Egalitarianism in Mechanism Design*

Geoffroy de Clippel†

This Version: January 2012

Abstract

This paper examines ways to extend egalitarian notions of fairness to mechanism design. It is shown that the classic properties of constrained efficiency and monotonicity with respect to feasible options become incompatible, even in quasi-linear settings. An interim egalitarian criterion is defined and axiomatically characterized. It is applied to find "fair outcomes" in classical examples of mechanism design, such as cost sharing and bilateral trade. Combined with ex-ante utilitarianism, the criterion characterizes Harsanyi and Selten's (1972) interim Nash product. Two alternative egalitarian criteria are proposed to illustrate how incomplete information creates room for debate as to what is socially desirable.

*The paper beneficiated from insightful comments from Kfir Eliaz, Klaus Nehring, David Perez-Castrillo, Kareen Rozen, and David Wettstein. Financial support from the National Science Foundation (grant SES-0851210) is gratefully acknowledged.

†Department of Economics, Brown University. Email: declippel@brown.edu

1. INTRODUCTION

Developments in the theory of mechanism design, since its origin in the seventies, have greatly improved our understanding of what is feasible in environments involving agents that hold private information. Yet little effort has been devoted to the discussion of socially desirable selection criteria, and the computation of incentive compatible mechanisms meeting those criteria in applications. In other words, extending the theory of social choice so as to make it applicable in mechanism design remains a challenging topic to be studied. The present paper makes some progress in that direction, with a focus on the egalitarian principle.

As a starting point, consider a classical social choice problem of complete information. Assume for simplicity that utilities are transferable in a weak sense: for any feasible utility profile that is individually rational, and any agent i, there exists an alternative feasible utility profile that makes i better off. Egalitarianism in that case means picking the unique feasible utility profile that is Pareto efficient and equalizes the participants' utilities. Kalai (1977) shows that this criterion is the only to satisfy the axioms of Pareto efficiency, anonymity and monotonicity. Pareto efficiency means that there is no waste, in that all the social surplus has been distributed. Anonymity requires that the selection criterion does not discriminate among agents based on their identity. Monotonicity means that no member of the society should be worse off when more collective decisions are available. In Section 3, I show that straightforward extensions of the Monotonicity and Efficiency criteria become incompatible in the more general context of mechanism design, even when utilities are fully transferable. This incompatibility will be shown to be due to the *combined* presence of incomplete information and incentive constraints.

In Section 4, I define a selection criterion that extends egalitarianism so as to be applicable in mechanism design, and provide a partial axiomatic characterization in terms of a sensible weakening of the monotonicity property that restricts the circumstances under which monotonicity applies. Let's take the point of view of a mechanism designer (e.g. benevolent social planner or impartial arbitrator). He ignores the agents' information, encoded as usual by types in the sense of Harsanyi (1967-68). To have as many options as possible to choose from, while guaranteeing that agents report their information truthfully, he must restrict attention to incentive compatible direct mechanisms (classical revelation principle). For each such mechanism and each profile of types, he can compute the interim utility that each agent would experience should that mechanism be implemented and should these types happen to be the correct ones. A mechanism fulfills the interim egalitarian criterion if it is constrained efficient in the sense of (interim) incentive efficiency (Holmström and Myerson (1983)), and all agents experience the same interim utilities independently of what the true information state might be. Not all mechanism design problems admit a mechanism that fulfills the interim egalitarian criterion. Those problems that do admit such a mechanism will be called simple. In Section 4, I introduce a list of axioms that are compatible with incentive efficiency and the weaker version of monotonicity on the set of all mechanism design problems. In addition, any solution satisfying these properties picks mechanisms that fulfill the interim egalitarian criterion in simple problems (see Theorem 2).

While I will continue interpreting social choice functions as selection criteria for uninformed mechanism designers throughout the paper, the interim egalitarian criterion and the underlying axioms also admit alternative interpretations in the absence of such third parties. The interim egalitarian criterion can indeed be seen alternatively as maximizing the expected-utility level that is commonly known to be reached by everyone. As for underlying axioms, Holmström and Myerson (1983) already observed, for instance, that incentive efficiency can be interpreted as an impossibility to find an alternative mechanism that is commonly known to bring higher expected utilities to all individuals. Monotonicity requires in that alternative interpretation of the model that it must be common knowledge that all individuals' expected utilities increase when more collective decisions become available.

Harsanyi (1963) (see also Shapley (1969) and Yaari (1981)) highlighted an intriguing relationship between the maximization of the Nash's (1950) product, on the one hand, and the egalitarian and utilitarian principles, on the other hand, in classical problems of complete information. Indeed, in any problem, there is a unique way of rescaling individual utilities so that the egalitarian and utilitarian principles coincide, and this coincidence happens exactly at the Nash solution. In Theorem 3, I extend this result by showing that there is a unique way of rescaling interim utilities in any mechanism design problem (not just simple ones) so that the interim egalitarian criterion and ex ante utilitarianism (shown to be an appropriate version of utilitarianism at the interim stage by Nehring (2004)) coincide. The resulting solution amounts to selecting the incentive compatible mechanism that is maximal according to Harsanyi and Selten's (1972) weighted Nash product (see Myerson (1979)). While Myerson (1984a) provided arguments against this solution when interpreted as a bargaining solution, the present result suggests that the maximization of Harsanyi and Selten's (1972) weighted Nash product may be more meaningful for social choice.

Section 6 is devoted to the computation of mechanisms that meet the interim egalitarian criterion in classical quasi-linear examples (sharing the cost of production of a public good, and determining the fair price in a bilateral trade problem). I conclude the paper by introducing and discussing by means of examples two other extensions of the egalitarian principle, one based on what the individuals themselves perceive as fair, and one based on a procedural approach to justice (see Section 7). The purpose of this last Section is to illustrate the richness of the topic of social choice under incomplete information, on which more work is needed.

Related Literature

Welfare economics under risk with symmetric uncertainty has been researched since Harsanyi's (1955) aggregation theorem. Ex-ante utilitarianism, as advocated in that work, has been criticized for its indifference to the distribution of utilities. Ex-ante egalitarianism is the first alternative to have been studied (advocated by Diamond (1967), Epstein and Segal (1992), among others). In those models, a state of nature will be revealed at some future date (e.g. who will suffer an accidental loss), and contingent contracts (e.g. insurance) are being selected at a time when individuals hold some prior beliefs on the relative likelihood of those states, but do not know which will occur. The model studied in the present paper is more general in that individuals may be asymmetrically informed, and it may be impossible to write contracts contingent on the state of nature, e.g. because it is not verifiable in court, or because it will never be disclosed (encoding individuals' preferences, for instance). The symmetric uncertainty model will thus be a special case, in which case the interim egalitarian principle boils down to ex-ante egalitarianism. It is worth noting that the mere presence of risk has already generated a rich literature on conflicting interpretations of the egalitarian principle and of social choice in general (see e.g. Fleurbaye (2010) and Grant et al. (2010)), particularly regarding a tension between ex-ante efficiency and ex-post social comparisons. Similar arguments should stimulate further research under asymmetric information as well. Section 7 highlights how the presence of asymmetric information leads to new conceptual questions that are likely to further enrich the debate on egalitarianism in mechanism design.

The only paper studying social choice criteria under incomplete information from an axiomatic perspective is Nehring (2004). He studies interim social welfare orderings that allow to compare profiles of ex-post utilities. Generalizing Harsanyi (1955), he shows that two axioms of consistency, one with the interim Pareto criterion and the other with uniform ex-post utilitarian comparisons, are compatible if and only if the individuals' beliefs can be derived from a common prior. In addition, the only interim social welfare ordering that satisfies these two axioms when a common prior exists, is the ex-ante utilitarian criterion. I show in de Clippel (2010) that Nehring's methodology – extending to the interim stage classical social welfare orderings by combining ex-post arguments with the interim Pareto criterion – essentially works only for the utilitarian criterion. Indeed, his two axioms become incompatible for most common priors once the utilitarian criterion is replaced by any other social welfare ordering that satisfies the strict Pigou-Dalton transfer principle (as does the egalitarian principle, for instance).

There is a small literature on axiomatic bargaining under asymmetric information that relates to the present paper. The model studied in the present paper is inspired from Myerson's (1984a) bargaining model. The virtual utility construction he elaborated in Myerson (1983, 1984a, and 1984b) will also prove useful to establish Theorem 2 below. As described earlier in this Introduction, the social choice criterion I derive by reconciling interim egalitarian and utilitarian criteria coincides with Myerson's (1979) variant of the bargaining solution proposed by Harsanyi and Selten (1972). Weidner (1992) reconstructs Harsanyi and Selten's characterization result of the weighted Nash product on Myerson's domain, and argues that their result then holds only when the players' types are independently distributed, an assumption that is not required in Theorem 3. The criticisms raised by Myerson (1984a) against Harsanyi and Selten (1972) and Myerson (1979) will be revisited in Section 5 in terms of social choice instead of bargaining.

There exists a slightly more extensive literature whose objective is to compute the optimal mechanism under a given social choice criterion, without discussing the normative appeal of the criterion itself. For instance, Myerson and Satterthwaite's (1983) apply the ex-ante utilitarian criterion to a bilateral trade problem. There is also a more recent literature that is developing at the intersection of computer science and economics that looks for strategy-proof mechanisms that maximize a worst-case scenario index, guaranteeing for instance a minimal percentage of the maximal total surplus in every possible realization of the types (see e.g. Guo and Conitzer (2009), Moulin (2009), de Clippel et al. (2009), and references therein). Though intuitively appealing, this criterion has not been axiomatically characterized yet. Another difference, of course, is that this approach is non-Bayesian. More importantly, we see that emphasis has so far been placed on the utilitarian principle.

Ray and Ueda (1996) study the interplay between incentives and welfare criteria from a different perspective. Collective decisions (sharing a collective output) are taken under complete information, after individuals decide on how much effort to perform. They show that the degree of inefficiency decreases in the extent of egalitarianism embodied in the social welfare function.

Finally, readers are referred to a companion paper (de Clippel et al., 2010) to see how one can apply the interim egalitarian criterion while avoiding interpresonal comparisons of utilities, by measuring utility gains endogenously in the tradition of Pazner and Schmeidler (1978).

2. MODEL

Social choice theory has been developed mostly in the space of utilities. Indeed, a social choice problem is a subset of $\mathbb{R}^I$ describing the utility profiles that can be achieved by underlying collective decisions (where I is the set of relevant individuals). It is tempting, at first, to describe a social choice problem under incomplete information as a collection of utility possibility sets, one for each possible information state. Yet this information is not sufficient, because it does not allow to keep track of incentive constraints. Indeed, the fact that a utility vector is feasible at some type profile does not allow to infer what would be the utility that an individual would get should he report a different type. One must thus consider a more detailed model, as studied under a different perspective in mechanism design.

A *mechanism design problem* is a sextuple $\mathcal{S} = (I, D, d^*, (T_i)_{i\in I}, p, (u_i)_{i\in I})$, where I is the finite set of *individuals*, D is the set of *collective decisions*, $d^* \in D$ is the *status-quo*, T_i is the finite set of individual i's possible *types*, $p \in \Delta(T)$ is the *common prior* determining the individuals' *beliefs* (where $T = \times_{i\in I} T_i$), and $u_i : D \times T \to \mathbb{R}$ is individual i's *utility function*, used to determine his interim preferences via the expected utility criterion. It will be assumed for notational convenience that $u_i(d^*, t) = 0$, for all $t \in T$. This is without loss of generality if utilities are understood as representing gains over the status-quo. I will also assume that T is the only nonempty common knowledge event. This is without loss of generality, as the results can be applied over minimal common knowledge events, and then merged so as to apply to the whole set of type profiles.

A (direct) *mechanism* for $\mathcal{S}$ is a function $\mu : T \to \Delta(D)$. If a mechanism μ is implemented truthfully, then individual i's expected utility when of type t_i is given by:

$$U_i(\mu|t_i) = \sum_{t_{-i}\in T_{-i}} p(t_{-i}|t_i) u_i(\mu(t), t).$$

If all the other individuals report their true type, while individual i reports t'_i instead of his true type t_i, then his expected utility is denoted $U_i(\mu, t'_i|t_i)$:

$$U_i(\mu, t'_i|t_i) = \sum_{t_{-i} \in T_{-i}} p(t_{-i}|t_i) u_i(\mu(t'_i, t_{-i}), t).$$

The mechanism μ is *incentive compatible* if $U_i(\mu|t_i) \geq U_i(\mu, t'_i|t_i)$, for each t_i, t'_i in T_i and each $i \in I$. The revelation principle (Myerson (1979)) implies that any agreement that is achievable through some form of communication can also be achieved through truth-telling in an incentive compatible direct mechanism. Hence one may restrict attention to those mechanisms without loss of generality.

An incentive compatible mechanism μ is *interim individually rational* if $U_i(\mu|t_i) \geq 0$, for all $t_i \in T_i$ and all $i \in I$. A mechanism is *feasible* if it is both incentive compatible and interim individually rational. The set of feasible mechanisms will be denoted $\mathcal{F}(\mathcal{S})$.

Let $\mathcal{U}(\mathcal{M})$ be the set of interim utilities that can be achieved via incentive compatible mechanisms that belong to a set $\mathcal{M}$:

$$\mathcal{U}(\mathcal{M}) = \{\mathfrak{u}(\mu) | \mu \in \mathcal{M}\},$$

where $\mathfrak{u}(\mu) = (U_i(\mu|t_i))_{t_i \in T_i, i \in I}$. It is easy to check that $\mathcal{U}(\mathcal{F}(\mathcal{S}))$ is a convex set, as first observed by Myerson (1979). For notational simplicity, I will restrict attention to mechanism design problems for which there exists $\mathfrak{u} \in \mathcal{U}(\mathcal{F}(\mathcal{S}))$ such that $\mathfrak{u} >> 0$, and for which $\mathcal{U}(\mathcal{F}(\mathcal{S}))$ is compact. This last assumption is true whenever D is finite and in usual applications that involve a continuum of collective decisions, but it is possible to construct mathematical examples that would violate it.

A *solution* is a correpondence Σ that associates a nonempty set of feasible mechanisms to each mechanism design problem: $\Sigma(\mathcal{S}) \subseteq \mathcal{F}(\mathcal{S})$, for each $\mathcal{S}$. Even though correspondences are allowed, solutions are assumed to be essentially single-valued, in the sense that all the individuals must be indifferent (whatever their private information) between any two mechanisms that belong to the solution of any problem $\mathcal{S} = (I, D, d^*(T_i)_{i \in I}, p, (u_i)_{i \in I})$:

$$(\forall \mu, \mu' \in \Sigma(\mathcal{S}))(\forall i \in I)(\forall t_i \in T_i) : U_i(\mu|t_i) = U_i(\mu'|t_i). \quad (1)$$

A mechanism design problem $\mathcal{S} = (I, D, d^*, (T_i)_{i \in I}, p, (u_i)_{i \in I})$ is *quasi-linear* if there exist a set $\hat{D}$ and functions $\hat{u}_i : \hat{D} \times T \to \mathbb{R}$ (one for each $i \in I$) such that $D = \hat{D} \times \mathbb{R}^I$ and $u_i((\hat{d}, m), t) = \hat{u}_i(\hat{d}, t) + m_i$, for all $(\hat{d}, m) \in \hat{D} \times \mathbb{R}^I$.

3. MONOTONICITY AND INCENTIVE EFFICIENCY

Monotonicity is a key property of the egalitarian solution (Kalai (1977)). No member of the society should be worse off (whatever his or her private information) when more collective decisions are available. Monotonicity properties of this type have a long tradition in the theories of social choice, distributive justice, and bargaining. First briefly discussed in the books of Luce and Raiffa (1957, pages 133 and 134) and Owen (1968), they have been systematically studied since then under the assumption of complete information, cf. Kalai and Smorodinsky (1975), Kalai (1977), Thomson and Myerson (1980), Kalai and Samet (1985), Young (1985a, 85b), Moulin and Thomson (1988), and Moulin (1992), to cite only a few references that illustrate the variety of environments where they have been studied.

In quasi-linear environments of complete information, selecting the collective decision that maximizes the sum of the gains and designing monetary compensations so as to equalize those gains across individuals, is indeed monotonic and consistent with the Pareto criterion. The purpose of the present section is to show that these two properties may become incompatible in the combined presence of incomplete information and incentive constraints.

The natural extension of the Pareto criterion was proposed by Holmström and Myerson (1983).

Incentive Efficiency (I-EFF) *Let $\mathcal{S}$ be a mechanism design problem, and let $\mu \in \Sigma(\mathcal{S})$. Then there does not exist a mechanism $\hat{\mu} \in \mathcal{F}(\mathcal{S})$ such that $U_i(\hat{\mu}|t_i) \geq U_i(\mu|t_i)$, for each $t_i \in T_i$, and each $i \in I$, with at least one of the inequalities being strict.*

I-EFF requires the solution to systematically exhaust the benefit of cooperation at the time of agreeing. The set of interim utilities associated to incentive efficient mechanisms will be denoted by $\mathcal{U}^{eff}(\mathcal{S})$:

$$\mathcal{U}^{eff}(\mathcal{S}) = \{\mathfrak{u} \in \mathcal{U}(\mathcal{F}(\mathcal{S}))| \nexists \mathfrak{u}' \in \mathcal{U}(\mathcal{F}(\mathcal{S})) \setminus \{\mathfrak{u}\} : \mathfrak{u}' \geq \mathfrak{u}\}.$$

There is also an obvious way to extend the monotonicity property to problems of incomplete information.

Monotonicity (MON) *Let $\mathcal{S}$ and $\mathcal{S}'$ be two mechanism design problems. Suppose that $\mathcal{S}'$ differs from $\mathcal{S}$ only in that more collective decisions are available: $I = I'$, $D \subseteq D'$, $T_i = T_i'$, and $u_i(d,t) = u_i'(d,t)$, for each $i \in I$, each $d \in D$, and each $t \in T$. If $\mu \in \Sigma(\mathcal{S})$ and $\mu' \in \Sigma(\mathcal{S}')$, then $U_i(\mu'|t_i) \geq U_i(\mu|t_i)$, for each $t_i \in T_i$, and each $i \in I$.*

Theorem 1 *There is no solution that satisfies both I-EFF and MON, even on the restricted class of quasi-linear mechanism design problems.*

Proof: The proof goes by way of example. Consider a mechanism design problem with two individuals,[1] 1 and 2, that can be of two types, L or H. Each individual knows only his own type, and believes that the two types of the other individual are equally likely. Each individual has up to 10 hours available to work, and his productivity per hour is 1 if his type is L, and 2 if his type is H. Allowing for any kind of transfers and free disposal, the set of decisions is thus $D = \{(\alpha_1, \alpha_2, m_1, m_2) \in [0,10]^2 \times \mathbb{R}^2 | m_1 + m_2 \leq 0\}$. The utility functions are given by the following expression: $u_i((\alpha, m), t) = \pi_i(t_i)\alpha_i + m_i$, for each $(\alpha, m) \in D$, each $i \in \{1,2\}$ and $t_i \in \{L, H\}$, with the convention $\pi_i(L) = 1$ and $\pi_i(H) = 2$, for each $i \in \{1,2\}$. One may think of each individual cultivating a similar field, their payoffs being the quantity produced on their own field, which depends on their productivity, modified by any kind of monetary subsidy and taxes. We will assume that $d^* = 0$.

Let's consider now a feasible mechanism[2] (α, m) that determines a decision

[1] The example extends in a straightforward way to accomodate any number of individuals, but at the cost of heavier notations. The qualitative results remain unchanged.

[2] To keep notations lighter, the same letters are used to denote a collective decision, and the mechanism that selects a collective decisions as a function of the reported types.

in D as a function of the individuals' reports.[3] The incentive constraints faced by the first individual can be written as follows:

$$\bar{m}_1(H) - \bar{m}_1(L) \leq \bar{\alpha}_1(L) - \bar{\alpha}_1(H) \leq \frac{\bar{m}_1(H) - \bar{m}_1(L)}{2} \tag{2}$$

where $\bar{\alpha}_1(L)$ (resp. $\bar{\alpha}_1(H)$) is the average quantity of time the first individual thinks he will have to work given the mechanism when of type L (resp. H), and $\bar{m}_1(L)$ (resp. $\bar{m}_1(H)$) is the average quantity of money the first individual thinks he will receive given the mechanism when of type L (resp. H), i.e.

$$\bar{\alpha}_1(L) = \frac{1}{2}(\alpha_1(L,L) + \alpha_1(L,H)) \text{ and } \bar{\alpha}_1(H) = \frac{1}{2}(\alpha_1(H,L) + \alpha_1(H,H)),$$

$$\bar{m}_1(L) = \frac{1}{2}(m_1(L,L) + m_1(L,H)) \text{ and } \bar{m}_1(H) = \frac{1}{2}(m_1(H,L) + m_1(H,H)).$$

Equation (2) implies that $\bar{m}_1(H) \leq \bar{m}_1(L)$ and $\bar{\alpha}_1(L) \leq \bar{\alpha}_1(H)$. If the mechanism is incentive efficient, then it must be that $\bar{\alpha}_1(H) = 10$. Otherwise, one could construct another feasible mechanism that interim Pareto dominates (α, m) by slightly increasing both $\bar{\alpha}_1(L)$ and $\bar{\alpha}_1(H)$ by a same amount, while keeping α_2 and m unchanged. Notice also that the second inequality in (2) must be binding if (α, m) is incentive efficient. Indeed, suppose on the contrary that the inequality is strict. Hence $\bar{\alpha}_1(L) < 10$ (as otherwise $\bar{\alpha}_1(L) = \bar{\alpha}_1(H)$, and (2) implies that $\bar{m}_1(L) = \bar{m}_1(H)$, which contradicts the fact that the second inequality is strict). Now one can construct another feasible mechanism that interim Pareto dominates (α, m) by increasing a bit $\bar{\alpha}_1(L)$, while keeping $\bar{\alpha}_1(H)$, α_2 and m unchanged. This contradicts the fact that (α, m) is incentive efficient, and thereby proves that the second inequality in (2) is indeed binding. Notice now that $\bar{\alpha}_1(L)$ must equal 10, as well. Otherwise, consider an alternative mechanism where $\bar{\alpha}_1(L)$ is increased by a small amount ϵ, while keeping $\bar{\alpha}_1(H)$ and $\bar{\alpha}_2$ constant, and changing monetary transfers as follows: $\Delta m(L,L) = \Delta m(L,H) = (-\epsilon, +\epsilon)$ and $\Delta m(H,L) = \Delta m(H,H) = (+\epsilon, -\epsilon)$. Since $\bar{\alpha}_1(L) < 10$ and the second inequality in (2) is binding, it must be that the first inequality in (2) is strict for the original mechanism. The change

[3] Utilities being linear in both time and money, there is no loss of generality in discussing only deterministic mechanisms.

makes lying for the first individual a bit more attractive when of a low type, but not enough for him to actually lie if ϵ is small enough. The incentive constraint remains binding when he is of a high type. As for the second individual, nothing changes for him, since he ignores the first individual's type, and the additional tax of ϵ when the first individual reports a high type is exactly compensated by the additional subsidy of ϵ when the first individual reports a low type. In terms of interim payoffs, both types of both individuals get at least as much with the new mechanism than with the original one, but the first individual gets strictly more when of a high type, thereby contradicting the fact that the original mechanism is incentive efficient. Hence $\bar{\alpha}_1(L) = \bar{\alpha}_1(H) = 10$, and (2) implies $\bar{m}_1(L) = \bar{m}_1(H)$. A similar reasoning applies to individual 2. Hence, if a mechanism (α, m) is incentive efficient, then there exists $(x_1, x_2) \in \mathbb{R}^2$ such that $x_1 + x_2 = 0$, and

$$U_i((\alpha, m)|L) = 10 + x_i \text{ and } U_i((\alpha, m)|H) = 20 + x_i,$$

for $i = 1, 2$. Conversely, any such interim payoffs can be achieved by an incentive compatible mechanism (α, m), where $\alpha_i(t) = 10$, $m_i(t) = x_i$, for each $t \in \{L, H\}^2$ and each $i \in \{1, 2\}$.

Consider now a similar problem, but where only a third field is available, with a total productivity of 3 per joint hour of work. Collective decisions determine how much time each individual devotes to the field, in which proportion to share the outpout, and possible monetary compensations. Formally,$D' = \{(\alpha'_1, \alpha'_2, s_1, s_2, m_1, m_2) \in [0, 10]^2 \times [0, 1]^2 \times \mathbb{R}^2 | s_1 + s_2 = 1 \text{ and } m_1 + m_2 \leq 0\}$, and the utility functions are given by the following expression: $u'_i((\alpha', s, m), t) = 3s_i \min\{\alpha'_1, \alpha'_2\} + m_i$, for each $(\alpha', s, m) \in D'$, each $i \in \{1, 2\}$ and $t_i \in \{L, H\}$. The status-quo is $d^* = 0$. Consider a mechanism (α', s, m') that is incentive efficient. Feasibility implies that

$$\sum_{i=1}^{2} \sum_{t_i \in \{L,H\}} U'_i(\alpha', s, m'|t_i) = \sum_{t \in \{L,H\}^2} \frac{1}{2} \sum_{i=1}^{2} u'_i((\alpha'(t), s(t), m'(t)), t) \leq 60$$

Since utilities are independent of the types, incentive constraints imply that the two types of each agent expect identical utility gains. Hence, if a mechanism

(α', s, m') is incentive efficient, then there exists $(x'_1, x'_2) \in \mathbb{R}^2$ such that $x'_1 + x'_2 = 0$, and

$$U'_i((\alpha', m')|L) = U'_i((\alpha', m')|H) \leq 15 + x'_i,$$

for $i = 1, 2$. The inequality must in fact be binding, since any interim payoff profile on the right-hand side can be achieved by an incentive compatible mechanism (α', s, m'), where $\alpha'_i(t) = 10$, $s_i(t) = 1/2$, $m'_i(t) = x'_i$, for each $t \in \{L, H\}^2$ and each $i \in \{1, 2\}$.

Finally, suppose that the impartial third party can choose to allocate the individuals' time between the three fields:

$$D'' = \{(\alpha_1, \alpha_2, \alpha'_1, \alpha'_2, s_1, s_2, m_1, m_2) \in [0, 10]^4 \times [0, 1]^2 \times \mathbb{R}^2 | \\ s_1 + s_2 = 1, \alpha_1 + \alpha'_1 \leq 10, \alpha_2 + \alpha'_2 \leq 10, \text{ and } m_1 + m_2 \leq 0\}$$

and the utility functions are given by the following expression:

$$u''_i((\alpha, \alpha', s, m), t) = \pi_i(t_i)\alpha_i + 3s_i \min\{\alpha'_1, \alpha'_2\} + m_i,$$

for each $(\alpha, \alpha', s, m) \in D''$, each $i \in \{1, 2\}$ and $t_i \in \{L, H\}$, with the convention $\pi_i(L) = 1$ and $\pi_i(H) = 2$, for each $i \in \{1, 2\}$. Notice that

$$\sum_{i=1}^{2} \sum_{t_i \in \{L,H\}} U''_i((\alpha, \alpha', m)|t_i) = \sum_{t \in \{L,H\}^2} \frac{1}{2} \sum_{i=1}^{2} u''_i((\alpha(t), \alpha'(t), s(t), m(t)), t) \leq 65,$$

for each $(\alpha, \alpha', s, m) \in \mathcal{F}(\mathcal{S}'')$, the last equality following from the fact that the maximal total surplus is 40 when both individuals' type is H and is 30 otherwise. Hence there is no way to find a feasible mechanism that gives an interim utility of at least $15 + x'_1$ and $15 + x'_2$ to the low-type individuals, and $20 + x_1$ and $20 + x_2$ to the high-type individuals, which contradicts MON, since $D \subseteq D''$, $D' \subseteq D''$, and both $u''_i(d, t) = u_i(d, t)$, for each $i \in I$, each $d \in D$, and each $t \in \{L, H\}^2$, and $u''_i(d', t) = u'_i(d', t)$, for each $i \in I$, each $d' \in D'$, and each $t \in \{L, H\}^2$. ■

I-EFF and MON would not be incompatible in quasi-linear environments of incomplete information in the absence of incentive constraints (such an informational assumption is maintained in some related models of bargaining and cooperative games, see e.g. Wilson (1978), de Clippel and Minelli (2004), de

Clippel (2007), and Kalai and Kalai (2010)). To illustrate this point,[4] suppose for instance that the output of both individuals could be observed at no cost in the first problem described in the proof of Theorem 1, and that contracts could be made contingent on that observed output. Then the social planner or the arbitrator could implement a mechanism that requires both individuals to work for ten hours on their own field, and for the individual with a high output to pay $10 to the individual with low output (if such a configuration occurs). The resulting interim utilities are 15 for both types of both individuals, which is also achievable in the last two problems described in the proof of Theorem 1, and hence I-EFF and MON are compatible in that example. Though the presence of incentive constraints is an important factor for the incompatibility of I-EFF and MON, the presence of incomplete information, i.e. asymmetric information at the time of making the collective choice, is also critical. The tension between I-EFF and MON would indeed disappear if the collective decision to be implemented at the interim stage was made at the ex-ante stage (i.e. before the agents learn their private information). To illustrate this point,[5] notice indeed that there exists an ex-ante incentive efficient mechanism in each of the three problems discussed in the proof of Theorem 1 that lead to an expected payoff of $15 to both individuals, and hence efficiency and monotonicity are compatible, even when incentive constraints are imposed, when considered at the ex-ante stage. Similarly, the impossibility result would not hold in the presence of moral hazard instead of adverse selection.

Theorem 1 thus shares some similitude with Myerson and Satterthwaite (1983), in that the incompatibility of their two properties, namely ex-post efficiency and interim individual rationality, also requires the presence of both incentive constraints and incomplete information. Indeed, if any mechanism could be implemented in Myerson and Satterthwaite's example without having to satisfy the incentive constraints, then any mechanism in Wilson's (1978) coarse core (which, he proved, is non-empty in even much more general exchange economies), is both interim individually rational and ex-post efficient

[4]The general argument is left to the reader.

[5]Again, the general argument is left to the reader.

(because of quasi-linearity). On the other hand, the first-best can be achieved in Myerson and Satterthwaite's example (and many other quasi-linear environments), even while imposing incentive constraints, if decisions are made at the ex-ante stage (see e.g. d'Aspremont and Gérard-Varet (1979, 1982)). It is then always possible to design some type-independent monetary transfer to meet the (ex-ante) individual rationality constraints. The common denominator between Myerson and Satterthwaite's (1983) impossibility result and Theorem 1 is that performing interim utility transfers across different "type-agents" (in the sense of Harsanyi (1967-68)) may be critically impeded by the presence of incentive constraints.

4. INTERIM EGALITARIAN CRITERION

As indicated in the Introduction, I think of an uninformed mechanism designer (e.g. benevolent social planner or impartial arbitrator) trying to have all individuals experience the same interim expected utility whatever the true type profile might be, while exhausting all the surplus under the constraint that individuals report their types truthfully. Formally, a mechanism μ satisfies the *interim egalitarian criterion* if it is incentive efficient and $U_i(\mu|t_i) = U_j(\mu|t_j)$, for all $i \neq j$, and for all $t \in T$.

As we will see in applications (or as can be checked in the first problem encountered in the proof of Theorem 1), mechanisms that pass the interim egalitarian criterion need not always exist. A mechanism design problem is *simple* if it admits at least one mechanism μ that passes the interim egalitarian criterion and that satisfies the following additional property:

$$(\exists \lambda \in \times_{i \in I} \mathbb{R}^{T_i}_{++}) : \mu \in \arg \max_{\nu \in \mathcal{F}(\mathcal{S})} \sum_{i \in I} \sum_{t_i \in T_i} \lambda_i(t_i) U_i(\nu|t_i).$$

This additional property is imposed for technical reasons that will become clear only when introducing the concept of virtual utility in the proof of Theorem 2. Observe though that it is very weak, and can often be dispensed with. Indeed, any mechanism that passes the interim egalitarian criterion is incentive efficient. A standard separation argument implies that there necessarily exists $\lambda \in \times_{i \in I} \mathbb{R}^{T_i}_{+} \setminus \{0\}$ such that $\mu \in \arg\max_{\nu \in \mathcal{F}(\mathcal{S})} \sum_{i \in I} \sum_{t_i \in T_i} \lambda_i(t_i) U_i(\nu|t_i)$. The additional property is thus a restriction only in that it requires the existence of a strictly positive λ, which is most often the case. It is always satisfied, for instance, (and can thus be dropped from the definition of simple problem) whenever D is finite.

I now provide a list of axioms – Restricted Monotonicity, Interim Welfarism, Exhaustivity, Irrelevant Splitting of a Type, Independance of Irrelevant Alternatives, and Anonymity – that will partly justify this interim egalitarian criterion, in that there exist solutions satisfying them on the whole domain of mechanism design problems, and any such solution must coincide with the interim egalitarian criterion for simple problems.

Efficiency and Monotonicity may also be incompatible under complete information when utilities are not transferable in any sense of the word. This is known at least since Luce and Raiffa (1957). It is helpful to understand why. Consider a first two-person problem where only one collective decision, d, is available to replace the status-quo, and assume further that the first participant is indifferent between the alternative and the status-quo, while the second strictly prefers the former. Efficiency will guide the social planner or the arbitrator in selecting the alternative over the status-quo. Consider then a second problem that is symmetric to the first, in that only one collective decision, d', is available to replace the status-quo, with the property that the second participant is indifferent between the alternative and the status-quo, while the first strictly prefers the former. Again, efficiency will guide the social planner or the arbitrator in selecting the alternative over the status-quo. Let then the third problem be the union of the first two: any lottery selecting either d or d' is feasible. Clearly, there is no way to solve uniquely that new problem so as to make both the first individual better off than with the solution to the second problem, and the second individual better off than with the solution to the first problem. It has long been understood that this kind of impossibility is due to an extreme lack of utility transferability. For instance, starting from d, there is no alternative decision in the first problem that would allow to make the first individual strictly better off, even if one is ready to make the second individual worse off in any amount. A usual way to deal with this complication is actually to avoid it altogether by restricting the class of acceptable social choice problems. A textbook example is Moulin's (1988, Theorem 3.2) presentation of Kalai's (1977, Theorem 1) axiomatic characterization of the egalitarian solution,[6] where social choice problems are assumed to satisfy a property of "minimal transferability,"[7] meaning that, at any feasible contract that is efficient and individually rational, and for any individual i, there exists an alternative feasible option that makes all the other individuals better off (thereby at the expense of i).[8] In particular, it rules out the possibility of satiation.

Theorem 1 shows that restricting the class of mechanism design problems is not anymore a practical way to resolve the incompatibility of Efficiency and Monotonicity under incomplete information, as this incompatibility already occurs on the very restrictive class of quasi-linear mechanism design problems (i.e. with unlimited one-to-one transferability). The reason is that incentive constraints may lead to feasible sets of interim utilities that are non-comprehensive (for instance, a type of an individual may beneficiate of an "informational rent" in any incentive efficient mechanism) and with the possibility of satiation, even in the simplest quasi-linear environments. The way asymmetric information and incentive constraints restrict what is feasible at the interim stage makes it significantly more complicate to derive axiomatic characterizations than when information is complete.

I will say that interim utilities are (weakly[9]) "transferable" at an incentive compatible mechanism μ if for any two participants i, j, and any possible pair of types, t_i for i and t_j for j, there exists an alternative incentive compatible

[6]To be precise, Kalai actually characterized the proportional solutions in his original paper, but of course the egalitarian solution is the only one to be anonymous in that class.

[7]Kalai's original result only required free disposal, which is weaker than the property of minimal transferability, but at the cost of accommodating only a weak form of Pareto efficiency, which is not very appealing, especially in social choice theory.

[8]Though not explicit in its name, the usual notion of transferable utility in cooperative games (or of quasi-linearity in mechanism design) requires in addition that, starting from *any* contract, utilities can be transferred at a *constant* rate of *one to one*. Minimal transferability is thus far weaker, to a point that it is almost unrelated.

[9]Again, transferability here should not be confused with its narrow meaning in cooperative games under complete information - cf. footnote 8.

mechanism ν that is better than μ for i when of type t_i, worse than μ for j of type t_j, and at least as good as μ for any combination of participant and type that is different from (i, t_i) and (j, t_j). Formally, for all $(i, j) \in I \times I$ with $i \neq j$ and all $(t_i, t_j) \in T_i \times T_j$ that comes with positive probability ($p(t_i, t_j) > 0$), there exists a feasible mechanism ν such that

$$U_i(\nu|t_i) > U_i(\mu|t_i),\ U_j(\nu|t_j) < U_j(\mu|t_j), \text{ and}$$
$$(\forall k \in I)(\forall t_k \in T_k) : (k, t_k) \notin \{(i, t_i), (j, t_j)\} \Rightarrow U_k(\nu|t_k) \geq U_k(\mu|t_k).$$

A social planner or an arbitrator may feel constrained at a mechanism where interim utilities are not transferable because he would like to pick an alternative mechanism that is more favorable to individual i, when of type t_i, at the expense of individual j, when of type t_j, but cannot do so because of the incentive and feasibility constraints. In such cases, having more collective decisions available may soften this constraint, and result in a mechanism that is less favorable to individual j of type t_j. This motivates the following axiom.

Restricted Monotonicity (R-MON) *Let $\mathcal{S}$ and $\mathcal{S}'$ be two mechanism design problem. Suppose that $\mathcal{S}'$ differs from $\mathcal{S}$ only in that more collective decisions are available: $I = I'$, $D \subseteq D'$, $T_i = T_i'$, and $u_i(d, t) = u_i'(d, t)$, for each $i \in I$, each $d \in D$, and each $t \in T$. Let $\mu \in \Sigma(\mathcal{S})$ be such that interim utilities are transferable at μ, and let $\mu' \in \Sigma(\mathcal{S}')$. Then $U_i(\mu'|t_i) \geq U_i(\mu|t_i)$, for each $t_i \in T_i$, and each $i \in I$.*

R-MON corresponds to MON applied only when starting from a mechanism at which interim utilities are transferable, and is thus weaker than MON. I am not claiming that monotonicity should systematically be violated when adding collective decisions to a problem whose solution is a mechanism at which interim utilities are not transferable. Instead I am arguing that these are cases where monotonicity *might* be problematic and less appealing. R-MON thus remains silent in those dubious cases.

Most results in social choice under complete information, including Kalai's, are phrased in the space of utilities. This is sometimes referred to as the welfarist assumption. Following the same approach is not exactly feasible under incomplete information, as one needs to know the underlying decisions in order to have the possibility of phrasing the incentive constraints. Yet, one can impose an axiom in that spirit after taking these constraints into account.

Interim Welfarism (I-WELF) *Let* $\mathcal{S} = (I, D, d^*, (T_i)_{i\in I}, p, (u_i)_{i\in I})$ *and* $\mathcal{S}' = (I, D', d^*, (T_i)_{i\in I}, p', (u_i')_{i\in I})$ *be two mechanism design problems. If* $\mathcal{U}(\mathcal{F}(\mathcal{S})) = \mathcal{U}(\mathcal{F}(\mathcal{S}'))$*, then* $\mathcal{U}(\Sigma(\mathcal{S})) = \mathcal{U}(\Sigma(\mathcal{S}'))$*.*

Of course, this definition boils down to the usual notion of welfarism under complete information, i.e. when each type set is a singleton.

The other axiom associated to the welfarist assumption requires that if a feasible mechanism generates the same interim utilities as another mechanism in the solution of a problem, then it also belongs to the solution of that problem.

Exhaustivity (EX) *Let* $\mathcal{S} = (I, D, d^*, (T_i)_{i\in I}, p, (u_i)_{i\in I})$ *be a mechanism design problem. If* $\mu \in \Sigma(\mathcal{S})$ *and* μ' *is a feasible mechanism such that* $U_i(\mu'|t_i) = U_i(\mu|t_i)$*, for all* $t_i \in T_i$ *and all* $i \in I$*, then* $\mu' \in \Sigma(\mathcal{S})$*.*

The next axiom is similar to Harsanyi and Selten's (1972) "Splitting Types" (see also Axiom 6 in Weidner (1992)). Let $\mathcal{S} = (I, D, d^*, (T_i)_{i\in I}, p, (u_i)_{i\in I})$ be a mechanism design problem, and let t_j be a type of an individual j. The mechanism design problem $\mathcal{S}' = (I, D, d^*, (T_i')_{i\in I}, p', (u_i')_{i\in I})$ obtained from S by *splitting* type t_j is defined as follows: $T_i' = T_i$, for all $i \in I \setminus \{j\}$, $T_j' = (T_j \setminus \{t_j\}) \cup \{t_{j1}, t_{j2}\}$, $p'(t') = p(t')$, for all $t' \in T'$ such that $t_j' \notin \{t_{j1}, t_{j2}\}$, $p'(t_{j1}, t_{-j}) = p'(t_{j2}, t_{-j}) = p(t)/2$, for all $t_{-j} \in T_{-j}$, $u_i'(d, t') = u_i(d, t')$, for all $i \in I$, all $d \in D$, and all $t' \in T'$ such that $t_j' \notin \{t_{j1}, t_{j2}\}$, $u_i'(d, t_{j1}, t_{-j}) = u_i'(d, t_{j2}, t_{-j}) = u_i(d, t)$, for all $i \in I$, all $d \in D$, and all $t_{-j} \in T_{-j}$. So individual j's types has been splitted into two sub-types that have the same information and the same preferences as when of type t_j in the original problem. As for j's other types, nothing has changed, and for the other individuals, they see j's two new subtypes half as likely as t_j in the original problem, and as having the exact same properties as t_j otherwise. Such a split is thus an irrelevant change in the way to model the situation at hand,[10] and it should be inconsequential

[10]If a mechanism $\mu : T \to \Delta(D)$ is feasible for $\mathcal{S}$, then $\mu' : T' \to \Delta(D)$ is feasible for $\mathcal{S}'$, where μ' is defined as follows: $\mu'(t') = \mu(t')$, for all $t' \in T'$ such that $t_j' \notin \{t_{j1}, t_{j2}\}$, and $\mu'(t_{j1}, t_{-j}) = \mu'(t_{j2}, t_{-j}) = \mu(t)$, for all $t_{-j} \in T_{-j}$. Conversely, if a mechanism $\mu' : T' \to$

on the solution of the problem. This is indeed the content of the following axiom.

Irrelevant Splitting of a Type (IST) *Let* $\mathcal{S} = (I, D, d^*, (T_i)_{i \in I}, p, (u_i)_{i \in I})$ *be a mechanism design problem, and let* $\mathcal{S}'$ *be the problem derived from* $\mathcal{S}$ *by splitting individual j's type* t_j *into two types* t_{j1} *and* t_{j2}*. If* $\mu \in \Sigma(\mathcal{S})$*, then* $\mu' \in \Sigma(\mathcal{S}')$*, where* $\mu'(t') = \mu(t')$*, for all* $t' \in T'$ *such that* $t'_j \notin \{t_{j1}, t_{j2}\}$*, and* $\mu'(t_{j1}, t_{-j}) = \mu'(t_{j2}, t_{-j}) = \mu(t)$*, for all* $t_{-j} \in T_{-j}$*.*

The next axiom captures the classical property of independence of irrelevant alternatives in our problems with incomplete information.

Independence of Irrelevant Alternatives (IIA) *Let* $\mathcal{S}$ *and* $\mathcal{S}'$ *be two social choice problem. Suppose that* $\mathcal{S}'$ *differs from* $\mathcal{S}$ *only in that more collective decisions are available:* $I = I'$*,* $D \subseteq D'$*,* $T_i = T'_i$*, and* $u_i(d, t) = u'_i(d, t)$*, for each* $i \in I$*, each* $d \in D$*, and each* $t \in T$*. If* $\mu \in \Sigma(\mathcal{S}') \cap \mathcal{F}(\mathcal{S})$*, then* $\mu \in \Sigma(\mathcal{S})$*.*

MON implies IIA. R-MON implies IIA only when interim utilities are transferable in the smaller problem. IIA is thus only a small addition to the list of axioms. IIA captures a property of rationality on the part of the uninformed third party making the collective decision. A violation of IIA would require a strong argument to justify such behavioral irrationality. MON (or R-MON), on the other hand, goes beyond IIA by imposing some principle of distributive justice, thereby narrowing the type of moral preference that this social planner or arbitrator is maximizing. Here, alternative properties might be meaningful as well, and leading to characterization of other moral preferences, or alternative characterizations of the same moral preferences.

The final axiom considered in this paper captures the idea that the solution should not be systematically biased in favor of some specific types of some specific individuals, and should not depend on the way collective decisions are labeled.

Anonymity (AN) *Let* $\mathcal{S} = (I, D, d^*, (T_i)_{i \in I}, p, (u_i)_{i \in I})$ *and* $\mathcal{S}' = (I, D, d^{**}, (T'_i)_{i \in I}, p', (u'_i)_{i \in I})$ *be two mechanism design problems. Suppose that there exist an iso-*

$\Delta(D)$ is feasible for $\mathcal{S}'$, then $U_j(\mu'|t_{j1}) = U_j(\mu'|t_{j2})$ and the mechanism $\mu : T \to \Delta(D)$ is feasible, where $\mu(t') = \mu'(t')$, for all $t' \in T$ such that $t'_j \neq t_j$, and either $\mu(t) = \mu'(t_{j1}, t_{-j})$, for all $t_{-j} \in T_{-j}$, or $\mu(t) = \mu'(t_{j2}, t_{-j})$, for all $t_{-j} \in T_{-j}$.

morphism $f : I \to I$, an isomorphism $g : D \to D$ with $g(d^*) = d^{**}$, and isomorphisms $h_i : T_i \to T'_{f(i)}$ (one for each $i \in I$) such that

1. $(\forall t \in T) : p(t) = p'(h(t))$, *and*

2. $(\forall t \in T)(\forall i \in I)(\forall d \in D) : u_i(d, t) = u'_{f(i)}(g(d), h(t))$,

with the convention $h(t) = (h_i(t_i))_{i \in I}$. Then $\mu' \in \Sigma(\mathcal{S}')$ if and only if $\mu \in \Sigma(\mathcal{S})$, where μ is the mechanism for $\mathcal{S}$ defined as follows: the probability of implementing $d \in D$ when first individual reports $t_1 \in T_1$, ..., and the I^{th} individual reports $t_I \in T_I$ is equal to the probability of implementing $g(d)$ under μ' when individual $f(1)$ reports $h_1(t_1)$, ..., and individual $f(I)$ reports $h_I(t_I)$.

Theorem 2 *There exists a solution satisfying I-EFF, R-MON, AN, I-WELF, EX, IST, and IIA. Any solution that satisfies these seven axioms must coincide with the interim egalitarian criterion on simple problems.*

Proof: A mechanism $\mu \in \mathcal{F}(\mathcal{S})$ belongs to the *interim lex-min solution* of the mechanism design problem $\mathcal{S} = (I, D, d^*, (T_i)_{i \in I}, p, (u_i)_{i \in I})$, $\mu \in \Sigma^{lex}(\mathcal{S})$, if and only if $\theta(\mathfrak{u}(\mu))$ maximizes $\theta(\mathfrak{u})$ according to the lexicographic order over all $\mathfrak{u} \in \mathcal{U}(\mathcal{F}(S))$, where $\theta : \times_{i \in I} \mathbb{R}_+^{T_i} \to \times_{i \in I} \mathbb{R}_+^{T_i}$ is the function that rearrange the components of a vector increasingly.

Let $\mathcal{S} = (I, D, d^*, (T_i)_{i \in I}, p, (u_i)_{i \in I})$ be a mechanism design problem. Observe that, if interim utilities are transferable at $\mu \in \Sigma^{lex}(\mathcal{S})$, then $U_i(\mu|t_i) = U_j(\mu|t_j)$, for all $t_i \in T_i$, all $t_j \in T_j$, and all $i, j \in I$. Indeed, otherwise, there exist (i, t_i) and (j, t_j) with $i \neq j$ such that $U_i(\mu|t_i) < U_j(\mu|t_j)$. Since interim utilities are transferable at μ, there exists a mechanism $\mu' \in \mathcal{F}(S)$ such that $U_i(\mu'|t_i) > U_i(\mu|t_i)$, $U_j(\mu'|t_j) < U_j(\mu|t_j)$, and $U_k(\mu'|t_k) \geq U_k(\mu|t_k)$, for all (k, t_k) different from (i, t_i) and (j, t_j). For each $\epsilon \in]0, 1[$, let $\mu^\epsilon := \epsilon\mu' + (1 - \epsilon)\mu \in \mathcal{F}(\mathcal{S})$. For ϵ small enough, $\theta(\mathfrak{u}(\mu^\epsilon))$ strictly dominates $\theta(\mathfrak{u}(\mu))$ according to the lexicographic order, thereby contradicting $\mu \in \Sigma^{lex}(\mathcal{S})$. So it must be indeed that all the components of $\mathfrak{u}(\mu)$ are identical. Consider now a larger mechanism design problem $\mathcal{S}'$, as defined in R-MON. Since $\mu \in \mathcal{F}(\mathcal{S}')$, it must be that the smallest component of $\mathfrak{u}(\mu')$ is larger of equal to the identical components of $\mathfrak{u}(\mu)$, for each $\mu' \in \Sigma^{lex}(\mathcal{S}')$. Hence Σ^{lex} satisfies R-MON, as desired.

The fact that Σ^{lex} satisfies IST follows from the fact that the set of interim utilities that can be achieved in the splitted problem is the set of interim utilities that can be achieved in the original problem, except that the utility associated to the (j, t_j) component now appears twice (once for (j, t_{j1}) and once for (j, t_{j2}), cf. footnote 10). The very definition of Σ^{lex} makes it straightforward to check that it satisfies the five other axioms. The fact that any solution that satisfies the axioms must coincide with the interim egalitarian criterion on simple problems is proved in the Appendix. A more informal roadmap of the argument is provided in the second supplemental appendix. ■

5. EGALITARIANISM AND UTILITARIANISM RECONCILED

The egalitarian solution requires the possibility of measuring the individuals' utility gains in some common units. Also, when such measurements are possible, the egalitarian principle often comes in conflict with other normative criteria, the most prominent alternative being the utilitarian principle. Various authors (see e.g. Harsanyi (1963), Shapley (1969), Yaari (1981)) showed that these two difficulties can be resolved simultaneously in the following sense. There is a unique way to rescale the participants' utilities so that there exists a feasible option that is optimal according to both the egalitarian and the utilitarian criteria (in the rescaled utilities). In addition, any such option is optimal according to Nash' product criterion (either in the original or in the rescaled utilities, since that criterion is invariant to linear transformations). Nehring (2004) has argued that maximizing the ex-ante sum of the individuals' utilities is the natural extension of the utilitarian criterion to problems with incomplete information. The analysis of the previous section suggests that the interim egalitarian criterion is a natural extension of the egalitarian solution to frameworks with incomplete information (at least for simple problems). One may thus wonder whether there is a way, in each mechanism design problem, to make both criteria compatible by rescaling the participants' interim utilities.

The next theorem answers positively (independently of whether the problem is simple or not). In addition, the resulting solution amounts to selecting the incentive compatible mechanism that is maximal according to Harsanyi and Selten's (1972) weighted Nash product (see Myerson (1979)). Weidner (1992) obtained a different characterization of that solution in a similar framework (based on axioms adapted from Nash (1950) and Harsanyi and Selten (1972)). He observes that his result is valid only when types are independent, an assumption that plays no role in the following theorem.

Theorem 3 *Let $\mathcal{S} = (I, D, d^*, (T_i)_{i \in I}, p, (u_i)_{i \in I})$ be a mechanism design problem, and let $\mu^* \in \mathcal{F}(\mathcal{S})$. Then*

$$\mu^* \in \arg\max_{\mu \in \mathcal{F}(\mathcal{S})} \Pi_{i \in I} \Pi_{t_i \in T_i} [U_i(\mu|t_i)]^{p(t_i)} \tag{3}$$

if and only if μ^* satisfies the two following conditions for some $\lambda \in \times_{i \in I} \mathbb{R}^{T_i}_{++}$:

1. $\mu^* \in \arg\max_{\mu \in \mathcal{F}(\mathcal{S})} \sum_{i \in I} \sum_{t_i \in T_i} p(t_i) U_i^{\lambda}(\mu|t_i)$,
2. $(\forall t \in T)(\forall i \in I)(\forall j \in I) : U_i^{\lambda}(\mu^*|t_i) = U_j^{\lambda}(\mu^*|t_j)$,

where $U_i^{\lambda}(\mu|t_i) := \lambda_i(t_i) U_i(\mu|t_i)$, for each $t_i \in T_i$ and each $i \in I$.

Proof: Let $W^* = \Pi_{i \in I} \Pi_{t_i \in T_i} [U_i(\mu^*|t_i)]^{p(t_i)}$. Since it is assumed that there exists at least one element of $\mathcal{U}(\mathcal{F}(\mathcal{S}))$ with only strictly positive components, it must be that $W^* > 0$ and $(U_i(\mu^*|t_i))_{t_i \in T_i, i \in I} >> 0$ under (3). The sets $\mathcal{U}(\mathcal{F}(\mathcal{S}))$ and

$$\{\mathfrak{u} \in \times_{i \in I} \mathbb{R}^{T_i}_{+} | \Pi_{i \in I} \Pi_{t_i \in T_i} \mathfrak{u}_i(t_i)^{p(t_i)} \geq W^*\}$$

are both closed and convex. Under (3), their intersection is the singleton $\{(U_i(\mu^*|t_i))_{t_i \in T_i, i \in I}\}$. Hence the separating hyperplane theorem implies that (3) is equivalent to the existence of a vector $l \in \times_{i \in I} \mathbb{R}^{T_i}$ for wich the two following conditions hold:

1. $\mu^* \in \arg\max_{\mu \in \mathcal{F}(\mathcal{S})} \sum_{i \in I} \sum_{t_i \in T_i} l_i(t_i) U_i(\mu|t_i)$,
2. l is proportional to the gradient of the curve $\{\mathfrak{u} \in \times_{i \in I} \mathbb{R}^{T_i}_{+} | \Pi_{i \in I} \Pi_{t_i \in T_i} \mathfrak{u}_i(t_i)^{p(t_i)} \geq W^*\}$ at $(U_i(\mu^*|t_i))_{t_i \in T_i, i \in I}$.

The second condition itself is equivalent to the existence of a strictly positive number α such that

$$l_i(t_i) = \frac{\alpha p(t_i)}{U_i(\mu^*|t_i)},$$

for all $t_i \in T_i$ and all $i \in I$. The result thus follows, by taking $\lambda_i(t_i) = l_i(t_i)/p(t_i)$, for all $t_i \in T_i$ and all $i \in I$. ■

While first embraced by Myerson (1979) as a reasonable bargaining solution, the criterion derived in Theorem 3 has been subsequently criticized by Myerson (1984a) because of its sensitivity to joint changes of the utility functions and the individuals' beliefs that leave interim preferences unchanged (see his Probability-Invariance axiom). Though not a bargaining solution, observe nevertheless that the interim egalitarian criterion does satisfy Myerson's probability invariance (I-WELF indeed implies it). In view of Theorem 3, the fact that the Harsanyi-Selten solution violates it can be traced back to the fact that the ex-ante utilitarian principle violates it as well. To the extent that ex-ante utilitarianism has some separate appeal and justification when understood as a social welfare criterion, Theorem 3 shows that the weighted Nash product derived by Harsanyi and Selten is perhaps more justified when understood as defining a social welfare ordering instead of a bargaining solution. Observe also that the interim egalitarian criterion being rooted in interim utilities, its definition (as well as its characterization in Theorem 2) extends to the case of non-common priors. This is not the case for the ex-ante utilitarian criterion, as shown by Nehring (2004), nor the Harsanyi-Selten criterion, as there is no unique way of defining marginal probabilities of types in the absence of a common prior.

6. APPLICATIONS

This section is devoted to the analysis of two classical quasi-linear examples in perspective of the analysis developed in previous sections. The first example is about the production of a public good and how to share its cost, while the second example is about fair terms of trade. The objective is to understand what the interim egalitarian criterion entails, when it exists, and to see how one can apply Theorem 3 to compute the Harsanyi/Selten/Myerson solution. We will also see that the interim lexmin solution introduced in the proof of Theorem 2 seems to select sensible mechanisms in problems that are not simple.

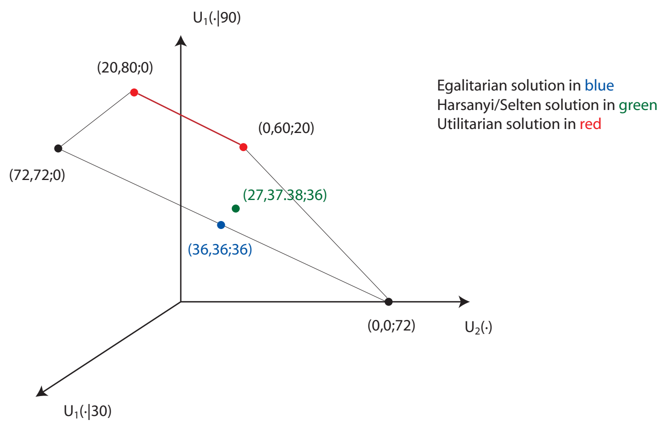


Figure 1: Illustration of Example 2 for p=1/10

Example 1 *Consider the following problem, which is similar*[11] *to the main example in Myerson (1979). Two individuals have the option to cooperate by investing in a public project that costs* \$100*, which is known to bring a satisfaction of* \$90 *to the second individual, and either a satisfaction of* \$30*, with probability* p*, or of* \$90*, with probability* $1-p$*, to the first individual. Formally,* $I = \{1, 2\}$*,* $D = \{(x, m) \in \{0, 1\} \times \mathbb{R}^2 | m_1 + m_2 \leq -100x\}$*, where* $x = 0$ *means that the project is not carried out,* $x = 1$ *means that the project is carried out,* $d^* = (0, (0, 0))$*,* $T_1 = \{30, 90\}$*,*[12] $u_1((x, m), 30) = 30x + m_1$*,* $u_1((x, m), 90) = 90x + m_1$*, and* $u_2((x, m), 30) = u_2((x, m), 90) = 90x + m_2$*.*

$\mathcal{U}(\mathcal{F}(\mathcal{S}))$ *thus coincides with the set of vectors*

$$(30\bar{x}(30)+\bar{m}_1(30), 90\bar{x}(90)+\bar{m}_1(90); 90(p\bar{x}(30)+(1-p)\bar{x}(90))+p\bar{m}_2(30)+(1-p)\bar{m}_2(90)),$$

that satisfy the following constraints:

$$\bar{m}_1(30) + \bar{m}_2(30) \leq -100\bar{x}(30) \text{ and } \bar{m}_1(90) + \bar{m}_2(90) \leq -100\bar{x}(90), \quad (4)$$

$$30(\bar{x}(90) - \bar{x}(30)) \leq \bar{m}_1(30) - \bar{m}_1(90) \leq 90(\bar{x}(90) - \bar{x}(30)), \quad (5)$$

where $\bar{x}(t_1)$ *denotes the expected probability of implementing the public project*

[11] The only difference is that I consider a fully quasi-linear problem, while Myerson allows for compensations only indirectly by allowing for randomization over three deterministic collective decisions, 'no public project,' 'public project funded by 1,' and 'public project funded by 2,' which prevents the possibility of compensations when the public project is not implemented.

[12] T_2 is ignored to make notations lighter. This is inconsequential, since T_2 is a singleton, the second individual having no private information.

when the first individual reports t_1*, and* $\bar{m}_i(t_1)$ *is the expected monetary payoff received by individual* i *when the first individual reports* t_1*. Indeed, a feasible mechanism must select lotteries defined over* D*, and taking expectations of the feasibility constraints will give (4). Conversely, any vector* $(\bar{x}(t_1), \bar{m}_1(t_1), \bar{m}_2(t_1)))$ *can be seen as the expectation of the lottery that picks* $[x = 0, m = (0, 0)]$ *with probability* $1 - \bar{x}(t_1)$*, and* $[x = 1, m = (\bar{m}_1(t_1)/\bar{x}(t_1), \bar{m}_2(t_1)/\bar{x}(t_1))]$ *with probability* $\bar{x}(t_1)$*, which selects elements of* D *whenever (4) is satisfied. Inequalities in (5), on the other hand, represent the incentive contraints. A traditional reasoning allows to conclude that incentive efficiency implies that* $\bar{x}(90) = 1$*, and both inequalities in (4), as well as the second inequality of (5) are binding. If* $p = 1/10$*, as in Myerson (1979), then simple computations imply that* $\mathcal{U}^{eff}(\mathcal{F}(\mathcal{S}))$ *is the set of vectors*

$$(30\bar{x}(30) + \bar{m}_1(30), 90\bar{x}(30) + \bar{m}_1(30); 72 - 82\bar{x}(30) - \bar{m}_1(30)), \qquad (6)$$

with $\bar{x}(30) \in [0, 1]$ *and* $\bar{m}_1(30) \in \mathbb{R}$*. It is then easy to check that a mechanism meets the interim egalitarian criterion if and only if* $\bar{x}(30) = 0$*,* $\bar{m}_1(30) = 36$*,* $\bar{m}_2(30) = -36$*,* $\bar{x}(90) = 1$*,* $\bar{m}_1(90) = -54$*, and* $\bar{m}_2(90) = -46$*, in which case both individuals get an expected utility of* 36*, independently of the true state. The problem is thus simple. Theorem 2 implies that any solution that satisfies the axioms must pick that specific mechanism in this numerical example. The ex-ante utilitarian solution, on the other hand, will contain only pooling mechanisms with* $\bar{x}(30) = \bar{x}(90) = 1$*,* $\bar{m}_1(90) = \bar{m}_1(30)$*, and* $\bar{m}_2(30) = \bar{m}_2(90) = -100 - \bar{m}_1(30)$*, leaving the choice of* $\bar{m}_1(30)$ *open. So, in this example, it does not refine the set of ex-ante incentive efficient mechanisms. The interim utilities are* $30 + \bar{m}_1(30)$ *for the first individual when he values the public project at* 30*,* $90 + \bar{m}_1(30)$ *for the first individual when he values the public project at* 90*, and* $90 - \bar{m}_1(30)$ *in expectation for the second individual. One can apply Theorem 3 to find Harsanyi and Selten's (1972) weighted Nash product over* $\mathcal{U}(\mathcal{F}(\mathcal{S}))$*. One concludes from (6) that the vector* $(2/15, 13/15; 1)$ *is orthogonal to* $\mathcal{U}^{eff}(\mathcal{F}(\mathcal{S}))$*. If the weighted Nash optimum is reached at a point in the interior of* $\mathcal{U}^{eff}(\mathcal{F}(\mathcal{S}))$*, then it must be for* $\lambda = (20/15, 26/27, 1)$*, given the first condition in Theorem 3.*

Solving then for the two equations implied by the second condition in Theorem 3, one obtains $\bar{x}(30) = 189/1092 \cong 0.173$*,* $\bar{m}_1(30) = 567/26 \cong 21.81$*,* $\bar{m}_2(30) = -1017/26 \cong -39.12$*,* $\bar{x}(90) = 1$*,* $\bar{m}_1(90) = -684/13 \cong -52.62$*, and* $\bar{m}_2(90) = -616/13 \cong -47.38$*, which turns out to be a feasible mechanism. The Harsanyi-Selten solution being unique, we are thus done solving the problem. The interim utilities are* 27 *for the first individual when he values the public project at* 30*,* $486/13 \cong 37.38$ *for the first individual when he values the public project at* 90*, and* 36 *in expectation for the second individual.*

The set of incentive efficient mechanisms and the three solutions are represented in the space of interim utilities on Figure 1. The utilitarian solution does not place any weight on the distribution of the gains from cooperation, and thereby allows to be more efficient in aggregate, the public project being implemented for sure independently of individual 1*'s type. On the other hand, this is achieved at the cost of equity, being too generous towards individual* 1 *when he values the public project at* \$90*, because one cannot rely on him to report his type for free. The second individual, for instance, cannot expect a gain larger than* \$20*, while the aggregate benefit is* \$80 *with probability* 9/10 *and* \$20 *with probability* 1/10*. The first individual's information rent vanishes at the egalitarian solution, in that both types of the first individual receive the same interim utility, but this comes at the cost of not implementing the public project when the first individual does not care much for it. Note that even though some mutually beneficial cooperative opportunities are not exploited, the mechanism itself, on the other hand, is incentive efficient. As already hinted by Theorem 3, the Harsanyi/Selten solution strikes a compromise between two conflicting points of view, allowing the public project to be realized with a positive probability when the first individual has the low type, but not systematically, so as to avoid being too soft on individual* 1 *when of the high type because of his informational advantage.*

Suppose now that there is a probability 9/10 *that the first individual values the public project at* \$30*, instead of* 1/10*. Starting from inequalities (4) and (5), it is easy to check that* $\mathcal{U}^{eff}(\mathcal{F}(\mathcal{S}))$ *is the set of vectors*

$$(30 + \bar{m}, 90 + \bar{m}; -10 - \bar{m}), \tag{7}$$

with $\bar{m} \in \mathbb{R}$. In other words, incentive efficient mechanisms must be pooling, with the public project being implemented for sure regardless of the first individual's report, and the expected monetary compensation being constant. In this case, the first individual has a larger payoff when of the high type than when of the low type at any incentive efficient mechanism, and the problem is not simple. The interim lex-min solution defined for all problems (see proof of Theorem 2) seems to make sense. Since individual 1 *systematically enjoys a payoff that is larger by a constant amount of* \$60 *when of a high type compared to the low type alternative, he is left aside and all what matters is to choose m so as to equalize the second individual's expected payoff with the first individual's payoff when he values the public project at* \$30. *Hence one must choose $\bar{m} = -\$20$, and the associated vector of interim utilities is $(10, 70; 10)$. The ex-ante utilitarian solution does not refine the set of incentive efficient mechanisms, while the Harsanyi-Selten solution selects a $\bar{m}$ that is slightly smaller than $-\$20$, thereby being slightly more generous towards the second individual.*[13]

More generally, the problem is simple whenever $p < 3/4$, and the interim egalitarian criterion will prescribe to implement the public project if and only if the first individual reports his high type. In that case, he pays \$90, *while the second individual pays* \$10. *In addition, a monetary transfer of* $\$(1-p)40$ *goes from the second to the first individual, independently of the type reported. The interim utility enjoyed by each type of each individual is $(1-p)40$. The problem is not simple when $p \geq 3/4$, in which case the interim lex-min solution has the public project implemented independently of the first individual's report, with* 1 *paying* 20% *of the cost, and* 2 *the remaining* 80%.

Example 2 *Consider the following classical bilateral trade problem, as in Myerson (1991, Section 10.3). A first individual owns one unit of a divisible good that is worth more to a second individual than to him. The good can be of relatively low quality, with probability p, in which case the good is worth*

[13] Not surprisingly, the Harsanyi-Selten solution converges towards the egalitarian solution, as p gets closer to 1, as the problem then approaches a problem with both complete information and transferable utilities, in which case the Nash and the egalitarian solutions coincide.

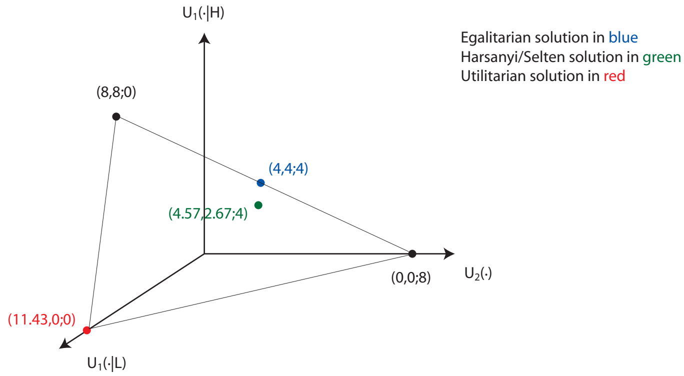


Figure 2: Illustration of Example 3 for p=4/5

\$20 per unit to the first individual and \$30 per unit to the second individual, or of relatively high quality, with probability $1-p$*, in which case the good is worth \$40 per unit to the first individual and \$50 per unit to the second individual. The true quality of the object is known to the seller only. Cooperating here means agreeing on a quantity to trade against some monetary compensation, as a function of what the seller reveals about the quality of the good he owns. The problem is to find a fair compensation scheme. Formally,* $I=\{1,2\}$*,* $D=\{(x,m)\in[0,1]\times\mathbb{R}^2|m_1+m_2\leq 0\}$*, where* x *represents the quantity traded,* $d^*=(0,(0,0))$*,* $T_1=\{L,H\}$*,*[14] $u_1((x,m),L)=m_1-20x$*,* $u_1((x,m),H)=m_1-40x$*,* $u_2((x,m),L)=30x+m_2$*, and* $u_2((x,m),H)=50x+m_2$*.*

$\mathcal{U}(\mathcal{F}(\mathcal{S}))$ *thus coincides with the set of vectors* $(m_1(L)-20x(L),m_1(H)-40x(H);p(30x(L)+m_2(L))+(1-p)(50x(H)+m_2(H)))$ *that satisfy the following constraints:*[15]

$$m_1(L)+m_2(L)\leq 0 \text{ and } m_1(H)+m_2(H)\leq 0, \tag{8}$$

$$20(x(L)-x(H))\leq m_1(L)-m_1(H)\leq 40(x(L)-x(H)), \tag{9}$$

where $(x(t_1),m(t_1))\in[0,1]\times\mathbb{R}$*, for each* $t_1\in\{L,H\}$*. Standard arguments imply that incentive efficiency require* $x(L)=1$*, and both inequalities in (8),*

[14] As in the previous example, T_2 is ignored to make notations lighter. This is inconsequential, since T_2 is a singleton, the second individual having no private information.

[15] Utilities being linear in both the good and money, there is no loss of generality in discussing only deterministic mechanisms.

as well as the first inequality of (9) to be binding. If $p = 4/5$*, as in Myerson (1991), then* $\mathcal{U}^{eff}(\mathcal{F}(\mathcal{S}))$ *is the set of vectors*

$$(m - 20x, m - 40x; 8 + 26x - m), \tag{10}$$

with $(x, m) \in [0, 1] \times \mathbb{R}$ *representing the quantity traded and the monetary compensation from* 2 *to* 1 *when the first individual reports* H*.* $\mathcal{U}^{eff}(\mathcal{F}(\mathcal{S}))$ *thus coincides with the triangle whose vertices are* $(80/7, 0; 0)$*,* $(8, 8, 0)$*, and* $(0, 0; 8)$*, as represented on Figure 2. The interim egalitarian criterion selects the mechanism with* $(x, m) = (0, 4)$*, in which case both individuals enjoy an interim utility of* 4 *whatever their types (this remains true for any* $p > 1/3$*). The problem is simple. Theorem 2 implies that any solution that satisfies the axioms must select that specific mechanism in this numerical example. The ex-ante utilitarian solution, on the other hand, selects the mechanism that is most advantageous to the first individual when he is of a low type -* $(x, m) = (4/7, 160/7)$ *leading to the extreme vector of interim utilities* $(80/7, 0; 0)$*. Contrarily to the previous example, the ex-ante utilitarian principle does refine the set of feasible mechanisms that are ex-ante efficient (which leads to interim utilities along the segment that joins* $(0, 0; 8)$ *to* $(80/7, 0; 0)$*). One can apply Theorem 3 to find Harsanyi and Selten's (1972) weighted Nash product over* $\mathcal{U}(\mathcal{F}(\mathcal{S}))$*. One concludes from (10) that the vector* $(7/10, 3/10; 1)$ *is orthogonal to* $\mathcal{U}^{eff}(\mathcal{F}(\mathcal{S}))$*. If the weighted Nash optimum is reached at a point in the interior of* $\mathcal{U}^{eff}(\mathcal{F}(\mathcal{S}))$*, then it must be for* $\lambda = (7/8, 3/2; 1)$*, given the first condition in Theorem 3. Solving then for the two equations implied by the second condition in Theorem 3, one obtains the feasible mechanism corresponding to the combination* $(x, m) = (2/21, 136/21)$*. The Harsanyi-Selten solution being unique, we are thus done solving the problem. The interim utilities are* $32/7 \cong 4.57$ *for the first individual when of a low type,* $56/21 \cong 2.67$ *for the first individual when of a high type, and* 4 *in expectation for the second individual.*

More generally, the problem is simple whenever $p > 1/3$*, and the interim egalitarian criterion prescribes trade in full against* \$34 *when the first individual's report is* L*, and no trade, but still with a transfer of* \$4 *from* 2 *to* 1 *when the report is* H*. The interim utilities are* $(4, 4; 4)$*. If, on the other hand, the* *high type is rather likely to occur ($p \leq 1/3$) then incentive efficiency occurs only at pooling mechanisms where full trade occurs independently of the first individual's report. In that case, the low type always gets a larger utility than the high type, and the problem is not simple. The interim lex-min solution defined for all problems (see proof of Theorem 2) then follows the pragmatic principle of equalizing the two remaining payoffs. This leads to trade in full with a transfer of* $\$45 - 10p$ *from* 2 *to* 1*, independently of individual* 1*'s report.*

7. DIRECTIONS FOR FUTURE RESEARCH

The purpose of this section is to present directions for future research on social choice in mechanism design. Of particular interest, I will argue that considering problems of incomplete information should lead to an even richer debate on what is equitable, due to the presence of multiple reasonable norms that differ only when information is asymmetrically distributed.

Mechanisms that meet the interim egalitarian criterion are obviously equitable in the following sense: an uninformed third party (social planner or arbitrator) can be sure that all the individuals enjoy the same expected benefits given their own private information, and this whatever the actual profile of types. Yet, I believe that there might be alternative appealing ways to proceed. Here is an example of solution that takes more into account what the individuals know to determine what is equitable. I will restrict attention to the class of quasi-linear problems. Given his own information, an individual i of type t_i can evaluate the total surplus achieved by an incentive compatible mechanism μ:

$$TS_{i,t_i}(\mu) = \sum_{t_{-i} \in T_{-i}} p(t_{-i}|t_i) \sum_{i \in I} u_i(\mu(t), t).$$

From his point of view, his share of the total surplus realized by μ is then

$$s_{i,t_i}(\mu) = \frac{U_i(\mu|t_i)}{TS_{i,t_i}(\mu)}$$

(with the convention $s_{i,t_i}(\mu) = 1/\#I$ if both $U_i(\mu|t_i)$ and $TS_{i,t_i}(\mu)$ are equal to zero). One may have to accept inefficiency in some type profiles in order to satisfy the incentive constraints. For instance, the public good is not always implemented at incentive efficient mechanisms in Example 1, when $t_1 = 30$

and $p < 3/4$, and trade does not always occur at incentive efficient mechanisms in Example 2, when $t_1 = 40$ and $p > 1/3$. More generally, we know from Myerson and Satterthwaite (1983) that ex-post efficiency, interim individual rationality and incentive compatibility may be incompatible. This is why it is more natural for the definition of the total surplus to be endogenous to the mechanism considered, instead of taking the maximal total surplus that could be achieved in the absence of incentive constraints. Another meaningful definition of egalitarianism under incomplete information would then be to select an interim individually rational and incentive efficient mechanism μ that equalizes the shares in each type profile: $s_{i,t_i}(\mu) = s_{j,t_j}(\mu)$, for all $i, j \in I$ and all $t \in T$. Perfect equalization being not always possible, as with the interim egalitarian criterion, it is natural to consider the weaker criterion of maximizing, according to the lexicographic ordering, the vector $\theta(s(\mu))$ over the set of individually rational and incentive efficient mechanisms μ. The associated solution will be denoted Σ^*. Notice that, if an interim individually rational and incentive efficient mechanism μ for a mechanism design problem $\mathcal{S}$ is such that $s_{i,t_i}(\mu) = s_{j,t_j}(\mu)$, for all $i, j \in I$ and all $t \in T$, then $\mu \in \Sigma^*(\mathcal{S})$ and $s_{i,t_i}(\mu) = 1/\#I$, for all $i \in I$ and $t_i \in T_i$. Indeed, if σ is the share that is common to all the individuals of any type, then we have:

$$\begin{aligned}\textstyle\sum_{t\in T} p(t) \sum_{i\in I} u_i(\mu(t), t) &= \textstyle\sum_{i\in I} \sum_{t_i\in T_i} p(t_i) U_i(\mu|t_i)\\ &= \textstyle\sum_{i\in I} \sum_{t_i\in T_i} p(t_i) \sigma TS_{i,t_i}(\mu)\\ &= \textstyle\sigma(\#I) \sum_{t\in T} p(t) \sum_{i\in I} u_i(\mu(t), t),\end{aligned}$$

which implies $\sigma = 1/\#I$. Any mechanism μ' such that $\theta(s(\mu'))$ lexicographically dominates $\theta(s(\mu))$ will lead to similar equations, except that the second equality is now changed into a strict inequality for $\sigma = 1/\#I$, thereby leading to a contradiction and showing that $\mu \in \Sigma^*(\mathcal{S})$. It also implies, conversely, that any interim individually rational and incentive efficient mechanism μ such that $s_{i,t_i}(\mu) = 1/\#I$, for all $i \in I$ and $t_i \in T_i$, must belong to $\Sigma^*(\mathcal{S})$. Of course, Σ^* coincides with the regular egalitarian criterion when information is complete (as does the interim egalitarian criterion), i.e. when the type sets contain a single element. Easy computations in Example 1 show that, for any $p < 3/4$, equalization of the shares is feasible and a mechanism belongs to Σ^* if and only if $\bar{x}(30) = 4/7$, $\bar{m}(30) = (-80/7, -320/7)$, $\bar{x}(90) = 1$, and $\bar{m}(30) = (-50, -50)$. Things become even clearer if the collective decision implemented when $t_1 = 30$ is decomposed as an explicit lottery over D: the public project is implemented with probability 4/7, in which case the first individual pays \$20 and the second individual pays \$80 (this distribution of the cost is ex-post egalitarian), while there is no payment and no transfer when the public project is not implemented. The interim utilities are 40/7 for the first individual of type $t_1 = 30$, 40 for the first individual of type $t_1 = 90$, and $40 - \frac{240}{7p}$ for the second individual. So, on the one hand, one could say that the mechanism is not equitable from the point of view of an uninformed third party (social planner or arbitrator), in that different individuals enjoy different levels of satisfaction, but on the other hand the mechanism looks equitable from the point of view of the individuals themselves, given their private information. One must resort to the lexicographic ordering when $p \geq 3/4$, in which case the public project is implemented regardless of the first individual's report, who has to pay $\$\frac{130-90p}{5-3p}$, while the second individual pays $\$\frac{370-210p}{5-3p}$. In Example 2, for any $p > 1/3$, equalization of the shares is feasible and a mechanism belongs to Σ^* if and only if $x(L) = 1$, $m(L) = (25, -25)$, $x(H) = 1/5$, and $\bar{m}(H) = (9, -9)$ (or, equivalently, the second individual pays \$45 per unit to the first conditional on him reporting 40 - again this is the ex-post egalitarian outcome). The interim utilities are 5 for the first individual of type $t_1 = L$, 1 for the first individual of type $t_1 = H$, and $1 + 4p$ for the second individual, which again is different from the prescription made by the interim egalitarian criterion. One must resort to the lexicographic ordering when $p \leq 1/3$, in which case the good is traded in full regardless of the first individual's report and he receives $\$45 - 10p$ from the second individual, which, this time, happens to coincide with the outcome of Σ^{lex} (defined in the proof of Theorem 2).

The fact that various extensions seem sensible is a feature that makes the subject of social choice under incomplete information richer. Additional and sharper axiomatic characterizations are needed to capture the essence of what distinguishes these various normative criteria. Since they all coincide in the special case of complete information, the key normative question one must address is how to treat information in determining what is equitable, and more specifically how to treat different types of a same individual. The interim egalitarian criterion (as well as the ex-ante utilitarian principle, and Harsanyi and Selten's (1972) weighted Nash product) ends up treating different types of individuals as different individuals in the way it is computed. Analogous treatments have already appeared in very different contexts, cf. the notion of "type-agent" introduced by Harsanyi (1967-68) to define Bayesian Nash equilibria, and that also played a key role in defining a notion of core under asymmetric information (de Clippel (2007)). Yet it seems that other inter-type compromises might have some normative appeal as well. This is the right place to mention a third notion of equity that would also coincides with egalitarianism in quasi-linear social choice problems under complete information (with risk-neutral individuals), but leads to different recommendations under incomplete information. Procedural justice offers an interesting alternative to the consequentialist approach that discusses equity exclusively in terms of the outcomes selected in various problems. What matters now is to let the individuals themselves select the social outcome via a procedure (or a game form) that is fair, in a broad sense of giving them equal opportunities. "Random Dictatorship," choosing with equal probability one of the individuals to act as a dictator, provides an example of such procedure, at least when the resulting equilibrium outcome is Pareto efficient (as it is under complete information if the problem is quasi-linear and individuals are risk-neutral). This example of fair procedure is a bit extreme, since the outcome when a dictator has been selected is clearly unfair, but at least all the individuals are in an equal position ex-ante. Notice that "Random Dictatorship" is also a corner-stone of Myerson's (1984a) theory of bargaining. Though trivial, there is an interesting equivalence under complete information between the egalitarian principle and the equilibrium outcome of that procedure in quasi-linear problems with risk-neutral individuals. This general equivalence breaks down when information is incomplete. A key insight from Myerson's work (1983 and 1984a) (see also Maskin and Tirole (1992)) is that being a dictator defines an implicit inter-type compromise in some problems, and I will simply observe now that this implicit compromise is incompatible with both the interim egalitarian criterion and Σ^* in both Examples 1 and 2.Indeed, it is easy to check in Example 1 that the principal-agent game when the first individual is a dictator admits a unique weak sequential equilibrium outcome, with the public project being implemented, the first individual paying \$10 and the second paying \$90, independently of the reported type. The associated interim utilities are $(20, 80; 0)$. The unique weak sequential equilibrium outcome when the second individual is the dictator and $p < 3/4$ implies that the project is implemented only when the reported type is 90, in which case the first individual pays \$90 and the second pays \$10. The associated interim utilities are $(0, 0; 80(1-p))$. When $p > 3/4$, the public project is always implemented, with the first individual paying \$30 and the second paying \$70. The associated interim utilities are $(0, 60; 20)$. Random Dictatorship thus leads to the interim utilities $(10, 40, 40(1-p))$ if $p < 3/4$ and $(10, 70, 10)$ if $p > 3/4$. We see that the interim egalitarian criterion coincides with the random dictatorship outcome if and only if $p > 3/4$. Σ^* differs both from the interim egalitarian criterion (or Σ^{lex} if the problem is not simple) and the random dictatorship outcome for any p. In Example 2, if $p > 1/3$, then the unique weak sequential equilibrium outcome when the first individual is the dictator leads to the good being traded at the highest possible price - \$30 per unit if the reported type is L, and \$50 per unit if the reported type if H - but only a third is traded when the reported type is H, while the good is traded in full when the reported type is H. If the second individual is the dictator with $p > 1/3$, then the good is traded if and only if the reported type is L. in which case it is traded in full at the lowest possible price - \$20 per unit. Random Dictatorship thus leads to the interim utilities $(5, 5/3, 5p)$, which differ from both Σ^{lex} and Σ^*. If $p < 1/3$, then trade occurs in full independently of the first individual's report. The price is $\$50 - 20p$ if the first individual is the dictator, and \$40 if the second individual is the dictator. Random Dictatorship thus leads to the interim utilities $(25 - 10p, 5 - 10p, 5 - 10p)$, as with Σ^{lex} and Σ^*.

The two alternative egalitarian criteria discussed in this last section also highlight the strength of I-WELF, as neither Σ^*, nor the Random Dictatorship solution, satisfy it. To see that, notice first that $\mathcal{U}(\mathcal{F}(\mathcal{S}))$ in Example 2 with $p = 4/5$ is the convex hull of $(0, 0; 0)$ and the three vectors shown on Figure 2, i.e. $(11.43, 0; 0)$, $(8, 8; 0)$, and $(0, 0; 8)$. This follows immediately from the previous characterization of the incentive efficiency frontier and the fact that $U_1(\mu|L) \geq U_1(\mu|H)$ at any incentive compatible mechanism (indeed, $m_1(L) - 20x(L) \geq m_1(H) - 20x(H) \geq m_1(H) - 40x(H)$, where the first inequality follows from the incentive constraint and the second follows from the fact that $x(H)$ is non-negative). Now, similar computations would show that the set of interim utilities that can be achieved by mechanisms that are individually rational and incentive compatible remains unchanged if p is 3/4 instead of 4/5, the buyer's value in the low type is \$92/3 instead of \$30, and the buyer's value in the high type is \$44 instead of \$50. Yet Σ^* now leads to the interim utilities $(16/3, 16/33; 136/33)$ instead of $(5, 1; 21/5)$, and Random Dictatorship now prescribes $(11/3, 8/9; 11/4)$ instead of $(5, 5/3; 4)$.[16]

References

d'Aspremont, C., and L.-A. Gérard-Varet, 1979. Incentives and Incomplete Information. *Journal of Public Economics* **11**, 25-45.

d'Aspremont, C., and L.-A. Gérard-Varet, 1982. Bayesian Incentive Compatible Beliefs. *Journal of Mathematical Economics* **10**, 83-103.

de Clippel, G., 2007. The Type-Agent Core for Exchange Economies with Asymmetric Information. *Journal of Economic Theory* **135**, 144-158.

de Clippel, G., 2010. A Comment on "The Veil of Public Ignorance." Brown University, Working Paper 2010-3.

de Clippel, G., and E. Minelli, 2004. Two-Person Bargaining with Verifiable Information. *Journal of Mathematical Economics* **40**, 799-813.

de Clippel, G., V. Naroditskiy, and A. Greenwald, 2009. Destroy to Save. *Proceedings of the 10th ACM Conference on Electronic Commerce.*

de Clippel, G., D. Perez-Castrillo, and D. Wettstein, 2010. Egalitarian Equivalence under Asymmetric Information. *Games and Economic Behavior*,

[16] Σ^* leads to the mechanism with $x(L) = 1$, $m(L) = (76/3, -76/3)$, $x(H) = 8/33$, and $m(H) = 42x(H)$. If individual 1 is a dictator, then the resulting equilibrium outcome corresponds to the mechanism with $x(L) = 1$, $m(L) = (92/3, -92/3)$, $x(H) = 4/9$, and $m(H) = 44x(H)$. If individual 2 is a dictator, then the resulting equilibrium outcome corresponds to the mechanism with $x(L) = 1$, $m(L) = (20, -20)$, $x(H) = 0$, and $m(H) = 0$.

forthcoming.
Diamond, P. A., 1967. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility: Comment. *Journal of Political Economy* **75**, 765-766.
Epstein, L. G., and U. Segal, 1992. Quadratic social welfare functions. *Journal of Political Economy* **100**, 691-712.
Fleurbaye, M., 2010. Assessing Risky Social Situations. *Journal of Political Economy* **118**, 649-680.
Guo, M., and V. Conitzer, 2009. Worst-Case Optimal Redistribution of VCG Payments in Multi-Unit Auctions. *Games and Economic Behavior* **67**, 69-98.
Grant, S., A. Kajii, B. Polak, and Z. Safra, 2010. Generalized Utilitarianism and Harsanyi's Impartial Observer Theorem. *Econometrica* **78**, 1939-1971.
Harsanyi, J. C., 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* **63**, 309-321.
Harsanyi, J. C., 1963. A Simplified Bargaining Model for the n-Person Cooperative Game. *International Economic Review* **4**, 194-220.
Harsanyi, J. C., 1967-68. Games with Incomplete Information Played by Bayesian Players. *Management Science* **14**, 159-182, 320-334, 486-502.
Harsanyi, J. C., and R. Selten, 1972. A Generalized Nash Solution for Two-Person Bargaining Games with Incomplete Information. *Management Science* **18**, 80-106.
Holmström, B., and R. B. Myerson, 1983. Efficient and Durable Decision Rules with Incomplete Information. *Econometrica* **51**, 1799-1819.
Kalai, E., 1977. Proportional Solutions to Bargaining Situations: Interpersonal Utility Comparisons. *Econometrica* **45**, 1623-1630.
Kalai, A. T., and E. Kalai, 2010. A Cooperative Value for Bayesian Games. Discussion Paper, Northwestern University.
Kalai, E., and D. Samet, 1985. Monotonic Solutions to General Cooperative Games. *Econometrica* **53**, 307-327.
Kalai, E., and M. Smorodinsky, 1975. Other Solutions to Nash's Bargaining Problem. *Econometrica* **43**, 513-518.
Luce, R.D., Raiffa, H., 1957. Games and Decisions. Wiley, New York.
Maskin, E. and J. Tirole, 1992. The Principal-Agent Relationship with an Informed Principal. II: Common Values. *Econometrica* **60**, 1-42.
Moulin, H., 1988. Axioms of Cooperative Decision Making. Cambridge University Press.
Moulin, H., 2009. Almost Budget-Balanced VCG Mechanisms to Assign Multiple Objects. *Journal of Economic Theory* **144**, 96-119.

Moulin, H., 1992. An application of the Shapley value to fair division with money. *Econometrica* **60**, 1331-1349.

Moulin, H., and W. Thomson, 1988. Can Everyone Benefit from Growth?: two Difficulties. *Journal of Mathematical Economics* **17**, 339-345.

Myerson, R. B., 1979. Incentive Compatibility and the Bargaining Problem. *Econometrica* **47**, 62-73.

Myerson, R. B., 1983. Mechanism Design by an Informed Principal. *Econometrica* **51**, 1767-1797.

Myerson, R. B., 1984a. Two-Person Bargaining Problems with Incomplete Information. *Econometrica* **52**, 461-487.

Myerson, R. B., 1984b. Cooperative Games with Incomplete Information. *International Journal of Game Theory* **13**, 69-96.

Myerson, R.B., 1991. Game Theory: Analysis of Conflict. Harvard University Press, Cambridge, MA.

Myerson, R. B. and M. Satterthwaite, 1983. Efficient Mechanisms for Bilateral Trading. *Journal of Economic Theory* **23**, 265-281.

Nehring, K., 2004. The Veil of Public Ignorance. *Journal of Economic Theory* **119**, 247-270.

Owen, G., 1968. Game Theory. W.B. Saunders, Philadelphia, PA.

Pazner, E., and D. Schmeidler, 1978. Egalitarian Equivalent Allocations: a New Concept of Economic Equity. *Quarterly Journal of Economics* **92**, 671-687.

Ray, D., and K. Ueda, 1996. Egalitarianism and Incentives. *Journal of Economic Theory* **71**, 324-348.

Shapley, L. S., 1969. Utility Comparisons and the Theory of Games. In: Guilbaud, G. Th. (Ed.), La Decision, CNRS, 251-263.

Thomson, W., and R. B. Myerson. Monotonicity and Independence Axioms. *International Journal of Game Theory* **9**, 37-49.

Weidner, F., 1992. The Generalized Nash bargaining Solution and Incentive Compatible Mechanisms. *International Journal of Game Theory* **21**, 109-129.

Wilson, R., 1978. Information, Efficiency, and the Core of an Economy. *Econometrica* **46**, 807816.

Yaari, M. E., 1981. Rawls, Edgeworth, Shapley, Nash: Theories of Distributive Justice Re-Examined. *Journal of Economic Theory* **24**, 1-39.

Young, P. H., 1985a. Producer Incentives in Cost Allocation. *Econometrica* **53**, 757-765.

Young, P. H., 1985b. Monotonic Solutions of Cooperative Games. *International Journal of Game Theory* **14**, 65-72.

Appendix: Proof of Theorem 2

(An informal roadmap of the argument is available in the second supplemental appendix.) Consider a solution Σ that satisfies I-EFF, R-MON, AN, I-WELF, IIA, EX, and IIA, and a problem $\mathcal{S} = (I, D, d^*, (T_i)_{i\in I}, p, (u_i)_{i\in I})$ that is simple. We have to prove that $\Sigma(\mathcal{S})$ coincides with the set of mechanisms meeting the interim egalitarian criterion. We first assume that $\#T_i = \#T_j$, for all i, j.

Let μ be a mechanism that passes the interim egalitarian criterion, and for which there exists $\lambda \in \times_{i\in I}\mathbb{R}^{T_i}_{++}$ such that $\mu \in \arg\max_{\nu\in\mathcal{F}(\mathcal{S})} W^\lambda(\mu) = \sum_{i\in I}\sum_{t_i\in T_i}\lambda_i(t_i)U_i(\nu|t_i)$. All the components of $\mathfrak{u}(\mu)$ are identical, by definition of the interim egalitarian criterion. Let ξ be this common component. The proof requires the consideration of other related mechanism design problems. The first alternative problem is obtained by changing both the common prior and the utility functions so as to keep the set of interim utilities that are achievable via feasible mechanisms unchanged: $S^1 = (I, D, d^*, (T_i)_{i\in I}, p^1, (u^1_i)_{i\in I})$, with p^1 being the uniform probability distribution on T, and $u^1_i(d,t) := \#(T_{-i})p(t_{-i}|t_i)u_i(d,t)$, for all (d,t) and all i.

Notice that this first modification does not change the individuals' interim evaluations of any mechanism since the products of the conditional probabilities with the state-contingent utilities remain constant (see Myerson, 1984a, Section 3). Hence $\mathcal{U}(\mathcal{F}(\mathcal{S}^1)) = \mathcal{U}(\mathcal{F}(\mathcal{S}))$, and λ is also orthogonal to $\mathcal{U}(\mathcal{F}(\mathcal{S}^1))$ at $\mathfrak{u}^1(\mu)$. Following Myerson's virtual utility construction (see also Lemma 4 in the first supplemental Appendix), it is possible to construct an auxiliary problem $\mathcal{S}^2 = (I, D^2, (T_i)_{i\in I}, p^1, (u^2_i)_{i\in I})$, where $D^2 = D \cup \{d_{i,t_i}|t_i \in T_i, i \in I\}$, $u^2_i(\cdot,t) = u^1_i(\cdot,t)$ on D, for all t and all i, and such that $\mathcal{U}(\mathcal{F}(\mathcal{S}^2))$ is the convex hull of the vectors 0 and $\mathfrak{u}^{i,t_i}$, for each $t_i \in T_i$ and each $i \in I$, where $\mathfrak{u}^{i,t_i}_j(t_j) = 0$, for all $(j,t_j) \neq (i,t_i)$, and $\mathfrak{u}^{i,t_i}_i(t_i) = W^\lambda(\mu)/\lambda_i(t_i)$. Notice that μ remains incentive efficient in $\mathcal{S}^2$.

Lemma 5 in the first supplemental Appendix shows that it is possible to define utility functions u^3 and, for each combination (i,t_i), a collective decision $\hat{d}_{i,t_i}$ such that $U^3_i(\hat{d}_{i,t_i}|t_i) = W^\lambda/\lambda_i(t_i)$, $U^3_i(\hat{d}_{i,t_i}|t'_i) = 0$ if $t'_i \neq t_i$, $U^3_j(\hat{d}_{i,t_i}|t_j) = 0$, for all $j \in N \setminus \{i\}$ and all $t_j \in T_j$, and

$$\sum_{j\in I}\frac{\lambda_j(t'_j)}{p^1(t'_j)}u^3_j(\hat{d}_{i,t_i},t')\leq \xi\sum_{j\in I}\frac{\lambda_j(t'_j)}{p^1(t'_j)},\text{ for all } t'\in T. \tag{11}$$

Let $D^3=\{d^*\}\cup\{\hat{d}_{i,t_i}|i\in I,t_i\in T_i\}$, and $S^3=(I,D^3,d^*,(T_i)_{i\in I},p^1,(u^3_i)_{i\in I})$. Clearly, $\mathcal{U}(\mathcal{F}(\mathcal{S}^3))=\mathcal{U}(\mathcal{F}(\mathcal{S}^2))$.

Let
$$0<\epsilon<\min_{(i,t)}\frac{\lambda_i(t_i)}{p^1(t_i)\sum_{j\in I}\frac{\lambda_j(t_j)}{p^1(t_j)}},$$
and $\mathcal{C}=\{x\in\mathbb{R}^I_+|(\forall i\in I):x_i+\epsilon\sum_{j\in I\setminus\{i\}}x_j\leq\xi(1+\epsilon(\#I-1))\}$. Notice that $\mathcal{C}$ is included in the half-space $\{x\in\mathbb{R}^I|\sum_{i\in I}\frac{\lambda_i(t_i)}{p^1(t_i)}x_i\leq\xi\sum_{i\in I}\frac{\lambda_i(t_i)}{p^1(t_i)}\}$, for all $t\in T$. Indeed, given t, let $x\in\mathcal{C}$ and let $J=\{i\in I|x_i>\xi\}$. If $J=\emptyset$, then $\sum_{i\in I}\frac{\lambda_i(t_i)}{p^1(t_i)}x_i$ is clearly lower or equal to $\xi\sum_{i\in I}\frac{\lambda_i(t_i)}{p^1(t_i)}$. Suppose then that J is nonempty, and let i be an element of J. The i-inequality in the definition of $\mathcal{C}$ implies that $x_i-\xi\leq\epsilon\sum_{j\in I\setminus J}(\xi-x_j)$. Multiplying by $\lambda_i(t_i)/p(t_i)$, and taking the sum over $i\in J$, one gets: $\sum_{i\in J}\frac{\lambda_i(t_i)}{p^1(t_i)}(x_i-\xi)\leq\sum_{j\in I\setminus J}\epsilon(\sum_{i\in J}\frac{\lambda_i(t_i)}{p^1(t_i)})(\xi-x_j)\leq\sum_{j\in I\setminus J}\epsilon(\sum_{i\in I}\frac{\lambda_i(t_i)}{p^1(t_i)})(\xi-x_j)\leq\sum_{j\in I\setminus J}\frac{\lambda_j(t_j)}{p^1(t_j)}(\xi-x_j)$, and hence $\sum_{i\in I}\frac{\lambda_i(t_i)}{p^1(t_i)}x_i\leq\xi\sum_{i\in I}\frac{\lambda_i(t_i)}{p^1(t_i)}$, as desired.

Let $\mathcal{S}^4=(I,D^4,d^*,(T_i)_{i\in I},p^1,(u^4_i)_{i\in I})$ be the mechanism design problem with $D^4=\mathcal{C}^T$ and $u^4_i(d^4,t)=d^4_i(t)$, for all $d^4\in D^4$, all $i\in I$, and all $t\in T$. Observe that the problem $\mathcal{S}^4$ is symmetric, and hence AN and (1) imply that the interim utility of any mechanism in the solution of that problem must give equal interim utility to all the individuals and whatever their private information. I-EFF and EX imply that the constant mechanism that selects $(\xi,\ldots,\xi)$ in $\mathcal{C}$, for all $t\in T$, belongs to $\Sigma(\mathcal{S}^4)$. It is easy to check that interim utilities are transferable at that constant mechanism. The combination of R-MON and EX imply that it also belongs to $\Sigma(\mathcal{S}^5)$, where $\mathcal{S}^5=(I,D^5,d^*,(T_i)_{i\in I},p^1,(u^5_i)_{i\in I})$ is the mechanism design problem with $D^5=D^3\cup D^4$, $u^5_i(d^5,t)=u^3_i(d^5,t)$ if $d^5\in D^3$ and $u^5_i(d^5,t)=u^4_i(d^5,t)$ if $d^5\in D^4$, for all $d^5\in D^5$, all $i\in I$, and all $t\in T$. Indeed, the way ϵ was chosen guarantees that the constant mechanism is incentive efficient in $\mathcal{S}^5$ (cf. (11)). EX implies that any feasible mechanism that gives the same vector of interim utilities also belongs to $\Sigma(\mathcal{S}^5)$. Let's choose one that can be expressed via lotteries on D^3, which is possible since the constant mechanisms that pick one of the decisions in D^3 independently of the individuals' reports generate all the extreme points of $\mathcal{U}^{eff}(\mathcal{F}(\mathcal{S}^5))$. R-MON and EX imply that it must remain a solution to $\mathcal{S}^3$. I-WELF imply that any mechanism in the solution of $\mathcal{S}^2$ must have the same interim utilities, and μ must thus belong to $\Sigma(\mathcal{S}^2)$, by EX. IIA implies that $\mu \in \Sigma(\mathcal{S}^1)$. Condition (1) implies that $\Sigma(\mathcal{S}^1)$ coincides with the set of mechanisms that meet the interim egalitarian criterion in $\mathcal{S}^1$. I-WELF and EX implies that $\Sigma(\mathcal{S})$ selects all mechanisms meeting the interim egalitarian criterion, as desired.

We now conclude the proof by dropping the assumption that $\#T_i = \#T_j$, for all i, j. As before, let μ be a mechanism that passes the interim egalitarian criterion. All the components of $\mathfrak{u}(\mu)$ are identical, by definition of the interim egalitarian criterion. Let ξ be this common component.

We now derive from $\mathcal{S}$ an alternative mechanism design problem where all the individuals have the same number of possible types. Let $\tau = \max_{i \in I} \#T_i$, let $h_i : T_i \to \{1, \ldots, \tau\}$ be an injective function, for each $i \in I$, let $\bar{t} \in T$, let $g_i : \{1, \ldots, \tau\} \to T_i$ be defined as follows: $g_i(t_i') = h_i^{-1}(t_i')$, if $t_i' \in Im(h_i)$, and $= \bar{t}_i$, otherwise, for each $t_i' \in \{1, \ldots, \tau\}$ and each $i \in I$. Let $\mathcal{S}' = (I, D, d^*, (T_i')_{i \in I}, p', (u_i')_{i \in I})$, where $T_i' = \{1, \ldots, \tau\}$, for each $i \in I$, $p'(t') = p(g(t'))/\Pi_{\{i \in I | g_i(t_i') = \bar{t}_i\}}(\tau + 1 - \#T_i)$, and $u_i'(d, t') = u_i(d, g(t'))$, for each $i \in I$, where $g(t') = (g_i(t_i'))_{i \in I}$. So $\mathcal{S}'$ differs from $\mathcal{S}$ only in that, for each $i \in I$, type $\bar{t}_i$ has been splitted sufficiently many times so that i has τ possible types.

The splitted version of μ (iteration of the definition in IST – see also footnote 10) meets the interim egalitarian criterion for $\mathcal{S}'$, and the associated vector of interim utilities is constant, with ξ being the common component. It is then easy to check that $\mathcal{S}'$ is simple, and hence the first part of the proof applies: any mechanism in $\Sigma(\mathcal{S}')$ generates a constant vector of interim utilities, with ξ being the common component.

Let $\nu \in \Sigma(\mathcal{S})$. IST implies that the splitted version of ν belongs to $\Sigma(\mathcal{S}')$. Condition (1) implies that the associated vector of interim utilities is constant, with each component being equal to ξ. Hence $U_i(\nu | t_i) = \xi = U_i(\mu | t_i)$, for all $t_i \in T_i$ and all $i \in I$. EX implies that $\mu \in \Sigma(\mathcal{S})$. EX and (1) implies that $\Sigma(\mathcal{S})$ coincides with the set of mechanisms that meet the interim egalitarian criterion, as desired. ■

First Supplemental Appendix (not for publication)

Results in this section are variants or reformulations of previous results established by Myerson (1983, 1984a, 1984b, 1991).

Lemma 1 *Let* $(T_i)_{i\in I}$ *be finite sets of types, let* p *be a probability distribution on* $T = \times_{i\in I} T_i$ *with full support, let* $v : T \to \mathbb{R}$*, and let* $\mathfrak{u} \in \times_{i\in I}\mathbb{R}^{T_i}$*. Then there exists* $x : T \to \mathbb{R}^I$ *such that*

1. $(\forall t \in T) : \sum_{i\in I} x_i(t) = v(t)$*, and*

2. $(\forall i \in I)(\forall t_i \in T_i) : \mathfrak{u}_i(t_i) = \sum_{t_{-i}} p(t_{-i}|t_i)x_i(t)$*,*

if and only if

$$\sum_{i\in I}\sum_{t_i\in T_i} p(t_i)\mathfrak{u}_i(t_i) = \sum_{t\in T} p(t)v(t).$$

<u>Proof</u>: If the two conditions are true, then

$$\begin{aligned}\textstyle\sum_{i\in I}\sum_{t_i\in T_i} p(t_i)\mathfrak{u}_i(t_i) &= \textstyle\sum_{i\in I}\sum_{t_i\in T_i} p(t_i)\sum_{t_{-i}} p(t_{-i}|t_i)x_i(t)\\ &= \textstyle\sum_{t\in T} p(t)\sum_{i\in I} x_i(t)\\ &= \textstyle\sum_{t\in T} p(t)v(t),\end{aligned}$$

where the first equality follows from *2*, the second equality follows from rearranging the terms, and the third equality follows *1*.

I now prove the converse, assuming for simplicity that $\sum_{t\in T} p(t)v(t) \neq 0$ (a straightforward translation argument implies that the result also holds when $\sum_{t\in T} p(t)v(t) = 0$). To do this, I will show that for any $(i, \bar{t}_i)$ there exists an $x : T \to \mathbb{R}^I$ that satisfies *1* and *2* for $\mathfrak{u}^{i.\bar{t}_i}$, where $\mathfrak{u}_i^{i.\bar{t}_i}(t_i) = \sum_{t\in T} p(t)v(t)/p(\bar{t}_i)$ and $\mathfrak{u}_j^{i.\bar{t}_i}(t_j) = 0$, for all $(j, t_j) \neq (i, \bar{t}_i)$. The result will indeed follow since any vector $\mathfrak{u}$ such that $\sum_{i\in I}\sum_{t_i\in T_i} p(t_i)\mathfrak{u}_i(t_i) = \sum_{t\in T} p(t)v(t)$ can be written as an affine combination of these vectors, and the equations in *1* and *2* are linear in x. So I now define a function $x : T \to \mathbb{R}^I$, and show that it satisfies the two sets of equations for $\mathfrak{u}^{i.\bar{t}_i}$:

$$x_i(t) = 0 \text{ and } x_j(t) = \frac{v(t)}{\#I - 1},$$

for all $t \in T$ such that $t_i \neq \bar{t}_i$, and

$$x_i(\bar{t}_i, t_{-i}) = v(\bar{t}_i, t_{-i}) - \sum_{j \in N \setminus \{i\}} x_j(\bar{t}_i, t_{-i})$$

$$x_j(\bar{t}_i, t_{-i}) = -\frac{1}{\#I - 1} \sum_{\tilde{t}_{-j} | \tilde{t}_i \neq \bar{t}_i} \frac{p(\tilde{t}_{-j}, t_j)}{p(\bar{t}_i, t_j)} v(t_j, \tilde{t}_{-j}).$$

for all $t_{-i} \in T_{-i}$. Equations in *1* and the equations in *2* for any (i, t_i) with $t_i \neq \bar{t}_i$ are trivially satisfied, by construction. I now check the equations in *2* for any (j, t_j) with $j \neq i$. For each $j \in I \setminus \{i\}$ and each $t_j \in T_j$, we have:

$$\sum_{t_{-j}} p(t_{-j}|t_j) x_j(t) = \sum_{t_{-ij}} p(t_{-ij}, \bar{t}_i | t_j) x_j(\bar{t}_i, t_{-i}) + \sum_{t_{-j} | t_i \neq \bar{t}_i} p(t_{-j}|t_j) x_j(t)$$

$$= -\frac{p(\bar{t}_i|t_j)}{\#I - 1} \sum_{t_{-j} | t_i \neq \bar{t}_i} \frac{p(t_{-j}, t_j)}{p(\bar{t}_i, t_j)} v(t) + \frac{1}{\#I - 1} \sum_{t_{-j} | t_i \neq \bar{t}_i} p(t_{-j}|t_j) v(t),$$

where the second equality follows from the definition of $\bar{x}$, and in particular the fact that $x_j(\bar{t}_i, t_{-i})$ does not depend on t_{-ij}. The last expression equals zero, as desired, because $p(\bar{t}_i|t_j) \frac{p(t_{-j}, t_j)}{p(\bar{t}_i, t_j)} = p(t_{-j}|t_j)$. Finally, since x satisfies the equations in *1*, we have that $\sum_{i \in I} \sum_{t_i \in T_i} p(t_i) \sum_{t_{-i}} p(t_{-i}|t_i) x_i(t) = \sum_{t \in T} p(t) v(t)$. This combined with what we just proved implies that $\sum_{t_{-i}} p(t_{-i}|\bar{t}_i) x_i(t) = \sum_{t \in T} p(t) v(t) / p(\bar{t}_i)$, as desired. ■

Lemma 2 *A feasible mechanism μ is incentive efficient if and only if there exist $\lambda \in \times_{i \in I} \mathbb{R}_+^{T_i} \setminus \{0\}$ and $\alpha \in \times_{i \in I} \mathbb{R}_+^{T_i \times T_i}$ such that*

1. $(\forall i \in I)(\forall (t_i, t_i') \in T_i \times T_i) : \alpha_i(t_i'|t_i)(U_i(\mu|t_i) - U_i(\mu, t_i'|t_i)) = 0$

2. $\sum_{i \in I} \sum_{t_i \in T_i} \lambda_i(t_i) U_i(\mu|t_i) = \sum_{t \in T} p(t) \max_{d \in D} \sum_{i \in I} v_i^{(\lambda, \alpha)}(d, t)$, *where the "virtual utility" functions $v_i^{(\lambda, \alpha)}$ are defined as follows:*

$$v_i^{(\lambda, \alpha)}(d, t) = \frac{1}{p(t_i)} [(\lambda_i(t_i) + \sum_{t_i'} \alpha_i(t_i'|t_i)) u_i(d, t) - \sum_{t_i'} \alpha_i(t_i|t_i') u_i(d, t_{-i}, t_i')],$$

for each $d \in D$, each $t \in T$, and each $i \in I$.

Proof: The vector λ is derived from a classical separation argument, using the fact that $\mathcal{U}(\mathcal{F}(\mathcal{S}))$ is closed and convex. The vector α specifies the dual variables of maximization of the weighted sum under the incentive constraints. Condition 1 is the usual condition stating that the dual variable associated to the constaint that individual i should not pretend to be of type t_i' when being actually of type t_i, is positive only if that constraint is binding. Condition 2 is obtained by rearranging the terms of the Lagrangean (more details available in Myerson (1991, Chapter 10), for instance). ■

Lemma 3 *Let $A = (a_{ij})_{1 \leq i,j \leq n}$ be a square matrix with non-positive elements off the diagonal (i.e. $a_{ij} \leq 0$ if $i \neq j$) and such that the sum of the elements in each column is strictly positive (i.e. $\sum_{i=1}^{n} a_{ij} > 0$, for each j). Then A is invertible.*

Proof: Let $x \in \mathbb{R}^n$ be such that $Ax = 0$. I will prove that $x \geq 0$. Suppose, on the contrary, that $J = \{j \in I | x_j < 0\} \neq \emptyset$. Then

$$\sum_{j \in J} \sum_{k=1}^{n} a_{jk} x_k = 0,$$

which is equivalent to

$$\sum_{j \in J} \sum_{k \in J} a_{jk} x_k + \sum_{j \in J} \sum_{k \in I \setminus J} a_{jk} x_k = 0.$$

Notice that $\sum_{j \in J} \sum_{k \in J} a_{jk} x_k = \sum_{k \in J} (\sum_{j \in J} a_{jk}) x_k < 0$, because $\sum_{j \in J} a_{jk} > 0$, for each $k \in J$, given the assumptions on A. All the coefficients of the second term fall off the diagonal of A and are thus negative, while the corresponding components of x are non-negative since $k \notin J$. Hence the second term is non-positive, reaching a contradiction. This shows that $x \geq 0$. A similar reasoning implies that $x \leq 0$, and x must actually be equal to zero. Hence A is invertible. ■

Lemma 4 *Let $\mathcal{S} = (I, D, (T_i)_{i \in I}, p, (u_i)_{i \in I})$ be a mechanism design problem, let μ be an incentive efficient mechanism, let $(\lambda, \alpha) \in (\times_{i \in I} \mathbb{R}_{++}^{T_i}) \times (\times_{i \in I} \mathbb{R}_{+}^{T_i \times T_i})$*

be a pair of vectors that satisfy the conditions of Lemma 2, let $j \in I$, let $\bar{t}_j \in T_j$, and let $W^\lambda = \sum_{i \in I} \sum_{t_i \in T_i} \lambda_i(t_i) U_i(\mu|t_i)$. Then it is possible to construct a decision $\hat{d}$ and utility functions $(\hat{u}_i)_{i \in I}$ defined on $(D \cup \{\hat{d}\}) \times T$ such that

1. *$\hat{u}_i(d,t) = u_i(d,t)$, for each $d \in D$, each $i \in I$, and each $t \in T$;*
2. *μ satisfies the conditions of Lemma 2 for (λ, α) in $\hat{\mathcal{S}} = (I, D \cup \{\hat{d}\}, (T_i)_{i \in I}, p, (\hat{u}_i)_{i \in I})$;*
3. *$\hat{U}_j(\hat{d}|\bar{t}_j) = W^\lambda / \lambda_j(\bar{t}_j)$;*
4. *$\hat{U}_j(\hat{d}|t'_j) = 0$ if $t'_j \in T_j \setminus \{t_j\}$;*
5. *$\hat{U}_i(\hat{d}|t_i) = 0$, for all $i \in N \setminus \{j\}$ and all $t_i \in T_i$.*

Proof: Let $f : \mathbb{R}^{I \times T} \to \mathbb{R}^{I \times T}$ be the linear transformation that maps any profile of ex-post utilities to its associated virtual utilities:

$$(f(u))_i(t) = \frac{1}{p(t_i)}[(\lambda_i(t_i) + \sum_{t'_i} \alpha_i(t'_i|t_i)) u_i(t) - \sum_{t'_i} \alpha_i(t_i|t'_i) u_i(t_{-i}, t'_i)],$$

for each $u \in \mathbb{R}^{I \times T}$, each $t \in T$, and each $i \in I$. Let $\hat{f} : \times_{i \in I} \mathbb{R}^{T_i}$ be an analogue transformation for interim utilities:

$$(f(\mathfrak{u}))_i(t_i) = \frac{1}{p(t_i)}[(\lambda_i(t_i) + \sum_{t'_i} \alpha_i(t'_i|t_i)) \mathfrak{u}_i(t_i) - \sum_{t'_i} \alpha_i(t_i|t'_i) \mathfrak{u}_i(t'_i)],$$

for each $\mathfrak{u} \in \times_{i \in I} \mathbb{R}^{T_i}$, each $t_i \in T_i$, and each $i \in I$. Lemma 3 implies that both f and $\hat{f}$ are invertible.

Let

$$v(t) = \max_{d \in D} \sum_{i \in I} v_i^{(\lambda,\alpha)}(d,t),$$

for each $t \in T$. Consider then the interim profile of virtual utilities $\mathfrak{u}$ defined as follows:

$$\mathfrak{u}_j(\bar{t}_j) = \frac{W^\lambda}{\lambda_j(\bar{t}_j)}$$

$$(\forall (i, t_i) \neq (j, \bar{t}_j)) : \mathfrak{u}_i(t_i) = 0,$$

and let $\mathfrak{v}$ be the associated profile of interim utilities, i.e. $\mathfrak{v} = \hat{f}(\mathfrak{u})$. Notice that

$$\sum_{i\in I}\sum_{t_i\in T_i} p(t_i)\mathfrak{v}_i(t_i) = \sum_{t_j\in T_j} p(t_j)(\hat{f}(\mathfrak{u}))_j(t_j) = W^\lambda = \sum_{t\in T} p(t)v(t).$$

Lemma 1 implies that there exists $x : T \to \mathbb{R}^I$ such that

1. $(\forall t \in T) : \sum_{i\in I} x_i(t) = v(t)$, and
2. $(\forall i \in I)(\forall t_i \in T_i) : \mathfrak{v}_i(t_i) = \sum_{t_{-i}} p(t_{-i}|t_i)x_i(t)$.

The conditions of the present lemma are then satisfied if one defines $\hat{d}$ and the utility functions such that $\hat{u}_i(\hat{d}, t) = (f^{-1}(x))_i(t)$, for all $i \in I$ and $t \in T$. Condition 1 is satisfied by definition, while condition 2 follows from Lemma 3 and the fact that $\sum_{i\in I} x_i(t) = v(t)$, $\forall t \in T$. The three remaining conditions follow from simple computations:

$$\sum_{t_{-i}\in T_{-i}} p(t_{-i}|t_i)u_i(\hat{d}, t) = \sum_{t_{-i}\in T_{-i}} p(t_{-i}|t_i)(f^{-1}(x))_i(t) = (\hat{f}^{-1}(\mathfrak{v}))_i(t_i) = \mathfrak{u}_i(t_i),$$

for each $t_i \in T_i$ and each $i \in I$. ■

Lemma 5 *Let I be a finite set of individuals, let $(T_i)_{i\in I}$ be a collection of type sets, let $p \in \Delta(T)$ be a common prior with full support, let $\lambda \in \times_{i\in I}\mathbb{R}^{T_i}_{++}$, and $\mathfrak{w} \in \times_{i\in I}\mathbb{R}^{T_i}_{++}$. Then there exists a collective decision d_{i,t_i}, for each combination (i, t_i), and a utility function $u_i : \{d_{j,t_j}|j \in I, t_j \in T_j\} \times T \to \mathbb{R}$, for each $i \in I$, such that the following conditions hold true for each (i, t_i):*

1. $U_i(d_{i,t_i}|t_i) = [\sum_{j\in I}\sum_{t_j\in T_j} \lambda_j(t_j)\mathfrak{w}_j(t_j)]/\lambda_i(t_i)$;
2. $U_i(d_{i,t_i}|t'_i) = 0$, *for all* $t'_i \in T_i \setminus \{t_i\}$;
3. $U_j(d_{i,t_i}|t_j) = 0$, *for all* $j \in N \setminus \{i\}$ *and all* $t_j \in T_j$;
4. $\sum_{j\in I} \frac{\lambda_j(t'_j)}{p(t'_j)} u_j(d_{i,t_i}, t') \leq \sum_{j\in I} \frac{\lambda_j(t'_j)}{p(t'_j)} \mathfrak{w}_j(t'_j)$, *for all* $t' \in T$

Proof: Let

$$v(t) = \sum_{j \in I} \frac{\lambda_j(t_j)}{p(t_j)} \mathfrak{w}_j(t_j),$$

for each $t \in T$. Notice that

$$\sum_{t \in T} p(t)v(t) = \sum_{t \in T} \sum_{j \in I} p(t_{-j}|t_j)\lambda_j(t_j)\mathfrak{w}_j(t_j) = \sum_{j \in I} \sum_{t_j \in T_j} \lambda_j(t_j)\mathfrak{w}_j(t_j).$$

Fix $i \in I$ and $\bar{t}_i \in T_i$. The previous computation implies that the conditions of Lemma 1 are satisfied for $\mathfrak{u} \in \times_{i \in I} \mathbb{R}^{T_i}$ defined as follows:

$$\mathfrak{u}_i(\bar{t}_i) = [\sum_{j \in I} \sum_{t_j \in T_j} \lambda_j(t_j)\mathfrak{w}_j(t_j)]/p(\bar{t}_i)$$

$$(\forall (j, t_j) \neq (i, \bar{t}_i)) : \mathfrak{u}_j(t_j) = 0.$$

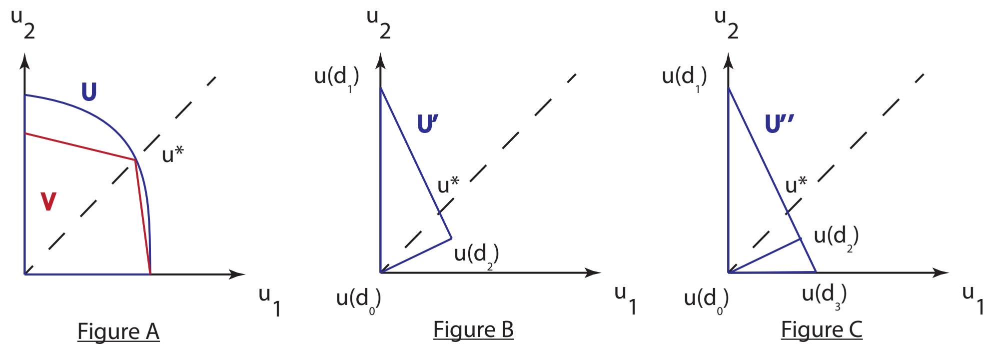
Let $x : T \to \mathbb{R}^I$ be a function that satisfies the two conditions of Lemma 1. It is then easy to check that the four conditions of the present lemma are satisfied if one defines $d_{i.\bar{t}_i}$ and the utility functions so that $u_j(d_{i.\bar{t}_i}, t) = p(t_j)x_j(t)/\lambda_j(t_j)$, for all $t_j \in T_j$ and all $j \in I$. ■

Second Supplemental Appendix (not for publication)

The purpose of this second supplemental appendix is to provide a non-technical and informal roadmap for the proof of Theorem 2 (found in the main Appendix of the paper), and to illustrate some of the difficulties one encounters when working under asymmetric information, making arguments significantly more intricate than in the special case of complete information.

Let's start by presenting Kalai's argument to show that any monotone, efficient, and anonymous solution must be egalitarian (under complete information).[17] Remember that, in line with classical models of bargaining and social choice at the time, his argument is phrased in the space of utilities. Let's assume that there are only two individuals so that we can rely on graphical representations. Consider a problem U, as depicted on Figure A. The egalitarian solution u^* falls on the 45-degree line. Consider then the symmetric problem V that is included in U and contains u^*. Given that the solution is anonymous, it must pick u^* as the solution for V. Monotonicity then implies that u^* must be the solution to U, as well, and we are done. This argument was possible because of the possibility of constructing a symmetric problem V that is included in U and that contains u^*. This is always possible given Kalai's assumption that problems are comprehensive. The argument to prove Theorem 2 in the Appendix aims at following a similar line of proof while working in the space of interim utilities. Yet here is a first difficulty: incentive constraints may make problems non-comprehensive even when utilities are fully transferable as in quasi-linear mechanism design problems. So, let's revisit Kalai's argument by considering another problem U' of complete information that is non-comprehensive, see Figure B. To make a better connection with the proof of Theorem 2, let's also consider the basic collective decisions underlying the problem. Remember indeed that one must consider these underlying collective decisions in order to write incentive constraints. So, suppose more precisely, that there are three collective decisions available, d_0, d_1, and d_2, and that util-

[17] To be more precise, Kalai (1977) offers a characterization of proportional solutions with an axiom of Homogeneity instead of Anonymity.



ity functions are such that the convex hull of the image of these three decisions in the utility space deliver U'. Notice that Kalai's original argument does not apply in this case, as there is no symmetric subset of U' that contains v^* in the relative interior of its Pareto frontier (in order to apply R-MON, given that MON is incompatible with Pareto efficiency on the expanded domain). One way to determine the solution to U' is to consider an expanded problem U'' (as in Figure C) by adding a collective decision d_3 whose image in the utility space expands U' into a right triangle. Kalai's argument applies to U'', implying that the solution to that problem is v^*. Applying IIA, one concludes that the solution to U' is v^*, as desired. Yet there is a substantial problem when trying to replicate that argument under incomplete information: constructing U'' from U' (in the space of interim utilities) is far from trivial. First notice that requiring I-WELF combined with R-MON does not imply that R-MON can be applied directly in the space of utilities. Indeed, one must keep the original set of collective decisions available when expanding a problem. This has no real consequence under complete information, but may have surprising effect on the set of achievable interim utilities under incomplete information, for the following two reasons:

1. Mechanism designers can build mechanisms that wisely combine additional decisions with existing ones in a way that provide mutually beneficial insurance to individuals. In other words, the impact of adding collective decisions (or equivalently utility vectors ex-post) may have complex implications at the interim stages because of expected utility, especially when participants hold different beliefs due to asymmetric in-

formation.

2. Mechanism designers can build mechanisms that wisely use additional decisions to better manage incentive constraints. Adding simple new collective decisions often dramatically moves the incentive efficiency frontier in a way that makes the application of IIA in a subsequent stage impossible.

This is where Myerson's virtual utility construct becomes most helpful. Indeed, his technique identify, for any incentive efficient mechanism, additional collective decisions that effectively linearize the set of interim utilities that are individually rational when added to the problem, while leaving the original mechanism incentive efficient in the enlarged problem. This explains the consideration of $\mathcal{S}^2$ in the proof of Theorem 2. Even so, the problem $\mathcal{S}^2$ may be highly asymmetric, and not contain any symmetric sub-problem. This is where I-WELF becomes handy, as it allows to construct $\mathcal{S}^3$, which is still typically asymmetric, but can be solved by considering an alternative fully symmetric problem where incentive constrainst are effectively non restrictive, cf. $\mathcal{S}^4$. The consideration of $\mathcal{S}^5$, whose set of collective decisions is the union of those available in either $\mathcal{S}^3$ or $\mathcal{S}^4$. This reasoning worked because $\mathcal{S}^4$ was fully symmetric, which required the consideration of a problem where all the individuals hold the same number of possible types and beliefs are derived by Bayesian updating from a uniform common prior. The second half of the proof of Theorem 2 (see page 40) shows how one can deal with problems that involve type sets of different cardinality by applying IST. In addition, I-WELF (or even the weaker property of probability invariance, as assumed by Myerson (1984a)) shows that there is no loss of generality in solving only problems with a uniform common prior when type sets have equal cardinality - see the construction of $\mathcal{S}^1$ at the beginning of the proof. This completes the argument.