ROBUST MECHANISM DESIGN

By

Dirk Bergemann and Stephen Morris

May 2003

COWLES FOUNDATION DISCUSSION PAPER NO. 1421

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Robust Mechanism Design*

Dirk Bergemann[†] Stephen Morris[‡]

First Version: December 2000
This Version: May 2003

Abstract

The mechanism design literature assumes too much common knowledge of the environment among the players and planner. We relax this assumption by studying implementation on richer type spaces, with more higher order uncertainty.

We study the "ex post equivalence" question: when is interim implementation on all possible type spaces equivalent to requiring ex post implementation on the space of payoff types? We show that ex post equivalence holds when the social choice correspondence is a function and in simple quasi-linear environments. When ex post equivalence holds, we identify how large the type space must be to obtain the equivalence. We also show that ex post equivalence fails in general, including in quasi-linear environments with budget balance.

For quasi-linear environments, we provide an exact characterization of when interim implementation is possible in rich type spaces. In this environment, the planner can fully extract players' belief types, so the incentive constraints reduce to conditions distinguishing types with the same beliefs about others' types but different payoff types.

KEYWORDS: Mechanism Design, Common Knowledge, Universal Type Space, Interim Equilibrium, Ex-Post Equilibrium, Dominant Strategies.
JEL CLASSIFICATION: C79, D82

---

*This research is supported by NSF Grant #SES-0095321. We are grateful for comments from seminar participants at many institutions. We thank Sandeep Baliga, Matt Jackson, Jon Levin, Bart Lipman, Eric Maskin, Zvika Neeman, Andrew Postlewaite, Ilya Segal and Tomas Sjöström for valuable discussions.

[†]Department of Economics, Yale University, 28 Hillhouse Avenue, New Haven, CT 06511, dirk.bergemann@yale.edu.

[‡]Department of Economics, Yale University, 30 Hillhouse Avenue, New Haven, CT 06511, stephen.morris@yale.edu.

> "Game theory has a great advantage in explicitly analyzing the consequences of trading rules that presumably are really common knowledge; it is deficient to the extent it assumes other features to be common knowledge, such as one player's probability assessment about another's preferences or information.
>
> I foresee the progress of game theory as depending on successive reductions in the base of common knowledge required to conduct useful analyses of practical problems. Only by repeated weakening of common knowledge assumptions will the theory approximate reality." Wilson (1987)

1. Introduction

The theory of mechanism design helps us understand institutions ranging from simple trading rules to political constitutions. We can understand institutions as the solution to a well defined planner's problem of achieving some objective or maximizing some utility function subject to incentive constraints. But a common criticism of mechanism design theory is that the optimal mechanisms solving the well defined planner's problem seem unreasonably complicated. Researchers have often therefore restricted attention to mechanisms that are "more robust", or less sensitive to the assumed structure of the environment.[1] However, if the optimal solution to the planner's problem is too complicated or sensitive to be used in practice, it is presumably because the original description of the planner's problem was itself flawed. We would like to see if improved modelling of the planner's problem endogenously generates the "robust" features of mechanisms that researchers have been tempted to assume.

As suggested by Robert Wilson in the above quote, the problem is that we make too many implicit common knowledge assumptions in our description of the planner's problem.[2] The modelling strategy must be to first make explicit the implicit common knowledge assumptions, and then weaken them. The approach to modelling incomplete information introduced by Harsanyi (1967/1968) and formalized by Mertens and Zamir (1985) is ideally suited to this task. In fact, Harsanyi's work was intended to address the then prevailing criticism of game theory that the very description of a game embodied common knowledge assumptions that could never prevail in practise. Harsanyi argued

[1] Discussions of this issue are an old theme in the mechanism design literature. Hurwicz (1972) discussed the need for "nonparametric" mechanisms (independent of parameters of the model). Wilson (1985) states that a desirable property of a trading rule is that it "does not rely on features of the agents' common knowledge, such as their probability assessments." Dasgupta and Maskin (2000) "seek auction rules that are independent of the details - such as functional forms or distribution of signals - of any particular application and that work well in a broad range of circumstances".

[2] An important paper of Neeman (2001) shows how rich type spaces can be used to relax implicit common knowledge assumptions in a mechanism design context. For other approaches to formalizing robust mechanism design, see Duggan and Roberts (1997), Eliaz (2002) and Lopomo (1998, 2000).

that by allowing an agent's type to include his beliefs about the strategic environment, his beliefs about other agents' beliefs, and so on, any environment of incomplete information could be captured by a type space. With this sufficiently large type space (including all possible beliefs and higher order beliefs), it is true (tautologically) that there is common knowledge among the agents of each agent's type space and each type's beliefs over the types of other agents. However, as a practical matter, applied economic analysis tends to assume much smaller type spaces than the universal type space, *and yet maintain the assumption that there is common knowledge among the agents of each agent's type spaces and each type's beliefs over the types of other agents.* In the small type space case, this is a very substantive restriction. There has been remarkably little work since Harsanyi checking whether analysis of incomplete information games in economics is robust to the implicit common knowledge assumptions built into small type spaces.[3] We will investigate the importance of these implicit common knowledge assumptions in the context of mechanism design.[4]

Formally, we fix a payoff environment, specifying a set of payoff types for each agent, a set of outcomes, utility functions for each agent and a social choice correspondence (SCC) mapping payoff type profiles into sets of acceptable outcomes. The planner *(partially) implements*[5] the social choice correspondence if there exists a mechanism and an equilibrium strategy profile of that mechanism such that equilibrium outcomes for every payoff type profile are acceptable according to the SCC.[6] While holding fixed this environment, we can construct many type spaces, where an agent's type specifies both his payoff type and his belief about other agents' types. Crucially, there may be many types of an agent with the same payoff type. The larger the type space, the harder it will be to implement the social choice correspondence, and so the more "robust" the resulting mechanism will be. The smallest type space we can work with is the "naive type space," where we set the possible types of each agent equal to the set of payoff types, and assume a common knowledge prior over this type space. This is the usual exercise performed in the mechanism design literature. The largest type space we can work with is the union of all possible type spaces that could have arisen from the payoff environment. This is equivalent to working with a "universal type space," in the sense of Mertens and Zamir (1985). There are many interesting type spaces in between the naive type space and the universal type space that are also interesting to study. For

[3] Battigalli and Siniscalchi (2003), Morris and Shin (2003).

[4] Neeman (2001) argued that small type space assumptions are especially important in the full surplus extraction results of Cremer and McLean (1985).

[5] "Partial implementation" is sometimes called "truthful implementation" or "incentive compatible implementation." Since we look exclusively at partial implementation in this paper, we will write "implement" instead of "partially implement".

[6] In a companion paper, Bergemann and Morris (2002), we use the framework of this paper to look at full implementation, i.e., requiring that every equilibrium delivers an outcome consistent with the social choice correspondence.

example, we can look at the union of all naive type spaces (so that the agents have common knowledge of a prior over payoff types but the mechanism designer does not); and we can look at the union of all type spaces where the common prior assumption holds.

For these different type spaces, we then ask how conditions for interim implementation compare with stronger equilibrium notions on the original type space of payoff types. We show that for social choice functions in general environments and for social choice correspondences in quasi-linear environments without balanced budget constraints there is a strong ex post equivalence result: interim implementation on all naive type spaces is equivalent to interim implementation on all type spaces which is equivalent to ex post implementation. However, these strong equivalences do not hold in general. We show by example that it is sometimes possible to interim implement on all type spaces, but not possible to ex post implement; and it is sometimes possible to interim implement on all *naive* type spaces, but not possible to interim implement on *all* type spaces.

Motivated by this gap, we suggest a weaker notion of ex post implementation, namely augmented ex post implementation, for which equivalence can be show to hold in general. Quasi-linear environments with budget balance constraints are an interesting class of models where the strong equivalence does not hold in general. We illustrate this by means of an example and then provide sufficient conditions for equivalence result. For example, there is an equivalence between ex post implementation and interim implementation on all type spaces if either there are only two agents, or if each agent has at most two payoff types.

Finally, for quasi-linear environments without budget balance constraints, we are able to provide an exact characterization of when interim implementation is possible on arbitrary type spaces. An agent's beliefs about other agents' types can always be fully extracted by standard arguments (Cremer and McLean (1985)), so incentive compatibility conditions reduce to distinguishing types with the same beliefs about others' types but different payoff types. This result parallels an observation of Neeman (2001) for revenue maximizing mechanisms. However, belief types may exhibit linear dependence, so a revenue maximizing seller might have to trade off the gains to extracting belief types against the costs of inducing agents to report their belief types truthfully. However, when the planner is only interested in efficiency (and does not care about transfers), any types with different beliefs can be distinguish at no cost. The interim implementation conditions we describe are automatically satisfied for a "generic" choice of prior on a fixed finite type space. However, we discuss problems with the standard notion of genericity and suggest that our conditions might be hard to satisfy in practise without unreasonable common knowledge assumptions.

In private values environments, ex post implementation is equivalent to dominant strategies implementation. Our positive and negative results all have counterparts in private values environments, and thus our results give sufficient conditions for (and counterexamples to) the equivalence of dominant strategies and Bayesian implementation. There was an early formal and informal debate on the relation between Bayesian and dominant strategies implementation. Consider the case where the prior on a fixed type space is common knowledge among the agents, but is not known to the planner. If we ask that the same direct mechanism Bayesian implement an SCC for every prior on that fixed type space, this is equivalent to dominant strategies implementation (Dasgupta, Hammond and Maskin (1979), Ledyard (1978, 1979) and Groves and Ledyard (1987)). But if the prior were truly common knowledge among the agents, then this particular weakening of the common knowledge assumptions is relatively easy to resolve, as information that is non-exclusive in the sense of Postlewaite and Schmeidler (1986), can always be truthfully elicited from the agents in an interim equilibrium.[7] But if this easily extracted information is used in designing the mechanism, then we lose the argument showing the equivalence of Bayesian and dominant strategies implementation. Our results identify some cases where Bayesian implementation for all priors implies dominant strategies implementation, even when the planner knows (or can easily extract) the true prior on a fixed type space. But in other cases, we show that more interesting relaxations of common knowledge assumptions are required to show the necessity of dominant strategies implementation. And in yet other cases, Bayesian implementation is possible on all type spaces even though dominant strategies implementation is impossible.

The rest of the paper proceeds as follows. Section 2 provides the setup, introduces the type spaces and provides the equilibrium notions. In Section 3 we present in some detail three examples which illustrate the role of type spaces in the implementation problem and point to the complex relationship between ex post implementation on the naive type space and interim implementation on larger type spaces. In Section 4 we present equivalence results for general social choice environments, also introducing our notion of augmented ex post implementability. The analysis specializes to the quasi-linear environment with and without balanced budget in Section 5. The interim implementability on arbitrary type spaces in the quasi-linear model is investigated in Section 6. We conclude with a discussion of further issues in Section 7.

2. Setup

2.1. Payoff Environment

We consider a finite set of agents $\mathcal{I} = \{1, 2, ..., I\}$. Agent i's *payoff type* is $\theta_i \in \Theta_i$, where Θ_i is a finite set. We write $\theta \in \Theta = \Theta_1 \times ... \times \Theta_I$. There is a set of outcomes A. Each agent has utility function $u_i : A \times \Theta \to \mathbb{R}$. A social correspondence $F : \Theta \to 2^A / \emptyset$. If

[7] See Choi and Kim (1999) for a formal use of this "folk" argument.

the true payoff type profile is θ, the planner would like the outcome to be an element of $F(\theta)$.

Throughout the paper, this environment is fixed and informally understood to be common knowledge. Note that we allow for interdependent types - one agent's payoff from a given outcome depends on other agents' payoff types. Also note that the payoff type profile is understood to contain all information that is relevant to whether the planner achieves his objective or not. For example, we do not allow the planner to trade off what happens in one state with what happens in another state. For the latter reason, this setup is somewhat restrictive. However, it incorporates many classic problems such as the efficient allocation of an object or the efficient choice of public good.

2.2. Type Spaces

While maintaining that the above payoff environment is common knowledge, we want to allow for agents to have all possible beliefs and higher order beliefs about other agents' types. A flexible framework for modelling such beliefs and higher order beliefs are type spaces.

A type space is a collection

$$\mathcal{T} = \left(T_i, \widehat{\theta}_i, \widehat{\pi}_i\right)_{i=1}^{I}.$$

Agent i's *type* is $t_i \in T_i$. A type of agent i must include a description of his payoff type. Thus there is a function

$$\widehat{\theta}_i : T_i \to \Theta_i,$$

with $\widehat{\theta}_i(t_i)$ being agent i's payoff type when his type is t_i. A type of agent i must also include a description of his beliefs about the types of the other agent. Write $\Delta(Z)$ for the space of probability measures on the Borel field of a measurable space Z, there is a function

$$\widehat{\pi}_i : T_i \to \Delta(T_{-i}),$$

with $\widehat{\pi}_i(t_i)$ being agent i's *belief type* when his type is t_i. Thus $\widehat{\pi}_i(t_i)[E]$ is the probability that type t_i of agent i assigns to other agents' types, t_{-i}, being an element of a measurable set $E \subseteq T_{-i}$. In the special case where each T_j is finite, we will abuse notation slightly by writing $\widehat{\pi}_i(t_i)[t_{-i}]$ for the probability that type t_i of agent i assigns to other agents having types t_{-i}.

Sometimes, we will be interested only in the beliefs of a type over the payoff types of other agents. Thus

$$\widehat{\psi}_i : T_i \to \Delta(\Theta_{-i})$$

represent agent i's beliefs about other agents' payoff types, i.e.,

$$\widehat{\psi}_i(t_i)[\theta_{-i}] = \sum_{\left\{t_{-i}:\widehat{\theta}_{-i}(t_{-i})=\theta_{-i}\right\}} \widehat{\pi}_i(t_i)[t_{-i}]. \tag{2.1}$$

2.3. Properties of Type Spaces

Global restrictions on the type space represent common knowledge assumptions among the agents. Some key properties are the following:

- Type Space $\mathcal{T}$ is "naive" if each $T_i = \Theta_i$ and each $\widehat{\theta}_i$ is the identity map.
- Type Space $\mathcal{T}$ is finite if each T_i is finite.
- Finite Type Space $\mathcal{T}$ has full support if $\widehat{\pi}_i(t_i)[t_{-i}] > 0$ for all i and t.
- Finite Type Space $\mathcal{T}$ satisfies the common prior assumption (with prior p) if there exists $p \in \Delta(T)$ such that

  $$\sum_{t_{-i}\in T_{-i}} p(t_i, t_{-i}) > 0 \text{ for all } i \text{ and } t_i$$

  and

  $$\widehat{\pi}_i(t_i)[t_{-i}] = \frac{p(t_i, t_{-i})}{\sum\limits_{t'_{-i}\in T_{-i}} p\left(t_i, t'_{-i}\right)}.$$

Now a canonical approach in the mechanism design literature is to restrict attention to a naive full support common prior type space. Thus it is assumed that there is common knowledge among the agents of a common prior over the payoff types. This assumption is not without loss of generality. The naive type space can be thought of the smallest type space embedding the payoff environment described above.

At the other extreme is the "universal type space" which allows for all possible beliefs or higher order beliefs about payoff types. This universal type space contains all possible type spaces that could have been constructed from the payoff environment. The existence of such a universal type space was proved constructively under a variety of topological assumptions by Mertens and Zamir (1985), Brandenburger and Dekel (1993) and other authors. The constructive argument fails without topological assumptions, but Heifetz and Samet (1998) showed the existence of a universal type space with no topological assumptions, i.e. a type space containing all type spaces of the measure theoretic form described above. For the purpose of this paper, it suffices to conceive the universal type space as the union of all type spaces we could possibly construct.[8]

As we relax implicit common knowledge assumptions in the standard mechanism design approach, we go from a naive full support common prior type space to the union of all type spaces. There are also some important intermediate type spaces; we will mention two here.

It is sometimes assumed that there is a true full support prior p over the payoff types, but the planner does not know what it is. (The complete information implementation literature can be subsumed in this specification.) We can represent this as follows. The type space is

$$T_i = \Delta_{++}(\Theta) \times \Theta_i,$$

with a typical element

$$t_i = (p_i, \theta_i).$$

The payoff type is defined in the natural way:

$$\widehat{\theta}_i(p_i, \theta_i) = \theta_i.$$

The belief type is defined on the assumption that there is common knowledge of the true prior among the agents:

$$\widehat{\pi}_i(p_i, \theta_i)\left[(p_j, \theta_j)_{j \neq i}\right] = \begin{cases} p_i(\theta_{-i} | \theta_i), \text{ if } p_j = p_i \text{ for all } j \neq i, \\ 0, \text{ otherwise.} \end{cases}$$

We will refer to this as the union of all full support common prior type spaces.

A second example of an intermediate type space is the union of all common prior type spaces. In the universal type space, there is no requirement that agents' beliefs be derived from some common prior. However, the common prior is an important economic assumption and it will sometime be interesting to look at the union of all type spaces satisfying the common prior assumption, or, equivalently, the subset of the universal type space where the common prior assumption holds.

[8] In an earlier version of this paper, we described an explicit construction of the universal type space for our environment, along the lines of Mertens and Zamir (1985) (see also Neeman (2001)). A slight variation in the construction arises from the product structure of the payoff type profiles and the maintained assumption that it is common knowledge that each agent knows his true payoff type. The union of all type spaces is potentially larger than this constructed space for two reasons. First, the constructed universal type space uses (and needs) topological assumptions on the underlying space. Second, it is possible to add types with different beliefs over others' types but identical beliefs and higher order beliefs about payoff types. Neither of these two differences matters for the positive or negative results we report in this paper.

2.4. Solution Concepts

Fix a payoff environment and a type space $\mathcal{T}$. A mechanism specifies a message set for each agent and a mapping from message profiles to outcomes. Social choice correspondence F is interim implementable if there exists a mechanism and an interim (or Bayesian) equilibrium of that mechanism such that outcomes are consistent with F. However, by the revelation principle, we can restrict attention to truth-telling equilibria of direct mechanisms.[9] A direct mechanism is a function $f : T \to A$.

Definition 2.1. *A direct mechanism $f : T \to A$ is interim incentive compatible on type space $\mathcal{T}$ if*

$$\int\limits_{t_{-i} \in T_{-i}} u_i \left(f\left(t_i, t_{-i}\right), \widehat{\theta}\left(t_i, t_{-i}\right) \right) d\widehat{\pi}_i\left(t_i\right) \geq \int\limits_{t_{-i} \in T_{-i}} u_i \left(f\left(t_i', t_{-i}\right), \widehat{\theta}\left(t_i, t_{-i}\right) \right) d\widehat{\pi}_i\left(t_i\right)$$

for all i, $t \in T$ and $t_i' \in T_i$.

The notion of interim incentive compatibility is often referred to as Bayesian incentive compatibility. We use the former terminology as there need not be a common prior on the type space.

Definition 2.2. *A direct mechanism $f : T \to A$ on $\mathcal{T}$ achieves F if*

$$f\left(t\right) \in F\left(\widehat{\theta}\left(t\right)\right)$$

for all $t \in T$.

It should be emphasized that a direct mechanism f can prescribe varying allocations for a given payoff profile θ as different types, t and t', may have an identical payoff profile $\theta = \widehat{\theta}\left(t\right) = \widehat{\theta}\left(t'\right)$.

Definition 2.3. *A social choice correspondence F is interim implementable on $\mathcal{T}$ if there exists $f : T \to A$ such that f is interim incentive compatible on $\mathcal{T}$ and f achieves F.*

We will be interested in comparing interim implementation with the stronger solution concept of ex post implementation. Ex post implementation uses the stronger solution concept of ex post equilibrium for incomplete information games.[10] By the revelation principle, it is again enough to verify ex post incentive compatibility.

[9] See Myerson (1991), Chapter 6.

[10] Ex post incentive compatibility was discussed as "uniform incentive compatibility" by Holmstrom and Myerson (1983). Ex post equilibrium is increasingly studied in game theory (see Kalai (2002)) and is often used in mechanism design as a more robust solution concept (Cremer and McLean (1985), Dasgupta and Maskin (2000), Perry and Reny (2002)).

Definition 2.4. *A direct mechanism $f : \Theta \to A$ is ex post incentive compatible if, for all i and $\theta \in \Theta$,*

$$u_i\left(f\left(\theta\right), \theta\right) \geq u_i\left(f\left(\theta_i', \theta_{-i}\right), \theta\right)$$

for all $\theta_i' \in \Theta_i$.

If there are private values (i.e., each $u_i\left(a, \theta\right)$ depends on θ only through θ_i), then ex post incentive compatibility is equivalent to dominant strategies incentive compatibility.

Definition 2.5. *A direct mechanism $f : \Theta \to A$ is dominant strategies incentive compatible if, for all i and $\theta \in \Theta$,*

$$u_i\left(f\left(\theta\right), \theta_i\right) \geq u_i\left(f\left(\theta_i', \theta_{-i}\right), \theta_i\right)$$

for all $\theta_i' \in \Theta_i$.

Definition 2.6. *A social choice correspondence F is ex post implementable if there exists $f : \Theta \to A$ such that f is ex post incentive compatible and*

$$f\left(\theta\right) \in F\left(\theta\right)$$

for all $\theta \in \Theta$.

2.5. Questions

For a fixed social choice correspondence F, we can ask the "ex post equivalence" question:

- when is ex post implementability of F equivalent to interim implementability on all type spaces?

We will provide a number of sufficient conditions for ex post equivalence, but in examples 1 and 2 in the next section, F is not ex post implementable but is interim implementable on any type space.

When ex post equivalence holds, we can ask how big the type space must be in order for interim implementability to be equivalent to ex post implementability? In particular, what is the relation between the following questions:

- is F interim implementable on all full support common prior naive type spaces?
- is F interim implementable on all common prior naive type spaces?
- is F interim implementable on all common prior type spaces?

- is F interim implementable on all type spaces?

For the results in this paper, full support and common prior assumptions are not important.[11] However, the naive type space restriction *is* important. In example 3 in the next section, it is possible to interim implement on any naive type space but not all type spaces.

3. Examples

This section presents three examples illustrating the relationship between interim implementation on different type spaces and ex post implementation.

The first two examples exhibit social choice correspondences that are interim implementable on all type spaces, but are not ex post implementable. The first example is very simple, but relies on (i) a restriction to deterministic allocations, (ii) a social choice correspondence that depends on only one agent's payoff type; and (iii) interdependent types. In the second example, we show how to dispense with all three features. Since this example has private values, we thus have an example where dominant strategies implementation is impossible but interim implementation is possible on any type space.

The third example exhibits a social choice correspondence that is interim implementable on all naive type spaces (with or without the common prior) but is not interim implementable on all type spaces.

3.1. F is Interim Implementable on All Type Spaces but not Ex Post Implementable

EXAMPLE 1. There are two agents. Each agent has two possible types: $\Theta_1 = \{\theta_1, \theta_1'\}$ and $\Theta_2 = \{\theta_2, \theta_2'\}$. There are three possible allocations: $A = \{a, b, c\}$. The payoffs of the two agents are given by the following tables (each box describes agent 1's payoff, then agent 2's payoff):

a	θ_2	θ_2'
θ_1	$1,0$	$-1,2$
θ_1'	$0,0$	$0,0$

b	θ_2	θ_2'
θ_1	$-1,2$	$1,0$
θ_1'	$0,0$	$0,0$

c	θ_2	θ_2'
θ_1	$0,0$	$0,0$
θ_1'	$1,1$	$1,1$

The social choice correspondence is given by

F	θ_2	θ_2'
θ_1	$\{a,b\}$	$\{a,b\}$
θ_1'	$\{c\}$	$\{c\}$

[11] However, the full support assumption is important when we look at full implementation and the common prior assumption is important when we look at revenue maximization.

These choices are maximizers of the sum of agents' utility. The key feature of this example is that the agents agree about the optimal choice when agent 1 is type θ_1'; when agent 1 is type θ_1, they agree that it is optimal to choose either a or b. But 1 prefers a when 2's type is θ_2, while 2 prefers a when his type is θ_2'.

We now show - by contradiction - that this correspondence is not ex post implementable. If F was implementable, we would have to have c chosen at profiles $\left(\theta_1', \theta_2\right)$ and $\left(\theta_1', \theta_2'\right)$; and either a or b chosen at profiles (θ_1, θ_2) and $\left(\theta_1, \theta_2'\right)$. But in order for type θ_1 to have an incentive to tell the truth when he is sure that agent 2 is type θ_2, we must have a chosen at profile (θ_1, θ_2); and in order for type θ_1 to have incentive to tell the truth when he is sure that agent 2 is type θ_2', we must have b chosen at profile $\left(\theta_1, \theta_2'\right)$. But if a is chosen at profile (θ_1, θ_2) and b is chosen at profile $\left(\theta_1, \theta_2'\right)$, then both types of agent 2 will have an incentive to misreport their types when they are sure that agent 1 is type θ_1.

However, the correspondence is interim implementable on any type space using the very simple mechanism of letting agent 1 pick the outcome. There is always an equilibrium of this mechanism where agent 1 will pick outcome a if his type is θ_1 and he assigns probability at least $\frac{1}{2}$ to the other agent being type θ_2; agent 1 will pick outcome b if his type is θ_1 and he assigns probability less than $\frac{1}{2}$ to the other agent being type θ_2; and agent 1 will pick outcome c if his type is θ_1'. By allowing the mechanism to depend on agent 1's beliefs about agent 2's type (something the planner does not care about intrinsically), the planner is able to relax incentive constraints that he cares about.

The failure of ex post implementation in this example relied on the assumption that only pure outcomes were chosen. This restriction can easily be dropped at the expense of adding a third payoff type for agent 1, so that the binding ex post incentive constraint for agent 1 is with a different type and outcome depending on 2's type. The example also had the social choice correspondence depending only on agent 1's payoff type and had interdependent values. We can mechanically change these two assumptions by letting the planner want different outcomes depending on agent 2's type. Now instead of having agent 1's utility depend on agent 2's type, it can depend on the planner's refined choice.

EXAMPLE 2. There are two agents. Agent 1 has three possible types, $\Theta_1 = \left\{\theta_1, \theta_1', \theta_1''\right\}$, and agent 2 has two possible types, $\Theta_2 = \left\{\theta_2, \theta_2'\right\}$. There are eight possible pure allocations, $\{a, b, c, d, a', b', c', d'\}$, and lotteries are allowed, so $A = \Delta\left(\{a, b, c, d, a', b', c', d'\}\right)$. The private value payoffs of agent 1 are given by the following table:

u_1	a	b	c	d	a'	b'	c'	d'
θ_1	1	-1	-1	$\frac{1}{3}$	-1	1	$\frac{1}{3}$	-1
θ_1'	0	0	3	0	0	0	3	0
θ_1''	0	0	0	3	0	0	0	3

The private value payoffs of agent 2 are given by the following table:

u_2	a	b	c	d	a'	b'	c'	d'
θ_2	0	2	0	0	0	2	0	0
θ_2'	2	0	0	0	2	0	0	0

The social choice correspondence F is described by the following table.

	θ_2	θ_2'
θ_1	$\{a, b\}$	$\{a', b'\}$
θ_1'	$\{c\}$	$\{c'\}$
θ_1''	$\{d\}$	$\{d'\}$

(3.1)

We now show - by contradiction - that this correspondence is not ex post implementable. Let p be the probability that a is chosen at profile (θ_1, θ_2) and let p' be the probability that a' is chosen at profile $\left(\theta_1, \theta_2'\right)$. In order for type θ_1 to have an incentive to tell the truth (and not report himself to be type θ_1'') when he is sure that agent 2 is type θ_2, we must have

$$p - (1 - p) \geq \frac{1}{3};$$

thus

$$p \geq \frac{2}{3}. \tag{3.2}$$

In order for type θ_1 to have an incentive to tell the truth (and not report himself to be type θ_1') when he is sure that agent 2 is type θ_2', we must have

$$-p' + \left(1 - p'\right) \geq \frac{1}{3};$$

thus

$$p' \leq \frac{1}{3}. \tag{3.3}$$

But in order for agent 2 to have an incentive to tell the truth when he is type θ_2 and he is sure that agent 1 is type θ_1, we must have

$$2\,(1 - p) \geq 2\left(1 - p'\right);$$

thus

$$p' \geq p. \tag{3.4}$$

However, (3.2), (3.3) and (3.4) generate a contradiction, so ex post implementation is not possible.

But it is straightforward to implement on any interim type space. Consider the following indirect mechanism for any arbitrary type space where individual 1 chooses a message $m_1 \in \{m_1^1, m_1^2, m_1^3, m_1^4\}$ and individual 2 chooses a message $m_2 \in \{m_2^1, m_2^2\}$ and let outcomes be chosen as follows:

	m_2^1	m_2^2
m_1^1	a	a'
m_1^2	b	b'
m_1^3	c	c'
m_1^4	d	d'

There is always an equilibrium where type θ_1 of agent 1 sends message m_1^1 if he believes agent 2 is type θ_2 with probability at least $\frac{1}{2}$ and message m_1^2 if he believes agent 2 is type θ_2 with probability less than $\frac{1}{2}$; type θ_1' always sends message m_1^3; and type θ_1'' always sends message m_1^4. Type θ_2 of agent 2 sends message m_2^1 and type θ_2 sends message m_2^2.

This private values example has the feature that dominant strategies implementation is impossible but interim implementation is possible on any type space, and seems to be the first example in the literature noting this possibility.[12]

As we will see in the next section, a necessary feature of the example is that we have a social choice correspondence (not function) that we are trying to implement. In the example, it was further key that there were aspects of the allocation that the planner did not care about but the agents did. In the example, this may look a little contrived but note that this a natural feature of quasi-linear environments where the planner wants to maximize the total welfare of agents. We will later present a quasi-linear utility example that delivers the same features as this example (Example 4).

3.2. F is Interim Implementable on All Naive Type Spaces but not Interim Implementable on All Type Spaces

EXAMPLE 3. This example has two agents, denoted by 1 and 2. Agent 1 has three possible types, $\Theta_1 = \{\theta_1^1, \theta_1^2, \theta_1^3\}$, and agent 2 has two possible types, $\Theta_2 = \{\theta_2^1, \theta_2^2\}$. The set of allocations is given by $\{a, b, c, d\}$ and each allocation can carry either the name of agent 1 or agent 2. The set of feasible deterministic allocations is therefore given by

$$A = \{a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2\}$$

and we shall allow for lotteries over these deterministic allocations. Each agent receives utility from the allocation and from the name of the allocation. The payoffs *before* the naming decision are given by:

[12] It is often noted that in public good problems with budget balance, dominant strategies implementation is impossible while Bayesian implementation is possible. However, the positive Bayesian implementation results (d'Aspremont and Gerard-Varet (1979) and d'Aspremont, Cremer and Gerard-Varet (2002)) hold only for "generic" priors on a fixed type space, not for all type spaces in our sense.

a	θ_2^1	θ_2^2
θ_1^1	0,2	0,2
θ_1^2	-4,0	1,0
θ_1^3	-3,0	-3,0

b	θ_2^1	θ_2^2
θ_1^1	0,0	0,0
θ_1^2	0,2	0,0
θ_1^3	-3,0	0,0

c	θ_2^1	θ_2^2
θ_1^1	0,0	-3,0
θ_1^2	0,0	0,2
θ_1^3	0,0	0,0

d	θ_2^1	θ_2^2
θ_1^1	-3,0	-3,0
θ_1^2	1,0	-4,0
θ_1^3	0,2	0,2

(3.5)

Each agent attaches an additional utility of 1 to the allocation bearing its name and conversely a disutility of -1 to the allocation bearing the name of the opponent. The additional naming decision on the final utility acts like a zero net transferable utility for the social problem.

The social choice correspondence F which maximizes the sum of the individual utilities is given by:

F	θ_2^1	θ_2^2
θ_1^1	$a_1,\ a_2$	$a_1,\ a_2$
θ_1^2	$b_1,\ b_2$	$c_1,\ c_2$
θ_1^3	$d_1,\ d_2$	$d_1,\ d_2$

(3.6)

With respect to the social choice correspondence F, we next make a few observations regarding the ex post incentive constraints for truthtelling. Starting with agent 2 we note that the efficient allocation always has a value 2, whereas every inefficient allocation has a value 0. In consequence, for all possible naming decisions, agent 2 will never have an incentive to misreport in order to generate a different allocation, but only in order to induce a different naming decision. As for agent 1, he values most alternatives in most states with 0, with a few exceptions. The negative entries -3 and -4 guarantee that he will not have incentive to misreport independent of the naming decision; and the positive valuation, 1, will require differential naming decisions to guarantee incentive compatibility.

First, we show that the social choice correspondence F is interim implementable by some selection $f \in F$ on any naive type space. If type θ_1^2 assigns probability at least $\frac{1}{2}$ to the other agent being type θ_2^1, then the following selection f attains F and is interim incentive compatible:

f	θ_2^1	θ_2^2
θ_1^1	a_1	a_1
θ_1^2	b_1	c_2
θ_1^3	d_2	d_2

(3.7)

We verify the interim incentive compatibility conditions for the social choice function f. We first observe that all the ex post incentive constraints hold except for agent 1 at type profile $\theta_1^2\theta_2^2$, where he has a profitable deviation by misreporting himself to be of type θ_1^1. Suppose then that type θ_1^2 assigns probability p to the other agent being type θ_2^1. His expected payoff to truth-telling, taking into account the naming decision, is

$$p\,(1) + (1-p)\,(-1)\,,$$

while his expected payoff to mis-reporting type θ_1^1 is

$$p\,(-4+1) + (1-p)\,(1+1)\,.$$

Thus truth-telling is optimal as long as

$$p\,(1) + (1-p)\,(-1) \geq p\,(-4+1) + (1-p)\,(1+1)\,,$$

or

$$p \geq \frac{3}{7}. \tag{3.8}$$

Conversely, if type θ_1^2 assigns probability less than $\frac{1}{2}$ to the other agent being type θ_2^1, then the following selection $f' \in F$ is interim incentive compatible:

f'	θ_2^1	θ_2^2
θ_1^1	a_2	a_2
θ_1^2	b_2	c_1
θ_1^3	d_1	d_1

(3.9)

The social choice functions f and f' differ in that all the naming decisions are reversed from f to f'. Again, we find that all ex post incentive constraints hold except for agent 1 at type profile $\theta_1^2\theta_2^1$, where he has a profitable deviation by misreporting himself to be of type θ_1^3. Suppose then that type θ_1^2 assigns probability p to the other agent being type θ_2^1. His expected payoff to truth-telling is

$$p\,(-1) + (1-p)\,(1)\,,$$

while his expected payoff to mis-reporting type θ_1^3 is

$$p\,(1+1) + (1-p)\,(-4+1)\,.$$

Thus truth-telling is optimal as long as

$$p\,(-1) + (1-p)\,(1) \geq p\,(1+1) + (1-p)\,(-4+1)\,,$$

or

$$p \leq \frac{4}{7}. \tag{3.10}$$

It follows from the inequalities (3.8) and (3.10) that if either p is large or if p is large, there is no problem interim implementing F.

However, on richer type spaces than the naive type space, there may be many types with payoff type θ_1^2, some of whom are sure that the other agent is type θ_2^1 while others are sure that he is type θ_2^2. That is the idea behind the following example of a "complete information" type space where F cannot be interim implemented. We consider the following type space:

	t_2^1	t_2^2	t_2^3	t_2^4	t_2^5	t_2^6	
t_1^1	$\frac{1}{6}$	0	0	0	0	0	θ_1^1
t_1^2	0	$\frac{1}{6}$	0	0	0	0	θ_1^2
t_1^3	0	0	$\frac{1}{6}$	0	0	0	θ_1^3
t_1^4	0	0	0	$\frac{1}{6}$	0	0	θ_1^3
t_1^5	0	0	0	0	$\frac{1}{6}$	0	θ_1^2
t_1^6	0	0	0	0	0	$\frac{1}{6}$	θ_1^1
	θ_2^1	θ_2^1	θ_2^1	θ_2^2	θ_2^2	θ_2^2	

Thus there are six types for each agent, t_1^k and t_2^l. The entries in the cell describe the probabilities of the common prior, which puts all probability mass on the diagonal. The payoff type corresponding to each type appears at the end of the row/column corresponding to that type. Thus, for example, type t_1^3 of agent 1 has payoff type θ_1^3 and believes that agent 2 has a payoff type θ_2^1 with probability one. It is in this sense, that we speak of complete information. We require that F is implemented even at "impossible" (zero probability) type profiles, but we could clearly adapt the example to have small probabilities off the diagonal.

Our impossibility argument will depend only on what happens at twelve critical type profiles: the diagonal profiles and the type profiles where agent 1 claims to be one type higher and agent 2 claims to be one type lower. In the next table, we note which pair of allocations at these twelve profiles are consistent with implementing F.

	t_2^1	t_2^2	t_2^3	t_2^4	t_2^5	t_2^6	
t_1^1	$\{a_1, a_2\}$					$\{a_1, a_2\}$	θ_1^1
t_1^2	$\{b_1, b_2\}$	$\{b_1, b_2\}$					θ_1^2
t_1^3		$\{d_1, d_2\}$	$\{d_1, d_2\}$				θ_1^3
t_1^4			$\{d_1, d_2\}$	$\{d_1, d_2\}$			θ_1^3
t_1^5				$\{c_1, c_2\}$	$\{c_1, c_2\}$		θ_1^2
t_1^6					$\{a_1, a_2\}$	$\{a_1, a_2\}$	θ_1^1
	θ_2^1	θ_2^1	θ_2^1	θ_2^2	θ_2^2	θ_2^2	

We observe that the incentive constraints for agent 1 and agent 2 form jointly a cycle through the type space. As we mentioned in the beginning of the example, we allow for random allocations. Consequently we write p_{kl} for the probability of the naming 1 and $1 - p_{kl}$ for the naming 2 when the type profile is $t = \left(t_1^k, t_2^l\right)$. The incentive constraints corresponding to types t_1^k mis-reporting to be type t_1^{k+1} (modulo 6) imply (for $k = 1, 2, .., 6$ respectively):

$$\begin{aligned}
&(1)\, p_{11} + (-1)\,(1 - p_{11}) \geq (1)\, p_{21} + (-1)\,(1 - p_{21}) \\
&(1)\, p_{22} + (-1)\,(1 - p_{22}) \geq (1+1)\, p_{32} + (1-1)\,(1 - p_{32}) \\
&(1)\, p_{33} + (-1)\,(1 - p_{33}) \geq (1)\, p_{43} + (-1)\,(1 - p_{43}) \\
&(1)\, p_{44} + (-1)\,(1 - p_{44}) \geq (1)\, p_{54} + (-1)\,(1 - p_{54}) \\
&(1)\, p_{55} + (-1)\,(1 - p_{55}) \geq (1+1)\, p_{65} + (1-1)\,(1 - p_{65}) \\
&(1)\, p_{66} + (-1)\,(1 - p_{66}) \geq (1)\, p_{16} + (-1)\,(1 - p_{16})
\end{aligned} \tag{3.11}$$

The incentive constraints corresponding to types t_2^l mis-reporting to be type t_2^{l-1} imply (for $l = 1, 2, .., 6$ respectively):

$$\begin{aligned}
&(2-1)\, p_{11} + (2+1)\,(1 - p_{11}) \geq (2-1)\, p_{16} + (2+1)\,(1 - p_{16}) \\
&(2-1)\, p_{22} + (2+1)\,(1 - p_{22}) \geq (2-1)\, p_{21} + (2+1)\,(1 - p_{21}) \\
&(2-1)\, p_{33} + (2+1)\,(1 - p_{33}) \geq (2-1)\, p_{32} + (2+1)\,(1 - p_{32}) \\
&(2-1)\, p_{44} + (2+1)\,(1 - p_{44}) \geq (2-1)\, p_{43} + (2+1)\,(1 - p_{43}) \\
&(2-1)\, p_{55} + (2+1)\,(1 - p_{55}) \geq (2-1)\, p_{54} + (2+1)\,(1 - p_{54}) \\
&(2-1)\, p_{66} + (2+1)\,(1 - p_{66}) \geq (2-1)\, p_{65} + (2+1)\,(1 - p_{65})
\end{aligned} \tag{3.12}$$

The inequalities (3.11) and (3.12) have a very simply structure. With very few exceptions, the payoffs appearing on the lhs and rhs of the inequalities are identical and only the probability weights differ. These inequalities are generated either by true and misreported types which induce only different naming decision but identical allocational decisions or different allocation decisions over which the agent is indifferent. The exceptions are the second and fifth inequality of agent 1, where a misreported type also leads to a different allocational decision. Re-arranging the inequalities, we obtain

$$\begin{array}{rcl@{\qquad}rcl}
0 & \geq & p_{21} - p_{11}, & 0 & \geq & p_{11} - p_{61} \\
-\frac{1}{2} & \geq & p_{32} - p_{22}, & 0 & \geq & p_{22} - p_{21} \\
0 & \geq & p_{43} - p_{33}, & 0 & \geq & p_{33} - p_{32} \\
0 & \geq & p_{54} - p_{44}, & 0 & \geq & p_{44} - p_{43} \\
-\frac{1}{2} & \geq & p_{65} - p_{55}, & 0 & \geq & p_{55} - p_{54} \\
0 & \geq & p_{16} - p_{66}, & 0 & \geq & p_{66} - p_{65}
\end{array}$$

When we sum these twelve constraints, the probabilities on the right hand side of the inequalities cancel out and we are left with the desired contradiction for any arbitrary choice of probabilities, namely $-1 \geq 0$. The probabilities cancelled out because the set of incentive constraints for agent 1 and agent 2 formed jointly a cycle through the type space.

4. Ex Post Equivalence Results for General Environments

Following the initial set of examples, we now present general results about the relationship between ex post implementability and interim implementability on larger type spaces. As suggested by the examples, the relationship between these implementation notions will depend on the nature of the implementation problem as represented by the social choice correspondence F. The first result is an immediate implication from the definition of ex post equilibrium.

Proposition 4.1. *If F is ex post implementable, then F is interim implementable on any type space.*

PROOF: If F is ex post implementable, then by hypothesis there exists $f^* : \Theta \to A$ with $f^*(\theta) \in F(\theta)$ for all θ, such that for all i, all θ and all θ_i':

$$u_i(f^*(\theta), \theta) \geq u_i\left(f^*\left(\theta_i', \theta_{-i}\right), \theta\right)$$

Consider then an arbitrary type space $\mathcal{T}$ and the direct mechanism $f : T \to A$ with $f(t) = f^*\left(\widehat{\theta}(t)\right)$. Incentive compatibility now requires

$$\begin{aligned} t_i &\in \arg\max_{t_i' \in T_i} \int_{t_{-i} \in T_{-i}} u_i\left(f\left(t_i', t_{-i}\right), \left(\widehat{\theta}_i(t_i), \widehat{\theta}_{-i}(t_{-i})\right)\right) d\widehat{\pi}_i(t_i) \\ &= \arg\max_{t_i' \in T_i} \int_{t_{-i} \in T_{-i}} u_i\left(f^*\left(\widehat{\theta}_i\left(t_i'\right), \widehat{\theta}_{-i}(t_{-i})\right), \left(\widehat{\theta}_i(t_i), \widehat{\theta}_{-i}(t_{-i})\right)\right) d\widehat{\pi}_i(t_i). \end{aligned}$$

This requires that

$$\widehat{\theta}_i(t_i) = \arg\max_{\theta_i \in \Theta_i} \int_{t_{-i} \in T_{-i}} u_i\left(f^*\left(\theta_i, \widehat{\theta}_{-i}(t_{-i})\right), \left(\widehat{\theta}_i(t_i), \widehat{\theta}_{-i}(t_{-i})\right)\right) d\widehat{\pi}_i(t_i).$$

But by hypothesis of ex post implementability, truthtelling is a best response for every possible profile θ_{-i}, and thus it remains a best response for arbitrary expectations over Θ_{-i}. ■

While Examples 1 and 2 in the previous section showed that the converse does not always hold, we can identify an important class of problems for which the equivalence can be established.

Proposition 4.2. *If F is single valued and F is interim implementable on every full support common prior naive type space $\mathcal{T}$, then F is ex post implementable.*

PROOF: Suppose therefore that the social choice function is not ex-post implementable, then there exists a payoff type profile θ, an agent i and a profitable deviation $\theta_i' \neq \theta_i$ such that

$$u_i\left(F\left(\theta_i, \theta_{-i}\right), \left(\theta_i, \theta_{-i}\right)\right) < u_i\left(F\left(\theta_i', \theta_{-i}\right), \left(\theta_i, \theta_{-i}\right)\right).$$

But now consider the naive type space with full support common prior p. One interim incentive constraint will be that

$$\sum_{\theta_{-i}'} p\left(\theta_{-i}'|\theta_i\right) u_i\left(F\left(\theta_i, \theta_{-i}'\right), \left(\theta_i, \theta_{-i}'\right)\right) \geq \sum_{\theta_{-i}'} p\left(\theta_{-i}'|\theta_i\right) u_i\left(F\left(\theta_i', \theta_{-i}'\right), \left(\theta_i, \theta_{-i}'\right)\right).$$

But this constraint will be violated if the full support prior p puts probability sufficiently close to 1 on θ. ■

This immediately implies the following strong equivalence result.

Proposition 4.3. *If F is single valued, then the following are equivalent:*

1. *F is interim implementable on all type spaces;*

2. *F is interim implementable on all common prior type spaces;*

3. *F is interim implementable on all common prior naive type spaces;*

4. *F is interim implementable on all common prior full support naive type spaces;*

5. *F is ex post implementable.*

PROOF. (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4) all follow by definition. In each case, we are asking for interim implementation on a smaller type space. By Proposition 4.2, (4) $\Rightarrow$ (5). By Proposition 4.1, (5) $\Rightarrow$ (1). ■

The argument here is straightforward. As long as we pin down what the mechanism does as a function of payoff types, then we can manipulate beliefs so that a failure of ex post incentive compatibility can be translated into a failure of interim incentive compatiblity. The logic of this argument goes back to Ledyard (1978, 1979) and Dasgupta, Hammond and Maskin (1979). If F is single-valued and F is implemented, it is true without loss of generality that any direct mechanism must depend only on payoff types. If F is a fat correspondence, and we restrict the planner to use a mechanism that depends only on payoff types, then the ex post equivalence results will go through. But if F is a fat correspondence, this assumption is *with* loss of generality. As we saw in examples 1 and 2, allowing the mechanism to depend on belief types may relax interim incentive constraints. In particular, if there happens to be common knowledge among the agents of a common prior on the naive type space, the planner might want to make the mechanism sensitive to the prior.

The gap between ex post and interim implementability in the general case of a correspondence leaves open the possibility that a weaker notion of ex post implementability could lead to an equivalence result. We introduce this weaker notion by allowing each agent to report a message $m_i \in M_i$ besides his payoff type. More precisely, an augmented direct mechanism is a game where each agent i reports a payoff type θ_i and in addition reports a message $m_i \in M_i$. Let

$$M_i^* = \{(\theta_i, m_i) : \theta_i \in \Theta_i, m_i \in M_i\}$$

be the set of extended reports which can be sent by agent i. Thus an augmented mechanism is parameterized by:

$$M^* = M_1^* \times ... \times M_I^*.$$

We first define the notion of augmented ex post implementability, where f is allowed to depend on M^*, or $f : M^* \to A$.

Definition 4.4. *A social choice rule $f : M^* \to A$ is augmented ex post incentive compatible given M^* if, for all i, $\theta_i \in \Theta_i$ and $\lambda_i \in \Delta\left(M_{-i}^*\right)$, there exists $m_i \in M_i$ such that*

$$(\theta_i, m_i) \in \underset{\left(\theta_i', m_i'\right)}{\arg\max} \int_{(\theta_{-i}, m_{-i}) \in M_{-i}^*} u_i\left(f\left(\left(\theta_i', \theta_{-i}\right), \left(m_i', m_{-i}\right)\right), (\theta_i, \theta_{-i})\right) d\lambda_i\left(\theta_{-i}, m_{-i}\right) \tag{4.1}$$

Definition 4.5. *A social choice correspondence F is augmented ex post implementable if there exists M^* and $f : M^* \to A$ such that f is augmented ex post incentive compatible and*

$$f(\theta, m) \in F(\theta)$$

for all $(\theta, m) \in M^*$.

The notion of augmented ex post implementability requires that each agent i has a best response for every possible distribution over M_{-i}^* which involves reporting his payoff type truthfully. In the special case where each M_i is a singleton, the notion of augmented ex post implementability reduces to ex post implementability. If M_i is a singleton, then the condition (4.1) simply states that truthtelling is a best response under the selection $f \in F$ for every possible distribution over the payoff types Θ_{-i} of the remaining agents. It is then sufficient to evoke the equivalence result in Proposition 4.2 to argue that the two notions, ex post and augmented ex post implementability, coincide.

Proposition 4.6. *If F is interim implementable on every $\mathcal{T}$, then F is augmented ex post implementable.*

PROOF: Suppose that F is interim implementable on every $\mathcal{T}$. It is a fortiori interim implementable on the universal type space. By hypothesis, there then exists a mapping $f : T \rightarrow A$ such that the interim incentive compatibility condition is satisfied for all i and all $t_i \in T_i$:

$$t_i \in \underset{t'_i \in T_i}{\arg\max} \int_{t_{-i} \in T_{-i}} u_i \left(f \left(t'_i, t_{-i} \right), \widehat{\theta} \left(t \right) \right) d\widehat{\pi}_i \left(t_i \right).$$

As every type t_i can be represented by a pair $t_i = \left(\widehat{\theta}_i \left(t_i \right), \widehat{\pi}_i \left(t_i \right) \right)$, consisting of its payoff type $\widehat{\theta}_i \left(t_i \right)$ and its belief type $\widehat{\pi}_i \left(t_i \right)$, we can rewrite the incentive compatibility condition as follows:

$$\left(\widehat{\theta}_i \left(t_i \right), \widehat{\pi}_i \left(t_i \right) \right) \in \underset{(\theta_i, \pi_i) \in \Theta_i \times \Delta(T_{-i})}{\arg\max}$$
$$\int_{t_{-i} \in T_{-i}} u_i \left(f \left(\left(\theta_i, \pi_i \right), \left(\widehat{\theta}_{-i} \left(t_{-i} \right), \widehat{\pi}_{-i} \left(t_{-i} \right) \right) \right), \widehat{\theta} \left(t \right) \right) d\widehat{\pi}_i \left(t_i \right)$$

Based on the universal type space we can then find a message space M_i such that F is augmented ex post implementable. Let $M_i = \Delta \left(T_{-i} \right)$ for all $\theta_i \in \Theta_i$ and consequently let $M_i^* = \Theta_i \times \Delta \left(T_{-i} \right)$. It remains to show that for all i, for all θ_i, and every $\lambda_i \in \Delta \left(M_{-i}^* \right)$, there exists m_i or equivalently π_i such that (4.1) is satisfied. By hypothesis, we have assumed implementability on the universal type space, and we can then use the fact that the universal type space is rich. More precisely for every distribution $\lambda_i \in \Delta \left(M_{-i}^* \right)$ and for every payoff type θ_i there is type t_i which shares the payoff type with θ_i, or $\theta_i = \widehat{\theta}_i \left(t_i \right)$ and shares the beliefs with λ_i, or:

$$\lambda_i = \widehat{\pi}_i \left(t_i \right).$$

It follows that M^* and f guarantee augmented ex post incentive compatibility of F. ■

The notion of augmented ex post implementability is also a sufficient condition for interim implementability, provided that the message space M^* and the type space $\mathcal{T}$ are finite spaces.

Proposition 4.7. *If F is augmented ex post implementable with a finite message space M, then F is interim implementable on every finite type space $\mathcal{T}$.*

PROOF. Suppose SCC F is augmented ex post implementable and fix an arbitrary finite type space $\mathcal{T}$. We show that there is an incomplete information game based on an indirect mechanism where truth-telling is an interim equilibrium. In the incomplete information game each agent i has to announce a payoff type and a message $m_i \in M_i$ and thus a strategy for agent i is given by $s_i : T_i \to \Delta(M_i^*)$. The outcome function g is assumed to be identical to the selection $f \in F$ which guarantees augmented ex post implementability by assumption. Under this indirect mechanism, consider first a restricted game in which each agent is forced to report his payoff type truthfully, or $\theta_i = \widehat{\theta}_i(t_i)$, but is unconstrained to report an arbitrary message $m_i \in M_i$. As the type space and the message space is finite, we know by standard existence arguments that the restricted game has an interim equilibrium. Next, consider removing the restriction of truthful reporting from agent i. By the assumption of augmented ex post implementability, we know that for an arbitrary distribution $\lambda_i(\theta_{-i}, m_{-i})$, there exists a best response under which the agent reports truthfully. Thus we can remove the restriction of truthful reporting from every agent and yet maintain the candidate strategy $s_i^* : T_i \to \Delta(M_i^*)$ as an interim equilibrium of the unrestricted game. ■

The finiteness of M^* and $\mathcal{T}$ is required to guarantee the existence of an interim equilibrium in a restricted game where each agent is forced to report his payoff truthfully, but is unconstrained with respect to his choice of a message $m_i \in M_i$. The basic argument for the sufficiency of augmented ex post implementability could be extended to larger type and message spaces provided we can still guarantee the existence of an interim equilibrium in the restricted game.[13]

5. Ex Post Equivalence for Quasi-Linear Environments

In this section we pursue the robustness of implementation problems for correspondences in quasi-linear environments. We consider allocation problems with and without the balanced budget requirement. The associated social choice mapping for these class of

[13] A weaker notion of augmented ex post implementability would require every agent i to have truth-telling as a best response only against all distributions $\lambda_i(m_{-i}) \in \Delta(M_{-i})$ for every payoff relevant profile θ_{-i}; or for all i, θ and all $\lambda_i(m_{-i}) \in \Delta(M_{-i})$:

$$(\theta_i, m_i) \in \underset{(\theta_i', m_i')}{\arg\max} \int_{m_{-i} \in M_{-i}} u_i\left(f\left(\left(\theta_i', \theta_{-i}\right), \left(m_i', m_{-i}\right)\right), \left(\theta_i, \theta_{-i}\right)\right) d\lambda_i(m_{-i}).$$

As augmented ex post implementability implies this weaker notion, Proposition 4.6 remains intact. We conjecture however that the converse, and hence Proposition 4.7 would not go through anymore with this weaker notion.

problems naturally takes the form of a correspondence as the set of permissible transfers is not unique.

5.1. The Quasi-Linear Environment

A quasi-linear environment takes the following special form. The outcome space A is the product of an allocation $z \in Z$ and transfers $y_i \in \mathbb{R}$ to every agent i: $A = Z \times \mathbb{R}^I$. The utility function $u_i(a, \theta)$ is additively separable, or

$$u_i((z, y), \theta) \triangleq v_i(z, \theta) + y_i,$$

for some $v_i : Z \times \Theta \to \mathbb{R}$ for each i. The social choice correspondence $F_\xi(\theta)$ is composed of a function $\xi : \Theta \to Z$ and arbitrary transfers $y \in \mathbb{R}^I$:

$$F_\xi(\theta) \triangleq \{(z, y) \in A : z = \xi(\theta)\}.$$

The correspondence $F_\xi(\theta)$ may represent the problem of implementing an efficient allocation without requiring a balanced budget.

We first express the possibility of implementation as a set of linear constraints. The only data of the problem that will interest us will be the incentive of a payoff type to manipulate the choice of $z \in Z$ by mis-reporting his payoff type. His ex post gain to reporting himself to be type θ_i' when he is type θ_i and he is sure that others have type profile θ_{-i} is:

$$\zeta_i\left(\theta_i, \theta_i', \theta_{-i}\right) \triangleq v_i\left(\xi\left(\theta_i', \theta_{-i}\right), \theta\right) - v_i\left(\xi\left(\theta_i, \theta_{-i}\right), \theta\right). \tag{5.1}$$

A set of transfer functions $y = (y_1, ..., y_I)$, each $y_i : \Theta \to \mathbb{R}$, satisfy ex post incentive compatibility if

$$y_i(\theta_i, \theta_{-i}) - y_i\left(\theta_i', \theta_{-i}\right) \geq \zeta_i\left(\theta_i, \theta_i', \theta_{-i}\right)$$

for all i, θ_i, θ_i' and θ_{-i}.

Proposition 5.1. *If F_ξ is interim implementable on every full support common prior naive type space $\mathcal{T}$, then F_ξ is ex post implementable.*

PROOF. Suppose that F_ξ is not ex post equilibrium implementable. Then for some i, there does not exist y_i such that for all θ :

$$y_i(\theta_i, \theta_{-i}) - y_i\left(\theta_i', \theta_{-i}\right) \geq \zeta_i\left(\theta_i, \theta_i', \theta_{-i}\right), \forall \theta_i' \in \Theta_i. \tag{5.2}$$

Consequently there exists at least one type profile $\theta_{-i} = \theta_{-i}^+$ such that no transfer function $y_i^+(\theta_i) \triangleq y_i\left(\theta_i, \theta_{-i}^+\right)$ can satisfy (5.2). In other words, the solution to the following maxmin problem:

$$\max_{\{y_i'(\cdot)\}} \left\{ \min_{(\theta_i, \theta_i') \in \Theta_i \times \Theta_i} \left\{ y_i'(\theta_i) - y_i'\left(\theta_i'\right) - \zeta_i\left(\theta_i, \theta_i', \theta_{-i}'\right) \right\} \right\}, \tag{5.3}$$

has a strictly negative solution, say $-\delta$. Without loss of generality, we may assume that the negative solution arises locally from the incentive constraint at θ_i^+ versus θ_i^-.

Now suppose that ξ is interim equilibrium implementable on the naive type space for all independent priors $p \in \Delta(\Theta)$. Consequently, for every p there must exist a set of transfers functions, $y_i^p : \Theta \to \mathbb{R}$, and associated interim payments:

$$y_i^p(\theta_i) \triangleq \sum_{\theta_{-i} \in \Theta_{-i}} y_i^p(\theta_i, \theta_{-i})\, p(\theta_{-i} | \theta_i),$$

such that $\forall i, \forall \theta_i, \theta_i'$:

$$y_i^p(\theta_i) - y_i^p(\theta_i') \geq \sum_{\theta_{-i} \in \Theta_{-i}} \zeta_i(\theta_i, \theta_i', \theta_{-i})\, p(\theta_{-i} | \theta_i).$$

In particular, the incentive constraints have to be satisfied at true type θ_i^+ vis-a-vis reported type θ_i^-, or:

$$y_i^p(\theta_i^+) - y_i^p(\theta_i^-) \geq \zeta_i(\theta_i^+, \theta_i^-, \theta_{-i}^+)\, p(\theta_{-i}^+) + \sum_{\theta_{-i} \neq \theta_{-i}^+} \zeta_i(\theta_i^+, \theta_i^-, \theta_{-i})\, p(\theta_{-i}).$$

For a given function ξ and a given $\delta > 0$, we can find $\varepsilon > 0$ such that for all probability distributions p with a marginal probability

$$p_{-i}(\theta_{-i}^+) > 1 - \varepsilon, \tag{5.4}$$

we have

$$\sum_{\theta_{-i} \neq \theta_{-i}^+} \zeta_i(\theta_i^+, \theta_i^-, \theta_{-i})\, p(\theta_{-i}) < \frac{\delta}{2}.$$

It follows that a necessary condition for interim implementation for all probability distributions p satisfying (5.4) is that there exist a transfer functions $y_i^p(\theta_i)$ such that

$$y_i^p(\theta_i^+) - y_i^p(\theta_i^-) - \zeta_i(\theta_i^+, \theta_i^-, \theta_{-i}^+)\, p(\theta_{-i}^+) > -\frac{\delta}{2}. \tag{5.5}$$

But for $p_{-i}(\theta_{-i}^+)$ arbitrarily close to 1, the condition (5.5) is eventually in contradiction with the solution to the max min problem represented in (5.3). This concludes the proof. ∎

A version of this result was reported in Bergemann and Valimaki (2002). While the argument is again straightforward, notice that it is distinct from the argument for Proposition 4.3. In particular, the argument for Proposition 5.1 allows the transfer payments to depend on the prior of the agents. The outcome function is therefore allowed to vary with the beliefs of the agents and we do not require the same selection of the social choice correspondence to work for all prior distributions.

We immediately also have ex post equivalence (and also equivalence for any intermediate type spaces):

Proposition 5.2. *F_ξ is interim implementable on all type spaces if and only if F_ξ is ex post implementable.*

This result shows that in the quasi-linear environment without balanced budget requirements, concerns about the richness of the type space are misplaced. Even if there is common knowledge of the naive type space, there is common knowledge among the agents of the prior on that type space *and* the designer knows what that prior is, implementation for every such common knowledge prior is equivalent to ex post implementation. Following Maskin (1992), a number of papers have examined ex post equilibrium implementation for environments with interdependent types in this quasi-linear setting.[14] The current equivalence result shows that ex post equilibrium implementation is also required simply to ensure interim implementation on the naive type space for all independent distributions over types. In the precise context of this result, the notion of an ex post equilibrium can thus be conceived as a robustness requirement for implementation problems embedded in a solution concept.

It is further an immediate consequence that the impossibility results in Jehiel and Moldovanu (2001) for ex post implementation with multi-dimensional signals extend to interim implementation when we impose robustness requirements as in the above proposition.

5.2. The Quasi-Linear Environment with Budget Balance

The social choice problem is now augmented by requiring that the allocation $z \in Z$ can be implemented with a balanced budget. Consequently, the social choice correspondence takes the form:

$$F_\xi(\theta) \triangleq \left\{ (z, y) \in A : z = \xi(\theta) \text{ and } \sum_{i=1}^{I} y_i(\theta) = 0 \right\}.$$

Our results for this environment will exploit a dual characterization of when implementation is possible. Our dual approach to analyzing this problem builds on the classic work of d'Aspremont and Gerard-Varet (1979) and the more recent works of d'Aspremont, Cremer and Gerard-Varet (1995, 2002). In contrast to these works who

[14] Dasgupta and Maskin (2000), Jehiel and Moldovanu (2001), Perry and Reny (2002), Bergemann and Valimaki (2002).

use the interim dual alone, we shall make use of both the ex post and the interim dual to establish relationships between these two different set of conditions.

The dual variables of our characterization will be the multipliers of the budget balance constraints, ν, and the multipliers of the incentive constraints, λ_i. In the case of ex post incentive constraints, we say that (ν, λ), with $\nu : \Theta \to \mathbb{R}$ and $\lambda_i : \Theta_i \times \Theta_i \times \Theta_{-i} \to \mathbb{R}_+$, satisfy the ex post flow condition (EF) if

$$\nu(\theta) = \sum_{\theta_i' \in \Theta_i} \lambda_i\left(\theta_i, \theta_i', \theta_{-i}\right) - \sum_{\theta_i' \in \Theta_i} \lambda_i\left(\theta_i', \theta_i, \theta_{-i}\right) \tag{5.6}$$

for all $\theta \in \Theta$ and all i; and satisfy the ex post weighting condition (EW) if

$$\sum_{i=1}^{I} \sum_{\theta \in \Theta} \sum_{\theta_i' \in \Theta_i} \lambda_i\left(\theta_i, \theta_i', \theta_{-i}\right) \zeta_i\left(\theta_i, \theta_i', \theta_{-i}\right) > 0. \tag{5.7}$$

By contrast, in the case of interim implementation, we face a set of interim incentive constraints on a general type space T. For ease of notation we write the beliefs of agent i with type t_i in this section as follows:

$$\widehat{\pi}_i(t_i)[t_{-i}] \triangleq \widehat{\pi}_i(t_{-i} | t_i).$$

The interim gain of mis-reporting t_i' by a true type t_i is then given by

$$\zeta_i\left(t_i, t_i'\right) \triangleq \sum_{t_{-i} \in T_{-i}} \zeta\left(t_i, t_i', t_{-i}\right) \widehat{\pi}_i\left(t_{-i} | t_i\right),$$

where the underlying ex post gain of mis-reporting is defined for general type spaces , as earlier in (5.1) for payoff types, by:

$$\zeta_i\left(t_i, t_i', t_{-i}\right) = v_i\left(\xi\left(\widehat{\theta}_i\left(t_i'\right), \widehat{\theta}_{-i}\left(t_{-i}\right)\right), \widehat{\theta}\left(t_i, t_{-i}\right)\right) - v_i\left(\xi\left(\widehat{\theta}_i\left(t_i\right), \widehat{\theta}_{-i}\left(t_{-i}\right)\right), \widehat{\theta}\left(t_i, t_{-i}\right)\right).$$

We then say that a set of multipliers (ν, λ), with $\nu : T \to \mathbb{R}$ and $\lambda_i : T_i \times T_i \to \mathbb{R}_+$, satisfy the interim flow condition (IF) if there exist, for all t and all i:

$$\nu(t) = \sum_{t_i' \in T_i} \lambda_i\left(t_i, t_i'\right) \widehat{\pi}_i\left(t_{-i} | t_i\right) - \sum_{t_i' \in T_i} \lambda_i\left(t_i', t_i\right) \widehat{\pi}_i\left(t_{-i} | t_i'\right) \tag{5.8}$$

and the interim weighting condition (IW):

$$\sum_{i=1}^{I} \sum_{t \in T} \sum_{t_i' \in T_i} \lambda_i\left(t_i, t_i'\right) \zeta_i\left(t_i, t_i'\right) > 0. \tag{5.9}$$

The precise dual characterization of the balanced budget implementation problem is given next.

Lemma 5.3 (Dual Characterization).

1. *The following are equivalent:*

   - *SCC F_ξ is ex post implementable;*
   - *there do not exist (ν, λ) satisfying EF and EW.*

2. *The following are equivalent:*

   - *SCC F_ξ is interim implementable;*
   - *there do not exist (ν, λ) satisfying IF and IW.*

PROOF: See appendix. ■

We proceed to establish equivalence between ex post and interim implementation by using the dual conditions. We show that if either there are only two agents or, for an arbitrary number of agents, the payoff space of each agent is binary then the equivalence between the implementation notions can be established. We show by means of an example with three agents and three states that the conditions for equivalence are rather tight.

The difficult part in the equivalence result is to show that interim implementation on a larger type space implies ex post implementation on the payoff type space. The proof proceeds by contrapositive in either of the two instances. We start with the failure of ex post implementation and appeal to the dual characterization to assert the existence of ex post multipliers which solve the flow and weight conditions. We then seek to find interim multipliers such that the interim flow conditions are met and the interim weight condition replicates the ex post weight condition. At first glance, this seems difficult to achieve. Notice that the ex post multipliers are defined for every type pair of agent i, $\left(\theta_i, \theta_i'\right)$, conditional on the type profile θ_{-i} of the other agents, or $\lambda_i\left(\theta_i, \theta_i', \theta_{-i}\right)$. In contrast the interim multipliers for every type pair of agent i, (t_i, t_i') have to be set without reference to the type profile of the other agents, or $\lambda_i\left(t_i, t_i'\right)$. But this apparent deficit in the ability to mimic the ex post weight condition can be compensated by (i) a large type space T_i relative to the payoff type space Θ_i, and (ii) belief types which generate either positive and equal or zero conditional probabilities. These two features allow the interim multipliers to replicate the ex post multipliers in a variety of circumstances.

Before we present the first result, it will be useful to record a simple fact regarding the nature of the dual solution. The ex post flow condition:

$$\forall \theta, \forall i, \quad v\left(\theta\right) = \sum_{\theta_i' \in \Theta_i} \lambda_i\left(\theta_i, \theta_i', \theta_{-i}\right) - \sum_{\theta_i' \in \Theta_i} \lambda_i\left(\theta_i', \theta_i, \theta_{-i}\right),$$

requires that the *net flow* of the incentive multipliers at every type profile is equalized across the agents. The next lemma states that if a dual solution exists it can always be strengthened so that the gross flows at every profile θ, i.e. the inflows and the outflows, are equalized across agents, or

$$\forall \theta, \forall i, j, \quad \sum_{\theta_i' \in \Theta_i} \lambda_i\left(\theta_i, \theta_i', \theta_{-i}\right) = \sum_{\theta_j' \in \Theta_j} \lambda_j\left(\theta_j, \theta_j', \theta_{-j}\right),$$

as well as

$$\forall \theta, \forall i, j, \quad \sum_{\theta_i' \in \Theta_i} \lambda_i\left(\theta_i', \theta_i, \theta_{-i}\right) = \sum_{\theta_j' \in \Theta_j} \lambda_j\left(\theta_j', \theta_j, \theta_{-j}\right).$$

The equal gross flow property also holds for the interim dual but is not used here.

Lemma 5.4 (Equal Gross Flows).
Suppose there exists a solution (ν, λ) to the ex post dual problem, then there exists an equal gross flow solution to the ex post dual problem.

PROOF: The proof is by construction. Consider an arbitrary θ where the gross flows are not equalized across agents. It follows that there is some agent i who has the largest gross flows, or

$$\forall j \neq i, \quad \sum_{\theta_i' \in \Theta_i} \lambda_i\left(\theta_i, \theta_i', \theta_{-i}\right) \geq \sum_{\theta_j' \in \Theta_j} \lambda_j\left(\theta_j, \theta_j', \theta_{-j}\right),$$

as well as

$$\forall j \neq i, \quad \sum_{\theta_i' \in \Theta_i} \lambda_i\left(\theta_i', \theta_i, \theta_{-i}\right) \geq \sum_{\theta_j' \in \Theta_j} \lambda_j\left(\theta_j', \theta_j, \theta_{-j}\right).$$

We can then create (or increase) for all j, $\lambda_j\left(\theta_j, \theta_j, \theta_{-j}\right)$ to $\lambda_j^+\left(\theta_j, \theta_j, \theta_{-j}\right)$ by setting:

$$\lambda_j^+\left(\theta_j, \theta_j, \theta_{-j}\right) \triangleq \lambda_j\left(\theta_j, \theta_j, \theta_{-j}\right) + \sum_{\theta_i' \in \Theta_i} \lambda_i\left(\theta_i, \theta_i', \theta_{-i}\right) - \sum_{\theta_j' \in \Theta_j} \lambda_j\left(\theta_j, \theta_j', \theta_{-j}\right).$$

It follows that at θ, the gross flows are now equalized and the inequality in the dual condition remained unchanged as $\xi_j\left(\theta_j, \theta_j, \theta_{-j}\right) = 0$ for all θ_j and θ_{-j}. Notice finally that the equalization can be performed at every θ independently. ■

The critical type space for the interim implementation in the next result is the complete information type space. We used this type space earlier in Example 3 and describe it now in the language of this paper more precisely. Let each $T_i = \Theta$ and hence a type of agent i will be written as $t_i = \theta^i \in \Theta$, where $\theta^i = \left(\theta_1^i, ..., \theta_I^i\right)$. We also write θ_{-i}^i for the vector θ^i excluding θ_i^i. We assume that $\widehat{\theta}_i\left(\theta^i\right) = \theta_i^i$ and $\widehat{\pi}_i$ satisfies

$$\sum_{\left\{t_{-i}:\widehat{\theta}_{-i}(t_{-i})=\theta_{-i}^i\right\}} \widehat{\pi}_i\left(\theta^i\right)[t_{-i}] = 1. \tag{5.10}$$

Thus we require that for each $\theta = (\theta_i, \theta_{-i})$, there is a type of agent i who has payoff type θ_i and assigns probability 1 to his opponents' having payoff type profile θ_{-i}. The complete information type space is $T = \times_{i=1}^I T_i = \left[\times_{i=1}^I \Theta_i\right]^I$.

Proposition 5.5 (Equivalence with Budget Balance: $I = 2$).
If $I = 2$, then F_ξ is ex post implementable if and only if F_ξ is interim implementable on all common prior type spaces.

Since the equivalence holds for all common prior type spaces, it must also hold for all type spaces. We will later discuss how example 3 shows that we do not also have equivalence with interim implementation on all naive type spaces.
PROOF: Clearly, if F_ξ^* is ex post implementable then F_ξ^* is interim implementable for all common prior type spaces. For the other direction, we argue by contrapositive. Suppose there does not exist an ex post implementation, then we can find a solution to the dual program with the multipliers λ and ν. By Lemma 5.4 a solution for the ex post dual has the gross flow property. Consider then the interim problem for the complete information type space $T_1 = T_2 = \Theta$. We seek to find multipliers $\lambda_i^* : \Theta \times \Theta \to \mathbb{R}_+$ and $\nu^* : \Theta \to \mathbb{R}$ such that the interim dual conditions (5.8) and (5.9) are met as well. For the purpose of this proof it will be convenient to have a running counter as superscript for the payoff states of agent i as follows:

$$\theta_i \in \Theta_i = \left\{\theta_i^1, \theta_i^2, .., \theta_i^{k_i}, ..., \theta_i^{K_i}\right\}.$$

By the gross flow equalization, we have for all $\theta = (\theta_i, \theta_j)$:

$$\sum_{\theta_i^{k_i} \in \Theta_i} \lambda_i\left(\theta_i^{k_i}, \theta_i, \theta_j\right) = \sum_{\theta_j^{k_j} \in \Theta_j} \lambda_j\left(\theta_j^{k_j}, \theta_j, \theta_i\right).$$

We can therefore always find, say by the greedy algorithm, for all θ an assignment:

$$\lambda_i^*\left(\theta_i^{k_i}\theta_j, \theta_i\theta_j^{k_j}\right) \in \mathbb{R}_+,\ \lambda_j^*\left(\theta_j^{k_j}\theta_i, \theta_j\theta_i^{k_i}\right) \in \mathbb{R}_+$$

such that for all $\theta_i^{k_i}$ and $\theta_j^{k_j}$, we have

$$\sum_{\theta_j^{k_j} \in \Theta_j} \lambda_i^*\left(\theta_i^k\theta_j, \theta_i\theta_j^{k_j}\right) = \lambda_i\left(\theta_i^{k_i}, \theta_i, \theta_j\right), \tag{5.11}$$

and

$$\sum_{\theta_i^{k_i}\in\Theta_i} \lambda_j^*\left(\theta_j^k\theta_i,\theta_j\theta_i^{k_i}\right)=\lambda_j\left(\theta_j^{k_j},\theta_j,\theta_i\right).$$

We next verify that the suggested definition of λ^* permits us to construct a solution for the equalities (5.8) and the inequality (5.9). Consider first the equalities. Here it is useful to distinguish between complete information types t, where $t_i = t_j$, and those t where $t_i \neq t_j$. Starting with the former, we have

$$\nu^*(t)=\sum_{t_i'\in T_i}\lambda_i^*\left(t_i,t_i'\right)$$

as $\widehat{\pi}_i\left(t_j|\,t_i\right)=1$ and $\widehat{\pi}_i\left(t_j'\middle|\,t_i\right)=0$. Writing it out more explicitly, we have

$$\nu^*\left(\theta_i\theta_j,\theta_i\theta_j\right)=\sum_{\theta_i^{k_i}\in\Theta_i}\sum_{\theta_j^{k_j}\in\Theta_j}\lambda_i^*\left(\theta_i\theta_j,\theta_i^{k_i}\theta_j^{k_j}\right),\text{ for all } i \text{ and } j, \tag{5.12}$$

and using (5.11), we know that

$$\sum_{\theta_j^{k_j}\in\Theta_j}\lambda_i^*\left(\theta_i\theta_j,\theta_i^{k_i}\theta_j^{k_j}\right)=\lambda_i\left(\theta_i,\theta_i^{k_i},\theta_j\right), \tag{5.13}$$

and hence (5.12) can be written as

$$\nu^*\left(\theta_i\theta_j,\theta_i\theta_j\right)=\sum_{\theta_i^{k_i}\in\Theta_i}\lambda_i\left(\theta_i,\theta_i^{k_i},\theta_j\right).$$

But as the gross flows are equalized in all states it follows from the existence of the ex post dual that we can find $\nu^*\left(\theta_i\theta_j,\theta_i\theta_j\right)$ such that the equality holds for i and j. Consider next types t with $t_i \neq t_j$. Here we have

$$\nu^*(t)=-\lambda_i^*\left(t_i',t_i\right),\text{ for all } i \tag{5.14}$$

as $\widehat{\pi}_i\left(t_j|\,t_i\right)=0$. By construction of the prior there exists only one type t_i', namely $t_i' = t_j$ such that $\widehat{\pi}_i\left(t_j|\,t_i'\right)>0$ and of course in this case it is $\widehat{\pi}_i\left(t_j|\,t_i'\right)=1$. Writing (5.14) more explicitly, we have

$$\nu^*\left(\theta_i^{k_i}\theta_j^{k_j},\theta_i\theta_j\right)=-\lambda_i^*\left(\theta_i\theta_j,\theta_i^{k_i}\theta_j^{k_j}\right),\text{ for all } i.$$

Finally consider the inequality (5.9) which we can write using the complete information types as follows:

$$\sum_{i=1}^{2}\sum_{t\in T}\sum_{t_i'\in T_i}\lambda_i^*\left(t_i,t_i'\right)\widehat{\pi}_i\left(t_{-i}|\,t_i\right)\zeta_i\left(\widehat{\theta}_i\left(t_i\right),\widehat{\theta}_i\left(t_i'\right),\widehat{\theta}_{-i}\left(t_{-i}\right)\right)>0,$$

or even simpler as

$$\sum_{i=1}^{2}\sum_{\theta\in\Theta}\sum_{t_i'\in T_i}\lambda_i^*\left(\theta,t_i'\right)\zeta_i\left(\theta_i,\widehat{\theta}_i\left(t_i'\right),\theta_{-i}\right)>0.$$

Finally using (5.13), we can replace $\sum_{t_i'\in T_i}\lambda_i\left(\theta,t_i'\right)$ by $\sum_{\theta_i^{k_i}\in\Theta_i}\lambda_i\left(\theta_i,\theta_i^{k_i},\theta_j\right)$, which leaves us precisely with the ex post inequality:

$$\sum_{i=1}^{2}\sum_{\theta\in\Theta}\sum_{\theta_i^{k_i}\in\Theta_i}\lambda_i\left(\theta_i,\theta_i^{k_i},\theta_j\right)\zeta_i\left(\theta_i,\theta_i^{k_i},\theta_{-i}\right)>0,$$

and this completes the proof. ■

The equivalence argument based on the complete information type space fails to extend to $I>2$ due to the very logic of the complete information type space. Suppose we were to construct an interim dual on the basis of the complete information types. Then there would have to exist some type pair of agent i, t_i, t_i', such that $\lambda_i\left(t_i,t_i'\right)>0$. Consider now the interim flow condition which is given by:

$$v\left(t\right)=\sum_{t_i'\in T_i}\lambda_i\left(t_i,t_i'\right)\widehat{\pi}_i\left(t_{-i}\left|t_i\right.\right)-\sum_{t_i'\in T_i}\lambda_i\left(t_i',t_i\right)\widehat{\pi}_i\left(t_{-i}\left|t_i'\right.\right).$$

Consider now a specific type profile (t_i',t_{-i}) with $t_i=\theta'$ and $t_i=t_j=\theta$ for all $j\neq i$. By definition of the complete information types, it follows that for all j and for all t_j,t_j':

$$\widehat{\pi}_j\left(t_i',t_{-ij}\left|t_j\right.\right)=\widehat{\pi}_j\left(t_i',t_{-ij}\left|t_j'\right.\right)=0.$$

This is simply because every type t_j places probability 1 on all other agents to have exactly the same type. However as only agent i has at (t_i',t_{-i}) a type different from all other agents, there exists a type $t_i=\theta$ such that $\widehat{\pi}_i\left(t_{-i}\left|t_i\right.\right)=1$. It follows that the flow condition for agent i at (t_i',t_{-i}) reads:

$$\nu\left(t_i',t_{-i}\right)=-\lambda_i\left(t_i,t_i'\right)\widehat{\pi}_i\left(t_{-i}\left|t_i\right.\right)=-\lambda_i\left(t_i,t_i'\right)<0.$$

In contrast for all $j \neq i$, it follows from the complete information type assumption that the flow condition reduces to

$$\nu\left(t_i', t_{-i}\right) = 0,$$

which obviously cannot constitute a solution to the dual.

This problem does not arise when the cardinality of the space of payoff types for each agent is at most two.

Proposition 5.6 (Equivalence with Budget Balance: $\#\Theta_i \leq 2$).
If $\#\Theta_i = 1$ for some i or if $\#\Theta_i \leq 2$ for all i, then F_ξ is ex post implementable if and only if F_ξ is interim implementable on all common prior naive type spaces.

As always, we immediately also have equivalence with interim implementability on larger type spaces.

PROOF: Clearly, if F_ξ is ex post implementable then F_ξ is interim implementable for all naive type spaces. For the other direction, consider first the case of $\#\Theta_i = 1$. By hypothesis F_ξ is interim implementable on all naive type spaces, and thus a fortiori F_ξ is interim implementable as well. By Proposition 5.2 it follows that F_ξ is ex-post implementable. We then consider the ex post implementation solution of F_ξ without a balanced budget and modify if necessary the transfers of agent i with $\#\Theta_i = 1$, say it is $i = 1$, to achieve budget balance. It suffices to have the transfer of agent 1 absorb the deficit or surplus of the remaining agents by letting

$$y_1(\theta) \triangleq -\sum_{i=2}^{I} y_i(\theta).$$

By hypothesis there are no incentive constraints to respect for agent 1 and such a modification is always feasible.

For the case of $\#\Theta_i = 2$ for all i, we shall analyze the implementation problem in its dual version and argue again by contrapositive. Suppose that ex post implementation is not feasible and by Lemma 5.3 there exist a solution (v, λ) to the ex post flow and weighting conditions. We find appropriate values for the interim dual variables, denoted by (v^*, λ^*), together with a prior p on the naive type space, i.e. $T_i = \Theta_i$ to obtain a solution to the interim dual (5.8) and (5.9). Thus fix the ex post variables (v, λ) and define a prior $p(\theta)$ through them:

$$p(\theta) \triangleq \frac{\sum_{i=1}^{I} \lambda_i\left(\theta_i, \theta_i', \theta_{-i}\right)}{\sum_{\theta\in\Theta}\sum_{i=1}^{I} \lambda_i\left(\theta_i, \theta_i', \theta_{-i}\right)}.$$

Since $\#\Theta_i = 2$, θ_i' is uniquely defined given θ_i and $\theta_i' \neq \theta_i$. By hypothesis, the multipliers

$\lambda_i(\cdot)$ constitute a solution to the ex post dual problem and hence

$$\sum_{\theta\in\Theta}\sum_{i=1}^{I}\lambda_i\left(\theta_i,\theta_i',\theta_{-i}\right)>0,$$

which guarantees that the prior $p(\cdot)$ is well defined. The associated interim belief is given by

$$p_i\left(\theta_{-i}\left|\theta_i\right.\right)\triangleq\frac{\lambda_i\left(\theta_i,\theta_i',\theta_{-i}\right)}{\sum_{\theta'_{-i}\in\Theta_{-i}}\lambda_i\left(\theta_i,\theta_i',\theta'_{-i}\right)},\tag{5.15}$$

provided that

$$\sum_{\theta'_{-i}\in\Theta_{-i}}\lambda_i\left(\theta_i,\theta_i',\theta'_{-i}\right)>0.$$

Otherwise we can, without loss of generality, set

$$p_i\left(\theta_{-i}\left|\theta_i\right.\right)\triangleq\frac{1}{2^{I-1}}.$$

Let

$$\lambda_i^*\left(\theta_i,\theta_i'\right)\triangleq\sum_{\theta'_{-i}\in\Theta_{-i}}\lambda_i\left(\theta_i,\theta_i',\theta'_{-i}\right)\tag{5.16}$$

and let

$$v^*(\theta)\triangleq v(\theta),\tag{5.17}$$

for all i,θ_i,θ_i' and θ. By inserting the rhs of (5.15)-(5.17) into (5.8)-(5.9), we get

$$\begin{aligned}\nu(\theta) &= \sum_{\theta_i'\in\Theta_i}\sum_{\theta'_{-i}\in\Theta_{-i}}\lambda_i\left(\theta_i,\theta_i',\theta'_{-i}\right)\times\frac{\lambda_i\left(\theta_i,\theta_i',\theta_{-i}\right)}{\sum_{\theta'_{-i}\in\Theta_{-i}}\lambda_i\left(\theta_i,\theta_i',\theta'_{-i}\right)}\\ &\quad-\sum_{\theta_i'\in\Theta_i}\sum_{\theta'_{-i}\in\Theta_{-i}}\lambda_i\left(\theta_i',\theta_i,\theta'_{-i}\right)\times\frac{\lambda_i\left(\theta_i',\theta_i,\theta_{-i}\right)}{\sum_{\theta'_{-i}\in\Theta_{-i}}\lambda_i\left(\theta_i',\theta_i,\theta'_{-i}\right)}\end{aligned}\tag{5.18}$$

and

$$\sum_{i=1}^{I}\sum_{\theta\in\Theta}\sum_{\theta_i'\in\Theta_i}\sum_{\theta'_{-i}\in\Theta_{-i}}\lambda_i\left(\theta_i,\theta_i',\theta'_{-i}\right)\zeta_i\left(\theta_i,\theta_i',\theta_{-i}\right)\frac{\lambda_i\left(\theta_i,\theta_i',\theta_{-i}\right)}{\sum_{\theta'_{-i}\in\Theta_{-i}}\lambda_i\left(\theta_i,\theta_i',\theta'_{-i}\right)}>0,\tag{5.19}$$

for all i,θ_i,θ_i' and θ. After eliminating the denominator in (5.18) and (5.19), respectively, through the obvious cancellations, the above interim dual conditions coincide with the ex post dual conditions (5.6) and (5.7). ■

The argument of the proof suggests an algorithm for constructing a prior over the naive type space. The prior displays correlation, but agent i with payoff type θ_i now assigns equal probability to 2^{I-1} possible payoff profiles of the remaining agent. The derived posterior has the further property that for every type profile θ, there exist some θ_i' such that $p\left(\theta_{-i}\left|\theta_i'\right.\right) > 0$ and thus evades the problem which arose in the complete information type space with many agents.

We conclude this section by discussing two examples illustrating the role of the assumptions in Proposition 5.5 and 5.6.

Recall that in Example 3, ex post implementation was not possible, interim implementation on all naive type spaces was possible, but interim implementation on all common prior type spaces was not possible. Recall that the naming of the allocation represented a fixed transfer of utility from one agent to the other agent. If arbitrary transfers of utility were allowed instead, the results would be unchanged: the transfers turn out to irrelevant in establishing the impossibility results and the positive results automatically go through if more instruments are allowed. This expanded version of Example 3 establishes that it is not possible to strengthen Proposition 5.5 to show the equivalence of ex post implementation and interim implementation on naive type spaces.

We already saw in Examples 1 and 2 examples where ex post implementation is impossible but interim implementation is possible on all type spaces. In order to establish the same conclusion in a quasi-linear environment, Propositions 5.5 and 5.6 show that the example must be somewhat complicated. Example 4 below represents a minimal departure from the assumptions in Propositions 5.5 and 5.6. It features three agents in which the first agent has a payoff type space of cardinality three, whereas the remaining two agents have binary payoff state spaces. We show in this example that ex post implementation is not possible but that interim implementation is possible on any type space. The example points to a failure of the equivalence results with a minimal relaxation of either one of the two distinct sets of sufficient conditions.

EXAMPLE 4. There are three agents. Agent 1 has three possible types:

$$\Theta_1 = \left\{\theta_1^+, \theta_1^0, \theta_1^-\right\},$$

while agents 2 and 3 each have two possible types,

$$\Theta_2 = \left\{\theta_2^+, \theta_2^-\right\} \quad \text{and} \quad \Theta_3 = \left\{\theta_3^+, \theta_3^-\right\}.$$

The social choice function is defined over three allocations $A = \{a, b, c\}$ and represented in the following table:

θ_3^+	θ_2^+	θ_2^-
θ_1^+	a	c
θ_1^0	b	a
θ_1^-	c	b

θ_3^-	θ_2^+	θ_2^-
θ_1^+	c	b
θ_1^0	a	c
θ_1^-	b	a

.

The gross payoffs of the agents from allocation a are given by:

θ_3^+	θ_2^+	θ_2^-
θ_1^+	$0, 0, 0$	$-1, -1, -1$
θ_1^0	$-1, -1, -1$	$0, 0, 0$
θ_1^-	$-1, -1, -1$	$-1, -1, \varepsilon$

θ_3^-	θ_2^+	θ_2^-
θ_1^+	$-1, -1, -1$	$\varepsilon, -1, -1$
θ_1^0	$0, 0, 0$	$-1, -1, -1$
θ_1^-	$-1, \varepsilon, -1$	$0, 0, 0$

from allocation b by:

θ_3	θ_2^+	θ_2^-
θ_1^+	$\varepsilon, -1, -1$	$-1, -1, -1$
θ_1^0	$0, 0, 0$	$-1, -1, -1$
θ_1^-	$-1, -1, -1$	$0, 0, 0$

θ_3	θ_2^+	θ_2^-
θ_1^+	$-1, -1, -1$	$0, 0, 0$
θ_1^0	$-1, -1, -1$	$-1, -1, -1$
θ_1^-	$0, 0, 0$	$-1, -1, -1$

and from allocation c by:

θ_3	θ_2^+	θ_2^-
θ_1^+	$-1, \varepsilon, \varepsilon$	$0, 0, 0$
θ_1^0	$\varepsilon, -1, -1$	$-1, -1, -1$
θ_1^-	$0, 0, 0$	$\varepsilon, \varepsilon, -1$

θ_3	θ_2^+	θ_2^-
θ_1^+	$0, 0, 0$	$-1, \varepsilon, \varepsilon$
θ_1^0	$-1, -1, -1$	$0, 0, 0$
θ_1^-	$\varepsilon, -1, \varepsilon$	$-1, -1, -1$

where $\varepsilon > 0$. This example has the feature that the social choice function and the payoffs are symmetric with respect to all the agents on the restricted domain:

$$\left\{\theta_1^+, \theta_1^-\right\} \times \left\{\theta_2^+, \theta_2^-\right\} \times \left\{\theta_3^+, \theta_3^-\right\}.$$

Moreover, social choice function and payoffs remain symmetric for agent 2 and 3 on the entire domain. In the absence of monetary transfer (i.e. zero transfers in all states) each agent i has exactly one profitable ex post deviation at every state θ_{-i}: either by mis-reporting θ_i^+ instead of truth-telling θ_i^- or mis-reporting θ_i^- instead of truthtelling θ_i^+. The only exception arises for agent 1 in the state $\theta_2^+\theta_3^+$, where it is not ex post profitable to misreport θ_1^- instead of truthtelling θ_1^+, but where it is profitable to misreport θ_1^0 instead of truthtelling θ_1^+ and to misreport θ_1^- instead of truthtelling θ_1^0.

The example on the symmetric and restricted domain does not satisfy ex post implementability and by Proposition 5.6 it then also fails to be interim implementable. The introduction of the additional state θ_1^0 together with the replacement of the profitable ex post deviation in the state $(\theta_2, \theta_3) = \left(\theta_2^+, \theta_3^+\right)$ from θ_1^+ to θ_1^- by the two profitable deviations θ_1^+ to θ_1^0 and from θ_1 to θ_1^0 leaves ex post implementation impossible, but opens the possibility for interim implementation.

We first verify the failure of ex post implementation by considering all states θ where the gains from deviating in terms of the gross payoffs are ε. The transfers $y_i(\theta)$ in these thirteen ex post incentive constraints then have to satisfy for agent 1 :

$$\begin{aligned}
y_1\left(\theta_1^+,\theta_2^+,\theta_3^+\right)-y_1\left(\theta_1^0,\theta_2^+,\theta_3^+\right)&\geq\varepsilon\\
y_1\left(\theta_1^0,\theta_2^+,\theta_3^+\right)-y_1\left(\theta_1^-,\theta_2^+,\theta_3^+\right)&\geq\varepsilon\\
y_1\left(\theta_1^+,\theta_2^-,\theta_3^-\right)-y_1\left(\theta_1^-,\theta_2^-,\theta_3^-\right)&\geq\varepsilon\\
y_1\left(\theta_1^-,\theta_2^+,\theta_3^-\right)-y_1\left(\theta_1^+,\theta_2^+,\theta_3^-\right)&\geq\varepsilon\\
y_1\left(\theta_1^-,\theta_2^-,\theta_3^+\right)-y_1\left(\theta_1^+,\theta_2^-,\theta_3^+\right)&\geq\varepsilon
\end{aligned}$$

and for agent 2 :

$$\begin{aligned}
y_2\left(\theta_1^+,\theta_2^+,\theta_3^+\right)-y_2\left(\theta_1^+,\theta_2^-,\theta_3^+\right)&\geq\varepsilon\\
y_2\left(\theta_1^-,\theta_2^+,\theta_3^-\right)-y_2\left(\theta_1^-,\theta_2^-,\theta_3^-\right)&\geq\varepsilon\\
y_2\left(\theta_1^-,\theta_2^-,\theta_3^+\right)-y_2\left(\theta_1^-,\theta_2^+,\theta_3^+\right)&\geq\varepsilon\\
y_2\left(\theta_1^+,\theta_2^-,\theta_3^-\right)-y_2\left(\theta_1^+,\theta_2^+,\theta_3^-\right)&\geq\varepsilon
\end{aligned}$$

and for agent 3 :

$$\begin{aligned}
y_3\left(\theta_1^+,\theta_2^+,\theta_3^+\right)-y_3\left(\theta_1^+,\theta_2^+,\theta_3^-\right)&\geq\varepsilon\\
y_3\left(\theta_1^-,\theta_2^-,\theta_3^+\right)-y_3\left(\theta_1^-,\theta_2^-,\theta_3^-\right)&\geq\varepsilon\\
y_3\left(\theta_1^+,\theta_2^-,\theta_3^-\right)-y_3\left(\theta_1^+,\theta_2^-,\theta_3^+\right)&\geq\varepsilon\\
y_3\left(\theta_1^-,\theta_2^+,\theta_3^-\right)-y_3\left(\theta_1^-,\theta_2^+,\theta_3^+\right)&\geq\varepsilon
\end{aligned}$$

We can now appeal to the balanced budget condition and observe that the sum of the left hand sides in the above inequalities is zero. This obviously leads to a contradiction with the right hand side which is positive. Thus ex post implementation is impossible.

We now show that interim implementation is possible on any type space. Consider the following indirect mechanism in which agent one sends a message besides his payoff type. Thus the message spaces are defined by $M_1=\Theta_1\times\{0,1\}$, with typical element $m_1=(\theta_1,\alpha)$; $M_2=\Theta_2$; and $M_3=\Theta_3$. If message profile $((\theta_1,\alpha),\theta_2,\theta_3)$ is sent, then allocation $\xi(\theta_1,\theta_2,\theta_3)$ is chosen and the following budget balance transfers are implemented:

$m_3=\theta_3^+$	θ_2^+	θ_2^-
$\left(\theta_1^+,0\right)$	$-6c,3c,3c$	$0,2c,-2c$
$\left(\theta_1^-,0\right)$	$6c,-3c,-3c$	$c,-2c,c$
$\left(\theta_1^0,0\right)$	$0,0,0$	$0,0,0$
$\left(\theta_1^+,1\right)$	$2b+2c,-b-c,-b-c$	$-a+b+2c,-b-2c,a$
$\left(\theta_1^-,1\right)$	$2b-2c,-b+c,-b+c$	$-a+b-3c,-b+2c,a+c$
$\left(\theta_1^0,1\right)$	$2b,-b,-b$	$-a+b,-b,a$

(5.20)

and

$m_3 = \theta_3^-$	θ_2^+	θ_2^-
$\left(\theta_1^+, 0\right)$	$0, -2c, 2c$	$2c, -c, -c$
$\left(\theta_1^-, 0\right)$	$c, c, -2c$	$0, 0, 0$
$\left(\theta_1^0, 0\right)$	$0, 0, 0$	$0, 0, 0$
$\left(\theta_1^+, 1\right)$	$-a + b + 2c, a, -b - 2c$	$-2a - 2c, a + c, a + c$
$\left(\theta_1^-, 1\right)$	$-a + b - 3c, a + c, -b + 2c$	$-2a, a, a$
$\left(\theta_1^0, 1\right)$	$-a + b, a, -b$	$-2a, a, a$

(5.21)

We assume that the parameters of the transfer payments satisfy the following conditions:

$$a \gg b \gg 1 \gg c > \varepsilon > 0. \tag{5.22}$$

Call the resulting mechanism $\mathcal{G}$ and fix any type space $\mathcal{T}$. We will argue that there exists an equilibrium of the incomplete information game $(\mathcal{G},\mathcal{T})$ of the following form:

$$s_1(t_1) = \begin{cases} \left(\widehat{\theta}_1(t_1), 0\right), \text{ if } \widehat{\psi}_1(t_1) \in \widehat{\Delta} \\ \left(\widehat{\theta}_1(t_1), 1\right), \text{ if } \widehat{\psi}_1(t_1) \notin \widehat{\Delta} \end{cases}$$

$$s_2(t_2) = \widehat{\theta}_2(t_2)$$

$$s_3(t_3) = \widehat{\theta}_3(t_3)$$

where $\widehat{\Delta}$ is a subset of $\Delta(\Theta_2 \times \Theta_3)$. We recall that $\widehat{\psi}_i(t_i)$ was defined in (2.1) as the belief of type t_i of agent i over the payoff types of the remaining agents.

The transfer payments suggested in the tables (5.20) and (5.21) together with the restrictions (5.22) have the following incentive properties. The incentive constraints for agents 2 and 3 are satisfied ex-post for all types and all reports by agent 1. In contrast, the ex-post incentive constraints for agent 1 are satisfied only partially. For $\alpha = 0$, the ex-post incentive constraints are satisfied for all $(\theta_2, \theta_3) \neq \left(\theta_2^+, \theta_3^+\right)$, and by contrast for $\alpha = 1$, the ex-post constraints are only satisfied at $(\theta_2, \theta_3) = \left(\theta_2^+, \theta_3^+\right)$. We now use these local properties to guarantee interim incentive compatibility for agent 1 by inducing him to choose α appropriately. The large rewards and penalties offered through a and b will induce him to the send the message $\alpha = 1$ if and only if he assigns a sufficiently high probability to the event that the payoff type of the other agents is $(\theta_2, \theta_3) = \left(\theta_2^+, \theta_3^+\right)$, otherwise he will have a preference to report $\alpha = 0$ in order to avoid the penalties imposed on all states $(\theta_2, \theta_3) \neq \left(\theta_2^+, \theta_3^+\right)$.

We begin with the ex-post incentive constraints for agent 2, which are given by:

$$y_2((\theta_1, \alpha), \theta_2, \theta_3) - y_2\left((\theta_1, \alpha), \theta_2', \theta_3\right) \geq \zeta_2\left(\theta_2, \theta_2', \theta_{-2}\right) \tag{5.23}$$

We first observe that the differential transfer implications of mis-reporting for agent 2, namely

$$y_2((\theta_1, \alpha), \theta_2, \theta_3) - y_2\left((\theta_1, \alpha), \theta_2', \theta_3\right), \quad \theta_2 \neq \theta_2'$$

given truthtelling by the other agents, are as follows:

θ_3^+	θ_2^+	θ_2^-
$(\theta_1^+, 0)$	c	$-c$
$(\theta_1^-, 0)$	$-c$	c
$(\theta_1^0, 0)$	0	0
$(\theta_1^+, 1)$	c	$-c$
$(\theta_1^-, 1)$	$-c$	c
$(\theta_1^0, 1)$	0	0

θ_3^-	θ_2^+	θ_2^-
$(\theta_1^+, 0)$	$-c$	c
$(\theta_1^-, 0)$	c	$-c$
$(\theta_1^0, 0)$	0	0
$(\theta_1^+, 1)$	$-c$	c
$(\theta_1^-, 1)$	c	$-c$
$(\theta_1^0, 1)$	0	0

As the message space for agent 2 contains only two elements, the entry in an arbitrary cell, say $\left(\left(\theta_1^+, 0\right), \theta_2^+, \theta_3^+\right)$, represents the corresponding difference in transfers:

$$y_2\left(\left(\theta_1^+, 0\right), \theta_2^+, \theta_3^+\right) - y_2\left(\left(\theta_1^+, 0\right), \theta_2^-, \theta_3^+\right),$$

as θ_2^- is the only message different from θ_2^+. The differential utility from misreporting, $\zeta_2\left(\theta_2, \theta_2', \theta_{-2}\right)$, is represented in the next table. By the binary nature of the state space, the true state identifies uniquely the misreported type, and the entry in an arbitrary cell, say $\left(\left(\theta_1^+, 0\right), \theta_2^+, \theta_3^+\right)$, represents $\zeta_2\left(\theta_2^+, \theta_2^-, \theta_1^+\theta_3^+\right)$:

$\theta_3 = \theta_3^+$	θ_2^+	θ_2^-
θ_1^+	ε	-1
θ_1^-	-1	ε
θ_1^0	-1	-1

$\theta_3 = \theta_3^-$	θ_2^+	θ_2^-
θ_1^+	-1	ε
θ_1^-	ε	-1
θ_1^0	-1	-1

Since $\varepsilon < c \ll 1$, it follows that the transfer gain from telling the truth always outweighs (ex post) the allocative gain from mis-reporting. As the ex post incentive constraints are satisfied, it is a fortiori true for the interim constraints. The suggested transfers maintain symmetry between agent 2 and 3 and hence the ex post incentive constraints are satisfied for agent 3 as well.

We next verify the interim incentive conditions for agent 1. We do this in three steps. First we identify sufficient conditions for agent 1 to report either $\alpha = 0$ or $\alpha = 1$. Second, we examine the truthtelling conditions for agent 1 given his choice of α. Third, we find joint conditions on a, b, c and ε such that truthtelling is satisfied after removing the conditioning on the choice of α.

In the first step, we assume that 1, c and ε are all infinitesimal compared with a and b. Suppose that agent 1 assigns probability p to $\left(\theta_2^+, \theta_3^+\right)$. Ignoring infinitesimal terms and assuming that all other agents report truthfully, we observe that agent's payoff to setting $\alpha = 1$ is at least

$$p\,(2b) + (1-p)\,(-2a) = 2p\,(a+b) - 2a$$

and is at most

$$p\,(2b)+(1-p)\,(-a+b)=p\,(a+b)-(a-b)\,,$$

where the bounds are obtained by varying the remaining probability $1-p$ over the other three possible states, $(\theta_2,\theta_3)\neq\left(\theta_2^+,\theta_3^+\right)$. Thus if

$$p>\frac{a}{a+b}$$

agent 1 will certainly set $\alpha=1$. If

$$p<\frac{a-b}{a+b}$$

agent 1 will certainly set $\alpha=0$.

In second step, we fix the choice of α and ask whether agent 1 will truthfully report his payoff type. Thus suppose that agent 1 sets $\alpha=0$. The only ex post profitable misrepresentations are for type θ_1^+ to claim to be θ_1^0 and for type θ_1^0 to claim to be θ_1^-, when $(\theta_2,\theta_3)=\left(\theta_2^+,\theta_3^+\right)$. In each case the allocative gain to lying is ε and otherwise it is -1. The gain in additional transfers from lying is at most $6c$ across all states, and thus a sufficient condition for no mis-reporting is

$$(1-p)\,(1-6c)-p\,(\varepsilon+6c)\geq 0\Leftrightarrow p\leq\frac{1-6c}{1+\varepsilon}.$$

Finally, suppose that agent 1 will set $\alpha=1$. We recall that the ex-post constraints of agent 1 are now satisfied in the state $(\theta_2,\theta_3)=\left(\theta_2^+,\theta_3^+\right)$ and hence the interim constraints will be satisfied for p in the neighborhood of 1. In particular, the gain from telling the truth, when $(\theta_2,\theta_3)=\left(\theta_2^+,\theta_3^+\right)$, is at least $2c-\varepsilon$. On the other hand, the gain from lying when $(\theta_2,\theta_3)\neq\left(\theta_2^+,\theta_3^+\right)$ is at most $\varepsilon+5c$. So interim constraints are satisfied as long as

$$p\,(2c-\varepsilon)-(1-p)\,(\varepsilon+5c)\geq 0\ \Leftrightarrow p\geq\frac{\varepsilon+5c}{7c}.$$

We can now choose a and b sufficiently large and c and ε sufficiently small (say $\varepsilon=\frac{1}{100}$, $c=\frac{1}{90}$, $b=9$, $a=100$) such that

$$\frac{\varepsilon+5c}{7c}<\frac{a-b}{a+b}<\frac{a}{a+b}<\frac{1-6c}{1+\varepsilon},\tag{5.24}$$

while preserving the inequalities (5.22). The inequalities in (5.24) allow us conclude that whenever $\alpha=0$ or $\alpha=1$ could conceivably be a best response, then reporting the payoff type truthfully is a best response as well. To see this, observe that if agent 1 ever sets $\alpha=0$, it must be that

$$p<\frac{a}{a+b},$$

but by (5.24), in this case we will also have

$$p < \frac{1-6c}{1+\varepsilon},$$

so he will truthfully report his payoff type. On the other hand, if agent 1 ever sets $\alpha = 1$, we must have

$$p > \frac{a-b}{a+b};$$

but again by (5.24), we will also have

$$p > \frac{\varepsilon+5c}{7c},$$

and thus agent 1 will again truthfully report his payoff type.

6. Interim Implementability on Arbitrary Type Spaces

In the special but important case of the quasi-linear environment from the previous section (without budget balance constraints), we establish a precise characterization of when interim implementation is possible for an arbitrary (finite) type space. Given risk neutral types, it is always possible to identify an agent's belief type via a set of bets, as suggested by d'Aspremont and Gerard-Varet (1979) and Myerson (1981), for example using a quadratic objective function. So the implementation problem reduces to finding transfers to distinguish the payoff types that are possible for a given belief type. Thus the characterization of interim implementability reduces to the type of conditions that would arise looking at implementation with independent types on the naive type space. After our main result, we provide a discussion of these conditions and how reasonable they are.

Our results follow an argument of Neeman (2001) concerning a revenue maximizing seller of a single object. Neeman noted that the full surplus extraction result of Cremer and McLean (1985) breaks down if we make the reasonable assumption that many payoff types are consistent with a given belief type. However, in that setting, one cannot assume that even the belief type is fully extracted, since there may be rents associated with beliefs of linearly dependent types (see Parreiras (2002)). However, when a planner does not care about transfers, belief types can be fully extracted giving an exact characterization.

To state our results, we use some conditions for implementation on the naive type space. Suppose it was common knowledge that agent i believed that others' payoff types were drawn according to distribution $\psi_i \in \Delta(\Theta_{-i})$. We define

$$\widehat{v}_i\left(\theta_i, \theta_i', \psi_i\right) \triangleq \sum_{\theta_{-i} \in \Theta_{-i}} \psi_i\left(\theta_{-i}\right) v_i\left(\xi\left(\theta_i', \theta_{-i}\right), \left(\theta_i, \theta_{-i}\right)\right),$$

as agent i's expected utility from allocation rule ξ given that he has beliefs ψ_i over his opponent's payoff type, he has payoff type θ_i, he reports himself to have payoff type θ_i', and he expects the social planner to behave as if there is truth-telling. Suppose further that the planner knew that agent i's type was an element of $\overline{\Theta_i} \subseteq \Theta_i$. The incentive constraints for individual i in this case become:

Definition 6.1. *Types $\overline{\Theta_i}$ are said to be ψ_i-*incentive compatible *if there exists $y : \overline{\Theta_i} \to \mathbb{R}$ such that*

$$\widehat{v}_i\left(\theta_i, \theta_i, \psi_i\right) + y\left(\theta_i\right) \geq \widehat{v}_i\left(\theta_i, \theta_i', \psi_i\right) + y\left(\theta_i'\right)$$

for all θ_i, $\theta_i' \in \overline{\Theta_i}$.

Thus if there were two agents with independent beliefs (ψ_1, ψ_2) over each others' types, a necessary and sufficient condition for interim implementation in this case would be that types Θ_i are ψ_i-incentive compatible. We will show that interim implementation on an arbitrary type space reduces to a collection of such conditions. For any beliefs π_i over other agents' types, the corresponding beliefs over other agents' payoff types are:

$$\widetilde{\psi}_i\left(\pi_i\right)\left[\theta_{-i}\right] \triangleq \sum_{\left\{t_{-i} : \widehat{\theta}_{-i}\left(t_{-i}\right) = \theta_{-i}\right\}} \widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right].$$

Now let Π_i^* be the collection of all possible belief types (i.e., the range of $\widehat{\pi}_i(\cdot)$); and let $\widetilde{\Theta}_i\left(\pi_i\right)$ be the collection of payoff types consistent with π_i, i.e.,

$$\widetilde{\Theta}_i\left(\pi_i\right) \triangleq \left\{\theta_i : \widehat{\pi}_i\left(t_i\right) = \pi_i \text{ and } \widehat{\theta}_i\left(t_i\right) = \theta_i \text{ for some } t_i\right\}.$$

Proposition 6.2 (Interim Implementation). *F_ξ is interim implementable if and only if for each i and $\pi_i \in \Pi_i^*$, types $\widetilde{\Theta}_i\left(\pi_i\right)$ are $\widetilde{\psi}_i\left(\pi_i\right)$- incentive compatible.*

Proof. (Necessity). Suppose that F_ξ is interim implementable and let $y_i : T \to \mathbb{R}$ be the transfer function for i in the direct truth-telling mechanism. Define $\widetilde{y}_i : T_i \times \Pi_i^* \to \mathbb{R}$ by

$$\widetilde{y}_i\left(t_i, \pi_i\right) \equiv \sum_{t_{-i} \in T_{-i}} \pi_i\left(t_{-i}\right) y_i\left(t_i, t_{-i}\right).$$

Thus $\widetilde{y}_i\left(t_i, \pi_i\right)$ is the expected transfer to agent i if he reports himself to be type t_i and his belief type is π_i. Now observe that the payoff to type t_i of agent i reporting himself to be type t_i' is:

$$\begin{aligned}
\zeta_i\left(t_i, t_i'\right) &\equiv \sum_{t_{-i} \in T_{-i}} \widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right]\left[v_i\left(\xi\left(\widehat{\theta}\left(t_i', t_{-i}\right)\right), \widehat{\theta}\left(t_i, t_{-i}\right)\right) + y_i\left(t_i', t_{-i}\right)\right], \\
&\equiv \widehat{v}_i\left(\widehat{\theta}_i\left(t_i\right), \widehat{\theta}_i\left(t_i'\right), \widetilde{\psi}_i\left(\widehat{\pi}_i\left(t_i\right)\right)\right) + \widetilde{y}_i\left(t_i', \widehat{\pi}_i\left(t_i\right)\right).
\end{aligned}$$

Now incentive compatibility requires that $\zeta_i(t_i, t_i) \geq \zeta_i(t_i, t_i')$ for all $t_i, t_i' \in T_i$. This implies that if $\widehat{\theta}_i(t_i) = \widehat{\theta}_i(t_i')$ and $\widehat{\pi}_i(t_i) = \widehat{\pi}_i(t_i') = \pi_i$, then

$$\widetilde{y}_i(t_i, \pi_i) = \widetilde{y}_i(t_i', \pi_i).$$

So $\widetilde{y}_i(t_i', \pi_i)$ can depend only on the belief and payoff types corresponding to t_i'. Writing $\widehat{y}_i(\theta_i', \pi_i', \pi_i)$ for the expected transfer to a type of agent i whose true belief over the opponent's type is π_i, but who reports himself to be a type with belief π_i' and payoff type θ_i', the incentive compatibility conditions become, for all $t_i, t_i' \in T_i$,

$$\begin{aligned} & v_i\left(\widehat{\theta}_i(t_i), \widehat{\theta}_i(t_i), \widetilde{\psi}_i(\widehat{\pi}_i(t_i))\right) + \widehat{y}_i\left(\widehat{\theta}_i(t_i), \widehat{\pi}_i(t_i), \widehat{\pi}_i(t_i)\right) \\ & \geq v_i\left(\widehat{\theta}_i(t_i), \widehat{\theta}_i(t_i'), \widetilde{\psi}_i(\widehat{\pi}_i(t_i))\right) + \widehat{y}_i\left(\widehat{\theta}_i(t_i'), \widehat{\pi}_i(t_i'), \widehat{\pi}_i(t_i)\right). \end{aligned}$$

Now if $\theta_i,\ \theta_i' \in \widetilde{\Theta}_i(\pi_i)$, this implies

$$\widehat{v}_i\left(\theta_i, \theta_i, \widetilde{\psi}_i(\pi_i)\right) + \widehat{y}_i(\theta_i, \pi_i, \pi_i) \geq \widehat{v}_i\left(\theta_i, \theta_i', \widetilde{\psi}_i(\pi_i)\right) + \widehat{y}_i(\theta_i', \pi_i, \pi_i).$$

So setting

$$y(\theta_i) = \widehat{y}_i(\theta_i, \pi_i, \pi_i),$$

we must have that types $\widetilde{\Theta}_i(\pi_i)$ are $\widetilde{\psi}_i(\pi_i)$- incentive compatible.

(Sufficiency) Suppose that there for each i and $\pi_i \in \Pi_i^*$, types $\widetilde{\Theta}_i(\pi_i)$ are $\widetilde{\psi}_i(\pi_i)$-incentive compatible. Then there exists $y_i^* : \Theta_i \times \Pi_i^* \to \mathbb{R}$ satisfying

$$\widehat{v}_i\left(\theta_i, \theta_i, \widetilde{\psi}_i(\pi_i)\right) + \widehat{y}_i(\theta_i, \pi_i,) \geq \widehat{v}_i\left(\theta_i, \theta_i', \widetilde{\psi}_i(\pi_i)\right) + \widehat{y}_i(\theta_i', \pi_i) \tag{6.1}$$

for all $\theta_i,\ \theta_i' \in \widetilde{\Theta}_i(\pi_i)$. Let

$$K = \max_{\theta_i, \theta_i' \in \widetilde{\Theta}_i(\pi_i),\ \pi_i \in \Pi_i^*} \left| \begin{array}{l} \widehat{v}_i\left(\theta_i, \theta_i, \widetilde{\psi}_i(\pi_i)\right) + y_i^*(\theta_i, \pi_i) \\ -\widehat{v}_i\left(\theta_i, \theta_i', \widetilde{\psi}_i(\pi_i)\right) - y_i^*(\theta_i', \pi_i') \end{array} \right|. \tag{6.2}$$

For each i, let

$$h_i(t_{-i}, \pi_i) = -\left[(1 - \pi_i(t_{-i}))^2 + \sum_{t_{-i}'} \left(\pi_i(t_{-i}')\right)^2\right].$$

This implies

$$\sum_{t_{-i} \in T_{-i}} \pi_i(t_{-i}) h_i(t_{-i}, \pi_i) - \sum_{t_{-i} \in T_{-i}} \pi_i(t_{-i}) h_i(t_{-i}, \pi_i') > 0, \tag{6.3}$$

for all $\pi_i' \neq \pi_i$. Let

$$\eta = \min_{i,\pi_i' \neq \pi_i} \sum_{t_{-i} \in T_{-i}} \pi_i (t_{-i}) \, h_i (t_{-i}, \pi_i) - \sum_{t_{-i} \in T_{-i}} \pi_i (t_{-i}) \, h_i \left(t_{-i}, \pi_i'\right), \tag{6.4}$$

Let

$$y_i (t) = y_i^* \left(\widehat{\theta}_i (t_i), \widehat{\pi}_i (t_i)\right) + \frac{K}{\eta} h_i (t_{-i}, \widehat{\pi}_i (t_i)).$$

Now if agent i is a type t_i with $\widehat{\theta}_i (t_i) = \theta_i$ and $\widehat{\pi}_i (t_i) = \pi_i$, and he reports himself to be a type t_i' with $\widehat{\theta}_i (t_i') = \theta_i'$ and $\widehat{\pi}_i (t_i') = \pi_i'$, his expected payoff is

$$\widehat{v}_i \left(\theta_i, \theta_i', \widehat{\psi}_i (\pi_i)\right) + y_i^* \left(\theta_i, \pi_i'\right) + \frac{K}{\eta} \sum_{t_{-i} \in T_{-i}} \pi_i (t_{-i}) \, h_i \left(t_{-i}, \pi_i'\right).$$

If he tells the truth, his payoff is

$$\widehat{v}_i \left(\theta_i, \theta_i, \widehat{\psi}_i (\pi_i)\right) + y_i^* (\theta_i, \pi_i) + \frac{K}{\eta} \sum_{t_{-i} \in T_{-i}} \pi_i (t_{-i}) \, h_i (t_{-i}, \pi_i).$$

His has an incentive to tell the truth if the latter expression exceeds the former, or:

$$\left\{ \begin{array}{l} \frac{K}{\eta} \left[\sum\limits_{t_{-i} \in T_{-i}} \pi_i (t_{-i}) \, h_i (t_{-i}, \pi_i) - \sum\limits_{t_{-i} \in T_{-i}} \pi_i (t_{-i}) \, h_i (t_{-i}, \pi_i') \right] \\ + \left[\widehat{v}_i \left(\theta_i, \theta_i, \widetilde{\psi}_i (\pi_i)\right) + y_i^* (\theta_i, \pi_i) - \widehat{v}_i \left(\theta_i, \theta_i', \widetilde{\psi}_i (\pi_i)\right) - y_i^* \left(\theta_i', \pi_i'\right) \right] \end{array} \right\} \geq 0.$$

If $\pi_i' \neq \pi_i$, the first term is strictly greater than K (by (6.4)) and the second term is less than or equal to K (by (6.2)), so the inequality holds. If $\pi_i = \pi_i'$, the first term is zero and the second term holds by (6.1). ■

The "only if" part of the result will continue to hold if there are an infinite set of types. For the "if" type, one could try to deal with infinite type spaces using the techniques of McAfee and Reny (1992), although one would probably require additional structure on the infinite type space. The result is sensitive to the assumptions of risk neutrality and no limited liability constraints (Robert (1991) made this argument in a fixed type space setting). But the key question is whether it is reasonable to assume that there are many possible belief types for each payoff type. We now discuss this issue.

Neeman (2001) emphasized that a key property in the surplus extraction results of Cremer and McLean (1985) is the following:

Definition 6.3. *Type space $\mathcal{T}$ satisfies the one-to-one property if*

$$\widehat{\pi}_i \left(t_i'\right) = \widehat{\pi}_i (t_i) \Rightarrow \widehat{\theta}_i \left(t_i'\right) = \widehat{\theta}_i (t_i).$$

This property implies that if the mechanism designer can find out the beliefs of a agent about other agent's types, then the mechanism designer can deduce his payoff type. If the one-to-one property holds, then the condition of Proposition 6.2 hold vacuously. A traditional justification for this assumption is that if we fixed a finite type space, and picked a generic common prior, and derived agents' beliefs from the common prior, the one-to-one property would automatically hold. However, it is unclear what this thought experiment is supposed to prove. What does it mean to assume that the set of possible types is common knowledge (already a very strong assumption), and then maintain common knowledge of the common prior *even as the common prior is varied*? It is more natural to ask if the one-to-one property is satisfied on large state spaces that relax common knowledge assumptions. In fact, the following very different property will hold on sufficiently rich type spaces.

Definition 6.4. *Type space $\mathcal{T}$ satisfies the product property if for all $\theta_i \in \Theta_i$ and $\pi_i \in \Delta(T_{-i})$ in the range of $\widehat{\pi}_i$, there exists $t_i \in T_i$ such that*

$$\widehat{\pi}_i(t_i) = \pi_i \ \text{ and } \ \widehat{\theta}_i(t_i) = \theta_i.$$

This property holds by construction on the universal type space (or the union of all possible type spaces). It seems like a natural property to assume when one does not want to make strong common knowledge assumptions about the players' higher order beliefs.[15] If one makes this assumption, Proposition 6.2 tells us the non-trivial incentive constraints that must then be satisfied.[16]

Further insight into the one-to-one and product properties can be obtained by restricting attention to the case where the common prior assumption holds. Under the common prior assumption, all type spaces have a certain conditional independence property of the prior (this was noted in Lemma 2 in Neeman (2001)). For simplicity, we state the result for finite type spaces, but this result should extend straightforwardly to arbitrary type spaces. Let Π_i and Π be the range of $\hat{\pi}_i(\cdot)$ and $\hat{\pi}(\cdot)$, respectively.

[15] While the product property holds automatically on the universal type space, it would be interesting to explore when it holds, or holds "typically", on large subsets of the universal type space where some common knowledge assumptions (e.g., the common prior assumption) are built in. Such an exercise is beyond the scope of this paper. In particular, it requires a notion of genericity for the universal type space. While the standard approach (fix types, vary the common knowledge common prior) is clearly flawed, it is not obvious what alternative to use (see Morris (2002) for a discussion of this issue).

[16] McLean and Postlewaite (2001) discuss an interesting environment where efficient implementation is possible even though the one-to-one property fails. The key to their results is that the payoff type information that cannot be extracted from belief types is exclusively private value information, and standard Vickrey auction arguments can be used to deal with this residual incentive problem.

Lemma 6.5 (Conditional Independence).
If $\mathcal{T}$ is a finite type space with a common prior $p(\cdot)$, then there exists $\beta \in \Delta(\Pi)$ and, for each i, $\nu_i : \Pi_i \to \Delta(\Theta_i)$, such that

$$\sum_{\left\{t \in T : \widehat{\pi}(t)=\pi \text{ and } \widehat{\theta}(t)=\theta\right\}} p(t) = \beta(\pi) \prod_{i=1}^{I} \nu_i(\theta_i | \pi_i). \tag{6.5}$$

Proof. Write $\phi_i(\theta_i | t_{-i}, \pi_i)$ for the probability that agent i has payoff type θ_i, contingent on agent i having belief type π_i and other agents having types t_{-i}. Abusing notation, write:

$$p(\pi_i) \equiv \sum_{t_{-i} \in T_{-i}} \sum_{\{t_i : \widehat{\pi}_i(t_i) = \pi_i\}} p(t),$$

$$p(\pi_i) \equiv \sum_{t_{-i} \in T_{-i}} \sum_{\{t_i : \widehat{\pi}_i(t_i) = \pi_i\}} p(t),$$

$$p(\theta_i | \pi_i) \equiv \frac{\sum_{t_{-i} \in T_{-i}} \sum_{\left\{t_i : \widehat{\pi}_i(t_i) = \pi_i \text{ and } \widehat{\theta}_i(t_i) = \theta_i\right\}} p(t)}{\sum_{t_{-i} \in T_{-i}} \sum_{\{t_i : \widehat{\pi}_i(t_i) = \pi_i\}} p(t)},$$

$$p(t_{-i} | \theta_i, \pi_i) \equiv \frac{\sum_{\left\{t_i : \widehat{\pi}_i(t_i) = \pi_i \text{ and } \widehat{\theta}_i(t_i) = \theta_i\right\}} p(t_i, t_{-i})}{\sum_{t'_{-i} \in T_{-i}} \sum_{\left\{t_i : \widehat{\pi}_i(t_i) = \pi_i \text{ and } \widehat{\theta}_i(t_i) = \theta_i\right\}} p\left(t_i, t'_{-i}\right)}.$$

We notice that by the definition of belief type:

$$p(t_{-i} | \theta_i, \pi_i) = \pi_i(t_{-i}),$$

for all θ_i and t_{-i}. So

$$\begin{aligned}
\phi_i(\theta_i | t_{-i}, \pi_i) &= \frac{p(\pi_i)\, p(\theta_i | \pi_i)\, p(t_{-i} | \theta_i, \pi_i)}{\sum_{\theta'_i \in \Theta_i} p(\pi_i)\, p\left(\theta'_i \middle| \pi_i\right) p\left(t_{-i} \middle| \theta'_i, \pi_i\right)}, \\
&= \frac{p(\pi_i)\, p(\theta_i | \pi_i)\, \pi_i(t_{-i})}{\sum_{\theta'_i \in \Theta_i} p(\pi_i)\, p\left(\theta'_i \middle| \pi_i\right) \pi_i(t_{-i})}, \\
&= p(\theta_i | \pi_i).
\end{aligned}$$

Thus

$$p(\theta|\pi) \equiv \frac{\sum\limits_{\left\{t:\widehat{\pi}(t)=\pi \text{ and } \widehat{\theta}(t)=\theta\right\}} p(t)}{\sum\limits_{\{t:\widehat{\pi}(t)=\pi\}} p(t)},$$

$$= \prod_{i=1}^{I} p(\theta_i|\pi_i).$$

By setting

$$\beta(\pi) \equiv \sum_{\{t:\widehat{\pi}(t)=\pi\}} p(t)$$

and

$$\nu_i(\theta_i|\pi_i) \equiv p(\theta_i|\pi_i),$$

we obtain the representation given in (6.5). ■

Now observe that condition (6.5) is trivially satisfied if the one-to-one property holds (since, for each π_i, there is a unique possible θ_i). The Lemma shows that if the one-to-one property fails, it must nonetheless be the case that a conditional independence property holds (in a non-trivial way) when agents' types are represented as the product of payoff and belief types. Just as independence is natural in some economic environments, we can tell stories why non-trivial conditional independence might arise naturally. Two that have been suggested are the following:

1. Suppose that agents' valuations of an object include a common value component and an idiosyncratic component, and agents observe a noisy signal of their common value component. Assuming that their idiosyncratic value components are independent, their valuations will be correlated but their signal about the common value component will be a sufficient statistic for an agent's beliefs about his opponent's type and will be the "belief type" in our language. Thus the characterization of Lemma 6.5 holds despite a failure of the one-to-one property. In fact, the product property will hold in this example. McLean and Postlewaite's (2001) analysis of efficient auction design with multidimensional types is an example where the one-to-one property is dropped but correlation of types is maintained, using this common value / idiosyncratic value motivation.

2. Suppose that agents are uncertain about the accuracy of each agent's signal. Thus an agent's type includes a parameter describing the accuracy of his signal and also the actual signal observed. Parreiras (2002) shows that this natural story leads to non-trivial conditional independence, and thus a breakdown of the one-to-one property.

7. Conclusion

This paper considered the robustness of general implementation problems. We formalized a notion of robustness by requiring interim implementation to be successful for large type spaces. We introduced successively richer type spaces, starting with the naive type all the way to the universal type space to represent strategic uncertainty. We investigated the idea that ex post equilibrium implementation is actually required if we were to demand interim implementation on large type spaces, such as the universal type space. An exact equivalence result between ex post equilibrium on the naive type space and interim implementation on common prior type spaces was shown to hold for many important implementation problems, such as social choice functions in general environments or efficient implementation in quasi-linear environments.

The exact equivalence did not extend to all social choice environments and we presented examples of correspondences (with and without transferable utility) in which the disparity between the ex post and interim implementation notions became apparent. In response to this gap, we suggested the notion of augmented ex post implementation to obtain a general equivalence result. Clearly, further research may yet identify necessary and sufficient conditions for general social choice environments without the augmented notion of ex post implementability.

The current results merely represent some initial steps to spell out how far the requirements of common knowledge can be weakened in the pursuit of mechanism design solutions. In this paper, we considered social choice problems where the payoff states did not include the beliefs of the agents or the designer. Yet, in many design problems, such as revenue maximizing auctions, the beliefs of the agents are clearly payoff relevant for the designer. It would be interesting to identify settings where even if the designer knew that the payoff types were correlated according to a particular distribution, ex post equilibrium implementation would be required to implement on richer type spaces consistent with that belief. This paper looked at partial (i.e., truthful / incentive compatible) implementation. In a companion paper we use the current framework to look at weak implementation and full implementation. In addition, in settings where we have a common prior over a type space or a subset thereof, we would like to consider the robustness of *virtual implementation* (Abreu and Matsushima (1992)) where we only require implementation with high ex ante probability (rather than for every possible type).

Finally, the introduction of large types spaces, naturally leads to the question as to whether single crossing properties on the naive type space have analogue properties on larger type spaces. For example, Dasgupta and Maskin (2000) identify a single crossing property that is sufficient for efficient ex post equilibrium implementation. In their leading example, they consider a two agent case where agent 1 observes a signal θ_1, agent 2 observes a signal θ_2 and their valuations of an object are $v_1 = a\theta_1 + \theta_2$ and

$v_2 = a\theta_2 + \theta_1$, respectively. Their single crossing sufficient condition in this example is that $a \geq 1$. We would like to examine the analogue to this property on larger type spaces. Their are at two ways of doing this. One is to take the signals θ_1 and θ_2 as the payoff-types in our larger type space constructions. But it is also natural to ask what would happen if we took agents' valuations to be their payoff-types in our construction. Now to capture the signal approach of Dasgupta and Maskin (2000), we would need to impose common knowledge restrictions on agents' beliefs about valuations. In their leading example, it is (implicitly) assumed that agent 1 knows for certain the value of $av_1 - v_2$ (it is always equal to $\left(a^2 - 1\right)\theta_1$). Just as a recent literature has tried to understand the meaning of the common prior assumption when expressed in the language of agents' higher order beliefs in the universal types space (see, e.g., Feinberg (2000)), we can also try and express implementability conditions in terms of agents' higher order beliefs about valuations. This would represent a further step to make the "Wilson doctrine", which prefaced this paper, precise and operational.

8. Appendix

PROOF of Lemma 5.3: Suppose initially that condition (1.) holds but (2.) is false. Then there exist $y_1, ..., y_I$, each $y_i : \Theta \to \mathbb{R}$ satisfying ex post incentive compatibility

$$y_i(\theta_i, \theta_{-i}) - y_i(\theta_i', \theta_{-i}) \geq \zeta_i(\theta_i, \theta_i', \theta_{-i}), \tag{8.1}$$

for all i, θ_i, θ_i' and θ_{-i} and budget balance

$$\sum_{i=1}^{I} y_i(\theta_i, \theta_{-i}) = 0; \tag{8.2}$$

and there exist $\nu : \Theta \to \mathbb{R}$, and $\lambda = (\lambda_1, ..., \lambda_I)$, each $\lambda_i : \Theta_i^2 \times \Theta_{-i} \to \mathbb{R}_+$ such that

$$\nu(\theta) = \sum_{\theta_i' \in \Theta_i} \lambda_i(\theta_i, \theta_i', \theta_{-i}) - \sum_{\theta_i' \in \Theta_i} \lambda_i(\theta_i', \theta_i, \theta_{-i}) \tag{8.3}$$

for all i and $\theta \in \Theta$; and

$$\sum_{i=1}^{I} \sum_{\theta \in \Theta} \sum_{\theta_i' \in \Theta_i} \lambda_i(\theta_i, \theta_i', \theta_{-i}) \zeta_i(\theta_i, \theta_i', \theta_{-i}) > 0. \tag{8.4}$$

But now (8.1) implies that

$$\begin{aligned} &\sum_{i=1}^{I} \sum_{\theta \in \Theta} \sum_{\theta_i' \in \Theta_i} \lambda_i(\theta_i, \theta_i', \theta_{-i}) \left[y_i(\theta_i, \theta_{-i}) - y_i(\theta_i', \theta_{-i})\right] \\ \geq\ & \\ &\sum_{i=1}^{I} \sum_{\theta \in \Theta} \sum_{\theta_i' \in \Theta_i} \lambda_i(\theta_i, \theta_i', \theta_{-i}) \zeta_i(\theta_i, \theta_i', \theta_{-i}). \end{aligned}$$

The left hand side can be written as

$$\sum_{i=1}^{I} \sum_{\theta \in \Theta} \left(\sum_{\theta_i' \in \Theta_i} \lambda_i(\theta_i, \theta_i', \theta_{-i}) - \sum_{\theta_i' \in \Theta_i} \lambda_i(\theta_i', \theta_i, \theta_{-i}) \right) y_i(\theta_i, \theta_{-i})$$

which by (8.3) is equal to

$$\sum_{i=1}^{I} \sum_{\theta \in \Theta} \nu(\theta) y_i(\theta_i, \theta_{-i}) = \sum_{\theta \in \Theta} \nu(\theta) \sum_{i=1}^{I} y_i(\theta_i, \theta_{-i}) = 0,$$

where the last equality comes from the budget balance condition (8.2). But this implies that

$$0 \geq \sum_{i=1}^{I} \sum_{\theta \in \Theta} \sum_{\theta'_i \in \Theta_i} \lambda_i \left(\theta_i, \theta'_i, \theta_{-i}\right) \zeta_i \left(\theta_i, \theta'_i, \theta_{-i}\right),$$

contradicting (8.4).

To show that (2.) implies (1.), we appeal to the following version of Farkas' Lemma and a Corollary.

Lemma 8.1 (Farkas' Lemma).
There exists $y \in \mathbb{R}^k$ such that

$$\sum_{k=1}^{K} a_{jk} y_k \geq b_j$$

for all $j = 1, ..., J$ if and only if there does not exist $\lambda \in \mathbb{R}^J_+$ such that

$$\sum_{j=1}^{J} a_{jk} \lambda_j = 0$$

and

$$\sum_{j=1}^{J} b_j \lambda_j > 0.$$

Corollary 8.2. *There exists $y \in \mathbb{R}^k$ such that*

$$\sum_{k=1}^{K} a_{jk} y_k \geq b_j$$

for all $j = 1, ..., J_0$ and

$$\sum_{k=1}^{K} a_{jk} y_k = b_j$$

for all $j = J_0 + 1, ..., J$ if and if there does not exist $\lambda \in \mathbb{R}^J$ such that

$$\lambda_j \geq 0,$$

for all $j = 1, ..., J_0$,

$$\sum_{j=1}^{J} a_{jk} \lambda_j = 0$$

and

$$\sum_{j=1}^{J} b_j \lambda_j > 0.$$

The interim dual characterization is established by the same argument, appropriately modified, as the ex post dual characterization. ■

References

[1] Abreu, D. and H. Matsushima. 1992. "Virtual Implementation in Iterative Undominated Strategies: Complete Information." *Econometrica* 60: 993-1008.

[2] d'Aspremont, C. and L.-A. Gerard-Varet. 1979. "Incentives and Incomplete Information." *Journal of Public Economics* 11: 25-45.

[3] d'Aspremont, C., J. Cremer and L.-A. Gerard-Varet. 1995. "Correlation, Independence and Bayesian Incentives."

[4] d'Aspremont, C., J. Cremer and L.-A. Gerard-Varet. 2002. "Balanced Bayesian Mechanisms."

[5] Battigalli, P. and M. Siniscalchi. 2003. "Rationalization and Incomplete Information."

[6] Bergemann, D. and S. Morris. 2002. "Robust Implementation: The Role of Rich Type Spaces."

[7] Bergemann, D. and J. Valimaki. 2002. "Information Acquisition and Mechanism Design." *Econometrica* 70: 1007-1033.

[8] Brandenburger, A. and E. Dekel. 1993. "Hierarchies of Beliefs and Common Knowledge." *Journal of Economic Theory* 59: 189-198.

[9] Choi, J. and T. Kim. 1999. "A Nonparametric, Efficient Public Good Decision Mechanism: Undominated Bayesian Implementation." *Games and Economic Behavior* 27: 64-85.

[10] Cremer, J. and R. McLean. 1985. "Optimal Selling Strategies Under Uncertainty for a Discriminating Monopolist when Demands are Interdependent." *Econometrica* 53: 345-361.

[11] Dasgupta, P., P. Hammond and E. Maskin. 1979. "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility." *Review of Economic Studies* 46: 185-216.

[12] Dasgupta, P. and E. Maskin. 2000. "Efficient Auctions." *Quarterly Journal of Economics* 115: 341-388.

[13] Duggan, J. and J. Roberts. 1997. "Robust Implementation." University of Rochester.

[14] Eliaz, K. 2002. "Fault Tolerant Implementation." *Review of Economic Studies* 69, 589-610.

[15] Feinberg, Y. 2000. "Characterizing Common Priors in the Form of Posteriors." *Journal of Economic Theory* 91: 127-179.

[16] Groves, T. and J. Ledyard 1987. "Incentive Compatibility Since 1972," in *Information, Incentives and Economic Mechanisms* (T. Groves, R. Radner, and S. Reiter, Eds.). University of Minnesota Press: Minneapolis.

[17] Harsanyi, J. 1967/68. "Games with Incomplete Information Played by Bayesian agents." *Management Science* 14, 159-182, 320-334, 486-502.

[18] Heifetz, A. and D. Samet. 1888. "Topology-Free Typology of Beliefs." *Journal of Economic Theory* 82: 324-341.

[19] Holmstrom, B. and R. Myerson (1983). "Efficient and Durable Decision Rules with Incomplete Information." *Econometrica* 51: 1799-1820.

[20] Hurwicz, L. 1972. "On Informationally Decentralized Systems," in *Decision and Organization* (R. Radner and C. McGuire, Eds.). North Holland: Amsterdam.

[21] Jehiel, P. and B. Moldovanu. 2001. "Efficient Design with Interdependent Valuations." *Econometrica* 65: 1237-1259.

[22] Kalai, E. 2002. "Large Robust Games." Northwestern University.

[23] Ledyard, J. 1978. "Incentive Compatibility and Incomplete Information." *Journal of Economic Theory* 18: 171-189.

[24] Ledyard, J. 1979. "Dominant Strategy Mechanisms and Incomplete Information." In *Aggregation and Revelation of Preferences*, ed. Jean-Jacques Laffont. Amsterdam: North Holland Chapter 17, pp. 309-319.

[25] Lopomo, G. 1998. "The English Auction is Optimal among Simple Sequential Auctions." *Journal of Economic Theory* 82: 144-166.

[26] Lopomo, G. 2000. "Optimality and Robustness of the English Auction." *Games and Economic Behavior* 36: 219-240.

[27] Maskin, E. 1992. "Auctions and Privatization." In: *Privatization: Symposium in Honor of Herbert Giersch*, ed. H. Siebert. Tuebingen: J.C.B.Mohr pp. 115-136.

[28] McAfee, P. and P. Reny. 1992. "Correlated Information and Mechanism Design." *Econometrica* 60: 395-421.

[29] McLean, R. and A. Postlewaite. 2001. "Efficient Auction Mechanisms with Multidimensional Signals." University of Pennsylvania.

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

[31] Morris, S. 2002. "Typical Types." http://www.econ.yale.edu/~sm326/typical.pdf.

[32] Morris, S. and H. S. Shin. 2003. "Global Games: Theory and Applications" in *Advances in Economics and Econometrics (Proceedings of the Eighth World Congress of the Econometric Society)* (M. Dewatripont, L. Hansen and S. Turnovsky, Eds). Cambridge, England: Cambridge University Press.

[33] Myerson, R. 1981. "Optimal Auction Design." *Mathematics of Operations Research* 6: 53-73.

[34] Myerson, R. 1991. *Game Theory.* Cambridge: Harvard University Press.

[35] Neeman, Z. 2001. "The Relevance of Private Information in Mechanism Design." Boston University.

[36] Parreiras, S. 2002. "Informational Rents and Mechanism Design." University of North Carolina at Chapel Hill.

[37] Perry, M. and P. Reny. 2002. "An Ex Post Efficient Auction." *Econometrica* 70: 1199-1212.

[38] Postlewaite, A. and D. Schmeidler. 1986. "Implementation in Differential Information Economies." *Journal of Economic Theory* 39: 14-33.

[39] Robert, J. 1991. "Continuity in Auction Design." *Journal of Economic Theory* 55: 169-179.

[40] Wilson, R. 1985. "Incentive Efficiency in Double Auctions." *Econometrica* 53: 1101-1115.

[41] Wilson, R. 1987. "Game-Theoretic Analyses of Trading Processes." In Advances in Economic Theory: Fifth World Congress, ed. Truman Bewley. Cambridge: Cambridge University Press chapter 2, pp. 33-70.