Utilitarian Cooperation under Incomplete Information*

Klaus Nehring

Department of Economics

University of California, Davis

e-mail: kdnehring@ucdavis.edu

First version: December 2000

This version: January 2002

*I thank audiences at the universities of Bielefeld, Bonn, Penn, Princeton, Rochester, and the Université de Quebec à Montréal for helpful comments. I am also grateful for the hospitality of Princeton University where this paper was completed.

ABSTRACT

A theory of cooperative choice under incomplete information is developed in which agents possess private information at the time of contracting. It is assumed that the group of cooperating agents has agreed on a utilitarian "standard of fairness" (group preference ordering) governing choices under complete information. The task is to extend this standard to choices whose consequences depend on agents' private information. It is accomplished by formulating appropriate axioms of Bayesian coherence at the group level.

Assuming the existence of a common prior, the first main result generalizes Harsanyi's (1955) classical characterization of utilitarian preference aggregation to incomplete information. We then show that Bayesian coherence of group preferences is compatible with Interim Pareto Dominance only if a common prior exists. This second result generalizes and corrects the classical literature on consistent Bayesian preference aggregation under complete information: allowing for incompleteness of information, consistent Bayesian aggregation turns out to be possible even if agents' beliefs differ, as long as differences in beliefs can be attributed to differences in information. We finally relax the assumption that the standard of fairness is complete. In the extreme case in which no interpersonal utility-comparisons are made, this leads to an ex-interim justification of ex-ante Pareto efficiency as a criterion of welfare evaluation.

1. INTRODUCTION

Traditionally, most theories of normative collective action are situated in contexts of certainty or assume at least "complete information", in the sense that agents have mutual knowledge of each others' beliefs. Yet it is by now a truism that allocation and collective choice problems are often shaped decisively by the existence of asymmetric information. In many cases, asymmetries of information exist already at the time when basic allocation decisions are to be made; such situations of "incomplete information" will occupy center stage of the current paper.[1] Examples abound: voters, in deciding on a voting rule (constitution), typically possess private information about their preferences already. So do couples, when compromising under conflicts of interest. Also, when a regulatory process is to be designed, the stakeholders (firms, customers, workers, recipients of externalities) possess private information about its costs and benefits, of the availability of alternative technologies and outside options, etc.). Indeed, whenever an incomplete contract needs to be renegotiated, incompleteness of information is likely to be an issue.

Specifically, we shall conceptualize collective choice in terms of a group of agents who jointly choose an allocation on the basis of some agreed-upon "criterion of group optimality" or "standard of fairness". Incompleteness of information complicates collective choice in two basic ways. It implies first that a group of agents cannot choose a (state-contingent) allocation directly, but only an allocation mechanism which induces a state-contingent allocation via the equilibrium play of the agents. This leads to the familiar incentive compatibility constraints on the set of feasible allocations. Yet incompleteness of information entails an additional, much less studied difficulty: by its very nature, the *collective* choice of the allocation mechanism must be based on publicly available (commonly known) information only. Hence any evaluation criterion underlying the group choice must reflect the group's ignorance of agents' private information. In particular, following Holmstrom-Myerson (1983) and others[2], the fundamental notion of Pareto efficiency needs to be formulated ex interim, at the stage when agents know their own information but not yet that of others; a state-contingent allocation f interim Pareto dominates another allocation g if it is *commonly known*[3] that every agent strictly prefers f to g.

[1] By contrast, when asymmetries of information arise only after the fundamental allocation decsions are taken, one speaks of "imperfect information".

[2] Independent originators of the notion of interim Pareto efficiency include Harris-Townsend (1981), Wilson (1978), and indeed, though lacking any discussion, Harsanyi-Selten (1972).

[3] An event E is commonly known if it is the case that E, that everyone knows that E,that everyone knows that everyone knows that E, etc. .

To go beyond interim Pareto dominance, we shall assume that the group has already accepted a "standard of fairness" as the basis for collective choice under complete information; examples of such standards are the utilitarian and the Rawlsian (leximin) orderings, as well as the Nash bargaining solution. We shall ask how these standards are to be extended to situations of incomplete information. In the present paper, we shall concentrate on the case of a utilitarian standard in the sense of Harsanyi's (1955) classic contribution, since it is the central, most frequently used and best-behaved criterion of social optimality under uncertainty. This focus allows us to zero in on the fundamental epistemic issues that arise form the nature of collective choice under incomplete information per se, for any standard. In a companion paper (Nehring 2002a), we plan to extend the analysis to non-utilitarian standards, a task which raises a number of additional, potentially controversial normative issues.

To get a better grasp of the issues involved, consider a situation a la Myerson-Satterthwaite (1983) in which two risk-neutral agents have the opportunity to trade an indivisible good. Assume that the agents already know their own reservation price for the good before selecting a trading mechanism,[4] and that they accept the utilitarian criterion as standard of fairness under complete information, evaluating outcomes in terms of the sum of utilities in the money metric. If, counterfactually, the two agents were able to choose the mechanism before knowing their reservation price, determining the optimal mechanism would simply be a matter of maximizing the expected total gains from trade ex ante as computed by Myerson and Satterthwaite. By contrast, under incomplete information, when the agents already know their reservation price before choosing the mechanism, the appropriate optimality criterion is no longer self-evident. Specifically, it is not clear on what basis a social expectation can be taken, since ex interim both agents have different information and thus different beliefs; moreover, these beliefs are not commonly known, and thus simply not available as the basis for cooperative, hence public, decision making. What is needed is a "group belief" that is derived from the agents' individual beliefs but relies on commonly known information only. Can such a "group belief" be meaningfully determined?

[4] This is in line with Myerson-Satterthwaite's crucial assumption of *ex-interim* participation constraints; I thank Urs Schweizer for pointing this out.

Constructing the Common Prior as the Group Belief

Whether and how a "group belief" can be determined depends in part on the structure of the agents' beliefs about each other. In this paper, we shall throughout make the standard assumption that agents' beliefs are consistent with the existence of a common prior.[5] In this case, we shall argue that the common prior is the desired group probability. Our argument will be decision-theoretic rather than epistemic. Indeed, a viable argument arguably has to take that form: given that groups have no minds of their own, it stands to reason that the very notion of a group probability must be decision-theoretic, that is, it must reflect the group's preferences[6] over bets. To make a decision-theoretic argument possible, we introduce a "type space in preferences" in which both individual and social preferences are defined over the domain of all acts (mappings from states of the type space to consequences); see section 5. Incompleteness of information is captured by making individual agents' preferences a random variable; by contrast, group preferences are assumed to be non-random (w.l.o.g.), since they must be based on commonly known information only.

The first main result of the paper, Theorem 3, the "Aggregation Theorem", identifies the common prior as the searched-for group belief, effectively generalizing Harsanyi's (1955) classical result to incomplete information. Specifically, it yields social preferences of the form

$$E_\mu \sum_{i \in I} U_i^f \,,$$

where U_i^f is agent i's utility derived from act f and viewed as a random variable, and E_μ denotes the expectation with respect to the common prior μ. A potential source of controversy is the use of a "subjective expected utility" functional form for group preferences; in deriving it via the Aggregation Theorem, we have striven to make do with minimal assumptions on the Bayesian rationality of group preference. Indeed, the only axiom with some Bayesian flavor is that of State Independence; neither Independence nor Completeness are assumed for group preferences over general acts.

In the context of contract theory, the Aggregation Theorem provides a justification for applying Myerson-Satterthwaite's (1983) computation of the optimal trading mechanism to situations in which the selection of the trading mechanism itself takes place ex interim. Moreover, in risk-

[5]That is to say, there exists a probability measure (the "common prior") on the state space (here: the type space) such that the beliefs of every type of every agent can be viewed as posteriors resulting from conditioning on the type. See section 2 for further discussion.

[6]In the following, we shall be using expressions such as "group preference", "social ordering", "standard of fairness" interchangeably.

neutral settings without participation constraints a la D'Aspremont-Gerard-Varet (1979) and Arrow (1979), in which there exist mechanisms that guarantee ex-post efficient allocations, the Aggregation Theorem justifies choosing such mechanisms over the others.[7]

When is "Bayesian Aggregation" Possible under Incomplete Information ?

The Aggregation Theorem assumes that agents' beliefs admit a common prior. This is necessary: the second main result of the paper, Theorem 4, the "Possibility Theorem", shows that it is possible for social preferences to jointly satisfy State Independence and Interim Pareto Dominance *only* if agents' beliefs admit a common prior. This result generalizes the existing (im)possibility results on Bayesian aggregation in the literature, all of which have been formulated in the context of complete information (see especially Hylland-Zeckhauser 1979, Hammond 1981, Broome 1990, Seidenfeld-Kadane-Shervish 1989 and Mongin 1995); that literature concluded that group preferences can satisfy Bayesian coherence together with ordinary Pareto Dominance only if agents' beliefs coincide. By contrast, according to Theorem 4, differences in beliefs among agents do not force a choice between Bayesian coherence and respect for consensual preference per se; the conflict arises only when these differences cannot be fully attributed to differences in information.[8] In a companion paper (Nehring 2002b), we shall extend the theory to the case of inconsistent beliefs in which there is no common prior.

Partial Utilitarianism: Allowing for Incomplete Interpersonal Comparisons of Utility

Among the manifold criticisms of utilitarianism, two stand out in particular, the assumption of the possibility of interpersonal utility comparisons, and the exclusion of fairness concerns about the distribution of utilities.[9] The latter issue shall be taken up in a companion paper, Nehring (2002a),

[7]Note that there typically are many mechanism that fail to guarantee ex-post efficiency but are not dominated ex interim by some mechanism that does; the selection of ex-post efficient mechanisms can thus not be justified on grounds of interim efficiency alone.

[8]In the sense that the different beliefs can be viewed as updates on a common prior. This interpretation of the common prior assumption is often referred to as the "Harsanyi doctrine". Its meaningfulness in a static setting is controversial. While Dekel-Gul (1997) have raised doubts, Bonanno-Nehring (1999) and Nehring (2001) endowed it with content even in a static setting in which there is no dynamic notion of receiving information.

[9]As a normative theory of group choice, utilitarianism is restrictive also by relying on preference (and group feasibility) information only; it is not a theory of cooperative bargaining, after all.

taking account in particular of Diamond's (1967) famous critique of Harsanyi's (1955) theorem. The former is addressed in section 9, where we allow the social standard to be incomplete while maintaining its additive character. In order to obtain a unique extension analogous to that of the Aggregation Theorem, the axioms of Bayesian rationality of the social ordering must be strengthened substantially. Besides assuming the Independence axiom, one must also require that group preferences over bets "reveal" a consistent likelihood ordering; this leads to a condition analogous to Savage's comparative likelihood axiom P4. The main result of section 9, Theorem 5, characterizes the unique extension of a partial utilitarian standard under these assumptions.

Of particular interest is the case in which the group abstains from *any* interpersonal utility comparisons, i.e. when Pareto Dominance is the sole criterion for group preference under complete information. At first sight, it may seem that one is then necessarily thrown back to interim Pareto efficiency, but this is not so! Indeed, Theorem 5 characterizes the extension of the "Pareto standard" as follows: the act f is weakly superior to the act g if every agent's imputed expected utility based on the common prior resulting from f exceeds that resulting from g, i.e. if

$$E_\mu U_i^f \geq E_\mu U_i^g \text{ for all agents } i. \tag{1}$$

We shall call an act (state-contingent allocation) "Bayesian interim efficient" if it is not dominated by some other feasible act in terms of the extended Pareto standard. In view of its characterization in terms of condition (1), Bayesian interim efficiency is equivalent to *ex-ante* Pareto efficiency, yet fully motivated from an interim perspective.

Bayesian interim efficiency reduces the problem of distributing utility to one among *agents* rather than among *types* of agents, as it would appear from the perspective of interim Pareto efficiency. The latter is eliminated, in effect, by assigning a definite weight to each type, namely its prior probability. This makes compelling conceptual sense, since, after all, only agents are real, and types are merely events denoting the epistemic possibility that an agent has particular preferences and beliefs.

Related Literature

While the literature on non-cooperative decision making under incomplete information is vast, cooperative decision making has received far less attention in this setting. The attention that it has received is almost exclusively in the positive, strategic vein; following the seminal paper by Wilson (1978), this literature focuses mainly on incomplete information versions of the core and its properties.[10] By contrast, we are aware of only two prior contributions to the theory of *"norm-based"* cooperative decision making under incomplete information with a motivation broadly comparable to ours, namely the early papers by Harsanyi-Selten (1972) and Myerson (1984) on the two-person Nash-bargaining solution under incomplete information. Since both of these propose extensions of a different, non-utilitarian standard of fairness, we shall defer a more detailed discussion of the specific proposals to Nehring (2002a), besides noting that while both assume interim Pareto efficiency, neither proposal yields Bayesian interim efficient allocations except under special circumstances.

Conceptually, neither of these two contributions foregrounds the epistemic issue of "group belief" that is central here. This may partly be the result of a different perspective: Myerson (1984) conceives of the intended solution as the recommendation of an impartial arbitrator (who is not formally modelled) without relevant private information of his own; while not very explicit about their vantage point, Harsanyi and Selten are probably also best read in this way.[11]

The common prior assumption plays a very different role in the two contributions. In Harsanyi-Selten (1972), it is central: indeed, their solution can be viewed as a non-symmetric cooperative Nash bargaining solution among the types of all players, in which types are weighted by their prior probability. Nonetheless, the relevance of the common prior becomes never an issue, since all beliefs are formulated as prior beliefs. Myerson (1984, p. 466), by contrast, invokes a "probability-invariance axiom" which states that "probabilities cannot be meaningfully defined separately from utilities, when state-dependent utilities are allowed"; by consequence, his solution does not separate between probabilities and utilities, either, and common priors do not matter. See section 6 for further discussion.

We have motivated this paper as dealing with norm-based cooperative decision-making by a group of agents. Alternatively, the paper can be viewed as developing a *welfare theory* under incomplete information that is essentially *Paretian* in that it respects agents' consensual preferences over acts and thus reflects their interim expectations. This *internal* approach to welfare theory is to be distinguished from an *external* one, on which welfare judgements are interpreted as those of an impartial outside observer, in other words: as judgments "for" the group but not "by" the group. Such an external approach naturally to takes the outsider's beliefs as the basis of welfare evaluation. While this sidesteps the issues arising from incompleteness of information, the outsider's ranking

[10] For a look at this currently quite active field, see the recent special issue of *Economic Theory* (2001, volume 18/2).

[11] The lack of an emphasis on epistemic issues probably also reflects the unvailability of crucial concepts at the time; indeed, Harsanyi-Selten (1972) predates even the fundamental concept of common knowledge (formalized first by Aumann 1976).

will be consistent with Interim Pareto Dominance only in special circumstances.[12]

Under uncertainty, the Pareto criterion is controversial; recent critics include Broome (1991), Mongin (1995,1998) and Gilboa-Samet-Schmeidler (2001). All of these criticisms apply to situations in which agents' beliefs differ and are commonly known; intuitively, one can argue that in such situations the beliefs of at least one agent must be "incorrect".[13] Clearly, this line of criticism does not undermine the normative appeal of Interim Pareto Dominance in the presence of a common prior, since in such situations differences in beliefs can be reconciled as reflecting differences in information.

Plan of the Paper

After introducing the formal framework in section 2, we illustrate some of the basic issues and concepts in section 3 through a voting example. From an epistemic point of view, the major issue is the relevance of the common prior at the interim stage. This issue becomes especially urgent if the meaningfulness of postulating a prior stage in which agents *had* complete information is denied, as in the critique of the common prior assumption by Dekel-Gul (1997) and Gul (1998). We summarize in section 4 the recent literature on the topic which characterizes both the existence of a common prior and the common prior itself purely in terms of interim beliefs; we also add a novel characterization of the common prior in terms of the interim betting behavior among risk-neutral agents. The results of section 4 form the "epistemic core" of the decision-theoretic results in the remainder of the paper and provide key tools for their proofs.

In section 5, we introduce a "type space in preferences" as the setting for the subsequent analysis; it amounts to an adaptation of the Anscombe-Aumann (1963) framework of decision-making under uncertainty to multiple agents with incomplete information. In section 6, the extension of a utilitarian standard to incomplete information is characterized (Theorem 3, the Aggregation Theorem). We then show that the key normative axioms underlying the Aggregation Theorem, Interim Pareto Dominance and State Independence, are compatible only if agents' beliefs admit a common prior (Theorem 4, the Possibility Theorem; Section 7). The interpretation of these results is discussed in

[12]Indeed, roughly speaking, this will only happen (in the case a utilitarian standard) if agents' beliefs are compatible with a common prior, and if the outsider's beliefs coincide with the common prior. The latter presumes that any relevant information by the outsider is commonly known to the group.

[13]This intuition underlies especially Gilboa-Samet-Schmeidler (2001) who formulate a restricted Pareto axiom that applies only to acts definable in terms of events over which agents' beliefs agree.

some detail in Section 8. Finally, in section 9, we generalize the Aggregation Theorem to the case of "partial utilitarian standards" characterized by limited interpersonal utility comparisons. Section 10 concludes. All proofs can be found in the appendix.

2. FORMAL FRAMEWORK

Definition 1 *A* **type space** *is a tuple* $\langle I, \Omega, \{p_i\}_{i \in I} \rangle$*, where*

- I *is a finite set of* agents.
- Ω *is a finite set of* states*; the subsets of* Ω *are called* events.
- *for every agent* $i \in I$*,* p_i *is a function that specifies, for each state* $\alpha \in \Omega$*, his probabilistic beliefs* $p_i^\alpha : 2^\Omega \to \mathbf{R}$ *at* α*.*

  These satisfy:
- *(Introspection)* For all $\alpha \in \Omega$ and all $i \in I$: $p_i^\alpha(\{\omega \in \Omega \mid p_i^\omega = p_i^\alpha\}) = 1$.
- *(Truth)* For all $\alpha \in \Omega$ and all $i \in I$: $p_i^\alpha(\{\alpha\}) > 0$.

The last two conditions may require some explanation. The first of them, "Introspection", states that agents know their own beliefs: at any state $\alpha \in \Omega$, any agent i is certain of the event that his belief p_i^ω is equal to his actual belief p_i^α. The second, "Truth", is equivalent to the condition that, for any state $\alpha \in \Omega$ and any agent i, if i is certain of some event $E \subseteq \Omega$, E is in fact the case at α; in particular, i cannot be certain of the false proposition "α does not occur" represented by the event $\Omega \backslash \{\alpha\}$. Truth is assumed for simplicity but is not unrestrictive.[14]

A type space can be viewed as a state space in which agents' beliefs are included in the description of the state. As a result, an agent's beliefs at a state describes not only his beliefs about facts of nature, but also his beliefs about other agents' (first-order) beliefs about states of nature, hence also his beliefs about agents' higher-order beliefs about states of nature, thus in effect: an entire belief hierarchy. A state in a type space can thus be thought of as a notational device for describing the belief hierarchies of each agent.[15]

[14]See Bonanno-Nehring (1999) for a detailed study of its relaxation.

[15]By results due to Armbruster-Boege (1979), Mertens-Zamir (1985), Brandenburger-Dekel (1993), any profile of probabilistic belief hierarchies has a type-space representation; the assumption that the state space Ω is finite is

For any $\alpha \in \Omega$, let $\Pi_i(\alpha) := \operatorname{supp} p_i^\alpha$. By Introspection and Truth, the family $\Pi_i := \{\Pi_i(\omega) \mid \omega \in \Omega\}$ is a partition of Ω, i's *type partition.* An agent "knows" an event E at α if he is certain of it, i.e. if $E \supseteq \Pi_i(\alpha)$. Let Π denote the meet (finest common coarsening) of the partitions $\{\Pi_i\}_{i \in I}$, with $\Pi(\alpha)$ denoting the cell of the meet containing state α. E is *common knowledge* if everybody knows that E, and if everybody knows that everybody knows that E, and so forth. Formally, E is "common knowledge" at α if $E \supseteq \Pi(\alpha)$. Since type spaces serve as a notational vehicle to represent hierarchies of (interim) beliefs, it can be assumed w.l.o.g. that $\Pi = \{\Omega\}$, i.e. that the universal event Ω is the smallest common knowledge event. Note that it is then unambiguous to speak of the common knowledge of an event without reference to the state.

A random variable Z is a real-valued function on Ω. For any random variable Z, agent i's expectation of that random variable, when viewed as a function of the state, is again a random variable $E_i Z : \alpha \mapsto E_i^\alpha Z$ with $E_i^\alpha Z = \sum_{\omega \in \Omega} p_i^\alpha(\omega) Z(\omega)$. For a probability measure μ on Ω, let $E_\mu Z$ denote the expectation of Z with respect to μ, $E_\mu Z = \sum_{\omega \in \Omega} \mu(\omega) Z(\omega)$. Finally, the indicator function associated with the event A is denoted by 1_A; constant random-variables are denoted in boldface.

The conventional interpretation of a type space is dynamic, describing a point in time (the "interim stage") at which agents have updated their prior beliefs upon receiving some private information signal, and where agents' prior beliefs had been commonly known at an "ex-ante" stage. This "*dynamic interpretation*" is appealed to in the standard narrative accompanying asymmetric information models, and the reader may assume it for the sake of familiarity. In formal terms, a dynamic interpretation assumes as primitives agents' priors q_i (with support Ω) and information partitions Π_i; an agents' type corresponds to the *signal received* $\Pi_i(\alpha)$, and the interim beliefs are derived as *conditional* probabilities, $p_i^\alpha = q_i(./\Pi_i(\alpha))$.

On closer reflection, assuming the existence of a prior stage at which beliefs were commonly known seems highly restrictive : in many situations in which agents' have private information about their own preferences or abilities at a given point in time, they *always* knew more about their own preferences or abilities than others, hence the posited prior stage never existed. Some authors have recently gone further and argued that postulating a prior complete information stage may

restrictive but standard.

Strictly speaking, to describe the actual interim situation fully, the formal primitive of the theory would need to a type space together with a state $\tau \in \Omega$ interpreted as the true state; see Bonanno-Nehring (1999) for an extended argument. Since here all assumptions and conclusions exclusively concern what is commonly known among agents, and since commonly known facts are independent of $\alpha \in \Omega$, it is not necessary to explicitly specify the true state τ. Thus, for convenience and familiarity, we have omitted it throughout.

not be meaningful even as a counterfactual; see in particular Dekel-Gul (1997) and Gul (1998). In the absence of such a prior stage, interim beliefs are unconditional probabilities describing agents' mutual uncertainty about each other. This "*static interpretation*" will be the "official" one adopted throughout the paper; nonetheless, we will typically use the dynamic ex-interim/ex-ante terminology due to its suggestiveness and entrenchment.

3. A VOTING EXAMPLE

To illustrate some of the central issues and concepts, we shall briefly study a highly stylized model of single-issue voting. Voters $i \in I$ have quasi-linear preferences over the adoption of a "project" $y \in Y = \{0,1\}$ and net transfers t_i $u_i(y,t_i) = y\theta_i + t_i$, where θ_i denotes i's benefit of the project. For maximal simplicity and not infrequent realism, we shall assume that transfers are de facto infeasible; their role is merely to measure voters' preference intensity. For the sake of comparison, we shall first describe the cooperative choice of a voting mechanism *ex ante,* i.e. under the assumption that voters have not yet received private information about their preferences. Thereafter, we shall look at the same cooperative choice problem *ex interim*, if the voting mechanism is to be chosen when voters already know their own preferences but not others.

A state $\omega \in \Omega$ is a profile of net benefits $\theta = (\theta_i)_{i \in I}$. Ex ante, the agents have a common prior μ over $\Omega = \mathbf{R}^I$; they anticipate that ex interim, they will be informed of their own net benefit θ_i, and of nothing else. An agent's "type" can thus be identified with his net benefit θ_i. For simplicity, assume that types are independent, i.e. that $\mu = \prod_{i \in I} \mu_i$. Voters must choose an incentive-compatible (direct) mechanism $f \in \mathcal{F}^{ic}$, i.e. a "state-contingent allocation" $f : \Omega \to \{0,1\}$ that depends only the sign of voters' net benefit: $f(\theta) = \widehat{f}(\sigma(\theta))$ with $\sigma(\theta) = (sgn\ \theta_i)_{i \in I}$, and such that $\widehat{f}$ is non-decreasing in $\sigma(\theta)$; note that incentive-compatibility is equivalent here to strategy-proofness.

Mechanism Selection Ex Ante

In deciding on an optimal voting mechanism, voters have agreed upon a utilitarian standard given by the summation of utilities in the money metric. Ex ante, its meaning is conceptually unproblematic: it amounts to choice of a mechanism $f \in \mathcal{F}^{ic}$ that maximizes $E_\mu \sum_{i \in I} U_i^f$, where $U_i^f(\theta) := u_i(f(\theta), 0)$.

Observation 1 *An optimal mechanism is given by*

$$f^*(\theta) := \begin{cases} 1 & \textit{if } \sum_{i\in I} E_{\mu_i}[\theta_i \mid sgn\ \theta_i] > 0 \\ 0 & \textit{otherwise} \end{cases}.$$

The optimal mechanism f^* is simply determined by maximizing $\sum_{i\in I} u_i(y)$ given knowledge of $\sigma(\theta)$. It can be interpreted as a voting mechanism in which voter i is given a choice between casting $E_{\mu_i}[\theta_i \mid sgn\ \theta_i > 0]$ votes *for* the project, $E_{\mu_i}[-\theta_i \mid sgn\ \theta_i < 0]$ votes *against* the project, or abstaining. Thus, if the expected intensity of a positive preference exceeds that of a negative preference, a voters' "pro" vote will weigh more heavily than his "con" vote. Moreover, if voters' expected intensities in both directions are equal, a voter's weight in the optimal mechanism will reflect their ex-ante expected strength of preference $E_{\mu_i}[|\theta_i| \mid sgn\ \theta_i]$.

To simplify even further, assume that the common prior is symmetric in voters, i.e. that $\mu_i = \mu_*$ for all $i \in I$. Then the optimal mechanism must be anonymous, and thus amount to "voting by quota". The voting-by-quota mechanism f_q (with quota q) is defined by setting

$$f_q(\theta) := \begin{cases} 1 & \text{if } \frac{\#\{i|\theta_i>0\}}{\#\{i|\theta_i\neq 0\}} > q, \\ 0 & \text{otherwise} \end{cases}.$$

Observation 2 *With a symmetric common prior, voting by quota is optimal,* $f^* = f_{q^*}$.
The optimal quota q^* *is equal to* $\frac{E_{\mu_*}[|\theta_i||\theta_i<0]}{E_{\mu_*}[|\theta_i||\theta_i<0]+E_{\mu_*}[|\theta_i||\theta_i>0]}$.

Typically, the optimal quota will differ from 0.5; for example, if pro voters care more than con voters in expectation, less than 50 % of pro-votes will be sufficient to accept the project.

With a symmetric common prior, it is also possible to justify the optimal mechanism by an efficiency argument. Indeed, ex ante, every agent prefers the quota q^* to any other, since for all $i \in I$ and $q \in [0,1] : E_\mu U_i^{f_{q^*}} \geq E_\mu U_i^{f_q}$. This leads immediately to the following observation.

Observation 3 f_{q^*} *is the unique anonymous mechanism that is ex-ante efficient among all incentive-compatible mechanisms* $f \in \mathcal{F}^{ic}$.

Mechanism Selection Ex Interim

Assume now that voters already know their own type (θ_i^τ) when selecting the voting mechanism (quota). Is the mechanism f_{q^*} still uniquely optimal ex-interim? And if so, in what sense? Does the argument from Pareto efficiency survive?

To address this question, one needs to recognize that under incomplete information, the relevant criterion of Pareto efficiency is that of interim Pareto efficiency due to Holmstrom-Myerson (1983): a mechanism f is *interim Pareto efficient* if there exists no other feasible mechanism g that is *commonly known* to be preferred by every agent, where each agent evaluates mechanisms based on his current (interim) beliefs. Formally, f **interim Pareto dominates** g if it is common knowledge that $E_i^\omega U_i^f > E_i^\omega U_i^g$ for all $i \in I$, where $E_i^\omega U_i^f$ is i's interim expected utility under f; see also section 6. In our example, *any* quota is interim efficient.

Observation 4 *For any $q \in [0, 1]$, f_q is interim Pareto efficient.*

To see this, consider a type with a positive net benefit $\theta_i > 0$. Such a type prefers mechanisms with a higher (interim) probability that the project will be accepted; this probability is given as $\prod_{j \neq i} \mu_j (\{\theta_{-i} \mid f(\theta_i, \theta_{-i}) = 1\})$. Evidently, this probability will be the larger, the smaller the quota q is; in particular, the quota $q = 0$ is his most preferred, as it guarantees acceptance of the project. Conversely, any type with $\theta_i < 0$ is interested in minimizing the chance that the project is accepted, with $q = 1$ as the best. This shows that the mechanisms based on extreme quotas f_0 and f_1 are interim Pareto efficient; a slightly more involved argument establishes this also for intermediate quotas. The example illustrates a general phenomenon: typically, there are far more interim efficient that ex-ante efficient allocations.

The argument from Pareto efficiency thus ceases to work ex interim. On the other hand, the basic intuition that the optimal mechanism should reflect the strength of preference indicated by a vote continues to apply. How can this intuition be captured? A straightforward but naive approach would be to maximize $\sum_{i \in I} E_i^\tau U_i^f$, where E_i^τ is agent i's expectation operator based on his true type θ_i^τ. The problem with this criterion is simply its inapplicability: no one knows its value, since every voter knows only the value of his own term in the sum. Intuitively, a viable criterion must in some manner "abstract" from agents' *private* expectations and extract a "publicly accessible version" of them to arrive at a decision criterion that can serve as the basis for collective decision making.

With independent types, the following heuristic consideration shows how this can be achieved. The key is the observation that, while ex-interim the value of $E_i U_i^f$ is known only to voter i, all other voters have the same interim belief about i's type (given by the commonly known probability measure μ). Thus others' expectations $E_j E_i U_i^f$ of $E_i U_i^f$ are commonly known; being independent of j, they can be written as $E_{\neq i} E_i U_i^f$. It follows that a viable interim utilitarian criterion is given by

$$\sum_{i \in I} E_{\neq i} E_i U_i^f.$$

This criterion evaluates mechanisms on the basis of the sum of agents' interim expected utilities, *as estimated by the others.* It is easily seen that in fact

$$\sum_{i \in I} E_{\neq i} E_i U_i^f = \sum_{i \in I} E_\mu E_i U_i^f = E_\mu \left(\sum_{i \in I} U_i^f \right).$$

Thus, in the special case of independent types, we have obtained a heuristic rationale for using the common prior as the basis for an interim group expectation. This heuristic interpretation can be extended to the non-independent case using Samet's (1998) characterization of the common prior which will be presented at the beginning of the next section. In section 6, we shall justify the common prior decision-theoretically together with the expectational functional form of the social ordering .

4. BACKGROUND: COMMON PRIORS FROM AN INTERIM PERSPECTIVE

As suggested in the previous section, the common prior's relevance to group decision making at the interim stage is not immediately clear; indeed, not even its content in terms of the agents' interim belief hierarchies is obvious. This issue has been the theme of a recent literature, whose main results will be briefly reviewed here. Since these results characterize the common prior fully in "interim" terms, they render it meaningful even if no ex-ante stage exists in which agents' beliefs were commonly known. Here, they will mainly serve as key lemmas in the proofs of the central decision-theoretic results.

Formally, a probability measure $\mu : 2^\Omega \to \mathbf{R}$ is a **common prior** if, for all $i \in I$, $\omega \in \Omega$ and $A \subseteq \Omega$, $p_i^\omega(A) = \mu^\alpha(A/\Pi_i(\omega))$ whenever $\mu(\omega) > 0$. In view of the partitional structure of the Π_i and the assumption that Ω is the only common knowledge event, it is easily verified that if a common prior exists, it is unique, hence commonly known, and assigns positive probability to every state.[16]

[16] To relate this definition to the usual dynamic definition in terms of a dynamically interpreted type space and the common prior, two adjustments need to be made. First, it is clear that interim beliefs allow inferences only about agents' priors conditional on the ex-interim common knowledge event; thus, on a dynamic interpretation, a common prior is to be understood as conditioned on this event. Moreover, prior beliefs that induce the same posterior at every state are clearly indistinguishable ex-interim. Hence the existence of a common prior can only be ascertained up to that equivalence.

Two approaches have been developed to clarify the content of the common prior assumption ex interim: variants of the no-betting characterization of Morris (1994) (Bonanno-Nehring 1999, Feinberg 2000, and others), and Samet's (1998) characterization directly in terms of agents' belief hierarchies. These are complementary: while the former spells out the behavioral content of the existence of a common prior, the latter is epistemically more transparent; a unified account has been given in Nehring (2001). Importantly for the present paper, Samet's (1998) approach yields an epistemic characterization of the common prior itself. We shall provide a counterpart to his result in terms of the betting behavior of risk-neutral agents; it will serve as a crucial lemma in the proof of the Aggregation Theorem.

Samet's (1998) approach, to be presented first, relies on agents' higher-order expectations and their limits. A sequence $s = (i_1, i_2, ...)$ of elements in I is an $I-$ *sequence* if, for every $i \in I$, $i_k = i$ for infinitely many k's.

Proposition 1 (Samet) *For any random-variable Z on Ω and $I-$sequence s, the limit of the iterated expectations $\lim_{k\to\infty}(E_{i_k}...E_{i_1}Z)$ exists and its value is common knowledge at each state.*

We will write E_sZ for $\lim_{k\to\infty}(E_{i_k}...E_{i_1}Z)$ and refer to it as the *asymptotic iterated expectation of* Z with respect to s. Proposition 1 has the following intuitive content. Under incomplete information, first-order and finitely iterated expectations typically reflect some private information; in particular, their value will not be common knowledge. However, each time expectations are taken with respect to the beliefs of some agent i_k, some of the private information about Z that may have remained after the first $k-1$ iterations among the agents i different from i_k is removed. In the limit of iterating expectations infinitely deeply, all private information is removed, hence the infinitely iterated expectation will be common knowledge. Proposition 1 leads directly to the following characterization of common priors and their existence also due to Samet (1998).

Theorem 1 *A common prior exists if and only if, for any random-variable Z on Ω and any two $I-$sequences s and s', it is common knowledge that $E_s^\omega Z = E_{s'}^\omega Z$.*[17]

The common prior μ is given by $\mu(A) = E_s^\alpha 1_A$ for all $A \subseteq \Omega$, for any $\alpha \in \Omega$ and $I-$sequence s.

Theorem 1 describes precisely where in the belief hierarchy the common prior resides, namely at the limit of infinitely deep expectations. The common prior thus appears far removed from any

[17] Note that, in view of the assumption that $\Pi_I = \{\Omega\}$, common knowledge that $E_s^\omega Z = E_{s'}^\omega Z$ is equivalent to the requirement that $E_sZ = E_{s'}Z$ as random-variables.

behavioral implications. To fill this gap, the following result characterizes common priors in terms of their betting implications. It requires the existence of a transferable currency with respect to which all agents are risk-neutral. The idea is to determine those random-variables (viewed as contingent payments to the group, "bets") which the group would be willing to bet on collectively using an appropriate sharing arrangement. As part of the "rules of the game", the group is constrained to employ sharing arrangements that are commonly known to be strictly acceptable to each agent; this ensures that acceptance of the bet does not reveal private information.

Definition 2 *Z is* **acceptable for** *I if there exist $Z_i : \Omega \to \mathbf{R}$ for $i \in I$ such that $Z = \sum_{i \in I} Z_i$ and such that it is common knowledge that $E_i^\omega Z_i > 0$ for all $i \in I$.*

Theorem 2 *i) A common prior exists if and only if 0 is not acceptable for I.*

ii) If a common prior μ exists, Z is acceptable for I if and only if $E_\mu Z > 0$.

While part i) is due to Morris (1994) and well-known, part ii) is novel. To illustrate its logic, consider the following example.

Example 1. Let $I = \{1, 2\}$, $\Omega = \{\tau, \beta, \gamma, \delta\}$, with τ as the true state. Let the agents' beliefs be given by $\Pi_1 = \{\{\tau, \beta\}, \{\gamma, \delta\}\}$, $\Pi_2 = \{\{\tau, \gamma\}, \{\beta, \delta\}\}$, with $\mu = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ as the common prior. Consider the group bet $Z = (-4, -4, -4, 20)$ with a prior expectation $E_\mu Z$ of 2. By Theorem 2ii), Z must be acceptable for $\{1, 2\}$. Indeed, decomposing Z into bets $Z_1 = (-2, 4, -8, 10)$ and $Z_2 = (-2, -8, 4, 10)$ ensures expected gains of $E_1 Z_1 = E_2 Z_2 = (1, 1, 1, 1)$ for both agents, thus verifying the Theorem. Note that at the true state τ, both agents are willing to accept their part of the bet, even though they would reject the bet Z individually. Indeed, at τ both *know* that together they loose by accepting the bet (whatever the side-payments), but this is not common knowledge.

5. TYPE SPACES IN PREFERENCES

Let X be a set of deterministic "social alternatives", with typical element x. Let $\mathcal{L} = \mathcal{L}(X)$ denote the set of all probability distributions on X, with typical element ℓ. If X is infinite, to avoid technicalities let $\mathcal{L}$ denote the set of "simple lotteries" (probability distributions with finite support). In the manner of Anscombe-Aumann (1963), a (social) *act* f maps states to probability distributions of social alternatives, $f : \Omega \longrightarrow \mathcal{L}$. In a multi-agent version of Anscombe-Aumann's "horse race" interpretation, a state $\omega \in \Omega$ describes the outcome of the horse race together with a profile of agents' belief hierarchies over that outcome. A social act is given by conducting a state-contingent "roulette lottery" which selects the social alternative x with conditional probability $f_x(\omega)$. It is understood that this conditional probability distribution is agreed upon by all agents, i.e. in this sense "objective".[18] Let $\mathcal{F} = \mathcal{F}(X)$ denote the set of all such acts.

Since individuals are mutually uncertain about each others' beliefs, they must be mutually uncertain about each others' preferences, too; formally, these are random variables $\alpha \longmapsto \succeq_i^\alpha$. Since the social ordering is derived from individual preferences, in principle it too is a random variable $\alpha \longmapsto \succeq_I^\alpha$. However, to serve as a basis for collective, hence public action, the social ordering itself needs to be public, that is: commonly known among agents; since we have assumed Ω to be the unique commonly known event, the social ordering can be treated as a constant $\succeq_I$.

Some additional notation is needed. Let $\mathcal{F}^{const}$ denote the set of constant acts $f : \Omega \longrightarrow \mathcal{L}$. These correspond to the playing of the same "roulette lottery" in every state, and will thus typically be referred to by the name of that lottery $\ell \in \mathcal{L}$. It will assumed to be common knowledge that individuals are "rational" in the sense of satisfying the Anscombe-Aumann axioms of Subjective Expected Utility theory.

Axiom 1 *(Weak Order) $\succsim$ is transitive and complete.*

Axiom 2 *(Independence) $f \succeq g$ iff $\lambda f + (1-\lambda)\, h \succeq \lambda g + (1-\lambda)\, h$, for all $f, g, h \in \mathcal{F}$ and $\lambda \in (0,1)$.*

Let $[\ell, f_{-\beta}]$ denote the act agreeing with f for $\omega \neq \beta$ and equal to ℓ for $\omega = \beta$. The state β is *non-null* if $[\ell, f_{-\beta}] \succ f$ for some $f \in \mathcal{F}$ and $\ell \in \mathcal{L}$.

Axiom 3 *(State Independence) If the state β is non-null, then $[\ell, f_{-\beta}] \succeq [\ell', f_{-\beta}]$ if and only if $\ell \succeq \ell'$, for all $f \in \mathcal{F}$ and $\ell, \ell' \in \mathcal{L}$.*

Evidently, if agent i is an expected-utility maximizer at α, a state ω is null with respect to $\succeq_i^\alpha$ if and only if $p_i^\alpha(\omega) = 0$. Thus, knowledge and common knowledge are well-defined in terms of agent's preferences: agent i knows the event E at state α if E^c is $\succeq_i^\alpha$-null.

Axiom 4 *(Continuity) For any $f \in \mathcal{F}$, the sets $\{g \mid f \succeq g\}$ and $\{g \mid f \preceq g\}$ are closed in the Euclidean topology on $\mathcal{F}$.*[19]

[18] This is the substantive restriction entailed by a multi-agent Anscombe-Aumann framework. The assumption and all the results could easily be recast in a multi-agent Savage framework. We have chosen the multi-agent Anscombe-Aumann framework for parsominy of exposition, and for its proximity to Harsanyi's original setup.

[19] If X is infinite, continuity of preference is to be understood in terms of pointwise convergence.

Axiom 5 *(Nontriviality) There exist $f, g \in \mathcal{F}$ such that $f \succ g$.*

We will also assume that agents' preferences over constant acts (lotteries) are commonly known (to be denoted by $\succeq_{i|\mathcal{L}}$), and that the agents have already agreed upon a "standard of fairness" $\succeq_{I|\mathcal{L}}$ over lotteries. The social standard $\succeq_{I|\mathcal{L}}$ is related to agents' preferences through the following Pareto condition.

Axiom 6 *(Pareto Dominance for Lotteries) $\ell \succeq_{I|\mathcal{L}} \ell'$ for any ℓ, ℓ' such that $\ell \succeq_{i|\mathcal{L}} \ell'$ for all $i \in I$.*

These data are summarized in the following definition of a type space in preferences.

Definition 3 *A* **type space in preferences** *is a tuple $\langle I, \Omega, \{\succeq_i\}_{i\in I}, \succeq_{I|\mathcal{L}}\rangle$, where*

1. *I is a finite set of agents;*

2. *Ω is a finite set of states.*

3. *For every agent $i \in I$, $\succeq_i$ is a mapping that specifies, for each state $\alpha \in \Omega$, a preference relation $\succeq_i^\alpha \subseteq \mathcal{F} \times \mathcal{F}$ that satisfies axioms 1 to 5.*

4. *(Introspection of Preferences and Truth of Beliefs). For every agent $i \in I$ and all states $\alpha, \beta \in \Omega$: β is non-null with respect to $\succeq_i^\alpha$ if and only if $\succeq_i^\beta = \succeq_i^\alpha$.*

5. *(Common Knowledge of Lottery Preferences). If $\ell \gtrsim_{i|\mathcal{L}}^\alpha \ell'$ at some $\alpha \in \Omega$, then it is common knowledge that $\ell \gtrsim_{i|\mathcal{L}}^\omega \ell'$, for any $\ell, \ell' \in \mathcal{L}$ and $i \in I$.*

6. *(Minimal Domain Richness) There exist lotteries $\{\ell_i\}_{i\in I}$ and ℓ_0 such that $\ell_i \succ_{i|\mathcal{L}} \ell_0$ and $\ell_i \sim_{j|\mathcal{L}} \ell_0$ for all $i \in I$ and $j \neq i$.*

7. *(Social Standard) $\succeq_{I|\mathcal{L}}$ is a transitive, reflexive, and continuous relation over $\mathcal{L}$ that respects Pareto Dominance For Lotteries.*

In the next two sections, we shall focus on the special case of standards $\succeq_{I|\mathcal{L}}$ that are complete and satisfy the Independence axiom; in view of Harsanyi's (1955) classic result, such standards will be called **utilitarian**.[20] The Anscombe-Aumann and Harsanyi theorems (appropriately adapted) yield the following representation result.

[20]Since the social standard $\succeq_{I|\mathcal{L}}$ is a datum in the following analysis, here Harsanyi's characterization plays the role of a qualitative *definition* of utilitarian orderings; it need not be assumed that Harsanyi's axioms *normatively justify* the class of such orderings (which is controversial). In particular, one could instead assume separability across individuals in the manner of Fleming (1952), and obtain independence as a derived property.

Proposition 2 *Let* $\langle I, \Omega, \{\succeq_i\}_{i\in I}, \succeq_{I|\mathcal{L}}\rangle$ *be a type space in preferences such that* $\succeq_{I|\mathcal{L}}$ *is utilitarian. Then there exist probability mappings* $p_i : \Omega \to \mathbf{R}^{\Omega}$ *and utility functions* $u_i : \mathbf{X} \to \mathbf{R}$ *such that, for all* $f, g \in \mathcal{F}$*,* $i \in I$*,* $\alpha \in \Omega$*,*

$$f \succeq_i^{\alpha} g \text{ if and only if } \sum_{\omega} \sum_{x\in X} p_i^{\alpha}(\omega) f_x(\omega) u_i(x) \geq \sum_{\omega} \sum_{x\in X} p_i^{\alpha}(\omega) g_x(\omega) u_i(x),$$

as well as, for all $\ell, \ell' \in \mathcal{L}$*,*

$$\ell \succeq_{I|\mathcal{L}} \ell' \text{ if and only if } \sum_{x\in X} \ell_x \sum_{i\in I} u_i(x) \geq \sum_{x\in X} \ell'_x \sum_{i\in I} u_i(x).$$

The probability mappings p_i *are unique, the utility functions* u_i *are unique up to joint positive affine transformations. Moreover, the* p_i *satisfy Introspection and Truth.*

To streamline notation, we shall write $u_i(\ell) = \sum_{x\in X} \ell_x u_i(x)$, and for any act h and agent i, and define a random variable U_i^h by setting $U_i^h(\omega) = u_i(h(\omega))$; U_i^h describes agent i's expected utility under h conditional on the state ω. With this notation, the representation of individual preferences can be simplified to

$$f \succeq_i^{\alpha} g \text{ if and only if } E_i^{\alpha} U_i^f \geq E_i^{\alpha} U_i^g.$$

6. UTILITARIANISM EX INTERIM

Almost by definition, a standard of cooperative choice will respect agents' consensual preference; under incomplete information, such consensus must be public. This is leads to the axiom of "Interim Pareto Dominance."[21]

Axiom 7 (Interim Pareto Dominance) $f \succeq_I g$ *(resp.* $f \succ_I g$*) whenever it is commonly known at* τ *that* $f \succeq_i^{\omega} g$ *(resp.* $f \succ_i^{\omega} g$*) for all* $i \in I$*.*

With respect to "Bayesian rationality" of the social ordering, it turns out to be sufficient to assume the social ordering to satisfy State Independence; this axiom entails that the utilitarian ranking $\succeq_{I|\mathcal{L}}$ can be employed state by state.

[21] Strict Interim Pareto Dominance is equivalent to "interim dominance" as defined in Holmstrom-Myerson (1983, p. 1805). However, their notion is defined without recourse to the notion of common knowledge; instead, it is motivated there informally by appealing to the idea that "a welfare economist, as an outsider, could not apply any concept of domination that depended on individuals' actual private information."

Axiom 8 (State Independence of Social Ordering) $[\ell, f_{-\alpha}] \succeq_I [\ell', f_{-\alpha}]$ *whenever* $\ell \succeq_{I|\mathcal{L}} \ell'$, *for all* $\ell, \ell' \in \mathcal{L}$, $f, g \in \mathcal{F}$ *and* $\alpha \in \Omega$.

To give full play to the force of these axioms, we need to assume that individual and group preferences are defined on a "rich" domain of acts; this assumption is discussed in detail in section 8 below. Note that if a domain of lotteries is rich, the set of social alternatives X must be infinite.

Definition 4 $\mathcal{L}$ *is* **rich** *if*

i) for all $i \in I$, $\ell, \ell' \in \mathcal{L}$ *and* $\alpha \in (0,1)$, *there exists* $\ell'' \in \mathcal{L}$ *such that* $\ell \sim_i \alpha\ell' + (1-\alpha)\ell''$.

ii) for any family of lotteries $\{\ell_i\}_{i \in I}$ *in* $\mathcal{L}$, *there exists* $\ell \in \mathcal{L}$ *such that* $\ell \sim_i \ell_i$ *for all* $i \in I$.

The following fact is easily verified.[22]

Fact 1 $\mathcal{L}$ *is rich if and only if, for any* $u \in \mathbf{R}^I$, *there exists* $\ell \in \mathcal{L}$ *such that* $u = (u_i(\ell))_{i \in I}$.

An **extension** $\succeq_I$ of the standard $\succeq_{I|\mathcal{L}}$ is a transitive, continuous relation such that $\succeq_I \supseteq \succeq_{I|\mathcal{L}}$. If agents' beliefs admit a common prior, there exists a unique extension of the utilitarian standard $\succeq_{I|\mathcal{L}}$; this extension maximizes the expected sum of agents' utilities based on the common prior.

Theorem 3 (Aggregation Theorem) *Let* $\langle I, \Omega, \{\succeq_i\}_{i \in I}, \succeq_{I|\mathcal{L}} \rangle$ *be a type space in preferences with a utilitarian standard* $\succeq_{I|\mathcal{L}}$, *a rich domain* $\mathcal{L}$, *and a common prior* μ. *Then there exists a unique extension* $\succeq_I$ *of* $\succeq_{I|\mathcal{L}}$ *satisfying Interim Pareto Dominance and State Independence; for this extension,* $f \succeq_I g$ *if and only if*

$$E_\mu \left(\sum_{i \in I} U_i^f \right) \geq E_\mu \left(\sum_{i \in I} U_i^g \right).$$

Example 2. In the context of Example 1 of section 4, assume that $X = \mathbf{R}^I$, that agents are risk-neutral with $u_i(x) = x_i$ for all $i \in I$, and that a utilitarian standard is given by $\sum_i u_i$. In natural notation, let f denote the act $(-2,-2;-2,-2;-2,-2;10,10)$, g the act $\mathbf{0}$, and h the act $(-2,-2;4,-8;-8,4;10,10)$. Note that, in the notation of example 1, h corresponds to the split of the bet Z into bets Z_1 and Z_2, while f corresponds to splitting up Z symmetrically. Since $E_\mu \left(\sum_{i \in I} U_i^f\right) > E_\mu \left(\sum_{i \in I} U_i^g\right)$, by Theorem 3, it must be the case that $f \succ_I g$. Indeed, by Interim Pareto Dominance, $h \succ_I g$; on the other hand, by State Independence, $h \sim_I f$. Thus, by transitivity, $f \succ_I g$.

[22] Recall that non-triviality of preferences is a maintained assumption.

It may seem paradoxical that f is ranked above g at the true state τ, even though at τ both agents know that they are made worse off by f than by g. However, this fact is only known but not *common knowledge* among the two agents, and thus not available to the construction of the extended standard $\succeq_I$.

Of course, this fact could be made available easily by communication among the two agents. For example, agent 1 might propose choosing g instead of f; at τ, agent 2 surely would accept this proposal. Note that if the agents stick with f in the other states, they will in effect implement the act $e = (0,0;-2,-2;-2,-2;10,10)$. Yet if this happens, e must have been feasible, after all; since e is ranked strictly above both f and g by the extended standard $\succeq_I$, e is recommended by $\succeq_I$ ahead of f, resolving the apparent paradox. The example illustrates the general point that any feasible and socially desirable communication will be realized by the optimal act (mechanism); see also remark 5 of the discussion in section 8.

The utilitarian ordering characterized in Theorem 3 describes subjective expected utility maximization at the social level. It performs a two-fold aggregation, of agents' utilities into a group utility by summation, and of agents' probabilities into a "consensus probability", the common prior. In view of Samet's characterization of the common prior (Theorem 1) and Proposition 1, one can restate the representation of the extended standard explicitly in terms of higher-order expectations as follows: $f \succ_I g$ if and only if, for some finite sequence $\{i_1, ..., i_k\}$, it is common knowledge that

$$E_{i_k}...E_{i_1}\left(\sum_{i\in I} U_i^f\right) > E_{i_k}...E_{i_1}\left(\sum_{i\in I} U_i^g\right).$$

As illustrated by the voting example of section 3, with independent types, the value of the second-order expectation $E_j E_k Z$ is commonly known for any $j \neq k$ and Z, and equal to $E_\mu Z$. Hence in this case, the extended utilitarian standard can be defined in terms of second-order expectations, i.e. $f \succeq_I g$ if and only if

$$E_j E_k\left(\sum_{i\in I} U_i^f\right) \geq E_j E_k\left(\sum_{i\in I} U_i^g\right) \text{ , for any } j \neq k.$$

7. THE COMPATIBILITY OF THE PARETO CRITERION AND STATE INDEPENDENCE

The Aggregation Theorem construes the extended social standard as subjective expected utility maximization at the social level; the common prior as the group's subjective probability measure can be viewed as an aggregating individual agents' belief from behind the "veil of public ignorance". If agents' beliefs fail to be consistent with a common prior, such aggregation is no longer possible. Not only does the common prior become unavailable as the canonical "consensus group belief", social expected utility maximization becomes incompatible with the Pareto principle. The source of the incompatibility is especially transparent in the risk-neutral case. Here, if agents' beliefs are inconsistent with a common prior, there exists a mutually profitable bet by Theorem 2. Interim Pareto Dominance requires a social preference for this bet over not betting; by contrast, State Independence entails social indifference between the two, since the bet involves a mere reshuffling of money=utility among agents in each state, a matter of indifference under a utilitarian standard.

The following result characterizes those type spaces for which "Bayesian preference aggregation" is possible. One condition can be formulated directly in terms of preference.

Axiom 9 *(No Free Lunch)* *For no acts $f, g \in \mathcal{F}$ such that $f(\omega) \sim_{I|\mathcal{L}} g(\omega)$ for all $\omega \in \Omega$, it is common knowledge that $f \succ_i^{\omega} g$ for all $i \in I$.*

"No Free Lunch" can be viewed as a preference analogue of No Betting; it asserts that it cannot be commonly known that everyone is made better off as the result of a mere reshuffling of utilities state-by-state (measured in terms of the basic ordering $\sim_{I|\mathcal{L}}$).

Theorem 4 (Possibility Theorem) *Let $\mathcal{T} = \langle I, \Omega, \{\succeq_i\}_{i \in I}, \succeq_{I|\mathcal{L}} \rangle$ be a type space in preferences such that $\succeq_{I|\mathcal{L}}$ is utilitarian. Then the following three statements are equivalent.*

1. *$\mathcal{T}$ admits an extension $\succeq_I$ that satisfies Interim Pareto Dominance and State-Independence.*
2. *Agents' beliefs admit a common prior μ.*
3. *$\mathcal{T}$ satisfies No Free Lunch.*

This result generalizes the existing (im)possibility results on Bayesian aggregation in the literature in the context of complete information (see Hylland-Zeckhauser 1979 and others quoted above) which concluded that Bayesian coherence is compatible with ordinary Pareto Dominance only if agents' beliefs are identical. According to the Possibility Theorem, however, differences in beliefs among agents do not force a choice between Bayesian coherence and respect for consensual preference per se; the conflict arises only when these differences cannot be fully attributed to differences in information in the sense that the different beliefs can be viewed as updates on a common prior. The Possibility Theorem thus removes the flair of "paradox" of the extant results; indeed, on a Harsanyian point of view on which agents' belief *should* admit a common prior as a matter of (intersubjective) rationality, it shows that Bayesian and Paretian aggregation is always possible whenever agents are fully rational in this sense.

This more positive conclusion is due not so much to incomplete information per se, but more specifically to *Interim* Pareto Dominance as the appropriate Pareto criterion; in the current type-space framework, the known impossibility results would reappear if instead the following condition of "joint improvement" was used.[23]

$$\text{For all } \alpha \in \Omega : f \succeq_I^\alpha g \text{ (resp. } f \succ_I^\alpha g) \text{ whenever } f \succeq_i^\alpha g \text{ (resp. } f \succ_i^\alpha g) \text{ for all } i \in I.$$

8. DISCUSSION

1. Assumptions of Bayesian Rationality.

Theorem 3 makes only one assumption on how group preferences deal with the "subjective" uncertainty that results from incompleteness of information, the State Independence axiom. In particular, the Theorem neither assumes completeness of the social ordering nor its satisfaction of the Independence axiom for general acts. Thus, the Theorem does not assume a priori that likelihood judgments are meaningful at a group level.[24]

2. Domain of Preferences

Theorem 3 pays a price for the weakness of its assumptions in terms of the requirement that the domain of preferences be rich. If one is willing to assume that group preferences satisfy the Independence axiom for general acts, it suffices to assume that the domain is minimally rich as shown by Corollary 1 in section 9.[25]

Moreover, the richness assumption may look stronger than it really is. For it is not necessary that agents' and the group's *actual* preferences are defined on a rich domain, only that these preferences

[23]To make the formal equivalence to the ordinary Pareto criterion exact, the individual and social rankings must be defined over the domain of all acts measurable with respect to (appropriately specified) "events of nature". Note that "joint improvement" is applicable to a social ranking that is a function of the state, and thus not appropriate here.

[24]For analyses of the role of different rationality assumptions at the individual and social level, see Mongin (1998) and Blackorby, Donaldson, Mongin (2000).

[25]There has been a fair amount of discussion of the role and appeal of richness assumptions in the context of Harsanyi's original Aggregation Theorem; see Blackorby-Donaldson-Weymark (1999) for a recent contribution.

are capable of being extended to such a domain. Specifically, it is only necessary to assume that agents' actual preferences and the agreed-upon utilitarian standard $\succeq_{I|\mathcal{L}(X)}$ be defined on a domain $\mathcal{L}(X)$ that is minimally rich, for some $X \subseteq \widehat{X}$. Now posit that agents' preferences are hypothetically extended to a larger, rich domain $\mathcal{L}(\widehat{X})$, with $X \subseteq \widehat{X}$. Then it is easily seen that $\succeq_{I|\mathcal{L}(X)}$ extends to a *unique* utilitarian standard $\succeq_{I|\mathcal{L}(\widehat{X})}$ on $\mathcal{L}(\widehat{X})$, which in turn uniquely extends to an extended standard $\succeq_I$ on $\mathcal{F}(\widehat{X})$ satisfying the assumptions of Theorem 3. The point is that the induced extended standard $\succeq_{I|\mathcal{F}(X)}$ on the actual domain $\mathcal{F}(X)$ does not depend on preferences outside X; thus only the possibility of such an embedding really matters.

3. The role of incentive-incompatible acts.

A crucial feature of our setup is the assumption that the individual and social preferences pertain to the set of *all* formally well-defined social acts $\mathcal{F}(X)$. Substantively, this means that each act is viewed as "conceivable", i.e. as consistent with a well-defined thought experiment. Including incentive incompatible acts in the domain in particular may be viewed by some as problematic. However, following Holmstrom-Myerson (1983), incentive-compatibility constraints are appropriately viewed as *feasibility* constraints of a particular sort, resulting from the factual infeasibility of getting individuals to report their beliefs truthfully. To view incentive-incompatible acts as *inconceivable,* rather than merely unfeasible, would surely misread their significance. Indeed, there *are* sensible thought-experiments supporting arbitrary acts. In particular, arbitrary acts can be construed as describing conditionals of the form "If individual *were to* report truthfully their belief hierarchies as ... and the outcome of the horse race is ..., then the following roulette lottery is run".

Moreover, infeasible acts are used in the construction of the extended social standard merely as "vehicles of reasoning". The ranking of all acts, feasible or not, is pinned down by the axioms and the standard of fairness; there is no need to concretely imagine and independently evaluate the counterfactual world corresponding to infeasible acts.[26] The role of the infeasible acts is to give full play to the normative content of the axioms; this indeed is the role of considering counterfactual situations in normative axiomatic theories quite generally.[27]

[26]For instance, in example 2, the act h may well be incentive-incompatible; it plays the role of a thought-experiment from which the group can deduce that the act f is socially preferable to g.

[27]An altogether different criticism might want to deny the normative validity of some of the axioms when incentive-incompatible acts are involved. Such a view would (implausibly) assume that incentive-compatibility matters directly to the social evaluation of social acts; in particular, it would not be compatible with viewing incentive-compatibility simply as a feasibility constraint.

4. Common Knowledge of Agents' Lottery Preferences and State Independence.

Condition 5 in the definition of a type space assumes agents' preferences over lotteries to be commonly known; it thus requires that all intersubjective uncertainty about other agents' preferences can be attributed to uncertainty about their beliefs. Without this assumption, State Independence of the social ordering would conflict with Interim Pareto Dominance.

This assumption is clearly restrictive, although much less so than it may seem at first sight. For example, causal factors of an agent's utility can frequently be recast as arguments; e.g., the utility of a certain amount of wealth given a state of health can be recast as the utility of a pair (amount of wealth, level of health), analogously to the standard way of accommodating prima-facie violations of state independence in single-person decision theory. Indeed, condition 5 can be viewed as state-independence of agents' (unconditional) preferences *across* the types of one agent, paralleling axiom 3 which requires state-independence of agents' (conditional) preferences *within* any type. Note in particular that under a dynamic asymmetric information interpretation, if condition 5 is violated for some agent, then this agent's ex-ante preference relation must be state-dependent.

Condition 5 and axiom 3 can be weakened substantially, as they need to hold only for a rich subdomain of lotteries $\mathcal{L}(X')$, with $X' \subseteq X$. Consider, for example, incomplete information about agents' quasi-linear valuation of a public project $y \in Y$: letting $X = Y \times \mathbf{R}^I$, assume that agents' preferences are given by a utility function $u_i^\alpha(y, t_i) = w_i^\alpha(y) + t_i$, and that, for any $y^* \in Y$, a utilitarian standard over $\mathcal{L}(y^* \times \mathbf{R}^I)$ is given which evaluates these lotteries in terms of $\sum_{i\in I} t_i$. Since agents' preferences over $\mathcal{L}(y^* \times \mathbf{R}^I)$ are commonly known, it makes sense to demand State Independence of the extended social standard $\succeq_{I|\mathcal{F}(y^* \times \mathbf{R})}$ on $\mathcal{F}(y^* \times \mathbf{R}^I)$. Theorem 3 leads to a utilitarian ranking on this subdomain of acts with the representation $E_\mu\left(\sum_{i\in I} t_i\right)$. A straightforward argument from interim Pareto indifference yields a uniquely defined extension on all of $\mathcal{F}(X)$ with the representation $E_\mu\left(\sum_{i\in I}(w_i(y) + t_i)\right)$.[28]

[28] It is worth comparing the crucial role of agent-level state-independence assumptions (condition 5 and axiom 3) in the present paper to their role in the closest contributions in the literature, Harsanyi-Selten's (1972) and Myerson's (1984). On one side, Harsanyi-Selten (1972) defined acts in utility-payoffs, and thus effectively build the agent-level state-independence assumptions into their framework but do not thematize them. By contrast, Myerson (1984) admits general state-dependent preferences; this indeed is central to his critique of the Harsanyi-Selten (1972) proposal, for Myerson argues that in a general state-dependent framework, probabilities cannot be defined separately from utilities, which leads him to require a "probability-invariance axiom". Since this requirement can only be formulated in a "multi-profile" (or more properly: multi-type-space) setting, it does not apply to the present analysis which deals with a particular, given type space only, just as Harsanyi's (1955) original Aggregation Theorem deals with a single preference profile.

5. Group Preference, Group Choice and the Role of Communication

The Aggregation Theorem delivers an axiomatic account of (utilitarian) group *preference* under incomplete information; group *choice* is to be understood simply as maximization of group preference under feasibility constraints. One scenario of group choice involves an impartial mediator who a) has complete control over all communication between agents, and b) who knows all that is common knowledge among the group. By (a), the mediator can select *any* incentive-compatible, individually-rational mechanism; by (b), the mediator can deduce the optimal mechanism from his information and run it. In other scenarios, a wide variety of further feasibility constraints may be relevant; these may result, for example, from incomplete control over agents' communication, the possibility of renegotiation, group participation constraints in the manner of the core, considerations of contract simplicity etc. .

The distinction between group preference, feasibility and group choice explains why little mention has been made before of communication among agents. This is indeed may come to mind as the natural response to agents' lack of knowledge about each other. However, one must bear in mind that such communication must be credible, that is: respect incentive-compatibility constraints. Moreover, any socially desirable, incentive-compatible communication will be built into the optimal mechanism as a matter of group *choice*.[29] It may, of course, simply happen sometimes that some agents try to "leak" some of their private information to others in the hope of influencing the group decision in a favorable manner; if they are successful, this just means that the informational basis for collective choice has changed.[30]

6. Group Choice versus Welfare Interpretation

Our results can also be viewed normatively as deriving "social welfare rankings" under incomplete information. Specifically, the Aggregation Theorem captures the spirit of the "*ex-ante* school" adapted to this context; the hallmark of the ex-ante school is the unrestricted acceptance of the Pareto principle applied to agents' preferences at the time of the group decision, and hence incorporating their current (here: their interim) beliefs. By contrast, the "ex-post school" subscribes to the Pareto principle only state by state; it typically interprets the social ordering as that of an "impartial observer" or "benevolent dictator" with beliefs of his own. For the ex-post school, incompleteness

[29] On the other hand, communication cannot be directly relevant to the "construction" of group *preferences*, since those preferences must be defined (as a matter of logic) relative to a given state of information, and communication works by changing the state of information.

[30] Whether or not successful "cheap talk" among agents is indeed possible in equilibrium depends on the situation and deserves further study; Holmstrom and Myerson (1983) discuss a related issue under the heading of "durability".

of information among agents presents no new normative issues of interest, since, in any case, it is the observer's beliefs that count.[31]

Interim welfare comparisons that respect Interim Pareto Dominance are naturally interpreted in terms of hypothetical group choices, i.e. as determining what the agents *should* choose *if* the state-contingent allocation was a matter of collective choice (rather than, say, non-cooperative interaction). Indeed, under incomplete information, an interpretation of interim welfare judgements in terms of *some* well-defined (hypothetical) choice situation under uncertainty seems necessary to pin down their meaning unambiguously by establishing a well-defined informational basis; otherwise, it would simply not be clear from what beliefs these welfare judgments could be derived. On the group choice interpretation, this informational basis is the sum total of everything that is commonly known by the group.

Due to these epistemic considerations, we view the group choice interpretation as the conceptually more fundamental one, and have adopted it as the "official" interpretation of the paper.

9. PARTIAL UTILITARIANISM

So far, we have maintained the assumption that the agreed-upon standard of fairness is complete. This is clearly demanding; indeed, it seems plausible that in many cases a group of agents' will only be able to agree upon some social comparisons, leaving many others undecided. In a utilitarian context, incompleteness of the social standard may result from limitations in the range of *consensual* interpersonal comparison of utilities. For example, while everyone may concede that it is socially preferable to allocate one marginal dollar to i (who has an income of \$10,000) rather than to j with an income of \$40,000, there may be disagreement as to whether it is socially preferable to give the dollar to i rather than to k with an income of \$20,000, and whether to give it to k rather than to j. In this section, we will therefore consider "partial utilitarian" standards and generalize the Aggregation Theorem to them. To achieve this, it is necessary to substantially strengthen the requirements of Bayesian rationality of the social ordering. Of special interest is the case in which the group abstains from any interpersonal utility comparison, i.e. in which $\succeq_{I|\mathcal{L}}$ is simply the Pareto standard (the Pareto dominance relation over lotteries) denoted by $\trianglerighteq_{I|\mathcal{L}}$; for this case, the generalized Aggregation

[31]The ex-post approach relies on a clearcut dichotomy between value and belief; as Broome (1991) and Mongin (1998) have pointed out among others, it seems doubtful whether value can ever purged from belief entirely, even in principle. The conceptual coherence of underwriting the Pareto principle ex post but not ex ante is thus put into question.

Theorem entails a justification of (as-if) *ex-ante* Pareto efficiency on ex-interim grounds.

Recall that any social standard is defined on a minimally rich domain and satisfies transitivity, reflexivity, continuity, as well as Pareto Dominance for Lotteries. In view of the following representation result, a social standard $\succeq_{I|\mathcal{L}}$ can thus be defined as **partial utilitarian** if it satisfies the Independence axiom as well. Partial utilitarian standards can be thought of as determining the scaling of an agents' utilities (more precisely: utility differences) in the social ordering up to a range of indeterminacy expressed by a closed, convex set of scaling vectors $\gamma \in \Delta(I)$.

Proposition 3 *Fix any vNM utility functions u_i for agents $i \in I$. Then the social standard $\succeq_{I|\mathcal{L}}$ is partial utilitarian if and only if there exist a (unique) closed convex set $\Gamma \subseteq \Delta(I)$ such that $\ell \succeq_{I|\mathcal{L}} \ell'$ if and only if $\sum_{i \in I} \gamma_i u_i(\ell) \geq \sum_{i \in I} \gamma_i u_i(\ell')$ for all $\gamma \in \Gamma$.*

Since the proof of this Proposition is analogous to Bewley's (1986) representation of incomplete SEU preferences in an Anscombe-Aumann (1963) setting, it is omitted here; note that the set Γ will depend on the choice of the u_i's. Note also that the Pareto standard $\trianglerighteq_{I|\mathcal{L}}$ corresponds to a maximally indeterminate set of scaling vectors $\Gamma = \Delta(I)$.

The natural generalization of the Theorem 3 yields the following "extended partial utilitarian" representation: $f \succeq_I g$ if and only if

$$E_\mu\left(\sum_{i \in I} \gamma_i U_i^f\right) \geq E_\mu\left(\sum_{i \in I} \gamma_i U_i^g\right) \text{ for all } \gamma \in \Gamma. \tag{2}$$

If $\Gamma = \Delta(I)$, (2) defines the "extended Pareto standard" to be denoted by $\trianglerighteq_I$. This standard is substantially more informative (that is: richer as a relation) than comparisons in terms of interim Pareto dominance. Evidently, on a dynamic interpretation of the common prior μ, the extended Pareto standard amounts to a comparison of acts in terms of *ex-ante* Pareto dominance.

To achieve the desired characterization, the assumptions on the Bayesian rationality of the social ordering need to be strengthened substantially. Indeed, it follows immediately from the last observation that one cannot characterize extended partial utilitarian standards by simply dropping the completeness assumption on the social standard in Theorem 3, since the Interim Pareto Dominance and State Independence axioms then no longer determine a unique extension. To see this, just observe that both the extended Pareto standard and the Interim Dominance relation extend the standard $\trianglerighteq_{I|\mathcal{L}}$ and satisfy State Independence.

While helpful, the Independence axiom is not enough, since Independence as well is satisfied by both the extended Pareto standard and the Interim Dominance relation. The following additional axiom does the trick; it can be viewed as an analogue of Savage's axiom P4 in an Anscombe-Aumann setting. For a random-variable $V : \Omega \to [0,1]$ and lotteries $\ell, \ell' \in \mathcal{L}$, denote by $V\ell + (1-V)\ell' \in \mathcal{F}$ the act f such that $f(\omega) = V(\omega)\ell + (1-V)(\omega)\ell'$ for $\omega \in \Omega$.

Axiom 10 (Comparative Likelihood) *For any random-variables* $V, W : \Omega \to [0,1]$ *and lotteries* $\ell, \ell', \ell'', \ell''' \in \mathcal{L}$ *such that* $\ell \succ_{I|\mathcal{L}} \ell'$ *and* $\ell'' \succ_{I|\mathcal{L}} \ell'''$: $V\ell + (1-V)\ell' \succeq_I W\ell + (1-W)\ell'$ *if and only if* $V\ell'' + (1-V)\ell''' \succeq_I W\ell'' + (1-W)\ell'''$.

The Comparative Likelihood axiom can be interpreted as follows. Consider the axiom first in the familiar context of individual decision making, in which the relation $\succeq_I$ describes the preferences of a certain Ms. I, and suppose, for the purpose of the argument, that one can one can meaningfully attribute beliefs to her that *underlie* her preferences and are *revealed* by them (i.e. by her choice behavior).[32] For two random variables V and W, compare the associated acts $V\ell + (1-V)\ell'$ and $W\ell+(1-W)\ell'$; these two acts are distinguished only in the likelihood with which they result in one of the two lotteries ℓ and ℓ' in different states. As a Bayesian decision maker, Ms. I will prefer the one to which she assigns the higher (marginal) likelihood of generating the superior lottery ℓ; thus, her preference for $V\ell+(1-V)\ell'$ over $W\ell+(1-W)\ell'$ reveals her belief that this likelihood is higher under the first act than under the second. Note that if Ms. I has subjective probability measure η, she will (weakly) prefer to bet on V rather than on W if and only if $E_\eta V \geq E_\eta W$. In the special case in which the random variables are indicator functions of the form $V = 1_A$ and $W = 1_B$, the preference of betting on 1_A over betting on 1_B reveals her belief that the event A is more likely than the event B. The Comparative Likelihood axiom just introduced simply requires that the revealed likelihood comparisons do not depend on the underlying lotteries.

In the present context, $\succeq_I$ refers to a social rather than individual ordering. It clearly does not make sense to posit a preexisting "group belief" that is merely *revealed* by the ordering $\succeq_I$. Instead, the Comparative Likelihood axiom needs to be read as follows. *If* the group comes to judge $V\ell + (1-V)\ell'$ as superior to $W\ell + (1-W)\ell'$, it *thereby judges* the likelihood of obtaining ℓ to be higher under the first act than under the second. On the basis of this *implied* likelihood judgment, it follows that $V\ell'' + (1-V)\ell'''$ should be ranked above $W\ell'' + (1-W)\ell'''$.

The following results delivers an Aggregation Theorem for partial utilitarian standards.

[32] We note that while Savage (1972) introduces the notion of personal probability along such lines, in the paraphrase of P4 he carefully avoids any reference to notions of "belief" and "revelation", eschewing any explicit articulation of the *rationale* for that axiom.

Theorem 5 *Let $\langle I, \Omega, \{\succeq_i\}_{i \in I}, \succeq_{I|\mathcal{L}} \rangle$ be a type space in preferences with common prior μ and a partial utilitarian standard $\succeq_{I|\mathcal{L}}$. Then the extended partial utilitarian standard $\succeq_I$ given by (2) is the unique extension satisfying Interim Pareto Dominance, Independence and Comparative Likelihood.*

While Comparative Likelihood does not entail State Independence formally, it clearly presupposes the latter conceptually; for if State Independence is violated, the preference between the lotteries considered by the Comparative Likelihood axiom would not be *exclusively* be a matter of comparing likelihoods. Conversely, if the ordering $\succeq_I$ is complete, Comparative Likelihood is entailed by State Independence and Independence. This is stated in the following Fact, which really is a result about single-agent decision making under uncertainty with incomplete preferences.

Fact 2 *If $\succeq_{I|\mathcal{L}}$ is complete, any extension $\succeq_I$ satisfying State Independence and Independence satisfies Comparative Likelihood as well.*

Fact 2 entails the following Corollary which shows that if one assumes Independence of the extension, one can make do with a minimally rich domain.

Corollary 1 *Let $\langle I, \Omega, \{\succeq_i\}_{i \in I}, \succeq_{I|\mathcal{L}} \rangle$ be a type space in preferences with a utilitarian standard $\succeq_{I|\mathcal{L}}$ and common prior μ. Then there exists a unique extension $\succeq_I$ of $\succeq_{I|\mathcal{L}}$ satisfying Interim Pareto Dominance, State Independence, and Independence, with $f \succeq_I g$ if and only if*

$$E_\mu \left(\sum_{i \in I} U_i^f \right) \geq E_\mu \left(\sum_{i \in I} U_i^g \right).$$

Simplifying the Distribution Problem: Bayesian Interim Efficiency

The extended Pareto standard $\trianglerighteq_I$ summarizes all welfare comparisons that can be made ex interim in the absence of utility comparisons across agents while exploiting the comparability of utility across the types of a given agent. This motivates the following notion of "Bayesian interim efficiency": a feasible f is **Bayesian interim efficient** if it is not strictly inferior to some feasible g in terms of the extended Pareto standard $\trianglerighteq_I$. Bayesian interim efficiency is clearly equivalent to ex-ante Pareto efficiency; Theorem 5 applied to the Pareto standard $\trianglerighteq_{I|\mathcal{L}}$ thus offers a justification for selecting ex-ante Pareto efficient allocations on the basis of an *ex-interim* rationale. While interim Pareto efficiency leaves open the distribution of utility among *types* of agents, Bayesian interim efficiency reduces the distribution problem to one among agents. The distribution problem among the different types of one agent is eliminated, in effect, by assigning a definite weight to each type, namely its prior probability, and by exploiting the built-in intrapersonal comparability of utility across types.

The move from interim Pareto efficiency to Bayesian interim efficiency has substantial economic content. For example, while property rights in the form of participation constraints can never lead to interim Pareto inefficiency, they easily entail Bayesian interim inefficiencies. Specifically, for any agent i, let g_i denote an act describing his best outside alternative, should he choose not to participate in the mechanism. Then f is **individually rational** if it is common knowledge that $f \succeq_i^\omega g_i$ for all $i \in I$. For a given collection $\{g_i\}_{i \in I}$, denote by $\mathcal{F}^{ir}$ the set of all individually rational acts, and $\mathcal{G}$ the set of acts satisfying all other feasibility constraints. If f is interim Pareto efficient in $\mathcal{G} \cap \mathcal{F}^{ir}$, then f is evidently interim Pareto efficient in $\mathcal{G}$ as well. By contrast, if f is Bayesian interim efficient in $\mathcal{G} \cap \mathcal{F}^{ir}$, then f need not be Bayesian interim efficient in $\mathcal{G}$, and indeed typically will not be. For example, in the context of trading an indivisible good among two risk-neutral agents, Myerson-Satterthwaite (1983) have shown that, except for degenerate cases, no acts (mechanisms) in $\mathcal{G} \cap \mathcal{F}^{ir}$ are ex-post efficient, with $\mathcal{G}$ denoting the class of all physically feasible and incentive compatible acts. On the other hand, it follows from the results of D'Aspremont-Gerard-Varet (1979) that all Bayesian interim efficient acts in $\mathcal{G}$ are ex-post efficient; hence no acts in $\mathcal{G} \cap \mathcal{F}^{ir}$ are Bayesian interim efficient in $\mathcal{G}$. In this way, the notion of Bayesian interim efficiency allows one to argue that the assignment of property rights can lead to welfare losses.

10. CONCLUSION

We have developed a theory of norm-based cooperation by a group of agents. The present contribution is limited by two assumptions in particular: that the standard of fairness is (partially) utilitarian, and that agents' beliefs admit a common prior. Under these assumptions, the Aggregation Theorem asserts that social evaluation ex interim is equivalent to a social evaluation from a (possibly fictitious) ex-ante viewpoint where all incompleteness information has been removed. This "ex-ante transformation" has been generalized to partially utilitarian standards in section 9, for which the Bayesian rationality of group preferences was crucial. The ex-ante transformation in effect eliminates the distribution problem among the types of any one agent. This makes compelling conceptual sense: indeed, already from a purely logical point of view it seems implausible that there should exist a genuine problem of distribution among types, for "types" really are events, not agent-like entities of some kind, and therefore do not represent genuine claimants of welfare.

If this is correct, it should be possible to generalize the "ex-ante transformation" of collective choice problems to situations in which the standard of fairness is non-utilitarian and/or in which beliefs are inconsistent with the common prior. Additional issues arise since the Aggregation and Possibility Theorems force a choice between social EU maximization and Interim Pareto Dominance in these cases. Moreover, if the Pareto criterion is maintained (as seems appropriate under a cooperative interpretation), it is not obvious how to modify/replace social EU maximization. These issues shall be tackled in future work on non-utilitarian standards and inconsistent beliefs, respectively (Nehring 2002a and Nehring 2002b).

APPENDIX: PROOFS

Proof of Theorem 2,ii).

To verify necessity, one infers from acceptability that $E_\mu Z_i > 0$ for all i, and thus $E_\mu Z = \sum_{i\in I} E_\mu Z_i > 0$.

For the converse, let $I = \{1, ..., n\}$ and define a sequence of $n-$tuples of random variables $\{(Z_i^k)_{i\in I}\}_{k=1,..,\infty}$ as follows. First, set $Z_1^1 = Z$ and $Z_i^1 = 0$ for $i \neq 1$, and let $i_k = k \bmod n$. Given $(Z_i^k)_{i\in I}$, define $(Z_i^{k+1})_{i\in I}$ by setting

$$Z_i^{k+1} := \begin{cases} Z_i^k - E_{i_k} E_{i_{k-1}} ... E_1 Z & \text{if } i = i_k \\ Z_i^k + E_{i_k} E_{i_{k-1}} ... E_1 Z & \text{if } i = i_{k+1} \\ Z_i^k & \text{if } i \neq i_k, i_{k+1} \end{cases}.$$

One easily verifies inductively that $\sum_{i\in I} Z_i^k = 0$ for all k.

Lemma 1 *i) For all i, k such that $i \neq i_k : E_i Z_i^k = 0$.*

ii) For all i, k such that $i = i_k : E_i Z_i^k = E_{i_k} E_{i_{k-1}} ... E_1 Z$.

We shall demonstrate the Lemma by induction.

For $k = 1$, it the Lemma holds by construction. Assume it to hold for some $k \geq 1$.

Part i) holds for $k+1$ and $i \neq i_k, i_{k+1}$, since then $Z_i^{k+1} = Z_i^k$ by construction, and thus $E_i Z_i^{k+1} = E_i Z_i^k = 0$ by validity of part i) for k.

Part i) holds for $k + 1$ and $i = i_k$, since then $Z_{i_k}^{k+1} = Z_i^k - E_{i_k} E_{i_{k-1}} ... E_1 Z$ by definition, hence

$$E_{i_k} Z_{i_k}^{k+1} = E_{i_k} Z_i^k - E_{i_k} E_{i_k} E_{i_{k-1}} ... E_1 Z = 0$$

by validity of part ii) for k.

Part ii) holds for $k + 1$ and $i = i_{k+1}$, since $Z_{i_{k+1}}^{k+1} = Z_{i_{k+1}}^k + E_{i_k} E_{i_{k-1}} ... E_1 Z$ by definition, hence

$$E_{i_{k+1}} Z_{i_{k+1}}^{k+1} = E_{i_{k+1}} Z_{i_{k+1}}^k + E_{i_{k+1}} E_{i_k} E_{i_{k-1}} ... E_1 Z = E_{i_{k+1}} E_{i_k} ... E_1 Z$$

by validity of part i) for k. $\square$

The sequence $\{i_k\}_{k=1,...,\infty}$ is an $I-$sequence. Hence, by Theorem 1, using linearity, $\lim_{k\to\infty} E_{i_k}^\omega E_{i_{k-1}} ... E_1 Z = E_\mu Z > 0$ for all $\omega \in \Omega$. It follows that if k is chosen suitably large (as equal to "K") and $\epsilon > 0$ sufficiently small, $E_{i_K}^\omega E_{i_{K-1}} ... E_1 Z > (n-1)\epsilon > 0$ for all $\omega \in \Omega$.

Set $\widetilde{Z}_i = \begin{cases} Z_i^K - (n-1)\epsilon\mathbf{1} & \text{if } i = i_K \\ Z_i^K + \epsilon\mathbf{1} & \text{if } i \neq i_K \end{cases}$.

Clearly $\sum_{i\in I}\widetilde{Z}_i=0$, since $\sum_{i\in I}Z_i^K=0$. In addition, by Lemma 1, it is common knowledge that $E_i\widetilde{Z}_i>0$ for all i, demonstrating the acceptability of Z. $\square$

Proof of Theorem 3.

Necessity is straightforward.

To prove sufficiency, consider two acts $f,g\in\mathcal{F}$ such that $E_\mu\sum_i U_i^f>E_\mu\sum_i U_i^g$.

Let $Z=\sum_i U_i^f-\sum_i U_i^g$; by Theorem 2, ii) there exist $\{Z_i\}_{i\in I}$ such that $\sum_i Z_i=Z$ and such that it is common knowledge that $E_iZ_i>\mathbf{0}$. Hence by domain richness and Fact 1, there exists some act $h\in\mathcal{F}$ such that $U_i^h=U_i^g+Z_i$ for $i\in I$. By Interim Pareto Dominance, evidently

$$h\succ_I g.$$

On the other hand, since $\sum_i U_i^h=\sum_i U_i^g+\sum_i Z_i=\sum_i U_i^f$, $h(\omega)\sim_{I|\mathcal{L}}f(\omega)$ for all $\omega\in\Omega$, hence by State Independence

$$h\sim_I f.$$

By transitivity,

$$f\succ_I g.$$

By continuity, this also implies that $f\succeq_I g$ whenever $E_\mu\sum_i U_i^f\geq E_\mu\sum_i U_i^g$. $\square$

Proof of Theorem 4.

It is straightforward that (2) implies (1). We shall show that the negation of (2) implies the negation of (3) which in turn implies the negation of (1).

Assume thus that agent's beliefs do not admit a common prior. By Theorem 2, i), there exists a vector $(Z_i)_{i\in I}$ such that a) $\sum_{i\in I}Z_i=0$ and such that b) it is common knowledge that $E_iZ_i>0$ for all $i\in I$. By Minimal Domain Richness, the range of $\ell\longmapsto(u_i(\ell))_{i\in I}$ has a non-empty interior. Hence it can be assumed w.l.o.g. that there exist acts f and g in $\mathcal{F}$ such that $U_i^f=Z_i$ and $U_i^g=\mathbf{0}$ for all $i\in I$. The pair f and g establishes the existence of a "free lunch" in view of a) and b).

Indeed, by b), it is common knowledge that $f\succ_i^\omega g$. Hence by Interim Pareto Dominance

$$f\succ_I g.$$

On the other hand, by a) and the existence of a utilitarian standard represented by $\sum_i u_i$, $f(\omega)\sim_{I|\mathcal{L}}g(\omega)$. Hence by State Independence

$$f\sim_I g.$$

Thus, this free lunch (and, obviously, the existence of any free lunch) entails the incompatibility of Interim Pareto Dominance and State Independence. $\square$

Proof of Fact 2.

Fact 2 can be obtained as a corollary from Bewley's (1986) SEU representation under the assumptions of the Proposition in terms of convex sets of probabilities. A direct proof is surprisingly involved. Below, we give only a semi-direct proof, appealing to the vNM representation of preferences of lotteries.

Consider random variables $V, W : \Omega \to [0,1]$ and lotteries $\ell, \ell', \ell'', \ell''' \in \mathcal{L}$ such that $\ell \succ_{I|\mathcal{L}} \ell'$, $\ell'' \succ_{I|\mathcal{L}} \ell'''$ and $V\ell + (1-V)\ell' \succeq_I W\ell + (1-W)\ell'$. We need to show that $V\ell'' + (1-V)\ell''' \succeq_I W\ell'' + (1-W)\ell'''$.

Since social preferences over lotteries satisfy Completeness , Continuity and Independence, they are represented by some social expected-utility functional $H : \mathcal{L} \to \mathbf{R}$. Define $q \in (0,1)$ implicitly by the equation

$$q\,(H(\ell) - H(\ell')) = (1-q)\left(H(\ell'') - H(\ell^{'''})\right).$$

By comparing social expected utilities, one easily verifies that, for any $v, w \in [0,1]$,

$$q\,[v\ell + (1-v)\ell'] + (1-q)\left[w\ell'' + (1-w)\ell^{'''}\right] \sim_{I|\mathcal{L}} q\,[w\ell + (1-w)\ell'] + (1-q)\left[v\ell'' + (1-v)\ell^{'''}\right].$$

Hence, by State Independence,

$$q\,[V\ell + (1-V)\ell'] + (1-q)\,[W\ell'' + (1-W)\ell'''] \sim_I q\,[W\ell + (1-W)\ell'] + (1-q)\,[V\ell'' + (1-V)\ell'''].$$

By the assumptions of the Fact and Independence, one has

$$q\,[W\ell + (1-W)\ell'] + (1-q)\,[V\ell'' + (1-V)\ell'''] \preceq_I q\,[V\ell + (1-V)\ell'] + (1-q)\,[V\ell'' + (1-V)\ell'''].$$

Hence, by Transitivity,

$$q\,[V\ell + (1-V)\ell'] + (1-q)\,[W\ell'' + (1-W)\ell'''] \preceq_I q\,[V\ell + (1-V)\ell'] + (1-q)\,[V\ell'' + (1-V)\ell'''].$$

By Independence again, one infers

$$[W\ell'' + (1-W)\ell'''] \preceq_I [V\ell'' + (1-V)\ell'''],$$

as needed to be shown. $\square$

Proof of Theorem 5.

Again, necessity is easily verified.

For the converse, the following lemma is crucial. It allows to take ex-ante expectations of binary acts.

Lemma 2 *For any* $V : \Omega \rightarrow [0,1]$ *and lotteries* $\ell, \ell' \in \mathcal{L}$ *such that* $\ell \succ_{I|\mathcal{L}} \ell'$:
$V\ell + (1-V)\ell' \sim_I (E_\mu V)\,\ell + (1 - E_\mu V)\ell'$.

Fix any lotteries $\ell, \ell' \in \mathcal{L}$ such that $\ell \succ_{I|\mathcal{L}} \ell'$. By Minimal Domain Richness, there also exist lotteries $\{\ell_i\}_{i \in I}$ and ℓ_0 such that $\ell_i \succ_i \ell_0$ and $\ell_i \sim_j \ell_0$ for all $i \in I$ and $j \neq i$. Take some $i_1 \in I$ and any $V : \Omega \rightarrow [0,1]$. By construction, it is commonly known that

$$V\ell_{i_1} + (1-V)\ell_0 \sim_j^\omega (E_{i_1} V)\,\ell_{i_1} + (1 - E_{i_1} V)\ell_0.$$

For $j = i_1$, this follows from the identity $E_{i_1} E_{i_1} V = E_{i_1} V$; for $j \neq i_1$, this follows from the fact that $\ell_{i_1} \sim_j \ell_0$.

Hence by Interim Pareto Dominance

$$V\ell_{i_1} + (1-V)\ell_0 \sim_I (E_{i_1} V)\,\ell_{i_1} + (1 - E_{i_1} V)\ell_0.$$

Since by Interim Pareto Dominance also $\ell_{i_1} \succ_{I|\mathcal{L}} \ell_0$, it follows from Comparative Likelihood that

$$V\ell + (1-V)\ell' \sim_I (E_{i_1} V)\,\ell + (1 - E_{i_1} V)\ell'.$$

By Transitivity, $k-$fold iteration of this argument implies for any sequence $\{i_1, ..., i_k\}$,

$$V\ell + (1-V)\ell' \sim_I (E_{i_k} ... E_{i_1} V)\,\ell + (1 - E_{i_k} ... E_{i_1} V)\ell'.$$

Hence, by Continuity and Proposition 1, for any $I-$sequence $s = \{i_1, i_2, ...\}$,

$$V\ell + (1-V)\ell' \sim_I (E_s V)\,\ell + (1 - E_s V)\ell'.$$

Since $E_s V = E_\mu V \cdot \mathbf{1}$ by Theorem 1, the Lemma follows. $\square$

Clearly, it is w.l.o.g. to assume that the lotteries ℓ_i defined above satisfy, for all $i \in I$, $u_i(\ell_i) = 1$ and $u_i(\ell_j) = 0$ for all $j \in I \cup \{0\}$ such that $j \neq i$.

Take any f and g in $\mathcal{F}$. By Independence, one can clearly assume w.l.o.g. that $0 \leq U_i^f(\omega) \leq \frac{1}{n}$ and $0 \leq U_i^g(\omega) \leq \frac{1}{n}$ for all ω.

Let $f_{[i]}$ denote the act $\left(nU_i^f\right)\ell_i + \left(1 - nU_i^f\right)\ell_0$; note that by construction, $U_i^{f_{[i]}} = nU_i^f$ and $U_j^{f_{[i]}} = \mathbf{0}$ for $j \neq i$. Hence it is common knowledge that $f \sim_j^\omega \sum_{i\in I} \frac{1}{n} f_{[i]}$, whence by Interim Pareto Dominance

$$f \sim_I \sum_{i\in I} \frac{1}{n} f_{[i]}.$$

Since by the Lemma $f_{[i]} \sim_I \left(E_\mu nU_i^f\right)\ell_i + \left(1 - E_\mu nU_i^f\right)\ell_0$, by $n-$fold application of Independence one obtains

$$f \sim_I \sum_{i\in I} \left(E_\mu U_i^f\right)\ell_i + \left(1 - \sum_{i\in I} E_\mu U_i^f\right)\ell_0, \tag{3}$$

as well as

$$g \sim_I \sum_{i\in I} \left(E_\mu U_i^g\right)\ell_i + \left(1 - \sum_{i\in I} E_\mu U_i^g\right)\ell_0 \tag{4}$$

by the same token.

Since $u_i(\sum_{j\in I} \left(E_\mu U_j^f\right)\ell_j + \left(1 - \sum_{j\in I} E_\mu U_j^f\right)\ell_0) = E_\mu U_i^f$, by Proposition 3,

$$\sum_{i\in I} \left(E_\mu U_i^f\right)\ell_i + \left(1 - \sum_{i\in I} E_\mu U_i^f\right)\ell_0 \succeq_{I|\mathcal{L}} \sum_{i\in I} \left(E_\mu U_i^g\right)\ell_i + \left(1 - \sum_{i\in I} E_\mu U_i^g\right)\ell_0$$

if and only if

$$\sum_{i\in I} \gamma_i E_\mu U_i^f \geq \sum_{i\in I} \gamma_i E_\mu U_i^g \text{ for all } \gamma \in \Gamma.$$

By Transitivity, we can thus infer from (3) and(4) that $f \succeq_I g$ if and only if $E_\mu\left(\sum_{i\in I} \gamma_i U_i^f\right) \geq E_\mu\left(\sum_{i\in I} \gamma_i U_i^g\right)$ for all $\gamma \in \Gamma$. $\square$

REFERENCES

[1] Anscombe, F. G. and R. Aumann (1963), A Definition of Subjective Probability, *Annals of Mathematical Statistics,* 34, 199-205.

[2] Armbruster, W. and W. Boege (1979), Bayesian Game Theory, in *Game Theory and Related Topics,* ed. by O. Moeschlin and D. Pallaschke, Amsterdam, North-Holland.

[3] Arrow, K. (1979), The Property Rights Doctrine and Demand Revelation under Incomplete Information, in M. Boskin (ed.), *Economics and Human Welfare,* Academic Press, New York.

[4] Aumann, R. (1976), Agreeing to Disagree, *Annals of Statistics,* 4, 1236-1239.

[5] Bewley, T. F. (1986): "Knightian Decision Theory, Part I," Cowles Foundation Discussion Paper No. 807.

[6] Blackorby, C., Donaldson, D. and P. Mongin (2000), Social Aggregation Without the Expected Utility Hypothesis, mimeo.

[7] Blackorby, C., Donaldson, D. and J. Weymark (1999), Harsanyi's Social Aggregation Theorem for State-Contingent Alternatives, *Journal of Mathematical Economics,* 32, 365-387.

[8] Bonanno, G. and K. Nehring (1999), How to Make Sense of the Common Prior Assumption under Incomplete Information, *International Journal of Game Theory,* 28, 409-434.

[9] Brandenburger, A. and E. Dekel (1986), Hierarchies of Beliefs and Common Knowledge, *Journal of Economic Theory,* 59, 189-198.

[10] Broome, J. (1991), *Weighing Goods,* Basil Blackwell, Oxford.

[11] D'Aspremont, C. and L. A. Gerard-Varet (1979), Incentives and Incomplete Information, *Journal of Public Economics,* 11, 25-45.

[12] Dekel, E. and F. Gul (1997), Rationality and Knowledge in Game Theory, in: Kreps D. M. and K.F. Wallis (eds.), *Advances in Economics and Econometrics*, vol. 1, Cambridge, Cambridge UP, 87-172.

[13] Diamond, P. (1967), Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility: Comment", *Journal of Political Economy,* 75, 765-766.

[14] Feinberg, Y. (2000), Characterizing the Existence of a Common Prior in the Form of Posteriors, *Journal of Economic Theory,* 91, 127-179.

[15] Fleming, M. (1952), A Cardinal Concept of Welfare, *Quarterly Journal of Economics,* 56, 366-384.

[16] Gilboa, I. , Samet, D. and D. Schmeidler (2001), Utilitarian Aggregation of Beliefs and Tastes, mimeo.

[17] Gul, F. (1998), A Comment on Aumann's View, *Econometrica,* 66, 923-927.

[18] Hammond, P. (1981), Ex Ante and Ex post Welfare Optimality under Uncertainty, *Economica* 48, 235-250.

[19] Harsanyi, J. (1967-68), Games with Incomplete Information Played by Bayesian Players, Parts I-III, *Management Science,* 8, 159-182, 320-334, 486-502.

[20] Harsanyi, J. (1955), Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility, *Journal of Political Economy,* 63, 309-321.

[21] Harsanyi, J. and R. Selten (1972), A Generalized Nash Solution for Two-Person Bargaining Games with Incomplete Information, *Management Science,* 18, P80-P106.

[22] Holmstrom, B. and R. Myerson (1983), Efficient and Durable Decision Rules with Incomplete Information, *Econometrica,* 51, 1799-1819.

[23] Hylland, A. and R. Zeckhauser (1979), The Impossibility of Bayesian Group Decision Making with Separate Aggregation of Beliefs and Values, *Econometrica,* 47, 1321-1336.

[24] Mertens, J.-F. and S. Zamir (1985), Formulation of Bayesian Analysis for Games with Incomplete Information, *International Journal of Game Theory,* 14, 1-29.

[25] Mongin, P. (1995), Consistent Bayesian Aggregation, *Journal of Economic Theory,* 66, 313-351.

[26] Mongin, P. (1998), The Paradox of the Bayesian Experts and State-Dependent Utility Theory, *Journal of Mathematical Economics,* 29, 331-361.

[27] Morris, S. (1994), Trade with Heterogeneous Prior Beliefs and Asymmetric Information, *Econometrica,* 62, 1327-1347.

[28] Myerson, R. (1984), Two-Person Bargaining Problems with Incomplete Information, *Econometrica,* 52, 461-487.

[29] Myerson, R. and M. Satterthwaite (1983), Efficient Mechanisms for Bilateral Trading, *Journal of Economic Theory,* 28, 265-281.

[30] Nehring, K. (2001), Common Priors under Incomplete Information: A Unification, *Economic Theory,* 18, 535-553.

[31] Nehring, K. (2002a), Bayesian Cooperation under Incomplete Information: Incorporating Fairness Concerns, in preparation.

[32] Nehring, K. (2002b), Utilitarianism without Common Priors, in preparation.

[33] Samet, D. (1998), Iterated Expectations and Common Priors, *Games and Economic Behavior* 24, 131-141.

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

[35] Wilson, R. (1978), Information, Efficiency and the Core of an Economy, *Econometrica* 46, 807-816.