Centre d'Economie de la Sorbonne
UMR 8174

Université Paris 1 – Panthéon - Sorbonne –

Cahiers de la MSE

The Ignorant Observer

Thibault GAJDOS

Feriel KANDIL

2006.41

Centre National de la Recherche Scientifique

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://mse.univ-paris1.fr/Publicat.htm
ISSN : 1624-0340

The Ignorant Observer*

Thibault Gajdos† Feriel Kandil‡

First draft: February 2005.
This draft: March 2006.

Abstract

Most prominent models of economic justice (and especially those proposed by Harsanyi and Rawls) are based on the assumption that impartiality is required for making moral decisions. However, although Harsanyi and Rawls agree on that, and furthermore agree on the fact that impartiality can be obtained under appropriate conditions of ignorance, they strongly disagree on the consequences of these assumptions. According to Harsanyi, they provide a justification for the utilitarian doctrine, whereas Rawls considers that they imply egalitarianism. We propose here an extension of Harsanyi's Impartial Observer Theorem, that is based on the representation of ignorance as the set of all possible probability distributions. We obtain a characterization of the observer's preferences that, under our most restrictive conditions, is a linear combination of Harsanyi's and Rawls' criteria. Furthermore, this representation is ethically meaningful, in the sense that individuals' utilities are cardinally measurable and unit comparable. This allows us to conclude that the impartiality requirement cannot be used to decide between Rawls' and Harsanyi's positions. Finally, we defend the view that a (strict) combination of Harsanyi's and Rawls' criteria provides a reasonable rule for social decisions.

Keywords: Impartiality, Justice, Decision under ignorance.
JEL Numbers: D63, D81.

1 Introduction

According to a long tradition among moral philosophers, moral judgements have to be made from the point of view of a rational, impartial and sympathetic observer. This idea is at the core of two prominent economic models of justice, namely Harsanyi's (1953; 1977) "utilitarianism" and Rawls' (1971) egalitarianism.

The fundamental insight put forward by Harsanyi and Rawls (and, independently, by Vickrey (1945)[1]), is that *impartiality* can be ensured if the observer is placed under appropriated

*We thank D. Bouyssou, A. Chateauneuf, M. Cohen, M. Fleurbaey, E. Karni, P. Mongin, J. Moreno-Ternero and John Weymark, as well as seminar audience at University Pompeu Fabra and University of Cergy-Pontoise for useful comments. Financial support from an ACI grant by the Ministry of Research is gratefully acknowledged.

†CNRS–CES, `gajdos@univ-paris1.fr`

‡CERC, `f_kandil@hotmail.com`

[1]See Mongin (2001) for a thorough comparison of Vickrey's and Harsanyi's approaches.

conditions of *ignorance* (the "veil of ignorance", in Rawls (1971)' terms). In particular, the observer should "*not know in advance* what his own social position would be in each social situation" (Harsanyi, (1977), p. 49).

A strong link is hence established between the theory of morality and the theory of decision making under ignorance. However, although Rawls and Harsanyi agree on the idea that fair rules are those chosen by rational individuals from behind the veil of ignorance, they strongly disagree on what these rules should actually be. Starting from similar assumptions, they end with opposite conclusions: according to Rawls, the impartiality requirement leads to the egalitarian (or maxmin) criterion, whereas according to Harsanyi the same requirement leads to utilitarianism.

Our aim is to figure out if the impartiality requirement, viewed as ignorance, implies Harsanyi's or Rawls' conclusion. There are two main difficulties here. First, whereas Harsanyi (1977) proposed a formal model of decision from behind the veil of ignorance (his celebrated "Impartial Observer Model"), Rawls only proposed informal arguments. We therefore need to build a model that can accommodate Rawls' views on impartial decisions. The second difficulty is related to a well-known weakness of Harsanyi's model: as shown by Sen (1976) and Weymark (1991), the weights attributed to individuals' utilities in Harsanyi's Impartial Observer Theorem are not meaningful, if these utilities are not cardinally comparable. We extend Harsanyi's model so that (*i*) Rawls' argument can be formalized and (*ii*) the conclusions we obtain are meaningful, i.e., make use of cardinally measurable and unit comparable individual utility functions.

To make our approach clear, let us briefly present Harsanyi's (1953; 1977) Impartial Observer Model. Let N be the set of individuals, and X be the set of social alternatives (both finite). Each individual has a preference relation $\succeq_i$ on the set Y of lotteries over X (social-alternative lotteries). Furthermore, these preferences are assumed to obey axioms of expected utility theory. The observer is assumed to be a rational individual, and be able to make judgements such as: "social-alternative lottery y is better for individual i than social-alternative lottery z for individual j". To formalize this idea, Harsanyi (1977) assumes that the observer has preferences on the set $\Delta(X \times N)$ of probability distributions over $X \times N$. Elements of $\Delta(X \times N)$ are called extended lotteries. The observer's preferences on extended lotteries are assumed to satisfy axioms of expected utility theory. Harsanyi then adds two axioms. The first one, known as the acceptance principle, states that whenever the observer has to rank two extended lotteries in which he is the same individual for sure, he does it the same way as that individual ranks the corresponding social-alternative lottery. This axiom is intended to capture the observer's sympathy towards individuals. The second axiom, which is intended to capture the observer's impartiality, states that the observer ranks two social-alternative lotteries as he would rank the extended lotteries in which there is an equal chance of being any individual, and all individuals face the same social-alternative lottery. In other words, Harsanyi represents ignorance by equiprobability. As a result, he obtains that the observer's preferences over social-alternative lotteries can be represented by the arithmetic mean of some adequately chosen individuals' von Neuman-Morgenstern utility functions.

Therefore, Harsanyi's conclusion is obtained under the assumption according to which *ignorance* can be represented by *equiprobability*. He justifies this assumption through the so-called Laplace's principle. This principle has raised numerous objections, as Mongin (2001) recalls, and therefore does not provide reasonable grounds for modeling impartiality. Furthermore, Harsanyi's theorem presents the following problem, raised by Sen (1976) and Weymark (1991) : even if one assumes that individuals' von Neumann-Morgenstern utilities have a cardinal signification (which is not the case in the standard expected utility theory), the choice of a *specific* representation of individuals' von Neumann-Morgenstern utilities implies that the weights that appear in the Harsanyi's theorem are not meaningful. The reason for that is that in Harsanyi's model, individuals' utilities are not cardinally comparable.

Karni (1998) and Mongin (2001) proposed a nice solution to escape both problems: they assume that the observer obeys the *subjective* expected utility theory. Therefore, provided one can identify the observer's subjective probabilities, the weights would be determined[2]. An important feature of these approaches is that they remain inside the Bayesian theory. But, precisely, Rawls (1971) explicitly rejected such an assumption. Therefore, if we want to take into account Rawls' arguments, we need a model that does not assume Bayesianism from the outset.

Note that Harsanyi and Rawls agree that probabilities should be taken into account, whenever they have some objective basis. This suggests that the decision maker's knowledge can be represented by a *set of probability distributions*, that describes all probability distributions that are *possible*, according to the decision maker's factual or logical knowledge. The first step in our reconstruction of Harsanyi's impartial observer model is therefore to provide an axiomatic foundation for the observer's preferences when his information takes the form of a set of probability distributions. Several axiomatizations of such preferences have been recently proposed[3].

[2]Karni (1998) obtains this result through a sophisticated construction, that can be interpreted in terms of impartiality, whereas Mongin (2001) proposes axioms of epistemic nature, related to those introduced by Karni and Schmeidler (1981).

[3]The idea of modeling information as a set of probability distributions seems to have first be proposed by Jaffray (1989), in a model where preferences were defined over belief functions. Wang (2003) and Gajdos, Talon and Vergnaud (2004b) considered information as a set of probability distributions together with an "anchor", i.e., a probability distribution that has particular salience. Finally, Hayashi (2005) proposed, in an independent work, a model essentially similar to that of Gajdos, Tallon and Vergnaud (2004a).

Among these models, Gajdos, Tallon and Vergnaud (2004a)'s one is the closest to the one we provide here. However, this model cannot be used for the observer's preferences, because it assumes state-independence (as do all models of this kind we are aware of), which would force all individuals to have the same preferences[4]. Furthermore, the axiomatization proposed by Gajdos, Tallon and Vergnaud (2004a) (which is stated in Savage's framework) requires an infinite state space, which would be difficult to justify in the present framework. Finally, this work was mainly concerned with a formulation of uncertainty aversion directly related to comparisons of sets of information, instead of the classical formulation in terms of randomization. However, in the present context, randomization has a natural interpretation from an ethical point of view, and we will therefore keep it explicitly in our model.

Viewing complete ignorance as equivalent as considering that *all* probability distributions are possible, we are then in position to reconstruct Harsanyi's impartial observer's theorem in an extended framework, that does not assume Bayesianism at the outset. We obtain our "Ignorant Observer" Theorem which, in his most precise formulation, asserts that the observer's preferences on social-alternatives lotteries can be represented by:

$$V(y) = \theta \min_{i \in N} V_i(y) + (1-\theta) \sum_{i \in N} \frac{1}{n} V_i(y)$$

where $\theta \in [0,1]$ is uniquely determined (for a given observer), and the utility functions V_i are cardinally measurable and unit comparable representations of individuals' preferences. More precisely, V_i $(i \in N)$ are chosen such that $V_i(Y) = V_j(Y)$ for all $i, j \in N$. The above result can also be written as:

$$V(y) = \theta \min_{i \in N} \frac{V_i(y) - \min_{w \in Y} V_i(w)}{\max_{w \in Y} V_i(w) - \min_{w \in Y} V_i(w)} + (1-\theta) \sum_{i \in N} \frac{1}{n} \frac{V_i(y) - \min_{w \in Y} V_i(w)}{\max_{w \in Y} V_i(w) - \min_{w \in Y} V_i(w)}$$

where again $\theta \in [0,1]$ is uniquely determined (for a given observer), but where the V_i are arbitrarily chosen von Neumann-Morgenstern utility functions representing individuals' preferences[5].

We therefore conclude that the impartiality requirement is compatible with both Harsanyi's and Rawls' views, for Harsanyi's criterion is obtained for $\theta = 0$, whereas Rawls' criterion is obtained for $\theta = 1$. Considering under which conditions this model specializes into Harsanyi's model or Rawls' one sheds some light on the debate between Rawls and Harsanyi. This leads us to defend the view that the (strict) combination of Harsanyi's and Rawls' criteria (i.e., setting $\theta \in (0,1)$) leads to a reasonable criterion for social decision.

Organization of the paper

[4] We are grateful to Edi Karni for having drawn our attention on this point.

[5] This result seems, at first sight, very similar to the one obtained by Karni (1998). There are, however, important differences between Karni's approach and ours. We elaborate on this in Sections 3 and 4.

The paper is organized as follows. First, we extend Harsanyi's framework by considering sets of probability distributions instead of lotteries on individual identities (Section 2). We then provide an axiomatic characterization of the observer's preferences in this extended framework (Section 3). In Section 4, we formalize the impartiality requirement as complete ignorance, in the sense that all lotteries on individual identities are considered as possible. We then reconstruct Harsanyi's impartial observer theorem under these hypothesis, assuming that individuals' preferences satisfy axioms of expected utility theory, and state our Ignorant Observer Theorems. Finally, we reconsider, in Section 5, the Harsanyi-Rawls debate in the light of our results, and defend the view that both solutions are unsatisfactory, whereas a mix of the two (i.e., with $\theta \in (0,1)$) is a reasonable criterion for social decision.

2 Modeling ignorance

Harsanyi and Rawls agree on the idea that decisions compatible with the principles of social justice can be thought of as *impartial* ones. Furthermore, they both consider that impartiality can be viewed as *ignorance* concerning the decision maker's position in society. This idea leads them to define fair allocation rules as the ones a rational agent, placed under a *veil of ignorance*, would choose.

However, Rawls and Harsanyi disagree on the treatment of the information available under the veil of ignorance. For Harsanyi, uncertainty can be reduced to equiprobability (using the Bayesian doctrine *and* Laplace's principle), whereas Rawls defends the idea that, under complete ignorance, the use of probabilities should be rejected. We propose here a model that can accommodate both Harsanyi's and Rawls' views. We will call *Ignorant Observers* rational agents placed under a veil of ignorance. This section is devoted to the description of our formal model of ignorance.

We consider a society made of a finite number of agents $N = \{1, \cdots, n\}$. Let X be a non-empty finite set of social alternatives (or consequences), and Y be the set of probability distributions over X (social-alternative lotteries). Following Harsanyi (1953) and (1977), individuals are assumed to have preferences on Y. These preferences are denoted $\succeq_i$ ($i \in N$). As customary we denote by $\sim_i$ and $\succ_i$ the symmetric and asymmetric components of $\succeq_i$.

An *observer* is someone able to make social judgements of the following kind: "social-alternative lottery y is better for individual i than social-alternative lottery z for individual j". In order to make such statement formally, Harsanyi (1977) assumed that the observer has preferences over the set of all *extended lotteries*, i.e., lotteries on $X \times N$. We will denote by $\mathcal{E}$ the set of such lotteries. An element of $\mathcal{E}$ is thus a function $\rho : X \times N \to [0,1]$, such that

$\sum_{x\in X}\sum_{i\in N}\rho(x,i) = 1$, where $\rho(x,i)$ is the probability of being in individual i's shoes, and getting x.

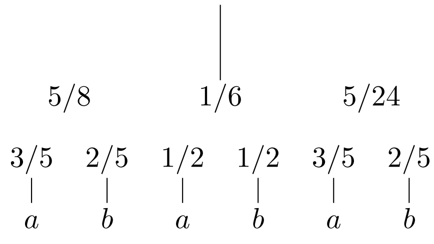
Karni and Weymark (1998) proposed the following illuminating interpretation of an extended lottery. Such a lottery can be viewed as a two stage lottery, where a first lottery on N determines which individual the observer is to be, and a second lottery on X then determines what the social state is. Formally, Karni and Weymark (1998) defined a "personal identity lottery" as a probability distribution p on N, and an "allocation" f as an assignment of a lottery on X to each individual. Let $\Delta(N)$ be the set of all probability distributions on N, and $\mathcal{A}$ be the set of allocations, i.e., the set of all functions from N to Y. Let $\mathcal{A}^c$ be the set of *constant* allocations, i.e., allocations f such that $f(i) = f(j)$ for all $i, j \in N$. As noted by Mongin and d'Aspremont (1998) and Karni and Weymark (1998), interpreting individuals as states of the nature, an allocation is an act in the Anscombe-Aumann (1963) model. The following example illustrates the correspondance between $\mathcal{E}$ and $\mathcal{A}\times\Delta(N)$. Assume that $N = \{1,2,3\}$, $X = \{a,b\}$, and consider the following extended lottery:

	1	2	3
a	3/8	1/12	1/8
b	1/4	1/12	1/12
$p(\rho(i))$	5/8	1/6	5/24

$y_i(\rho)$ is then:

	1	2	3
a	3/5	1/2	3/5
b	2/5	1/2	2/5

Finally, (f,p) can be represented as follows:



Our formulation of the observer's preferences will be based on this observation. More formally, an extended lottery ρ can be identified to a couple $(f,p) \in \mathcal{A}\times\Delta(N)$ as follows. Let $p(\rho)(i) = \sum_{x\in X}\rho(x,i)$, for all $i \in N$, and whenever $p(\rho)(i) > 0$, let $y_i(\rho)(x) = \frac{\rho(x,i)}{\sum_{z\in X}\rho(z,i)}$ for all $x \in X$, with $f(\rho)(i) = y_i(\rho)$. If $p(\rho)(i) = 0$, let $y_i(\rho)$ be an arbitrary element of Y. Finally, define $p(\rho) \otimes f(\rho)$ by $(p(\rho) \otimes f(\rho))(x, i) = p(\rho)(i) y_i(\rho)(x)$. Clearly, $\rho = p(\rho) \otimes f(\rho)$.

Let $\mathbb{P}$ be the set of a non-empty, closed sets of probability distributions on N. A generic element of $\mathbb{P}$ will be denoted by $\mathcal{P}$. Furthermore, for all $\mathcal{P} \in \mathbb{P}$, $co(\mathcal{P})$ denotes the convex hull of $\mathcal{P}$, and $\mathbb{P}_C$ denotes the set of all non-empty, closed and convex sets of probability distributions on N. Finally, δ_i is the probability distribution on N defined by $\delta_i(i) = 1$.

The observer's preferences $\succeq$ will be defined on the product $\mathcal{A} \times \mathbb{P}$ ($\sim$ and $\succ$ will, as usual, denote the symmetric and asymmetric component of $\succeq$, respectively). The couple $(f, \mathcal{P})$ can be interpreted as follows: the observer knows that the probability distribution on N according to which his identity will be chosen is in the set $\mathcal{P}$, but has no further information on the process assigning identities. Observe that an extended lottery ρ can be identified to $(f(\rho), \{p(\rho)\})$. It is important to note that a set of probability distribution $\mathcal{P} \in \mathbb{P}$ is here thought of as an *objective* data of the decision problem in hand.

This representation of information is compatible both with Harsanyi's and Rawls' views on decision under uncertainty. Indeed, Harsanyi himself insisted that the *principle of best information* is one major basis for rational decisions:

> (...) in any practical application of the Bayesian approach here is always an implicit recognition of the principle that the decision maker must choose his subjective probability in a rational manner, i.e., in the light of the best information available to him.
>
> Harsanyi, (1977), 3.6, p. 47

A point on which Rawls agrees:

> (...) I have simply assumed that judgements of probability, if they are to be grounds of rational decision, must have an objective basis, that is, a basis in knowledge of particular facts (or in reasonable beliefs).
>
> Rawls, (1971), Section 28

Harsanyi assumes that the observer is Bayesian *and* that being completely ignorant about the probability distribution that governs the individual lottery is equivalent to know for sure that the individual lottery has a uniform distribution. In other words, Harsanyi makes the following assumptions: (i) for all $f \in \mathcal{A}$, and all $\mathcal{P} \in \mathbb{P}$, there exists $p \in \mathcal{P}$ such that $(f, \mathcal{P}) \sim (f, \{p\})$, and (ii) for all $f \in \mathcal{A}$, $(f, \Delta(N)) \sim (f, \{\mu\})$, where μ is the uniform distribution on N. The first assumption is a version of Bayesianism, whereas the second one is what Harsanyi (1977) calls the"Equal Chance Principle", and is a mere reformulation of the Laplace principle. These assumptions seem highly dubious, and are actually related to the decision maker's attitude towards imprecise information, and not to his supposed rationality: they have a psychological meaning (namely: neutrality towards uncertainty). Note that, in this respect, Rawls' approach is not more convincing, since he assumes from the outset that the decision maker has an extreme aversion towards uncertainty. Our aim is therefore to propose a general decision model on $\mathcal{A} \times \mathbb{P}$ that lets more open the decision maker's attitude towards information imprecision.

3 The Observer's preferences

We now turn to the observer's preferences on $\mathcal{A} \times \mathbb{P}$. Our model is similar in spirit to the one axiomatized by Gajdos, Tallon and Vergnaud (2004a). However, this last model is state-independent, which would yield, in our framework, to uniform individuals' preferences. We therefore need to relax the sate-independence assumption. Furthermore, the above mentioned paper aimed at defining an imprecision aversion concept directly related to comparisons of sets of information. But, as we will show, the classical definition of uncertainty aversion through preference for randomization can be easily interpreted in the social choice framework. We will therefore keep such a definition of uncertainty aversion. Finally, Gajdos, Tallon and Vergnaud (2004a) axiomatization relies on operations on the state space that are difficult to interpret in this framework, where states are individuals. We will therefore avoid them. Let us note that Gilboa and Schmeidler (1989)'s maxmin model cannot be used here, for two reasons. First, because it is state-independent, it would lead to uniform individuals' preferences. Second, this model does not allow to take into account objective information. Indeed, in Gilboa and Schmeidler's (1989) model, objects of choices are acts (i.e., elements of $\mathcal{A}$). There is therefore no way to say, for instance, that the decision maker prefers the act f together with an information set $\mathcal{P}$ over an act g, together with an information set $\mathcal{Q}$. The (unique) set of probability distributions the decision maker uses to evaluate acts in Gilboa and Schmeidler's (1989) model is fixed and of purely subjective nature: it *only* depends on the decision maker's preferences[6].

We start by three quite standard Axioms, that require the preference relation $\succeq$ on $\mathcal{A} \times \mathbb{P}$ to be complete, transitive, non-degenerate and continuous.

Axiom 1 (Ordering)**.** *$\succeq$ is a reflexive, complete and transitive binary relation on $\mathcal{A} \times \mathbb{P}$.*

Axiom 2 (Non-degeneracy)**.** *For all $\mathcal{P} \in \mathbb{P}$, there exist $f,g \in \mathcal{A}$ such that $(f,\mathcal{P}) \succ (g,\mathcal{P})$.*

Axiom 3 (Continuity)**.** *For all f, g, $h \in \mathcal{A}$, and all $\mathcal{P} \in \mathbb{P}$, if $(f,\mathcal{P}) \succ (g,\mathcal{P}) \succ (h,\mathcal{P})$, then there exist α and β in $(0,1)$ such that :*

$$(\alpha f + (1-\alpha)h, \mathcal{P}) \succ (g,\mathcal{P}) \succ (\beta f + (1-\beta)h, \mathcal{P}).$$

[6]Note that this would also make difficult a state-dependent extension of Gilboa and Schmeidler's model, since one would then face the problem of the identification of state-dependent utilities on Y.

The following notion of mixture of *sets* of probability distributions will be extensively used in the sequel.

Notation 1. *For all $\mathcal{P}, \mathcal{Q} \in \mathbb{P}$, and all $\alpha \in [0,1]$, the $\alpha-$mixture of $\mathcal{P}$ and $\mathcal{Q}$ is defined by:*

$$\alpha\mathcal{P} + (1-\alpha)\mathcal{Q} = \{p \in \Delta(N) | p = \alpha p_1 + (1-\alpha)p_2 \ p_1 \in \mathcal{P},\, p_2 \in \mathcal{Q}\}.$$

An element $(f, \alpha\mathcal{P}+(1-\alpha)\mathcal{Q})$ of $\mathcal{A}\times\mathbb{P}$ can be interpreted as a compound lottery, in which in the first stage $(\mathcal{P}, f)$ and $(\mathcal{Q}, f)$ are obtained with probabilities α and $(1-\alpha)$, respectively. Let us consider $(f, \mathcal{P}_1)$, $(g, \mathcal{Q}_1)$, $(f, \mathcal{P}_2)$ and $(g, \mathcal{Q}_2)$ such that the Observer prefers $(f, \mathcal{P}_1)$ to $(g, \mathcal{Q}_1)$, and $(f, \mathcal{P}_2)$ to $(g, \mathcal{Q}_2)$. Assume, now, that the observer faces a choice between $(f, \alpha\mathcal{P}_1+(1-\alpha)\mathcal{Q}_1)$ and $(g, \alpha\mathcal{Q}_1+(1-\alpha)\mathcal{Q}_2)$. He might reason as follows: with probability α, I would obtain $(f, \mathcal{P}_1)$ if I have chosen $(f, \alpha\mathcal{P}_1+(1-\alpha)\mathcal{Q}_1)$, and $(g, \mathcal{Q}_1)$ if I have chosen $(g, \alpha\mathcal{Q}_1+(1-\alpha)\mathcal{Q}_2)$. Since I prefer $(f, \mathcal{P}_1)$ over $(g, \mathcal{Q}_1)$, it is better for me to choose $(f, \alpha\mathcal{P}_1+(1-\alpha)\mathcal{Q}_1)$, conditional on the realization of the event whose probability is α. Similarly, since I prefer $(f, \mathcal{Q}_1)$ over $(g, \mathcal{Q}_2)$, it is better for me to choose $(f, \alpha\mathcal{P}_1+(1-\alpha)\mathcal{Q}_1)$, conditional on the realization of the event whose probability is $(1-\alpha)$. Thus, I prefer *unconditionally* $(f, \alpha\mathcal{P}_1+(1-\alpha)\mathcal{Q}_1)$ over $(g, \alpha\mathcal{Q}_1+(1-\alpha)\mathcal{Q}_2)$. This leads us to the following Axiom, which is a mere extension of the "Constrained Independence Axiom" proposed by Karni and Safra (2000)[7].

Axiom 4 (Set-Mixture Independence)**.** *For all $\mathcal{P}_1, \mathcal{Q}_1, \mathcal{P}_2, \mathcal{Q}_2 \in \mathbb{P}$, and for all $f, g \in \mathcal{A}$,*

$$\left.\begin{array}{l}(f,\mathcal{P}_1) \succeq (\succ)(g,\mathcal{Q}_1)\\ (f,\mathcal{P}_2) \succeq (g,\mathcal{Q}_2)\end{array}\right\} \Rightarrow (f, \alpha\mathcal{P}_1+(1-\alpha)\mathcal{P}_2) \succeq (\succ)(g, \alpha\mathcal{Q}_1+(1-\alpha)\mathcal{Q}_2).$$

The next axiom concerns comparisons of information sets. It states that if an allocation f is judged better than another allocation g according to any probability distribution in $\mathcal{P}$, then $(f, \mathcal{P})$ is judged better than $(g, \mathcal{P})$. Formally:

Axiom 5 (Dominance)**.** *For all $\mathcal{P} \in \mathbb{P}$, if for all $p \in \mathcal{P}$, we have $(f, \{p\}) \succeq (g, \{p\})$, then $(f, \mathcal{P}) \succeq (g, \mathcal{P})$.*

Now, because we do not want to impose state-independence (that would lead in our framework to the conclusion that all individuals' preferences on Y are identical), we need to construct allocations that would play the role constant allocations usually play. To do so, we define the set $\mathcal{A}^{cv}$ of constant-valued allocations. These allocations are characterized by the fact that the observer is indifferent between being individual i or individual j for sure, for all couple (i, j). Formally,

$$\mathcal{A}^{cv} = \{f \in \mathcal{A} | (f, \{\delta_i\}) \sim (f, \{\delta_j\}), \forall i, j \in N\}.$$

[7] The above interpretation is exactly the one proposed by Karni and Safra (2000), where sets of probability distributions are considered instead of probability distributions.

The next axiom is a classical boundedness requirement, with respect to the set of constant-valued allocations. In particular, this axiom guarantees that $\mathcal{A}^{cv}$ is not empty.

Axiom 6 (Boundedness)**.** *For all $\mathcal{P} \in \mathbb{P}$ and $f \in \mathcal{A}$, there exist $\bar{f}$ and $\underline{f}$ in $\mathcal{A}^{cv}$ such that:*

$$(\bar{f}, \mathcal{P}) \succeq (f, \mathcal{P}) \succeq (\underline{f}, \mathcal{P}).$$

A similar axiom can be found, for instance, in Luce and Krantz (1971)'s axiomatization of state-dependent expected utility. Essentially, it amounts to assume that from the observer's point of view, the range of utility over social lotteries, conditionally on being individual i for sure, is the same as the range of utility over social lotteries, conditionally on being individual j for sure, for all $i, j \in N$.

The next axiom is the analogous of the $C-$independence Axiom of Gilboa and Schmeidler (1989), where the set of constant valued allocations replaces the set of constant allocations. It states that if $(f, \mathcal{P})$ is judged better than $(g, \mathcal{Q})$, then this relation is preserved if one mixes f and g with a same constant-valued allocation h.

Axiom 7 ($\mathcal{A}^{cv}$-Independence)**.** *For all $f, g \in \mathcal{A}$, $h \in \mathcal{A}^{cv}$, $\mathcal{P}, \mathcal{Q} \in \mathbb{P}$, and all $\alpha \in (0,1)$,*

$$(f, \mathcal{P}) \succeq (g, \mathcal{Q}) \Leftrightarrow (\alpha f + (1-\alpha)h, \mathcal{P}) \succeq (\alpha g + (1-\alpha)h, \mathcal{Q}).$$

The next axiom simply states that information on possible probability distributions is irrelevant when one compares constant-valued allocations.

Axiom 8 (Information indifference on $\mathcal{A}^{cv}$)**.** *For all $h \in \mathcal{A}^{cv}$, and all $\mathcal{P}, \mathcal{Q} \in \mathbb{P}$, $(h, \mathcal{P}) \sim (h, \mathcal{Q})$.*

We will also assume that only the part of the allocation on which there is a positive probability for some probability distribution in the information set matters. Formally, let $S(\mathcal{P})$ be the subset of N defined by $S(\mathcal{P}) = \{i \in N | \exists p \in \mathcal{P} \text{ s.t. } p(i) > 0\}$. In other words, $S(\mathcal{P})$ is the union of the supports of all probability distributions in the information set. For any subset E of N, and pair of allocations (f, g), we define the allocation $f_E g$ as $(f_E g)(i) = f(i)$ if $i \in E$, and $(f_E g)(i) = g(i)$ is $i \in N \setminus E$.

Axiom 9 (Equivalence)**.** *For all $f, g \in \mathcal{A}$, $\mathcal{P} \in \mathbb{P}$, $(f, \mathcal{P}) \sim (f_{S(\mathcal{P})} g, \mathcal{P})$.*

Finally, our last axiom is a version of the uncertainty aversion axiom of Schmeidler (1984), Chateauneuf (1991) and Gilboa and Schmeidler (1989). In the framework of decision under uncertainty, it simply stipulates that the decision maker exhibits (weak) preference for hedging. This axiom may also be interpreted from an ethical point of view. We defer this discussion to Section 4.

Axiom 10 (Uncertainty aversion)**.** *For all $f, g \in \mathcal{A}$, $\mathcal{P} \in \mathbb{P}$, $\alpha \in (0,1)$, $(f,\mathcal{P}) \sim (g,\mathcal{P})$ implies $(\alpha f + (1-\alpha)g, \mathcal{P}) \succeq (f,\mathcal{P})$.*

We recall that $\hat{\succeq}_i$ is the restriction of $\succeq$ on $\mathcal{A} \times \{\delta_i\}$. We say that a function $\hat{V}_i : Y \to \mathbb{R}$ *represents* $\hat{\succeq}_i$ if for all $f, g \in \mathcal{A}$:

$$(f, \{\delta_i\}) \succeq (g, \{\delta_i\}) \Leftrightarrow \hat{V}_i(f(i)) \geq \hat{V}_i(g(i)).$$

We can now state the following Theorem.

Theorem 1. *Axioms 1 to 10 hold if, and only if, there exist affine functions $\hat{V}_i : Y \to \mathbb{R}$, $i \in N$, representing $\hat{\succeq}_i$ such that $\hat{V}_i(Y) = \hat{V}_j(Y)$, for all $i, j \in N$, and a function $\mathcal{F} : \mathbb{P} \to \mathbb{P}_C$ satisfying, for all $\mathcal{P}, \mathcal{Q} \in \mathbb{P}$:*

1. *$\mathcal{F}(\mathcal{P}) \subseteq co(\mathcal{P})$*
2. *For all $\alpha \in [0,1]$, $\mathcal{F}(\alpha\mathcal{P} + (1-\alpha)\mathcal{Q}) = \alpha\mathcal{F}(\mathcal{P}) + (1-\alpha)\mathcal{F}(\mathcal{Q})$*

such that for all $(f,\mathcal{P}), (g,\mathcal{Q}) \in \mathcal{A} \times \mathbb{P}$, $(f,\mathcal{P}) \succeq (g,\mathcal{Q})$ if, and only if,

$$\min_{p \in \mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(f(i)) \geq \min_{p \in \mathcal{F}(\mathcal{Q})} \sum_i \hat{V}_i(g(i)).$$

Furthermore, $\mathcal{F}$ is unique, and $\{\hat{V}_i\}_{i \in N}$ are unique up to a common positive affine transformation.

Proof. See the Appendix. □

Finally, a more precise representation might be obtained in some cases, if one adds the two following axioms.

The first one may be interpreted as an anonymity requirement. The idea is that the only things that matter for the observer, besides the information set, are the utility *levels* he obtains, conditionally on being each individual. For any $f \in \mathcal{A}$, and any permutation $\varphi : N \to N$, define $\mathcal{A}(f^\varphi) = \left\{ g \in \mathcal{A} \,\middle|\, (g, \{\delta_i\}) \sim (f, \{\delta_{\varphi^{-1}(i)}\}), \forall i \in n \right\}$. Hence, for all $g \in \mathcal{A}(f^\varphi)$, individual i gets the same utility level as individual $\varphi^{-1}(i)$ under allocation f. Now, let $\mathcal{P}^\varphi$ be defined by $\mathcal{P}^\varphi = \{p^\varphi | p \in \mathcal{P}\}$, where for all $p \in \mathcal{P}$, $p^\varphi(i) = p(\varphi^{-1}(i))$ for all $i \in N$. Essentially, $(f,\mathcal{P})$ and $(g,\mathcal{P}^\varphi)$ (with φ a permutation and $g \in \mathcal{A}(f^\varphi)$) only differ by the names of the states. The following axiom states that the observer is indifferent between them.

Axiom 11 (Anonymity)**.** *For all $(f,\mathcal{P}) \in \mathcal{A} \times \mathbb{P}$ and all permutation $\varphi : N \to N$, and all $g \in \mathcal{A}(f^\varphi)$, $(f,\mathcal{P}) \sim (g,\mathcal{P}^\varphi)$.*

The next axiom states that whenever two allocations share the same worst probability distribution in $\mathcal{P}$, a mixture of these allocations will not reduce the degree of uncertainty, and therefore will not lead to an improvement, whenever $\mathcal{P}$ is the available information set. Recall that in Axiom 10, we interpreted preference for hedging (or mixture) as uncertainty aversion. With this in mind, indifference for hedging would be interpreted as *neutrality* towards uncertainty. The following axiom can therefore be viewed as a restricted neutrality towards uncertainty[8].

Axiom 12 (Restricted Mixture Neutrality). *For all $\mathcal{P} \in \mathbb{P}$, and all $f, g \in \mathcal{A}$, if there exists $p^* \in \mathcal{P}$ such that $(f, \{p\}) \succeq (f, \{p^*\})$ and $(g, \{p\}) \succeq (g, \{p^*\})$ for all $p \in \mathcal{P}$, then for all $\alpha \in [0,1]$,*

$$(f, \mathcal{P}) \sim (g, \mathcal{P}) \Leftrightarrow (\alpha f + (1-\alpha) g, \mathcal{P}) \sim (g, \mathcal{P}).$$

Before stating our next result, we need some additional piece of notation. For any $S \subseteq N$, let $\Delta(S)$ be the set of all probability distributions with support in S. Let:

$$\mathbb{B} = \{\mathcal{P} \in \mathbb{P} | \exists (\alpha_t)_{t=1,\dots,r} \in [0,1] \text{ and } (S_t)_{t=1,\dots,r} \in N \text{ s.t. } \sum_t \alpha_t = 1, \mathcal{P} = \sum_t \alpha_t \Delta(S_t)\}.$$

Hence, an element of $\mathbb{B}$ is a linear combination of simplices with supports in N. Furthermore, for $\mathcal{P} = \sum_t \alpha_t \Delta(S_t)$, let $c(\mathcal{P}) = \sum_t \alpha_t c(\Delta(S_t))$, where $c(\Delta(S_t))$ is the probability distribution defined by $c(\Delta(S_t))(s) = \frac{1}{|S_t|}$ (hence, $c(\Delta(S_t))$ is the uniform distribution on S_t).

We then obtain the following representation when the information belongs to $\mathbb{B}$[9].

Theorem 2. *Under the assumptions of Theorem 1, Axioms 11 and 12 hold if, and only if, there exist affine functions $\hat{V}_i : Y \to \mathbb{R}$, $i \in N$, representing $\hat{\succeq}_i$ such that $\hat{V}_i(Y) = \hat{V}_j(Y)$, for all $i, j \in N$, and $\theta \in [0,1]$ such that for all $\mathcal{P}, \mathcal{Q} \in \mathbb{B}$, and all $f, g \in \mathcal{A}$, $(f, \mathcal{P}) \succeq (g, \mathcal{Q})$ if, and only if:*

$$\theta \min_{p \in \mathcal{P}} \sum_i p(i) \hat{V}_i(f(i)) + (1-\theta) \sum_i c(\mathcal{P})(i) \hat{V}_i(f(i)) \geq \theta \min_{p \in \mathcal{Q}} \sum_i p(i) \hat{V}_i(g(i)) + (1-\theta) \sum_i c(\mathcal{Q})(i) \hat{V}_i(g(i)).$$

Furthermore, θ is unique, and $\{\hat{V}_i\}_{i \in N}$ are unique up to a common positive affine transformation.

Proof. See the appendix. □

[8] Mixture neutrality can be considered as the analogue of the *betweeness* property for preferences under risk (on the betweeness property, see e.g., Chew (1983; 1989) and Dekel (1986)). Attitude towards mixture is considered in details in Klibanoff (2001).

[9] The set $\mathbb{B}$ is known in decision theory as the set of cores of beliefs. Actually, the following representation theorem, which is only given when the information takes the form of a core of a belief, can be extended to convex combinations of symmetric polytopes. However, for the problem under consideration, such an extension would not be a great improvement.

4 The Ignorant Observer Theorem

We now define individuals' preferences $\succeq_i$ on Y. As Harsanyi does, we assume that individuals obey von Neumann and Morgenstern's axiom. It should be noted that this assumption is by no means in conflict with Rawls' views. Indeed, Rawls only rejected the use of the Bayesian doctrine from behind the veil of ignorance. Formally, these axioms are as follows.

Axiom 13 (Ordering)**.** *$\succeq_i$ is a reflexive, complete and transitive binary relation on Y.*

Axiom 14 (Continuity)**.** *For all $w, y, z \in Y$ such that $w \succ_i y \succ_i z$, there exists and $\alpha \in (0, 1)$ such that $\alpha w + (1 - \alpha)z \sim_i y$.*

Axiom 15 (Independence)**.** *For all $w, y, z \in Y$ and all $\alpha \in [0, 1]$, $y \succeq_i z$ implies $\alpha y + (1-\alpha)w \succeq_i \alpha z + (1 - \alpha)w$.*

As well known (see e.g. Fishburn (1970)) a preference relation satisfies these three axioms if, and only if, it can be represented by an Expected Utility functional. This is stated formally in the following theorem.

Theorem 3. *Axiom 13 to 15 hold if, and only if, there exists a linear real-valued function U_i on Y such that for all $y, z \in Y$, $y \succeq_i z$ if, and only if, $U_i(y) \geq U_i(z)$. Furthermore, such a representation is unique up to a positive affine transformation.*

Our aim is to deduce, from individuals' preferences $\succeq_i$ on Y and the Observer's preferences $\succeq$ on $\mathcal{A} \times \mathbb{P}$, a "social preference" $\succeq^*$ on Y. In order to do so, one needs to specify how these preferences interact.

Individuals' preferences and the observer's ones are linked by the so-called "acceptance principle" (see Harsanyi (1977)) which states that if the observer is *sure* to be i, her choices should be the same as those of i. Let δ_i be the probability distribution on N defined by $\delta_i(i) = 1$. The acceptance axiom can be restated as follows in our framework.

Axiom 16 (Acceptance)**.** *For all $i \in N$, and all $f, g \in \mathcal{A}$, $(f, \{\delta_i\}) \succeq (g, \{\delta_i\})$ if and only if $f(i) \succeq_i g(i)$.*

We now turn to the link between social preferences and the observer's ones. The fundamental idea of the veil of ignorance is that (fair) social preferences are those of an observer who is *totally ignorant* about the position he will eventually get in the society. There are two issues here. First, social preferences are defined over social alternatives in Y, whereas the observer's preferences are defined on the product $\mathcal{A} \times \mathbb{P}$ of allocations and information sets. For all $y \in Y$, let $k_y \in \mathcal{A}$ be the allocation defined by $k_y(i) = y$ for all $i \in N$. It is natural to define the observer's preferences

$\succeq_Y$ on $Y \times \mathbb{P}$ as follows: $(y, \mathcal{P}) \succeq_Y (z, \mathcal{Q})$ iff $(k_y, \mathcal{P}) \succeq (k_z, \mathcal{Q})$. Now, we should formalize the idea that the observer is totally ignorant about his position. In our framework, this is captured by the fact that the information set is $\Delta(N)$, the set of all probability distributions over the set N of individuals. This leads us to the following axiom.

Axiom 17 (Ignorance)**.** *For all $y, z \in Y$, $y \succeq^* z$ if, and only if, $(k_y, \Delta(N)) \succeq (k_z, \Delta(N))$.*

We can now state a first representation Theorem.

Theorem 4. *Assume that the observer's preferences satisfy all axioms of Theorem 1, and individuals' preferences satisfy all the axioms of Theorem 3. Then Axioms 16 and 17 hold if, and only if, there exist affine functions $V_i : Y \to \mathbb{R}$, $i \in N$, representing $\succeq_i$, and a function $\mathcal{F} : \mathbb{P} \to \mathbb{P}_C$ satisfying, for all $\mathcal{P}, \mathcal{Q} \in \mathbb{P}$:*

1. *$\mathcal{F}(\mathcal{P}) \subseteq co(\mathcal{P})$*
2. *For all $\alpha \in [0,1]$, $\mathcal{F}(\alpha\mathcal{P} + (1-\alpha)\mathcal{Q}) = \alpha\mathcal{F}(\mathcal{P}) + (1-\alpha)\mathcal{F}(\mathcal{Q})$*

such that for all $y, z \in Y$, $y \succeq^* z$ if, and only if,

$$\min_{p \in \mathcal{F}(\Delta(N))} \sum_i p(i) \frac{V_i(y) - \min_{w \in Y} V_i(w)}{\max_{w \in Y} V_i(w) - \min_{w \in Y} V_i(w)} \geq \min_{p \in \mathcal{F}(\Delta(N))} \sum_i p(i) \frac{V_i(z) - \min_{w \in Y} V_i(w)}{\max_{w \in Y} V_i(w) - \min_{w \in Y} V_i(w)}.$$

Furthermore, $\mathcal{F}$ is unique.

Proof. This is a straightforward corollary of Theorem 1. Axiom 16 implies $\succeq_i = \hat{\succeq}_i$. Let V_i be an affine representation of $\succeq_i$. By Theorem 3, V_i is unique up to a positive affine transformation. Therefore, $V_i^*(y) = \frac{V_i(y) - \min_{w \in Y} V_i(w)}{\max_{w \in Y} V_i(w) - \min_{w \in Y} V_i(w)}$ is also an affine representation of $\succeq_i$. Furthermore, $V_i^*(Y) = V_j^*(Y)$ for all $i, j \in N$. Define $\tilde{V}_i(f) = V_i^*(f(i))$ for all $f \in \mathcal{A}$. Clearly, $\tilde{V}_i$ are affine representations of $\hat{\succeq}_i$ and satisfy $\tilde{V}_i(Y) = \tilde{V}_j(Y)$ for all $i, j \in N$. The result then follows from Theorem 1. $\square$

It should be noted that, although the *impartial* observer only compares allocations under complete ignorance, the above theorem rests on a primitive ordering on the entire set $\mathcal{A} \times \mathbb{P}$. In other words, the domain of the observer's preference is in some sense too large. This is certainly a drawback, which is common to Harsanyi's Theorem and ours (Harsanyi's observer's preferences are defined over *all* extended lotteries, although only constant impartial lotteries are eventually considered by the impartial observer). Karni and Weymark (1998) pointed out this difficulty, and solved it for Harsanyi's Theorem, providing an informationally parsimonious version of Harsanyi's Impartial Observer Theorem. It is an open question whether a similar improvement could be made in our case. However, we did not feel that this question was essential, and left it aside.

Observe that the exact form of $\mathcal{F}(\Delta(N))$ will actually depend on the decision maker *uncertainty aversion.* A more precise form may be obtained if we use Axioms 11 and 12. Indeed, Theorem 2 leads to the following result.

Theorem 5. *Assume that the observer's preferences satisfy all axioms of Theorem 2, and individuals' preferences satisfy all the axioms of Theorem 3. Then Axioms 16 and 17 hold if, and only if, there exist affine functions* $V_i : Y \to \mathbb{R}$, $i \in N$, *representing* $\succeq_i$, *and a real number* $\theta \in [0,1]$, *such that for* $y, z \in Y$, $y \succeq^* z$ *if, and only if:*

$$\theta \min_{i\in N} \frac{V_i(y) - \min_{w\in Y} V_i(w)}{\max_{w\in Y} V_i(w) - \min_{w\in Y} V_i(w)} + (1-\theta) \sum_{i\in N} \frac{1}{n} \frac{V_i(y) - \min_{w\in Y} V_i(w)}{\max_{w\in Y} V_i(w) - \min_{w\in Y} V_i(w)}$$
$$\geq \theta \min_{i\in N} \frac{V_i(z) - \min_{w\in Y} V_i(w)}{\max_{w\in Y} V_i(w) - \min_{w\in Y} V_i(w)} + (1-\theta) \sum_{i\in N} \frac{1}{n} \frac{V_i(z) - \min_{w\in Y} V_i(w)}{\max_{w\in Y} V_i(w) - \min_{w\in Y} V_i(w)}.$$

Furthermore, θ *is unique.*

Proof. Similar to that of Theorem 4. □

Such a criterion can be easily interpreted: it consists in a weighted average between Harsanyi's utilitarian criterion, and Rawls' maxmin criterion. It should be noted that this specific functional form comes from Axioms 11 and 12. Axiom 11 plays a transparent role: it ensures that the symmetry of the set of probability distributions that represents the available information will be preserved in the decision rule. Axiom 12 forces the set of probability distributions used in the decision rule to have a similar shape as the set of probability distributions that represents the available information.

Let us note that, in both Theorem 4 and 5, individuals' utility functions $\frac{V_i(y) - \min_{w\in Y} V_i(w)}{\max_{w\in Y} V_i(w) - \min_{w\in Y} V_i(w)}$ are cardinally measurable and unit comparable. Therefore, the weights assigned to these functions cannot be manipulated, and are meaningful. As mentioned above, the normalization of individuals' utility functions we obtained is very similar to that obtained by Karni (1998), Dhillon and Mertens (1999), Segal (2000) and Moreno-Ternero and Roemer (2005)[10] : In all these works, the weight of each individual depends on the diameter of the range of her utility function[11]. Since Karni (1998) is the paper to which we are the closest, let us insist on a significant difference between Karni's approach and the one we followed here. First, Karni assumes that the observer's preferences are defined on extended lotteries, in Harsanyi's sense[12]. This

[10]Although in a very restrictive case (i.e., when individuals' preferences are risk-isomorphic).
[11]See also Karni (2003) for a related approach.
[12]Although he uses Anscombe and Aumann's (1963) formalism.

implies that he faces the well-known problem of the determination of the weights the observer attaches to each individual. This is where his "Impartiality Axiom", plays a key role. Our approach is rather different: since the weights that appear in our representation theorem are based on objective information (i.e., since probability distributions on individual lotteries are part of the model), we do not face the problem of their determination. Hence, the normalization of individuals' utilities can be seen as coming from a strictly epistemic axiom (namely, Axiom 6), which should not be interpreted in terms of impartiality: our impartiality requirement lies actually in the nature of the information the observer can use to make her decisions.

5 The justification problem

5.1 Under what conditions is the observer utilitarian or rawlsian?

According to our Ignorant Observer theorem, there is not one but a plurality of decision rules complying with the impartiality requirement (as formalized by the veil of ignorance). These include both Harsanyi's utilitarian criterion and Rawls' maximin principle.

For instance, assuming that ignorant observer is an expected utility maximizer, one would obviously obtain the utilitarian rule. This requirement would take, in our framework, the following form, which is a strengthening of Axiom 10:

Axiom 18 (Neutrality towards uncertainty)**.** *For all* $f, g \in \mathcal{A}$ *and all* $\mathcal{P} \in \mathbb{P}$*,* $(f, \mathcal{P}) \sim (g, \mathcal{P})$ *implies, for all* $\alpha \in (0,1)$*,* $(\alpha f + (1-\alpha)g, \mathcal{P}) \sim (f, \mathcal{P})$*.*

On the other hand, the ignorant observer would be rawlsian if, and only if, she obeys the following axiom.

Axiom 19 (Extreme aversion towards uncertainty)**.** *For all* $f \in \mathcal{A}$*,* $\mathcal{P} \in \mathbb{P}$ *and* $p \in \mathcal{P}$*,* $(f, \{p\}) \succeq (f, \mathcal{P})$*.*

How can such axioms be justified? We have two possibilities. We may use a decision theoretic semantics, interpreting these axioms from an epistemic point of view. But we may also use a social choice semantics, interpreting these axioms from an ethical point of view. In the next two subsections, we investigate these interpretations. This investigation aims at clearly identifying the ethical and epistemic foundations of the veil of ignorance. It can be seen as a continuation of Mongin's (2001) attempt to clarify the foundations of Harsanyi's approach.

5.2 An epistemic view on Rawls-Harsanyi's debate

The model shows that the rational choice of the ignorant observer depends on her attitude towards uncertainty. If she is neutral to uncertainty she will use the expected utility criterion, whereas if she is extremely averse to uncertainty, she will use the maximin criterion. But, from an epistemic point of view, it is reasonable to suppose that she is none of the two. Hence, it seems difficult to defend Axioms 18 and 19 on an epistemic basis.

Consider first Axiom 18. In a decision theoretic semantics, this axiom can be interpreted as neutrality towards uncertainty. As shown by Ellsberg (1961), such a neutrality towards uncertainty is very unlikely. Consider an individual facing an urn, containing 100 balls, either black or white, in unknown proportion. The decision maker has the choice between betting on black (then receiving 100€ if the drawn ball is black, and 0€ if it is white), and betting on white (then receiving 100€ if the drawn ball is white, and 0€ if it is black). Much likely, the decision maker would be indifferent between these two bets. Now, one proposes to this decision maker the following bet: win 50€ whatever the color of the drawn ball will be. Most of the decision makers would prefer such a bet to any of the two first ones. Hence, there are good reasons to dismiss this axiom from a decision theoretic point of view.

Is Axiom 19 more appealing? This axiom requires in particular the observer to prefer (weakly) *any* lottery on the set of individuals to complete ignorance. In particular, she must prefer (weakly) to be the worst-off individuals for sure, rather than facing complete ignorance concerning her identity. Such a requirement seems very unlikely from a decision theoretic point of view.

Actually, Rawls added two epistemic arguments for the maximin criterion. The first one relates to the "very considerable normal risk-aversion (given the special features of the original position)" (Rawls (1974a), p.144). He pointed out "the crucial nature of the decision in the original position" (id., p.143). It is a *solemn* choice which consequences are decisive not only for the decision maker, but also for her family and her descendants. For, she has to choose which type of society they will leave in now and forever. This argument implies that the *importance* of the decision under consideration should be taken into account. But this raises new difficulties, for it is clear that the degree of importance of a decision is something essentially subjective. All in all, this is a completely different model, that is far away from existing models of decision making under uncertainty. And the main insight of the veil of ignorance – namely, that assuming ignorance is enough to know what a fair decision should be, without any other assumptions specific to the fact that the problem under consideration is a social choice problem – would be lost.

The second epistemic argument Rawls raised in favor of the maximin relates to the "contractual" meaning of the Original Position. The partners are to favor the principles that any of them would agree with (in Rawls' terminology, there should be a public and unanimous agreement). This entails, says Rawls, that any of the partners has a veto. Taking into account this veto, the rational solution is to be prudent, and therefore to use the maximin criterion, for "the maximin criterion assures the less favored that inequalities work to their advantage" (id.). Actually, some authors, (see, e.g., Binmore ((1989)) did propose a model of collective bargaining or a voting model including a veto procedure from behind the veil of ignorance. But, at best, this argument leads to conclude that *contracting* from behind the veil of ignorance leads to a maxmin allocation rule. It cannot be used to prove that an impartial observer should apply the maxmin criterion.

We shall therefore conclude that the Rawls-Harsanyi debate cannot be resolved in favor of any of the protagonists on an epistemic basis. It is therefore necessary to contemplate an ethical interpretation of Axioms 18 and 19. Actually we shall show, in the next subsection, that it is even more difficult to provide an ethical justification either for Axiom 18 or for Axiom 19.

5.3 An ethical view on Rawls-Harsanyi's debate

Axiom 18 is relatively easy to interpret from an ethical point of view, since it is directly related to a kind of neutrality towards inequalities. To grasp the intuition behind the axiom, let us consider the following simple example.

Assume that the society is composed of two individuals, 1 and 2, and let $\mathcal{P} = \Delta(\{1,2\})$, $\Delta(\{1,2\})$ is the set of all probability distributions on $\{1,2\}$. Let f and g be two allocations, whose incomes (in terms of utilities, represented by a function representing $\succeq_i$) are as follows:

	1	2
f	1	0
g	0	1
h	$\frac{1}{2}$	$\frac{1}{2}$

Assume that the ignorant observer is indifferent between $(f, \Delta(\{1,2\}))$ and $(g, \Delta(\{1,2\}))$ (this would be the case, in particular, if Axiom 11 holds). Observe that f and g are highly unequal. Now, consider h, which is simply defined as $\frac{1}{2}f + \frac{1}{2}g$. Obviously, h is less unequal than f and g. It makes sense, therefore, to assume that the ignorant observer will prefer h to f and g. However, Axiom 18 forces the observer to be *indifferent* between f and g. Actually, this is the main point made by Diamond (1967)'s famous objection to Harsanyi. Such an argument can also be found, e.g., in Epstein and Segal (1992).

We now turn to Axiom 19. This axiom says that the Ignorant Observer gives consideration only to the worst consequences. This assumption is inconsistent with Rawls' own interpretation of the Original Position. According to him (Rawls, (1974a; 1975), the decision maker placed under the veil of ignorance should consider all members of the society as free and equal moral persons. Then, her decision should comply with two requirements dealing with the society members' opportunities. First, "those able to gain from their good fortune do so in ways agreeable to those less favoured". (Rawls, (1974a), p.145). Secondly, "natural variations (in talents) are recognised as an opportunity, particularly since they are often complementary and form a basis for social ties" (ibid.). The first requirement imposes a Pareto-optimality condition. The second requirement imposes a condition of "fair opportunity" for each member of the society (ibid.). It is obvious that such conditions contradict with Axiom 19.

5.4 Arguments for a compromise

Thus, none of the axioms 18 or 19 are ethically appealing. We would like, here, to give some arguments in favor of the compromise suggested by Theorem 5. Once one is convinced that both Rawls' and Harsanyi's criteria are unappealing, adoption of the Axioms of Theorem 5 would indeed lead to a strict compromise between Rawls' egalitarianism and Harsanyi's utilitarianism.

The two key axioms that allow to obtain Theorem 5 are Axioms 11 and 12. The first one is a standard anonymity assumption, that simply states that individuals' names are irrelevant when one compares two allocations. Axiom 12 is less usual, and can actually be viewed as a restriction of Axioms 18, that escapes Diamond's critique. Indeed, consider again Diamond's example presented in section 5.3. It is clearly not the case that f and g share the same worst probability distribution in $\Delta(\{1,2\})$, since the worst case if f is chosen is to be individual 2 for sure, whereas the worst case if g is chosen is to be individual 1 for sure. Therefore, one can not conclude that $h \sim f$. Now, consider the following variation of Diamond's example, with 3 individuals.

	1	2	3
f	1	0	0
g	1	1	0
h	1	$\frac{1}{2}$	0

Assume, now, that the observer is indifferent between $(f, \Delta(\{1,2,3\}))$ and $(g, \Delta(\{1,2,3\}))$. Since δ_3 is the worst probability distribution for both f and g, Axiom 12 leads to the conclusion that $(h, \Delta(\{1,2,3\})) \sim (f, \Delta(\{1,2,3\})) \sim (g, \Delta(\{1,2,3\}))$. Is this reasonable? Yes, insofar as the observer's indifference between $(f, \Delta(\{1,2,3\}))$ and $(g, \Delta(\{1,2,3\}))$ indicates that he cares a lot about the worst-off individual, and this individual is the same whatever the allocation chosen (namely, 3), and has always the same utility.

This leads us to believe that the allocation rule proposed in Theorem 5 (with $\theta \in (0,1)$) is a reasonable one. Of course, this rule is in sharp conflict with Harsanyi's view, since it is not compatible with utilitarianism, as soon as $\theta > 0$. On the other hand, it is not sure that Rawls would have been unpleased with it. Although after *A Theory of Justice* was published, he wrote some articles to defend the maximin (see e.g., Rawls (1974a), (1974b)), his arguments were often conflicting, and even contradicting, as we have seen, his own interpretation of the veil of ignorance. We take these contradictions as an evidence that his main purpose was not to defend a specific criterion of rational decision under ignorance. Since his project was first to propose a theory of social justice alternative to Utilitarianism, his main purpose with the veil of ignorance model was to acknowledge a solution excluding Utilitarianism. As himself wrote:

> "But I do not wish to overemphasize this criterion: a deeper investigation (...) may show that some other conception of justice is more reasonable."
>
> Rawls, (1974a), p.145

Appendix

Proof of Theorem 1

The necessity part of the Theorem is easily checked. We therefore only prove the sufficiency part. The proof goes through several claims. Although not explicitly stated in the claims, all the assumptions of Theorem 1 are made throughout this subsection.

Claim 1. *$\mathcal{A}^{cv}$ is convex.*

Proof. Let $f, g \in \mathcal{A}^{cv}$, and $\alpha \in [0,1]$. By definition of $\mathcal{A}^{cv}$, $(f, \{\delta_i\}) \sim (f, \{\delta_j\})$ for all $i, j \in N$. Therefore, by Axiom 7, $(\alpha f + (1-\alpha)g, \{\delta_i\}) \sim (\alpha f + (1-\alpha)g, \{\delta_j\})$ for all $i, j \in N$. Hence, $\alpha f + (1-\alpha)g \in \mathcal{A}^{cv}$, which proves that $\mathcal{A}^{cv}$ is convex. □

Claim 2. *For all $i \in N$, all $f, g, h \in \mathcal{A}$, and all $\alpha \in (0,1)$,*

$$(f, \{\delta_i\}) \succeq (g, \{\delta_i\}) \Leftrightarrow (\alpha f + (1-\alpha)h, \{\delta_i\}) \succeq (\alpha g + (1-\alpha)h, \{\delta_i\}).$$

Proof. Let $f, g, h \in \mathcal{A}$. By Axiom 6, there exist $\bar{h}, \underline{h} \in \mathcal{A}^{cv}$ such that $(\bar{h}, \{\delta_i\}) \succeq (h, \{\delta_i\}) \succeq (\underline{h}, \{\delta_i\})$. Hence, by Axioms 1 and 3, there exists $\theta \in [0,1]$ such that $(h, \{\delta_i\}) \sim (\theta\bar{h} + (1-\theta)\underline{h}, \{\delta_i\})$. Let $\hat{h} = \theta\bar{h} + (1-\theta)\underline{h}$. By Claim 1, $\hat{h} \in \mathcal{A}^{cv}$. Next, let $\tilde{h}$ be defined by $\tilde{h}(i) = h(i)$, and $\tilde{h}(j) = \hat{h}(j)$ for all $j \neq i$. We then have, by Axiom 9, $(\tilde{h}, \{\delta_i\}) \sim (h, \{\delta_i\}) \sim (\hat{h}, \{\delta_i\})$. Furthermore, for all $j \neq i$, $(\tilde{h}, \{\delta_j\}) \sim (\tilde{h}, \{\delta_i\})$ by construction. Therefore, $\tilde{h} \in \mathcal{A}^{cv}$. By Axiom 7, for all $\alpha \in (0,1)$,

$$(f, \{\delta_i\}) \succeq (g, \{\delta_i\}) \Leftrightarrow (\alpha f + (1-\alpha)\tilde{h}, \{\delta_i\}) \succeq (\alpha g + (1-\alpha)\tilde{h}, \{\delta_i\}).$$

But, by Axiom 9, $(\alpha f + (1-\alpha)\tilde{h}, \{\delta_i\}) \sim (f + (1-\alpha)h, \{\delta_i\})$, and $(\alpha g + (1-\alpha)\tilde{h}, \{\delta_i\}) \sim (g + (1-\alpha)h, \{\delta_i\})$. Therefore,

$$(f, \{\delta_i\}) \succeq (g, \{\delta_i\}) \Leftrightarrow (\alpha f + (1-\alpha)h, \{\delta_i\}) \succeq (\alpha g + (1-\alpha)h, \{\delta_i\}),$$

the desired result. □

Claim 3. *For all i, there exits an affine function $\hat{V}_i : Y \to \mathbb{R}$ such that for all $f, g \in \mathcal{A}$,*

$$(f, \{\delta_i\}) \succeq (g, \{\delta_i\}) \Leftrightarrow \hat{V}_i(f(i)) \geq \hat{V}_i(g(i)).$$

Furthermore, $\hat{V}_i$ is unique up to a positive affine transformation.

Proof. Let i be fixed in N, and $\hat{\succeq}_i$ be the restriction of $\succeq$ to $\mathcal{A} \times \{\delta_i\}$. By Axioms 1, 3 and Claim 2, $\hat{\succeq}_i$ satisfies von Neumann-Morgenstern Axioms. Therefore there exists an affine function $U : \mathcal{A} \to \mathbb{R}$, unique up to a positive affine transformation, such that for all $f, g \in \mathcal{A}$,

$$(f, \{\delta_i\}) \succeq (g, \{\delta_i\}) \Leftrightarrow U_i(f) \geq U_i(g).$$

By Axiom 9, for all $f, f' \in \mathcal{A}$ such that $f(i) = f'(i)$, $(f, \{\delta_i\}) \sim (f', \{\delta_i\})$. Therefore, defining $\hat{V}_i : Y \to \mathbb{R}$ by $\hat{V}_i(f(i)) = U_i(f')$ for all $f' \in \mathcal{A}$ such that $f'(i) = f(i)$, one obtains the desired result. $\square$

In the sequel, we will make the following slight abuse of notation: we will denote by $\hat{V}_i$ both the function $\hat{V}_i : Y \to \mathbb{R}$ defined in Claim 3 and the function $\tilde{V}_i : \mathcal{A} \to \mathbb{R}$ defined by $\tilde{V}_i(f) = \hat{V}_i(f(i))$.

Claim 4. *For all $\mathcal{P} \in \mathbb{P}$, $f, g \in \mathcal{A}$,*

$$\begin{cases} (f, \{\delta_i\}) \succ (g, \{\delta_i\}), \forall i \in N \Rightarrow (f, \mathcal{P}) \succ (g, \mathcal{P}) \\ (f, \{\delta_i\}) \succeq (g, \{\delta_i\}), \forall i \in N \Rightarrow (f, \mathcal{P}) \succeq (g, \mathcal{P}). \end{cases}$$

Proof. Let $f, g \in \mathcal{A}$ be such that $(f, \{\delta_i\}) \succ (g, \{\delta_i\})$ for all $i \in N$. By Axiom 4, for all $(\alpha_1, \cdots, \alpha_n)$ such that $\alpha_i \geq 0$ for all i, and $\sum_{i \in N} \alpha_i = 1$, we have $(f, \sum_{i \in N} \alpha_i \{\delta_i\}) \succ (g, \sum_{i \in N} \alpha_i \{\delta_i\})$. Therefore, for all $p \in \Delta(N)$, $(f, \{p\}) \succ (g, \{p\})$. Hence, by Axiom 5, for all $\mathcal{P} \in \mathbb{P}$, $(f, \mathcal{P}) \succ (g, \mathcal{P})$. The second part of the Claim is proved using the same argument. $\square$

Claim 5. *There exist f and g in $\mathcal{A}^{cv}$ such that for all $i \in N$, $(f, \{\delta_i\}) \succ (g, \{\delta_i\})$.*

Proof. Let $\mathcal{P} \in \mathbb{P}$ be fixed. By Axiom 2, there exists $\hat{f}, \hat{g} \in \mathcal{A}$ such that $(\hat{f}, \mathcal{P}) \succ (\hat{g}, \mathcal{P})$. By Axiom 6, there exist $f, g \in \mathcal{A}^{cv}$ such that $(f, \mathcal{P}) \succeq (\hat{f}, \mathcal{P})$ and $(\hat{g}, \mathcal{P}) \succeq (g, \mathcal{P})$. Therefore, $(f, \mathcal{P}) \succ (g, \mathcal{P})$. Since f and g belong to $\mathcal{A}^{cv}$, Axiom 8 implies that $(f, \{\delta_i\}) \succ (g, \{\delta_i\})$ for all $i \in N$. $\square$

Claim 6. *For all $\mathcal{P} \in \mathbb{P}$, there exists an $\mathcal{A}^{cv}-$affine functional $V_{\mathcal{P}} : \mathcal{A} \to \mathbb{R}$ such that for all $f, g \in \mathcal{A}$:*

$$(f, \mathcal{P}) \succeq (g, \mathcal{P}) \Leftrightarrow V_{\mathcal{P}}(f) \geq V_{\mathcal{P}}(g).$$

Furthermore, $V_{\mathcal{P}}$ is unique up to a positive affine transformation, and $V_{\mathcal{P}}(\mathcal{A}) = V_{\mathcal{P}}(\mathcal{A}^{cv})$.

Proof. This result follows from Claim 1, Axioms 1, 3, 6, 7 and Corollary 2 in Castagnoli, Marinacci and Maccherroni (2003). □

For all $h \in \mathcal{A} \setminus \mathcal{A}^{cv}$, let $\mathcal{A}_h = co\{h, \mathcal{A}^{cv}\}$.

Claim 7. *There exist affine representations $\hat{V}_i$ $(i \in N)$ of $\succeq$ on $\mathcal{A} \times \{\delta_i\}$, satisfying $\hat{V}_i(\mathcal{A}) = \hat{V}_j(\mathcal{A})$, for all $i, j \in N$, such that, for all $\mathcal{P} \in \mathbb{P}$, $h \in \mathcal{A} \setminus \mathcal{A}^{cv}$, there exits an $\mathcal{A}^{cv}-$affine representation $V_{\mathcal{P}}$ of $\succeq$ on $\mathcal{A} \times \{\mathcal{P}\}$ and non-negative numbers $\lambda_1(h, \mathcal{P}), \cdots, \lambda_n(h, \mathcal{P})$, not all equal to zero, and summing up to one, such that, for all $f \in \mathcal{A}_h$,*

$$V_{\mathcal{P}}(f) = \sum_{i \in N} \lambda_i(h, \mathcal{P}) \hat{V}_i(f).$$

Moreover, $\{\hat{V}_i\}_{i \in N}$ are unique up to a common positive affine transformation.

Proof. For all i, let f_i^* be such that $f_i^* \hat{\succeq}_i f$ for all $f \in \mathcal{A}$, and f_{*i} be such that $f \hat{\succeq}_i f_{*i}$ for all $f \in \mathcal{A}$. These allocations are well defined, since $\mathcal{A}$ is a compact set. Define f^* and f_* by $f^*(i) = f_i^*(i)$ and $f_*(i) = f_{*i}(i)$ for all $i \in N$.

By Axiom 9, $f^* \hat{\sim}_i f_i^*$, and $f_* \hat{\sim}_i f_{*i}$ for all $i \in N$. Therefore, by Claim 4, $f^* \hat{\succeq}_i f \hat{\succeq}_i f_*$ for all $f \in \mathcal{A}$ and all $i \in N$. Let $\hat{V}_i$ $(i \in N)$ be affine representations of $\hat{\succeq}_i$, as defined in Claim 3, and $V_{\mathcal{P}}$ be an $\mathcal{A}^{cv}-$affine representation of $\succeq$ on $\mathcal{A} \times \{\mathcal{P}\}$, as defined in Claim 6. Without loss of generality, since the $\hat{V}_i$ and $V_{\mathcal{P}}$ are defined up to a positive affine transformation, we can choose $\hat{V}_i$ such that $\hat{V}_i(f^*) = V_{\mathcal{P}}(f^*) = 1$ and $\hat{V}_i(f_*) = V_{\mathcal{P}}(f_*) = -1$ for all $i \in N$. We thus have $\hat{V}_i(\mathcal{A}) = [-1, 1]$ for all $i \in N$.

We now show that for all $f \in \mathcal{A}^{cv}$, and all $i, j \in N$, $\hat{V}_i(f) = \hat{V}_j(f)$. Let $f \in \mathcal{A}^{cv}$. Three cases are possible.

Case 1. $f^* \hat{\succeq}_i f \hat{\succeq}_i f_*$, for all $i \in N$.

By Axioms 1 and 3, for all $\mathcal{P} \in \mathbb{P}$, there exists $\alpha_{\mathcal{P}}$ such that $(f, \mathcal{P}) \sim (\alpha_{\mathcal{P}} f^* + (1 - \alpha_{\mathcal{P}}) f_*, \mathcal{P})$. By Claim 1, $\alpha_{\mathcal{P}} f^* + (1 - \alpha_{\mathcal{P}}) f_* \in \mathcal{A}^{cv}$. Therefore, by Axiom 8, $\alpha_{\mathcal{P}} f^* + (1 - \alpha_{\mathcal{P}}) f_* \hat{\sim}_i f$ for all $i \in N$. Hence, for all $i \in N$, $\hat{V}_i(f) = \alpha_{\mathcal{P}} \hat{V}_i(f^*) + (1 - \alpha_{\mathcal{P}}) \hat{V}_i(f_*) = 2\alpha_{\mathcal{P}} - 1$, which proves that $\hat{V}_i(f) = \hat{V}_j(f)$ for all $i, j \in N$.

Case 2. $f \hat{\succeq}_i f^*$, for all $i \in N$.

Again, by Axioms 1 and 3, for all $\mathcal{P} \in \mathbb{P}$, there exists $\alpha_{\mathcal{P}}$ such that $(f^*, \mathcal{P}) \sim (\alpha_{\mathcal{P}} f + (1 - \alpha_{\mathcal{P}}) f_*, \mathcal{P})$. By Claim 1, $\alpha_{\mathcal{P}} f + (1 - \alpha_{\mathcal{P}}) f_* \in \mathcal{A}^{cv}$. Therefore, by Axiom 8, $\alpha_{\mathcal{P}} f + (1 - \alpha_{\mathcal{P}}) f_* \hat{\sim}_i f^*$ for all $i \in N$. Hence, for all $i \in N$, $\hat{V}_i(f^*) = \alpha_{\mathcal{P}} \hat{V}_i(f) + (1 - \alpha_{\mathcal{P}}) \hat{V}_i(f_*)$. Therefore, $\hat{V}_i(f) = \frac{2 - \alpha_{\mathcal{P}}}{\alpha_{\mathcal{P}}}$, which proves that $\hat{V}_i(f) = \hat{V}_j(f)$ for all $i, j \in N$.

Case 3. $f_* \hat{\succeq}_i f$, for all $i \in N$.

This case is hold as Case 2 above.

We thus have $\hat{V}_i(f) = \hat{V}_j(f)$, for all $i, j \in N$ and $f \in \mathcal{A}^{cv}$.

Assume now that $\{\tilde{V}_i\}_{i \in N}$ is another set of affine functions representing $\{\hat{\succeq}_i\}_{i \in N}$ and satisfying $\tilde{V}_i(\mathcal{A}) = \tilde{V}_j(\mathcal{A})$ for all $i, j \in N$. This implies that $\tilde{V}_i(f^*) = \tilde{V}_j(f^*)$ and $\tilde{V}_i(f_*) = \tilde{V}_j(f_*)$ for all $i, j \in N$. Therefore, by the same argument as above, we must have $\tilde{V}_i(f) = \tilde{V}_j(f)$ for all $i, j \in N$, and all $f \in \mathcal{A}^{cv}$. Because $\tilde{V}_i$ are unique up to an affine transformation, there must exist $a_1, \cdots, a_n > 0$ and $b_1, \cdots, b_n \in \mathbb{R}$ such that for all i, $\tilde{V}_i = a_i \hat{V}_i + b_i$. But as we have shown, one must also have: $\tilde{V}_i(f) = \tilde{V}_j(f)$ for all $f \in \mathcal{A}^{cv}$. Hence, $a_i \hat{V}_i(f) + b_i = a_j \hat{V}_j(f) + b_j$ for all $i, j \in N$, $f \in \mathcal{A}^{cv}$. Since $\hat{V}_i(f) = \hat{V}_j(f)$ for all $f \in \mathcal{A}^{cv}$ and all $i, j \in N$, and $\hat{V}_i$ is not constant on $\mathcal{A}^{cv}$, this implies $a_i = a_j$ and $b_i = b_j$.

Let $F : \mathcal{A} \to \mathbb{R}^{n+1}$ be defined by $F(f) = (V_{\mathcal{P}}(f), \hat{V}_1(f), \cdots, \hat{V}_n(f))$. Let $K = F(\mathcal{A}_h)$. By Claims 1, 3 and 6, K is convex. Therefore, by Claims 4 and 5, we can apply proposition 2 in De Meyer and Mongin (1995) : there exist non-negative numbers $\lambda_1(h, \mathcal{P}), \cdots, \lambda_n(h, \mathcal{P})$, and a real number $\mu(h, \mathcal{P})$ such that for all $f \in \mathcal{A}_h$,

$$V_{\mathcal{P}}(f) = \sum_{i \in N} \lambda_i(h, \mathcal{P}) \hat{V}_i(f) + \mu(h, \mathcal{P}).$$

We hence have:

$$\begin{cases} V_{\mathcal{P}}(f^*) = 1 = \sum_{i \in N} \lambda_i(h, \mathcal{P}) + \mu(h, \mathcal{P}) \\ V_{\mathcal{P}}(f_*) = -1 = -\sum_{i \in N} \lambda_i(h, \mathcal{P}) + \mu(h, \mathcal{P}). \end{cases}$$

Therefore, $\sum_i \lambda_i(h, \mathcal{P}) + \mu(h, \mathcal{P}) = \sum_{i \in N} \lambda_i(h, \mathcal{P}) - \mu(h, \mathcal{P})$, which implies $\mu(h, \mathcal{P}) = 0$, and therefore, $\sum_{i \in N} \lambda_i(h, \mathcal{P}) = 1$.

□

Claim 8. *For all $\mathcal{P} \in \mathbb{P}$, there exists a unique compact and convex set $\mathcal{F}(\mathcal{P}) \in \mathbb{P}$ such that, for all $f, g \in \mathcal{A}$,*

$$(f, \mathcal{P}) \succeq (g, \mathcal{P}) \Leftrightarrow \min_{p \in \mathcal{F}(\mathcal{P})} \sum_i p(i) \hat{V}_i(f(i)) \geq \min_{p \in \mathcal{F}(\mathcal{P})} \sum_i p(i) \hat{V}_i(f(i)),$$

where the $\hat{V}_i$ are affine representations of $\hat{\succeq}_i$ such that $\hat{V}_i(\mathcal{A}) = \hat{V}_j(\mathcal{A})$ for all $i, j \in N$. Moreover, $\{\hat{V}_i\}_{i \in N}$ are unique up to a common positive affine transformation.

Proof. Let $\mathcal{P} \in \mathbb{P}$ be fixed, and $V_{\mathcal{P}}$, $\hat{V}_i$ $(i \in N)$ be defined as in Claim 7.

Let $\tilde{B} = \{(\hat{V}_1(f), \cdots, \hat{V}_n(f)) | f \in \mathcal{A}\}$, and $B(N, V_{\mathcal{P}}(\mathcal{A}))$ be the set of functions from N to $V_{\mathcal{P}}(\mathcal{A})$. One has obviously $\tilde{B} \subseteq B(N, V_{\mathcal{P}}(\mathcal{A}))$. We show now that $B(N, V_{\mathcal{P}}) \subseteq \tilde{B}$ is also true. Let $\varphi : N \to V_{\mathcal{P}}(\mathcal{A})$. By Claim 6, $V_{\mathcal{P}}(\mathcal{A}) = V_{\mathcal{P}}(\mathcal{A}^{cv})$. Therefore, for all $i \in N$, there exists $h_i \in \mathcal{A}^{cv}$ such that $\varphi(i) = V_{\mathcal{P}}(h_i)$. Let $f \in \mathcal{A}$ be defined by $f(i) = h_i(i)$ for all $i \in N$. We then have $\hat{V}_i(f) = \hat{V}_i(h_i)$ and, since $h_i \in \mathcal{A}^{cv}$, $\hat{V}_i(h_i) = V_{\mathcal{P}}(h_i)$. Therefore, $(\hat{V}_1(f), \cdots, \hat{V}_n(f)) = \varphi$.

Let $I : B(N, V_{\mathcal{P}}(\mathcal{A})) \to \mathbb{R}$ be defined by:

$$I(\varphi) = V_{\mathcal{P}}(f), \text{ if } \varphi = (\hat{V}_1(f), \cdots, \hat{V}_n(f)).$$

By Claim 4, I is monotone, and by Claim 7, $I(0) = 0$ and $I(1) = 1$. and I is $V_{\mathcal{P}}(\mathcal{A}^{cv})$-affine. Furthermore, Axiom 10 implies that I is concave. Therefore, its positive homogeneous extension J to $B(N)$, the set of all functions from N to $\mathbb{R}$ is monotone, superlinear, and such that $J(\varphi + k) = J(\varphi) + k$ for all $k \in \mathbb{R}$. Therefore, by a classical result (see, e.g., the "Fundamental Lemma" in Chateauneuf (1991), and Lemma 3.5 in Gilboa and Schmeidler (1989)), there exists a unique compact convex set $\mathcal{F}(\mathcal{P}) \in \mathbb{P}$ such that:

$$J(\varphi) = \min_{p \in \mathcal{F}(\mathcal{P})} \sum_{i \in N} p(i)\varphi(i).$$

Therefore, for all $f \in \mathcal{A}$,

$$(f, \mathcal{P}) \succeq (g, \mathcal{P}) \Leftrightarrow \min_{p \in \mathcal{F}(\mathcal{P})} p(i)\hat{V}_i(f) \geq \min_{p \in \mathcal{F}(\mathcal{P})} p(i)\hat{V}_i(g)$$

which is equivalent to

$$(f, \mathcal{P}) \succeq (g, \mathcal{P}) \Leftrightarrow \min_{p \in \mathcal{F}(\mathcal{P})} p(i)\hat{V}_i(f(i)) \geq \min_{p \in \mathcal{F}(\mathcal{P})} p(i)\hat{V}_i(g(i)).$$

Finally, that $\{\hat{V}_i\}_{i \in N}$ are unique up to a common positive affine transformation follows from Claim 7.

□

Claim 9. *For all $\mathcal{P}, \mathcal{Q} \in \mathbb{P}$, there exist unique compact and convex sets $\mathcal{F}(\mathcal{P}), \mathcal{F}(\mathcal{Q}) \in \mathbb{P}$ such that, for all $f, g \in \mathcal{A}$,*

$$(f, \mathcal{P}) \succeq (g, \mathcal{Q}) \Leftrightarrow \min_{p \in \mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(f(i)) \geq \min_{p \in \mathcal{F}(\mathcal{Q})} \sum_i p(i)\hat{V}_i(f(i)),$$

where the $\hat{V}_i$ are affine representations of $\hat{\succeq}_i$ such that $\hat{V}_i(\mathcal{A}) = \hat{V}_j(\mathcal{A})$ for all $i, j \in N$. Moreover, $\{\hat{V}_i\}_{i \in N}$ are unique up to a common positive affine transformation.

Proof. Let $f, g \in \mathcal{A}$ and $\mathcal{P}, \mathcal{Q} \in \mathbb{P}$. Let $\mathcal{F}(\mathcal{P})$ and $\mathcal{F}(\mathcal{Q})$ defined as in Claim 7. Assume that $(f, \mathcal{P}) \succeq (g, \mathcal{Q})$. By Axiom 6, there exist $\bar{f}, \underline{f}, \bar{g}, \underline{g} \in \mathcal{A}^{cv}$ such that:

$$\begin{cases} (\bar{f}, \mathcal{P}) \succeq (f, \mathcal{P}) \succeq (\underline{f}, \mathcal{P}) \\ (\bar{g}, \mathcal{Q}) \succeq (g, \mathcal{Q}) \succeq (\underline{g}, \mathcal{Q}). \end{cases}$$

Without loss of generality, assume that $(\bar{f}, \{\delta_j\}) \succeq (\bar{g}, \{\delta_j\})$ and $(\bar{f}, \{\delta_k\}) \succeq (\bar{g}, \{\delta_k\})$ for some $j, k \in N$. Then, by Claim 4, $(\bar{f}, \mathcal{Q}) \succeq (\bar{g}, \mathcal{Q})$ and $(\underline{f}, \mathcal{P}) \succeq (\underline{g}, \mathcal{P})$. Therefore,

$$\begin{cases} (\bar{f}, \mathcal{P}) \succeq (f, \mathcal{P}) \succeq (\underline{g}, \mathcal{P}) \\ (\bar{f}, \mathcal{Q}) \succeq (g, \mathcal{Q}) \succeq (\underline{g}, \mathcal{Q}). \end{cases}$$

Hence, by Axioms 1 and 3, there exist $\lambda, \mu \in [0,1]$ such that $(f, \mathcal{P}) \sim (\lambda \bar{f} + (1-\lambda)\underline{g}, \mathcal{P})$ and $(g, \mathcal{Q}) \sim (\mu \bar{f} + (1-\mu)\underline{g}, \mathcal{Q})$. Hence $(\lambda \bar{f} + (1-\lambda)\underline{g}, \mathcal{P}) \succeq (\mu \bar{f} + (1-\mu)\underline{g}, \mathcal{Q})$. Observe that by Claim 1, $\lambda \bar{f} + (1-\lambda)\underline{g} \in \mathcal{A}^{cv}$ and $\mu \bar{f} + (1-\mu)\underline{g} \in \mathcal{A}^{cv}$. Therefore, by Axiom 8, $(\lambda \bar{f} + (1-\lambda)\underline{g}, \{\delta_1\}) \succeq (\mu \bar{f} + (1-\mu)\underline{g}, \{\delta_1\})$, and therefore, $\hat{V}_1(\lambda \bar{f}(1) + (1-\lambda)\underline{g}(1)) \geq \hat{V}_1(\mu \bar{f}(1) + (1-\mu)\underline{g}(1))$.

Since $(\lambda \bar{f} + (1-\lambda)\underline{g}, \mathcal{P}) \sim (f, \mathcal{P})$, we have by Claim 8:

$$\begin{aligned} \min_{p \in \mathcal{F}(\mathcal{P})} \sum_i p(i) \hat{V}_i(f(i)) &= \min_{p \in \mathcal{F}(\mathcal{P})} \sum_i p(i) \hat{V}_i(\lambda \bar{f}(i) + (1-\lambda)\underline{g}(i)) \\ &= \hat{V}_1(\lambda \bar{f}(1) + (1-\lambda)\underline{g}(1)). \end{aligned}$$

Similarly,

$$\begin{aligned} \min_{p \in \mathcal{F}(\mathcal{Q})} \sum_i p(i) \hat{V}_i(g(i)) &= \min_{p \in \mathcal{F}(\mathcal{P})} \sum_i p(i) \hat{V}_i(\mu \bar{f}(i) + (1-\mu)\underline{g}(i)) \\ &= \hat{V}_1(\mu \bar{f}(1) + (1-\mu)\underline{g}(1)). \end{aligned}$$

Since $\hat{V}_1(\lambda \bar{f}(1) + (1-\lambda)\underline{g}(1)) \geq \hat{V}_1(\mu \bar{f}(1) + (1-\mu)\underline{g}(1))$, we finally obtain:

$$\min_{p \in \mathcal{F}(\mathcal{P})} \sum_i p(i) \hat{V}_i(f(i)) \geq \min_{p \in \mathcal{F}(\mathcal{Q})} \sum_i p(i) \hat{V}_i(g(i)),$$

the desired result. $\square$

Hence, there exists a unique function $\mathcal{F} : \mathbb{P} \to \mathbb{P}_C$, such that $\mathcal{F}(\mathcal{P})$ is compact and convex for all set $\mathcal{P} \in \mathbb{P}$, such that for all $f, g \in \mathcal{A}$, $\mathcal{P}, \mathcal{Q} \in \mathbb{P}$,

$$(f, \mathcal{P}) \succeq (g, \mathcal{Q}) \Leftrightarrow \min_{p \in \mathcal{F}(\mathcal{P})} \sum_i p(i) \hat{V}_i(f(i)) \geq \min_{p \in \mathcal{F}(\mathcal{Q})} \sum_i p(i) \hat{V}_i(f(i)),$$

where the $\hat{V}_i$ are affine representations of $\hat{\succeq}_i$ such that $\hat{V}_i(\mathcal{A}) = \hat{V}_j(\mathcal{A})$, for all $i, j \in N$. Moreover, $\{\hat{V}_i\}_{i \in N}$ are unique up to a common positive affine transformation.

It remains to show that $\mathcal{F}(\mathcal{P}) \subseteq co(\mathcal{P})$ for all $\mathcal{P} \in \mathbb{P}$, and that for all $\alpha \in (0,1)$, all $\mathcal{P}, \mathcal{Q} \in \mathbb{P}$, $\mathcal{F}(\alpha \mathcal{P} + (1-\alpha)\mathcal{Q}) = \alpha \mathcal{F}(\mathcal{P}) + (1-\alpha)\mathcal{F}(\mathcal{Q})$. This is done in the two following claims.

Claim 10. *For all* $\mathcal{P} \in \mathbb{P}$, $\mathcal{F}(\mathcal{P}) \subseteq co(\mathcal{P})$.

Proof. Assume there exists $\mathcal{P} \in \mathbb{P}$ such that $\mathcal{F}(\mathcal{P}) \nsubseteq co(\mathcal{P})$. Then, there exists $p^* \in \mathcal{F}(\mathcal{P})$ such that $p^* \notin co(\mathcal{P})$. Since $\mathcal{F}(\mathcal{P})$ and $co(\mathcal{P})$ are convex sets, a separation argument implies that there exists a function $\phi : N \to \mathbb{R}$ such that $\sum_i p^*(i)\phi(i) < min_{p \in \mathcal{F}(\mathcal{P})} \sum_i p(i)\phi(i)$. There exist numbers a, b with $a > 0$ such that $a\phi(i) + b \in \hat{V}_i(\mathcal{A})$ for all i (choosing a sufficiently close to zero will ensure that $a_j\phi_j(i) + b_j \in [-1, 1]$ for all $i \in N$). Hence, for all i, there exists $y_i \in Y$ such that $a\phi(i) + b = \hat{V}_i(y_i)$. Define f by $f(i) = y_i$ for all $i \in N$. Note that $\min_{p \in co(\mathcal{P})} \sum_i p(i) \hat{V}_i(f(i)) \in$

$\hat{V}_i(\mathcal{A}^{cv})$, and therefore there exists $h \in \mathcal{A}^{cv}$ such that $\hat{V}_i(h(i)) = \min_{p\in co(\mathcal{P})} \sum_i p(i)\hat{V}_i(f(i))$ for all $i \in N$. We have, for all $p \in \mathcal{P}$:

$$\sum_i p(i)\hat{V}_i(f(i)) \geq \min_{p\in co(\mathcal{P})} \sum_i p(i)\hat{V}_i(f(i)) = \min_{p\in co(\mathcal{P})} \sum_i p(i)(a\phi(i)+b) = \hat{V}_i(h(i))$$

Therefore, for all $p \in \mathcal{P}$, $(f,\{p\}) \succeq (h,\{p\})$. But:

$$\begin{aligned}
\min_{p\in\mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(f(i)) &\leq \sum p^*(i)\hat{V}_i(f(i)) \\
&= \sum_i p^*(i)(a\phi(i)+b) < \min_{p\in co(\mathcal{P})} \sum_i p(i)(a\phi_i+b) = \hat{V}_i(h(i)) \\
&= \min_{p\in\mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(h(i)).
\end{aligned}$$

Hence, $(h,\mathcal{P}) \succ (f,\mathcal{P})$, which is a violation of Axiom 5. $\square$

Claim 11. *For all* $\mathcal{P},\mathcal{Q} \in \mathbb{P}$, *all* $\alpha \in (0,1)$, $\mathcal{F}(\alpha\mathcal{P}+(1-\alpha)\mathcal{Q}) = \alpha\mathcal{F}(\mathcal{P}) + (1-\alpha)\mathcal{F}(\mathcal{Q})$.

Proof. Let $\mathcal{P},\mathcal{Q} \in \mathbb{P}$, and $\alpha \in (0,1)$. For all $f \in \mathcal{A}$, Let $p^*(f) \in \arg\min_{p\in\mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(f(i))$, and $q^*(f) \in \arg\min_{p\in\mathcal{F}(\mathcal{Q})} \sum_i p(i)\hat{V}_i(f(i))$ (since by Claim 9, $\mathcal{F}(\mathcal{P})$ and $\mathcal{F}(\mathcal{Q})$ are compact, $p^*(f)$ and $q^*(f)$ are well-defined). By Claims 9 and 10, $(f,\mathcal{P}) \sim (f,\{p^*(f)\})$ and $(f,\mathcal{Q}) \sim (f,\{q^*(f)\})$. By Axiom 4, this implies:

$$(f,\alpha\mathcal{P}+(1-\alpha)\mathcal{Q}) \sim (f,\alpha\{p^*(f)\}+(1-\alpha)\{q^*(f)\}).$$

Hence, by Claims 9 and 10:

$$\begin{aligned}
\min_{p\in\mathcal{F}(\alpha\mathcal{P}+(1-\alpha)\mathcal{Q})} \sum_i p(i)\hat{V}_i(f(i)) &= \sum_i (\alpha p^*(f)(i) + (1-\alpha)q^*(f)(i))\hat{V}_i(f(i)) \\
&= \alpha \sum_i p^*(f)(i)\hat{V}_i(f(i)) + (1-\alpha)\sum_i q^*(f)(i)\hat{V}_i(f(i)) \\
&= \alpha \min_{p\in\mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(f(i)) + (1-\alpha) \min_{p\in\mathcal{F}(\mathcal{Q})} \sum_i p(i)\hat{V}_i(f(i)) \\
&= \min_{p\in\alpha\mathcal{F}(\mathcal{P})+(1-\alpha)\mathcal{F}(\mathcal{Q})} \sum p(i)\hat{V}_i(f(i)).
\end{aligned}$$

Since this holds for all $f \in \mathcal{F}$, unicity of $\mathcal{F}$ (see Claim 9) implies that $\mathcal{F}(\alpha\mathcal{P}+(1-\alpha)\mathcal{Q}) = \alpha\mathcal{F}(\mathcal{P}) + (1-\alpha)\mathcal{F}(\mathcal{Q})$, the desired result. $\square$

Proof of Theorem 2

The necessity part of the Theorem is easily checked. We therefore only prove the sufficiency part.

We will use the following notations. For all subset T of N, let $\Delta(T)$ be the simplex over T. Let $c_T \in \Delta(T)$ be defined by $c_T(s) = \frac{1}{|T|}$ for all $s \in T$. Finally, let $\mathcal{H} : \mathbb{P} \times \Delta(N) \times [0,1] \to \mathbb{P}$ be defined by :

$$\mathcal{H}(\mathcal{Q}, c, \theta) = \{p \in \Delta(N) \,|\exists q \in \mathcal{Q} \text{ s.t. } p = \theta q + (1-\theta)c\}.$$

By Theorem 1, there exist affine functions $\hat{V}_i : Y \to \mathbb{R}$, $i \in N$, representing $\hat{\succeq}_i$ and such that $\hat{V}_i(\mathcal{A}) = \hat{V}_j(\mathcal{A})$, for all $i, j \in N$, and a function $\mathcal{F} : \mathbb{B} \to \mathbb{B}$ satisfying for all $\mathcal{P}, \mathcal{Q} \in \mathbb{B}$,

1. $\mathcal{F}(\mathcal{P}) \subseteq co(\mathcal{P})$
2. For all $\alpha \in [0,1]$, $\mathcal{F}(\alpha\mathcal{P} + (1-\alpha)\mathcal{Q}) = \alpha\mathcal{F}(\mathcal{P}) + (1-\alpha)\mathcal{F}(\mathcal{Q})$

such that for all $\mathcal{P}, \mathcal{Q} \in \mathbb{B}$, and all $f, g \in \mathcal{A}$, $(f, \mathcal{P}) \succeq (g, \mathcal{Q})$ if, and only if:

$$\min_{p \in \mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(f(i)) \geq \min_{p \in \mathcal{F}(\mathcal{Q})} \sum_i \hat{V}_i(g(i)).$$

Furthermore, $\mathcal{F}$ is unique, and $\{\hat{V}_i\}_{i \in N}$ are unique up to a common positive affine transformation.

As in the proof of Theorem 1, we will use the following slight abuse of notation: for all $i \in N$ and $f \in \mathcal{A}$ we define $\hat{V}_i(f)$ as $\hat{V}_i(f(i))$.

The proof goes through several claims. Although not explicitly stated in the claims, all the assumptions of Theorem 2 are made throughout this subsection.

Claim 12. *For all $\mathcal{P} \in \mathbb{B}$, and all bijection $\varphi : N \to N$, $\mathcal{F}(\mathcal{P}^\varphi) = (\mathcal{F}(\mathcal{P}))^\varphi$.*

Proof. Let $\mathcal{P} \in \mathbb{B}$, and φ be a permutation on N. We will prove that $\mathcal{F}(\mathcal{P})^\varphi \subseteq \mathcal{F}(\mathcal{P}^\varphi)$. Assume that such is not the case, i.e., there exists $p^* \in \mathcal{F}(\mathcal{P})^\varphi$ such that $p^* \notin \mathcal{F}(\mathcal{P}^\varphi)$. Since $\mathcal{F}(\mathcal{P}^\varphi)$ and $\mathcal{F}(\mathcal{P})^\varphi$ are convex sets, a standard separation argument implies that there exists a function $\phi : N \to \mathbb{R}$ such that: $\sum_i p^*(i)\phi(i) < \min_{p \in \mathcal{F}(\mathcal{P}^\varphi)} \sum_i p(i)\phi(i)$. There exist numbers a, b with $a > 0$ such that $a\phi(i) + b \in \hat{V}_i(\mathcal{A})$ for all i. Hence, for all i, there exists $y_i \in Y$ such that $a\phi(i) + b = \hat{V}_i(y_i)$ (choosing a sufficiently close to zero will ensure that $a_j\phi_j(i) + b_j \in [-1, 1]$ for all $i \in N$). Define f by $f(i) = y_i$ for all $i \in N$. We then have: $\sum_i p^*(i)\hat{V}_i(f) < \min_{p \in \mathcal{F}(\mathcal{P}^\varphi)} \sum_i p(i)\hat{V}_i(f)$. Axiom 6 and Claim 1 in the proof of Theorem 1 imply that $\mathcal{A}(h^\psi) \neq \emptyset$, for all permutation $\psi : N \to N$, and all $h \in \mathcal{A}$. Let $g \in \mathcal{A}(f^{\varphi^{-1}})$. For all $p \in \mathcal{F}(\mathcal{P})$, $\sum_i p(i)\hat{V}_i(g) = \sum_i p^\varphi(i)\hat{V}_i(f)$. Therefore, $\min_{p \in \mathcal{F}(\mathcal{P})^\varphi} \sum_i p(i)\hat{V}_i(f) = \min_{p \in \mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(g)$. Therefore,

$$\min_{p \in \mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(g) = \min_{p \in \mathcal{F}(\mathcal{P})^\varphi} \sum_i p(i)\hat{V}_i(f) \leq \sum_i p^*(i)\hat{V}_i(f) < \min_{p \in \mathcal{F}(\mathcal{P}^\varphi)} \sum_i p(i)\hat{V}_i(f).$$

Therefore, $(f, \mathcal{P}^\varphi) \succ (g, \mathcal{P})$, a contradiction with Axiom 11. The inclusion $\mathcal{F}(\mathcal{P}^\varphi) \subseteq \mathcal{F}(\mathcal{P})^\varphi$ can be proved using a similar argument. Therefore, $\mathcal{F}(\mathcal{P}^\varphi) = \mathcal{F}(\mathcal{P})^\varphi$. $\square$

Claim 13. *Let $\mathcal{P} \in \mathbb{B}$, I a subset of N, and $f_k \in \mathcal{A}$ ($k \in I$). Then:*

$$\bigcap_{k\in I}\left(\arg\min_{p\in\mathcal{P}}\sum_i p(i)\hat{V}_i(f_k)\right) \neq \emptyset \Rightarrow \bigcap_{k\in I}\left(\arg\min_{p\in\mathcal{F}(\mathcal{P})}\sum_i p(i)\hat{V}_i(f_k)\right) \neq \emptyset.$$

Proof. Let $\mathcal{P} \in \mathbb{B}$, $I = \{1, \cdots, m\}$, and $f_k \in \mathcal{A}$ ($k \in I$) such that:

$$\bigcap_{k\in I}\left(\arg\min_{p\in\mathcal{P}}\sum_i \hat{V}_i(f_k)\right) \neq \emptyset.$$

Assume for instance that $(f_1, \mathcal{P}) \succeq (f_k, \mathcal{P})$, for all $k \in I$. By Axiom 6, there exists $\bar{f} \in \mathcal{A}^{cv}$ such that $(\bar{f}, \mathcal{P}) \succeq (f_1, \mathcal{P})$. Hence, for all $k \in I \setminus \{1\}$, $(\bar{f}, \mathcal{P}) \succeq (f_1, \mathcal{P}) \succeq (f_k, \mathcal{P})$. Hence, by Axioms 1 and 3, for all $k \neq 1$, there exist $\alpha_k \in [0, 1]$ such that $(\alpha_k \bar{f} + (1-\alpha_k) f_k, \mathcal{P}) \sim (f_1, \mathcal{P})$. But observe that since $\bar{f} \in \mathcal{A}^{cv}$, $\arg\min_{p\in\mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(\alpha_k \bar{f} + (1-\alpha_k) f_k) = \arg\min_{p\in\mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(f_k)$. Therefore, there is no loss in generality if we assume $(f_1, \mathcal{P}) \sim (f_k, \mathcal{P})$, for all $k \in I$, an assumption we maintain through this proof.

We now proceed by induction. Let

$$\mathbf{P}(r) = \left\{\bigcap_{k\in I}\left(arg\min_{p\in\mathcal{P}}\sum_i p(i)\hat{V}_i(f_k)\right) \neq \emptyset \Rightarrow \bigcap_{k\in\{1,\cdots,r\}}\left(\arg\min_{p\in\mathcal{F}(\mathcal{P})}\sum_i p(i)\hat{V}_i(f_k)\right) \neq \emptyset\right\}.$$

We first prove that $\mathbf{P}(2)$ is true. Assume that:

$$\left(\arg\min_{p\in\mathcal{F}(\mathcal{P})}\sum_i p(i)\hat{V}_i(f_1)\right) \cap \left(\arg\min_{p\in\mathcal{F}(\mathcal{P})}\sum_i \hat{V}_i(f_2)\right) = \emptyset.$$

Let $p^* \in \arg\min_{p\in\mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(f_1)$, and $\hat{p} \in \arg\min_{p\in\mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(f_2)$. By assumption,

$$\sum_i \hat{p}(i)\hat{V}_i(f_2) < \sum_i p^*(i)\hat{V}_i(f_2),$$

and:

$$\sum_i p^*(i)\hat{V}_i(f_1) = \sum_i \hat{p}(i)\hat{V}_i(f_2).$$

By Axiom 12, for all $\alpha \in (0, 1)$, $(\alpha f_1 + (1-\alpha) f_2, \mathcal{P}) \sim (f_1, \mathcal{P})$. Therefore:

$$\min_{p\in\mathcal{F}(\mathcal{P})}\sum_i p(i)\hat{V}_i(\alpha f_1 + (1-\alpha) f_2) = \min_{p\in\mathcal{F}(\mathcal{P})}\sum_i p(i)\hat{V}_i(f_1) = \sum_i p^*(i)\hat{V}_i(f_1).$$

Therefore there exists $\tilde{p} \in \mathcal{F}(\mathcal{P})$ such that:

$$\sum_i \tilde{p}(i)\hat{V}_i(\alpha f_1 + (1-\alpha) f_2) = \sum_i p^*(i)\hat{V}_i(f_1),$$

i.e.,

$$\alpha \sum_i \tilde{p}(i)\hat{V}_i(f_1) + (1-\alpha)\sum_i \tilde{p}(i)\hat{V}_i(f_2) = \sum_i p^*(i)\hat{V}_i(f_1).$$

Therefore, since $\sum_i p^*(i)\hat{V}_i(f_1) = \sum_i \hat{p}(i)\hat{V}_i(f_2)$:

$$\alpha \sum_i \tilde{p}(i)\hat{V}_i(f_1) + (1-\alpha)\sum_i \tilde{p}(i)\hat{V}_i(f_2) = \alpha\sum_i p^*(i)\hat{V}_i(f_1) + (1-\alpha)\sum_i \hat{p}(i)\hat{V}_i(f_2),$$

and therefore $\tilde{p} \in \left(\arg\min_{p\in\mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(f_1)\right) \cap \left(\arg\min_{p\in\mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(f_2)\right)$, a contradiction. Therefore, $\mathbf{P}(2)$ is true.

We now assume that $\mathbf{P}(r-1)$ is true, with $r-1<m$, and prove that then $\mathbf{P}(r)$ is also true. Assume that $\bigcap_{k\in\{1,\cdots,r\}} \left(\arg\min_{p\in\mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(f_k)\right) = \emptyset$. By the induction assumption, there exists $p^* \in \bigcap_{k\in\{1,\cdots,r-1\}} \left(\arg\min_{p\in\mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(f_k)\right)$, and for all $\hat{p} \in \left(\arg\min_{p\in\mathcal{F}(\mathcal{P})} \sum_i p(i)\hat{V}_i(f_r)\right)$,

$$\sum_i \hat{p}(i)\hat{V}_i(f_r) < \sum_i \hat{p}(i)\hat{V}_i(f_r)$$

and

$$\sum_i \hat{p}(i)\hat{V}_i(f_r) = \sum_i \hat{p}(i)\hat{V}_i(f_k),\ \forall k\in\{1,\cdots,r-1\}.$$

Let $h = \sum_{k=1}^{r-1} \frac{1}{r-1} f_k$. We then have:

$$\begin{aligned}\arg\min_{p\in\mathcal{P}} \sum_i p(i)\hat{V}_i(h) &= \arg\min_{p\in\mathcal{P}} \sum_i p(i)\hat{V}_i\left(\sum_{i=1}^{r-1}\frac{1}{r-1}k_i\right) \\ &= \arg\min_{p\in\mathcal{P}} \sum_i p(i)\left(\sum_{i=1}^{r-1}\hat{V}_i(f_k)\right),\end{aligned}$$

the last equality being implied by the fact that the $\hat{V}_i$ are affine.

Let $p_1 \in \arg\min_{p\in\mathcal{P}} \sum_i p(i)\hat{V}_i(h)$, and assume that $p_1 \notin \bigcap_{k=1}^{r-1}\left(\arg\min_{p\in\mathcal{P}} \sum_i \hat{V}_i(f_k)\right)$. Then, for all $p_2 \in \bigcap_{k=1}^{r-1}\left(\arg\min_{p\in\mathcal{P}} \sum_i p(i)\hat{V}_i(f_k)\right)$, and all $k\in\{1,\cdots,r-1\}$, $\sum_i p_2(i)\hat{V}_i(f_k) \le \sum_i p_1(i)\hat{V}_i(f_k)$, with a strict inequality for some $j\in\{1,\cdots,r-1\}$. But, then, $\sum_i p_2(i)\hat{V}_i(h) < \sum_i p_1(i)\hat{V}_i(h)$, a contradiction with the fact that $p_1 \in \arg\min_{p\in\mathcal{P}} \sum_i p(i)\hat{V}_i(h)$. Therefore, $\arg\min_{p\in\mathcal{P}} \sum_i p(i)\hat{V}_i(h) \subseteq \bigcap_{i=1}^{r-1}\left(\arg\min_{p\in\mathcal{P}} \sum_i p(i)\hat{V}_i(f_k)\right)$. Conversely, let:

$$q_1 \in \bigcap_{i=1}^{r-1}\left(\arg\min_{p\in\mathcal{P}} \sum_i p(i)\hat{V}_i(f_k)\right).$$

Then, by definition, $q_1 \in \arg\min_{p\in\mathcal{P}} \sum_i p(i)\hat{V}_i(f_k)$, for all $k\in\{1,\cdots,r-1\}$. Therefore, $q_1 \in \arg\min_{\mathcal{P}} \sum_i p(i)\left(\sum_{k=1}^{r-1}\hat{V}_i(f_k)\right) = \arg\min_{p\in\mathcal{P}} \sum_i p(i)\hat{V}_i(h)$. Therefore,

$$\bigcap_{k=1}^{r-1}\left(\arg\min_{p\in\mathcal{P}} \sum_i p(i)\hat{V}_i(f_k)\right) \subseteq \arg\min_{p\in\mathcal{P}} \sum_i p(i)\hat{V}_i(h),$$

which implies $\bigcap_{k=1}^{r-1}\left(\arg\min_{p\in\mathcal{P}}\sum_i p(i)\hat{V}_i(f_k)\right) = \arg\min_{p\in\mathcal{P}}\sum_i p(i)\hat{V}_i(h)$.

Hence, $\left(\arg\min_{p\in\mathcal{P}}\sum_i p(i)\hat{V}_i(h)\right)\cap\left(\arg\min_{p\in\mathcal{P}}\sum_i p(i)\hat{V}_i(f_r)\right)\neq\emptyset$. Furthermore, by assumption, $p^*\in\arg\min_{p\in\mathcal{F}(\mathcal{P})}\sum_i p(i)\hat{V}_i(h) = \arg\min_{p\in\mathcal{P}}\sum_i p(i)\left(\sum_{k=1}^{r-1}\frac{1}{r-1}\hat{V}_i(f_k)\right)$.

Therefore, Axiom 12 implies for all $\alpha\in[0,1]$, $(\alpha h+(1-\alpha)f_r,\mathcal{P})\sim(h,\mathcal{P})$. Therefore,

$$\min_{p\in\mathcal{F}(\mathcal{P})}\sum_i p(i)\hat{V}_i(\alpha h+(1-\alpha)f_r) = \min_{p\in\mathcal{F}(\mathcal{P})}\sum_i p(i)\hat{V}_i(h) = \sum_i p^*(i)\hat{V}_i(h).$$

Hence, there exists $\tilde{p}\in\mathcal{F}(\mathcal{P})$ such that:

$$\sum_i \tilde{p}(i)\hat{V}_i(\alpha h+(1-\alpha)f_r) = \sum_i p^*(i)\hat{V}_i(h),$$

i.e., since $\sum p^*(i)\hat{V}_i(h) = \sum_i \hat{p}(i)\hat{V}_i(f_r)$:

$$\alpha\sum_i \tilde{p}(i)\hat{V}_i(h) + (1-\alpha)\sum_i \tilde{p}(i)\hat{V}_i(h) = \alpha\sum_i p^*(i)\hat{V}_i(h) + (1-\alpha)\sum_i \hat{p}(i)\hat{V}_i(f_r),$$

which implies that $\tilde{p}\in\left(\arg\min_{p\in\mathcal{P}}\sum_i p(i)\hat{V}_i(h)\right)\cap\left(\arg\min_{p\in\mathcal{P}}\sum_i p(i)\hat{V}_i(f_r)\right)$. But:

$$\left(\arg\min_{p\in\mathcal{P}}\sum_i p(i)\hat{V}_i(h)\right)\subseteq\bigcap_{k\in\{1,\cdots,r-1\}}\left(\arg\min_{p\in\mathcal{F}(\mathcal{P})}\sum_i p(i)\hat{V}_i(f_k)\right).$$

Therefore, $\tilde{p}\in\bigcap_{l\in\{1,\cdots,r\}}\left(\arg\min_{p\in\mathcal{F}(\mathcal{P})}\sum_i p(i)\hat{V}_i(f_k)\right)$, a contradiction. □

Claim 14. *Let $S_1,S_2\subset N$ such that $S_1\neq\emptyset$, $S_2\neq\emptyset$. Then, for all $\alpha\in[0,1]$, there exists $\theta\in[0,1]$ such that $\mathcal{F}(\alpha\Delta(S_1)+(1-\alpha)\Delta(S_2)) = \mathcal{H}(\alpha\Delta(S_1)+(1-\alpha)\Delta(S_2),\alpha c_{S_1}+(1-\alpha)c_{S_2},\theta)$.*

Proof. Let $\alpha\in[0,1]$, and $S_1,S_2\subset N$ such that $S_1\neq\emptyset$, $S_2\neq\emptyset$. To simplify notation, let $c_1=c_{S_1}$ and $c_2=c_{S_2}$, $\Delta_1=\Delta(S_1)$, $\Delta_2=\Delta(S_2)$ and $\Delta=\alpha\Delta_1+(1-\alpha)\Delta_2$. We first consider the case $|S_1|>1$ and $|S_2|>1$.

We know by Theorem 1 that $\mathcal{F}(\Delta)=\alpha\mathcal{F}(\Delta_1)+(1-\alpha)\mathcal{F}(\Delta_2)$. On the other hand, since $\Delta_i^\varphi=\Delta_i$, for all permutation $\varphi: N\to N$ such that $\varphi(S_i)=S_i$, Claim 12 implies that $\mathcal{F}(\Delta_i)^\varphi=\mathcal{F}(\Delta_i)$ for all such permutations. Hence, $c_i\in\mathcal{F}(\Delta_i)$, and therefore, $\alpha c_1+(1-\alpha)c_2\in\alpha\mathcal{F}(\Delta_1)+(1-\alpha)\mathcal{F}(\Delta_2)=\mathcal{F}(\Delta)$. Let $c=\alpha c_1+(1-\alpha)c_2$

Since Δ is a polytope, it has a finite number of facets and vertices. Let $\{F_1,\cdots,F_K\}$ be the set of all facets of Δ, and $\Pi=\{\pi_1^*,\cdots,\pi_M^*\}$ be the set of its vertices. Furthermore, for all $m\in\{1,\cdots,M\}$, let J_m be a subset of $\{1,\cdots,K\}$ such that $\cap_{j\in J_m}F_j=\{\pi_m^*\}$.

For each facet F_j, let q_j be in the (relative) interior of F_j. Let Ψ_j be the (unique) hyperplane supporting Δ at q_j, defined by $\Psi_j=\{p\,|\,\phi_j(p)=\mu_j\}$, with $\mu_j\in\mathbb{R}$, and ϕ_j a linear function (observe that $F_j\subset\Psi_j$). Since Ψ_j is a separating hyperplane for Δ, ϕ_j can be chosen such that

$\phi_j(p) \geq \mu_j$ for all $p \in \Delta$. Let $\phi_j(p) = \sum_i \phi_j(i)p(i)$. There exist numbers a_j and b_j, with $a_j > 0$ such that $a_j\phi_j(i) + b_j \in \hat{V}_i(\mathcal{A})$ for all $i \in N$ (choosing a sufficiently close to zero will ensure that $a_j\phi_j(i) + b_j \in [-1, 1]$ for all $i \in N$). Therefore, there exist $y_i^j \in Y$ such that $a_j\phi_j(i) + b_j = \hat{V}_i(y_i^j)$ for all $i \in N$. Define f_j by $f_j(i) = y_i^j$ for all $i \in N$.

Observe that $\mathcal{H}(\Psi_j \cap \Delta, c, \theta)$ are the sets for which $\sum_i p(i)\hat{V}_i(f_j)$ is constant and smaller than $\sum_i c(i)\hat{V}_i(f_j)$. Furthermore, for all $p \in \mathcal{H}(\Psi_j \cap \Delta, c, \theta)$, $p' \in \mathcal{H}(\Psi_j, c, \theta')$, $\theta < \theta'$ if, and only if, $\sum_i p(i)\hat{V}_i(f_j) > \sum_i p'(i)\hat{V}_i(f_j)$. Therefore, it is the case that, for some $\hat{\theta}_j$,

$$\arg\min_{p \in \mathcal{F}(\Delta)} \sum_i p(i)\hat{V}_i(f) \subseteq \mathcal{H}(\Psi_j \cap \Delta, c, \hat{\theta}_j).$$

Now, assume that there are two facets F_ℓ and F_j such that $\hat{\theta}_\ell \neq \hat{\theta}_j$. Without loss of generality, we can take F_ℓ and F_j such that $F_\ell \cap F_j \neq \emptyset$, and $\hat{\theta}_\ell > \hat{\theta}_j$. Finally, consider $\pi \in F_\ell \cap F_j$. There exist numbers $\mu_1, \mu_2, \lambda_1, \lambda_2$ such that[13] $\mu_1\phi_\ell(i) + \lambda_1 \in \hat{V}_i(\mathcal{A})$ for all $i \in N$, $\mu_2\phi_j(i) + \lambda_2 \in \hat{V}_i(\mathcal{A})$ for all $i \in N$, and

$$\begin{cases} \sum_i \pi(i)(\mu_1\phi_\ell(i) + \lambda_1) = \sum_i \pi(i)(\mu_2\phi_j(i) + \lambda_2) \\ \sum_i c(i)(\mu_1\phi_\ell(i) + \lambda_1) = \sum_i c(i)(\mu_2\phi_j(i) + \lambda_2). \end{cases}$$

Therefore, there exit $\tilde{y}_i^j$ and $\tilde{y}_i^\ell$ in Y such that $\hat{V}_i(\tilde{y}_i^\ell) = \mu_1\phi_\ell(i) + \lambda_1$ and $\hat{V}_i(\tilde{y}_i^j) = \mu_2\phi_j(i) + \lambda_2$ for all $i \in N$. Let g_ℓ be defined by $g_\ell(i) = \tilde{y}_i^\ell$ for all i, and g_j be defined by $g_j(i) = \tilde{y}_i^j$ for all i. Observe that:

$$\arg\min_{p \in \Delta} \sum_i p(i)\hat{V}_i(g_j) = \arg\min_{p \in \Delta} \sum_i p(i)\hat{V}_i(f_j)$$

and

$$\arg\min_{p \in \Delta} \sum_i p(i)\hat{V}_i(g_\ell) = \arg\min_{p \in \Delta} \sum_i p(i)\hat{V}_i(f_\ell).$$

Therefore,

$$\arg\min_{p \in \Delta} \sum_i p(i)\hat{V}_i(g_j) \cap \arg\min_{p \in \Delta} \sum_i p(i)\hat{V}_i(g_\ell) \neq \emptyset.$$

Hence, by Axiom 12, $(g_j, \Delta) \sim (g_\ell, \Delta)$, i.e.,

$$\min_{p \in \mathcal{F}(\Delta)} \sum_i p(i)\hat{V}_i(g_j) = \min_{p \in \mathcal{F}(\Delta)} \sum_i p(i)\hat{V}_i(g_\ell).$$

But we also have

$$\min_{p \in \mathcal{F}(\Delta)} \sum_i p(i)\hat{V}_i(g_j) = \sum_i q_j(i)\hat{V}_i(g_j)$$

for any $q_j \in \mathcal{H}(\Psi_j \cap \Delta, c, \hat{\theta}_j)$, and similarly,

$$\min_{p \in \mathcal{F}(\Delta)} \sum_i p(i)\hat{V}_i(g_\ell) = \sum_i q_\ell(i)\hat{V}_i(g_\ell)$$

[13] Again, choosing these numbers sufficiently close to zero will ensure that $\mu_1\phi_\ell(i) + \lambda_1$ and $\mu_2\phi_j(i) + \lambda_2$ belong to $[-1, 1]$.

for any $q_\ell \in \mathcal{H}(\Psi_\ell \cap \Delta, c, \hat{\theta}_\ell)$. So, in particular, we have

$$\min_{p\in\mathcal{F}(\Delta)} \sum_i p(i)\hat{V}_i(g_j) = \sum_i \mathcal{H}(\pi, c, \hat{\theta}_j)(i)\hat{V}_i(g_j)$$

and

$$\min_{p\in\mathcal{F}(\Delta)} \sum_i p(i)\hat{V}_i(g_\ell) = \sum_i \mathcal{H}(\pi, c, \hat{\theta}_\ell)(i)\hat{V}_i(g_\ell).$$

Therefore, we must have:

$$\begin{aligned}
\sum_i \mathcal{H}(\pi, c, \hat{\theta}_j)(i)\hat{V}_i(g_j) &= \sum_i \mathcal{H}(\pi, c, \hat{\theta}_\ell)(i)\hat{V}_i(g_\ell) \\
\sum_i (\hat{\theta}_j\pi(i) + (1-\hat{\theta}_j)c(i))\hat{V}_i(g_j) &= \sum_i (\hat{\theta}_\ell\pi(i) + (1-\hat{\theta}_\ell)c(i))\hat{V}_i(g_\ell) \\
\hat{\theta}_j \sum_i \pi(i)\hat{V}_i(g_j) + (1-\hat{\theta}_j)\sum_i c(i)\hat{V}_i(g_j) &= \hat{\theta}_\ell \sum_i \pi(i)\hat{V}_i(g_\ell) + (1-\hat{\theta}_\ell)\sum_i c(i)\hat{V}_i(g_\ell).
\end{aligned}$$

But $\sum_i \pi(i)\hat{V}_i(g_j) = \sum_i \pi(i)\hat{V}_i(g_\ell)$, $\sum_i c(i)\hat{V}_i(g_j) = \sum_i c(i)\hat{V}_i(g_\ell)$ and $\sum_i \pi(i)\hat{V}_i(g_\ell) < \sum_i c(i)\hat{V}_i(g_\ell)$, $\sum_i \pi(i)\hat{V}_i(g_j) < \sum_i (i)\hat{V}_i(g_j)$. Therefore $\hat{\theta}_\ell > \hat{\theta}_j$ implies

$$\hat{\theta}_j \sum_i \pi(i)\hat{V}_i(g_j) + (1-\hat{\theta}_j)\sum_i c(i)\hat{V}_i(g_j) > \hat{\theta}_\ell \sum_i \pi(i)\hat{V}_i(g_\ell) + (1-\hat{\theta}_\ell)\sum_i c(i)\hat{V}_i(g_\ell),$$

a contradiction. Therefore, $\hat{\theta}_j = \hat{\theta}_\ell$. Let $\hat{\theta} = \hat{\theta}_k$ for all $k \in \{1, \ldots, K\}$. Observe that since $\mathcal{H}(\Psi_j, c, \hat{\theta})$ are supporting hyperplanes of $\mathcal{F}(\Delta)$ for all $j \in \{1, \cdots, K\}$, $\mathcal{F}(\Delta) \subseteq \mathcal{H}(\Delta, c, \hat{\theta})$.

Now, consider any vertex π^*_m of Δ. Because $\pi^*_m \in \bigcap_{k\in J_m} F_k$,

$$\bigcap_{k\in J_m} \left(\arg\min_{p\in\Delta} \sum_i p(i)\hat{V}_i(f_k) \right) \neq \emptyset.$$

Then, Claim 13 implies:

$$\bigcap_{k\in J_m} \left(\arg \min_{p\in\mathcal{F}(\Delta)} \sum_i p(i)\hat{V}_i(f_k) \right) \neq \emptyset.$$

But:

$$\bigcap_{k\in J_m} \mathcal{H}(F_k, c, \hat{\theta}) = \mathcal{H}(\pi^*_m, c, \hat{\theta}).$$

We thus have:

$$\emptyset \neq \bigcap_{k\in J_m} \left(\arg \min_{p\in\mathcal{F}(\Delta)} \sum_i p(i)\hat{V}_i(f_k) \right) \subseteq \bigcap_{k\in J_m} \mathcal{H}(F_k, c, \hat{\theta}) = \mathcal{H}(\pi^*_m, c, \hat{\theta}).$$

Therefore, $\mathcal{H}(\pi^*_m, c, \hat{\theta}) \in \mathcal{F}(\Delta)$. Now consider any other vertex π^*_r of Δ. Then there exists a permutation $\varphi : N \to N$ (that depends of π^*_r) satisfying $\varphi(S_1) = S_1$ and $\varphi(S_2) = S_2$, such that $\pi^*_r = (\pi^*_m)^\varphi$. Since $\Delta = \Delta^\varphi$ for any such permutation, Claim 12 implies that:

$$\left(\mathcal{H}(\pi^*_m, c, \hat{\theta})\right)^\varphi = \mathcal{H}((\pi^*_m)^\varphi, c, \hat{\theta}) = \mathcal{H}(\pi^*_r, c, \hat{\theta}) \in \mathcal{F}(\Delta).$$

Thus, for any vertex π^* of Δ, $\mathcal{H}(\pi^*, c, \hat{\theta}) \in \mathcal{F}(\Delta)$. Since Δ is polyhedral, $\mathcal{H}(\Delta) = co\{\mathcal{H}(\pi^*_m, c, \hat{\theta})|\pi^*_m \in \Pi\}$. Therefore $\mathcal{H}(\Delta, c, \hat{\theta}) \subseteq \mathcal{F}(\Delta)$. Since we proved that $\mathcal{F}(\Delta) \subseteq \mathcal{H}(\Delta, c, \hat{\theta})$, we finally obtain $\mathcal{H}(\Delta, c, \hat{\theta}) = \mathcal{F}(\Delta)$.

It remains to consider the case $|S_i| = 1$ for some $i \in \{1, 2\}$ (if $|S_1| = |S_2| = 1$, the result follows trivially from Theorem 1). Assume without loss of generality, that $|S_1| = 1$. Then $\Delta = \alpha\delta_k + (1-\alpha)\Delta_2$, for some $k \in N$, where δ_k is the probability distribution on N defined by $\delta_k(k) = 1$. By Theorem 1, $\mathcal{F}(\Delta) = \alpha\delta_k + (1-\alpha)\mathcal{F}(\Delta_2)$. But by the above result, $\mathcal{F}(\Delta_2) = \mathcal{H}(\Delta_2, c_2, \hat{\theta})$. Hence, $\mathcal{F}(\Delta) = \alpha\delta_k + (1-\alpha)\mathcal{H}(\Delta_2, c_2, \hat{\theta}) = \mathcal{H}(\alpha\delta_k + (1-\alpha)\Delta_2, \alpha\delta_k + (1-\alpha)c_2, \hat{\theta})$, which proves the desired result. □

Claim 15. *There exists $\theta \in [0, 1]$ such that for all subset S of N, $\mathcal{F}(\Delta(S)) = \mathcal{H}(\Delta(S), c_S, \theta)$.*

Proof. Let S_1 and S_2 be two subsets of N. We use the same notation as in Claim 14. By Claim 14, we know that there exist $\theta_1, \theta_2 \in [0, 1]$ such that $\mathcal{F}(\Delta_i) = \mathcal{H}(\Delta_i, c_i, \theta_i)$. What remains to be proved is that $\theta_1 = \theta_2$. Let $\alpha \in (0, 1)$, and $\Delta = \alpha\Delta_1 + (1-\alpha)\Delta_2$. By Claim 14, we also know that there exists $\theta_3 \in [0, 1]$ such that:

$$\mathcal{F}(\Delta) = \mathcal{H}(\Delta, \alpha c_1 + (1-\alpha)c_2, \theta_3) = \alpha\mathcal{H}(\Delta_1, c_1, \theta_3) + (1-\alpha)\mathcal{H}(\Delta_2, c_2, \theta_3).$$

Finally, by Theorem 1,

$$\mathcal{F}(\Delta) = \alpha\mathcal{F}(\Delta_1) + (1-\alpha)\mathcal{F}(\Delta_2)$$

Therefore,

$$\alpha\mathcal{H}(\Delta_1, c_1, \theta_1) + (1-\alpha)\mathcal{H}(\Delta_2, c_2, \theta_2) = \alpha\mathcal{H}(\Delta_1, c_1, \theta_3) + (1-\alpha)\mathcal{H}(\Delta_2, c_2, \theta_3)$$

which implies $\theta_1 = \theta_2 = \theta_3$, the desired result. □

References

Anscombe, F., and R. Aumann (1963): "A definition of subjective probability," *Annals of Mathematical Statistics*, 34, 199–205.

Binmore, K. (1989): "Social Contract I: Harsanyi and Rawls," *Economic Journal*, 99(395), 84–102.

Castagnoli, E., F. Maccheroni, and M. Marinacci (2003): "Expected Utility with Multiple Priors," *Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications held in Lugano, Switzerland*, pp. 121–132.

Chateauneuf, A. (1991): "On the use of capacities in modeling uncertainty aversion and risk aversion," *Journal of Mathematical Economics*, 20, 343–369.

Chew, S. (1983): "A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox," *Econometrica*, 51, 1065–1092.

——— (1989): "Axiomatic utility theories with the betweenness property," *Annals of operations research*, 19, 273–298.

De Meyer, B., and P. Mongin (1995): "A Note on Affine Aggregation," *Economics Letters*, 47(2), 177–183.

Dekel, E. (1986): "An axiomatic characterization of preferences under uncertainty: weakening the independence axiom," *Journal of Economic Theory*, 40, 304–318.

Dhillon, A., and J. Mertens (1999): "Relative utilitarianism: An improved axiomatization," *Econometrica*, 67, 471–498.

Diamond, P. (1967): "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility : Comment," *Journal of Political Economy*, 75, 765–766.

Ellsberg, D. (1961): "Risk, ambiguity, and the Savage axioms," *Quarterly Journal of Economics*, 75, 643–669.

Epstein, L., and U. Segal (1992): "Quadratic social welfare functions," *," Journal of Political Economy*, 99, 263–286.

Fishburn, P. (1970): *Utility theory for decision making*. Wiley.

Gajdos, T., J.-M. Tallon, and J.-C. Vergnaud (2004a): "Coping with imprecise information: A decision theoretic apporach," Document de travail du crest, n. 2004-14.

——— (2004b): "Decision making with imprecise probabilistic information," *Journal of Mathematical Economics*, (40), 647–681.

Gilboa, I., and D. Schmeidler (1989): "Maximin Expected Utility With a Non-Unique Prior," *Journal of Mathematical Economics*, 18, 141–153.

Harsanyi, J. (1953): "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking," *Journal of Political Economy*, 61, 434–435.

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

Hayashi, T. (2005): "Information, Subjective Belief and Preference," mimeo, University of Rochester.

Jaffray, J.-Y. (1989): "Linear utility for belief functions," *Operations Research Letters*, 8, 107–112.

Karni, E. (1998): "Impartiality: Definition and representation," *Econometrica*, 66, 1405–1515.

——— (2003): "Impartiality and interpresonal comparisons of variations in well-being," *Social Choice and Welfare*, 21, 95–111.

Karni, E., and Z. Safra (2000): "An extension of a theorem of von Neumann and Morgenstern with an application to social choice theory," *Journal of Mathematical Economics*, 34, 315–327.

Karni, E., and D. Schmeidler (1981): "An expected utility theory for state-dependent preferences," *Working paper (Foerder Institute for Economic Research, Tel Aviv University, Tel Aviv*, 48.

Karni, E., and J. A. Weymark (1998): "An informationally parsimonious impartial observer theorem," *Social Choice and Welfare*, 15(3), 321–332.

Klibanoff, P. (2001): "Characterizing uncertainty aversion through preference for mixtures," *Social Choice and Welfare*, 18, 289–301.

Luce, R., and D. Krantz (1971): "Conditional Expected Utility," *Econometrica*, 39, 253–271.

Mongin, P. (2001): "The Impartial Observer Theorem of Social Ethics," *Journal of Mathematical Economics*, 17, 147–179.

Mongin, P., and C. d'Aspremont (1998): "Utility Theory and Ethics," in *Handbook of Utility Theory, I*, ed. by S. Barberà, P. Hammond, and C. Seidl, pp. 233–289. Reidel, Dordrecht.

Moreno-Ternero, J., and J. Roemer (2005): "Objectivity, priority and the veil of ignorance," Core discussion paper 2005-81.

Rawls, J. (1971): *A Theory of Justice*. Harvard University Press, Cambridge, Mass.

——— (1974a): "Concepts of distributional equity: Some reasons for the maximin criterion," *American Economic Review*, 64(2), 141–146.

——— (1974b): "Reply to Alexander and Musgrave," *Quartely Journal of Economics*, 88(4), 633–655.

——— (1975): "A Kantian Conception of Equality," *Cambridge Review*, 96(28), 94–99.

Schmeidler, D. (1984): "Subjective Probability and Expected Utility Without Additivity," *IMA preprint series*.

Segal, U. (2000): "Let's agree that all dictatorship are equally bad," *Journal of Political Economy*, 108, 569–589.

Sen, A. (1976): "Welfare Inequalities and Rawlsian Axiomatics," *Theory and Decision*, 7, 243–262.

Vickrey, W. (1945): "Measuring marginal utility by reaction to risk," *Econometrica*, 13, 319–333.

Wang, T. (2003): "Two classes of multiple priors," Discussion paper, University British Columbia.

Weymark, J. A. (1991): "A Reconsideration of the Harsanyi-Sen Debate on Utilitarianism," in *Interpersonnal Comparisons of Well-Being*, ed. by J. Elster, and J. Roemer, pp. 255–320, Cambridge. Cambridge University Press.