\hat{u}_s}(\alpha), \Psi^{u_s,\hat{u}_s}(\bar{\alpha})) < \varepsilon$.

Consider $\alpha \neq \bar{\alpha}$. Let $P(\Psi^{u_s,\hat{u}_s}(\bar{\alpha}))$ be the orthogonal projection of $\Psi^{u_s,\hat{u}_s}(\bar{\alpha})$ on the segment $[u_s, u_s^\alpha]$. Consider the triangle formed by the three points $\Psi^{u_s,\hat{u}_s}(\alpha)$, $\Psi^{u_s,\hat{u}_s}(\bar{\alpha})$ and $P(\Psi^{u_s,\hat{u}_s}(\bar{\alpha}))$ and let $\theta$ be the angle corresponding to vertex $\Psi^{u_s,\hat{u}_s}(\alpha)$. A standard trigonometric argument yields
\[
    d(\Psi^{u_s,\hat{u}_s}(\alpha), \Psi^{u_s,\hat{u}_s}(\bar{\alpha})) = \frac{d(P(\Psi^{u_s,\hat{u}_s}(\alpha), \Psi^{u_s,\hat{u}_s}(\bar{\alpha}))}{\sin \theta}
\]

It can be showed (see Figure 1) that $d(P(\Psi^{u_s,\hat{u}_s}(\alpha), \Psi^{u_s,\hat{u}_s}(\bar{\alpha})) \leq d(u_s^\alpha, u_s^\bar{\alpha}) = |\bar{\alpha} - \alpha| \times d(u_s, \hat{u}_s)$, and that there exists $\theta > 0$ such that $\theta < \theta < \frac{\pi}{2}$. Hence,
For any $u_s^\circ$, $\Psi^u_s,\hat{u}_s$.

Figure 1: Definition and continuity of function $\Psi^u_s,\hat{u}_s$.

Note: $P(u_s^\circ)$ is the orthogonal projection of $u_s^\circ$ on the line $(u_s, u_s^\circ)$. Thus $d(u_s^\circ, P(u_s^\circ)) \leq d(u_s^\circ, u_s^\circ)$. The geometry of the triangle also implies that $d(P(\Psi^u_s,\hat{u}_s(\alpha)), \Psi^u_s,\hat{u}_s(\alpha)) \leq d(u_s^\circ, P(u_s^\circ))$, which yields $d(P(\Psi^u_s,\hat{u}_s(\alpha)), \Psi^u_s,\hat{u}_s(\alpha)) \leq d(u_s^\circ, u_s^\circ)$. It must also be the case that the angle $\theta$ is greater than $\theta = \min(\beta, \gamma) > 0$.

$d(\Psi^u_s,\hat{u}_s(\alpha), \Psi^u_s,\hat{u}_s(\alpha)) < |\hat{\alpha} - \alpha| \times \frac{d(u_s,\hat{u}_s)}{\sin \theta}$. For any $\varepsilon > 0$, defining $\delta = \frac{\sin \theta \times \varepsilon}{d(u_s,\hat{u}_s)}$, we obtain that $|\hat{\alpha} - \alpha| < \delta$ implies $d(\Psi^u_s,\hat{u}_s(\alpha), \Psi^u_s,\hat{u}_s(\alpha)) < \varepsilon$. The function $\Psi^u_s,\hat{u}_s$ is continuous.

The construction of function $\Psi^u_s,\hat{u}_s$ can be done for any $u_s$ and $\hat{u}_s$ in $Ind_{w_s}(w)$. Hence $Ind_{w_s}(w)$ is path-connected. The reasoning is true for any $w \in Im_{w_s}$.

**Step 2:**

Now we go back to the ordering $\succeq$ defined on $U^o$. To show that the indifference curves are connected, we will show like in Step 1 that they are path-connected.

For any $y \in Im_G = \{ y \in \mathbb{R} : \exists u \in U^o, G(u) = y \}$, we denote $Ind(y) = \{ u \in U^o : G(u) = y \}$ the indifference curve of level $y$. For any $u$ and $\hat{u}$ in $Ind(y)$, we need to define a continuous function $\Lambda : [0, 1] \rightarrow Ind(y)$ such that $\Lambda(0) = u$ and $\Lambda(1) = \hat{u}$.

To do so, we first distinguish three subsets of $S$:

- $S^+ = \{ s \in S : w_s(u_s^1, \ldots, u_s^N) > w_s(\hat{u}_s^1, \ldots, \hat{u}_s^N) \}$
- $S^- = \{ s \in S : w_s(u_s^1, \ldots, u_s^N) < w_s(\hat{u}_s^1, \ldots, \hat{u}_s^N) \}$
\[ S^- = \left\{ s \in S : w_s(u^1_s, \ldots, u^n_s) = w_s(\hat{u}^1_s, \ldots, \hat{u}^N_s) \right\} \]

Clearly, \((S^+, S^-, S^-)\) constitutes a complete partition of \(S\). Note also that, due to the monotonicity assumption, whenever \(S^+\) is non-empty, \(S^-\) must also be non-empty, and reciprocally.

Using the notation of Step 1, we introduce the following vectors in \(U^0_s\). For \(s \in S^+\), \(\bar{u}_s = (\bar{u}^1_s, \ldots, \bar{u}^N_s)\) is the intersection between the line \((\underline{u}_s, u_s)\) and the social indifference curve of level \(\bar{w}_s = w_s(\hat{u}^1_s, \ldots, \hat{u}^N_s)\). Similarly, for \(s \in S^-\), \(\bar{u}_s\) is the intersection between the line \((\underline{u}_s, \bar{u}_s)\) and the social indifference curve of level \(w_s = w_s(u^1_s, \ldots, u^N_s)\). By an argument similar to the one used for \(\Psi_{u_s, \hat{u}_s}(\alpha)\) in Step 1, we have that \(\bar{u}_s\) is well defined and unique. Further more, it is such that \(\bar{u}_s < u_s\) for \(s \in S^+\) and \(\bar{u}_s < \hat{u}_s\) for \(s \in S^-\).

Now, for any \(\alpha\) and \(\beta\) in \([0, 1]\), we define \(u(\alpha, \beta)\) in the following way:

- For all \(s \in S^+\) and for all \(i \in I\), \(u(\alpha, \beta)_s^i = (1 - \alpha)u^i_s + \alpha \bar{u}^i_s\).
- For all \(s \in S^-\) and for all \(i \in I\), \(u(\alpha, \beta)_s^i = (1 - \beta)\bar{u}^i_s + \beta \hat{u}^i_s\).
- For all \(s \in S^-\) and for all \(i \in I\), \(u(\alpha, \beta)_s^i = \hat{u}^i_s\).

Because each \(U^0_s\) is convex, \(u(\alpha, \beta) \in U^0\). In addition, from the definition of the function \(G\), we have that, for any \(\alpha \in (0, 1)\), \(G(u(\alpha, 0)) < G(u) < G(u(\alpha, 1))\), provided that \(S^+\) and \(S^-\) are non-empty. Hence, by continuity and monotonicity, since \(\bar{u}_s < \hat{u}_s\) for \(s \in S^-\), there exists a unique \(\tilde{\beta} \in (0, 1)\) such that \(G(u^0_{\tilde{\beta}}) = G(u)\). We denote this \(\tilde{\beta}\) by \(\kappa(\alpha)\). Setting \(\kappa(0) = 0\) and \(\kappa(1) = 1\), it is clear that the function \(\kappa\) is continuous on the interval \([0, 1]\). We denote \(u(\alpha) \equiv u(\alpha, \kappa(\alpha))\). Clearly, \(u(\alpha) \in Ind(y)\) and it changes continuously with \(\alpha\).

It remains to define the function \(\Lambda\) in the following way:\(^8\)

- For \(\alpha \in [0, \frac{1}{3}]\), \(\Lambda(\alpha) = v\) with \(v^s = \Psi_s(3\alpha)\) for all \(s \in S^- \cup S^0\) and \(v^s = u^s\) for all \(s \in S^+\).
- For all \(\alpha \in \left[\frac{1}{3}, \frac{2}{3}\right]\), \(\Lambda(\alpha) = u(3\alpha - \frac{1}{3})\).
- For \(\alpha \in (\frac{2}{3}, 1]\), \(\Lambda(\alpha) = v\) with \(v^s = \Psi_s\left(3\left(\alpha - \frac{2}{3}\right)\right)\) for all \(s \in S^+\) and \(v^s = \hat{u}^s\) for all \(s \in S^- \cup S^0\).

\(^8\)The functions \(\Psi_s\) are defined like in Step 1 for the vectors: 1/ \(\hat{u}_s\) and \(\hat{u}_s\) for \(s \in S^+\); 2/ \(u_s\) and \(\bar{u}_s\) for \(s \in S^-\); 3/ \(u_s\) and \(\hat{u}_s\) for \(s \in S^0\).
We hence have that $\Lambda(\alpha) \in Ind(y)$, $\Lambda(0) = u$ and $\Lambda(1) = \hat{u}$. Furthermore, given that the functions $\kappa$ and $\Psi_s$ are continuous, the function $\Lambda$ is also continuous.

The reasoning applies to any $u$ and $\hat{u}$ in $Ind(y)$, whatever $y \in Im_G$.

Q.E.D.

References


