Bayesian Implementation

Thesis by

John Duggan

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1994

(Submitted June 13, 1994)

To Nancy

Acknowledgements

Any redeeming aspects of this thesis reflect the incredibly free, yet supportive environment in which it was written. I thank my advisors, chiefly Tom Palfrey, John Ledyard, Kim Border, and Richard McKelvey, who encouraged me to explore my own interests and always offered guidance when I became lost. As role models, they have each contributed to the formation of my academic character, and I can only hope to one day match the scientific quality and creativity of their work. To John Ledyard I extend special thanks for his encouragement and support throughout my stay at Caltech.

I gladly acknowledge Mark Fey's special and important role in the development of this thesis. As a loyal officemate, he has willingly subjected himself to more than his share of obscure technical details as I've tried to forge them into theorems. Amazingly, he was able to comprehend these details and help me make sense of them.

I am grateful to my family, for their thoughts have been with me since the beginning of this enterprise and have been a source of strength. In this reference I include the Seidmans, who have warmly welcomed me as a member of their clan, and Miller, who has long served as my best reality check.

I cannot adequately acknowledge the love and understanding I have received from my wife, Nancy, but I will try. If I had my life to do over again, and I could choose any field of study at any school on any planet in any solar system, I would have you with me. Life wouldn't be the same without you. Thanks.

Abstract

In Chapter 1, I briefly survey the literature on Bayesian implementation, discuss its shortcomings, and summarize the contribution of this thesis. In Chapter 2, I formally state the implementation problem, making no assumptions about the agents' sets of types, preferences, or beliefs, and I prove Jackson's (1991) necessity and sufficiency results for environments satisfying two weak conditions called "invariance" and "independence." In short, incentive compatibility and Bayesian monotonicity are necessary for Bayesian implementability, and incentive compatibility and monotonicity-no-veto are sufficient. I prove Jackson's result that, for environments with conflict of interest, Bayesian monotonicity and monotonicity-no-veto are equivalent, but I show that conflict-of-interest places an unnatural restriction on agents' beliefs when the set of states is uncountable. I note that, when agents have uncountable sets of types, preferences over social choice functions derived from conditional expected utility calculations will generally be incomplete, and I show that this incompleteness sometimes leads to implausible Bayesian equilibrium predictions. I propose an extension of expected utility preferences that preserves the properties of invariance and independence.

In Chapter 3, I consider environments satisfying invariance and a condition called "interiority," and I show that incentive compatibility and an extension of Bayesian monotonicity are necessary and sufficient for Bayesian implementability. Using the extension of expected utility preferences proposed in Chapter 2

and assuming best-element-private values, I then show that interiority is satisfied in two important classes of environments: it holds in private and public good economies, and it holds in lottery environments, for which the set of outcomes is the set of probability measures over a measurable space of pure outcomes.

In Chapter 4, I consider lottery environments satisfying best-element-private values and a condition called "strict separability," and I use the results of Chapter 3 to show that incentive compatibility is necessary and sufficient for virtual Bayesian implementability. I then show that strict separability is satisfied for a suitably large class of environments. It holds when private values and value-distinguished types are satisfied and the set of pure outcomes is finite, and it holds when private values and value-distinguished types are satisfied and the set of pure outcomes is a finite set crossed with an open set of allocations of a transferable private good.

Contents

Chapter 1

Introduction

The canonical implementation problem is the problem facing a planner who must choose a social alternative when the desirability of alternatives depends on the pooled information of individuals. The planner could simply ask individuals to reveal their information and, assuming individuals are truthful, use their replies to determine the most desirable alternative. It is likely, however, that this simple institution will give some individuals incentives to report their information falsely, leading the planner to choose an alternative she wouldn't have chosen were she fully informed. That is, some individual might have information that leads him to believe he can do better by misleading the planner than by reporting honestly. If the moral fibre of society cannot be counted on to restrain the self-interest of individuals, the planner must impose an institution that gives individuals the incentives—whatever their information may be—to take actions leading to desirable alternatives. Other examples of implementation problems abound. Instead of a planner and a society, a regulator may seek to overcome externalities in production by devising a system of fines that gives a collection of firms incentives to produce at levels maximizing total surplus. Or a monopsonistic seller may wish to organize an auction that maximizes revenue.

Underlying these examples is a common structure that lends itself to formal analysis. By treating this structure formally, it is possible to prove theorems that apply to all of the examples at once, as well as the multitude of examples not listed above. I show next how the elements of the planner's problem are represented abstractly, using $\rightarrow$ to indicate the relationship of thing to mathematical object. Once understood for this example, the mapping should be obvious for the problems of the regulator and auctioneer. The elements are represented thusly,

society	$\rightarrow$	set of agents
the social alternatives	$\rightarrow$	a set of outcomes
an individual's information	$\rightarrow$	an agent's type
the pooled information of individuals	$\rightarrow$	a profile of types (state of the world)
an individual's conditional beliefs	$\rightarrow$	probability measure on a σ-algebra of states
the desires of the planner	$\rightarrow$	a social choice function
an agent's motivations	$\rightarrow$	type-contingent preferences

over social choice functions,

where a social choice function is defined as a mapping from states to outcomes. Note that I take agents' preferences over social choice functions as primitives, whereas agents are usually supposed to have preferences over outcomes that induce preferences over social choice functions. It is convenient to distinguish an implementation problem from a *Bayesian environment*, or simply *environment*, which is a specification of sets of agents, outcomes, types, agents' beliefs, a σ-algebra of states, and agents' interim preferences. An *implementation problem* is then an environment together with a social choice function.

Omitted from the above list is the institution put in place by the planner. This is represented abstractly by a mechanism, or game form, which permits a set of actions for each agent and specifies an outcome for every combination of agents' actions. Together with an environment, a mechanism induces a game of incomplete information in the obvious way. Given an implementation problem, a mechanism is said to *Bayesian implement* the social choice function at hand if there exists a Bayesian equilibrium of the induced game of incomplete information, and the outcomes of every such Bayesian equilibrium coincide with the outcomes of the social choice function. A social choice function is *Bayesian implementable* if there exists a mechanism that Bayesian implements it. *Bayesian implementation* is the design of mechanisms to solve implementation problems in this sense, and the theory of Bayesian implementation, the topic of this thesis, seeks to understand the conditions under which implementation problems have solutions.

The work on Bayesian implementation is abundant and notably includes Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1989), Mookherjee and Reichelstein (1990), and Jackson (1991).[1] These papers contribute to the theory of Bayesian implementation by isolating properties of implementation problems that are necessary for the existence of solutions, and by isolating properties of implementation problems that are sufficient for the existence of solutions. Managable conditions that are both necessary and sufficient for Bayesian implementability are prized but rare. Jackson (1991) supplies the sharpest known characterizations: the only implementation problems with solutions satisfy incentive compatibility and Bayesian monotonicity; and for environments with conflict of interest, all implementation problems satisfying incentive compatibility and Bayesian monotonicity have solutions.[2] For other environments, he shows that incentive compatibility and monotonicity-no-veto are sufficient for the existence of a solution to an implementation problem.

The results of Bayesian implementation, including Jackson's, are limited in two respects. First, it is universally assumed that interim preferences over social choice functions are derived from measurable utility functions over outcomes by calculating conditional expected payoffs. Second, it is assumed that each agent's

[1] See Palfrey (1992) and Palfrey and Srivastava (1993) for surveys of the literature.

[2] Jackson actually considers the implementation of social choice sets, which are collections of social choice functions.

set of types, and therefore the set of states, is finite.[3] This implies that each agent's set of possible preferences over outcomes is finite, a particularly strong restriction when the set of outcomes is infinite. For example, consider a two consumer, two commodity exchange economy in which consumer i's type is a real number t_i between zero and one, a state is a pair (t_1, t_2) of such numbers, and consumer i's utility function over commodity bundles is Cobb-Douglas with parameter t_i. While it may be reasonable to suppose that any utility function for consumer i with $0 < t_i < 1$ is possible, perhaps uniformly distributed over the interval, the assumption that i's set of types is finite precludes this. Consequently, there are many interesting environments, when formulated naturally to allow for infinite sets of types, to which existing results on Bayesian implementation do not apply.

In both respects, the results on Bayesian implementation are less general than the well-known results on Nash implementation (see Maskin, 1986). Nash environments can be formulated as Bayesian environments in which agents' types are perfectly correlated and their type-contingent beliefs place probability one on the realized state, and the results for Nash implementation do not rely on either the assumption that agents' sets of types are finite or that agents' preferences over outcomes have any utility representation, measurable or otherwise. The results of Chapter 2 extend Jackson's to the implementation of social

[3]An exception is Palfrey and Srivastava (1991), who consider compact metric spaces of types. The authors do not, however, explore the problems of measurability in full detail.

choice functions in environments for which neither of these assumptions need hold, therefore obtaining the results on Nash implementation as special cases. Implicit in my formulation of the implementation problem is less structure than is commonly regarded as definitional. The set of outcomes is arbitrary, the set of states is the cross product of arbitrary sets of types, the σ-algebra of states need not have a product structure, conditional beliefs of agents need not be derived from prior beliefs about the distribution of states, interim preferences over social choice functions need not be derived from preferences over outcomes, and these preferences need satisfy only reflexivity. In particular, the set of states may be uncountable and there may exist non-measurable sets of states. I show that Jackson's necessity result holds for environments satisfying "invariance," and that his general sufficiency result holds for environments satisfying invariance and "independence." The conditions of invariance and independence are extremely weak, the former stipulating that outcomes on sets of zero conditional measure are irrelevant for an agent's comparison of two social choice functions, and the latter stipulating that agents can compare the outcomes of social choice functions on proper subsets of states. I also prove Jackson's sufficiency result for environments with conflict of interest, but I show by way of an example that his condition places unnatural restrictions on the beliefs of agents when the set of states is uncountable.

Allowing for uncountable sets of states raises two important issues not encountered in the work on Bayesian implementation. First, and easiest to address, modulo games and integer games no longer possess the properties required of them. Jackson's proof uses a version of the modulo game such that, given the strategies of other agents, any agent can choose a strategy that wins the contest at every state. While this device is extremely useful for eliminating unacceptable strategy profiles as equilibria when the set of states is finite, it is ineffective when any agent's set of types is infinite. I construct a version of the modulo game, called the name recognition contest, that is effective for arbitrary sets of types. The contest specifies a very large set of possible names, and it asks each agent to report a name and a subset of names that is restricted in size but larger, in a sense, than the union of all agents' sets of types. The name is interpreted as the agent's name, and the subset of names is interpreted as the set of names recognizable to the agent. If one agent recognizes the reported names of all other agents but is recognized by no other agent, then the one agent wins the contest. The set of possible names and the admissible subsets of recognizable names are specified so that, given the strategies of other agents, any agent can report an unrecognizable name and a set of recognizable names owning the reported names of all types of other agents.

Second, and more fundamental, when there exist non-measurable sets of states, there will generally be social choice functions for which conditional expected payoffs cannot be calculated, and in such cases interim preferences derived from conditional expected utility calculations will be incomplete. Note that an outcome function composed with a strategy profile is exactly a social choice function, so agents' interim preferences over social choice functions naturally induce interim preferences over strategy profiles. When these interim preferences are incomplete, there are two equally viable notions of strategic stability: according to one view a strategy profile is stable when for no type of any agent does there exist a strictly preferred unilateral deviation; and according to another view a strategy profile is stable when it is weakly preferred by every type of every agent to every unilateral deviation. I adopt the first, permissive view of Bayesian equilibrium, but the results of Chapter 2 hold—with a slightly different meaning—when reformulated in accordance with the second, restrictive view as well.

Although the above results hold in any case, I argue that the incompleteness of interim preferences derived from conditional expected utility calculations is often implausible. Specifically, I offer an example in which this implausibility is reflected in Bayesian equilibrium strategy profiles that are clearly unstable. This particular manifestation of the problem of incompleteness is due to my permissive definition of Bayesian equilibrium, but I show that the restrictive definition leads to unsatisfactory Bayesian equilibrium predictions in the same example. That is, strategy profiles that are clearly stable are not equilibria. The issue of incompleteness is often circumvented by simply restricting agents to subsets of strategies for which conditional expected payoffs can be calculated, but this is unsatisfactory for two reasons. First, and especially relevant to the theory of Bayesian implementation, there may be no natural subset of strategies to which agents should be restricted. Even if this is not a problem, there remains the difficult task of formalizing the notion of "natural" in this context. Second, I show that in some games there exist strategy profiles that are clearly stable even though no agent can calculate conditional expected payoffs for them. Again, this objection is particularly relevant for the theory of Bayesian implementation, since its goal is to design mechanisms admitting only special equilibria. Unacceptable strategy profiles should be eliminated by well-designed mechanisms—not disqualified by ad hoc assumptions.

I propose an extension of expected utility interim preferences that avoids these pathologies by approximating conditional expected payoffs of social choice functions. More precisely, a social choice function composed with an agent's state-contingent utility function yields a mapping from states to the real numbers, and when this function is bounded there will exist a measurable function that dominates it pointwise. Integrals can be calculated for each such function with respect to an agent's conditional beliefs, and the conditional expected payoff of the social choice function can be approximated by the infimum of these integrals. This approximation is defined for every bounded function from states to the real numbers, and I refer to it as the upper integral. Assuming that agents' utility functions are bounded, I use the upper integral to extend expected utility interim preferences over social choice functions to complete binary relations in such a way that environments with interim preferences given by this extension satisfy invariance and independence.

Since conflict-of-interest is unduly restrictive, the only results with wide applicability to environments with uncountable sets of states are then the partial characterizations of Chapter ??. In Chapter 3, I provide a full characterization of Bayesian implementability applicable in environments satisfying invariance and a condition called "interiority." For such environments, incentive compatibility and an extension of Bayesian monotonicity are necessary and sufficient for Bayesian implementability. Interiority requires, roughly, that there exist an "interior" set of outcomes such that no social choice function with interior values is best for any type of any agent, and that any strict preference for one social choice function over another can be replaced by a strict preference for an interior social choice function over the other. This greatly simplifies the elimination of undesirable strategy profiles, and the results of Chapter 2 can be tightened accordingly. I then show that many environments of interest satisfy interiority. Assuming best-element-private values, I show that interiority is satisfied by the continuous environments, which are defined by three properties: agents have extended expected utility interim preferences, agents' utility functions over outcomes are continuous, and the set of outcomes that are best for no type of any agent is dense. The weakness of interiority follows upon consideration of the environments satisfying these conditions. They include the private good economies with continuous, monotonic preferences, and since conflict-of-interest is irrelevant here, they include the pure public good economies as well. These environments also include the lottery environments, for which the set of outcomes is the set of probability measures on an underlying measurable space of pure outcomes, and for which a weak no-indifference assumption holds.

Extended Bayesian monotonicity is even stronger than Bayesian monotonicity, which is known to be restrictive in some environments, so rather few implementation problems can be expected to have solutions in the sense of Bayesian implementability. It is therefore of interest to explore weaker, yet acceptable standards for the existence of a solution. One way to weaken the requirements of Bayesian implementability in lottery environments is to require only that a social choice function have arbitrarily close (in an appropriate metric) Bayesian implementable neighbors. This is the notion of virtual Bayesian implementability, first introduced by Matsushima (1988) in Nash environments and subsequently investigated by Abreu and Sen (1991) in Nash environments. Abreu and Matsushima (1990b) consider Bayesian environments in which each agent's set of types is finite and which satisfy a very weak condition reminiscent of what I call "strict separability." They prove that incentive compatibility and a technical measurability condition are necessary and sufficient for virtual implementability in iteratively undominated strategies, and inspection of their sufficiency proof reveals that these conditions are also sufficient for Bayesian implementability. Matsushima (1993) considers Bayesian environments in which there are at least three agents, the set of pure outcomes is a finite set crossed with an open set of allocations of a transferable private good, and each agent's set of types is finite. Assuming a weak form of value-distinguished types, he shows that strict incentive compatibility is sufficient for virtual Bayesian implementability. These characterizations are both partial in nature and both rely on the assumption that the set of states is finite.

In (Duggan, 1994), I consider lottery environments satisfying best-element-private values and strict separability, assuming only that agents' prior beliefs agree on sets of measure zero and that each agent's set of types is a Hausdorff topological space.[4] I show that in such environments incentive compatibility is necessary and sufficient for virtual Bayesian implementability. The framework in which I prove these results differs from that of Chapters 2 and 3 in two important respects. First, I restrict agents' strategies to fairly natural subsets for which expected payoffs can be calculated. Second, I argue that this restriction is problematic in the common interim formulation of Bayesian equilibrium, so I use the ex ante formulation of Bayesian equilibrium, whereby agents choose strategies before they learn their types and they seek to maximize their ex ante expected payoffs. Measurability issues aside, the interim formulation of Bayesian equilibrium is somewhat more desirable than the ex ante, since there may be ex ante equilibria with some agents not best-responding on sets of types with zero prior probability.

In Chapter 4, I prove the characterization result of (Duggan, 1994) for virtual Bayesian implementability in the interim framework of the two previous chap-

[4] I actually use the even weaker assumption that the diagonal of the set of states is measurable with respect to the product σ-algebra derived from the σ-algebras on the agents' sets of types.

ters. The proof uses the characterization of Chapter 3 and proceeds by showing that, in lottery environments satisfying best-element-private values and strict separability, every implementation problem satisfying incentive compatibility has arbitrarily close approximations satisfying incentive compatibility and extended Bayesian monotonicity. Strict separability demands, roughly, that there exist a social choice function such that each type of each agent is best off when the social choice function uses none other than the agent's true type, regardless of which types of other agents are used. If the social choice function is employed as a mechanism in which agents simply report their types, this is tantamount to requiring that truth be a strict dominant strategy equilibrium of the induced game. This is apparently a strong condition, but I show that it is satisfied in a suitably large class of environments. It holds in lottery environments satifying private values and value-distinguished types when the set of pure outcomes is finite, or when the set of pure outcomes is a finite set crossed with an open set of allocations of a transferable private good.[5]

The results of Chapters 2, 3, and 4 represent a substantial contribution to the theory of Bayesian implementation. In Chapter 2, I extend the most powerful existing characterization results, which rely on the assumptions that the set of states is finite and that agents have preferences over outcomes with a measurable utility representation, to environments satisfying the much weaker conditions of

[5]I should take this opportunity to point out the weakness of Abreu and Matsushima's measurability condition: it is always satisfied in environments satsifying private values and value-distinguished types.

invariance and independence. After extending Jackson's full characterization of Bayesian implementability for environments with conflict of interest, I note that it's applicability is limited when the set of states is uncountable, and in Chapter 3 I offer an alternative full characterization that applies to every environment satisfying invariance and interiority, regardless of the set of states. These results are summarized in Figure 1, where arrows are labeled by the conditions needed for the implication, and in Figure 2 I illustrate the logical relationships between several interesting classes of environments and these conditions, where undirected lines represent conjunction. From Figures 1 and 2 can be derived a number of easy corollaries, only some of which are stated formally in the sequel. In Chapter 4, I replace the existing results for virtual Bayesian implementability, which are partial in nature and rely on the assumption that the set of states is finite, by a full characterization of virtual Bayesian implementability for lottery environments satisfying best-element-private values and strict separability, regardless of the set of states. I show that these conditions, strict separability in particular, hold in a suitably large class of environments.

Technically, these results are contributions to game theory, but they can be applied in specific examples of implementation problems to the extent that Bayesian equilibrium describes the behavior of individuals, firms, bidders, and so on. As with other results on Bayesian implementation, the mechanisms in my sufficiency proofs may be exceedingly difficult to use in practice and the Bayesian equilibria of their induced games may fare poorly as predictions of be-

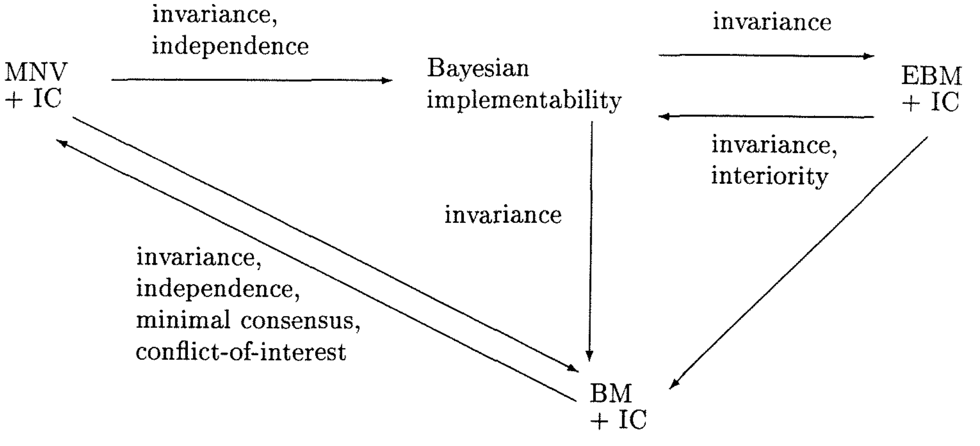

Figure 1.1: Results for Implementability

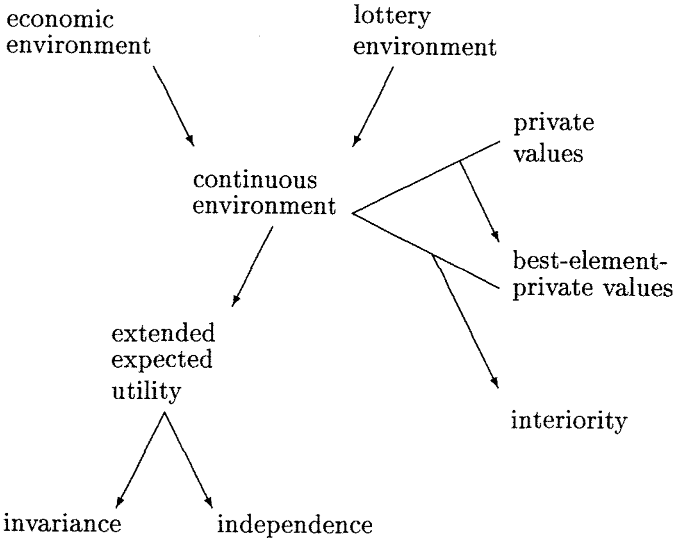

Figure 1.2: Conditions for Environments

havior, and my results are therefore subject to the usual criticism. But it seems to me that this criticism is misplaced. These mechanisms are mathematical constructions designed solely for the purpose of proving game-theoretic propositions under the weakest possible assumptions, and as such they can hardly be expected to serve as institutions, systems of fines, or auctions in real implementation problems. The theory of Bayesian implementation seeks to understand when implementation problems, abstractly formulated, have solutions, thereby informing the planner, regulator, or monopsonist as to whether it is theoretically possible to create institutions with proper incentive properties. In case it is possible, the institutions actually imposed will have to rely on whatever environment-specific structure is available, and practical aspects will have to be considered.

An appropriate target of criticism may be, however, the notion of Bayesian implementation itself. That is, it may be objected that the theoretical possibility of implementing social choice functions using unrealistic mechanisms is uninformative in specific examples of implementation problems. Jackson (1992) argues that to each solution concept there should correspond a class of admissible mechanisms, and that the mechanisms used in sufficiency proofs should be admissible for their corresponding solution concept. He proposes, in effect, to replace the notion of implementation with a stronger notion of admissible implementation. Another alternative, and the one I favor, is to replace the solution concept of Bayesian equilibrium with weaker, more compelling solution concepts. The ultimate solution concept would reflect difficulty of play and would be powerful in the statistical sense—it would predict any strategy profile that could reasonably arise in a game—even at the cost of high type 1 error. Some possibilities in this direction are explored by Abreu and Matsushima (1990a) for iterative removal of weakly dominated strategies, Abreu and Matsushima (1990b, 1992) for iterative removal of strictly dominated strategies, and Jackson (1992) for one-stage removal of weakly dominated strategies. Jackson shows, however, that even this very weak solution concept is not weak enough. I leave further considerations of these issues for future work.

Chapter 2

Bayesian Implementability in Arbitrary Environments

In Section 2.1, I supply the notation and definitions required for the formal treatment of Bayesian implementation in arbitrary environments. I impose less structure on the components of an implementation problem than is usually presumed definitional. In particular, I take agents' interim preferences over social choice functions as primitives rather than assuming they are derived from conditional expected utility calculations, and I place no restrictions on the size of agents' sets of types. In Section 2.2, I note that the integer game and modulo game are ineffective when the set of states is infinite, and I offer the name recognition contest in its place. This contest possesses the crucial property of the integer game and modulo game, it is applicable for arbitrary sets of types, and it does not rely on the axiom of choice. In Section 2.3, I prove Jackson's necessity result for environments satisfying invariance, and I prove his general sufficiency result for environments satisfying invariance and independence. I also prove the equivalence of Bayesian monotonicity and monotoncity-no-veto for environments satisfying conflict-of-interest, with an extension of Jackson's full characterization as a corollary. I show by way of an example, however, that conflict-of-interest restricts agents' beliefs in an unnatural way when the set of states is uncountable. When this is the case, there will generally be social choice functions for which conditional expected payoffs cannot be calculated and expected utility interim preferences will be incomplete. In Section 2.4, I show that this incompleteness leads to Bayesian equilibria that are clearly unstable, and I propose an extension of expected utility interim preferences that satisfies invariance and independence.

2.1 Notation and Definitions

A completely general approach to Bayesian implementation would consider implementation problems with no restrictions on the set of agents, the set of outcomes, the sets of types for each agent, the beliefs of agents, the preferences of agents, or the social choice function to be implemented. Existing results on Bayesian implementation with incomplete information are quite general with respect to the social choice function, often implementing collections of social choice functions called "social choice sets." But they are less than general in other respects. It is universally assumed that the set of agents is finite and that the preferences of agents over social choice functions are derived from expected utility calculations, and it is nearly always assumed that the sets of types for each agent are finite and that agents' conditional beliefs are derived from a common prior. When this last assumption is not made, it is replaced by a weaker assumption regarding common support of beliefs.

My formulation of the implementation problem is completely general in all but two respects. I assume that the set of agents is finite with at least two members, and I consider the implementation of single social choice functions rather than social choice sets. The latter restriction is made for convenience, and the results below would likely hold for social choice sets with three or more agents—once considerations of closure (see Jackson, 1991) are made. Beyond these assumptions, the restrictions I impose are weak enough to be considered definitional. An implementation problem is an ordered pair (e, f) consisting of an environment e and a social choice function f, which is just a mapping from states to outcomes. Let F denote the collection of social choice functions for e with generic elements $f, h \in F$, and let $\mathcal{P}(F \times F)$ denote the set of binary relations on F. Formally, an environment e is a sequence $(I, O, T, \mathcal{T}, \mu, R)$, where

$I = \{1, \ldots, n\}$	set of agents with $n \geq 2$ and elements i, j

O	set of possible outcomes with elements x, y
T_i	set of possible types for agent i with elements t_i
$T = \times_{i \in I} T_i$	set of possible states with elements $t = (t_1, \ldots, t_n)$
$\mathcal{T}$	σ-algebra on T
$\mu : I \times T \times \mathcal{T} \to \Re$	agents' conditional beliefs
$R : I \times T \to \mathcal{P}(F \times F)$	agents' interim preferences over social choice functions.

This formulation departs from the standard formulation in several ways: I require nothing of the sets of outcomes or types; $\mathcal{T}$ need not have a product structure; conditional beliefs need not be derived from prior beliefs on $(T, \mathcal{T})$; and the $R_i(t_i)$ need not be derived from conditional expected utility calculations, nor need they satisfy completeness or transitivity.

I do require that $\mathcal{T}$ contains all singleton cylinder sets of the form $\{t_i\} \times T_{-i}$, where I use the notation t_{-i} and T_{-i} in the usual way. Agents' conditional beliefs depend only on their own types, that is, $\mu_i(.|t_i, t_{-i}) = \mu_i(.|t_i, t'_{-i})$ for all $i \in I$, all $t \in T$, and all $t'_{-i} \in T_{-i}$, and I will write simply $\mu_i(.|t_i)$ for the beliefs of agent i at type t_i. Of course, $\mu_i(\{t_i\} \times T_{-i}|t_i) = 1$ for all $i \in I$ and all $t_i \in T_i$. For all $i \in I$ and all $t_i \in T_i$, let $\mu_i^*(.|t_i)$ denote the outer measure defined by

$$\mu_i^*(S|t_i) = \inf\{\mu_i(Q|t_i) | S \subseteq Q, Q \in \mathcal{T}\},$$

for all $S \subseteq T$, and let μ^* denote the set function defined by

$$\mu^*(S) = \sup\{\mu_i^*(S|t_i)|i \in I, t_i \in T_i\}$$

for all $S \subseteq T$. The next proposition shows that μ^* is an outer measure, so that all sets with μ^*-measure zero are μ^*-measurable, and that μ^* is a measure on the σ-algebra of μ^*-measurable sets. Of course, this holds for each $\mu_i^*(.|t_i)$.

Proposition 1 *$\mu^*(\emptyset) = 0$; for all $S^1, S^2 \subseteq T$, $S^1 \subseteq S^2$ implies $\mu^*(S^1) \leq \mu^*(S^2)$; and for all $S^1, S^2 \subseteq T$, $\mu^*(S^1 \cup S^2) \leq \mu^*(S^1) + \mu^*(S^2)$.*

Proposition 1 is proved, with all other propositions, in the appendix. I will write $f \sim f^*$ if $\mu^*(\{t \in T | f(t) \neq f^*(t)\}) = 0$. The next proposition shows that $\sim$ is an equivalence relation.

Proposition 2 *$\sim$ is reflexive, symmetric, and transitive.*

Let $[f]$ denote the set of social choice functions f^* such that $f \sim f^*$.

I also require that agents' interim preferences over social choice functions depend only on their own types, that is, $R_i(t) = R_i(t_i, t'_{-i})$ for all $i \in I$, all $t \in T$, and all $t'_{-i} \in T_{-i}$, and I write simply $R_i(t_i)$ to denote the interim preferences of agent i at type t_i. These relations are meant to represent agents' weak interim preferences over social choice functions and therefore must satisfy reflexivity, but they need not satisfy completeness or transitivity. As usual, I write $f\ P_i(t_i)\ h$ if and only if $f\ R_i(t_i)\ h$ and $\neg h\ R_i(t_i)\ f$. The relations $P_i(t_i)$ are asymmetric and represent the agents' strict preferences. In the standard formulation of the implementation problem, interim preferences are derived from conditional expected payoffs determined by the integrals of state-contingent utility functions $u_i(.|.) : O \times T \to \mathbb{R}$. Denoting these preferences by $\tilde{R}$, $f\ \tilde{R}_i(\tilde{t}_i)\ h$ if and only if

$$\int_T u_i(f(t)|t)\, d\mu_i(t|\tilde{t}_i) \geq \int_T u_i(h(t)|t)\, d\mu_i(t|\tilde{t}_i).$$

When T is countable and $\mathcal{T}$ is the power set of T these interim preferences on F are complete and transitive, but when T is uncountable and $\mathcal{T}$ is a proper subset of the power set of T they will in general be incomplete, for it may be that $u_i(f(.)|.)$ is non-measurable for some social choice function. Formulating a Nash environment as a Bayesian environment, interim preferences $\hat{R}_i(t_i)$ over social choice functions are given by agents' preferences over outcomes, and they are complete as well as transitive. I will proceed without making any specific assumptions about the nature of interim preferences over social choice functions, and my results consequently apply to these environments as special cases. I return to these issues in Section 2.4.

Regardless of the set of states, environments with preferences given by $\tilde{R}$ or $\hat{R}$ satisfy the two weak conditions defined next. Given an environment e and two social choice functions f and h, let $f/_S h$ denote the *splicing of f with h along S*. That is, $f/_S h$ is defined by

$$(f/_S h)(t) = \begin{cases} h(t) & \text{if } t \in S \\ f(t) & \text{else} \end{cases}$$

for all $t \in T$.

Definition 1 *An environment e satisfies* **invariance** *if for all $i \in I$, all $t_i \in T_i$, and all $f^1, f^2, h^1, h^2 \in F$,*

$$f^1 \; R_i(t_i) \; h^1 \quad \textit{and} \; \& \quad \begin{array}{l} \mu_i^*(\{t \in T | f^1(t) \neq f^2(t)\} | t_i) = \\ \mu_i^*(\{t \in T | h^1(t) \neq h^2(t)\} | t_i) = 0 \end{array}$$

imply $f^2 \; R_i(t_i) \; h^2$.

Definition 2 *An environment e satisfies* **independence** *if for all $i \in I$, all $t_i \in T_i$, all $f^1, f^2, h^1, h^2 \in F$, and all $S \subseteq T$,*

$$f^1/_S h^1 \; R_i(t_i) \; f^2/_S h^1 \; \textit{implies} \; f^1/_S h^2 \; R_i(t_i) \; f^2/_S h^2.$$

Invariance requires only that an agent's preferences between two social choice functions are unaffected by changes on sets of states with zero outer measure. In particular, if $t_i' \neq t_i''$ then the outcomes of f and h on the set $\{t_i''\} \times T_{-i}$ are irrelevant for agent i's interim preferences at t_i'. This is often taken for granted, but I make the assumption explicit. Independence, in effect, allows agents to compare the outcomes of social choice functions on proper subsets of states. When the preference $f^1/_S h^1 \; R_i(t_i) \; f^2/_S h^1$ is independent of the specification of h^1, as independence requires, it makes sense to say that agent i weakly prefers the outcomes of f^1 on $T \setminus S$ to those of f^2.

A game of incomplete information in its most abstract form is a quadruple (I, T, Σ, Π), where

Σ_i	set of strategies available to agent I
$\Sigma = \times_{i \in I} \Sigma_i$	set of profiles of available strategies

$\Pi : I \times T \to \mathcal{P}(\Sigma \times \Sigma)$ interim preferences over strategy profiles.

Elements of Σ_i are denoted σ_i, strategy profiles are denoted σ, and I use the notation σ_{-i} in the usual way. I require that agents' interim preferences over strategy profiles depend only on their own types, that is, $\Pi_i(t) = \Pi_i(t_i, t'_{-i})$ for all $i \in I$, all $t \in T$, and all $t'_{-i} \in T_{-i}$, and I write simply $\Pi_i(t_i)$ to denote agent i's interim preferences at type t_i. I interpret $\Pi_i(t_i)$ as a strict preference relation and therefore require it to be asymmetric. A mechanism is an ordered pair (M, g), where

M_i	set of possible messages for agent i
$M = \times_{i \in I} M_i$	set of possible message profiles
$g : M \to O$	outcome function.

Elements of M_i are denoted m_i, profiles of messages are denoted m, and I use the notation m_{-i} in the usual way. Let $\Sigma_i(M_i)$ denote the set of the functions $\sigma_i : T_i \to M_i$, and let $\Sigma(M) = \times_{i \in I} \Sigma_i(M_i)$ denote the set of profiles of such functions. In an environment e, a mechanism (M, g) induces a game of incomplete information with the set $\Sigma_i = \Sigma_i(M_i)$ of strategies available to each agent i and interim preferences $\Pi_i(t_i)$ over strategy profiles given by interim preferences over the corresponding social choice functions. That is, $\sigma \; \Pi_i(t_i) \; \tilde{\sigma}$ if and only if

$$g \circ \sigma \; P_i(t_i) \; g \circ \tilde{\sigma}.$$

I conform with the literature on Bayesian implementation by considering only induced games with pure strategies rather than behavioral strategies, which map from an agent i's types to probability measures over M_i equipped with some σ-algebra.

Work on Bayesian implementation defines the Bayesian equilibrium strategy profiles of a game of incomplete information as those for which no type of any agent can gain from a unilateral deviation. This is satisfactory when agents' interim preferences are complete, but the possibility that the $R_i(t_i)$ are incomplete raises an interesting issue in the definition of Bayesian equilibrium. Is a strategy profile stable whenever it is weakly preferred to every unilateral deviation, or is it stable whenever there is no strictly preferred unilateral deviation? I take the second, more permissive approach, but the results of this chapter continue to hold when reformulated according to the restrictive approach.[1] I next define Bayesian equilibrium, given a mechanism and an environment, for an induced game of incomplete information, and this leads to the definition of Bayesian implementability.

Definition 3 *For a mechanism (M, g) and environment e, let $B_{(M,g,e)}$ denote the set of* **Bayesian equilibrium** *strategy profiles of the game of incomplete information in e induced by (M, g). Then $\sigma \in B_{(M,g,e)}$ if $\sigma \in \Sigma(M)$ and*

$$\neg g \circ (\tilde{\sigma}_i, \sigma_{-i}) \; P_i(t_i) \; g \circ \sigma$$

[1] More precisely, the statements of the definitions and theorems below would be unchanged, though their content would reflect the different meaning of Bayesian equilibrium under the restrictive approach.

for all $i \in I$, all $t_i \in T_i$, and all $\tilde{\sigma}_i \in \Sigma_i(M_i)$.

Definition 3 is stated in terms of social choice functions, reflecting their primitive status, but it is equivalent to the usual definition (see Myerson, 1991) when T is countable, $\mathcal{T}$ is the power set of T, and interim preferences over social choice functions are given by $\tilde{R}$.

Definition 4 *A mechanism (M, g)* **Bayesian implements** *the social choice function f in e if $B_{(M,g,e)} \neq \emptyset$ and for all $\sigma \in B_{(M,g,e)}$ $g \circ \sigma \in [f]$. A social choice function is* **Bayesian implementable** *in e if there exists a mechanism that Bayesian implements it in e.*

A social choice function f is Bayesian implementable in e if there exists a mechanism with at least one Bayesian equilibrium and with the property that the outcomes of every Bayesian equilibrium coincide with the outcomes of f at all but a μ^*-outer measure zero set of states. Such a mechanism effectively solves the implementation problem (e, f).

In Section 2.3, I will show how the implementation problems with solutions are related to those satisfying incentive compatibility and Bayesian monotonicity, which I state next for a given an implementation problem (e, f). Following the above convention, $\Sigma_i(T_i)$ will denote the set of all functions from T_i to T_i, and $\Sigma(T)$ will denote the set of all profiles of such functions. Let $\tau \in \Sigma(T)$ denote the truthful strategy profile defined by $\tau_i(t_i) = t_i$ for all $i \in I$ and all $t_i \in T_i$.

Definition 5 *An implementation problem* (e, f) *satisifes* **incentive compatibility** *if* $\tau \in B_{(T,f,e)}$.

For all $i \in I$, let $\tilde{F}_i$ denote the set of social choice functions $\tilde{f}$ such that, for all $t_i \in T_i$ and all $\alpha_i \in \Sigma_i(T_i)$,

$$\neg \tilde{f} \circ (\alpha_i, \tau_{-i}) \; P_i(t_i) \; f.$$

The set $D_{(e,f)}$ of deceptions is the set of $\alpha \in \Sigma(T)$ such that $f \circ \alpha \notin [f]$.

Definition 6 *An implementation problem* (e, f) *satisfies* **Bayesian monotonicity** *if for all* $\tilde{\alpha} \in D_{(e,f)}$ *there exist* $j \in I$, $\tilde{t}_j \in T_j$, *and* $\tilde{f} \in \tilde{F}_j$ *such that*

$$\tilde{f} \circ \tilde{\alpha} \; P_j(\tilde{t}_j) \; f \circ \tilde{\alpha}.$$

Incentive compatibility requires of a social choice function f that truthful reporting is a Bayesian equilibrium in e of the game induced by the mechanism (T, f), and Bayesian monotonicity corresponds Maskin monotonicity (see Maskin, 1986) in Bayesian environments with incomplete information. It applies when agents use deceptive strategies, which are reports of types that lead under the mechanism (T, f) to outcomes other than social choices on a set of positive μ^*-outer measure. Bayesian monotonicity requires for each deception the existence of an agent j, a type $\tilde{t}_j$, and a social choice function $\tilde{f}$ such that, when other agents report deceptively, type $\tilde{t}_j$ of agent j prefers the outcomes under $\tilde{f}$ to those under f. The additional requirement that $\tilde{f} \in \tilde{F}_j$ entails that, when other agents report truthfully, no type of agent j prefers the outcomes under $\tilde{f}$ to those under f.

Jackson (1991) introduces monotonicity-no-veto, a property of implementation problems which in a sense combines Bayesian monotonicity with no veto power. For a given implementation problem (e, f), I next state a simplified version of Jackson's condition. For all $j \in I$ and all $Q = \bigcup_{i \in I} Q_i \times T_{-i}$, let $Q_{\neq i} = \bigcup_{j \neq i} Q_j \times T_{-j}$. I refer to such a set Q as a *cross*.

Definition 7 *An implementation problem* (e, f) *satisfies* **monotonicity-no-veto** *if for all* $\alpha \in \Sigma(T)$, *all* $Q = \bigcup_{i \in I} Q_i \times T_{-i}$, *and all* $\hat{f} \in F$ *such that* $(f \circ \alpha)/_Q \hat{f} \notin [f]$,

$$\forall i \in I, t_i \in T_i, h \in F$$

$$\neg (f \circ \alpha)/_{Q_{\neq i}}(h \circ \alpha)\ P_i(t_i)\ (f \circ \alpha)/_{Q_{\neq i}} \hat{f}$$

implies

$$\exists j \in I, \tilde{t}_j \in T_j \setminus Q_j, \tilde{f} \in \tilde{F}_j \qquad (\tilde{f} \circ \alpha)/_{Q_{\neq j}} \hat{f}\ P_j(\tilde{t}_j)\ (f \circ \alpha)/_{Q_{\neq j}} \hat{f}.$$

Monotonicity-no-veto is not entirely transparent. It is stronger than Bayesian monotonicity because it considers crosses Q and social choice functions $\hat{f}$ that, roughly speaking, do best for each agent i on $Q_{\neq i}$. Monotonicity-no-veto requires, for each deception α and every such Q and $\hat{f}$, the existence of $j \in I$, $\tilde{t}_j \in T_j \setminus Q_j$, and $\tilde{f} \in \tilde{F}_j$ satisfying the following property. When outcomes are given off Q by the social choice function f with agents reporting α and on Q by $\hat{f}$, type $\tilde{t}_j$ of agent j would prefer to have outcomes off Q determined by the social choice function $\tilde{f}$ with agents reporting α. That is, type $\tilde{t}_j$ of agent

j must have a strict preference when outcomes are changed on only the subset $T \setminus Q$ of states. Setting $Q = \emptyset$, this additional requirement is unrestrictive, and monotonicity-no-veto reduces to Bayesian monotonicity.

Jackson observes that monotonicity-no-veto is actually equivalent to Bayesian monotonicity in environments with a certain conflict of interest.[2] Moreover, Jackson observed that, when the set of states is finite, this conflict of interest is evidenced by every economic environment with a private good. The following is a somewhat weakened version of Jackson's condition.

Definition 8 *An environment e satisfies* **conflict-of-interest** *if, for all $i \in I$, all $t_i \in T_i$, all $f \in F$, and all $\alpha \in \Sigma(T)$, there exist $j_1, j_2 \in I$, $t_{j_1} \in T_{j_1}$, $t_{j_2} \in T_{j_2}$, $h^1, h^2 \in F$ with $j_1 \neq j_2$ such that*

$$f/_{\{t_i\} \times T_{-i}}(h^1 \circ \alpha)\ P_{j_1}(t_{j_1})\ f$$

and

$$f/_{t_i \times T_{-i}}(h^2 \circ \alpha)\ P_{j_2}(t_{j_2})\ f.$$

This version is weaker than Jackson's in two ways. First, Jackson's holds for arbitrary sets $S \subseteq T$ rather than for cylinder sets $\{t_i\} \times T_{-i}$, and second, Jackson's version requires the existence of constant social choice functions h^1 and h^2 satisfying the above condition. Nonetheless, my version of conflict-of-interest is strong enough to deliver the equivalence of Bayesian monotonicity and monotonicity-no-veto, even when the set of states is uncountable.

[2]Jackson refers to these environments as "economic environments."

2.2 The Name Recognition Contest

Proving that an implementation problem has a solution usually requires the construction of a mechanism that solves the problem. The mechanisms used in the Bayesian implementation literature rely on the existence of a contest that is a *free-for-all*, in the sense that every agent can win at every state if the strategies of the other agents are fixed. Precisely, a contest is a pair (C, w) where $C = \times_{i\in I} C_i$, C_i is a message space for agent i, and $w : C \to I$ specifies a winner for every profile $c = (c_1, \ldots, c_n)$ of messages. The required property is the following: for all $j \in I$ and all $\gamma \in \Sigma(C)$ there exists $\tilde{c}_j \in C_j$ such that, for all $t \in T$, $w(\tilde{c}_j, \gamma_{-j}(t_{-j})) = j$. An example of such a contest is the well-known integer game, which has each agent reporting an integer with a prize going to the owner of the highest number. That is, $C = \times_{i\in I}\mathbb{Z}$ with elements $z = (z_1, \ldots, z_n)$ and $w(z) = \min(\arg\max_{i\in I} z_i)$. When each agent's set of types is finite this contest is a free-for-all, since an agent can simply report one plus the highest integer reported by any agent at any state. When agents have infinite sets of types this construction becomes problematic, since there may not be a highest integer reported by other agents.

Another example is an adaptation due to Jackson (1991) of the modulo game. Jackson's construction has each agent report $n+1$ vectors of integers in the set $V = \{0, 1, \ldots, n\overline{T}^2\}$, where $\overline{T} = \max_{i\in I} |T_i|$. The first n vectors of each agent's report have length $\overline{T}$, and each agent's $n+1$th vector has length one.[3] An agent

[3]This is actually somewhat simplified. By increasing the size of V, I am able to use reports

i wins the contest if, for all $j \neq i$, j's $n+1$th vector appears as a component in i's jth vector and i's $n+1$th vector does not appear in j's ith vector. If no such agent exists then agent 1 wins the contest by default. More precisely, each $C_i = V_1 \times \cdots \times V_n \times V$, where for all $i \in I$, $V_i = \{0, 1, \ldots, n\overline{T}^2\}^{\overline{T}}$. Let c_i^j denote the jth vector in i's report c_i, and let $c_i^{j,k}$ denote the kth component of i's jth vector. Finally, let $w(c) = i$ if, for all $j \neq i$, $c_j^{n+1} = c_i^{j,k}$ for some component k and $c_i^{n+1} \neq c_j^{i,k}$ for no component k. If no such i exists then let $w(c) = 1$.

To see that this contest is a free-for-all, fix the strategies γ_{-i} of agents other than i. Agent i must set $\tilde{c}_i^j$ so that, for all $j \neq i$, each element of the set $\{c_j^{n+1} | \exists t_j \in T_j \; c_j = \gamma_j(t_j)\}$ appears as a component of $\tilde{c}_i^j$. Since j has at most $\overline{T}$ types and i's jth vector has $\overline{T}$ components, i can pick $\tilde{c}_j$ to satisfy this condition. Also i must set $\tilde{c}_i^{n+1}$ so that it does not appear as a component in any vector in the set $\{c_j^i | \exists j \in I, t_j \in T_j \; c_j = \gamma_j(t_j)\}$. This set contains at most $n\overline{T}$ vectors each with length $\overline{T}$. The set of components appearing in these vectors therefore has size no greater than $n\overline{T}^2$. Since $\tilde{c}_i^{n+1}$ can take on $n\overline{T}^2 + 1$ values, i can pick $\tilde{c}_i^{n+1}$ to satisfy the condition. This shows that the adapted modulo game has the desired property. Although difficult to describe, the adapted modulo game has the advantage that when T is finite so is C, a property not shared with the integer game. As with the integer game, however, the construction breaks down when T is infinite.

In this section I construct a contest, the name recognition contest, that

---

such that each agent's $n+1$th vector has length 1 rather than the length n used by Jackson.

is very close to the adapted modulo game but has the free-for-all property for arbitrary sets of types for each agent. Each agent picks a name from a very large set of possible names and picks a subset of familiar names that is restricted in size but still larger than the disjoint union of the agents' sets of types. If an agent i is familiar with the names of all other agents and no other agent recognizes i's name then i wins the contest, and if this holds for no agent then the winner is the lowest indexed agent. Any agent can win the contest given the strategies of other agents by reporting a set of familiar names that includes the names of all types of all other agents and picking a name that is not familiar to any type of any other agent.

Before formally defining the name recognition contest, some preliminaries are in order. For two arbitrary sets X and Y, I will write $X \trianglerighteq Y$ if there exists a mapping from X onto Y, and I will write $X \triangleright Y$ if $X \trianglerighteq Y$ and not $Y \trianglerighteq X$. This relation captures a notion similar to the cardinalities of X and Y, and coincides with that notion when the axiom of choice holds. To see that $X \triangleright Y \trianglerighteq Z$ implies $X \triangleright Z$, let f_{XY} denote a mapping from X onto Y, and let f_{YZ} denote a mapping from Y onto Z. Then $f_{YZ} \circ f_{XY}$ is a mapping from X onto Z. Suppose there exists a mapping f_{ZX} from Z onto X. Then $f_{ZX} \circ f_{YZ}$ is a mapping from Y onto X, a contradiction.

As no confusion will result, I will denote the name recognition contest by

(C, w). Let $\overline{T} = \bigcup_{i \in I} T_i \times \{i\}$, let the set A denote the power set of

$$\mathbb{T} = \bigcup_{i \in I} \bigcup_{t_i \in T_i} \{i\} \times \{t_i\} \times \overline{T},$$

and note that $A \rhd \mathbb{T}$. Let $\mathcal{A}$ denote the collection of subsets B of A such that $\overline{T} \trianglerighteq B$, and note that $\mathcal{A}$ is non-empty. When T is infinite, for example, $\mathcal{A}$ contains all finite subsets of A. Intuitively, A is the set of names agents have to choose from, and $\mathcal{A}$ is the collection of sets of familiar names agents have to choose from. For all $i \in I$, let $C_i = A \times \mathcal{A}$ with elements $c_i = (a_i, B_i)$, so each agent i reports a name $a_i \in A$ and a set $B_i \in \mathcal{A}$ of familiar names. Let $w(c) = i$ if $a_i \in A \setminus \bigcup_{j \neq i} B_j$ and $\bigcup_{j \neq i} \{a_j\} \subseteq B_i$, and if this holds for no i then $w(c) = 1$. The next theorem shows that the name recognition contest is a free-for-all.

Theorem 1 *For all $j \in I$ and all $\gamma \in \Sigma(C)$ there exists $\tilde{c}_j \in C_j$ such that, for all $t \in T$, $w(\tilde{c}_j, \gamma_{-j}(t_{-j})) = j$.*

Proof: I will sometimes write an agent i's strategy $\gamma_i \in \Sigma_i(A \times \mathcal{A})$ as (ν_i, β_i). To see that

$$\mathbb{T} \trianglerighteq \bigcup_{i \in I} \bigcup_{t_i \in T_i} \beta_i(t_i),$$

note that, for all $i \in I$ and all $t_i \in T_j$, $\overline{T} \trianglerighteq \beta_i(t_i)$ implies the existence of a mapping ϕ_{i,t_i} from $\overline{T}$ onto $\beta_i(t_i)$. Then define $\phi : \mathbb{T} \to \bigcup_{i \in I} \bigcup_{t_i \in T_i} \beta_i(t_i)$ by $\phi(i, t_i, t_j, j) = \phi_{i,t_i}(t_j, j)$. Take $b \in \beta_i(t_i)$ for some $i \in I$ and $t_i \in T_i$. Since ϕ_{i,t_i} is onto $\beta_i(t_i)$, there exists $(t_j, j) \in \overline{T}$ such that $\phi_{i,t_i}(t_j, j) = b$. Then $\phi(i, t_i, t_j, j) = b$, and it follows that ϕ is onto $\bigcup_{i \in I} \bigcup_{t_i \in T_i} \beta_i(t_i)$.

Now take $j \in I$ and fix the strategies γ_{-j} of the other agents. Let $\tilde{B}_j = \{b \in A | \exists i \in I, t_i \in T_i \ \nu_i(t_i) = b\}$. That $\overline{T} \trianglerighteq \tilde{B}_j$ follows since the mapping $\psi : \overline{T} \to \tilde{B}_j$ defined by $\psi(t_j) = \nu_j(t_j)$ is clearly onto. Let $\tilde{b}$ be an element of $A \setminus \bigcup_{i \in I} \bigcup_{t_i \in T_i} \beta_i(t_i)$, which is non-empty since

$$A \triangleright \mathbb{T} \trianglerighteq \bigcup_{i \in I} \bigcup_{t_i \in T_i} \beta_i(t_i).$$

Let $\tilde{a}_j = \tilde{b}$ for all $t_j \in T_j$. Then when j uses $\tilde{c}_j = (\tilde{a}_j, \tilde{B}_j)$, j recognizes the names of all other agents at every state, while no agents recognize j's name. Therefore, $w(\tilde{c}_j, \gamma_{-j}(t_{-j})) = j$ for all $t \in T$. ∎

There are other contests, similar in spirit to the integer game, that also have the free-for-all property. For example, assume that T is infinite and well-order the set $2^{\overline{T}}$ by $\preceq$, let 0 denote the least element of $2^{\overline{T}}$, and let $[0, \omega]$ denote the set of elements between 0 and ω according to $\preceq$. The contest has each agent report an element of $[0, \omega_1)$, where ω_1 is the least element of $2^{\overline{T}}$ with the property that $[0, \omega_1]$ has greater cardinality than $\overline{T}$.[4] The winner of the contest is the agent who reports the highest element according to $\preceq$. The set of reports of types of agents other than j will have cardinality no greater than that of $\overline{T}$, so it has a least upper bound $\overline{\omega} \in [0, \omega_1)$. Note that $(\overline{\omega}, \omega_1)$ is non-empty, for otherwise $[0, \overline{\omega}]$ has cardinality greater than $\overline{T}$, contradicting the choice of ω_1. Then agent

[4] That such an element exists can be assumed without loss of generality. If not, consider the set $\overline{T}' = \overline{T} \cup \{\omega'\}$ and extend $\preceq$ to $\preceq'$, according to which ω' is the unique greatest element. Then $\preceq'$ is a well-ordering and the set of ω such that $[0, \omega]$ has cardinality greater than $\overline{T}$ owns ω'. By the definition of a well-ordering, there is a least such element ω_1.

j can report $\omega \in (\overline{\omega}, \omega_1)$, thereby winning the contest at every state.[5] Unlike the name recognition contest, this contest uses infinite message spaces even when T is finite and it relies on the axiom of choice to ensure the existence of $\preceq$.

2.3 Characterization Results

The necessity of incentive compatibility for the existence of a solution to an implementation problem is known as the revelation principle, first acknowledged in the Bayesian framework by d'Aspremont and Gérard-Varet (1979) and Myerson (1979). The necessity of monotonicity for implementation in Nash environments was recognized by Maskin (see Maskin, 1986), and was established for a class of Bayesian environments Postlewaite and Schmeidler (1986). Jackson (1991) was the first to note the implications of types with zero probability. In this case, an implementation problem has a solution only if there exists a μ^*-equivalent implementation problem satisfying incentive compatibility and Bayesian monotonicity. It is not necessary that the original problem satisfy these conditions. The next theorem shows that Jackson's result holds for every environment satisfying invariance. Note that independence is not needed.

Theorem 2 *Assume e satisfies invariance. f is Bayesian implementable in e only if (e, f^*) satisfies incentive compatibility and Bayesian monotonicity for some $f^* \in [f]$.*

[5] I am indebted to Kim Border for this construction.

Proof: Assume that the mechanism (M, g) Bayesian implements f in e, and take $\sigma^* \in B_{(M,g,e)}$. Let $f^* = g \circ \sigma^* \in [f]$. Then for all $i \in I$, all $t_i \in T_i$, and all $\alpha_i \in \Sigma_i(T_i)$

$$\neg f^* \circ (\alpha_i, \tau_{-i}) \; P_i(t_i) \; f^*$$

if and only if

$$\neg f^* \circ (\sigma_i^* \circ \alpha_i, \sigma_{-i}^*) \; P_i(t_i) \; g \circ \sigma^*,$$

which holds since $\sigma^* \in B_{(M,g,e)}$. This implies that $\tau \in B_{(T,f^*,e)}$ and that (e, f^*) satisfies incentive compatibility.

Now take any $\tilde{\alpha} \in D_{(e,f^*)}$, so that $f^* \circ \alpha \notin [f^*]$. Then Proposition 2 implies that $f^* \circ \alpha \notin [f]$, and since (M, g) Bayesian implements f in e it follows that $\sigma^* \circ \tilde{\alpha} \notin B_{(M,g,e)}$. So there exist $j \in I$, $\tilde{t}_j \in T_j$, and $\tilde{\sigma}_j \in \Sigma_j(T_j)$ such that

$$g \circ (\tilde{\sigma}_j, \sigma_{-j}^* \circ \tilde{\alpha}_{-j}) \; P_j(\tilde{t}_j) \; g \circ \sigma^* \circ \tilde{\alpha}.$$

By invariance, this implies that

$$g \circ (\tilde{m}_j, \sigma_{-j}^* \circ \tilde{\alpha}_{-j}) \; P_j(\tilde{t}_j) \; g \circ \sigma^* \circ \tilde{\alpha},$$

where $\tilde{m}_j = \tilde{\sigma}_j(\tilde{t}_j)$. Define $\tilde{f} = g \circ (\tilde{m}_j, \sigma_{-j}^*)$, so $\tilde{f} \circ \tilde{\alpha} = g \circ (\tilde{m}_j, \sigma_{-j}^*) \circ \tilde{\alpha} = g \circ (\tilde{m}_j, \sigma_{-j}^* \circ \tilde{\alpha}_{-j})$ and $f^* \circ \tilde{\alpha} = g \circ \sigma^* \circ \tilde{\alpha}$ imply

$$\tilde{f} \circ \tilde{\alpha} \; P_j(\tilde{t}_j) \; f^* \circ \tilde{\alpha},$$

as desired. To see that $\tilde{f} \in \tilde{F}_j$, take any $t_j \in T_j$ and $\alpha_j \in \Sigma_j(T_j)$, and note that invariance implies

$$\neg \tilde{f} \circ (\alpha_j, \tau_{-j}) \; P_j(t_j) \; f^*$$

if and only if

$$\neg g \circ (\tilde{m}_j, \sigma^*_{-j}) \; R_j(t_j) \; g \circ \sigma^*,$$

which holds since $\sigma^* \in B_{(M,g,e)}$. ∎

Implementation problems satisfying incentive compatibility and Bayesian monotonicity are particularly tractable, for a rather intuitive reason. To solve such an implementation problem (e, f), a mechanism can have agents report their types and use these reports to pick the outcome determined by the social choice function f. Since incentive compatibility is satisfied, truthful reporting will be a Bayesian equilibrium of the mechanism with outcomes that coincide everywhere with f, so the only remaining difficulty is the possibility that some equilibria lead to outcomes that do not coincide μ^*-almost everywhere with f. One way to deal with this is to enrich the message spaces of agents, allowing each agent i to indicate that other agents are reporting deceptively and to impose the outcomes of some social choice function $\tilde{f} \in \tilde{F}_i$ using the reported types of other agents. No agent can gain by imposing these outcomes when other agents report truthfully, but when the mechanism at every state uses deceptive reports of agents to pick outcomes determined by f, Bayesian monotonicity ensures that some type of some agent j can gain by defecting and imposing the outcomes of some $\tilde{f} \in \tilde{F}_j$.

Incentive compatibility and Bayesian monotonicity are insufficient, however, for the existence of a solution to an implementation problem. Once the message spaces of agents are so enriched, a mechanism need not at every state use agents' reports to pick the outcomes determined by f. In this case the opportunity to defect may not be enough to eliminate all unwanted equilibria, but the attractiveness of defecting is greatly enhanced for implementation problems satisfying monotonicity-no-veto. Take a strategy profile, suppose that off some cross Q of states the mechanism uses the deceptive reports of agents to pick the outcomes determined by f, and suppose that on Q the mechanism picks outcomes in some other way. If the strategy profile is a Bayesian equilibrium and the mechanism gives each agent i the opportunity to impose any social choice function on the band $Q_{\neq i}$ and any social choice function $\tilde{f} \in \tilde{F}_i$ elswhere, then the antecedent of monotonicity-no-veto must hold. The condition ensures that some $i \in T_i \setminus Q_i$ can gain by defecting and imposing the outcomes of some $\tilde{f} \in \tilde{F}_i$ on the set $T \setminus Q$. Jackson showed that incentive compatibility and monotonicity-no-veto are, in fact, sufficient for the solution to an implementation problem, and the next theorem shows that a version of this result holds for every environment satisfying invariance and independence.

Theorem 3 *Assume e satisfies invariance and independence. Then f is Bayesian implementable in e if (e, f^*) satisfies incentive compatibility and monotonicity-no-veto for some $f^* \in [f]$.*

Proof: It suffices to find a mechanism (M, g) such that $B_{(M,g,e)} \neq \emptyset$ and $g \circ \sigma \in [f^*]$ for all $\sigma \in B_{(M,g,e)}$, since this implies $g \circ \sigma \in [f]$ by Proposition 2. Have each agent i report $\hat{m}_i = (\hat{t}_i, \hat{f}_i, \hat{h}_i, \hat{l}_i, \hat{k}_i, \hat{c}_i, \hat{e}_i) \in M_i = T_i \times \tilde{F}_i \times F \times \{0,1\} \times \{0,1\} \times C_i \times C_i$, where (C, w) denotes the name recognition contest. I will sometimes represent a strategy $\hat{\sigma}_i \in \Sigma_i(M_i)$ by the sequence $(\hat{\alpha}_i, \hat{\phi}_i, \hat{\psi}_i, \hat{\lambda}_i, \hat{\kappa}_i, \hat{\gamma}_i, \hat{\epsilon}_i)$ of component functions. Before defining the outcome function, I must adapt the name recognition contest so that it may be played by any two agents. Let $w_{i,j}(c_i, c_j) = i$ if $a_i \in A \backslash B_j$ and $a_j \in B_i$, and similarly for j. If neither condition holds then $w_{i,j}(c_i, c_j) = \emptyset$. Given a message profile $\hat{m}$, let $I(\hat{k}) = \{i \in I | \hat{k}_i = 1\}$, let $I(\hat{l}) = \{i \in I | \hat{l}_i = 1\}$, and order $I(\hat{l})$ as follows. If there exists $i \in I(\hat{l})$ such that, for all $j \in I(\hat{l}) \setminus \{i\}$, $w_{i,j}(\hat{e}_i, \hat{e}_j) = i$, then i is top ranked in $I(\hat{l})$. Use the same procedure to determine the top ranked agent of $I(\hat{l}) \setminus \{i\}$, who will be ranked second in $I(\hat{l})$. Iterate this procedure until there is no agent who beats the remaining agents in the name recognition contest, and then order them according to their indices. Let $L(\hat{l}, \hat{e})$ denote the lowest ranked agent of $I(\hat{l})$ according to the ordering derived from $\hat{e}$.

Now partition M into $n + 2$ sets

$$\begin{aligned} M^0 &= \{\hat{m} \in M | \exists i, j \in I \text{ such that } i \neq j, i \in I(\hat{k}), \\ &\qquad j \in I(\hat{k}) \cup I(\hat{l}), i = w_{i,j}(\hat{e}_i, \hat{e}_j)\} \\ M^i &= \{\hat{m} \in M \setminus M^0 | L(\hat{l}, \hat{e}) = i\} \\ M^{n+1} &= M \setminus \bigcup_{i=0}^{n} M^i, \end{aligned}$$

and note that $\hat{l} \neq 0$ implies that $\hat{m} \in M^0 \cup \bigcup_{i \in I} M^i$. It is straightforward to check that $\hat{k} \neq 0$ and $\hat{m} \in M^{n+1}$ if and only if $\hat{l} = 0$ and there do not exist distinct $i, j \in I(\hat{k})$ such that $i = w_{i,j}(\hat{e}_i, \hat{e}_j)$. Define the outcome function as

$$g(\hat{m}) = \begin{cases} \hat{h}_{w(\hat{c})}(\hat{t}) & \text{if } \hat{m} \in M^0 \\ \hat{f}_i(\hat{t}) & \text{if } \hat{m} \in M^i \\ f^*(\hat{t}) & \text{if } \hat{m} \in M^{n+1}. \end{cases}$$

Unlike most mechanisms, each agent reports two social choice functions, two integers, and two entries in the name recognition contest. These complications are needed to ensure, given a strategy profile and the cross Q consisting of the states at which at least one agent reports a positive integer, that any agent j can impose the outcomes of social choice functions in F on $Q_{\neq j}$ without affecting the outcomes off $Q_{\neq j}$, and that every type $t_j \in T_j \setminus Q_j$ of agent j can impose the outcomes of social choice functions in $\tilde{F}_j$ off $Q_{\neq j}$ without affecting outcomes on $Q_{\neq j}$.

To anticipate the arguments below, suppose that outcomes are given by the strategy profile $\hat{\sigma}$. The mechanism is constructed so that agent j can impose the outcomes of $\tilde{h} \in F$ on $Q_{\neq j}$ by reporting $\tilde{h}$ as j's second social choice function, reporting one as j's second integer, and submitting entries that ensure victory in both name recognition contests. The second integer acts aggressively, triggering the $\hat{c}$-name recognition contest whenever another agent i reports a positive integer and loses to j in the pairwise $\hat{e}$-name recognition contest. Leaving other reports the same, it is easy to see that this strategy will not affect the outcomes off $Q_{\neq j}$. A type $t_j \in T_j \setminus Q_j$ of agent j can impose the outcomes of $\tilde{f} \in \tilde{F}_j$ off $Q_{\neq j}$ by reporting $\tilde{f}$ as j's first social choice function, reporting one as j's first integer, and submitting an entry in the $\hat{e}$-name recognition contest to ensure victory. Leaving other reports the same, I claim that this strategy changes only the outcomes off $Q_{\neq j}$. If $\hat{\sigma}(t) \in M^0$ then the only consequence of j's switch is to include j in the set of agents reporting a positive passive integer, which will not move the reported message profile. Since j's second social choice function and entry in the $\hat{c}$-name recognition contest are the same, the outcome is unaffected. If $\hat{\sigma}(t) \in M^i$ for some i then j's switch will not move the reported message profile. Since j wins the $\hat{e}$-name recognition contest, the outcome is unaffected. If $\hat{\sigma}(t) \in M^{n+1}$ then j's switch will not move the reported message profile. Since j reports the same type, the outcome is unaffected.

The proof consists of two steps. The first step is to establish the existence of $\sigma^* \in B_{(M,g,e)}$ such that $g \circ \sigma^* \in [f^*]$. The second step is to show that this equality holds for all $\sigma \in B_{(M,g,e)}$.

Step 1 : Consider the strategy profile σ^* defined by $\sigma_i^*(t_i) = (t_i, \bar{f}_i, \bar{h}, 0, 0, \bar{c}_i)$ for all $i \in I$ and all $t_i \in T_i$, where $\bar{f}_i \in \tilde{F}_i$, $\bar{h} \in F$, and $\bar{c}_i \in C_i$ are arbitrary constants. To see that $\sigma^* \in B_{(M,g,e)}$, take any $i \in I$ and $\tilde{\sigma}_i \in \Sigma_i(M_i)$, and partition T_i into $S_i^1 = \{t_i \in T_i | \tilde{\lambda}_i(t_i) > 0\}$ and $S_i^2 = T_i \setminus S_i^1$. For $t_i \in S_i^2$

$$\neg g \circ (\tilde{\sigma}_i, \sigma_{-i}^*) \; P_i(t_i) \; g \circ \sigma^*$$

if and only if

$$\neg f^* \circ (\tilde{\alpha}_i, \tau_{-i}) \; P_i(t_i) \; f^*,$$

which holds since (e, f^*) satisfies incentive compatibility. For $t_i \in S_i^1$,

$$\neg g \circ (\tilde{\sigma}_i, \sigma^*_{-i})\ P_i(t_i)\ g \circ \sigma^*$$

if and only if

$$\neg \tilde{f} \circ (\tilde{\alpha}_i, \tau_{-i})\ P_i(t_i)\ f^*$$

which holds since $\tilde{f} = \tilde{\phi}_i(t_i) \in \tilde{F}_i$. Therefore $\sigma^* \in B_{(M,g,e)}$, and clearly $\mu^*(\{t \in T | g(\sigma^*(t)) \neq f^*(t)\}) = \mu^*(\emptyset) = 0$.

Step 2 : Now take $\sigma^* \in B_{(M,g,e)}$, and suppose that $g \circ \sigma^* \notin [f^*]$. Let $Q_i = \{t_i \in T_i | \kappa_i^*(t_i) + \lambda_i^*(t_i) > 0\}$, let $Q = \bigcup_{i \in I} Q_i \times T_{-i}$, and let $Q_{\neq i} = \bigcup_{j \neq i} Q_j \times T_{-j}$ for all $i \in I$. Note that $g(\sigma^*(t)) = f^*(\alpha^*(t))$ for all $t \in T \setminus Q$, which is to say that

$$g \circ \sigma^* = (f^* \circ \alpha^*)/_Q (g \circ \sigma^*). \tag{2.1}$$

To see that

$$\forall i \in I, t_i \in T_i, h \in F$$

$$\neg((f^* \circ \alpha^*)/_{Q \setminus Q_{\neq i}} (g \circ \sigma^*))/_{Q_{\neq i}} (h \circ \alpha^*)\ P_i(t_i)\ (f^* \circ \alpha^*)/_Q (g \circ \sigma^*)$$

take $j \in I$ and $\tilde{t}_j \in T_j$ and suppose that

$$((f^* \circ \alpha^*)/_{Q \setminus Q_{\neq i}} (g \circ \sigma^*))/_{Q_{\neq j}} (\tilde{h} \circ \alpha^*)\ P_j(\tilde{t}_j)\ (f^* \circ \alpha^*)/_Q (g \circ \sigma^*) \tag{2.2}$$

for some $\tilde{h} \in F$. I claim that type $\tilde{t}_j$ of agent j can gain from a unilateral deviation, contradicting the assumption that $\sigma^* \in B_{(M,g,e)}$.

Consider the strategy $\tilde{\sigma}_j \in \Sigma_j(M_j)$ defined by $\tilde{\sigma}(t_j) = (\alpha_j^*(t_j), \phi_j^*(t_j), \tilde{h}, \lambda_j^*(t_j), 1, \tilde{c}_j, \tilde{e}_j)$, where $\tilde{c}_j$ and $\tilde{e}_j$ guarantee that j wins both name recognition contests at all states. This strategy changes the outcomes of $g \circ \sigma^*$ only on the set $Q_{\neq j}$, where agent j wins the name recognition contest and $\tilde{h}$ determines the outcome. Then

$$g \circ (\tilde{\sigma}_j, \sigma^*_{-j}) = (g \circ \sigma^*)/_{Q_{\neq j}}(\tilde{h} \circ \alpha^*)) \tag{2.3}$$

$$= ((f^* \circ \alpha^*)/_{Q \setminus Q_{\neq j}}(g \circ \sigma^*))/_{Q_{\neq j}}(\tilde{h} \circ \alpha^*) \tag{2.4}$$

where (2.4) follows from (2.3) by substituting from (2.1). Then substituting (2.1) and (2.4) into (2.2) yields

$$g \circ (\tilde{\sigma}_j, \sigma^*_{-j}) \; P_j(\tilde{t}_j) \; g \circ \sigma^*,$$

contradicting the assumption that $\sigma^* \in B_{(M,g,e)}$. Therefore,

$$\forall i \in I, t_i \in T_i, h \in F$$

$$\neg((f^* \circ \alpha^*)/_{Q \setminus Q_{\neq i}}(g \circ \sigma^*))/_{Q_{\neq i}}(h \circ \alpha^*) \; P_i(t_i) \; (f^* \circ \alpha^*)/_Q(g \circ \sigma^*).$$

By independence, this implies

$$\forall i \in I, t_i \in T_i, h \in F$$

$$\neg(f^* \circ \alpha^*)/_{Q_{\neq i}}(h \circ \alpha^*) \; P_i(t_i) \; (f^* \circ \alpha^*)/_{Q_{\neq i}}(g \circ \sigma^*)$$

so the antecedent of monotonicity-no-veto is fulfilled.

By monotonicity-no-veto, there exist $j \in I$, $\tilde{t}_j \in T_j \setminus Q_j$, and $\tilde{f} \in \tilde{F}_j$ such that

$$(\tilde{f} \circ \alpha^*)/_{Q_{\neq j}}(g \circ \sigma^*) \quad P_j(\tilde{t}_j) \quad (f^* \circ \alpha^*)/_{Q_{\neq j}}(g \circ \sigma^*) \tag{2.5}$$

I claim that type $\tilde{t}_j$ of agent j can gain from a unilateral deviation, contradicting the assumption that $\sigma^* \in B_{(M,g,e)}$. Note that $\kappa_j^*(t_j) = 0$, and consider the strategy $\tilde{\sigma}_j \in \Sigma_j(M_j)$ defined by

$$\tilde{\sigma}_j(t_j) \quad = \quad \begin{cases} (\alpha_j^*(t_j), \tilde{f}, \psi_j^*(t_j), 1, 0, \gamma_j^*(t_j), \tilde{e}_j) & \text{if } t_j = \tilde{t}_j \\ \sigma_j^*(t_j) & \text{else} \end{cases}$$

for all $t_j \in T_j$, where $\tilde{e}_j$ guarantees that j wins the second name recognition contest at all states. This strategy changes the outcomes of $g \circ \sigma^*$ only on the set $(\{\tilde{t}_j\} \times T_{-j}) \cap (T \setminus Q_{\neq j})$, where the outcome is determined by $\tilde{f} \circ \alpha^*$. Then

$$\mu_j^*(\{t \in T | g \circ (\tilde{\sigma}_j, \sigma_{-j}^*)(t) \neq (\tilde{f} \circ \alpha)/_{Q_{\neq j}}(g \circ \sigma^*)(t)\} | \tilde{t}_j) = 0,$$

so (2.5) and invariance imply

$$g \circ (\tilde{\sigma}_j, \sigma_{-j}^*) \quad P_j(\tilde{t}_j) \quad (f^* \circ \alpha^*)/_{Q_{\neq j}}(g \circ \sigma^*). \tag{2.6}$$

Since $\tilde{t}_j \notin Q_j$, (2.1), (2.6), and invariance imply

$$g \circ (\tilde{\sigma}_j, \sigma_{-j}^*) \quad P_j(\tilde{t}_j) \quad g \circ \sigma^*,$$

contradicting the assumption that $\sigma^* \in B_{(M,g,e)}$. Therefore, $g \circ \sigma^* \in [f^*]$. ∎

Unlike Theorem 2, Theorem 3 holds only for environments satisfying invariance and independence. Weak as these conditions are, independence can be dropped by suitably restating the property of monotonicity-no-veto. Changing the antecedent to

$$\forall i \in I, t_i \in T_i, h \in F, \tilde{h} \in \tilde{F}_i$$

$$\neg((f \circ \alpha)/_{Q \setminus Q_{\neq i}}(\tilde{h} \circ \alpha))/_{Q_{\neq i}}(h \circ \alpha) \ P_i(t_i) \ (f \circ \alpha)/_Q \hat{f}$$

suffices for this purpose. Little is gained by dropping independence, so I use the original statement of monotonicity-no-veto, which is closer to Jackson's.

Jackson (1991) shows that Bayesian monotonicity implies monotonicity-no-veto for environments satisfying conflict-of-interest, and it follows as a corollary that incentive compatibility and Bayesian monotonicity are necessary and sufficient for Bayesian implementability in such environments. Since, in Jackson's framework, these environments include all economies with a private good and selfish, monotonic utility functions over outcomes, this full characterization applies to many implementation problems of practical interest. I next prove Jackson's theorem for arbitrary environments satisfying conflict-of-interest, but I argue that the corollary on Bayesian implementability has limited applicability when the set of states is uncountable.

Theorem 4 *Assume an environment e satisfies conflict-of-interest. Then* (e, f) *satisfies Bayesian monotonicity if and only if it satisfies monotonicity-no-veto.*

Proof: That monotonicity-no-veto implies Bayesian monotonicity follows simply by setting $Q = \emptyset$. Now assume that (e, f) satisfies Bayesian monotonicity. Take $\alpha \in \Sigma(T)$, $Q = \bigcup_{i \in I} Q_i \times T_{-i}$, and $\hat{f} \in F$ such that $(f \circ \alpha)/_Q \hat{f} \notin [f]$ and

$$\forall i \in I, t_i \in T_i, h \in F$$

$$\neg (f \circ \alpha)/_{Q_{\neq i}}(h \circ \alpha) \; P_i(t_i) \; (f \circ \alpha)/_{Q_{\neq i}} \hat{f}.$$

Suppose that $Q \neq \emptyset$, so there exists $i \in I$ and $t_i \in T_i$ with $t_i \in Q_i$. Consider agents $j_1 \neq j_2$ corresponding to i, t_i, $f \circ \alpha$, and α in the definition of conflict-of-interest. If $j_1 \neq i$ then $\{t_i\} \times T_{-i} \subseteq Q_{\neq j_1}$ implies that

$$\forall h \in F \ \neg(f \circ \alpha)/_{\{t_i\}\times T_{-i}}(h \circ \alpha) \ P_{j_1}(t_{j_1}) \ f \circ \alpha,$$

contradicting the choice of j_1. Therefore, $j_1 = i$. The same argument establishes that $j_2 = i$, so that $j_1 = j_2$. This again contradicts the choice of j_1 and j_2, so the original supposition must be wrong. That is, $Q = \emptyset$. It follows that

$$(f \circ \alpha)/_Q \hat{f} = f \circ \alpha$$

and $f \circ \alpha \notin [f]$. By Bayesian monotonicity, there exists $j \in I$, $\tilde{t}_j \in T_j$, and $\tilde{f} \in \tilde{F}_j$ such that

$$\tilde{f} \circ \alpha \ P_j(\tilde{t}_j) \ f \circ \alpha.$$

Then $Q = \emptyset$ implies $\tilde{t}_j \in T_j \setminus Q_j$ and

$$(\tilde{f} \circ \alpha)/_{Q_{\neq j}} \hat{f} \ P_j(\tilde{t}_j) \ (f \circ \alpha)/_{Q_{\neq j}} \hat{f},$$

which establishes monotonicity-no-veto. ∎

The next corollary follows easily from Theorems 2, 3, and 4.

Corollary 1 *Assume e satisfies invariance, independence, minimal consensus, and conflict-of-interest. Then f is Bayesian implementable in e if and only if (e, f^*) satisfies incentive compatibility and Bayesian monotonicity for some $f^* \in [f]$.*

While conflict-of-interest is satisfied by a large and interesting class of environments when the set of states is finite, or even countably infinite, it becomes very restrictive when the set of states is uncountable. Intuitively, when agents' beliefs about the realization of the state are continuous, the sets $\{t_i\} \times T_{-i}$ will be negligible to all agents other than i and it will be impossible to find agents j_1 and j_2 as in the definition of conflict-of-interest. I next formalize this intuition with a simple example.

Example 1 Assume that $n = 3$, that each $T_i = [0, 1]$, and that $\mathcal{T} = *_{i=1}^{3} \mathcal{T}_i$, where each $\mathcal{T}_i$ is the Borel σ-algebra on $[0, 1]$. Assume that agents' conditional beliefs are derived from a common prior represented by a density function ϕ on T. Each agent i's beliefs conditional on type t_i are then represented by a conditional density $\phi(.|t_i)$. I leave the set of outcomes and preferences unspecified, assuming only that invariance is satisfied. Take any $i \in I, t_i \in T_i, f \in F, \alpha \in \Sigma(T), j_1 \neq j_2, t_{j_1} \in T_{j_1}, t_{j_2} \in T_{j_2}$, and $h^1, h^2 \in F$. Without loss of generality, assume $j_1 \neq i$. Then since $\mu^*_{j_1}(\{t_i\} \times T_i | t_{j_1}) = 0$, invariance implies

$$\neg f /_{\{t_i\} \times T_{-i}} (h^1 \circ \alpha) \; P_{j_1}(t_{j_1}) \; f,$$

so conflict-of-interest cannot hold. □

In this example I assume nothing about the set of outcomes or about the interim preferences of agents, so that the example holds for even the most natural choices of these parameters. In particular, the set of outcomes may be a set of allocations of private goods and interim preferences may be derived from expected utility calculations with strictly monotonic, selfish utility functions for each agent. The example shows that, when the set of states is uncountable, Corollary 1 applies only if the beliefs of agents are rather special in the sense that for each type t_i of agent i there exists a type t_j of agent $j \neq i$ whose conditional beliefs place positive probability on the lower-dimensional event $\{t_i\} \times T_{-i}$. Theorems 2 and 3 avoid this limitation, but they provide only a partial characterization of Bayesian implementability.

2.4 Extended Expected Utility

The standard formulation of the implementation problem specifies state-contingent utility functions over outcomes for agents that, together with the assumption that each agent's set of types is finite, induce complete interim preferences over social choice functions given by the magnitude of their conditional expected payoffs. I formally define the notion of an expected utility environment next, without the assumption of finitude.

Definition 9 *$\tilde{e} = (I, O, T, \mathcal{T}, \mu, \tilde{R})$ is an* **expected utility environment** *if there exists $u : I \times T \times O \to \mathbb{R}$ such that for all $i \in I$, all $\tilde{t}_i \in T_i$, and all $f, h \in F$,*

$$f \ \tilde{R}_i(\tilde{t}_i) \ h$$

if and only if

$$\int_T u_i(f(t)|t)\, d\mu_i(t|\tilde{t}_i) \geq \int_T u_i(h(t)|t)\, d\mu_i(t|\tilde{t}_i).$$

I have omitted some natural structure from this definition, including a σ-algebra on the set of outcomes and measurability of agents' state-contingent utility functions with respect to the product σ-algebra on O and T, but this is unimportant for what follows.

When T is uncountable and $\mathcal{T}$ is not equal to the power set of T, there will in general exist social choice functions for which the above integrals are not defined, and $\tilde{R}_i(t_i)$ will be incomplete. The results of Section 2.3 go through in any case, but the next example shows that this form of incompleteness has unsatisfactory consequences for the set of Bayesian equilibrium strategy profiles for some games. In particular, some strategy profiles will qualify as Bayesian equilibria simply because their conditional expected payoffs cannot be calculated and unilateral deviations cannot be compared to them according to $\tilde{R}$.

Example 2 For an expected utility environment $\tilde{e}$, assume that $n = 2$, that $O = \{0,1\} \times \{0,1\}$, that $T_1 = T_2 = [0,1]$, and that $\mathcal{T} = \mathcal{T}_1 * \mathcal{T}_2$, where $\mathcal{T}_1$ and $\mathcal{T}_2$ are equal to the Borel σ-algebra on $[0,1]$. Agent's conditional beliefs are derived from a common uniform prior on T. For all $t \in T$ and all $x = (x_1, x_2) \in O$, assume $u_1(x|t) = x_1$ and $u_2(x|t) = x_2$. Consider the following mechanism. Let $M_1 = M_2 = 2^{[0,1]} \times [0,1]$ with elements $\hat{m}_i = (\hat{t}_i, \hat{S}_i)$, and define the outcome function $g = (g_1, g_2)$ by

$$g_1(\hat{m}_1, \hat{m}_2) = \begin{cases} 1 & \text{if } \hat{t}_2 \in \hat{S}_1 \\ 0 & \text{else} \end{cases}$$

for all $\hat{m} \in M$, and similarly for g_2. The induced game of incomplete information is strategically simple. Each agent i should report $\hat{S}_i = [0,1]$, thereby receiving a payoff of 1 at every state. Now let S denote a non-measurable subset of

$[0, 1/2]$, and note that the strategy profile σ that has each agent i report i's true type and the set S at every type is a Bayesian equilibrium. This follows since $u_1(g(\sigma(.))|.)^{-1}(\{1\}) = T_1 \times S \notin \mathcal{T}$, so that agent 1 can calculate the integral $\int_T u_1(g(\sigma(t))|t)\, d\mu_1(t|\tilde{t}_1)$ at no type $\tilde{t}_1$. Therefore, there is no unilateral deviation $\tilde{\sigma}$ such that $g \circ (\tilde{\sigma}_1, \sigma_2)\ \tilde{P}_1(\tilde{t}_1)\ g \circ \sigma$. A similar argument for agent 2 establishes that $\sigma \in B_{(M,g,\tilde{e})}$. □

The predictions of Bayesian equilibrium in the above example are disturbing, since σ is qualified as a Bayesian equilibrium of the induced game simply because no type of any agent can compare the social choice function $g \circ \sigma$ to the social choice function induced by any unilateral deviation. Considerations of payoffs are irrelevant. In fact, both agents receive one unit of utility at less than half of the possible states, though each agent i could receive one unit of utility at every state simply by reporting $\hat{S}_i = [0, 1]$. The particular perverse predictions of Bayesian equilibrium in this example are due to my permissive definition of Bayesian equilibrium, but equally perverse predictions are generated by the restrictive definition. An intuitively obvious candidate for a Bayesian equilibrium in the above game is the strategy profile $\hat{\sigma}$ that has each agent i report i's true type and the set $[0, 1]$ at every type, but it is not an equilibrium under the restrictive definition of Bayesian equilibrium, because $\neg g \circ \hat{\sigma}\ \tilde{R}_1(\tilde{t}_1)\ g \circ (\tilde{\sigma}_1, \hat{\sigma}_2)$, where $\tilde{\sigma}_1$ is defined in the example. The problem is not the definition of Bayesian equilibrium, but rather the incompleteness of interim preferences given by $\tilde{R}$.

These issues might be dealt with by simply restricting each agent i to the use of a subset $\Sigma'_i \subseteq \Sigma_i(T_i)$ of admissible strategies such that each $u_i(g(\sigma'(.))|.)$ is $\mathcal{T}$-measurable for every $\sigma' \in \Sigma'$, but this merely transforms the problem to one of determining which subsets are admissible. If each agent's message space has a natural σ-algebra (and utility functions and outcome functions are appropriately measurable) then the admissible subsets might be just the measurable mappings from types to messages, but this approach is also problematic. Even if the notion of "natural" could be formalized, it is unlikely that there would always exist unique natural σ-algebras, and in case there were a unique natural σ-algebra it would be difficult to justify this restriction on the grounds of payoffs. Indeed, the next example shows that there are simple games with strategy profiles for which no type of any agent can calculate conditional expected payoffs, but for which each agent is clearly made worse off by unilaterally deviating. The assumption that such strategy profiles will never arise is ad hoc and particularly inappropriate in the context of Bayesian implementation.

Example 3 For an expected utility environment $\tilde{e}$, assume $n = 2$, $O = \{x, y, z\}$, $T_1 = T_2 = [0, 1]$, and $\mathcal{T} = \mathcal{T}_1 * \mathcal{T}_2$, where $\mathcal{T}_1$ and $\mathcal{T}_2$ are equal to the Borel σ-algebra on $[0, 1]$. Agents' conditional beliefs are derived from a common prior according to which, with probability $2/3$, states are drawn from a uniform distribution on the diagonal of T. For all $t \in T$, let $u_1(x|t) = u_2(x|t) = 1$, let $u_1(y|t) = u_2(y|t) = 3/4$, and let $u_1(z|t) = u_2(z|t) = 0$. Consider the following mechanism. Let $M_1 = M_2 = \{H, T\}$, and define the outcome function g by

$$g(\hat{m}) = \begin{cases} x & \text{if } \hat{m} = (H, H) \\ y & \text{if } \hat{m} = (T, T) \\ z & \text{else} \end{cases}$$

for all $\hat{m} \in M$. This mechanism induces a game of incomplete information in $\tilde{e}$ similar to the matching pennies game, the only difference being that matching tails to tails results in a slightly lower payoff than matching heads to heads. Let S be a non-measurable subset of $[0,1]$. Note that agents cannot calculate conditional expected payoffs from the strategy profile σ that has both agents reporting heads when their types are in S and tails otherwise, for $u_1(g(\sigma(.))|.)^{-1}(\{1\}) = S \times S \notin \mathcal{T}$, and similarly for agent 2. This strategy profile cannot, however, be discounted as an equilibrium, for it has the property that any agent who unilaterally deviates is clearly worse off. To see this, suppose type $s \in S$ of agent 1 reports tails instead of heads. The set of states at which agent 2 reports heads contains the singleton set $\{(s, s)\}$, which according to agent 1's conditional beliefs at type s, has probability $2/3$. The most optomistic approximation of the payoff of this deviation would presume that type s of agent 1 loses one unit of utility only at the state (s, s) and gains $3/4$ units of utility elsewhere, but even then it is apparent that agent 1 is worse off, since $(1/3)(3/4) < 2/3$. Now suppose type $s \in [0,1] \setminus S$ of agent 1 reports heads instead of tails. Again, the most optimistic approximation of the payoff to this deviation would presume that type s of agent 1 loses $3/4$ units of utility only at state (s, s), which has conditional probability $2/3$, and gains one unit of utility elsewhere. Again, agent 1 is worse off, since $1/3 < (2/3)(3/4)$. A similar argument for agent 2 establishes the claim. □

This example shows that comparisons between social choice functions are often possible even when conditional expected payoffs cannot be calculated. For another, less specific example, suppose that $u_i(f(.)|.)$ and $u_i(h(.)|.)$ are non-measurable but that there exists a $\mathcal{T}$-measurable function $v : T \to \mathbb{R}$ such that, for all $t \in T$, $u_i(f(t)|t) - u_i(h(t)|t) \leq v(t)$ and

$$\int_T v(t)\, d\mu_i(t|\tilde{t}_i) \leq 0.$$

Then it is clear that type $\tilde{t}_i$ of agent i should weakly prefer h to f. Unfortunately, this comparison will not in general complete $\tilde{R}_i(\tilde{t}_i)$. The extension of $\tilde{R}_i(\tilde{t}_i)$ to a complete preference relation on F requires comparisons that are not as clear but are nonetheless plausible. Such an extension should be interpreted as one among many possible postulates of behavior in environments with non-measurable subsets of types.

To this end, I define an extension of the integral to non-measurable functions $v : T \to \mathbb{R}$ by calculating the integrals of measurable functions that are above v but very close. I write $w \succeq v$ if, for all $t \in T$, $w(t) \geq v(t)$, and I formally define the upper integral of v with respect to $\mu_i(.|\tilde{t}_i)$ as

$$\int^* v(t)\, d\mu_i(t|\tilde{t}_i) = \inf\left\{ \int_T w(t)\, d\mu_i(t|\tilde{t}_i) \,\middle|\, w \succeq v, w\ \mathcal{T} - \text{mble} \right\}$$

for all $i \in I$, all $\tilde{t}_i$, and all bounded functions $v : T \to \mathbb{R}$. I next formally define an extended expected utility environment as an expected utility environment with interim preferences over social choice functions extended to complete preferences using the upper integral in a way that satisfies the technical requirements of Chapters 3 and 4 and admits an intuitive interpretation. The additional requirement that u is bounded is necessary to prove the existence of the upper integral.

Definition 10 $e^* = (I, O, T, \mathcal{T}, \mu, R^*)$ *is an* **extended expected utility environment** *if there exists* $u : I \times O \times T \to \mathbb{R}$ *bounded in absolute value such that for all* $i \in I$, *all* $\tilde{t}_i \in T_i$, *and all* $f, h \in F$,

$$f \; R_i^*(\tilde{t}_i) \; h$$

if and only if

$$\int^* \{u_i(f(t)|t) - u_i(h(t)|t)\} \, d\mu_i(t|\tilde{t}_i) \\ - \int^* \{u_i(h(t)|t) - u_i(f(t)|t)\} \, d\mu_i(t|\tilde{t}_i) \geq 0.$$

Thus, when the upper integral of one differential is non-negative and the integral of the other is non-positive, so that a preference is clear, $R_i^*(\tilde{t}_i)$ agrees with that preference. And when the upper integrals of both differentials have the same sign, f is weakly preferred to h according to $R_i^*(\tilde{t}_i)$ if and only if the upper integral of the first differential gives at least as strong an indication of preference as the integral of the second.

The upper integral has several properties that will be important for the sequel. I state these in the form of propositions and then summarize their implications for extended expected utility environments in Theorem 5. In the proofs of the following propositions, relegated to the appendix, I will often refer to a $\mathcal{T}$-measurable function $w : T \to \mathbb{R}$ such that $w \succeq v$ and $\int_T w(t)\, d\,\mu_i(t|\tilde{t}_i) = \int^* v(t)\, d\mu_i(t|\tilde{t}_i)$ as an upper approximation of v. Let V denote the set of functions $v : T \to \mathbb{R}$ bounded in absolute value.

Proposition 3 *For all $i \in I$, all $\tilde{t}_i \in T_i$, and all $v \in V$, $\int^* v(t)\, d\mu_i(t|\tilde{t}_i)$ exists.*

Proposition 4 *For all $i \in I$, all $\tilde{t}_i \in T_i$, and all $v \in V$, there exists a $\mathcal{T}$-measurable function $w \in V$ such that $w \succeq v$ and*

$$\int_T w(t)\, d\mu_i(t|\tilde{t}_i) = \int^* v(t)\, d\mu_i(t|\tilde{t}_i).$$

Proposition 5 *For all $i \in I$, all $\tilde{t}_i \in T_i$, and all $v_1, v_2 \in V$, $\mu_i^*(\{t \in T | v_1(t) \neq v_2(t)\}|\tilde{t}_i) = 0$ implies*

$$\int^* v_1(t)\, d\mu_i(t|\tilde{t}_i) = \int^* v_2(t)\, d\mu_i(t|\tilde{t}_i).$$

Proposition 6 *For all $i \in I$, all $\tilde{t}_i \in T_i$, and all $v_1, v_2 \in V$, $v_2 \succeq v_1$ implies*

$$\int^* v_2(t)\, d\mu_i(t|\tilde{t}_i) \geq \int^* v_1(t)\, d\mu_i(t|\tilde{t}_i).$$

Proposition 7 *For all $i \in I$, all $\tilde{t}_i \in T_i$, all $v_1, v_2 \in V$, and all $S \subseteq T$ with $\mu_i^*(S|\tilde{t}_i) > 0$,*

$$v_2 \succeq v_1 \text{ and } \forall t \in S\ v_2(t) > v_1(t)$$

implies

$$\int^* \{v_2(t) - v_1(t)\}\, d\mu_i(t|\tilde{t}_i) > 0.$$

Proposition 8 *For all $i \in I$, all $\tilde{t}_i \in T_i$, all $c \in \mathbb{R}$, and all $v \in V$,*

$$\int^* \{v(t) + c\}\, d\mu_i(t|\tilde{t}_i) = c + \int^* v(t)\, d\mu_i(t|\tilde{t}_i).$$

Proposition 9 *For all $i \in I$, all $\tilde{t}_i \in T_i$, all $a, b \in \mathbb{R}_+$, and all $v_1, v_2 : T \to \mathbb{R}$,*

$$\int^* \{av_1(t) + bv_2(t)\}\, d\mu_i(t|\tilde{t}_i) \leq a \int^* v_1(t)\, d\mu_i(t|\tilde{t}_i) + b \int^* v_2(t)\, d\mu_i(t|\tilde{t}_i).$$

Proposition 10 *For all $i \in I$, all $\tilde{t}_i \in T_i$, and all $v \in V$,*

$$-\int^* v(t)\, d\mu_i(t|\tilde{t}_i) \leq \int^* -v(t)\, d\mu_i(t|\tilde{t}_i).$$

Proposition 11 *For all $i \in I$ and all $\tilde{t}_i \in T_i$, if $\{v_k\}$ is a sequence of functions in V with $v_k \to v \in V$ then*

$$\int^* v(t)\, d\mu_i(t|\tilde{t}_i) \leq \lim \inf_{k\to\infty} \int^* v_k(t)\, d\mu_i(t|\tilde{t}_i).$$

Proposition 12 *For all $i \in I$ and all $\tilde{t}_i \in T_i$, if $\{v_k\}$ is a sequence of functions in V with $v_k \to v \in V$ uniformly then*

$$\int^* v(t)\, d\mu_i(t|\tilde{t}_i) \geq \lim \sup_{k\to\infty} \int^* v_k(t)\, d\mu_i(t|\tilde{t}_i).$$

Propositions 3 and 5 establish the existence of the upper integral for bounded functions from states to the real numbers and its invariance with respect to changes on sets of states with outer measure zero. Proposition 4 shows that every bounded function has an upper approximation, a property useful in the proofs of several propositions. Propositions 6 and 7 establish two monotonicity properties of the upper integral. Propositions 8, 9, and 10 establish several weak additivity properties of the upper integral. Finally, Propositions 11 and 12 establish two weak continuity properties of the upper integral that are not used in the sequel, but are stated here as a matter of interest.

The interim preferences $R_i^*(t_i)$ are *complete* if, for all $f, h \in F$, either $f\ R_i^*(t_i)\ h$ or $h\ R_i^*(t_i)\ f$ or both. I refer to $R_i^*(t_i)$ as *strictly monotonic* if two conditions hold. First, $u_i(f(.)|.) \succeq u_i(h(.)|.)$ implies $f\ R_i^*(t_i)\ h$, and second, $f\ P_i^*(t_i)\ h$ is implied by $u_i(f(.)|.) \succeq u_i(h(.)|.)$, $\mu_i^*(S|t_i) > 0$, and, for all $t \in S$, $u_i(f(t)|t) > u_i(h(t)|t)$.

Theorem 5 *Every extended expected utility environment e^* satisfies invariance and independence. Moreover, each $R_i^*(t_i)$ is complete and strictly monotonic.*

Proof: Proposition 5 easily implies that every extended expected utility environment satisfies invariance. That these environments also satisfy independence follows since, for all $i \in I$, all $f^1, f^2, h^1, h^2 \in F$, and all $S \subseteq T$, both functions $u_i((f^1/_S h^1)(.)|.) - u_i((f^2/_S h^1)(.)|.)$ and $u_i((f^1/_S h^2)(.)|.) - u_i((f^2/_S h^2)(.)|.)$ are equal to zero on the set S. Proposition 3 shows that, for all $i \in I$, all $\tilde{t}_i \in T_i$, and all $f, h \in F$,

$$\int^* \{u_i(f(t)|t) - u_i(h(t)|t)\}\, d\mu_i(t|\tilde{t}_i) - \int^* \{u_i(h(t)|t) - u_i(f(t)|t)\}\, d\mu_i(t|\tilde{t}_i)$$

is well-defined. That $R_i^*(\tilde{t}_i)$ is complete follows by the symmetry of the definition of weak preference. Proposition 6 shows that, for all $i \in I$, all $\tilde{t}_i \in T_i$, and all $f, h \in F$, $u_i(f(.)|.) \succeq u_i(h(.)|.)$ implies $f\ R_i^*(\tilde{t}_i)\ h$. Propositions 6 and 7 together show that $u_i(f(.)|.) \succeq u_i(h(.)|.)$, with strict inequality on a set of positive $\mu_i^*(.|\tilde{t}_i)$-outer measure, implies $f\ P_i^*(\tilde{t}_i)\ h$. This establishes the strict monotonicity of $R_i^*(t_i)$. ∎

2.5 Conclusion

In this chapter, I formulate the implementation problem for arbitrary environments, and I extend the results of Jackson (1991) to environments satisfying the weak conditions of invariance and independence. In particular, the set of states may be uncountable. I take agents' interim preferences as primitives, assuming only reflexivity, thereby obtaining as special cases results for Nash environments and for environments with preferences derived from conditional expected utility calculations. Of technical interest is the name recognition contest, used to prove the sufficiency result for monotonicity-no-veto, which possesses the crucial properties of the modulo game but which is effective for arbitrary sets of types.

I then prove the equivalence of Bayesian monotonicity and montonicity-no-veto in environments satisfying conflict-of-interest, a relationship noted by Jackson, with the obvious full characterization of Bayesian implementability as a corollary. I offer an example showing that Jackson's conflict-of-interest condition is quite strong when the set of states is uncountable, entailing an unnatural restriction on the beliefs of agents regarding the realized state, thereby limiting the applicability of the corollary in such environments.

I end with an example showing that interim preferences derived from conditional expected utility calculations do not accurately reflect the motivations of reasonable agents when the set of states is uncountable. The reason for this is that, in such environments, these interim preferences will generally be incomplete. Another example shows that the usual way of completing expected utility interim preferences, effectively restricting agents to use strategy profiles that induce comparable social choice functions, is unsatisfactory. Specifically, there exist games with strategy profiles for which no agent can calculate conditional expected payoffs but which are clearly stable. I offer an intuitively appealing and technically managable extension of expected utility interim preferences using upper integrals.

Chapter 3

A Full Characterization of Bayesian Implementability in Very General Environments

In Section 3.1, I add to the conceptual apparatus set forth in Chapter 2. In Section 3.2, I show that the conjunction of incentive compatibility and extended Bayesian monotonicity is necessary and sufficient for the implementation of social choice functions in environments satisfying invariance and interiority. In

Section 3.3, I show that the class of environments satisfying interiority is large. Assuming best-element-private values, interiority is satisfied by every continuous environment, a class that includes the lottery environments and the economic environments. Lottery environments are extended expected utility environments for which the set of outcomes is the set of probability measures over a measurable space of pure outcomes, and the economic environments include every economy for which agents have continuous, monotonic preferences over commodity bundles. Among these are the pure public goods economies, which do not typically satisfy conflict-of-interest and to which Theorem 4 does not apply. Section 3.4 concludes the chapter.

3.1 Notation and Definitions

Given an environment $e = (I, O, T, \mathcal{T}, \mu, R)$, a set $\hat{O}$ of outcomes is *interior* if for all $i \in I$, all $t_i \in T_i$, all $S \subseteq T$ with $\mu_i^*(S|t_i) > 0$, and all $f \in F$ with $f(S) \subseteq \hat{O}$, there exists $h \in F$ such that

$$f/_S h \; P_i(t_i) \; f.$$

For an interior set $\hat{O}$ of outcomes, let $\hat{F}$ denote the set of social choice functions $\hat{h}$ with $\hat{h}(T) \subseteq \hat{O}$. Roughly, $\hat{F}$ is the set of social choice functions that are not maximal for any type of any agent on any set of states with positive outer measure. If e is a private good exchange economy, for example, then $\hat{F}$ is the set of social choice functions that do not allocate the entire endowment to any agent at any set of states with positive outer measure.

In Section 3.2, I consider environments for which there exists an interior set of outcomes such that any strict preference for one social choice function over another can be replaced by a strict preference for an interior social choice function over the other. I state this condition formally next. Given an environment e and a cross $Q = \bigcup_{i \in I} Q_i \times T_{-i}$, let $Q_{\neq j} = \bigcup_{i \neq j} Q_i \times T_{-i}$, let $Q_i^+ = Q \setminus Q_{\neq i}$, and let $Q^+ = \bigcup_{i \in I} Q_i^+$. That is, Q^+ is the set of states in one and only one set Q_i. I will sometimes refer to $Q \setminus Q^+$ as the *center* of the cross, to Q^+ as the *arms* of the cross Q, and to Q_i^+ as the *ith arm* of the cross. Note that the Q_i^+ are pairwise disjoint.

Definition 11 *An environment e satisfies* **interiority** *if there exists an interior set $\hat{O}$ of outcomes such that for all $i \in I$, all $t_i \in T_i$, all $\alpha \in \Sigma(T)$, all $S \subseteq T$, and all $f^1, f^2, f^3, f^4 \in F$*

$$(f^1 \circ \alpha)/_S f^2 \; P_i(t_i) \; (f^3 \circ \alpha)/_S f^4$$

implies

$$\exists h \in \hat{F} \; (f^1 \circ \alpha)/_S (h \circ \alpha) \; P_i(t_i) \; (f^3 \circ \alpha)/_S f^4.$$

Suppose a type t_i of agent i prefers one social choice function to another and that both social choice functions use a deception α off a set S. Then interiority stipulates that there exists a social choice function h with interior values such that the agent's preference is unaffected when the first social choice function is spliced with $h \circ \alpha$ along S. Note that this condition involves something like the continuity of $P_i(t_i)$ and denseness of $\hat{F}$ in F. Moreover, since h uses α and f^2 does not, the condition hints of best-element-private values, stated precisely in Section 3.3 for extended expected utility environments, which requires roughly that the outcomes that give an agent high utility are independent of other agents' types. Theorem 9 explores these intuitions and shows that interiority is satisfied in a large class of environments.

For environments e satisfying interiority, the results of Section 3.2 characterize the implementation problems (e, f) for which there exists a solution. They are just the ones satisfying incentive compatibility and extended Bayesian monotonicity, defined formally next. Given an implementation problem (e, f), recall that $\tilde{F}_i$ denotes the set of social choice functions $\tilde{f}$ such that $\neg \tilde{f} \circ (\alpha_i, \tau_{-i})\ P_i(t_i)\ f$ for all $t_i \in T_i$ and all $\alpha_i \in \Sigma_i(T_i)$. In Section 2.1 I defined the splicing of one social choice function with another along a set of states. In this chapter I deal with the splicing of one social choice function with a collection $\{f_i\}_{i \in I}$ of social choice functions along disjoint collections $\{Q_i^+\}_{i \in I}$ of states, and the following shorthand will be useful. Let

$$f//_{Q_i^+} f_i = (\cdots ((f/_{Q_1^+} f_1)/_{Q_2^+} f_2) \cdots)/_{Q_n^+} f_n$$

denote the result of splicing f with each f_i along the corresponding Q_i^+. For a set $H_i \subseteq F$, a function $\alpha \in \Sigma(T)$, and a function $\psi_i \in \Sigma_i(H_i)$, let $\langle \psi_i, \alpha \rangle$ denote the social choice function defined by $\langle \psi_i, \alpha \rangle(t) = \psi_i(t_i)(\alpha(t))$ for all $t \in T$. Let $H = \times_{i \in I} H_i$, and let $\Sigma(H)$ denote the set of profiles $\psi = (\psi_1, \ldots, \psi_n)$ of functions $\psi_i \in \Sigma_i(H_i)$.

Definition 12 *An implementation problem* (e, f) *satisfies* **extended Bayesian monotonicity** *if there exists a collection* $\{H_i\}_{i\in I}$*, with each* $H_i \subseteq \tilde{F}_i$*, such that for all* $\alpha \in \Sigma(T)$*, all* $Q = \bigcup_{i\in I} Q_i \times T_{-i}$ *with* $\mu^*(Q \setminus Q^+) = 0$*, and all* $\psi \in \Sigma(H)$*,*

$$\forall i \in I, t_i \in T_i, h \in H_i, \hat{f} \in F$$

$$\neg (h \circ \alpha)/_{Q_{\neq i}} \hat{f}\ P_i(t_i)\ (f \circ \alpha)//_{Q_i^+} \langle \psi_i, \alpha \rangle$$

implies

$$(f \circ \alpha)//_{Q_i^+} \langle \psi_i, \alpha \rangle \in [f].$$

To see that this condition implies Bayesian monotonicity, take $\alpha \in D_{(e,f)}$, and let each $Q_i = \emptyset$ in the statement of extended Bayesian monotonicity. Then each $Q_i^+ = \emptyset$ and each $Q_{\neq i} = \emptyset$, so

$$(f \circ \alpha)//_{Q_i^+} \langle \psi_i, \alpha \rangle = f \circ \alpha \notin [f].$$

Then there exists $j \in I$, $\tilde{t}_j \in T_j$, and $\tilde{h} \in H_j \subseteq \tilde{F}_j$ such that

$$(\tilde{h} \circ \alpha)\ P_j(\tilde{t}_j)\ (f \circ \alpha),$$

which establishes Bayesian monotonicity.

For implementation problems (e, f) satisfying extended Bayesian monotonicity, there exists a collection $\{H_i\}_{i\in I}$ of sets of social choice functions with the following property. Suppose the center of a cross Q has zero μ^*-outer measure, the outcomes on each arm Q_i are given by social choice functions in H_i and α, and outcomes off Q are given by f and α. If no type of any agent i would prefer to have outcomes determined on the ith arm by any social choice function in

H_i and outcomes determined elsewhere by any social choice function in F, then the outcomes described above must coincide with the outcomes of f at all but a μ^*-outer measure zero set of states.

3.2 Characterization Results

The most powerful full characterization of Bayesian implementability is Jackson's (1991) result for environments satisfying conflict-of-interest, but he assumes that the set of states is finite. I show in Section 2.3 that Jackson's result is true even for uncountable sets of states, but that in such environments conflict-of-interest implies an unnatural restriction on the beliefs of agents. Since the advantages of conflict-of-interest are limited to a rather small class of environments, in this section I explore the possibilities of Bayesian implementation in environments satisfying interiority instead. I find that environments with this structure are quite tractable: the conjunction of incentive compatibility and extended Bayesian monotonicity is necessary and sufficient for the existence of a solution to an implementation problem. The next theorem establishes the necessity of these conditions, and does not rely on the assumption of interiority.

Theorem 6 *Assume e satisfies invariance. Then f is Bayesian implementable in e only if (e, f^*) satisfies incentive compatibility and extended Bayesian monotonicity for some $f^* \in [f]$.*

Proof: Assume that f is Bayesian implementable by the mechanism (M, g) and take $\sigma^* \in B_{(M,g,e)}$. Let $f^* = g \circ \sigma^* \in [f]$. Incentive compatibility of (e, f^*) follows from the proof of Theorem 2. For each i define

$$H_i = \{g \circ (m_i, \sigma^*_{-i}) | m_i \in M_i\},$$

where $g \circ (m_i, \sigma^*_{-i})$ is defined by $g \circ (m_i, \sigma^*_{-i})(t) = g(m_i, \sigma^*_{-i}(t_{-i}))$ for all $t \in T$. Now take $\alpha \in \Sigma(T)$, $Q = \bigcup_{i \in I} Q_i \times T_{-i}$ with $\mu^*(Q \setminus Q^+) = 0$, and $\psi \in \Sigma(H)$. Assume

$$\forall i \in I, t_i \in T_i, h \in H, \hat{f} \in F$$
$$\neg (h \circ \alpha)/_{Q_{\neq i}} \hat{f} \; P_i(t_i) \; (f^* \circ \alpha)//_{Q_i^+} \langle \psi_i, \alpha \rangle.$$

For all $i \in I$ and all $t_i \in T_i$, let $m_i^{t_i}$ satisfy $\psi_i(t_i) = g \circ (m_i^{t_i}, \sigma^*_{-i})$. Then define $\hat{\sigma} \in \Sigma(M)$ by

$$\hat{\sigma}_i(t_i) = \begin{cases} m_i^{t_i} & \text{if } t_i \in Q_i \\ \sigma_i^*(\alpha_i(t_i)) & \text{else,} \end{cases}$$

for all $i \in I$ and all $t_i \in T_i$. To see that

$$(f^* \circ \alpha)//_{Q_i^+} \langle \psi_i, \alpha \rangle \in [g \circ \hat{\sigma}], \tag{3.1}$$

take $t \notin Q \setminus Q^+$. First suppose $t \in Q^+$. Since the Q_i^+ are pairwise disjoint, there is exactly one $j \in I$ such that $t \in Q_j^+$. Then $\hat{\sigma}_j(t_j) = m_i^{t_j}$ and $\hat{\sigma}_{-j}(t_{-j}) = \sigma^*_{-j}(\alpha_{-j}(t_{-j}))$, so

$$g(\hat{\sigma}(t)) = g(m^{t_j}, \sigma^*_{-j}(\alpha_{-j}(t_{-j})))$$

$$\begin{aligned} &= (g \circ (m^{t_j}, \sigma^*_{-j}))(\alpha(t)) \\ &= \psi_j(t_j)(\alpha(t)) \\ &= \langle \psi_j, \alpha \rangle (t) \\ &= ((f^* \circ \alpha)//_{Q_i^+} \langle \psi_i, \alpha \rangle)(t). \end{aligned}$$

Now suppose $t \in T \setminus Q$. Then $\hat{\sigma}(t) = \sigma^*(\alpha(t))$ and

$$g(\hat{\sigma}(t)) = g(\sigma^*(\alpha(t))) = f^*(\alpha(t)) = (f^* \circ \alpha)//_{Q_i^+} \langle \psi_i, \alpha \rangle (t),$$

so

$$\{t \in T | (g \circ \hat{\sigma})(t) \neq ((f^* \circ \alpha)//_{Q_i^+} \langle \psi_i, \alpha \rangle)(t)\} \subseteq Q \setminus Q^+.$$

The claim follows, since $\mu^*(Q \setminus Q^+) = 0$.

To see that $\hat{\sigma} \in B_{(M,g,e)}$, take $j \in I$, $\tilde{t}_j \in T_j$, and $\tilde{\sigma}_j \in \Sigma_j(M_j)$. By invariance, it follows that

$$\neg g \circ (\tilde{\sigma}_j, \hat{\sigma}_{-j}) \; P_j(\tilde{t}_j) \; g \circ \hat{\sigma}$$

if and only if

$$\neg g \circ (\tilde{m}_j, \hat{\sigma}_{-j}) \; P_j(\tilde{t}_j) \; g \circ \hat{\sigma}$$

where $\tilde{m}_j = \tilde{\sigma}_j(\tilde{t}_j)$. Let $I_j(t) = \{i \neq j | t_i \in Q_i\}$, $J_j(t) = \{i \neq j | t_i \notin Q_i\}$, and define $h \in H_j$ and $\hat{f} \in F$ by

$$\begin{aligned} h(t) &= g \circ (\tilde{m}_j, \sigma^*_{-j})(t) \\ \hat{f}(t) &= g \circ (\tilde{m}_j, (m_i^{t_i})_{i \in I_j(t)}, (\sigma_i^*(\alpha_i(t_i)))_{i \in J_j(t)}) \end{aligned}$$

for all $t \in T$. It is straightforward to check that

$$g \circ (\tilde{m}_j, \hat{\sigma}_{-j}) = (h \circ \alpha)/_{Q_{\neq j}} \hat{f}. \tag{3.2}$$

Then invariance, (3.1), (3.2), and

$$\neg(h \circ \alpha)/_{Q_{\neq j}}\hat{f}\ P_j(\tilde{t}_j)\ (f^* \circ \alpha)//_{Q_i^+}\langle \psi_i, \alpha \rangle$$

imply

$$g \circ (\tilde{m}_j, \hat{\sigma}_{-j})\ P_j(\tilde{t}_j)\ g \circ \hat{\sigma}.$$

Therefore, $\hat{\sigma} \in B_{(M,g,e)}$. Since (M, g) Bayesian implements f, it follows that $g \circ \hat{\sigma} \in [f]$. Finally,

$$(f^* \circ \alpha)//_{Q_i^+}\langle \psi_i, \alpha \rangle \sim g \circ \hat{\sigma} \sim f \sim f^*$$

and Proposition 2 imply

$$(f^* \circ \alpha)//_{Q_i^+}\langle \psi_i, \alpha \rangle \in [f^*],$$

establishing that (e, f^*) satisfies extended Bayesian monotonicity. ∎

The next theorem establishes the sufficiency of incentive compatibility and extended Bayesian monotonicity in environments satisfying invariance and interiority. Interiority is used to show that the center of a certain cross has μ^*-outer measure zero, so that extended monotonicity may be applied. This is done by having agents play the name recognition contest, constructed in Section 2.2, and allowing the winner to impose on the center of the cross the outcomes of any social choice function with interior values. Interiority implies that for no type of any agent is there such a social choice function that does best when the center of the cross has positive outer measure. No conflict of interest is needed.

Theorem 7 *Assume e satisfies invariance and interiority. Then f is Bayesian implementable in e if (e, f^*) satisfies incentive compatibility and extended Bayesian monotonicity for some $f^* \in [f]$.*

Proof: It suffices to find a mechanism (M, g) such that $B_{(M,g,e)} \neq \emptyset$ and $g \circ \sigma \in [f^*]$ for all $\sigma \in B_{(M,g,e)}$, since this implies $g \circ \sigma \in [f]$. Have each agent i report $\hat{m} = (\hat{t}_i, \hat{f}_i, \hat{h}_i, \hat{l}_i, \hat{k}_i, \hat{p}_i, \hat{c}_i) \in M_i = T_i \times \hat{F} \times H_i \times \{0,1\} \times \{0,1\} \times \{0,1\} \times C_i$, where (C, w) denotes the name recognition contest, constructed in Section 2.2. I will sometimes represent a strategy $\hat{\sigma}_i \in \Sigma_i(M_i)$ by the sequence $(\hat{\alpha}_i, \hat{\phi}_i, \hat{\psi}_i, \hat{\lambda}_i, \hat{\kappa}_i, \hat{\pi}_i, \hat{\gamma}_i)$ of component functions. Partition M into the following sets

$$
\begin{aligned}
M^{i,i} &= \{\hat{m} \in M | \hat{l}_i = 1, \hat{l}_{-i} = \hat{k}_{-i} = \hat{p}_{-i} = 0\} \\
M^{i,j} &= \{\hat{m} \in M | i \neq j, \hat{p}_i > 0, \hat{l}_j + \hat{k}_j + \hat{p}_j > 0, \hat{l}_{-i,j} = \hat{k}_{-i,j} = \hat{p}_{-i,j} = 0\} \\
M^{\circ} &= \{\hat{m} \in M | \exists i_1, i_2, i_3 \in I, i_1 \neq i_2 \neq i_3 \neq i_1, \hat{k}_{i_1} > 0, \\
&\quad \hat{l}_{i_2} + \hat{k}_{i_2} + \hat{p}_{i_2} > 0, \hat{l}_{i_3} + \hat{k}_{i_3} + \hat{p}_{i_3} > 0\} \\
M^* &= M \setminus \left(M^{\circ} \cup \bigcup_{i,j \in I} M^{i,j} \right),
\end{aligned}
$$

and define the outcome function as

$$
g(\hat{m}) = \begin{cases} \hat{h}_i(\hat{t}) & \text{if } \hat{m} \in M^{i,i} \\ \hat{f}_{w(\hat{c})}(\hat{t}) & \text{if } \hat{m} \in M^{\circ} \cup \bigcup_{i \neq j} M^{i,j} \\ f^*(\hat{t}) & \text{if } \hat{m} \in M^*. \end{cases}
$$

The mechanism is somewhat different than the one used to prove Theorem 3, because now agents report a single entry in the name recognition contest and three integers: an aggressive integer $\hat{k}_i$, a passive integer $\hat{l}_i$, and an intermediate integer $\hat{p}_i$. An agent i's passive integer allows the agent to impose any social choice function in H_i whenever no other agent reports a positive integer; i's intermediate integer triggers the name recognition contest between i and j whenever exactly one other agent j reports a positive integer; and i's aggressive integer triggers the name recognition contest whenever two or more other agents report positive integers. The proof consists of two steps. The first step is to establish the existence of $\sigma^* \in B_{(M,g,e)}$ such that $g \circ \sigma^* \in [f]$, and the second step is to show that the equality holds for all $\sigma \in B_{(M,g,e)}$.

Step 1 : Consider the strategy profile σ^* defined by $\sigma_i^*(t_i) = (t_i, \bar{f}_i, \bar{h}_i, 0, 0, 0, \bar{c}_i)$ for all $i \in I$ and all $t_i \in T_i$, where $\bar{f}_i \in \hat{F}$, $\bar{h}_i \in H_i$, and $\bar{c}_i \in C_i$ are arbitrary constants. To see that $\sigma^* \in B_{(M,g,e)}$, take any $i \in I$ and $\tilde{\sigma}_i \in \Sigma_i(M_i)$, and partition T_i into $S_i^1 = \{t_i \in T_i | \tilde{\lambda}(t_i) = 1\}$ and $S_i^2 = T_i \setminus S_i^1$. For $t_i \in S_i^2$,

$$\neg g \circ (\tilde{\sigma}_i, \sigma_{-i}^*) \; P_i(t_i) \; g \circ \sigma^*$$

if and only if

$$\neg f^* \circ (\tilde{\alpha}_i, \tau_{-i}) \; P_i(t_i) \; f^*$$

which holds since (e, f^*) satisfies incentive compatibility. For $t_i \in S_i^1$, invariance implies that

$$\neg g \circ (\tilde{\sigma}_i, \sigma_{-i}^*) \; P_i(t_i) \; g \circ \sigma^*$$

if and only if

$$\neg \tilde{h}_i \circ (\tilde{\alpha}_i, \tau_{-i}) \; P_i(t_i) \; f^*$$

which holds since $\tilde{h}_i = \tilde{\psi}_i(t_i) \in H_i \subseteq \tilde{F}_i$. Therefore, $\sigma^* \in B_{(M,g,e)}$, and clearly $\mu^*(\{t \in T | g(\sigma^*(t)) \neq f^*(t)\}) = \mu^*(\emptyset) = 0$.

Step 2 : Now take $\sigma^* \in B_{(M,g,e)}$. Let $Q_i = \{t_i \in T_i | \lambda_i^*(t_i) + \kappa_i^*(t_i) + \pi_i^*(t_i) > 0\}$, let $Q = \bigcup_{i \in I} Q_i \times T_{-i}$, let $Q_{\neq i} = \bigcup_{j \neq i} Q_j \times T_{-j}$, let $Q_i^+ = Q \setminus Q_{\neq i}$, and let $Q^+ = \bigcup_{i \in I} Q_i^+$. To see that $\mu^*(Q \setminus Q^+) = 0$, suppose not. Then there exists $j \in I$ and $\tilde{t}_j \in T_j$ such that $\mu_j^*(Q \setminus Q^+ | \tilde{t}_j) > 0$. Note that

$$g(\sigma^*(t)) = \phi^*_{w(\gamma^*(t))}(\alpha^*(t)) \in \hat{O}$$

for all $t \in Q \setminus Q^+$, where $\hat{O}$ is the set of interior outcomes in the statement of interiority. Then there exists $h \in F$ such that

$$(g \circ \sigma^*)/_{Q \setminus Q^+} h \; P_j(\tilde{t}_j) \; g \circ \sigma^*,$$

and by interiority, there exists $\tilde{f} \in \hat{F}$ such that

$$(g \circ \sigma^*)/_{Q \setminus Q^+} (\tilde{f} \circ \alpha^*) \; P_j(\tilde{t}_j) \; g \circ \sigma^*. \tag{3.3}$$

I claim that type $\tilde{t}_j$ of agent j can gain from a unilateral deviation, contradicting the assumption that $\sigma^* \in B_{(M,g,e)}$. Consider the strategy $\tilde{\sigma}_j \in \Sigma_j(M_j)$ defined by

$$\tilde{\sigma}_j(t_j) \quad = \quad (\alpha_j^*(t_j), \tilde{f}, \psi_j^*(t_j), \lambda_j^*(t_j), 1, \tilde{\pi}_j(t_j), \tilde{c}_j)$$

for all $t_j \in T_j$, where $\tilde{c}_j$ guarantees that j wins the name recognition contest at all states and $\tilde{\pi}_j(t_j) = \lambda_j^*(t_j)$. Suppose for the moment that this strategy changes the outcomes of $g \circ \sigma^*$ only on the set $Q \setminus Q^+$, where j wins the name recognition contest and $\tilde{f}$ determines the outcome with α^*. This implies that

$$g \circ (\tilde{m}_j, \sigma^*_{-j}) = (g \circ \sigma^*)/_{Q \setminus Q^+}(\tilde{f} \circ \alpha^*),$$

which yields

$$g \circ (\tilde{m}_j, \sigma^*_{-j}) \; P_j(\tilde{t}_j) \; g \circ \sigma^*,$$

after substituting into (3.3). This contradiction shows that $\mu^*(Q \setminus Q^+) = 0$.

To see that $\tilde{\sigma}_j$ does indeed change the outcomes of $g \circ \sigma^*$ only on the set $Q \setminus Q^+$, take $t \in T$. There are three cases to consider. First, suppose $t \in T \setminus Q$. Then $(\tilde{\lambda}_j(t_j), \lambda^*_{-j}(t_{-j})) = \lambda^*(t) = 0$, $(\tilde{\kappa}_j(t_j), \kappa^*_{-j}(t_{-j})) = (1, 0, \ldots, 0)$, and $(\tilde{\pi}_j(t_j), \pi^*_{-j}(t_{-j})) = (\lambda^*_j(t_j), 0, \ldots, 0) = 0$, so

$$\begin{aligned} g(\tilde{\sigma}_j(t_j), \sigma^*_{-j}(t_{-j})) &= f^*(\tilde{\alpha}_j(t_j), \alpha^*_{-j}(t_{-j})) \\ &= f^*(\alpha^*(t)) \\ &= g(\sigma^*(t)), \end{aligned}$$

as desired. Second, suppose $t \in Q^+$, so there is exactly one agent i who reports a positive integer. If $i = j$ and $\lambda^*_j(t_j) = 1$ then $(\tilde{\lambda}_j(t_j), \lambda^*_{-j}(t_{-j})) = \lambda^*(t) = (1, 0, \ldots, 0)$, $\kappa^*_{-j}(t_{-j}) = 0$, and $\pi^*_{-j}(t_{-j}) = 0$, so

$$\begin{aligned} g(\tilde{\sigma}_j(t_j), \sigma^*_{-j}(t_{-j})) &= \tilde{\psi}_j(t_j)(\tilde{\alpha}_j(t_j), \alpha^*_{-j}(t_{-j})) \\ &= \psi^*_j(t_j)(\alpha^*(t)) \\ &= g(\sigma^*(t)), \end{aligned}$$

as desired. If $i = j$ and $\lambda_j^*(t_j) = 0$ then $(\tilde{\lambda}_j(t_j), \lambda_{-j}^*(t_{-j})) = \lambda^*(t) = 0$, and $\kappa_{-j}^*(t_{-j}) = \pi_{-j}^*(t_{-j}) = 0$, so

$$\begin{aligned} g(\tilde{\sigma}_j(t_j), \sigma_{-j}^*(t_{-j})) &= f^*(\tilde{\alpha}_j(t_j), \alpha_{-j}^*(t_{-j})) \\ &= f^*(\alpha^*(t)) \\ &= g(\sigma^*(t)), \end{aligned}$$

as desired. If $i \neq j$ and $\lambda_i^*(t_i) = 1$ then $(\tilde{\lambda}_j(t_j), \lambda_{-j,i}^*(t_{-j,i})) = \lambda_{-i}^*(t_{-i}) = 0$, $(\tilde{\kappa}_j(t_j), \kappa_{-j,i}^*(t_{-j,i})) = \kappa_{-i}^*(t_{-i}) = 0$, and $(\tilde{\pi}_j(t_j), \pi_{-j,i}^*(t_{-j,i})) = (\lambda_j^*(t_j), 0, \ldots, 0) = 0$, so

$$\begin{aligned} g(\tilde{\sigma}_j(t_j), \sigma_{-j}^*(t_{-j})) &= \psi_i^*(t_i)(\tilde{\alpha}_j(t_j), \alpha_{-j}^*(t_{-j})) \\ &= \psi_i^*(t_i)(\alpha^*(t)) \\ &= g(\sigma^*(t)), \end{aligned}$$

as desired. If $j \neq i$ and $\lambda_i^*(t_i) = 0$ then $(\tilde{\lambda}_j(t_j), \lambda_{-j}^*(t_{-j}) = 0$, so

$$\begin{aligned} g(\tilde{\sigma}_j(t_j), \sigma_{-j}^*(t_{-j})) &= f^*(\tilde{\alpha}_j(t_j), \alpha_{-j}^*(t_{-j})) \\ &= f^*(\alpha^*(t)) \\ &= g(\sigma^*(t)), \end{aligned}$$

as desired. Third, suppose $t \in Q \setminus Q^+$. Then there exists $i \neq j$ such that $\lambda_i^*(t_i) + \kappa_i^*(t_i) + \pi_i^*(t_i) > 0$, so $\tilde{\kappa}_j(t_j) = 1$ and $\tilde{\gamma}_j(t_j) = \tilde{c}_j$ imply that

$$\begin{aligned} g(\tilde{\sigma}_j(t_j), \sigma_{-j}^*(t_{-j})) &= \tilde{\phi}_j(t_j)(\tilde{\alpha}_j(t_j), \alpha_{-j}^*(t_{-j})) \\ &= \tilde{f}(\alpha^*(t)), \end{aligned}$$

as desired.

Next, I show that

$$(f^* \circ \alpha^*)//_{Q_i^+}\langle \psi_i^*, \alpha^* \rangle \in [g \circ \sigma^*]. \tag{3.4}$$

To this end take $t \in Q^+$. Since the Q_i^+ are disjoint, there is exactly one j such that $t \in Q_j^+$, and then

$$\begin{aligned} g(\sigma^*(t)) &= \psi_j^*(t_j)(\alpha^*(t)) \\ &= \langle \psi_j^*, \alpha^* \rangle(t) \\ &= ((f^* \circ \alpha^*)//_{Q_i^+}\langle \psi_i^*, \alpha^* \rangle)(t). \end{aligned}$$

Now take $t \in T \setminus Q$. Then

$$g(\sigma^*(t)) = f^*(\alpha^*(t)),$$

and the claim follows since $\mu^*(Q \setminus Q^+) = 0$. To apply extended Bayesian monotonicity, I must show that

$$\begin{aligned} &\forall i \in I, t_i \in T_i, h \in H_i, \hat{f} \in F \\ &\qquad \neg(h \circ \alpha^*)/_{Q_{\neq i}}\hat{f}\ P_i(t_i)\ (f^* \circ \alpha^*)//_{Q_i^+}\langle \psi_i^*, \alpha^* \rangle. \end{aligned}$$

To this end, take $j \in I$, $\tilde{t}_j \in T_j$, $h \in H_j$, and $\hat{f} \in F$, and suppose that

$$(h \circ \alpha^*)/_{Q_{\neq j}}\hat{f}\ P_j(\tilde{t}_j)\ (f^* \circ \alpha^*)//_{Q_i^+}\langle \psi_i^*, \alpha^* \rangle. \tag{3.5}$$

I claim that this allows type $\tilde{t}_j$ of agent j to gain from a unilateral deviation, contradicting the assumption that $\sigma^* \in B_{(M,g,e)}$.

Note that interiority and (3.5) imply the existence of $\tilde{f} \in \hat{F}$ such that

$$(h \circ \alpha^*)/_{Q_{\neq j}}(\tilde{f} \circ \alpha^*)\ P_j(\tilde{t}_j)\ (f^* \circ \alpha^*)//_{Q_i^+}\langle \psi_i^*, \alpha^* \rangle. \tag{3.6}$$

Now consider the strategy $\tilde{\sigma}_j \in \Sigma_j(M_j)$ defined by

$$\tilde{\sigma}_j = (\alpha_j^*(t_j), \tilde{f}, h, 1, 1, 1, \tilde{c}_j)$$

for all $t_j \in T_j$, where $\tilde{c}_j$ guarantees that j wins the name recognition contest at all states. This strategy changes the outcomes of $g \circ \sigma^*$ on $Q_{\neq j}$, where j wins the name recognition contest and imposes the outcomes of $\tilde{f}$ using the agents' reports α^*, and it changes outcomes off $Q_{\neq j}$, where j imposes the outcomes of h using α^*. That is,

$$g \circ (\tilde{m}_j, \sigma_{-j}^*) = (h \circ \alpha^*)/_{Q_{\neq j}}(\tilde{f} \circ \alpha^*). \tag{3.7}$$

Then (3.4), (3.6), (3.7), and invariance imply

$$g \circ (\tilde{m}_j, \sigma_{-j}^*)\ P_j(\tilde{t}_j)\ g \circ \sigma^*,$$

a contradiction. Extended Bayesian monotonicity then implies that

$$(f^* \circ \alpha^*)//_{Q_i^+}\langle \psi_i^*, \alpha^* \rangle \in [f^*]. \tag{3.8}$$

Finally, (3.4), (3.8), and Proposition 2 imply that $g \circ \sigma^* \in [f^*]$, as desired. ∎

3.3 Interiority

Although interiority is not implied by conflict-of-interest, in this section I show that, assuming best-element-private values, many interesting extended expected utility environments satisfying conflict-of-interest also satisfy interiority. Interiority is satisfied, for example, in economies with a private good and monotonic preferences. But interiority is also satisfied in environments that do not typically satisfy conflict-of-interest. It holds in pure public good economies with free disposal, and it holds in lottery environments, for which the set of outcomes is the set of probability measures over a set of pure outcomes. The structure common to economic environments and lottery environments that makes them especially tractable is that agents' state-contingent utility functions are continuous and, in a certain sense, monotonic: agent's best outcomes, if they exist, are extreme points of the set of outcomes. More to the point, the best outcomes of any type of any agent can be approximated by outcomes that are best for no type of any agent. Extended expected utility environments in which agents' best outcomes range over a Euclidean space of outcomes, as in the spatial political model (Ordeshook, 1986), do not satisfy interiority. The results for economic environments and lottery environments follow from a more general theorem for continuous environments, defined next.

Definition 13 e^* *is a* **continuous environment** *if it is an extended expected utility environment,* O *is a topological space, each* $u_i(.|t)$ *is continuous, and*

$$O^* = \{x \in O | \forall i \in I, t \in T \ \exists y \in O \ u_i(y|t) > u_i(x|t)\}$$

is dense in O.

I next define best-element-private values for extended expected utility environments with topological spaces of outcomes, which include the continuous environments. Versions of the condition have appeared in the Bayesian implementation literature, (see Palfrey and Srivastava, 1993) but for purposes of comparison I also supply the more common condition of private values.

Definition 14 *An environment e satisfies* **best-element-private values** *if for all* $i \in I$, *all* $t_i \in T_i$, and all $k \in \mathbb{Z}_{++}$ *there exists a non-empty open set* $B_k \subseteq O$ *such that for all* $t_{-i} \in T_{-i}$ *and all* $x \in B_k$

$$u_i(x|t) > \sup_{y \in O} u_i(y|t) - \frac{1}{k}.$$

Definition 15 *An environment e satisfies* **private values** *if for all* $i \in I$ *and all* $t, t' \in T$,

$$t_i = t_i' \text{ implies } u_i(.|t) = u_i(.|t').$$

Roughly, best-element-private values requires that, given an agent and a type, there exist open sets of outcomes that give the agent arbitrarily high utility, independent of other agents' types. Unlike its relatives in the literature, best-element-private values does not require for each type of each agent the existence of a best outcome independent of other agents' types, but my version of the condition relies on topological properties of the set of outcomes and is therefore not strictly weaker. Private values requires simply that each agent's state-contingent utility function is independent of other agents' types. For such environments, I will write $u_i(.|.)$ as a function of outcomes and i's type only.

Theorem 8 shows how the continuous environments satisfying private values are related to those satisfying best-element-private values.

Theorem 8 *Every continuous environment satisfies private values only if it satisfies best-element-private values.*

Proof: Take $i \in I$, $t_i \in T_i$, and a positive integer k. By the definition of a supremum, there exists $x \in O$ with $u_i(x|t_i) > \sup_{y \in O} u_i(y|t_i) - 1/k$. Since O^* is dense in O, there is a net $\{x_\alpha\}$ in O^* converging to x. Since $u_i(.|t_i)$ is continuous, there exists α such that

$$\sup_{y \in O} u_i(y|t_i) > u_i(x_\alpha|t_i) > \sup_{y \in O} u_i(y|t_i) - \frac{1}{k}.$$

Continuity of $u_i(.|t_i)$ then implies that these inequalities hold for an open set $B_k \subseteq O$ owning x_α. Since private values is satisfied, these inequalities are independent of t_{-i}. ∎

The next theorem shows how the continuous environments for which best-element-private values is satisfied are related to those satisfying interiority. This result, together with Theorems 6 and 7, yields as a corollary for these environments a full characterization of the implementation problems with solutions. Note that the proof of Theorem 9 relies on the axiom of choice.

Theorem 9 *If a continuous environment satisfies best-element-private values then it satisfies interiority.*

Proof: Let $\hat{O} = O^*$. To see that $\hat{O}$ is an interior set of outcomes, take $\hat{f}, \hat{h} \in F$ with $\hat{h}(T) \subseteq \hat{O}$, $i \in I$, $t_i \in T_i$, and $S \subseteq T$ with $\mu_i^*(S|t_i) > 0$. By the choice of $\hat{h}$, for all $t \in S$ there exists $x_t \in O$ such that

$$u_i(x_t|t) > u_i(\hat{h}(t)|t).$$

Since $u_i(.|t)$ is continuous, for all $t \in S$ there exists an open set $B_t \subseteq O$ such that this inequality holds for every element of B_t. Since $\hat{O}$ dense in O, each B_t must contain an element of $\hat{O}$. Using the axiom of choice, it is possible to define the social choice function h such that, for all $t \in S$, $h(t)$ is such an outcome. Then, by Theorem 5

$$\hat{f}/_S h \; P_i^*(t_i) \; \hat{f}/_S \hat{h},$$

as desired.

Now take $i \in I$, $\tilde{t}_i \in T_i$, $\alpha \in \Sigma(T)$, $S \subseteq T$, and $f^1, f^2, f^3, f^4 \in F$ such that

$$(f^1 \circ \alpha)/_S f^2 \; P_i^*(\tilde{t}_i) \; (f^3 \circ \alpha)/_S f^4.$$

Define $v : T \to \mathbb{R}$ by

$$v(t) = u_i(((f^1 \circ \alpha)/_S f^2)(t)|t) - u_i(((f^3 \circ \alpha)/_S f^4)(t)|t)$$

for all $t \in T$, and define $v_k : T \to \mathbb{R}$ by $v_k(t) = v(t) - 1/k$ for all $t \in T$. Note that, for all $t \in S$ and all $k \in \mathbb{Z}_{++}$,

$$v_k(t) = v(t) - \frac{1}{k} \leq \sup_{y \in O} u_i(y|t) - u_i(f^4(t)|t) - \frac{1}{k}.$$

Best-element-private values then implies that there exists an open set $B_k \subseteq O$ such that, for all $x \in B_k$ and all $t \in S$ with $t_i = \tilde{t}_i$,

$$u_i(x|t) - u_i(f^4(t)|t) > v_k(t),$$

and since B_k is open and $\hat{O}$ is dense, there exists $x_k \in \hat{O} \cap B_k$ for which this inequality holds. From Theorem 5 it follows that, for all k,

$$\begin{aligned}&\int^* \{u_i(((f^1 \circ \alpha)/_S \overline{x}_k)(t)|t) - u_i(((f^3 \circ \alpha)/_S f^4)(t)|t)\}\, d\mu_i(t|\tilde{t}_i)\\ &\quad > \int^* v_k(t)\, d\mu_i(t|\tilde{t}_i)\end{aligned}$$

and

$$\begin{aligned}&\int^* \{u_i(((f^3 \circ \alpha)/_S f^4)(t)|t) - u_i(((f^1 \circ \alpha)/_S \overline{x}_k)(t)|t)\}\, d\mu_i(t|\tilde{t}_i)\\ &\quad \leq \int^* -v_k(t)\, d\mu_i(t|\tilde{t}_i),\end{aligned}$$

where I write $\overline{x}_k$ for the constant social choice function that picks x_k at every state. These inequalities imply that, for all k,

$$\begin{aligned}&\int^* \{u_i(((f^1 \circ \alpha)/_S \overline{x}_k)(t)|t) - u_i(((f^3 \circ \alpha)/_S f^4)(t)|t)\}\, d\mu_i(t|\tilde{t}_i)\\ &\qquad - \int^* \{u_i(((f^3 \circ \alpha)/_S f^4)(t)|t) - u_i(((f^1 \circ \alpha)/_S \overline{x}_k)(t)|t)\}\, d\mu_i(t|\tilde{t}_i)\\ &\quad > \int^* v_k(t)\, d\mu_i(t|\tilde{t}_i) - \int^* -v_k(t)\, d\mu_i(t|\tilde{t}_i)\\ &\quad = \int^* \{u_i(((f^1 \circ \alpha)/_S f^2)(t)|t)\\ &\qquad -u_i(((f^3 \circ \alpha)/_S f^4)(t)|t)\}\, d\mu_i(t|\tilde{t}_i)\\ &\qquad - \int^* \{u_i(((f^3 \circ \alpha)/_S f^4)(t)|t)\\ &\qquad -u_i(((f^1 \circ \alpha)/_S f^2)(t)|t)\}\, d\mu_i(t|\tilde{t}_i) - \frac{2}{k},\end{aligned}$$

where the last equality follows from Proposition 8. Since

$$(f^1 \circ \alpha)/_S f^2 \; P_i^*(\tilde{t}_i) \; (f^3 \circ \alpha)/_S f^4,$$

it follows that

$$\begin{aligned}\liminf_{k\to\infty} & \int^* \{u_i(((f^1 \circ \alpha)/_S \overline{x}_k)(t)|t) - u_i(((f^3 \circ \alpha)/_S f^4)(t)|t)\} \, d\mu_i(t|\tilde{t}_i) \\ & - \int^* \{u_i(((f^3 \circ \alpha)/_S f^4)(t)|t) - u_i(((f^1 \circ \alpha)/_S \overline{x}_k)(t)|t)\} \, d\mu_i(t|\tilde{t}_i) \\ > \;\; & 0.\end{aligned}$$

In particular, this inequality holds for some $x_k \in \hat{O}$. This establishes interiority.

∎

Corollary 2 *Assume e^* is a continuous environment satisfying best-element-private values. Then f is Bayesian implementable in e^* if and only if (e^*, f^*) satisfies incentive compatibility and extended Bayesian monotonicity for some $f^* \in [f]$.*

The continuous environments, as the name suggests, include many environments of interest. I next give a general definition of economic environments that encompasses private good production economies and mixed production economies. Assuming free disposal, it also encompasses the pure public good economies, which do not typically satisfy conflict-of-interest. The price of this generality is, however, the complexity of the definition. Let $\mathbb{R}^\Lambda$ denote Euclidean space with basis $\{e^\lambda | \lambda \in \Lambda\}$, where each e^λ is a unit coordinate vector.

For subsets $O \subseteq \mathbb{R}^\Lambda$ and $\Lambda_i \subseteq \Lambda$ let O^{Λ_i} denote the projection of O onto the subspace generated by $\{e^\lambda | \lambda \in \Lambda_i\}$. For $x \in O$, let $x^{\Lambda_i} \in O^{\Lambda_i}$ denote the projection of x onto O^{Λ_i}. For $x, y \in O$, I write $x >^{\Lambda_i} y$ if each component of x^{Λ_i} is at least as great as the corresponding component of y^{Λ_i} with strict inequality for some component.

Definition 16 *e^* is an* **economic environment** *if it is an extended expected utility environment such that*

1. *O is a convex subset of $\mathbb{R}^\Lambda$ with $|\Lambda| < \infty$;*
2. *each $u_i(.|t)$ is continuous on O;*
3. *for all $i \in I$ there exists $\Lambda_i \subseteq \Lambda$ such that $0 \in O^{\Lambda_i}$, $O^{\Lambda_i} \cap \mathbb{R}^{\Lambda_i}_{++} \neq \emptyset$, and $\Lambda = \bigcup_{i \in I} \Lambda_i$;*
4. *for all $i, j \in I$ and all $\tilde{x}, \hat{x} \in O$,*

$$u_i(\hat{x}|t) > u_i(\tilde{x}|t) \text{ and } u_j(\tilde{x}|t) < \sup_{y \in O} u_j(y|t)$$

is implied by $\hat{x} >^{\Lambda_i} \tilde{x} >^{\Lambda_i} 0$.

To see how this definition applies to private good economies, consider an extended expected utility environment with continuous utility functions over a set O of outcomes with the following structure. There is a convex set X of technologically feasible allocations of a finite set K of commodities, and each agent i's set of consumable allocations is $\mathbb{C}^i = \times_{k \in K} \mathbb{R}_+$. Define $O = X \cap \times_{i \in I} \mathbb{C}^i$ and let O^{Λ_i} denote the projection of O onto $\mathbb{C}^i$. Assuming that O contains a strictly positive vector and that each agent i has strictly monotonic and selfish preferences with respect to O^{Λ_i}, items 1, 2, and 3 in Definition 16 are immediately apparent. To see item 4, note that $\tilde{x} >^{\Lambda_i} 0$ implies that agent i is consuming a positive amount of some commodity at allocation $\tilde{x}$, and because other agents have monotonic preferences $\tilde{x}$ is maximal for none of them. Therefore, this is an economic environment. Definition 16 applies to pure public good economies by letting $O^{\Lambda_i} = O$ for all $i \in I$.

The next theorem shows how the economic environments are related to the continuous environments. Once it is proved, corollaries for interiority and Bayesian implementability follow immediately.

Theorem 10 *Every economic environment is a continuous environment.*

Proof: With the Euclidean topology on O, all that needs to be shown is that O^* is dense in O. Take $x \in O \setminus O^*$, so there exists $i \in I$ such that at least one component of x^{Λ_i} is greater than zero. Consider a sequence $\{y_k\}$ in O with $y_k^{\Lambda_i} = (1 - 1/k)x^{\Lambda_i}$ for all k, which exists since O^{Λ_i} is convex and owns zero. It may not be that $y_k \to x$, so define the sequence $\{x_k\}$ in O by $x_k = (1-\epsilon_k)x+\epsilon_k y_k$ for all k, where $\epsilon_k > 0$ is small enough to ensure $||x_k - x|| < 1/k$. This is possible since O is convex. Note that $x_k \to x$ and, for all $j \in I$, all $t \in T$, and all $k \in \mathbb{Z}_{++}$,

$$x >^{\Lambda_i} x_k >^{\Lambda_i} 0$$

implies $u_j(x_k|t) < \sup_{y \in O} u_j(y|t)$. Therefore, $x_k \in O^*$ for all k and it follows that O is contained in the closure of O^*. ∎

Corollary 3 follows from Theorems 9 and 10. Then Corollary 4 follows from Corollaries 2 and 3.

Corollary 3 *If an economic environment satisfies best-element-private values then it satisfies interiority.*

Corollary 4 *Assume e^* is an economic environment satisfying best-element-private values. Then f is Bayesian implementable in e^* if and only if (e^*, f^*) satisfies incentive compatibility and extended Bayesian monotonicity for some $f^* \in [f]$.*

Another important type of environment is that for which the set of outcomes is the set of probability measures over a measurable space of pure outcomes. I refer to these environments as lottery environments and offer a formal definition next. For a measurable space $(A, \mathcal{A})$, let $\Delta(A, \mathcal{A})$ denote the set of probability measures on $(A, \mathcal{A})$. I will write a for a generic element of A and $x(X)$ for the measure of the set $X \in \mathcal{A}$ under the probability measure $x \in O$.

Definition 17 *e^* is a* **lottery environment** *if it is an extend expected utility environment such that $O = \Delta(A, \mathcal{A})$ for some measurable space $(A, \mathcal{A})$, and there exists a function $v : I \times A \times T \to \mathbb{R}$ bounded in absolute value such that each $v_i(.|t)$ is $\mathcal{A}$-measurable and, for all $x \in O$,*

$$u_i(x|t) = \int_A v_i(a|t)\, dx(a).$$

Furthermore, there exists a (possibly finite) countable set $\{a_k\} \subseteq A$ such that for all $i \in I$ and all $t \in T$ there exist K and L such that $v_i(a_K|t) > v_i(a_L|t)$.

The last requirement of a lottery environment, that there exist a countable set over which no agent is indifferent at any state, is rather weak. It is satisfied, for example, when A is a separable topological space, each $v_i(.|t)$ is continuous, and no agent is indifferent over the entire set A of pure outcomes at any state.

The next theorem shows how the lottery environments are related to the continuous environments. Again, corollaries for interiority and Bayesian implementability follow immediately once it is proved.

Theorem 11 *Every lottery environment is a continuous environment.*

Proof: Let $\bar{v} > 0$ bound v in absolute value, and let W denote the set of $\mathcal{A}$-measurable functions $w : A \to \mathbb{R}$ bounded in absolute value by $\bar{v}$. I will construct a suitable topology on O by defining a metric $d_0 : O \times O \to \mathbb{R}$. Let

$$d_0(x,y) = \sup\left\{\left|\int_A w(a)\,dx(a) - \int_A w(a)\,dy(a)\right| \,\middle|\, w \in W\right\}$$

for all $x, y \in O$. Non-negativity follows easily, and the triangle inequality follows since

$$\begin{aligned}
d_0(x,y) &= \sup\left\{\left|\int_A w(a)\,dx(a) - \int_A w(a)\,dy(a)\right| \,\middle|\, w \in W\right\} \\
&= \sup\left\{\left|\int_A w(a)\,dx(a) - \int_A w(a)\,dz(a)\right.\right. \\
&\qquad \left.\left. + \int_A w(a)\,dz(a) - \int_A w(a)\,dy(a)\right| \,\middle|\, w \in W\right\} \\
&\leq \sup\left\{\left|\int_A w(a)\,dx(a) - \int_A w(a)\,dz(a)\right|\right. \\
&\qquad \left. + \left|\int_A w(a)\,dz(a) - \int_A w(a)\,dy(a)\right| \,\middle|\, w \in W\right\} \\
&\leq \sup\left\{\left|\int_A w(a)\,dx(a) - \int_A w(a)\,dz(a)\right| \,\middle|\, w \in W\right\}
\end{aligned}$$

$$+\sup\left\{\left|\int_A w(a)\,dz(a) - \int_A w(a)\,dy(a)\right| \;\middle|\; w \in W\right\}$$
$$= \quad d_0(x,z) + d_0(z,y)$$

for all $x, y, z \in O$. Now suppose that $x \neq y$, so there exists $X \in \mathcal{A}$ such that $x(X) \neq y(X)$. Define the function $w \in W$ by

$$w(t) \quad = \quad \begin{cases} \overline{v} & \text{if } t \in X \\ 0 & \text{else} \end{cases}$$

for all $t \in T$. Then

$$\begin{aligned} \left|\int_A w(a)\,dx(a) - \int_A w(a)\,dy(a)\right| &= \overline{v}|x(X) - y(X)| \\ &> 0, \end{aligned}$$

so $d_0(x,y) > 0$. This establishes that d_0 is indeed a metric.

That each $u_i(.|.)$ is bounded in absolute value follows since v is bounded in absolute value by $\overline{v}$. To see that each $u_i(.|t)$ is continuous, take $i \in I$, $t \in T$, and $\{x_k\} \subseteq O$ such that $x_k \to x$, and note that $v_i(.|t) \in W$. Then

$$|u_i(x_k|t) - u_i(x|t)| = \left|\int_A v_i(a|t)\,dx_k(a) - \int_A v_i(a|t)\,dx(a)\right| \leq d_0(x_k, x) \to 0,$$

as desired.

To see that O^* is dense in O, let y be an outcome that places positive probability on every a_k appearing in the definition of a lottery environment. If $\{a_k\}$ is countably infinite, let

$$y(X) = \sum_{a_k \in X} \frac{1}{k^2}$$

for all $X \in \mathcal{A}$, and if $\{a_k\}$ is finite then weight each pure outcome by $1/|\{a_k\}|$. Let $(1-\epsilon)f \oplus \epsilon h$ denote the social choice function defined by

$$((1-\epsilon)f \oplus \epsilon h)(t)(X) = (1-\epsilon)f(t)(X) + \epsilon h(t)(X)$$

for all $t \in T$ and all $X \in \mathcal{A}$. Take $x \in O$ and consider the sequence $\{y_k\}$ defined by

$$y_k = \left(1 - \frac{1}{k}\right) x \oplus \frac{1}{k} y$$

for all k. Each element of this sequence is contained in O^*, since no $v_i(.|t)$ is constant on $\{a_k\}$ and each pure outcome a_k has positive probability according to y_k. That $y_k \to x$ follows since

$$\begin{aligned} d_O(y_k, x) &= \sup \left\{ \left| \int_A w(a)\, dy_k(a) - \int_A w(a)\, dx(a) \right| \,\middle|\, w \in W \right\} \\ &\leq \sup \left\{ \left| \frac{1}{k} \int_A w(a)\, dy(a) - \frac{1}{k} \int_A w(a)\, dx(a) \right| \,\middle|\, w \in W \right\} \\ &\leq \frac{2}{k} \overline{v}, \end{aligned}$$

so the closure of O^* includes O. ∎

Corollary 5 follows from Theorems 9 and 11. Then Corollary 6 follows from Corollaries 2 and 5.

Corollary 5 *If a lottery environment satisfies best-element-private values then it satisfies interiority.*

Corollary 6 *Assume e^* is a lottery environment satisfying invariance and best-element-private values. Then f is Bayesian implementable in e^* if and only if*

(e^*, f^*) *satisfies incentive compatibility and extended Bayesian monotonicity for some* $f^* \in [f]$.

3.4 Conclusion

In Section 2.3, I extended Jackson's full characterization of Bayesian implementability for environments satisfying conflict-of-interest, and I argued that the result has limited applicability when the set of states is uncountable due to the restrictiveness of conflict-of-interest in these environments. In this chapter, I offer an alternative, full characterization of Bayesian implementability for environments satisfying invariance and interiority, and I show that this class of environments is large, including many of those satisyfing conflict-of-interest. Assuming best-element-private values, interiority is satisfied by the continuous environments, which are extended utility environments with continuous preferences and a dense set of outcomes that are best for no agent at any state. Examples of continuous environments are private good economic environments, which satisfy conflict-of-interest, and public good environments, which do not. Other examples that will be of great importance in the next chapter are the lottery environments, for which the set of outcomes is the set of probability measures on a measurable space of pure outcomes. These are the environments in which virtual implementability is usually considered, and the full characterization of Bayesian implementability proved in this chapter will be instrumental in proving a corresponding result for virtual Bayesian implementability.

Chapter 4

A Full Characterization of Virtual Bayesian Implementability in Quite General Environments

In Section 4.1, I add to the conceptual apparatus developed in Chapters 2 and 3. In particular, I define value-distinguished types and strict separability, and in the context of lottery environments, I define virtual Bayesian implementability. In Section 4.2, I use the full characterization of Bayesian implementability proved in

Chapter 3 to show that, for lottery environments satisfying best-element-private values and strict separability, a social choice function is virtually implementable in an environment if and only if the corresponding implementation problem satisfies incentive compatibility. In Section 4.3, I show that the class of lottery environments satisfying strict separability is suitably large. It includes the lottery environments satisfying private values and value-distinguished types when the set of pure outcomes is finite, and it includes those when the set of pure outcomes is a finite set crossed with an open set of allocations of a transferable private good. Section 4.4 concludes the chapter.

4.1 Notation and Definitions

In Sections 4.2 and 4.3, I consider environments for which there exists a social choice function that, in a sense, separates the types of each agent. More precisely, there must exist a social choice function such that each type of each agent strictly prefers the outcomes determined using the agent's true type to those determined using any other type, regardless of which types of other agents are used. I formally define this condition next for arbitrary environments, and in Section 4.2, I offer an equivalent statement for lottery environments.

Definition 18 *An environment e satisfies* **strict separability** *if there exists $h^* \in F$ such that for all $i \in I$, all $t_i \in T_i$, all $\alpha \in \Sigma(T)$ with $\alpha_i(t_i) \neq t_i$, all*

$S \subseteq T$ with $\mu_i^*(S|t_i) > 0$, and all $h \in F$,

$$h/_S(h^* \circ (\tau_i, \alpha_{-i}))\ P_i(t_i)\ h/_S(h^* \circ \alpha).$$

This is even stronger than requiring that truth be a strict dominant strategy of the game induced by the mechanism (T, h^*), for strict preference must hold not only for $S = \{t_i\} \times T_{-i}$ but for all S with positive outer measure. Nonetheless, I show in Section 4.3 that this condition is satisfied for a suitably large class of environments.

In Chapters 2 and 3, I characterize the implementation problems with solutions in the strong sense that there exist a mechanism with Bayesian equilibria that coincide almost everywhere with the social choice function at hand. In any practical implementation problem, however, the predictions of Bayesian equilibrium must be accepted as approximations of behavior, so this standard may be unreasonably strong. It is therefore of interest to consider the implementation problems for which there exist mechanisms with Bayesian equilibria that are arbitrarily close to the social choice function at hand, or put differently. This is the idea of virtual Bayesian implementability, formalized next. Let $\overline{v}$ denote an absolute upper bound of v, and let W denote the set of $\mathcal{A}$-measurable functions $w : A \to \mathbb{R}$ bounded in absolute value by $\overline{v}$.

Definition 19 *A social choice function f is ϵ-***virtually Bayesian implementable** *in e^* if there exists $h \in F$ such that h is Bayesian implementable in e^*,*

(e^*, h) is incentive compatible, and

$$\forall t \in T, w \in W, \alpha \in \Sigma(T)$$
$$\left|\int_A w(a)\, df(\alpha(t))(a) - \int_A w(a)\, dh(\alpha(t))(a)\right| \leq \epsilon.$$

f is **virtually Bayesian implementable** *in e^* if, for all $\epsilon > 0$, it is ϵ-virtually Bayesian implementable in e^*.*

In other words, f is virtually Bayesian implementable in e^* if and only if there exists a sequence $\{h^k\}$ in F such that each h^k is Bayesian implementable in e^*, each (e^*, h^k) is incentive compatible, and $h^k \to f$ in the metric d_1, defined by

$$d_1(f^1, f^2) = \sup\{\left|\int_A w(a)\, df^1(\alpha(t))(a) - \int_A w(a)\, df^2(\alpha(t))(a)\right| \; |t \in T, w \in W, \alpha \in \Sigma(T)\}$$

for all $f^1, f^2 \in F$. This metric represents a fairly strong notion of closeness of social choice functions. It is straightforward to check that convergence in d_1 implies convergence in the metric d_2, defined by

$$d_2(f^1, f^2) = \sup\left\{|f^1(t)(X) - f^2(t)(X)| \; |t \in T, X \in \mathcal{A}\right\}$$

for all $f^1, f^2 \in F$. And when A is finite, so $\Delta(A, \mathcal{A})$ is a finite dimensional simplex in Euclidean space, convergence in d_2 implies uniform convergence in the Euclidean metric.

Note that the social choice function h in Definition 19 must be not only Bayesian implementable in e^*, but (e^*, h) must be incentive compatible as well. In environments with countable sets of states each with positive probability according to a common prior, the first condition implies the second, and the qualification of incentive compatibility is redundant. Theorem 2 shows that, in more complex environments, Bayesian implementability of h implies incentive compatibility of (e^*, h') only for a μ^*-equivalent social choice function h'. This weaker condition is not strong enough to prove the necessity of incentive compatibility for virtual Bayesian implementability, stated in Theorem 12. The qualification is therefore important when the set of states is uncountable.

4.2 Characterization Results

In this section, I use the results of Section 3.2 to improve the existing results on virtual Bayesian implementability due to Abreu and Matsushima (1990b) and Matsushima (1993), who offer partial characterizations relying on the assumption that the set of states is finite. I first show that incentive compatibility of an implementation problem must be satisfied whenever the social choice function at hand is virtually Bayesian implementable in a lottery environment. The intuition for this is straightforward. Suppose that f is virtually Bayesian implementable in e^* but that truthful reporting is not a Bayesian equilibrium of the mechanism (T, f). Then there exists an agent, a type, and a false report such that the agent's type strictly prefers the outcomes determined using the false report to those determined using the truth. For a suitably specified metric, this strict preference should hold for all social choice functions sufficiently close to

f, but by virtual Bayesian implementability there exist arbitrarily close social choice functions for which the strict preference does not hold. Therefore, (e^*, f) must satisfy incentive compatibility. The proof is not much more complicated than this intuition but considerations of strict preference involve manipulations of upper integrals, which must be treated carefully.

Theorem 12 *Assume e^* is a lottery environment. Then f is virtually Bayesian implementable in e^* only if (e^*, f) satisfies incentive compatibility.*

Proof: Assume that f is virtually Bayesian implementable in e^*. Note that (e^*, f) satisfies incentive compatibility if and only if, for all $i \in I$, all $\tilde{t}_i \in T_i$, and all $\tilde{\alpha}_i \in \Sigma_i(T_i)$,

$$\begin{aligned} &\int^* \{u_i(f(t)|t) - u_i(f(\tilde{\alpha}_i(t_i), t_{-i})|t)\}\, d\mu_i(t|\tilde{t}_i) \\ &\qquad - \int^* \{u_i(f(\tilde{\alpha}_i(t_i), t_{-i})|t) - u_i(f(t)|t)\}\, d\mu_i(t|\tilde{t}_i) \\ &\geq 0. \end{aligned}$$

Since f is virtually Bayesian implementable in e^*, for all $k \in \mathbb{Z}_{++}$ there exists $h^k \in F$ such that (e^*, h^k) is incentive compatible and

$$\begin{aligned} &\forall t \in T, w \in W, \alpha \in \Sigma(T) \\ &\qquad \left| \int_A w(a)\, df(\alpha(t))(a) - \int_A w(a)\, h^k(\alpha(t))(a) \right| \leq \frac{1}{k}. \end{aligned}$$

Setting $\alpha = (\tilde{\alpha}_i, \tau_{-i})$ and, for each $t \in T$, $w = v_i(.|t)$, this yields

$$\forall t \in T \;\; \left| u_i(f(\tilde{\alpha}_i(t_i), t_{-i})|t) - u_i(h^k(\tilde{\alpha}_i(t_i), t_{-i})|t) \right| \leq \frac{1}{k},$$

and setting $\alpha = \tau$, it yields

$$\forall t \in T \ \left|u_i(f(t)|t) - u_i(h^k(t)|t)\right| \leq \frac{1}{k}.$$

These inequalities together with Propositions 6 and 8 imply that, for all k,

$$\begin{aligned}
&\int^* \{u_i(f(t)|t) - u_i(f(\tilde{\alpha}_i(t_i), t_{-i})|t)\}\, d\mu_i(t|\tilde{t}_i) \\
&\qquad - \int^* \{u_i(f(\tilde{\alpha}_i(t_i), t_{-i})|t) - u_i(f(t)|t)\}\, d\mu_i(t|\tilde{t}_i) \\
= &\int^* \{u_i(f(t)|t) - u_i(h^k(t)|t) + u_i(h^k(t)|t) \\
&-u_i(h^k(\tilde{\alpha}_i(t_i), t_{-i})|t) + u_i(h^k(\tilde{\alpha}_i(t_i), t_{-i})|t) \\
&-u_i(f(\tilde{\alpha}_i(t_i), t_{-i})|t)\}\, d\mu_i(t|\tilde{t}_i) \\
&- \int^* \{u_i(f(\tilde{\alpha}_i(t_i), t_{-i})|t) - u_i(h^k(\tilde{\alpha}_i(t_i), t_{-i})|t) \\
&+u_i(h^k(\tilde{\alpha}_i(t_i), t_{-i})|t) - u_i(h^k(t)|t) \\
&+u_i(h^k(t)|t) - u_i(f(t)|t)\}\, d\mu_i(t|\tilde{t}_i) \\
\geq &\int^* \{u_i(h^k(t)|t) - u_i(h^k(\alpha_i(t_i), t_{-i})|t) - \frac{2}{k}\}\, d\mu_i(t|\tilde{t}_i) \\
&- \int^* \{u_i(h^k(\alpha_i(t_i), t_{-i})|t) - u_i(h^k(t)|t) + \frac{2}{k}\}\, d\mu_i(t|\tilde{t}_i) \\
= &\int^* \{u_i(h^k(t)|t) - u_i(h^k(\alpha_i(t_i), t_{-i})|t)\}\, d\mu_i(t|\tilde{t}_i) \\
&- \int^* \{u_i(h^k(\alpha_i(t_i), t_{-i})|t) - u_i(h^k(t)|t)\}\, d\mu_i(t|\tilde{t}_i) - \frac{4}{k}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\int^* \{u_i(f(t)|t) - u_i(f(\tilde{\alpha}_i(t_i), t_{-i})|t)\}\, d\mu_i(t|\tilde{t}_i) \\
&\qquad - \int^* \{u_i(f(\tilde{\alpha}_i(t_i), t_{-i})|t) - u_i(f(t)|t)\}\, d\mu_i(t|\tilde{t}_i) \\
\geq &\sup_{k\to\infty} \int^* \{u_i(h^k(t)|t) - u_i(h^k(\alpha_i(t_i), t_{-i})|t)\}\, d\mu_i(t|\tilde{t}_i)
\end{aligned}$$

$$- \int^{*} \{u_i(h^k(\alpha_i(t_i), t_{-i})|t) - u_i(h^k(t)|t)\} \, d\mu_i(t|\tilde{t}_i) - \frac{4}{k}$$

$$\geq \; 0,$$

where the second inequality follows from the incentive compatibility of each (e^*, h^k). This establishes that (e^*, f) satisfies incentive compatibility. ∎

The proof of Theorem 13 relies on a formulation of strict separability that, for lottery environments, is equivalent to the statement in Section 4.1. This formulation is stated in the next proposition and the equivalence is proved in the appendix.

Proposition 13 *An extended expected utility environment satisfies strict separability if and only if there exists $h^* \in F$ such that, for all $i \in I$, all $\tilde{t}_i \in T_i$, and all $\alpha \in \Sigma(T)$ with $\alpha_i(\tilde{t}_i) \neq \tilde{t}_i$,*

$$\mu_i^* \left(\{t \in T | u_i((h^* \circ (\tau_i, \alpha_{-i}))(t)|t) > u_i((h^* \circ \alpha)(t)|t)\} \, |\tilde{t}_i \right) = 1.$$

That is, strict separability is equivalent to the existence of a social choice function h^* such that each agent i's utility is strictly greater at μ^*-almost every state when h^* uses i's true type rather than a false one. Recall that the notation $(1-\epsilon)f \oplus \epsilon h$ denotes the social choice function defined by

$$((1-\epsilon)f \oplus \epsilon h)(t)(X) = (1-\epsilon)f(t)(X) + \epsilon h(t)(X)$$

for all $t \in T$ and all $X \in \mathcal{A}$. In lottery environments, each $u_i(x|t)$ is given by the integral of $v_i(.|t)$ with respect to the probability measure x, and it follows from the linearity of the integral that

$$u_i(((1-\epsilon)f \oplus \epsilon h)(.)|.) = (1-\epsilon)u_i(f(.)|.) + \epsilon u_i(h(.)|.)$$

for all $f, h \in F$. The proof of Theorem 13 proceeds by showing that, when strict separability is satisfied in a lottery environment e^*, every social choice function f for which (e^*, f) is incentive compatible has arbitrarily close neighbors h in the d_1-metric such that (e^*, h) satisfies incentive compatibility and extended Bayesian monotonicity. It then follows from Theorem 7 that h is Bayesian implementable, and that f is virtually Bayesian implementable in e^*. I construct the social choice functions h by augmenting f with the social choice function h^* appearing in the statement of strict separability. Specifically, I define h as $(1-2\epsilon)f \oplus \epsilon h^* \oplus \epsilon\overline{y}$, where $\overline{y}$ is a constant social choice function that places positive probability on each pure outcome a_k appearing in the Definition 17 at every state. Linearity of $u_i(.|t)$ is crucial in showing that the resulting implementation problems (e^*, h) satisfy incentive compatibility and extended Bayesian monotonicity. Note that the proof of Theorem 13 relies on the axiom of choice.

Theorem 13 *Assume a lottery environment e^* satisfies best-element-private values and strict separability. Then f is virtually Bayesian implementable in e^* if (e^*, f) satisfies incentive compatibility.*

Proof: Take $\epsilon > 0$, and define $h \in F$ by

$$h = (1-2\epsilon)f \oplus \epsilon h^* \oplus \epsilon\overline{y}$$

for all $t \in T$, where $\overline{y}$ is the constant social choice function that chooses y at every state and y places positive probability on every a_k. If $\{a_k\}$ is countably infinite, for example, y may be defined by

$$y(X) = \sum_{a_k \in X} \frac{1}{k^2}$$

for all $X \in \mathcal{A}$. If $\{a_k\}$ is finite then the weights on each pure outcome can be set at $1/|\{a_k\}|$. To see that h satisfies the conditions of $4\epsilon\overline{v}$-virtual Bayesian implementability of f in e^*, take $t \in T$, $w \in W$, and $\alpha \in \Sigma(T)$, and note that

$$\begin{aligned}
&\left|\int_A w(a)\, df(\alpha(t))(a) - \int_A w(a)\, dh(\alpha(t))(a)\right| \\
&= \left|\int_A w(a)\, df(\alpha(t))(a) - (1-2\epsilon)\int_A w(a)\, df(\alpha(t))(a)\right. \\
&\quad \left. -\epsilon \int_A w(a)\, dh^*(\alpha(t))(a) - \epsilon \int_A w(a)\, d\overline{y}(a)\right| \\
&\leq 2\epsilon\left|\int_A w(a)\, df(\alpha(t))(a)\right| + \epsilon\left|\int_A w(a)\, dh^*(\alpha(t))(a)\right| \\
&\quad +\epsilon\left|\int_A w(a)\, d\overline{y}(a)\right| \\
&\leq 4\epsilon\overline{v}.
\end{aligned}$$

I have left to show only that (e^*, h) is incentive compatible and that h is Bayesian implementable in e^*, since ϵ may be chosen arbitrarily small.

To see that (e^*, h) is incentive compatible, take $i \in I$, $\tilde{t}_i \in T_i$, and $\alpha_i \in \Sigma_i(T_i)$. If $\alpha_i(\tilde{t}_i) = \tilde{t}_i$ then invariance immediately implies that $h\ R_i^*(\tilde{t}_i)\ h \circ (\alpha_i, \tau_{-i})$, so assume that $\alpha_i(\tilde{t}_i) \neq \tilde{t}_i$, and let S denote the set of states t such that $u_i(h^*(t)|t) > u_i(h^*(\alpha_i(t_i), t_{-i})|t)$. Note that $\mu_i^*(S|\tilde{t}_i) = 1$ by Proposition

13. Then

$$\begin{aligned}
&\int^* \{u_i((h \circ (\alpha_i, \tau_{-i}))(t)|t) - u_i(h(t)|t)\}\, d\mu_i(t|\tilde{t}_i) \\
&= \int^* \{(1-2\epsilon)[u_i(f(\alpha_i(t_i), t_{-i})|t) - u_i(f(t)|t)] \\
&\quad + \epsilon[u_i(h^*(\alpha_i(t_i), t_{-i})|t) - u_i(h^*(t)|t)]\}\, d\mu_i(t|\tilde{t}_i) \\
&= \int^* \{(1-2\epsilon)[u_i(f(\alpha_i(t_i), t_{-i})|t) - u_i(f(t)|t)] \\
&\quad + \epsilon[u_i(f/_S(h^* \circ (\alpha_i, \tau_{-i}))(t)|t) - u_i((f/_S h^*)(t)|t)]\}\, d\mu_i(t|\tilde{t}_i) \\
&\leq \int^* \{(1-2\epsilon)[u_i(f(\alpha_i(t_i), t_{-i})|t) - u_i(f(t)|t)]\}\, d\mu_i(t|\tilde{t}_i) \\
&= (1-2\epsilon) \int^* \{u_i(f(\alpha_i(t_i), t_{-i})|t) - u_i(f(t)|t))\}\, d\mu_i(t|\tilde{t}_i)
\end{aligned}$$

where the second equality follows from Proposition 5, the inequality follows from Proposition 6, and the last equality follows from Proposition 9. A similar argument uses Proposition 7 instead of Proposition 6 to show that

$$\begin{aligned}
&\int^* \{u_i(h(t)|t) - u_i((h \circ (\alpha_i, \tau_{-i}))(t)|t)\}\, d\mu_i(t|\tilde{t}_i) \\
&\geq (1-2\epsilon) \int^* \{u_i(f(t)|t) - u_i(f(\alpha_i(t_i), t_{-i})|t)\}\, d\mu_i(t|\tilde{t}_i).
\end{aligned}$$

Then

$$\begin{aligned}
&\int^* \{u_i(h(\alpha_i(t_i), t_{-i})|t) - u_i(h(t)|t)\}\, d\mu_i(t|\tilde{t}_i) \\
&\quad - \int^* \{u_i(h(t)|t) - u_i(h(\alpha_i(t_i), t_{-i})|t)\}\, d\mu_i(t|\tilde{t}_i) \\
&\leq (1-2\epsilon) \int^* \{u_i(f(\alpha_i(t_i), t_{-i})|t) - u_i(f(t)|t))\}\, d\mu_i(t|\tilde{t}_i) \\
&\quad -(1-2\epsilon) \int^* \{u_i(f(t)|t) - u_i(f(\alpha_i(t_i), t_{-i})|t)\}\, d\mu_i(t|\tilde{t}_i) \\
&\leq 0,
\end{aligned}$$

where the second inequality follows since (e^*, f) satisfies incentive compatibility. Therefore, $h\ R_i^*(\tilde{t}_i)\ h \circ (\alpha_i, \tau_{-i})$, and (e^*, h) satisfies incentive compatibility.

Corollary 6 shows that h is Bayesian implementable in e^* if (e^*, h) satisfies extended Bayesian monotonicity in addition to incentive compatibility. For all $i \in I$, let

$$H_i = \left\{(1-2\epsilon) f \circ (t'_i, \tau_{-i}) \oplus \frac{\epsilon}{z+1} h^* \circ (t'_i, \tau_{-i}) \right.$$
$$\left. \oplus \frac{\epsilon z}{z+1} h^* \circ (t''_i, \tau_{-i}) \oplus \epsilon y \,\middle|\, t'_i, t''_i \in T_i, z \in \mathbb{Z}_+ \right\},$$

and let $h(.|t'_i, t''_i, z)$ denote the element of H_i specified by the parameters t'_i, t''_i, and z. To see that each $H_i \subseteq \tilde{F}_i$, take $i \in I$, $\tilde{t}_i \in T_i$, $\alpha_i \in \Sigma_i(T_i)$, and $h(.|t'_i, t''_i, z) \in H_i$. If $t'_i = t''_i = \tilde{t}_i$ then invariance implies that $h\ R_i^*(\tilde{t}_i)\ h(.|t'_i, t''_i, z) \circ (\alpha_i, \tau_i)$, so assume that $t'_i \neq \tilde{t}_i$ or $t''_i \neq \tilde{t}_i$. Let S' denote the set of states t such that $u_i(h^*(t)|t) > u_i((h^* \circ (t'_i, \tau_{-i}))(t)|t)$, let S'' denote the set of states t such that $u_i(h^*(t)|t) > u_i((h^* \circ (t''_i, \tau_{-i}))(t)|t)$, and note that $\mu_i^*(S' \cup S''|\tilde{t}_i) = 1$ by Proposition 13. Then

$$\int^* \{u_i(h(\alpha_i(t_i), t_{-i}|t'_i, t''_i, z)|t) - u_i(h(t)|t)\}\, d\mu_i(t|\tilde{t}_i)$$
$$= \int^* \{(1-2\epsilon)[u_i(f(t'_i, t_{-i})|t) - u_i(f(t)|t)]$$
$$+ \frac{\epsilon}{z+1}[u_i(h^*(t'_i, t_{-i})|t) - u_i(h^*(t)|t)]$$
$$+ \frac{\epsilon z}{z+1}[u_i(h^*(t''_i, t_{-i})|t) - u_i(h^*(t)|t)]\}\, d\mu_i(t|\tilde{t}_i)$$
$$= \int^* \{(1-2\epsilon)[u_i(f(t'_i, t_{-i})|t) - u_i(f(t)|t)]$$
$$+ \frac{\epsilon}{z+1}[u_i(f/_{S' \cup S''}(h^* \circ (t'_i, \tau_{-i}))(t)|t)$$

$$
\begin{aligned}
& -u_i((f/_{S' \cup S''} h^*)(t)|t)] \\
& + \frac{\epsilon z}{z+1}[u_i(f/_{S' \cup S''}(h^* \circ (t_i'', \tau_{-i}))(t)|t) \\
& -u_i((f/_{S' \cup S''} h^*)(t)|t)]\} \, d\mu_i(t|\tilde{t}_i) \\
\leq \; & \int^* \{(1-2\epsilon)[u_i(f(t_i', t_{-i})|t) - u_i(f(t)|t)]\} \, d\mu_i(t|\tilde{t}_i) \\
= \; & (1-2\epsilon) \int^* \{u_i(f(t_i', t_{-i})|t) - u_i(f(t)|t)\} \, d\mu_i(t|\tilde{t}_i),
\end{aligned}
$$

where the second equality follows from Proposition 5, the inequality follows from Proposition 6, and the last equality follows from Proposition 9. A similar argument uses Proposition 7 instead of Proposition 6 to show that

$$
\begin{aligned}
& \int^* \{u_i(h(t)|t) - u_i(h(\alpha_i(t_i), t_{-i}|t_i', t_i'', z)|t)\} \, d\mu_i(t|\tilde{t}_i) \\
& \quad \geq \; (1-2\epsilon) \int^* \{u_i(f(t)|t) - u_i(f(t_i', t_{-i})|t)\} \, d\mu_i(t|\tilde{t}_i).
\end{aligned}
$$

Then

$$
\begin{aligned}
& \int^* \{u_i(h(\alpha_i(t_i), t_{-i}|t_i', t_i'', z)|t) - u_i(h(t)|t)\} \, d\mu_i(t|\tilde{t}_i) \\
& \qquad - \int^* \{u_i(h(t)|t) - u_i(h(\alpha_i(t_i), t_{-i}|t_i', t_i'', z)|t)\} \, d\mu_i(t|\tilde{t}_i) \\
& \leq \; (1-2\epsilon) \int^* \{u_i(f(t_i', t_{-i})|t) - u_i(f(t|t))\} \, d\mu_i(t|\tilde{t}_i) \\
& \qquad -(1-2\epsilon) \int^* \{u_i(f(t)|t) - u_i(f(t_i', t_{-i})|t)\} \, d\mu_i(t|\tilde{t}_i) \\
& \leq \; 0,
\end{aligned}
$$

where the second inequality follows since (e^*, f) satisfies incentive compatibility. Therefore, $h \; R_i^*(\tilde{t}_i) \; h(.|t_i', t_i'', z)(\alpha_i, \tau_{-i})$, and $H_i \subseteq \tilde{F}_i$.

Take $\alpha \in \Sigma(T)$, $Q = \bigcup_{i \in I} Q_i \times T_{-i}$ with $\mu^*(Q \setminus Q^+) = 0$, and $\psi \in \Sigma(H)$, and assume that

$$\forall i \in I, t_i \in T_i, \hat{h} \in H_i, \hat{f} \in F$$

$$(h \circ \alpha)//_{Q_i^+}\langle \psi_i, \alpha \rangle \; R_i^*(t_i) \; (\hat{h} \circ \alpha)/_{Q_{\neq i}} \hat{f}.$$

To show that

$$(h \circ \alpha)//_{Q_i^+}\langle \psi_i, \alpha \rangle \in [h],$$

I will first establish that, for all $i \in I$ and all $t_i \in T_i \setminus Q_i$, $\alpha_i(t_i) = t_i$. To see this, suppose there exists $j \in I$ and $\tilde{t}_j \in T_j \setminus Q_j$ such that $\alpha_j(\tilde{t}_j) \neq \tilde{t}_j$. I use the axiom of choice to define the social choice function f_j as follows. Note that y places positive probability on each pure outcome a_k in the statement of strict separability, and that for every $t \in T$ there exist $K(t)$ and $L(t)$ such that $v_j(a_{K(t)}|t) > v_j(a_{L(t)}|t)$. Let $f_j(t)$ be the outcome that transfers the probability on the pure outcome $a_{L(t)}$ to the pure outcome $a_{K(t)}$. That is, $f_j(t)(\{a_{K(t)}\}) = y(\{a_{K(t)}\} \cup \{a_{L(t)}\})$, $f_j(t)(\{a_{L(t)}\}) = 0$, and $f_j(t)(\{a_k\}) = y_j(\{a_k\})$ for all other k. Clearly, for all $t \in T$,

$$u_j(f_j(t)|t) = \int_A v_j(a|t)\, df_j(t)(a) > \int_A v_j(a|t)\, dy(a) = u_j(y|t).$$

Now take a social choice function h_j such that, for all $i \in I$ and all $t \in Q_i^+$,

$$h_j(t) = (1 - 2\epsilon) f \circ (t_i', \alpha_{-i}) \oplus \frac{\epsilon}{z+1} h^* \circ (t_i', \alpha_{-i})$$

$$\oplus \frac{\epsilon z}{z+1} h^* \circ (t_i'', \alpha_{-i}) \oplus f_j$$

where $h(.|t_i', t_i'', z) = \psi_i(t_i)$. In words, h_j alters the outcome of each $\psi_i(t_i)(\alpha(t))$ only by transferring probability according to f_j between the appropriate pure outcomes, raising j's payoff at every state in Q^+. I claim that

$$(h(.|\alpha_j(\tilde{t}_j), \tilde{t}_j, 1) \circ \alpha)/_{Q_{\neq j}} h_j \; P_j^*(\tilde{t}_j) \; (h \circ \alpha)//_{Q_i^+} \langle \psi_i, \alpha \rangle,$$

a contradiction. To see this, let S denote the set of states $t \in \{\tilde{t}_j\} \times T_{-j}$ such that $u_j(h^*(\tilde{t}_j, \alpha_{-j}(t_{-j}))|t) > u_j(h^*(\alpha(t))|t)$, and take any state $t \in S \cap ((T \setminus Q) \cup Q^+)$. Note that $\mu_j^*(S \cap ((T \setminus Q) \cup Q^+)|\tilde{t}_j) = 1$ by Proposition 13 and the assumption that $\mu^*(Q \setminus Q^+) = 0$. If $t \in S \cap (T \setminus Q)$ then

$$\begin{aligned}
& u_j(h(\alpha(t)|\alpha_j(\tilde{t}_j), \tilde{t}_j, 1)|t) - u_j(h(\alpha(t))|t) \\
= \; & \frac{\epsilon}{2}[u_j(h^*(\tilde{t}_j, \alpha_{-j}(t_{-j}))|t) - u_j(h^*(\alpha(t))|t)] \\
+ \; & \epsilon[u_j(f_j(t)|t) - u_j(y|t)] \\
> \; & 0,
\end{aligned}$$

where the inequality follows since $t \in S$ and by the construction of f_j. If $t \in S \cap Q_i^+$ for some i then

$$\begin{aligned}
u_j(h_j(t)|t) - u_j(\psi_i(t_i)(\alpha(t))|t) &= \epsilon(u_j(f_j(t)|t) - u_j(y|t) \\
&> 0,
\end{aligned}$$

which follows by the construction of f_j. Then, noting that $\mu_j^*(S \cap ((T \setminus Q) \cup Q^+)|\tilde{t}_j) = 1$,

$$\int^* \{u_j((h(.|\alpha_j(\tilde{t}_j), \tilde{t}_j, 1) \circ \alpha)/_{Q_{\neq j}} h_j(t)|t)$$

$$-u_j((h \circ \alpha)//_{Q_i^+}\langle\psi_i,\alpha\rangle(t)|t)\}\,d\mu_j(t|\tilde{t}_j)$$
$$-\int^* \{u_j((h \circ \alpha)//_{Q_i^+}\langle\psi_i,\alpha\rangle(t)|t)$$
$$-u_j((h(.|\alpha_j(\tilde{t}_j),\tilde{t}_j,1) \circ \alpha)/_{Q_{\neq j}} h_j(t)|t)\}\,d\mu_j(t|\tilde{t}_j)$$
$$> \quad 0,$$

follows by application of Propositions 5, 6, and 7, establishing the claim. Therefore, $\alpha_j(\tilde{t}_j) = \tilde{t}_j$.

Next, I will show that, for all $i \in I$ and all $t_i \in Q_i$, there exists $z \in \mathbb{Z}_+$ such that $\psi_i(t_i) = h(.|t_i,t_i,z)$. To see this, suppose there exists $j \in I$ and $\tilde{t}_j \in Q_j$ such that $\psi_j(\tilde{t}_j) = h(.|t'_j,t''_j,z)$ with $t'_j \neq \tilde{t}_j$ or $t''_j \neq \tilde{t}_j$. Note that $\mu_j^*(Q_{\neq j}|\tilde{t}_j) = 0$, since $\mu^*(Q \setminus Q^+) = 0$. Consider the social choice function $h(.|t'_j,\tilde{t}_j,z+1) \in H_j$, let S' denote the set of states t such that $u_j(h^*(\tilde{t}_j,\alpha_{-j}(t_{-j}))|t) > u_j(h^*(t'_j,\alpha_{-j}(t_{-j}))|t)$, and let S'' denote the set of states t such that $u_j(h^*(\tilde{t}_j,\alpha_{-j}(t_{-j}))|t) > u_j(h^*(t''_j,\alpha_{-j}(t_{-j}))|t)$. Note that $\mu_i^*(S' \cup S''|\tilde{t}_j) = 1$. Then

$$\begin{aligned}
&\int^* \{u_j(h(\alpha(t)|t'_j,t''_j,z+1)|t) - u_j(h(\alpha(t)|t'_j,t''_j,z)|t)\}\,d\mu_j(t|\tilde{t}_j)\\
&= \int^* \left\{\frac{\epsilon}{z+2}u_j(h^*(t'_j,\alpha_{-j}(t_{-j}))|t) + \frac{\epsilon(z+1)}{z+2}u_j(h^*(\tilde{t}_j,\alpha_{-j}(t_{-j}))|t)\right.\\
&\quad \left. -\frac{\epsilon}{z+1}u_j(h^*(t'_j,\alpha_{-j}(t_{-j}))|t) - \frac{\epsilon z}{z+1}u_j(h^*(t''_j,\alpha_{-j}(t_{-j}))|t)\right\}\,d\mu_j(t|\tilde{t}_j)\\
&= \int^* \left\{\frac{\epsilon(z^2+z)}{(z+1)(z+2)}[u_j(h^*(\tilde{t}_j,\alpha_{-j}(t_{-j}))|t) - u_j(h^*(t''_j,\alpha_{-j}(t_{-j}))|t)]\right.\\
&\quad \left. +\frac{\epsilon}{(z+1)(z+2)}[u_j(h^*(\tilde{t}_j,\alpha_{-j}(t_{-j}))|t) - u_j(h^*(t'_j,\alpha_{-j}(t_{-j}))|t)]\right\}\,d\mu_i(t|\tilde{t}_j)\\
&> \quad 0,
\end{aligned}$$

where the inequality follows by application of Propositions 5 and 7. A similar argument uses Proposition 6 instead of Proposition 7 to show

$$\int^* \{u_j(h(\alpha(t)|t'_j, t''_j, z)|t) - u_j(h(\alpha(t)|t'_j, t''_j, z+1)|t)\}\, d\mu_j(t|\tilde{t}_j) \leq 0,$$

which implies $h(.|t'_j, \tilde{t}_j, z+1)\ P^*_j(\tilde{t}_j)\ h(.|t'_j, t''_j, z)$. This contradiction shows that $t'_j = t''_j = \tilde{t}_j$, as claimed.

Now take any $t \in (T \setminus Q) \cup Q^+$. If $t \in T \setminus Q$ then $\alpha(t) = t$ and $(h \circ \alpha)//_{Q_i^+}\langle\psi_i, \alpha\rangle(t) = h(t)$. If $t \in Q^+$ then there is exactly one $j \in I$ such that $t \in Q_j^+$. Then $\alpha_{-j}(t_{-j}) = t_{-j}$ and

$$\begin{aligned} (h \circ \alpha)//_{Q_i^+}\langle\psi_i, \alpha\rangle(t) &= \psi_j(t_j)(\alpha(t)) \\ &= h(\alpha(t)|t_j, t_j, z) \\ &= h(t). \end{aligned}$$

Since $\mu^*(Q \setminus Q^+) = 0$, this establishes that

$$(h \circ \alpha)//_{Q_i^+}\langle\psi_i, \alpha\rangle \in [h].$$

Therefore, (e^*, h) satisfies extended Bayesian monotonicity, and f is virtually Bayesian implementable in e^*. ∎

4.3 Strict Separability

Strict separability is, by all accounts, a strong condition. As noted above, it is even stronger for an environment e^* than the existence of a social choice function

h^* such that (e^*, h^*) satisfies strict dominant strategy incentive compatibility, which would require for extended expected utility environments that, for all $i \in I$, all $\tilde{t}_i \in T_i$, and all $\alpha \in \Sigma(T)$ with $\alpha_i(\tilde{t}_i) \neq \tilde{t}_i$,

$$\int^* \{u_i((h^* \circ (\tau_i, \alpha_{-i}))(t)|t) - u_i((h^* \circ \alpha)(t))\}\, d\mu_i(t|\tilde{t}_i) \\ - \int^* \{u_i((h^* \circ \alpha)(t)|t) - u_i((h^* \circ (\tau_i, \alpha_{-i}))(t)|t)\}\, d\mu_i(t|\tilde{t}_i) > 0.$$

Theorem 5 and Propositions 5 and 13 clearly show that this is implied by strict separability. Of course, strict dominant strategy incentive compatibility is even stronger than dominant strategy incentive compatibility, which is referred to as strategy-proofness in a large literature founded by Gibbard (1973) and Satterthwaite (1975) that seeks to characterize the implementation problems satisfying this condition. The most powerful theorems in this area are due to Roberts (1979) and Hylland (1980), who implicitly consider environments satisfying two conditions that figure prominently in this section. One is private values, stated in Section 3.3, and the other is value-distinguished types, stated next for an extended expected utility environment.

Definition 20 *An environment e satisfies* **value-distinguished types** *if for all $i \in I$, all $t, t' \in T$, all $a \in \mathbb{R}$, and all $b \in \mathbb{R}_{++}$,*

$$u_i(.|t) = a + b u_i(.|t') \text{ implies } t_i = t'_i.$$

Value-distinguished types is common in the literature (see Palfey and Srivastava, 1993), requiring that no two of any agent's types induce the same utility function over outcomes. The conjunction of value-distinguished types and private values amounts to identifying each agent i's types t_i with unique, distinct utility functions $u_i(.|t_i)$ over outcomes. In lottery environments, these utility functions correspond to distinct utility functions $v_i(.|t_i)$ over pure outcomes.

Hylland considers lottery environments with a finite set of pure outcomes satisfying an unrestricted strict domain condition. I next define a slightly larger class of environments to which Theorem 14 applies.

Definition 21 e^* *is a* **finite lottery environment** *if it is a lottery environment with A finite and $\mathcal{A} = 2^A$ such that for all $i \in I$ and all $t \in T$ there exist pure outcomes $a_1, a_2 \in A$ such that $v_i(a_1|t) > v_i(a_2|t)$.*

Hylland shows that, assuming private values, value-distinguished types, and a weak citizen sovereignty condition, an implementation problem satisfies dominant strategy incentive compatibility if and only if the social choice function at hand is a probability combination of dictatorial social choice functions, also referred to as a random dictatorship. It follows that, for these environments, random dictatorship is implied by strict dominant strategy incentive compatibility, but it is straightforward to check that these conditions are inconsistent. That is, when e^* is a finite lottery environment satisfying unrestricted domain, there exists no social choice function h^* such that (e^*, h^*) satisfies strict dominant strategy incentive compatibility and citizen sovereignty. This appears to contradict Theorem 14, which states that, assuming private values and value-distinguished types, every finite lottery environment satisfies strict separability.

Of course, there is not actually a contradiction here. For a lottery environment e^* satisfying private values and value-distinguished types, I construct a social choice function h^* such that (e^*, h^*) satisfies strict dominant strategy incentive compatibility but fails citizen sovereignty. This condition is eminantly reasonable for Hylland's purposes but irrelevant, and in fact detrimental, for mine.

Theorem 14 *Assume e^* is a finite lottery environment satisfying private values and value-distinguished types. Then e^* satisfies strict separability.*

Proof: Let $A = \{a_0, \ldots, a_K\}$. Private values and value-distinguished types imply the existence of a one-to-one correspondence between each agent i's types t_i and i's utility function $u_i(.|t_i)$ over outcomes, which is given by $v_i(.|t_i)$. Moreover, each $v_i(.|t_i)$ can be viewed as a vector $v^i(t_i) \in \mathbb{R}^A$ with components $v^i_k(t_i) = v_i(a_k|t_i)$. I will therefore identify agent i's type t_i with the vector $v^i = v^i(t_i)$, type $\tilde{t}_i$ with $\tilde{v}^i = v_i(\tilde{t}_i)$, and so on. It follows that each T_i is a subset of $K+1$-dimensional Euclidean space.

Obviously, A is itself a countable set over which no agent is indifferent at any state. By Proposition 13, to establish strict separability it suffices to exhibit a social choice function h^* such that, for all $i \in I$, all $\tilde{t}_i, \hat{t}_i \in T_i$ such that $\tilde{t}_i \neq \hat{t}_i$, and all $t_{-i} \in T_{-i}$,

$$u_i(h^*(\tilde{t}_i, t_{-i})|\tilde{t}_i) > u_i(h^*(\hat{t}_i, t_{-i})|\tilde{t}_i),$$

or equivalently,

$$\tilde{v}^i \cdot h^*(\tilde{v}^i, v^{-i}) > \tilde{v}^i \cdot h^*(\hat{v}^i, v^{-i}).$$

Value-distinguished types implies that the normalization $\nu : I \times \mathbb{R}^A \to \mathbb{R}^A$ defined by

$$\nu(i, v^i) = \frac{v^i - v_0^i(1, \ldots, 1)}{||v^i||}$$

is a one-to-one mapping, and it clearly does not affect the above inequalities. I therefore assume without loss of generality that, for all $i \in I$ and all $v^i \in T_i$, $v_0^i = 0$ and $||v^i|| = 1$.

Define h^* by

$$h_0^*(v) = \frac{1}{2} - \sum_{i \in I} \sum_{k=1}^{K} \frac{v_k^i}{2nK}$$

and, for all $k \geq 1$,

$$h_k^*(v) = \frac{1}{2K} + \sum_{i \in I} \frac{v_k^i}{2nK}$$

for all $v \in T$. Note that, for all $v \in T$, $h^*(v) \in \Delta(A, \mathcal{A})$. Then, for all $i \in I$, all $\tilde{v}^i, \hat{v}^i \in T_i$, and all $v^{-i} \in T_{-i}$,

$$\begin{aligned}
\tilde{v}^i \cdot h^*(\hat{v}, v_{-i}) &= \tilde{v}_0^i h_0^*(\hat{v}^i, v^{-i}) + \sum_{k=1}^{K} \tilde{v}_k^i h_k^*(\hat{v}^i, v^{-i}) \\
&= \sum_{k=1}^{K} \tilde{v}^i \left(\frac{1}{2K} + \frac{\hat{v}_k^i}{2nK} + \sum_{j \neq i} \frac{v_k^j}{2nK} \right) \\
&= \left(\sum_{k=1}^{K} \frac{\tilde{v}_k^i \hat{v}_k^i}{2nK} \right) + C \\
&= \frac{\tilde{v}^i \cdot \hat{v}^i}{2nK} + C,
\end{aligned}$$

where C is a term that does not depend on $\hat{v}^i$. Given $\tilde{v}^i$, clearly $\hat{v}^i = \tilde{v}^i$ is a unique maximum of this expression, as desired. ∎

The following corollary is an easy consequence of Theorems 8, 13, and 14.

Corollary 7 *Assume e^* is a finite lottery environment satisfying private values and value-distinguished types. Then f is virtually Bayesian implementable in e^* if and only if (e^*, f) satisfies incentive compatibility.*

Hylland (1980) also considers lottery environments with infinite sets of pure outcomes, which allows for the possibility of a finite environment with a transferable private good. I formally define this class of environments next. For a subset Q of Euclidean space, let $\mathcal{Q}_B$ denote the Borel σ-algebra on Q.

Definition 22 *e^* is a* **lottery environment with transfers** *if it is a lottery environment with $A = X \times Q$, where X is finite and Q is a non-empty open subset of $\mathbb{R}^I$, $\mathcal{A} = 2^X * \mathcal{Q}_B$, and for all $i \in I$ and all $t_i \in T_i$, there exist functions $c_i : T_i \rightarrow \mathbb{R}$, $d_i : T_i \rightarrow \mathbb{R}_{++}$, and $w_i : X \times T_i \rightarrow \mathbb{R}$ bounded in absolute value such that, for all $(x_k, q) \in A$, $v_i(x_k, q|t_i) = c_i(t_i) + d_i(t_i)(w_i(x_k|t_i) + q_i)$.*

In these environments, however, Hylland's assumption of unrestricted domain is untenable and his result has few ramifications for strict separability. Roberts (1979) considers environments with sets of pure outcomes consisting of a finite set crossed with a set of allocations of a transferable private good for which agents have increasing, additively separable preferences, but he does not allow for the possibility of randomizing over these pure outcomes. This possibility is, of course, crucial for the proof of strict separability in the next theorem.

Theorem 15 *Assume e^* is a lottery environment with transfers satisfying private values and value-distinguished types. Then e^* satisfies strict separability.*

Proof: Let $X = \{x_1, \ldots, x_K\}$ and let Q be an open set of $\mathbb{R}^I$ with elements $q = (q_1, \ldots, q_n)$. Private values and value-distinguished types imply the existence of a one-to-one correspondence between each agent i's types t_i and i's utility functions $u_i(.|t_i)$ over outcomes, which are given by $v_i(.|t_i) = c_i(t_i) + d_i(t_i)(w_i(.|t_i) + q_i)$. Moreover, each $v_i(.|t_i)$ can be viewed as a vector $(v^i(t_i), d_i(t_i)) \in \mathbb{R}_{K+2}$ with first $K+1$ components given by $v^i_k(t_i) = c_i(t_i) + d_i(t_i)w_i(x_k|t_i)$. I will therefore identify each type t_i with the vector $(v^i, d_i) = (v^i(t_i), d_i(t_i))$, each type $\tilde{t}_i$ with the vector $(\tilde{v}^i, \tilde{d}_i) = (v^i(\tilde{t}_i), d_i(\tilde{t}_i))$, and so on. The normalization $\nu : I \times \mathbb{R}^{K+2} \to \mathbb{R}^{K+2}$ defined by

$$\nu(i, v^i) = \frac{v^i - v^i_0(1, \ldots, 1)}{||d_i||}$$

is a one-to-one mapping, and as in the proof of Theorem 14, I can therefore assume without loss of generality that, for all $i \in I$ and all $v^i \in T_i$, $v^i_0 = 0$ and $d_i = 1$. I will further identify each $(v^i, 1)$ with v^i, so each T_i is a subset of $K+1$-dimensional Euclidean space, as in the proof of Theorem 14. Note that the components of the v^i are uniformly bounded, since w is bounded, but that it is not necessarily the case that $||v^i|| = 1$. It is for this reason that the social choice function constructed in the proof of Theorem 14 is inadequate for lottery environments with transfers.

A countable set of pure outcomes over which no agent is indifferent at any state is $\{(x_0, q), (x_0, q')\}$, where $q_i > q'_i$ for all $i \in I$. By Proposition 13, to establish strict separability it suffices to exhibit a social choice function h^* such that, for all $i \in I$, all $\tilde{t}_i, \hat{t}_i \in T_i$ with $\tilde{t}_i \neq \hat{t}_i$, and all $t_{-i} \in T_{-i}$,

$$u_i(h^*(\tilde{t}_i, t_{-i})|\tilde{t}_i) > u_i(h^*(\hat{t}_i, t_{-i})|\tilde{t}_i).$$

I construct a social choice function that, for all $t \in T$, picks a probability measure on A with degenerate marginal probability on Q. Letting $\rho^*(t)$ represent the allocation $q \in Q$ such that $h^*(X \times \{q\}) = 1$, and denoting the marginal probability measure of $h^*(t)$ on X by $\phi^*(t)$, h^* can be represented by the pair (ϕ^*, ρ^*). Specifically,

$$h^*(t)(A') = \sum_{\{x \in X | (x, \rho^*(t)) \in A'\}} \phi^*(t)(\{x\})$$

for all $t \in T$ and all $A' \in \mathcal{A}$. It is straightforward to check that, for all $i \in I$ and all $t \in T$,

$$u_i(h^*(t)|t_i) = v^i \cdot \phi^*(t) + \rho_i^*(t).$$

Then to establish strict separability, it suffices to construct a social choice function h^* such that, for all $\tilde{v}^i, \hat{v}^i \in T_i$ with $\tilde{v}^i \neq \hat{v}^i$ and for all $v^{-i} \in T_{-i}$,

$$\tilde{v}^i \cdot \phi^*(\tilde{v}^i, v^{-i}) + \rho_i^*(\tilde{v}^i, v^{-i}) > \tilde{v}^i \cdot \phi^*(\hat{v}^i, v^{-i}) + \rho_i^*(\hat{v}^i, v^{-i}).$$

In defining such a social choice function, I initially assume that $Q = \mathbb{R}^I$. Once this is done, I show how the construction can be easily adapted when Q is a proper open subset of $\mathbb{R}^I$.

Intuitively, h^* initially allots to each agent $1/2nK$ probability for each x_k, $k \geq 1$. These allotments are then adjusted by positive or negative amounts according to a specified cost function with all leftover probability placed on x_0. The cost function $C : (-1/2nK, 1/2nK) \to \mathbb{R}$ is defined by

$$C(p_k) \quad = \quad \begin{cases} \frac{p_k^2}{\frac{1}{2nK} - p_k} & \text{if } p_k \geq 0 \\ \frac{p_k^2}{\frac{1}{2nK} + p_k} & \textit{else.} \end{cases}$$

The important properties of C are $C'(0) = 0$,

$$\lim_{p_k \to 1/2nK} C'(p_k) = - \lim_{p_k \to -1/2nK} C'(p_k) = \infty,$$

and $C''(p_k) > 0$ for all $p_k \in (-1/2nK, 1/2nK)$. The adjustment to each agent's allotment is optimal given that agent's type. That is, the h^* solves

$$\max_{p_k \in (-1/2nK, 1/2nK) k \geq 1} p \cdot v^i - \sum_{k=1}^{K} C(p_k)$$

for all $i \in I$. Continuity of the objective function and boundedness of the T_i ensure that these problems have solutions, and strict concavity of C implies that the solution is unique and given by

$$\hat{v}_k^i = C'(p_k)$$

for $k \geq 1$. h^* then places, in addition to i's initial allotment, probability $C'^{-1}(v_k^i)$ on x_k at a cost of $C(C'^{-1}(v_k^i))$ for all $k \geq 1$. Note that this additional probability is negative if $v_k^i < 0$, in which case i prefers x_0 over x_k and would pay to have probability alloted from x_k to x_0.

Define (ϕ^*, ρ^*) as follows. I will write each $\phi^*(v)$ as the sum $\sum_{i\in I} \psi^*(v^i)$, where the function $\psi^* : \mathbb{R}^{K+1} \to \mathbb{R}^{K+1}$ is defined by

$$\psi^*_k(v^i) = \frac{1}{2nK} + C'^{-1}(v^i_k)$$

for $k \geq 1$, and

$$\psi^*_0(v^i) = \frac{1}{n} - \sum_{k=1}^{K} \psi^*_k(v^i).$$

Using the convention that $n+1 = 1$, define $\rho^* = (\rho^*_1, \ldots, \rho^*_n)$ by

$$\rho^*_i(v^i) = -\sum_{k=1}^{K} C(C'^{-1}(v^i_k)) + \sum_{k=1}^{K} C(C'^{-1}(v^{i+1}_k)).$$

Obviously, ρ^* maps to $Q = \mathbb{R}^I$, and in fact ρ^* is budget balancing. To see that ϕ^* maps to $\Delta(X, 2^X)$, note that

$$\begin{aligned}
\phi^*_0(v) &= \sum_{i\in I} \psi^*_0(v^i) = \left(\frac{1}{n} - \sum_{k=1}^{K} \psi^*_k(v^i)\right) \\
&= 1 - \sum_{i\in I}\sum_{k=1}^{K} \left(\frac{1}{2nK} + C'^{-1}(v^i_k)\right) \\
&= \frac{1}{2} - \sum_{i\in I}\sum_{k=1}^{K} C'^{-1}(v^i_k) \\
&> \frac{1}{2} - \frac{1}{2} = 0,
\end{aligned}$$

and

$$\begin{aligned}
\phi^*_k(v) &= \sum_{i\in I} \psi^*_k(v^i) = \sum_{i\in I} \left(\frac{1}{2nK} + C'^{-1}(v^i_k)\right) \\
&> \frac{1}{2K} - \frac{1}{2K} = 0.
\end{aligned}$$

Lastly,

$$\sum_{k=0}^{K} \phi^*_k(v) = 1 - \sum_{i\in I}\sum_{k=1}^{K} \left(\frac{1}{2nK} + C'^{-1}(v^i_k)\right)$$

$$+\sum_{i\in I}\sum_{k=1}^{K}\left(\frac{1}{2nK}+C'^{-1}(v_k^i)\right)=1,$$

as claimed.

Now take $i \in I$, $\tilde{v}^i, \hat{v}^i \in T_i$, and $v^{-i} \in T_{-i}$. Note that

$$\begin{aligned}
&\tilde{v}^i \cdot \phi^*(\hat{v}^i, v_{-i}) + \rho_i^*(\hat{v}^i, v^{-i}) \\
&= \tilde{v}_0^i \psi_0^*(\hat{v}^i) + \left(\sum_{k=1}^{K} \tilde{v}_k^i \psi_k^*(\hat{v}^i)\right) - \left(\sum_{k=1}^{K} C(C'^{-1}(\hat{v}_k^i))\right) + C_0 \\
&= \left(\sum_{k=1}^{K} \tilde{v}_k^i \left(\frac{1}{2nK} + C'^{-1}(\hat{v}_k^i)\right)\right) - \left(\sum_{k=1}^{K} C(C'^{-1}(\hat{v}_k^i))\right) + C_0,
\end{aligned}$$

where C_0 is a term that does not depend on $\hat{v}^i$. Then

$$\tilde{v}^i \cdot \phi^*(\tilde{v}^i, v^{-i}) + \rho_i^*(\tilde{v}^i, v^{-i}) > \tilde{v}^i \cdot \phi^*(\hat{v}^i, v^{-i}) + \rho_i^*(\hat{v}^i, v^{-i})$$

if and only if

$$\sum_{k=1}^{K} \tilde{v}_k^i C'^{-1}(\tilde{v}_k^i) - C(C'^{-1}(\tilde{v}_k^i)) > \sum_{k=1}^{K} \tilde{v}_k^i C'^{-1}(\hat{v}_k^i) - C(C'^{-1}(\hat{v}_k^i)).$$

This follows unless $\hat{v}^i = \tilde{v}^i$, since by construction

$$\tilde{v}_k^i - C(C'^{-1}(\tilde{v}_k^i)) > \tilde{v}_k^i p_k - C(C'^{-1}(p_k))$$

for all $k \geq 1$ and all $p_k \neq C'^{-1}(\tilde{v}_k^i)$. This establishes the desired result when $Q = \mathbb{R}^I$.

When Q is a proper open subset of $\mathbb{R}^I$ the same social choice function works, with some minor adaptations. Take $q \in Q$ and $\epsilon > 0$ such that $B_\epsilon(q) \subseteq Q$, and let $\overline{v}$ bound the absolute value of the components of all $v \in T$. It follows that each ρ_i^* is bounded in absolute value by $\overline{C} = KC(C'^{-1}(\overline{v}))$. Set $\delta = \epsilon/\overline{C}\sqrt{n}$.

It is straightforward to check that the social choice function represented by the pair $((1-\delta)\overline{x}_0 \oplus \delta\phi^*, q + \delta\rho^*)$ satisfies the conditions of strict separability. ∎

The following corollary is an easy consequence of Theorems 8, 13, and 15.

Corollary 8 *Assume e^* is a lottery environment with transfers satisfying private values and value distinguished types. Then f is virtually Bayesian implementable in e^* if and only if (e^*, f) satisfies incentive compatibility.*

4.4 Conclusion

In this chapter, I weaken the requirements of Bayesian implementability in lottery environments by considering the virtually Bayesian implementable social choice functions, for which there exist arbitrarily close Bayesian implementable neighbors. Since the predictions of Bayesian equilibrium must be accepted as approximations of behavior in any practical implementation problem, it should be sufficient to implement something close to a given social choice function. Thus, virtual Bayesian implementability is an acceptable standard. Surprisingly, it is also much weaker than Bayesian implementability in environments satisfying best-element-private values and strict separability. While Corollary 6 shows that incentive compatibility and extended Bayesian monotonicity are both necessary for Bayesian implementability in such environments, Theorem 13 shows that incentive compatibility is by itself sufficient for virtual implementability. This result improves on the existing characterizations of virtual Bayesian implementability, which are partial in nature and rely on the assumption that the set of states is finite.

The result also has an interpretation that sheds light on the condition of extended Bayesian monotonicity. Rephrasing, it shows that, for lottery environments satisfying best-element-private values and strict separability, the implementation problems satisfying extended Bayesian monotonicity are dense (in the d_1-metric) in the set of implementation problems satisfying incentive compatibility. This starkly contrasts the results of Muller and Satterthwaite (1977) and Duggan and Schwartz (1993) for Nash environments, which show that monotonicity is very restrictive.

Theorems 8 and 11 show that every lottery environment satisfying private values, a rather weak condition, also satisfies best-element-private values. Strict separability is, however, a strong condition not easily satisfied. Nonetheless, I show that, assuming private values and value-distinguished types, it is satisfied in finite lottery environments and in lottery environments with transfers. This class includes many environments of interest and entails no restrictions on the size of the set of states.

Appendix A

Proofs of Propositions

Proposition 1 *$\mu^*(\emptyset) = 0$; for all $S^1, S^2 \subseteq T$, $S^1 \subseteq S^2$ implies $\mu^*(S^1) \leq \mu^*(S^2)$; and for all $S^1, S^2 \subseteq T$, $\mu^*(S^1 \cup S^2) \leq \mu^*(S^1) + \mu^*(S^2)$.*

Proof: That $\mu^*(\emptyset) = 0$ follows since $\mu_i^*(\emptyset|t_i) = 0$ for all $i \in I$ and $t_i \in T_i$. Take $S^1, S^2 \subseteq T$ with $S^1 \subseteq S^2$. Suppose that $\mu^*(S^1) > \mu^*(S^2)$. Then there exists $i \in I$ and $t_i \in T_i$ such that $\mu_i^*(S^1|t_i) > \mu^*(S^2)$. But then

$$\mu^*(S^2) \geq \mu_i^*(S^2|t_i) \geq \mu_i^*(S^1|t_i) > \mu^*(S^2),$$

a contradiction. Now take $S^1, S^2 \subseteq T$ and suppose that $\mu^*(S^1 \cup S^2) > \mu^*(S^1) + \mu^*(S^2)$. Then there exists $i \in I$ and $t_i \in T_i$ such that $\mu_i^*(S^1 \cup S^2|t_i) > \mu^*(S^1) + \mu^*(S^2)$.

$$\mu_i^*(S^1 \cup S^2|t_i) \leq \mu_i^*(S^1|t_i) + \mu^*(S^2|t_i) \leq \mu^*(S^1) + \mu^*(S^2) < \mu_i^*(S^1 \cup S^2|t_i),$$

a contradiction. ∎

Proposition 2 *$\sim$ is reflexive, symmetric, and transitive.*

Proof: Reflexivity and symmetry are obvious. Now assume $f^1 \sim f^2 \sim f^3$, and note that

$$\{t \in T | f^1(t) \neq f^3(t)\} \subseteq \{t \in T | f^1(t) \neq f^2(t)\} \cup \{t \in T | f^2(t) \neq f^3(t)\}.$$

Since μ^* is an outer measure and $\mu^*(\{t \in T | f^1(t) \neq f^2(t)\}) = \mu^*(\{t \in T | f^2(t) \neq f^3(t)\}) = 0$, it follows that

$$\mu^*(\{t \in T | f^1(t) \neq f^2(t)\} \cup \{t \in T | f^2(t) \neq f^3(t)\}) = 0,$$

which implies $\mu^*(\{t \in T | f^1(t) \neq f^3(t)\}) = 0$. Therefore, $f^1 \sim f^3$. ∎

Proposition 3 *For all $i \in I$, all $\tilde{t}_i \in T_i$, and all $v \in V$, $\int^* v(t)\, d\mu_i(t|\tilde{t}_i)$ exists.*

Proof: Let $\overline{v}$ be an upper bound of v. Note that the set $\{w \in V | w \succeq v, w\ \mathcal{T} - \text{mble}\}$ contains the constant function with value $\overline{v}$, and is therefore non-empty. The set

$$\left\{ \int_T w(t)\, d\mu_i(t|\tilde{t}_i) \middle| w \succeq v, w\ \mathcal{T} - \text{mble} \right\}$$

is a set of real numbers bounded below by $-\overline{v}$, and therefore has a unique infimum. ∎

Proposition 4 *For all $i \in I$, all $\tilde{t}_i \in T_i$, and all $v \in V$, there exists a $\mathcal{T}$-measurable function $w \in V$ such that $w \succeq v$ and*

$$\int^* w(t)\, d\mu_i(t|\tilde{t}_i) = \int^* v(t)\, d\mu_i(t|\tilde{t}_i).$$

Proof: Consider following monotone decreasing sequence $\{w_k\}$ of $\mathcal{T}$-measurable functions with integrals converging to

$$\int^* v(t)\, d\mu_i(t|\tilde{t}_i).$$

By definition, for every positive integer k there exists a $\mathcal{T}$-measurable function w'_k such that $w'_k \succeq v$ and

$$\int_T w'_k(t)\, d\mu_i(t|\tilde{t}_i) - \int^* v(t)\, d\mu_i(t|\tilde{t}_i) \leq \frac{1}{k}.$$

Let $w_1 = w'_1$, and for all $k > 1$ define

$$w_k(t) = \min\{w_{k-1}(t), w'_k(t)\}$$

for all $t \in T$. Each w_k is $\mathcal{T}$-measurable, $w_k \succeq w_{k+1} \succeq v$, and their integrals converge to the desired quantity. Set $w = \lim_{k\to\infty} w_k$. Then w is $\mathcal{T}$-measurable, $w \succeq v$, and by the monotone convergence theorem,

$$\int_T w(t)\, d\mu_i(t|\tilde{t}_i) = \lim_{k\to\infty} \int_T w_k(t)\, d\mu_i(t|\tilde{t}_i) = \int^* v(t)\, d\mu_i(t|\tilde{t}_i),$$

as desired. ∎

Proposition 5 *For all $i \in I$, all $\tilde{t}_i \in T_i$, and all $v_1, v_2 \in V$, $\mu_i^*(\{t \in T | v_1(t) \neq v_2(t)\}|\tilde{t}_i) = 0$ implies*

$$\int^* v_1(t)\, d\mu_i(t|\tilde{t}_i) = \int^* v_2(t)\, d\mu_i(t|\tilde{t}_i).$$

Proof: Assume $\mu_i^*(\{t \in T | v_1(t) \neq v_2(t)\}|\tilde{t}_i) = 0$. Let w_1 be an upper approximations of v_1, and let

$$S' = \{t \in T | v_1(t) \neq v_2(t)\}.$$

It may be that $S' \notin \mathcal{T}$, but there exists $S \in \mathcal{T}$ such that $S' \subseteq S$ and $\mu_i(S|\tilde{t}_i) = 0$. Define $w_2 : T \to \mathbb{R}$ by

$$w_2(t) = \begin{cases} w_1(t) & \text{if } t \in T \setminus S \\ \overline{u} & \text{if } t \in S \end{cases}$$

for all $t \in T$. Then w_2 is $\mathcal{T}$-measurable, $w_2 \succeq v_2$, and

$$\begin{aligned} \int^* v_1(t)\, d\mu_i(t|\tilde{t}_i) &= \int_T w_1(t)\, d\mu_i(t|\tilde{t}_i) \\ &= \int_T w_2(t)\, d\mu_i(t|\tilde{t}_i) \\ &\geq \int^* v_2(t)\, d\mu_i(t|\tilde{t}_i). \end{aligned}$$

A symmetric argument establishes the opposite inequality, and yields the desired result. ∎

Proposition 6 *For all $i \in I$, all $\tilde{t}_i \in T_i$, and all $v_1, v_2 \in V$, $v_2 \succeq v_1$ implies*

$$\int^* v_2(t)\, d\mu_i(t|\tilde{t}_i) \geq \int^* v_1(t)\, d\mu_i(t|\tilde{t}_i).$$

Proof: Let w_2 be an upper approximation of v_2, and note that $w_2 \succeq v_1$, which immediately implies the desired inequality. ∎

Proposition 7 *For all $i \in I$, all $\tilde{t}_i \in T_i$, all $v_1, v_2 \in V$, and all $S \subseteq T$ with $\mu_i^*(S|\tilde{t}_i) > 0$,*

$$v_2 \succeq v_1 \text{ and } \forall t \in S\ v_2(t) > v_1(t)$$

implies

$$\int^* v_2(t) - v_1(t)\, d\mu_i(t|\tilde{t}_i) > 0.$$

Proof: Assume $v_2 \succeq v_1$ and, for all $t \in S$, $v_2(t) > v_1(t)$. Let w be an upper approximation of $v_2 - v_1$, let $S^+ = \{t \in T | v_2(t) - v_1(t) > 0\}$, and let $S^- = \{t \in T | v_2(t) - v_1(t) = 0\}$. Note that $T = S^+ \cup S^-$ and that $S \subseteq S^+$ implies $\mu_i^*(S^+|\tilde{t}_i) > 0$. Since w is an upper approximation, it must be that $w^{-1}((0,\infty)) \in \mathcal{T}$ and $S^+ \subseteq w^{-1}((0,\infty))$, which implies that $\mu_i(w^{-1}((0,\infty))|\tilde{t}_i)$ > 0 and

$$\int^* v_2(t) - v_1(t)\, d\mu_i(t|\tilde{t}_i) = \int_T w(t)\, d\mu_i(t|\tilde{t}_i) > 0,$$

as desired. ∎

Proposition 8 *For all $i \in I$, all $\tilde{t}_i \in T_i$, all $c \in \mathbb{R}$, and all $v \in V$ such that $v + c \in V$,*

$$\int^* v(t) + c\, d\mu_i(t|\tilde{t}_i) = c + \int^* v(t)\, d\mu_i(t|\tilde{t}_i).$$

Proof: Let w be an upper approximation of $v + c$, and note that $w - c$ is an upper approximation of v. Therefore,

$$\begin{aligned} \int^* v(t)\, d\mu_i(t|\tilde{t}_i) &= \int_T w(t) - c\, d\mu_i(t|\tilde{t}_i) \\ &= c + \int_T w(t)\, d\mu_i(t|\tilde{t}_i) \\ &= c + \int^* v(t) - c\, d\mu_i(t|\tilde{t}_i), \end{aligned}$$

establishing the desired result. ∎

Proposition 9 *For all $i \in I$, all $\tilde{t}_i \in T_i$, all $a, b \in \mathbb{R}_+$, and all $v_1, v_2 : T \to \mathbb{R}$,*

$$\int^* av_1(t) + bv_2(t)\, d\mu_i(t|\tilde{t}_i) \leq a \int^* v_1(t)\, d\mu_i(t|\tilde{t}_i) + b \int^* v_2(t)\, d\mu_i(t|\tilde{t}_i).$$

Proof: Let w be an upper approximation of $av_1 + bv_2$, let w_1 be an upper approximation of v_1, and let w_2 be an upper approximation of v_2. Clearly, $aw_1 + bw_2 \succeq av_1 + bv_2$, so

$$\begin{aligned}
\int^* av_1(t) + bv_2(t)\, d\mu_i(t|\tilde{t}_i) &\leq \int_T aw_1(t) + bw_2(t)\, d\mu_i(t|\tilde{t}_i) \\
&= a\int_T w_1(t)\, d\mu_i(t|\tilde{t}_i) + b\int_T w_2(t)\, d\mu_i(t|\tilde{t}_i) \\
&= a\int^* v_1(t)\, d\mu_i(t|\tilde{t}_i) + b\int^* v_2(t)\, d\mu_i(t|\tilde{t}_i),
\end{aligned}$$

establishing the desired result. ∎

Proposition 10 *For all $i \in I$, all $\tilde{t}_i \in T_i$, and all $v \in V$,*

$$-\int^* v(t)\, d\mu_i(t|\tilde{t}_i) \leq \int^* -v(t)\, d\mu_i(t|\tilde{t}_i).$$

Proof: By Proposition 9, it follows that

$$0 = \int^* v(t) - v(t)\, d\mu_i(t|\tilde{t}_i) \leq \int^* v(t)\, d\mu_i(t|\tilde{t}_i) + \int^* -v(t)\, d\mu_i(t|\tilde{t}_i),$$

which establishes the desired result. ∎

Proposition 11 *For all $i \in I$ and all $\tilde{t}_i \in T_i$, if $\{v_k\}$ is a sequence of functions in V with $v_k \to v \in V$ then*

$$\int^* v(t)\, d\mu_i(t|\tilde{t}_i) \leq \lim\inf_{k\to\infty} \int^* v_k(t)\, d\mu_i(t|\tilde{t}_i).$$

Proof: Let $\{w_k\}$ denote a sequence of upper approximations of $\{v_k\}$, and let $w = \lim\inf_{k\to\infty} w_k$. Then w is $\mathcal{T}$-measurable and

$$\int_T w(t)\, d\mu_i(t|\tilde{t}_i) \leq \lim\inf_{k\to\infty} \int_T w_k(t)\, d\mu_i(t|\tilde{t}_i)$$

$$= \liminf_{k\to\infty} \int^* v_k(t)\, d\mu_i(t|\tilde{t}_i)$$

where the first inequality follows from Fatou's Lemma. To see that $w \succeq v$, take $t \in T$ and suppose that $v(t) - w(t) = \epsilon > 0$. Take k such that $v(t) - v_k(t) < \epsilon/2$ and $w_k(t) - w(t) < \epsilon/2$, which is possible since $v_k(t)$ converges to $v(t)$. Then

$$\begin{aligned} v(t) - w(t) &= (v(t) - v_k(t)) + (v_k(t) - w_k(t)) + (w_k(t) - w(t)) \\ &< \epsilon, \end{aligned}$$

a contradiction. Therefore, $w \succeq v$, and it follows that

$$\int^* v(t)\, d\mu_i(t|\tilde{t}_i) \leq \int_T w(t)\, d\mu_i(t|\tilde{t}_i),$$

which establishes the desired result. ∎

Proposition 12 *For all $i \in I$ and all $\tilde{t}_i \in T_i$, if $\{v_k\}$ is a sequence of functions in V with $v_k \to v \in V$ uniformly then*

$$\int^* v(t)\, d\mu_i(t|\tilde{t}_i) \geq \limsup_{k\to\infty} \int^* v_k(t)\, d\mu_i(t|\tilde{t}_i).$$

Proof: Let $\{w_k\}$ be a sequence of upper approximations of $\{v_k\}$, let $w = \limsup_{k\to\infty} w_k$, and let w' be an upper approximation of v. Note that

$$\begin{aligned} \limsup_{k\to\infty} \int^* v_k(t)\, d\mu_i(t|\tilde{t}_i) &= \limsup_{k\to\infty} \int_T w_k(t)\, d\mu_i(t|\tilde{t}_i) \\ &\leq \int_T w(t)\, d\mu_i(t|\tilde{t}_i), \end{aligned}$$

where the inequality follows from a variant of Fatou's Lemma, so it suffices to show that

$$\int_T w(t)\, d\mu_i(t|\tilde{t}_i) \geq \int_T w'(t)\, d\mu_i(t|\tilde{t}_i).$$

If not, then $S = \{t \in T | w(t) > w'(t)\} \in \mathcal{T}$ and $\mu_i(S|\tilde{t}_i) > 0$. Note that $S = \bigcup_{z \in \mathbb{Z}_{++}} S_z$, where $S_z = \{t \in T | w(t) > w'(t) + 1/z\}$. By countable additivity of $\mu_i(.|\tilde{t}_i)$, there exists $z \in \mathbb{Z}_{++}$ such that $\mu_i(S_z|\tilde{t}_i) > 0$. Since $v_k \to v$ uniformly, there exists K such that, for all $k \geq K$ and all $t \in S_z$, $|v_k(t) - v(t)| < 1/3z$.

Now take any $k \geq K$ and note that $S_z^k = \{t \in S_z | w_k(t) - v_k(t) > 1/3z\} \in \mathcal{T}$ has $\mu_i(.|\tilde{t}_i)$-measure zero, for otherwise the $\mathcal{T}$-measurable function $w'_k \in V$ defined by

$$w'_k(t) = \begin{cases} w_k(t) - \frac{1}{3z} & \text{if } t \in S_z^k \\ w_k(t) & \text{else} \end{cases}$$

for all $t \in T$ satisfies $w'_k \succeq v_k$ and

$$\int_T w'_k(t)\, d\mu_i(t|\tilde{t}_i) < \int_T w_k(t)\, d\mu_i(t|\tilde{t}_i).$$

Note also that $Q_z^k = \{t \in S_z | w_k(t) > w(t) - 1/3z\} \subseteq S_z^k$, since

$$w_k(t) > w(t) - \frac{1}{3z} > w'(t) + \frac{2}{3z} \geq v(t) + \frac{2}{3z} > v_i(t) + \frac{1}{3z}$$

for all $t \in Q_z^k$. Of course, $Q_z^k \in \mathcal{T}$, and it follows that $\mu_i(Q_z^k|\tilde{t}_i) = 0$. But

$$S_z \subseteq \bigcup_{k=K}^{\infty} Q_z^k,$$

which implies that $\mu_i(S_z|\tilde{t}_i) = 0$. This contradiction establishes the desired result. ∎

Proposition 13 *An extended expected utility environment satisfies strict separability if and only if there exists $h^* \in F$ such that, for all $i \in I$, all $\tilde{t}_i \in T_i$, and all $\alpha \in \Sigma(T)$ with $\alpha_i(\tilde{t}_i) \neq \tilde{t}_i$,*

$$\mu_i^* \left(\{ t \in T | u_i((h^* \circ (\tau_i, \alpha_{-i}))(t)|t) > u_i((h^* \circ \alpha)(t)|t)\} | \tilde{t}_i \} \right) = 1.$$

Proof: First consider the necessity of this condition. Let S denote the set of states at which the outcomes of $h^* \circ (\tau_i, \alpha_{-i})$ are preferred to the outcomes of $h^* \circ \alpha$. Suppose that $\mu_i^*(S|\tilde{t}_i) < 1$, so that $\mu_i^*(T \setminus S|\tilde{t}_i) > 0$. Then Proposition 6 implies

$$\int^* u_i((h/_{T\setminus S}(h^* \circ (\tau_i, \alpha_{-i})))(t)|t) - u_i((h/_{T\setminus S}(h^* \circ \alpha))(t)|t)\, d\mu_i(t|\tilde{t}_i) \leq 0.$$

This implies, by Proposition 10 and the comments in Section 2.4, that

$$h/_{T\setminus S}(h^* \circ \alpha)\ \tilde{R}_i(\tilde{t}_i)\ h/_{T\setminus S}(h^* \circ (\tau_i, \alpha_{-i})),$$

contradicting strict separability.

To see sufficiency, take $S \subseteq T$ with $\mu_i^*(S|\tilde{t}_i) > 0$, and let $S' = \{t \in T | u_i((h^* \circ (\tau_i, \alpha_{-i}))(t)|t) - u_i((h^* \circ \alpha)(t)|t)\}$. Note that $\mu_i^*(S'|\tilde{t}_i) = \mu_i^*(S|\tilde{t}_i) > 0$. Then by Propositions 5 and 7,

$$\begin{aligned}
&\int^* u_i((h/_S(h^* \circ (\tau_i, \alpha_{-i})))(t)|t) - u_i((h/_S(h^* \circ \alpha))(t)|t)\, d\mu_i(t|\tilde{t}_i) \\
&= \int^* u_i((h/_{S'}(h^* \circ (\tau_i, \alpha_{-i})))(t)|t) - u_i((h/_{S'}(h^* \circ \alpha))(t)|t)\, d\mu_i(t|\tilde{t}_i) \\
&> 0,
\end{aligned}$$

and by Propositions 5 and 6,

$$\int^* u_i((h/_S(h^* \circ \alpha))(t)|t) - u_i((h/_S(h^* \circ (\tau_i, \alpha_{-i})))(t)|t)\, d\mu_i(t|\tilde{t}_i)$$

$$= \int^* u_i((h/_{S'}(h^* \circ \alpha))(t)|t) - u_i((h/_{S'}(h^* \circ (\tau_i, \alpha_{-i})))(t)|t)\, d\mu_i(t|\tilde{t}_i)$$
$$\leq 0.$$

These two inequalities imply

$$\int^* u_i((h/_S(h^* \circ (\tau_i, \alpha_{-i})))(t)|t) - u_i((h/_S(h^* \circ \alpha))(t)|t)\, d\mu_i(t|\tilde{t}_i)$$
$$- \int^* u_i((h/_S(h^* \circ \alpha))(t)|t) - u_i((h/_S(h^* \circ (\tau_i, \alpha_{-i})))(t)|t)\, d\mu_i(t|\tilde{t}_i)$$
$$> 0,$$

so that $h/_S(h^* \circ (\tau_i, \alpha_{-i}))\ P_i^*(\tilde{t}_i)\ h/_S(h^* \circ \alpha)$. ∎

Bibliography

[1] Abreu, Dilip and Hitoshi Matsushima. (1990) "Exact Implementation." mimeograph.

[2] Abreu, Dilip and Hitoshi Matsushima. (1990, revised 1992) "Virtual Implementation in Iteratively Undominated Strategies: Incomplete Information." mimeograph.

[3] Abreu, Dilip and Hitoshi Masushima. (1992) "Virtual Implementation in Iteratively Undominated Strategies: Complete Information." *Econometrica* 60: 993-1008.

[4] Abreu, Dilip and Arunava Sen. (1991) "Virtual Implementation in Nash Equilibrium." *Econometrica* 59: 997-1021.

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

[6] Duggan, John. (1994) "Virtual Implementation in Bayesian Equilibrium with Infinite Sets of Types." mimeograph.

[7] Duggan, John and Thomas Schwartz. (1993) "Bogosity Results for Maskin Monotonic Social Choice Rules." mimeograph.

[8] Gibbard, Allan. (1973) "Manipulation of Voting Schemes: A General Result." *Econometrica* 41: 587-601.

[9] Hylland, Aanund. (1980) "Strategy Proofness of Voting Procedures with Lotteries as Outcomes and Infinite Sets of Strategies." mimeograph.

[10] Jackson, Matthew O. (1991) "Bayesian Implementation." *Econometrica* 59: 461-477.

[11] Jackson, Matthew O. (1992) "Implementation in Undominated Strategies: A Look at Bounded Mechanisms." *Review of Economic Studies* 59: 757-775.

[12] Kingman, J. F. C. and S. J. Taylor. (1966) *Introduction to Measure and Probability.* Cambridge Press.

[13] Maskin, Eric S. (1986) "The Theory of Implementation in Nash Equilibrium: A Survey." In Leonid Hurwicz, David Schmeidler, and Hugo Sonnenschein, eds., *Social Goals and Social Organization.* New York: Cambridge.

[14] Matsushima, Hitoshi. (1988) "A New Approach to the Implementation Problem." *Journal of Economic Theory* 45: 128-144.

[15] Matsushima, Hitoshi. (1993) "Bayesian Monotonicity with Side Payments." *Journal of Economic Theory* 59: 107-121.

[16] Mookherjee, Dilip and Stefan Reichelstein. (1990) "Implementation via Augmented Revelation Mechanisms." *Review of Economic Studies* 57: 453-475.

[17] Muller, Eitan and Mark A. Satterthwaite. (1977) "The Equivalence of Strong Positive Association and Strategy-Proofness." *Journal of Economic Theory* 14: 412-418.

[18] Myerson, Roger B. (1979) "Incentive Compatibility and the Bargaining Problem." *Econometrica* 47: 61-73.

[19] Myerson, Roger B. (1991) *Game Theory: Analysis of Conflict.* Cambridge: Harvard.

[20] Ordeshook, Peter C. (1986) *Game Theory and Political Theory.* New York: Cambridge.

[21] Palfrey, Thomas R. (1992) "Implementation in Bayesian Equilibrium: The Multiple Equilibrium Problem in Mechanism Design." In Jean-Jacques Laffont, ed., *Advances in Economics Theory, Sixth World Congress, vol. 1.* Cambridge Press.

[22] Palfrey, Thomas R. and Sanjay Srivastava. (1989) "Implementation with Incomplete Information in Exchange Economies." *Econometrica* 57: 115-134.

[23] Palfrey, Thomas R. and Sanjay Srivastava. (1991) "Efficient Trading Mechanisms with Pre-play Communication." *Journal of Economic Theory* 55: 17-40.

[24] Palfrey, Thomas R. and Sanjay Srivastava. (1993) *Bayesian Implementation.* Fundamentals of Pure and Applied Economics 53. Langhorne PA: Harwood.

[25] Postlewaite, Andrew and David Schmeidler. (1986) "Implementation in Differential Information Economies." *Journal of Economic Theory* 39: 14-33.

[26] Roberts, Kevin. (1979) "The Characterization of Implementable Choice Rules." In Jean-Jacques Laffont, ed., *Aggregation and Revelation of Preferences.* North-Holland.

[27] Satterthwaite, Mark Allen. (1975) "Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions." *Journal of Economic Theory* 10: 187-217.