Foundations and Trends® in
Microeconomics
Vol. 8, No. 3 (2012) 169–230

DOI: 10.1561/0700000057

now
the essence of knowledge

An Introduction to Robust Mechanism Design

By Dirk Bergemann and Stephen Morris

Contents

Foundations and Trends® in
Microeconomics
Vol. 8, No. 3 (2012) 169–230

DOI: 10.1561/0700000057

now
the essence of knowledge

An Introduction to Robust Mechanism Design*

Dirk Bergemann[1] and Stephen Morris[2]

[1] *Department of Economics, Yale University, New Haven, USA dirk.bergemann@yale.edu.*
[2] *Department of Economics, Princeton University, Princeton, USA smorris@princeton.edu.*

Abstract

This essay provides an introduction to our recent work on robust mechanism design. The objective is to provide an overview of the research agenda and its results. We present the main results and illustrate many of them in terms of a common and canonical example, the single unit auction with interdependent values. In addition, we provide an extended discussion about the role of alternative assumptions about type spaces in our work, and the literature at large, in order to explain

* We would like to thank Eric Maskin for inviting us to publish the work covered in this survey in a collection of the World Scientific Series in Economic Theory edited by Eric. An early version of this essay appeared as an introduction in Bergemann and Morris (2012b). We would like to thank our co-authors Hanming Fang, Moritz Meyer-ter-Vehn, Karl Schlag, Satoru Takahashi and Olivier Tercieux in this research agenda and Nemanja Antic, Andreas Blume, Tilman Borgers, Jacques Cremer, Moritz Meyer-ter-Vehn, Phil Reny and Olivier Tercieux for comments on this essay. We had the opportunity to deliver the present material at a number of invited lectures, notably at Boston University, Northwestern University and the European and North American Econometric Society Meetings and a set of slides which cover and accompany this essay can be found at http://dirkbergemann.commons.yale.edu/files/2010/12/robustmechanismdesign1.pdf.

the common logic of the informational robustness approach that unifies the work.

Keywords: Mechanism design; robust mechanism design; common knowledge; universal type space; interim equilibrium; ex post equilibrium; dominant strategies; rationalizability; partial implementation; full implementation; robust implementation.

JEL Codes: C79, D82

1

Introduction

This essay brings together and presents a number of results on the theme of robust mechanism design and robust implementation that we have been working on in the past decade. This work examines the implications of relaxing the strong informational assumptions that drive much of the mechanism design literature. It discusses joint work of the two of us with each other and with co-authors Hanming Fang, Moritz Meyer-ter-Vehn, Karl Schlag, Satoru Takahashi, and Olivier Tercieux.

The objective of this essay is to provide the reader with an overview of the research agenda pursued in these papers. We present the main results of these papers and illustrate many of them in terms of a common and canonical example, the single unit auction with interdependent values. It is our hope that the use of this example facilitates the presentation of the results and that it brings the main insights within the context of an important economic mechanism, the generalized second price auction. In addition, we include an extended discussion about the role of alternative assumptions about type spaces in our work and the literature, in order to explain the common logic of the informational robustness approach that unifies the work surveyed in this essay.

The mechanism design literature of the last thirty years has been a huge success on a number of different levels. There is a beautiful theoretical literature that has shown how a wide range of institutional design questions can be formally posed as mechanism design problems with a common structure. Elegant characterizations of optimal mechanisms have been obtained. Market design has become more important in many economic arenas, both because of new insights from theory and developments in information and computing technologies, which enable the implementation of large scale trading mechanisms. A very successful econometric literature has tested auction theory in practise.

However, there has been an unfortunate disconnect between the general theory and the applications/empirical work: mechanisms that work in theory or are optimal in some class of mechanisms often turn out to be too complicated to be used in practise. Practitioners have then often been led to argue in favor of using simpler but apparently sub-optimal mechanisms. It has been argued that the optimal mechanisms are not "robust" — i.e., they are too sensitive to fine details of the specified environment that will not be available to the designer in practise. These concerns were present at the creation of the theory and continue to be widespread today.[1] In response to the concerns, researchers have developed many attractive and influential results by imposing (in a somewhat ad hoc way) stronger solution concepts and/or simpler mechanisms motivated by robustness considerations. Our starting point is the influential concern of Wilson (1987) regarding the robustness of the game theoretic analysis to the common knowledge assumptions:

> "Game theory has a great advantage in explicitly analyzing the consequences of trading rules that presumably are really common knowledge; it is deficient to the extent it assumes other features to be common

[1] Hurwicz (1972) discussed the need for "non-parametric" mechanisms wich are independent of the distributional assumptions regarding the willingness-to-pay of the agents. Wilson (1985) states that trading rules should be "belief-free" by requiring that they "should not rely on features of the agents' common knowledge, such as their probability assessments." Dasgupta and Maskin (2000) seek "detail-free" auction rules "that are independent of the details — such as functional forms or distribution of signals - of any particular application and that work will in a broad range of circumstances."

> knowledge, such as one agent's probability assessment about another's preferences or information. I foresee the progress of game theory as depending on successive reductions in the base of common knowledge required to conduct useful analyses of practical problems. Only by repeated weakening of common knowledge assumptions will the theory approximate reality."

Wilson emphasized that as analysts we are tempted to assume that too much information is common knowledge among the agents, and suggested that more robust conclusions would arise if researchers were able to relax those common knowledge assumptions. Harsanyi (1967–68) had the original insight that relaxing common knowledge assumptions is equivalent to working with a type space which is larger if there is less common knowledge. A natural theoretical question then is to ask whether it is possible to explicitly model the robustness considerations in such a way that stronger solution concepts and/or simpler mechanisms emerge endogenously. In other words, if the optimal solution to the planner's problem is too complicated or too sensitive to be used in practice, it is presumably because the original description of the planner's problem was itself flawed. We would like to investigate if improved modelling of the planner's problem endogenously generates the "robust" features of mechanisms that researchers have been tempted to assume. Our research agenda in robust mechanism design is therefore to *first* make explicit the implicit common knowledge assumptions and then *second* to weaken them.

Thus, formally, our approach suggests asking what happens to the conventional insights in the theory of mechanism design when confronted with larger and richer type spaces with weaker requirements regarding the common knowledge of between the designer and the agents. In this respect, a very important contribution is due to Neeman (2004) who showed that the small type space assumption is of special importance for the full surplus extraction results, as in Myerson (1981) and Cremer and McLean (1988). In particular, he showed that the full surplus extraction results fail to hold if agents' private information doesn't display a one-to-one relationship between each agent's beliefs

about the other agents and his preferences (valuation). The extended dimensionality relative to the standard model essentially allows for a richer set of higher-order beliefs.

Similarly, in an analysis of the first price auction, Fang and Morris (2006) look at the role of richer type spaces by considering private values but allowing for multidimensional types. There, each bidder observes his own private valuation and a noisy signal of his opponent's private valuation. This model of private information stands in stark contrast to the standard analysis of auctions with private values, where each agent's belief about his competitor is simply assumed to coincide with the common prior. In the presence of multidimensional private signals, Fang and Morris (2006) find that the celebrated revenue equivalence result between the first and the second price auction fails to hold. With the richer type space, it is not even possible to rank the auction format with respect to their expected revenue.

2

Leading Example: Allocating a Private Good with Interdependent Values

It is the objective of this essay to present the main themes and results of our research on robust mechanism design through a prominent example, namely the efficient allocation of a single object among a group of agents. We are considering the following classic single good allocation problem with interdependent values. There are I agents. Each agent i has a "payoff type" $\theta_i \in \Theta_i = [0,1]$. Write $\Theta = \Theta_1 \times \cdots \times \Theta_I$. Each agent i has a quasi-linear utility function and attaches monetary value $v_i : \Theta \to \mathbb{R}$ to getting the object, where the valuation function v_i has the following linear form:

$$v_i(\theta) = \theta_i + \gamma \sum_{j \neq i} \theta_j.$$

The parameter γ is a measure of the interdependence in the valuations. If $\gamma = 0$, then we have the classic private values case. If $\gamma > 0$, we have positive interdependence in values, if $\gamma < 0$, we have negative interdependence. If $\gamma = 1$, then we have a model of common values.

In this setting, a social choice function must specify the allocation of the object and the (expected) payments that agents make as a function of the payoff type profile. Thus a social choice function f can be written as $f(\theta) = (q(\theta), y(\theta))$ where the allocation rule determines the

probability $q_i(\theta)$ that agent i gets the object if the type profile is θ, with $q(\theta) = (q_1(\theta), \ldots, q_I(\theta))$; and transfer function, $y(\theta) = (y_1(\theta), \ldots, y_I(\theta))$, where $y_i(\theta)$ determines the payment that agent i makes to the planner.

If $\gamma < 1$, then the socially efficient allocation is to give the object to an agent with the highest payoff type θ_i, provided that $v_i(\theta) \geq 0$, for otherwise the socially efficient allocation is not to assign the object at all. With this caveat, *an* efficient allocation rule is given by:

$$q_i^*(\theta) = \begin{cases} \dfrac{1}{\#\{k : \theta_k \geq \theta_j \text{ for all } j\}}, & \text{if } \theta_i \geq \theta_j \text{ for all } j; \\ 0, & \text{otherwise.} \end{cases}$$

The specific form of the tie-breaking rule, here simply assumed to be uniform by construction of $q_i^*(\theta)$, is without importance. If $\gamma = 1$, there are common values and all allocations are efficient, but the above q^* continues to form an efficient allocation rule. While the papers surveyed in this essay deal with general allocation problems — and in particular, are not restricted to quasi-linear environments — this survey is explicitly using this example and focussing on the efficient allocation rule.

Now let us consider mechanisms for allocating the object. Suppose for the moment that we were in a private value environment, i.e., $\gamma = 0$. Then a well-known mechanism to achieve the efficient allocation is the second price sealed bid auction. Here, each player i announces a "bid" $b_i \in [0, 1]$, and the object is allocated to the highest bidder who pays the second highest bid. Each agent has a dominant strategy to bid his true payoff type θ_i and the object is allocated efficiently. The second price sealed bid mechanism is a specific instance of a Vickrey–Clarke–Groves mechanism which are known to achieve efficiency and incentive compatibility in dominant strategies for a large class of allocation problems in private value environments with quasi-linear utility.

Maskin (1992) introduced a suitable generalization of the Vickrey–Clarke–Groves mechanism to an environment with interdependent values. With interdependence, that is for $\gamma \neq 0$, the "generalized" Vickrey–Clarke–Groves mechanism asks each agent i to report, or "bid" $b_i \in [0, 1]$, but now the rule of the "generalized" second price sealed bid auction is that agent i with the highest report or "bid" wins, and pays

the second highest bid plus γ times the sum of the bid of others:

$$y_i^*(b) = \left(\max_{k \neq i} \{b_k\} + \gamma \sum_{j \neq i} b_j \right) q_i^*(b).$$

We observe that if $\gamma = 0$, then the payment rule of the "generalized" second price sealed bid auction reduces to the familiar rule of the second price sealed bid auction. If agents bid "truthfully," setting their bid b_i equal to their payoff type θ_i, then the generalized second price auction leads to the realization of the social choice function $f^*(\theta) = (q^*(\theta), y^*(\theta))$.

As long as parameter of interdependence is $\gamma \leq 1$, ensuring that a single crossing property is satisfied, this social choice function is "ex post incentive compatible." That is, if an agent expected other agents to report their types truthfully, he has an incentive to report his type truthfully. Conditioning on truthtelling by the other agents, the utility of a winning bidder who tells the truth is

$$\left(\theta_i + \gamma \sum_{j \neq i} \theta_j \right) - \left(\max_{k \neq i} \{\theta_k\} + \gamma \sum_{j \neq i} \theta_j \right).$$

This expression is greater than 0 if $\theta_i > \max_{k \neq i} \{\theta_k\}$ and less than 0 if $\theta_i < \max_{k \neq i} \{\theta_k\}$.

We observe that the winning bidder cannot affect the transfer through this report; this is the VCG aspect of the generalized second price auction. Now if his payoff type is larger than the payoff type of everybody else, he would like to win the object, and thus he cannot do better than bid his true value. On the other hand, if agent i's payoff type is lower than the highest payoff type among the remaining bidders, then he would have to report a higher type to receive object, but as $\theta_i < \max_{k \neq i} \{\theta_k\}$, the resulting net utility for bidder i would be negative. If $\theta_i = \max_{k \neq i} \{\theta_k\}$, the agent would be indifferent between winning the object or not. We have thus established that the efficient allocation is implemented with ex post incentive compatibility conditions. Thus the generalized second price auction ensured that, for all beliefs and higher-order beliefs, there is an equilibrium that leads to the efficient allocation.

This mechanism is "robust" in the sense that as long as there is common knowledge of the environment and payoffs as we described them, there will be an equilibrium where the efficient allocation rule is followed whatever the beliefs and higher-order beliefs of the agents about the payoff types of the other agents. Ex post incentive compatibility is clearly *sufficient* for "partial robust implementation," i.e., the existence of a mechanism with the property that, whatever agents' beliefs and higher-order beliefs, there is an equilibrium giving rise to the efficient allocation. In Bergemann and Morris (2005), we study when the existence of an ex post incentive compatible direct mechanism is *necessary* for partial robust implementation. But formalizing this question is delicate, and has been the subject of some confusion in the literature. In the next section, we will discuss how the language of type spaces can be used to formalize this and other questions and to highlight some subtleties in the formalization.

3

Type Spaces

We will be interested in situations where there is common knowledge of the structure of the environment described in the previous section, but the planner may not know much about each agent's beliefs or higher-order beliefs about other agents' types. Thus rather than making the usual "Bayesian" assumption that the planner knows some true common prior over $\Theta = \Theta_1 \times \cdots \times \Theta_I$, we want to be able to capture the planner's uncertainty about agents' types, and what each agent believes about other agents' types, by allowing richer type spaces.

It is important to study type spaces that are richer than Θ, because we want to allow for the possibility that two types of an agent may be identical from a payoff type perspective, but have different beliefs about, say, the payoff types of other agents. In addition, we want to allow for interim type spaces, where there are no restrictions on a type's interim belief about other agents' types. Requiring that types' interim beliefs be derived from some prior probability distribution on the type space, in other words that the type space constitutes a common prior type space, will then represent an important special case. In what follows, we will focus on finite type spaces but our results readily extend to infinite type spaces and some of the work to be discussed explicitly considers such infinite type spaces.

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

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

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

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

with $\widehat{\pi}_i(t_i)$ being agent is *belief type* when his type is t_i. Thus $\widehat{\pi}_i(t_i)[E]$ is the probability that type t_i of agent i assigns to other agents' types, t_{-i}, being an element of $E \subseteq T_{-i}$. We will abuse notation slightly by writing $\widehat{\pi}_i(t_i)[t_{-i}]$ for the probability that type t_i of agent i assigns to other agents having types t_{-i}. Now a *type space* is a collection

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

The standard approach in the mechanism design literature is to assume a common knowledge prior, $p \in \Delta(\Theta)$, on the set of payoff types Θ. This standard approach can be modelled in our language by identifying the set of types T_i with the payoff types Θ_i and defining beliefs by

$$\widehat{\pi}_i(\theta_i)[\theta_{-i}] \triangleq \frac{p(\theta_i, \theta_{-i})}{\sum_{\theta'_{-i} \in \Theta_{-i}} p(\theta_i, \theta'_{-i})}.$$

It is useful to distinguish two distinct, critical and strong, assumptions embedded in the standard approach. First, it is assumed that there is a unique belief type associated with each payoff type. More precisely, we will say that a type space $\mathcal{T}$ is a *payoff type space* if each $\widehat{\theta}_i$ is a bijection, so that the set of possible types is identified with the set of payoff types. While often motivated by analytic convenience, when maintained in particular applications, this assumption is often strong and unjustified. This assumption need not be paired with the common prior assumption, but it often is. Type space $\mathcal{T}$ is a *common prior type space* if there exists $\pi \in \Delta(T)$ such that

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

and

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

Thus the standard approach consists of requiring both that $\mathcal{T}$ is a payoff type space and that $\mathcal{T}$ is a common prior type space. We can think of this as the smallest type space that is used in the Bayesian analysis that embeds the payoff environment described above. The standard approach makes strong common knowledge assumptions of the type that Wilson (1987) and others have argued should be expunged from mechanism design. For example, a well known implication of the standard approach is that if the common prior p is picked generically (under Lebesgue measure), the seller is able to fully extract the agents' surplus (Myerson, 1981; Cremer and McLean, 1988). While the insight that correlation in agents' types can be exploited seems to be an economically important one, it is clear that full surplus extraction is not something which can be carried out in practise. While a number of assumptions underlying the model of full surplus extraction,[1] Neeman (2004) highlights the role of the implausible assumption that "beliefs determine preferences" (BDP), i.e., that there is a common knowledge of a mapping that identifies a unique possible valuation associated with any given belief over others' types. The innocuous looking "genericity" assumption obtains its bite by being combined with the strong common knowledge assumptions entailed by the payoff type space restriction.

To illustrate the role of richer type spaces, let us consider an example from Fang and Morris (2006). Suppose there are two agents, i and j, whose valuations of the object are either low (v_l) or high (v_h), with each valuation equally likely. In addition, each agent observes a low (l) or high (h) signal, $s \in \{l, h\}$, which correctly reflects the other agent's valuation with probability $q \geq \frac{1}{2}$. This situation is modelled in the

[1] Robert (1991), Laffont and Martimort (2000), and Peters (2001) highlight the importance of risk neutrality and unlimited liability, absence of collusion and absence of competition, respectively.

language of this essay by setting $I = 2$; $\Theta_i = \{v_l, v_h\}$; $T_i = \{v_l, v_h\} \times \{l, h\}$; $\widehat{\theta}_i(\theta_i, s_i) = \theta_i$; writing $s_j \simeq \theta_i$ if $(\theta_i, s_j) = (v_l, l)$ or (v_h, h),

$$\widehat{\pi}_i((\theta_i, s_i))[(\theta_j, s_j)] = \begin{cases} q^2, & \text{if } s_i \simeq \theta_j \text{ and } s_j \simeq \theta_i; \\ q(1-q), & \text{if } s_i \simeq \theta_j \text{ but not } s_j \simeq \theta_i; \\ q(1-q), & \text{if } s_j \simeq \theta_i \text{ but not } s_i \simeq \theta_j; \\ (1-q)^2, & \text{if neither } s_i \simeq \theta_j \text{ nor } s_j \simeq \theta_i. \end{cases}$$

In this type space, there are independent private values as represented by the payoff types, but there are multidimensional types. The BDP property ("beliefs determine preferences") fails because an agent's beliefs about others' types depend only on his signal and thus reveal no information about his valuation.

At the other extreme from the payoff type space is the largest type space embedding the payoff relevant environment described above which places no restrictions on agents' beliefs or higher-order beliefs about other agents' payoff types, allowing for any beliefs and higher-order beliefs about payoff types. This is the universal type space of Harsanyi (1967–68) and Mertens and Zamir (1985), allowing players to hold all possible beliefs and higher-order beliefs about others' payoff types.[2] In much of the present work, we will study a number of classic mechanism problems allowing for all possible beliefs and higher-order beliefs or, equivalently, the universal space.[3] By re-working key results in the literature under this admittedly extreme assumption, we hope to highlight the importance of informational robustness.

However, we believe that the future of work on robust mechanism design will consist of exploring type spaces which are intermediate between payoff type spaces and the universal type space. Such intermediate type spaces embody intermediate common knowledge

[2] The universal space is an infinite type space, so the language in this section must be extended appropriately to incorporate it. In the exposition here, we maintain common certainty that each agent is certain of his own payoff type and that preferences are pinned down by a profile of payoff types. These assumptions are not present in the standard settings where universal type spaces are developed. But the standard construction can be straightforwardly adapted to incorporate these assumptions — see, e.g., the discussion in Section 2.5 of Bergemann and Morris (2005) and Heifetz and Neeman (2006).

[3] As discussed in Section 2.5 of Bergemann and Morris (2005), there is a small gap between the union of all possible type spaces and the universal space that arises from "redundant" types. We will ignore this distinction for purposes of this introductory essay.

assumptions about higher-order beliefs. In the remainder of this section, we discuss examples of intermediate type spaces that are discussed in our work and the literature at large.

In some strands of the implementation literature, it is explicitly or implicitly assumed that there is a true prior p over the payoff types which is common knowledge among the agents, but which the planner does not know. The complete information implementation literature can subsumed in this specification. We can represent this as follows. The type space is $T_i \triangleq \Delta(\Theta) \times \Theta_i$, with a typical element $t_i = (p_i, \theta_i)$. The payoff type is defined in the natural way, $\widehat{\theta}_i(p_i, \theta_i) \triangleq \theta_i$. The belief type is defined on the assumption that there is common knowledge of the true prior among the agents:

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

Choi and Kim (1999) is a representative example of a contribution that explicitly works with this class of type space in an *incomplete information* setting. Choi and Kim (1999) is also discussed in Bergemann and Morris (2005), where we show that in a quasi-linear environment with budget balance and two agents, we can always partially implement allocation rules on the above type spaces, even though it is not possible to partially implement on all type spaces.

A second classic intermediate type space is the common prior universal type space. In the universal type space, there is no requirement that agents' beliefs be derived from some common prior. But it makes sense to discuss the subset of the universal type space where a common prior assumption holds. As described formally above, a type space is a common prior type space if there is a probability measure on the type space such that the players' beliefs over other players' types are conditional beliefs under that common prior. The common prior universal type space embeds all such common prior type spaces. In particular, the results on partial implementation in Bergemann and Morris (2005) do not depend on whether the common prior assumption is imposed or not, but the results on full implementation in Bergemann and Morris (2009a,b, 2011b) do. Bergemann and Morris (2008b) examine the implications for robust full implementation of restricting

attention to common prior type spaces: the results are unchanged if there are strategic complementarities in the direct mechanism (which is true under negative interdependence in preferences, i.e., $\gamma < 0$ in the single good example), but are drastically changed if there are strategic substitutes (which happens with positive interdependence, i.e., $\gamma > 0$ in the single good example).

A third natural class of models to study is when many but not all beliefs are consistent with a given payoff type. In particular, we can assume that there is a benchmark belief corresponding to each payoff type and his true belief must be within a small neighborhood of that benchmark belief. More generally, suppose that if agent i is type θ_i, then his beliefs over the payoff types of others are contained in a set $\Psi_i(\theta_i) \subseteq \Delta(\Theta_{-i})$. A local robustness condition is the requirement that there is common knowledge that all types of all agents have beliefs over others' payoff types within such set. Thus we fix, for each agent i, $\Psi_i : \Theta_i \to 2^{\Delta(\Theta_{-i})}/\varnothing$. Now, in any type space, an agent's beliefs over others' payoff types are implicitly defined and by writing $\psi_i(t_i)$ for those beliefs, we have that:

$$\psi_i(t_i)[\theta_{-i}] = \sum_{\{t_{-i}:\widehat{\theta}_{-i}(t_{-i})=\theta_{-i}\}} \widehat{\pi}_i(t_i)[t_{-i}].$$

Now suppose we restrict attention to type spaces with the property that $\psi_i(t_i) \in \Psi_i(\widehat{\theta}_i(t_i))$ for all agents i and types t_i. If we require each payoff type to have only a single possible belief about others' payoff types (i.e., each $\Psi_i(\theta_i)$ is a singleton), this reduces to the payoff type restriction above. If we put no restrictions on beliefs (i.e., each $\Psi_i(\theta_i) = \Delta(\Theta_{-i})$), then we have the universal type space. A natural "local robustness" approach is to allow Ψ_i to consist of a benchmark belief and a small set of beliefs which are close and versions of this approach have been pursued in a number of settings. Lopomo et al. (2009) and Jehiel et al. (2010) examine local robust implementation of social choice functions. Artemov et al. (2010) examine locally robust (full) virtual implementation of social choice functions. In Bergemann and Morris (2009b), we report on the effect of local robustness considerations in the context of virtual implementation.

These three classes of restrictions are merely representative. Other results in the literature can be understood as reflecting intermediate classes of type spaces in between payoff type spaces and the universal type space. Gizatulina and Hellwig (2010) consider all type spaces with the restriction that agents are *informationally small* in the sense of McLean and Postlewaite (2002); they show that notwithstanding a failure of the BDP property highlighted by Neeman (2004), it is possible to extract almost the full surplus in quasilinear environments. We follow Ledyard (1979) in restricting attention to *full support type spaces* in Bergemann and Morris (2005).

Other results in the literature can be understood as allowing richer type spaces, by allowing payoff perturbations outside the payoff type environment, but then imposing restrictions on beliefs and higher-order beliefs about (perturbed) payoffs types. Type spaces which maintain *approximate common knowledge* of benchmark type spaces are studied by Chung and Ely (2003) and Aghion et al. (2012) as well as in Meyer-Ter-Vehn and Morris (2011). Oury and Tercieux (2012) can be interpreted as a study of type spaces which are *close in the product topology* to some set of benchmark type spaces. Allowing perturbations outside the specified payoff type environment are important in these results.

A final class of restrictions imposed on type spaces are those labelled "generic." As noted above, a classic argument that full surplus extraction is possible on finite type spaces relies on a restriction to "generic" common priors to ensure that the "beliefs determine preferences" property holds (Cremer and McLean, 1988; Neeman, 2004). Here, genericity is applied to finite payoff type spaces (McAfee and Reny (1992) report an extension to infinite payoff type spaces). Since the payoff type space restriction entails such strong common knowledge assumptions and the BDP property seems unnatural it is interesting to ask if the BDP property holds generically for richer type spaces. It is important to note first of all that the property will fail dramatically if we look at the (payoff type) universal type space: by construction, every combination of payoff type and beliefs about others' types are possible, and thus BDP fails. Therefore a small literature has examined whether BDP holds if we

look at the common prior universal type space (the full surplus extraction question is not well posed without the common prior assumption). Unfortunately, there is no agreement or naturally compelling definition of "typical" or "generic" properties in infinite type spaces. Bergemann and Morris (2001) noted that among the (infinite) space of all finite common prior types within the universal type space, one can always perturb a BDP type by a small amount in the product topology and get a non-BDP type and conversely perturb the non-BDP type by a small amount to get back to a BDP type. For topological notions of genericity, answers depend on the topology adopted and the topological definition of genericity employed (see results in Dekel et al. (2006), Barelli (2009), Chen and Xiong (2010), Chen and Xiong (2013) and Gizatulina and Hellwig (2011)).[4] Heifetz and Neeman (2006) report an approach based on alternative geometric and generalized measure theoretic views of genericity for infinite state spaces. We do not consider restrictions based on "genericity" notions in our work. The work on genericity is important but complements rather than substitutes for work which highlights transparently the implicit common knowledge assumptions built into type spaces (such as the BDP property) and judges the relevance of the type spaces for economic analysis based on the plausibility and relevance of those assumptions directly.

We conclude this section by emphasizing that the "payoff type" framework described above is not without loss of generality. In particular, it is assumed that all agents' utility depends only on a vector of payoff types with the property that each element of the vector is known by each agent. Put differently, it is assumed that the join of agents' information fully determines all agents' preferences. This assumption is natural for private value environments and captures important interdependent value environments, but it is restrictive. To see this, consider the single good environment where each agent is valuation of the object is given by

$$v_i = \theta_i + \gamma \sum_{j \neq i} \theta_j. \tag{3.1}$$

[4] But see Chen and Xiong (2013) for a problem in the analysis of Barelli (2009).

We maintain common knowledge that each agent i knows his own payoff type θ_i. What is the content of this assumption? Summing (3.1) across agents gives

$$\sum_{i=1}^{I} v_i = (1 + \gamma(I-1)) \sum_{i=1}^{I} \theta_i,$$

and re-arranging then gives

$$\sum_{j \neq i} \theta_j = \left(\frac{1}{(1 + \gamma(I-1))} \sum_{i=1}^{I} v_i \right) - \theta_i. \tag{3.2}$$

Substituting (3.2) into (3.1) gives

$$\begin{aligned} \theta_i &= v_i - \gamma \sum_{j \neq i} \theta_j \\ &= v_i - \gamma \left(\left(\frac{1}{(1 + \gamma(I-1))} \sum_{i=1}^{I} v_i \right) - \theta_i \right), \end{aligned}$$

which implies

$$\begin{aligned} \theta_i &= \frac{1}{1-\gamma} \left(v_i - \frac{\gamma}{(1 + \gamma(I-1))} \sum_{i=1}^{I} v_i \right) \\ &= \frac{1}{1-\gamma} \left(\left(1 - \frac{\gamma}{(1 + \gamma(I-1))} \right) v_i + \frac{\gamma}{(1 + \gamma(I-1))} \sum_{j \neq i} v_j \right). \end{aligned} \tag{3.3}$$

Thus common knowledge of the payoff type environment implicitly entails the extreme sounding assumption that there there is common knowledge that each agent i knows a particular linear combination of the agents' values, as expressed in (3.3).

We nonetheless maintain the payoff type environment throughout the present work because we are focussed on classical questions about implementing social choice functions (and correspondences) which would be impossible if knowing the join of agents' information is not sufficient to implement the social choice functions. In Bergemann

et al. (2010), we introduce a language for characterizing interdependent types in terms of revealed preference in strategic settings. This richer language can be used to explore settings beyond the payoff type environment. In related work, Bergemann et al. (2012), we illustrate how the language of payoff and belief types can be usefully employed to obtain necessary and sufficient conditions for partial implementation in a large number of settings, such as private value, interdependent value and belief extraction environments.

4

Robust Foundations for Dominant and Ex Post Incentive Compatibility

In Bergemann and Morris (2005), we ask whether a planner can design a mechanism with the property that for any belief and higher-order beliefs that the agents may have, there exists a Bayesian equilibrium of the corresponding incomplete information game where an acceptable outcome is chosen. If we can find such a mechanism, then we say that we have a solution to the robust mechanism design problem. The construction of an ex post incentive compatible mechanism that delivers an acceptable outcome is clearly sufficient, but is it necessary? We call this the ex post equivalence question.

In the special case of private values, ex post incentive compatibility reduces to dominant strategies incentive compatibility. There has been an extended debate, going back to the very beginnings of the development of mechanism design, about whether dominant strategies incentive compatibility should be required or whether Bayesian incentive compatibility is sufficient. Scholars have long pointed out that — as a practical matter — the planner was unlikely to know the "true prior"

over the type space. Therefore, it would be desirable to have a mechanism which was going to work independent of the prior. For a private value environment, Dasgupta et al. (1979), Ledyard (1978), and Ledyard (1979) observed that if a direct mechanism was going to implement a social choice correspondence for every prior on a fixed type space, then there must be dominant strategies implementation. Other scholars pointed out that if the planner did not know the prior (and the agents do) then we should not restrict attention to direct mechanisms. Rather, we should allow the mechanism to elicit reports of the true prior from the agents. After all, since this information is *non-exclusive* in the sense of Postlewaite and Schmeidler (1986), this elicitation will not lead to any incentive problems. A formal application of this folk argument appears in the work of Choi and Kim (1999).

Bergemann and Morris (2005) provide a resolution of this debate by carefully formalizing — using the type space language above — what is and is not being assumed about what is common knowledge about beliefs. This leads to a more nuanced answer to the prior debate about the necessity of dominant strategies incentive compatibility, as well as the extension to an environment with interdependent values. In particular, we show that under some circumstances, even if the planner is able to let the mechanism depend on the agents' beliefs and higher-order beliefs (and thus elicit any knowledge that agents may have about priors on a fixed type space), it is still true that ex post incentive compatibility is necessary for Bayesian implementation for all possible beliefs. This is true if the planner is trying to implement a social choice correspondence which is "separable," a property that is automatically satisfied by social choice *functions*. But for some multi-valued social choice correspondences, it is impossible to identify an ex post incentive compatible selection from a social choice correspondence; but nonetheless, it is possible to find a mechanism with an acceptable equilibrium on any type space. We can illustrate both of these points with the single good allocation example.

Let us first consider the case of a social choice function $f(\theta) = (q(\theta), y(\theta))$ specifying the allocation and transfers in our single good environment. For a given (large) type space $\mathcal{T}$ and a given social choice function f, interim incentive compatibility on a type space $\mathcal{T}$ requires

that:

$$\sum_{t_{-i}\in T_{-i}}\left[\left(\widehat{\theta}_i(t_i)+\gamma\sum_{j\neq i}\widehat{\theta}_j(t_j)\right)q_i(\widehat{\theta}(t))-y_i(\widehat{\theta}(t))\right]\widehat{\pi}_i(t_i)[t_{-i}]$$

$$\geq\sum_{t_{-i}\in T_{-i}}\left[\left(\widehat{\theta}_i(t_i)+\gamma\sum_{j\neq i}\widehat{\theta}_j(t_j)\right)q_i(\widehat{\theta}(t_i',t_{-i}))-y_i(\widehat{\theta}(t_i',t_{-i}))\right]$$
$$\times\widehat{\pi}_i(t_i)[t_{-i}]$$

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

We refer here to "interim" rather than "Bayesian" incentive compatibility to emphasize that the beliefs of agent i, $\widehat{\pi}_i(t_i)[t_{-i}]$, are interim beliefs (without the necessity of a common prior). Now, intuitively, the larger the type space of each agent, the more incentive constraints there are to satisfy, and the harder it becomes to implement a given social choice function. As we consider larger type spaces, that is as we move from the smallest type space, the payoff type space, to the largest type space, the universal type space, the incentive problems become successively more difficult.

It is then natural to ask whether there is a "belief free" solution concept that can guarantee that a reporting strategy profile of the agents remains an equilibrium for all possible beliefs and higher-order beliefs. A social choice function $f(\theta)=(q(\theta),y(\theta))$ is ex post incentive compatible if, for all i, $\theta\in\Theta$, $\theta_i'\in\Theta_i$:

$$\left(\theta_i+\gamma\sum_{j\neq i}\theta_j\right)q_i(\theta)-y_i(\theta)\geq\left(\theta_i+\gamma\sum_{j\neq i}\theta_j\right)q_i(\theta_i',\theta_{-i})-y_i(\theta_i',\theta_{-i}).$$

Under "ex post incentive compatibility" each payoff type of each agent has an incentive to tell the truth *if* he expects all other agents to tell the truth (whatever his beliefs about others' payoff types). Now, given the above definitions, it is apparent that a sufficient condition for robust truthful implementation is that there exists an allocation rule as a function of agents' payoff types that is "ex post incentive compatible," i.e., in a payoff type direct mechanism, each agent has an incentive to announce his type truthfully whatever his beliefs about

others' payoff types. In Bergemann and Morris (2005), we show that a social choice function f is interim incentive compatible on every type space $\mathcal{T}$ if and only if f is ex post incentive compatible.

The above discussion applied to social choice *functions*. Does it extend to social choice correspondences, where multiple outcomes are acceptable for the planner for any given profile of payoff types? Suppose that the planner wanted to implement an allocation rule q but did not care about transfers — i.e., the usual setting in which efficient allocations are studied. Then we would allow for more general payment rules $\widetilde{y} = (\widetilde{y}_1, \ldots, \widetilde{y}_I)$ that could depend on agents' beliefs and higher-order beliefs, with each $\widetilde{y}_i : T \to \mathbb{R}$. Thus we would ask whether for a fixed allocation rule q, we could find for every type space $\mathcal{T}$ payment rules $(\widetilde{y}_1, \ldots, \widetilde{y}_I)$ such that the incentive compatibility condition

$$\sum_{t_{-i} \in T_{-i}} \left[\left(\widehat{\theta}_i(t_i) + \gamma \sum_{j \neq i} \widehat{\theta}_j(t_j) \right) q_i(\widehat{\theta}(t_i, t_{-i})) - \widetilde{y}_i(t_i, t_{-i}) \right] \widehat{\pi}_i(t_i)[t_{-i}]$$

$$\geq \sum_{t_{-i} \in T_{-i}} \left[\left(\widehat{\theta}_i(t_i) + \gamma \sum_{j \neq i} \widehat{\theta}_j(t_j) \right) q_i(\widehat{\theta}(t_i', t_{-i})) - \widetilde{y}_i(t_i, t_{-i}) \right]$$

$$\times \widehat{\pi}_i(t_i)[t_{-i}]$$

holds for all i, $t \in T$ and $t_i' \in T_i$. By allowing the transfers to depend on the beliefs and higher order beliefs we weaken the incentive constraints.

Now, the criticism of the classical justification of dominant strategies discussed above argued that Dasgupta et al. (1979), Ledyard (1978), and Ledyard (1979) were flawed because they did not allow transfers to depend on beliefs. However, in this single good environment, it turns out that allowing transfers to depend on higher-order beliefs does not help. In fact, ex post equivalence continues to hold in this environment and holds more generally in quasi-linear environments where a planner has a unique acceptable outcome (not specifying transfers) but does not care about transfers. Such a correspondence is a leading of example of what we call a "separable" correspondence.

In view of these results, the notion of ex post equilibrium may be viewed as incorporating concern for robustness to beliefs and higher-order beliefs. This "ex post equivalence" result also suggest that the

robustness requirement imposes a striking simplicity on the implementing mechanism. The language of large, and larger, type spaces would suggest that we have to solve successively more difficult incentive problems. After all, as we demand robustness with respect to some or all beliefs and higher-order beliefs, the number of incentive constraints are increasing. But we make the problem more difficult, we eventually have to solve the incentive constraints at every profile θ exactly, without reference to any expectation over payoff profiles. Thus, while the incentive constraints per se are demanding, the set of constraints reduces and hence the solution becomes substantially easier to compute as we only need to verify the incentive constraints at the exact payoff type profiles θ rather than the much larger set of possible types t.

But ex post equivalence does not hold in general in the case of general correspondences. In Bergemann and Morris (2005), we give some abstract examples to make this point. In particular, we describe a private values example with the feature that dominant strategies implementation is impossible but interim implementation is possible on any type space, and this seems to be the first example in the literature noting this possibility. The example points to the fact that interim incentive compatibility can occur for all type spaces, using mechanisms that elicit and respond to the beliefs of the agents, even if ex post incentive compatibility is impossible. Here, let us report an interdependent values example due to Jehiel et al. (2006) which makes the same point in the single good allocation problem.

Suppose now that the payoff type of agent i is given by $\theta_i = (\theta_{i1}, \theta_{i2}) \in [0,1]^2$ and that the value of the object to agent i is then

$$v_i(\theta) = \theta_{i1} + \gamma \sum_{j \neq i} \theta_{j1} + \varepsilon \prod_{j=1}^{I} \theta_{j2}, \tag{4.1}$$

with $\varepsilon > 0$. In the two agent case where $\gamma = 0$, this example was analyzed by Jehiel et al. (2006). In this case, the only ex post incentive compatible social choice functions are trivial ones where the allocation of the object is independent of all agents' types. Under the assumption that the object must always be allocated to one of the two agents, this example thus illustrates the general result of Jehiel et al. (2006) that in generic quasi-linear environments with interdependent values and

multidimensional types, ex post implementation of non-trivial social choice functions is impossible.[1] But it is straightforward to see that an almost efficient allocation of the object can be robustly implemented, since if the object is sold by a second price auction, each agent will have an incentive to bid within ε of θ_{i1}. This observation can be extended to interdependent values if interdependence is not too large, with $0 < \gamma < \frac{1}{I-1}$; in this case, then the generalized second price auction would implement the correspondence of almost efficient allocations. We postpone an explanation of why we need $\gamma < \frac{1}{I-1}$ and how much this argument generalizes until our discussion of Bergemann and Morris (2009a) and Meyer-Ter-Vehn and Morris (2011).

An important class of economic environments where our separability condition fails are quasi-linear environments where transfers are required to be budget balanced. We show in Bergemann and Morris (2005) that ex post equivalence holds nonetheless in some special cases: if there are two agents (Proposition 2) or if each agent has at most two types (Proposition 6). The latter result highlights the importance of allowing rich type spaces: Example 3 shows that we can have partial robust implementation on all payoff type spaces but not on the universal type space. This environment is important because it includes the classic public good problem. In general ex post equivalence fails in this environment (a detailed example was presented in a working paper version of Bergemann and Morris (2005)). Thus there is not a general equivalence between dominant strategy implementation and robust implementation for public good problems.

One implication of our results in Bergemann and Morris (2005) is that we can distinguish settings where a restriction to dominant strategies equilibrium (under private values) or ex post equilibrium

[1] If there are more than two agents, or if the object is not allocated to either of two agents, then agents are assumed indifferent between outcomes which violates the genericity condition in the impossibility result of Jehiel et al. (2006). Bikhchandani (2006) discusses non-trivial ex post incentive compatible allocations that arise if the object need not be allocated to any agent.

See Jehiel and Moldovanu (2001) for an analysis of how multidimensional types already limit the possibility of implementing efficient social choice rules in standard Bayesian settings and Eso and Maskin (2002) and Jehiel et al. (2008) for more on settings with non-trivial ex post implementation with multidimensional type spaces in environments failing the genericity conditions of Jehiel et al. (2006).

in mechanism design problems can or cannot be justified by informational robustness arguments. Thus Dasgupta and Maskin (2000) and Perry and Reny (2002) use ex post equilibrium as a solution concept in studying efficient auctions with interdependent values. This is equivalent to robust partial implementation.

Our analysis in Bergemann and Morris (2005) is limited to asking whether a fixed social choice correspondence — mapping payoff type profiles to sets of possible allocations — can or cannot be robustly partially implemented. Thus we focus on a "yes or no" question. Many of the most interesting questions involve asking what happens when we consider what is the best mechanism for the universal type space when we are interested in a finer objective, and a number of recent papers have addressed this question. Chung and Ely (2007) consider the objective of revenue maximization for the seller of a single object (under the seller's beliefs about agents' valuations), allowing all possible beliefs and higher-order beliefs of the agents, and show conditions under which the seller cannot do better than using a dominant strategy mechanism. The best mechanism from the point of view of the seller would generally allow many outcomes for any given profile of payoff type profiles, and will not in general be separable, and thus the results of Bergemann and Morris (2005) do not apply. Smith (2010) and Börgers and Smith (2012) study the classic problems of public good provision and general social choice with rich private preferences (i.e., the Gibbard–Satterthwaite question) respectively. They identify simple mechanisms that perform better than dominant strategy mechanisms — in the sense of providing weakly better outcomes on all type spaces and strictly better outcomes on some type spaces — for each of these two problems. Yamashita (2011) identifies a mechanism that performs better than any dominant strategy mechanism in the classic bilateral trading problem (the notion of robustness is different from that considered in Bergemann and Morris (2005) but similar results would hold with our notion of robustness). Finally, Bierbrauer and Hellwig (2011) combine the informational robustness approach studied here with a requirement that the social objective be collusion-proof and then obtain restrictions on the social choice function which satisfy both desiderata.

An interesting question for further analysis is the extent to which the results of Bergemann and Morris (2005) continue to hold for more local versions of robustness. Lopomo et al. (2009) identify settings where local robust implementation of a social choice function is equivalent to ex post implementation. Jehiel et al. (2010) give examples illustrating when this equivalence doesn't hold, but nonetheless show that local robust implementation is a very strong and, in particular, generically impossible with multidimensional payoff types.

5

Full Implementation

All of the above results are phrased in terms of incentive compatibility, and by use of the revelation principle, are therefore statements about the existence of a truthtelling equilibrium in the direct mechanism. The construction of the truthtelling equilibrium of course presumes that when we verify the truthtelling constraint of agent i that the other agents are telling the truth as well. This does not address — let alone exclude — the possibility of other equilibria in the direct mechanism; equilibria in which the agents are not telling the truth, and importantly, in which the social choice function is not realized.

As private information may enable the agents to coordinate behavior in many different ways, the designer has to be concerned with the fact that there may exist equilibrium behavior by the agents which does not realize his objective. The notion of *full implementation*, in contrast to *truthful* or *partial implementation,* addresses this by requiring that every equilibrium in the mechanism attains the social objective.[1]

[1] There is a large literature in economic theory — much of it building on the work of Maskin (1999) — devoted to the problem of full implementation: When is it the case that there is a mechanism such that every equilibrium in this mechanism is consistent with a given social choice correspondence? While elegant characterizations of implementability

In Bergemann and Morris (2008a), we restrict attention to the solution concept of ex post equilibrium, and ask what conditions are required for full ex post implementation, i.e., all ex post equilibria to deliver outcomes in the social choice correspondence? In Bergemann and Morris (2009a), we move on to ask when is it possible interim implement a social choice correspondence for all possible higher-order beliefs. In general, the latter is a more stringent requirement. We say that a social choice correspondence that is interim implementable for all possible type spaces is *robustly implementable*.

5.1 Ex Post Implementation

Bergemann and Morris (2005) required — for any beliefs and higher-order beliefs — *an* equilibrium that delivered the right outcome. This required ex post incentive compatibility or — equivalently — that truth-telling is an ex post equilibrium of the "direct" mechanism where agents just report their payoff types. Now, in Bergemann and Morris (2008a), we ask: if we take ex post equilibrium as the primitive solution concept, when can we design a mechanism such that, not only does an ex post equilibrium deliver the right outcome, but also *every* ex post equilibrium delivers the right outcome. Thus there is *full* implementation under the solution concept of ex post equilibrium — and we call this *ex post implementation*. We show that — in addition to ex post incentive compatibility — an ex post monotonicity condition is necessary and almost sufficient. The ex post monotonicity condition neither implies nor is implied by Maskin monotonicity (necessary and almost sufficient for implementation under complete information). By "almost sufficient," we mean sufficient in economic environments and after an additional no veto condition also sufficient in general environments.

In a direct mechanism, such as the generalized second price auction, undesirable behavior by agent i is easiest interpreted as a *misreport* or

---

were developed, the "augmented" mechanisms required to achieve positive results were complex and seemed particularly implausible. While the possibility of multiple equilibria does seem to be a relevant one in practical mechanism design problems, particularly in the form of collusion and shill bidding, the theoretical literature so far has not developed practical insights, with a few recent exceptions such as Ausubel and Milgrom (2005) and Yokoo et al. (2004).

deception θ'. In a direct revelation mechanism, if agents misreport θ' rather than truthfully report θ, then the resulting social outcome is given by $f(\theta')$ rather than $f(\theta)$. The notion of ex post monotonicity guarantees that (i) a whistle-blower (among the agents) will alert the principal of deceptive reporting θ' by receiving a reward and (ii) a whistle-blower will not falsely report a deception.

The social choice function $f = (q, y)$ satisfies ex post monotonicity if for every θ, θ' with $f(\theta) \neq f(\theta')$, there exist i, $\widehat{q}_i \in [0,1]$ and $\widehat{y}_i \in \mathbb{R}$ such that

$$\left(\theta_i + \gamma \sum_{j \neq i} \theta_j\right) \widehat{q}_i - \widehat{y}_i > \left(\theta_i + \gamma \sum_{j \neq i} \theta_j\right) q_i(\theta'_i, \theta'_{-i}) - y_i(\theta'_i, \theta'_{-i}),$$

while

$$\left(\theta''_i + \gamma \sum_{j \neq i} \theta'_j\right) q_i\left(\theta''_i, \theta'_{-i}\right) - y_i(\theta''_i, \theta'_{-i}) \geq \left(\theta''_i + \gamma \sum_{j \neq i} \theta'_j\right) \widehat{q}_i - \widehat{y}_i$$

for all $\theta''_i \in \Theta_i$.

Proposition 3 in Bergemann and Morris (2008a) then establishes that the social choice function implied by the generalized second price auction satisfies the ex post monotonicity condition. Moreover, due to the quasi-linearity of the utility function, it also represents an economic environment, and hence can be fully implemented in an ex post equilibrium, provided there are three or more bidders. In fact, with interdependent values, or $\gamma \neq 0$, the implementation can be achieved in the direct mechanism itself and does not need to make use of an augmented mechanism. In other words, the direct mechanism is shown to have a unique ex post equilibrium if $\gamma \neq 0$. This three or more player result contrasts with the observation of Birulin (2003) that, with only two players, there are a continua of undominated ex post equilibria in the direct mechanism of the single good allocation problem.

5.2 Robust Implementation in the Direct Mechanism

But can the planner design a mechanism with the property that for any beliefs and higher-order beliefs that the agents may have, *every* equilibrium has the property that an acceptable outcome is chosen? We call

this "robust implementation" and investigate the possibility of robust implementation in Bergemann and Morris (2009a) and Bergemann and Morris (2011a). We should immediately emphasize that the question of robust implementation is *not* the same as the ex post implementation question analyzed in Bergemann and Morris (2008a): to rule out bad equilibria there, it was enough to make sure you could not construct a "bad" ex post equilibrium; for robust implementation, we must rule out bad Bayesian, or interim, equilibria on all type spaces. In Bergemann and Morris (2009a), we consider a well-behaved environment with payoff type spaces represented by intervals of the real line and "aggregator single crossing" preferences. In this environment, we give a "contraction property" — equivalent to not too much interdependence in types — and show that if *strict* ex post incentive compatibility and the contraction property hold, then robust implementation is possible in the direct mechanism. If either fails, robust implementation is impossible in *any* mechanism.

To describe the results in more detail we return to the (generalized) second price auction. We start with the private value environment, where it is well-known that the second price auction has many equilibria in which the agents do not tell the truth, and in consequence the allocation is not guaranteed to be efficient. The reason is that truthtelling is only a weak best response and hence just a dominant strategy, but not a strictly dominant strategy. The good news is that we can easily modify the original auction so that truthful bidding becomes a strictly dominant strategy. Fix $\varepsilon > 0$. Now, with probability $1 - \varepsilon$, let us allocate the object to the highest bidder and have him pay the second highest bid. With the complementary probability ε, let us randomly and uniformly pick an agent, and allocate the object to that agent with probability b_i, a probability that is proportional to his bid. Thus the ε-allocation rule (parameterized by ε) is defined by

$$q_i^{**}(\theta) \triangleq (1 - \varepsilon)\, q_i^{*}(\theta) + \varepsilon q_i(\theta), \tag{5.1}$$

with

$$q_i(\theta) \triangleq \frac{\theta_i}{I}.$$

This *modified* generalized second price auction is supported by an associated set of (expected) transfers conditional on the reported type profile θ:

$$y_i^{**}(\theta) = \frac{\varepsilon}{2I}\theta_i^2 + (1-\varepsilon)\left(\max_{k\neq i}\{\theta_k\}\right) q_i^*(\theta). \tag{5.2}$$

The transfer rule $y_i^{**}(\theta)$ supports truthtelling as an equilibrium in strictly dominant strategies, that is $b_i = \theta_i$ forms a *strictly* dominant strategy in this mechanism. The strictness is established by making the allocation responsive to the bid of agent i even if agent i is not the highest bidder. It follows that whatever agent is beliefs or higher-order beliefs about θ_{-i} are, he will have a *strictly* dominant strategy to set $b_i = \theta_i$. In our language, for any $\varepsilon > 0$, we can guarantee the *robust implementation* of the almost efficient, or ε-efficient allocation rule q^{**}.[2]

Now consider the case of interdependent values $\gamma \neq 0$. We can modify the generalized second price sealed bid auction to turn the ex post equilibrium into a strict ex post equilibrium, just as we modified the second price sealed bid auction. We construct the following allocation rule $q_i^{**}(\theta)$. With probability $1-\varepsilon$, we have the winning bidder i pay:

$$\max_{j\neq i}\{\theta_j\} + \gamma\sum_{j\neq i}\theta_j,$$

and with probability ε, we randomly and uniformly pick an agent, and allocate the object to that agent with probability b_i, a probability that is proportional to his bid. In the event that agent i is assigned the object, he then pays:

$$\frac{1}{2}\theta_i + \gamma\sum_{j\neq i}\theta_j.$$

[2] In related modifications of the second price auction in a private value environment, Plum (1992) considers a convex combination of a first-price and a second-price auction (with a small weight on the former) and Blume and Heidhues (2004) introduce a small reserve price in the second price auction. Either of these modifications render the equilibrium outcome unique, but in contrast to the present formulation, these modifications do not strengthen truth-telling from a weakly dominant to a strictly dominant strategy.

Now, in this modification of the generalized second price auction, the associated transfers can be written as, in a generalization of (5.2):

$$y_i^{**}(\theta) = \frac{\varepsilon}{2I}\theta_i^2 + \frac{\varepsilon\theta_i}{I}\left(\gamma\sum_{j\neq i}\theta_j\right) + (1-\varepsilon)\left(\max_{k\neq i}\left\{\theta_k + \gamma\sum_{j\neq i}\theta_j\right\}\right)q_i^*(\theta). \tag{5.3}$$

The social choice function in this modified generalized second price auction is given by a pair of allocation and transfer functions: $f^{**}(\theta) = (q^{**}(\theta), y^{**}(\theta))$. The net utility of agent i, given a true payoff profile θ and reported payoff profile θ', is explicitly given by

$$\left(\theta_i + \gamma\sum_{j\neq i}\theta_j\right)\left(\frac{\varepsilon}{I}\theta_i' + (1-\varepsilon)q_i^*(\theta')\right) - \frac{\varepsilon}{2I}\theta_i'^2 - \frac{\varepsilon\gamma\theta_i'}{I}\sum_{j\neq i}\theta_j' - (1-\varepsilon)\left(\max_{k\neq i}\left\{\theta_k' + \gamma\sum_{j\neq i}\theta_j'\right\}\right)q_i^*(\theta').$$

The net utility function is a linear combination of the efficient allocation rule and the proportional allocation rule. It is straightforward to compute the best response of each agent i, given a point belief about the payoff type profile θ and reported profile θ'_{-i} of the remaining agents. The best response is linear in the true valuation and in the size of the misrepresentation $(\theta_j - \theta_j')$, downwards or upwards, of the other agents:

$$\theta_i' = \theta_i + \gamma\sum_{j\neq i}(\theta_j - \theta_j'). \tag{5.4}$$

From here, it follows that the reports of agent i and agent j are strategic substitutes if $\gamma > 0$ and strategic complements if $\gamma < 0$. For example, with $\gamma > 0$, if agent j increases his report, then in response agent i optimally chooses to lower his report.

From (5.4), we can conclude that truthtelling indeed forms a strict ex post equilibrium. But even though we have a strict ex post incentive

compatible mechanism, we cannot guarantee the robust implementation of q^{**}. In fact, we shall now show that the direct mechanism robustly implements the efficient outcome if and only if the interdependence is moderate,[3] or

$$|\gamma| < \frac{1}{I-1}.$$

Moreover, no mechanism, whether it is the direct mechanism or an augmented mechanism is able to robustly implement the efficient outcome if the interdependence is too large, or

$$|\gamma| \geq \frac{1}{I-1}.$$

This necessary and sufficient condition for robust implementation should be compared with the necessary and sufficient condition for robust partial implementation, which we earlier showed to require the single crossing condition, namely

$$\gamma \leq 1.$$

As we analyzed truthtelling in the direct mechanism for all possible beliefs and higher-order beliefs, all we had to do was to guarantee the incentives to reveal the private information, agent by agent, while presuming truthtelling by other agents. Now, as we seek robust implementation, we cannot suppose the truthtelling behavior of the other agents but rather have to guarantee it. We shall obtain this guarantee by identifying restrictions on the rational behavior of each agent, and then use these restrictions to inductively obtain further restrictions. More formally, we shall analyze the outcome of the mechanism under rationalizability with incomplete information. An action, which in the direct mechanism, simply constitutes a reported payoff type, is called incomplete information rationalizable if it survives the process of iteratively elimination of dominated strategies. As rationalizability with complete information, rationalizability under incomplete information

[3] The importance of this moderate interdependence condition arose earlier in the work of Chung and Ely (2001) who showed that it was sufficient for implementing the efficient outcome in the unperturbed generalized second price auction under iterated deletion of weakly dominated strategies.

defines an inductive process: first suppose that every payoff type θ_i could send any message m_i; then, second, delete those messages m_i that are not a best response to some conjecture over pairs of payoff type and messages (θ_{-i}, m_{-i}) of the opponents that have not yet been deleted. The inductive procedure is then to repeat the second step until convergence is achieved.

We observe that the notion of incomplete information rationalizability is belief free as the candidate action needs only to be a best response to some beliefs about the other agents actions and payoff types. We can focus on the notion of incomplete information rationalizability because of the following epistemic result: a message m_i can be sent by an agent with payoff type θ_i in an interim equilibrium on some type space if and only if m_i is "incomplete information rationalizable" for payoff type θ_i. The equivalence between robust and rationalizable implementation is an incomplete generalization of Brandenburger and Dekel (1987) and can be seen as a special case of the incomplete information results of Battigalli and Siniscalchi (2003). It illustrates a general point well-known from the literature on epistemic foundations of game theory: that equilibrium solution concepts only have bite if we make strong assumptions about type spaces, i.e., we assume small type spaces where the common prior assumption holds.

We now describe the inductive argument for rationalizability in the direct mechanism for the single-unit auction. For concreteness, we shall assume here positive interdependence, $\gamma > 0$, but all the relevant arguments go through with negative interdependence, after suitably reversing the signs. In the direct mechanism a message m_i is simply a reported payoff type θ_i'. Each agent i has some conjecture about the other agents true type profile θ_{-i} and their reported type profile θ'_{-i}. We denote such a conjecture by λ_i:

$$\lambda_i(\theta_{-i}, \theta'_{-i}) \in \Delta(\Theta_{-i} \times \Theta_{-i}).$$

We can then ask what is the set of reports that agent i might send for some conjecture $\lambda_i(\theta_{-i}, \theta'_{-i})$ over his opponents' types θ_{-i} and reports θ'_{-i} in the k-th step of the inductive procedure. We denote this set by $\beta_i^k(\theta_i)$. We restrict the conjectures $\lambda_i(\theta_{-i}, \theta'_{-i})$ of agent i in step k to be of the form that type θ_j can only be conjectured to send message θ'_j if it was rationalizable at step $k-1$, i.e., if $\theta'_j \in \beta_i^{k-1}(\theta_j)$.

We initialize the inductive process at step $k = 0$ by allowing all possible reports $\beta_i^0(\theta_i) = [0, 1]$. In the context of the almost efficient allocation rule $f^{**}(\theta) = (q^{**}(\theta), y^{**}(\theta))$ and the associated ex post compatible transfer $y_i^{**}(\theta)$, the expected payoff of agent i is quadratic in his report θ_i'. It follows that the best response of agent i to a probability one conjecture about his opponents true type and reported type profiles to be $(\theta_{-i}, \theta'_{-i})$, is given by the linear best response θ_i' with

$$\theta_i' = \theta_i + \gamma \sum_{j \neq i} (\theta_j - \theta_j'). \tag{5.5}$$

Thus if he expects the other agents to underreport their type, i.e., $\theta_j - \theta_j' > 0$, then the best response of agent i is to correct this by overreporting his type. We notice that the best response has a self-correcting property. With the correction induced by the reported type θ_i' in (5.5), the reported valuation of agent i actually equals his true valuation:

$$\theta_i' + \gamma \sum_{j \neq i} \theta_i' = \theta_i + \gamma \sum_{j \neq i} (\theta_j - \theta_j') + \gamma \sum_{j \neq i} \theta_i' = \theta_i + \gamma \sum_{j \neq i} \theta_j.$$

The best response (5.5) of agent i only corrects the valuation of agent i, the other reported valuation continue to differ from the true valuations under the best response of agent i. The linear best response property then leads to a *set* of best responses $\beta_i^k(\theta_i)$ in step k, which can be characterized in terms of a lower and upper bound:

$$\beta_i^k(\theta_i) = [\underline{\beta}_i^k(\theta_i), \overline{\beta}_i^k(\theta_i)].$$

With the inductive procedure, the bounds $\{\underline{\beta}_i^k(\theta_i), \overline{\beta}_i^k(\theta_i)\}$ in step k are determined by the restrictions identified in round $k - 1$:

$$\{(\theta_{-i}, \theta'_{-i}) : \theta_j' \in \beta_j^{k-1}(\theta_j),\ \forall j \neq i\}.$$

The upper bound $\overline{\beta}_i^k(\theta_i)$ identifies the largest rationalizable report by agent i with payoff type θ_i. It is obtained by identifying a feasible point conjecture at which the sum of underreports of the other agents, $\sum_{j \neq i}(\theta_j - \theta_j')$, is maximized:

$$\overline{\beta}_i^k(\theta_i) = \theta_i + \ \gamma \max_{\{(\theta'_{-i}, \theta_{-i}) : \theta_j' \in \beta_j^{k-1}(\theta_j),\ \forall j \neq i\}} \ \sum_{j \neq i} (\theta_j - \theta_j')\}.$$

The largest rationalizable report for agent i, given his payoff type θ_i, arises under the conjecture that the remaining agents maximally underreport relative to their true payoff type. But the lowest reported type of payoff type θ_j is given by the lower bound obtained in the preceding step $k-1$, and thus using the lower bound $\underline{\beta}_j^{k-1}(\theta_j)$ from step $k-1$ explicitly, we get:

$$\overline{\beta}_i^k(\theta_i) = \theta_i + \ \gamma \max_{\theta_{-i} \in \Theta_{-i}} \sum_{j \neq i} (\theta_j - \underline{\beta}_j^{k-1}(\theta_j)))\}.$$

Similarly, the lowest possible report of payoff type θ_i, the "maximal" underreport, emerges from the point conjecture that the remaining agents are "maximally" overreporting relative to their true type, thus:

$$\underline{\beta}_i^k(\theta_i) = \theta_i + \ \gamma \max_{\theta_{-i} \in \Theta_{-i}} \sum_{j \neq i} (\theta_j - \overline{\beta}_j^{k-1}(\theta_j))\}.$$

Given the compactness of the payoff type set, in fact $\Theta_i = [0,1]$, we obtain explicit expressions for the lower and upper bounds. In step $k=1$, the conjectures about the other players are unrestricted, and so for every j:

$$\max_{\theta_j \in \Theta_j} \ (\theta_j - \underline{\beta}_j^0(\theta_j))) = 1 - 0 = 1,$$

and hence

$$\overline{\beta}_i^1(\theta_i) = \theta_i + \gamma(I-1),$$

and more generally we find that in step k the upper bound is given by

$$\overline{\beta}_i^k(\theta_i) = \theta_i + (\gamma(I-1))^k, \tag{5.6}$$

and likewise the recursion for the lower bound yields:

$$\underline{\beta}_i^k(\theta_i) = \theta_i - (\gamma(I-1))^k. \tag{5.7}$$

We thus find that a reported payoff type θ_i', different from the true type θ_i, can be eliminated for sufficiently large k from the best response set, or

$$\theta_i' \neq \theta_i \Rightarrow \theta_i' \notin \beta_i^k(\theta_i),$$

provided that:

$$|\gamma|(I-1) < 1 \Leftrightarrow |\gamma| < \frac{1}{I-1}. \tag{5.8}$$

We then have a sufficient condition for robust implementation, which requires that the interdependence among the agents is only moderate in this sense of the above inequality. The next question then is whether the above sufficient condition is also a necessary condition for robust implementation. Indeed, suppose that the parameter of interdependence, γ, were larger than the inequality (5.8) requires, or:

$$\gamma \geq \frac{1}{I-1}.$$

We can use the richness of the possible type space $\mathcal{T}$ to identify specific types, in particular specific belief types, under which the interim expected valuations of any two payoff types θ_i and θ_i', with $\theta_i \neq \theta_i$, are indistinguishable. Thus suppose that each payoff type θ_i is convinced, i.e., has the point conjecture that the payoff type θ_j of agent j is given by:

$$\theta_j \triangleq \frac{1}{2} + \frac{1}{\gamma(I-1)}\left(\frac{1}{2} - \theta_i\right), \quad \forall j.$$

If we now compute the interim expected value of the object for i under the above belief, we find that the interim expected value of the object for agent i is in fact independent of θ_i:

$$\theta_i + \gamma(I-1)\left(\frac{1}{2} + \frac{1}{\gamma(I-1)}\left(\frac{1}{2} - \theta_i\right)\right) = \frac{1}{2}(1 + \gamma(I-1)).$$

It then follows immediately that the payoff types cannot be distinguished in the direct mechanism, as each payoff type θ_i assigns the same expected value to the object given his private information. We say that the payoff types are *indistinguishable*, and in fact they are indistinguishable in any, direct or indirect, mechanism. We have thus established that in the single unit auction, robust implementation is possible (using the modified generalized VCG mechanism) if

$$|\gamma| < \frac{1}{I-1}, \tag{5.9}$$

and conversely that robust implementation is impossible (in *any* mechanism) if

$$|\gamma| \geq \frac{1}{I-1}.$$

This result has to be contrasted with robust incentive compatibility condition, namely the ex post incentive compatibility, which required (only) that $\gamma < 1$.

In Bergemann and Morris (2009a), we generalize the property of moderate interdependence (5.9) and refer to it more generally as a "contraction property," as it is suggested by the contraction like property of the lower and upper bounds, (5.6) and (5.7), respectively. We assume that preferences are single crossing with respect to a one dimensional aggregator of agents' types. A "deception" specifies for each payoff type of each agent, a set of payoff types that might be misreported. Our contraction property requires that for any deception, there is at least one misreport of one type of one "whistleblowing" agent for whom the misreports of others will not reverse the sign of the impact of the whistleblower's misreport on his preferences. The robust implementation result that we established above in the context of the single unit auction can now be stated for the general environment as follows. Robust implementation is possible in the *direct* (or any augmented) mechanism if and only if strict ex post incentive compatibility and the contraction property hold.

A noteworthy aspect of the above result is that the strict separation between possibility and impossibility not only holds for the direct mechanism but for any other, possibly augmented mechanism. To wit, the literature on implementation frequently uses "augmented mechanism" to obtain sufficient conditions for implementation. Here, the robustness requirement implies that augmented mechanisms, relative to the simple mechanism in the form of the direct mechanism, lose their force. Hence, the more stringent requirements of robust implementation reduce the role of complex and overly sensitive mechanisms.

The above analysis also demonstrates that while robust implementation is a strong requirement, it is weaker than dominant strategy implementation. After all, in the environment with interdependent

values, a dominant strategy equilibrium does not even exist, nonetheless truthtelling in the direct mechanism is an ex post equilibrium, and as we showed is indeed the unique incomplete information rationalizable outcome.

As we saw in the example of the single unit auction, the "contraction property" had a natural interpretation in a linear valuation environment. This interpretation remains, even in a *nonlinear utility environment*, provided that the aggregator remains linear. For example, a linear aggregator for each agent i might be of the form:

$$h_i(\theta) = \theta_i + \sum_{j \neq i} \gamma_{ij} \theta_j,$$

where each weight γ_{ij} measures the importance of payoff type j for preference of agent i. In the case of the linear aggregator, we can form an interaction matrix based on the weights γ_{ij} across all agent pairs i and j:

$$\Gamma \triangleq \begin{bmatrix} 0 & |\gamma_{12}| & \cdots & |\gamma_{1I}| \\ |\gamma_{21}| & 0 & & \vdots \\ \vdots & & \ddots & |\gamma_{I-1I}| \\ |\gamma_{I1}| & \cdots & |\gamma_{II-1}| & 0 \end{bmatrix}.$$

We can then give a generalized version of the moderate interdependence condition in terms of the interaction matrix Γ. In Bergemann and Morris (2009a), we show that the interaction matrix has the contraction property if and only if largest eigenvalue of the interaction matrix is less than 1.

5.3 The Robustness of Robust Implementation

Meyer-Ter-Vehn and Morris (2011) show that if there is a approximate common knowledge that we are in an environment close to a strict version of that of Bergemann and Morris (2009a) (i.e., with one dimensional interdependent values under an aggregator function and a uniformly strict contraction property, and uniformly strict ex post incentive compatibility), then the social choice correspondence consisting of almost efficient allocations can be robustly implemented.

This result can be illustrated by the two dimensional perturbation of the single good allocation problem we discussed in Section 4. Thus suppose again that the payoff type of agent i is given by $\theta_i = (\theta_{i1}, \theta_{i2}) \in [0,1]^2$ and that the value of the object to agent i is, as earlier in (4.1):

$$v_i(\theta) = \theta_i^1 + \gamma \sum_{j \neq i} \theta_{j1} + \varepsilon \prod_{j=1}^{I} \theta_{j2},$$

with $\varepsilon > 0$ and $\gamma < \frac{1}{I-1}$. It is an implication of the lower hemicontinuity of rationalizable outcomes that in the modified generalized second price auction of Bergemann and Morris (2009a), types $(\theta_{i1}, \theta_{i2})$ will have an incentive to bid something in the neighborhood of θ_{i1}. The social choice correspondence of almost efficient allocations of the private good is therefore almost robustly (fully) implemented.

While Meyer-Ter-Vehn and Morris (2011) delivers a robust full implementation result — by generalizing arguments in Bergemann and Morris (2009a) — the purpose of Meyer-Ter-Vehn and Morris (2011) is only to deliver a partial implementation result. This raises the question of whether it is possible to get partial robust implementation of the almost efficient allocations (without full robust implementation) without the moderate interdependence condition of $\gamma < \frac{1}{I-1}$. While the argument presented here rely directly on the arguments of Bergemann and Morris (2009a), there is an important connection between partial and full implementation identified by Oury and Tercieux (2012), which might indicate that there is a strong link between partial and full implementation. They show that requiring continuous, but partial, implementation in complete or incomplete information settings implies the necessity of full implementation.

5.4 Robust Implementation in General Mechanisms

Section 3 restricted attention to a class of well-behaved environments. In contrast, in Bergemann and Morris (2011a), we characterize robust implementation in general environments with general mechanisms. By robust implementation we mean that *every* equilibrium on *every* type space $\mathcal{T}$ generates outcomes consistent with the social choice function f. As we seek to identify necessary and sufficient conditions for robust

implementation, conceptually there are (at least) two approaches to obtain the conditions. One approach would be to simply look at the interim implementation conditions for every possible type space $\mathcal{T}$ and then try to characterize the intersection or union of these conditions for all type spaces. But in Bergemann and Morris (2011a), we focus our analysis on a second, more elegant, approach. We first establish an equivalence between robust and *rationalizable implementation* and then derive the conditions for robust implementation as an implication of rationalizable implementation. The advantage of the second approach is that after establishing the equivalence, we do not need to argue in terms of large type spaces, but rather derive the results from a novel argument using the iterative deletion procedure associated with rationalizability. This equivalence was already used in Bergemann and Morris (2009a), but for the arguments in Bergemann and Morris (2011a) we allow for general (perhaps infinite and non-compact mechanisms), and thus new versions of the equivalence results must be developed.

As suggested by the analysis in the direct mechanism, ex post incentive compatibility and a robust monotonicity condition are necessary and almost sufficient for robust implementation. And, in the aggregator single crossing environment of Bergemann and Morris (2009a), robust monotonicity is equivalent to the contraction property.

5.5 Rationalizable Implementation

In Bergemann and Morris (2009a, 2011a), we establish necessary and sufficient conditions for "robust implementation" in *environments with incomplete information*. In particular, we showed that a social choice function f can be interim (or Bayesian) equilibrium implemented for all possible beliefs and higher-order beliefs if and only if f is implementable under an incomplete information version of rationalizability. These results prompted us to refine and further develop the rationalizability arguments in *environments with complete information*. In Bergemann et al. (2011), we establish stronger necessary and sufficient conditions than in the incomplete information environment and show that these conditions are almost equivalent to the Nash equilibrium implementation conditions when the social choice function is responsive (a social

choice function is responsive if distinct states imply distinct social choices). With respect to the necessary conditions, we strengthen the monotonicity condition, due to Maskin (1999), from a weak inequality to a strict inequality.

Writing the strict Maskin monotonicity condition in the context of the single good example, we say that a social choice function f satisfies strict Maskin *monotonicity* if $f(\theta) \neq f(\theta')$ implies that for some $i, \widehat{q}_i$ and $\widehat{y}_i$,

$$\left(\theta'_i + \gamma \sum_{j \neq i} \theta'_j\right) \widehat{q}_i - \widehat{y}_i > \left(\theta'_i + \gamma \sum_{j \neq i} \theta'_j\right) q_i(\theta) - y_i(\theta),$$

and

$$\left(\theta_i + \gamma \sum_{j \neq i} \theta_j\right) q_i(\theta) - y_i(\theta) > \left(\theta_i + \gamma \sum_{j \neq i} \theta_j\right) \widehat{q}_i - \widehat{y}_i.$$

The latter condition requires that if the socially desired alternatives differ in state θ and θ', then there must exist an agent i and a reward allocation $(\widehat{q}_i, \widehat{y}_i)$ such that if the true state were θ' and agent i were to expect the other agents to claim that the state is θ, i could be offered a reward $(\widehat{q}_i, \widehat{y}_i)$ that would give him a strict incentive to "report" the deviation of the other agents, but that the reward y would not tempt him if the true state were in fact θ. The strengthening of the monotonicity condition, commonly referred to as Maskin monotonicity, that we require is that the reward y gives agent i a strict incentive to "report truthfully" if the true state were θ. In the single good example, the efficient allocation rule $f^*(\theta) = (q^*(\theta), y^*(\theta))$ fails Maskin monotonicity and thus strict Maskin monotonicity, because the allocation is on the boundary, but nearly efficient rules such as $f^{**}(\theta) = (q^{**}(\theta), y^{**}(\theta))$, defined earlier by (5.1) and (5.2), are both Maskin monotonic and strict Maskin monotonic.

Given that we are stating the result in terms of a social choice function, rather than a social choice correspondence, the notion of full implementation is akin to requiring that the game (generated by the mechanism) has a unique equilibrium (outcome). The implementation results in Bergemann et al. (2011) then suggest that sufficient

conditions to get a unique rationalizable outcome are similar to those required for a unique Nash equilibrium outcome, provided that the social choice function is responsive. This is noteworthy as the necessary and almost sufficient condition of Maskin monotonicity is much weaker than the well-known conditions under which there are close relationships between the uniqueness of Nash equilibrium and the uniqueness of the rationalizable outcomes, such as supermodular or concave games. The present results indicate the strength of the implementation approach to reduce the number of equilibria. By using infinite message spaces and stochastic allocations, we strengthen the positive implementation results under Nash equilibrium to the weaker solution concept of rationalizability.

The techniques by which we identify necessary and almost sufficient "monotonicity" conditions for robust implementation under incomplete information in Bergemann and Morris (2011a) and rationalizable implementation under complete information in Bergemann et al. (2011) can be extended to identify necessary and almost sufficient monotonicity conditions for implementation in rationalizable strategies in standard incomplete information environments. These conditions are related to but stronger than the Bayesian monotonicity conditions identified by Postlewaite and Schmeidler (1986) and Jackson (1991) for equilibrium implementation under incomplete information. These conditions are developed and used in Oury and Tercieux (2012).

5.6 The Role of the Common Prior

In the presentation of the results thus far, we did not place any restrictions on the agents' beliefs and higher-order beliefs. Bergemann and Morris (2008b), we investigate the impact of restricting attention to common prior type spaces.

We recall that in the single unit auction model, the best response of agent i was of the linear form:

$$\theta_i' = \theta_i + \gamma \sum_{j \neq i} (\theta_j - \theta_j'). \tag{5.10}$$

If $\gamma < 0$, the negative informational interdependence gives rise to strategic complementarity in the reporting of the payoff types in the

direct mechanism. Conversely, positive informational interdependence in agents' types, or $\gamma > 0$ gives rise to strategic substitutability in direct mechanism.

The relationship between the informational interdependence and the nature of the strategic interaction then allows us to offer sharp predictions on the role of the common prior. With strategic complements, we know that in games of complete information, there is no gap between Nash equilibrium and rationalizable actions in the sense that there are multiple equilibria if and only if there are multiple rationalizable actions. This is a well-known result which appeared prominently in Milgrom and Roberts (1990). Now, with the linear best response property (5.10), this result remains true with the appropriate solution concepts for games with incomplete information. In particular, given the restriction to a common prior type space, the behavior under incomplete information rationalizability is equivalent to behavior in the incomplete information correlated equilibrium. In other words, there is a unique Bayes Nash equilibrium if and only if there is unique incomplete information rationalizable outcome. Thus, provided that we are considering mechanism with strategic complementarities, whether or not we restrict attention to common prior type spaces makes no difference, and in particular the contraction property continues to play the same role as a necessary and sufficient condition for robust implementation, as described earlier.

On the other hand, if we consider environments that give rise to strategic substitutability in the direct mechanism, then the presence of a common prior facilitates the robust implementation. Here, it is possible to robustly implement the allocation problem, even if the contraction property fails. In particular, in the single unit auction model we can allow the parameter of interdependence γ to satisfy:

$$\frac{1}{I-1} < \gamma < 1,$$

and still guarantee robust implementation in the direct mechanism if we restrict attention to type spaces satisfying the common prior assumption. This leads to the following result in Bergemann and

Morris (2008b). If the reports are strategic complements, then robust implementation with a common prior implies robust implementation without a common prior. If the reports are strategic substitutes, then robust implementation with a common prior fails to imply robust implementation without a common prior.

5.7 Dynamic Mechanisms

All the results discussed so far have dealt with static mechanisms. In Bergemann and Morris (2007), we analyze the modified generalized second price auction in a dynamic mechanism. We consider the ascending (or English) auction in a *complete information* environment. We ask whether the sequential mechanism offers advantages relative to the static, direct revelation, mechanism in terms of achieving robust implementation. The advantage of the sequential mechanism is the ability to reveal and communicate private information in the course of the mechanism. The revelation of private information can decrease the uncertainty faced by the bidders and ultimately improve the final allocation offered by the mechanism. In auctions, the source of the uncertainty can either be payoff uncertainty (uncertainty about payoff relevant information) or strategic uncertainty (uncertainty about the bids of the other agents). We show that the efficient outcome is fully implemented even when

$$\frac{1}{I-1} < \gamma < 1.$$

Recall that in this setting, we know that full robust implementation (with incomplete information) is not possible under any mechanism and that full implementation does not occur in this direct mechanism even with complete information. Thus we show that in at least some settings, sequential refinements help achieve full implementation.

This result is in the spirit of the classical results of Moore and Repullo (1988) showing the possibility of full implementation of social choice functions even when Maskin monotonicity fails, if subgame perfection is used as a solution concept within the dynamic mechanism.

Aghion et al. (2012) show that full implementation is no longer possible, even under subgame perfection, if the mechanism is required to work also for types close to complete information.

More closely related to our work in Bergemann and Morris (2007), Mueller (2009), and Penta (2011) examine the robustness of dynamic mechanisms in environments with incomplete information. The results are sensitive to the sequential refinement used in this context, with Mueller (2009) obtaining very permissive results with a stronger refinement and Penta (2011) getting less permissive results with a weaker refinement. Our approach uses a version of Penta's weaker refinement, but results in Penta (2011) suggest that our positive results in Bergemann and Morris (2007) do rely heavily on the complete information assumption.

5.8 Virtual Implementation

In complete as well as in incomplete information settings, the relaxation from "exact" implementation to "virtual" implementation leads to a significant weakening of the necessary conditions for implementation. Virtual implementation, as initially defined by Matsushima (1988) and Abreu and Sen (1991), requires that the social choice function arises with probability arbitrarily close to 1, but not necessarily equal to 1. In Bergemann and Morris (2009b), we characterize robust virtual implementation in general environments with well-behaved, finite or compact, mechanisms. We show that ex post incentive compatibility and a robust measurability condition are necessary and almost sufficient for robust virtual implementation. Robust measurability can also be naturally interpreted as a restriction on the amount of interdependence of agents' types. But it neither implies nor is implied by robust monotonicity. However, in the aggregator environment of Bergemann and Morris (2009a), robust measurability and robust monotonicity are both equivalent to the contraction property and the only impact of relaxing "exact" to "virtual" robust implementation is the relaxation from *strict* ex post incentive compatibility in Bergemann and Morris (2009a) to *weak* ex post incentive compatibility.

With respect to our leading example, the single unit auction, the transition from the generalized second price auction to the modified second price auction, can now be interpreted as the virtual implementation of the generalized second price auction. After all, in the modified generalized second price auction, the allocation of the generalized second price auction is only chosen with probability $1 - \varepsilon$, for some $\varepsilon > 0$. The key result in Bergemann and Morris (2009b) is a characterization of when two payoff types are *strategically distinguishable* in the sense that they can be guaranteed to behave differently in some mechanism. The condition of robust measurability now requires that *strategically indistinguishable* types are treated the same by the social choice function.

We now provide an exact characterization of strategic distinguishability in the context of the single-unit auction. If we have sets of payoff types, $\Psi_1 \subset \Theta_1$ and $\Psi_2 \subset \Theta_2$, of agents 1 and 2, respectively, we say that the set Ψ_2 separates the set Ψ_1 if knowing agent 1s preferences and knowing that agent 1 is sure that agent 2s type is in Ψ_2, we can rule out at least some payoff type of agent 1 in Ψ_1. Now consider an iterative process where we start, for each agent, with all subsets of his payoff type set, namely the power set of Θ_1, 2^{Θ_1}, and — at each stage — delete subsets of payoff types that are separated by every remaining subset of types of his opponents. A pair of types are said to be *pairwise inseparable* if the set consisting of that pair of types survives this process. We show that two types are strategically indistinguishable if and only if they are pairwise inseparable.

If there are private values and every payoff type is value distinguished, then every pair of types will be pairwise separable and thus strategically distinguishable. Thus strategic indistinguishability arises only when the degree of interdependence in preferences is large. We can illustrate this within the context of our single-unit auction example. As the utility function $u_i(\cdot)$ is linear in the monetary transfer for all types and all agents, the separability must come from different valuations of the object. For given type set profile Ψ_{-i} of all agents but i, we can identify the set of possible (expected) valuations of agent i with type θ_i

by writing:

$$
\begin{aligned}
&V_i(\theta_i, \Psi_{-i}) \\
&\quad = \left\{ v_i \in \mathbb{R}_+ \,\middle|\, \exists \lambda_i \in \Delta(\Psi_{-i}) \text{ s.t. } v_i = \theta_i + \gamma \sum_{\theta_{-i} \in \Psi_{-i}} \lambda_i(\theta_{-i}) \sum_{j \neq i} \theta_j \right\} \\
&\quad = \left[\theta_i + \gamma \sum_{j \neq i} \min \Psi_j,\ \theta_i + \gamma \sum_{j \neq i} \max \Psi_j \right], \qquad (5.11)
\end{aligned}
$$

where we write with minor abuse of notation, $\min \Psi_j$ and $\max \Psi_j$ to identify the smallest and largest real number in the set Ψ_j, respectively. Now we say that Ψ_{-i} separates Ψ_i if and only if

$$\bigcap_{\theta_i \in \Psi_i} V_i(\theta_i, \Psi_{-i}) = \varnothing.$$

By the linearity of the valuation, this is equivalent to requiring that

$$V_i(\max \Psi_i, \Psi_{-i}) \cap V_i(\min \Psi_i, \Psi_{-i}) = \varnothing.$$

By (5.11), this will hold if and only if

$$\max \Psi_i + \gamma \sum_{j \neq i} \min \Psi_j > \min \Psi_i + \gamma \sum_{j \neq i} \max \Psi_j.$$

We can rewrite the inequality as

$$\max \Psi_i - \min \Psi_i > \gamma \sum_{j \neq i} (\max \Psi_j - \min \Psi_j).$$

Thus Ψ_{-i} separates Ψ_i if and only if the difference between the smallest and the largest element in the set Ψ_i is larger than the weighted sum of the differences of the smallest and the largest element in the remaining sets Ψ_j for all $j \neq i$. Conversely, Ψ_{-i} does not separate Ψ_i if the above inequality is reversed, i.e.,

$$\max \Psi_i - \min \Psi_i \leq \gamma \sum_{j \neq i} (\max \Psi_j - \min \Psi_j). \qquad (5.12)$$

We write Ξ_i^k for the kth level inseparable sets of player i, and we have:

$$\Xi_i^0 = 2^{\Theta_i},$$

and define an inductive process by:

$$\Xi_i^{k+1} = \{\Psi_i \in \Xi_i^k \,|\, \Psi_{-i} \text{ does not separate } \Psi_i, \text{ for some } \Psi_{-i} \in \Xi_{-i}^k\},$$

and a (finite) limit type set profile is defined by:

$$\Xi_i^* = \bigcap_{k \geq 0} \Xi_i^k.$$

Now, we can identify the kth level inseparable set for the single unit auction example as follows. By (5.12), we have

$$\Xi_i^k = \left\{ \Psi_i \in \Xi_i^k \middle| \max \Psi_i - \min \Psi_i \leq \gamma \sum_{j \neq i} \max_{\Psi_j \in \Xi_j^k} (\max \Psi_j - \min \Psi_j) \right\},$$

Now by induction, we have that

$$\Xi_i^{k+1} = \{\Psi_i | \max \Psi_i - \min \Psi_i \leq (\gamma(I-1))^k\}.$$

Thus if $\gamma(I-1) < 1$, Ξ_i^* consists of singletons, $\Xi_i^* = (\{\theta_i\})_{\theta_i \in [0,1]}$, while if $\gamma(I-1) \geq 1$, Ξ_i^* consists of all subsets, $\Xi_i^* = 2^{[0,1]}$. In consequence, we find that if $\gamma < \frac{1}{I-1}$, so that interdependence is not too large, every distinct pair of types are pairwise separable. If on the other hand, $\gamma \geq \frac{1}{I-1}$, then every pair of payoff types are pairwise inseparable.

While our sufficiency argument for robust virtual implementation builds on Abreu and Matsushima (1992), the interpretation of our results ends up being rather different. In a standard Bayesian setting, the measurability condition of Abreu and Matsushima (1992) is arguably a weak technical requirement. As a result, the "bottom line" of the virtual implementation literature has been that full implementation, i.e., getting rid of undesirable equilibria, does not impose any substantive constraints beyond incentive compatibility, i.e., the existence of desirable equilibria. By requiring the more demanding, but more plausible, robust formulation of incomplete information, we end up with a condition that is substantive (imposing significantly more structure in interdependent value environments than incentive compatibility) and easily interpretable.

A conclusion that emerges from Bergemann and Morris (2009a,b, 2011a), and that we developed here within the single good example,

is that we keep ending up with the same moderate interdependence condition, $\gamma < \frac{1}{I-1}$, as a necessary and sufficient condition for full implementation. In general, though, the robust monotonicity condition of Bergemann and Morris (2011a) (and its contraction property version in Bergemann and Morris (2009a)) neither implies nor is implied by robust measurability, as we show by examples in Bergemann and Morris (2009b). Kunimoto and Serrano (2010) present a detailed discussion of these conditions and develop an argument as to why robust monotonicity should be seen as the weaker of the independent conditions. Artemov et al. (2010) characterize an analogue of robust measurability under local robustness conditions and argue that it is a weak condition.

In Section 6.3 of Bergemann and Morris (2009b), we do briefly reconsider the single good example under a local robustness condition: suppose that in the single good example, each agent puts probability mass $1 - \delta$ on a uniform distribution over the payoff types of other agents, but that δ probability may be allocated to any beliefs. Thus if $\delta = 0$, we have a standard payoff type space with independent types and if $\delta = 1$ we have the universal type space that is the focus of the present research. We show that virtual implementation is possible if and only if $\delta\gamma < \frac{1}{I-1}$. In this sense, at least within the single good example, the path to global robustness through local robustness is smooth.

6

Open Issues

In most of the work discussed, we defined the allocation problem in terms of social choice function or correspondence, which specified for each profile of payoff types θ a specific allocation. Importantly, the social choice function was defined independent of the beliefs of the agents and/or the principal. While this specification accommodates many allocation problems, in particular the socially efficient allocation, it cannot represent others, such as revenue maximizing allocations. Here, the allocation rule typically depends on the beliefs of the principal or the agents, as the optimal allocation relies on trading off outcomes across different states, where the trade-offs have to be evaluated with the likelihood of each state, and hence requires the use of beliefs. Bergemann and Schlag (2008) suggest a possible approach to analyze revenue maximization problems in the absence of prior beliefs. They consider the classic monopoly problem of a seller who offers a homogenous good to buyers with privately known valuations. In the absence of a (common) prior, we require the seller to minimize his expected regret through an optimal pricing policy. The resulting pricing policy hedges against the uncertainty with respect to the true distribution through a uniquely determined randomized pricing policy. And

while the resulting mixed strategy can be interpreted as the optimal pricing rule against a specific prior distribution, a random pricing policy is never the uniquely optimal policy given a known prior. In fact, against a known prior, there always exists an optimal pricing rule that is deterministic. Bergemann and Schlag (2011) consider the problem of optimal pricing when the seller has *some* prior information. In this version of the problem the seller knows that demand will be in a *small* neighborhood of a *given* model distribution. We characterized the optimal pricing policy under two distinct, but related, decision criteria with multiple priors: (*i*) maximin expected utility and (*ii*) minimax expected regret. The resulting model can be interpreted as a locally robust version of the classic problem of optimal monopoly pricing.

A second, and related, limitation of the present work is that we were mostly concerned with "global" notions of robustness. We allowed for any beliefs and higher-order beliefs consistent with the existing model. It would be of interest to look at "local" notions of robustness, where more limited perturbation of the types and information structures are considered. For example, in ongoing work, Bergemann and Morris (2011b), we consider games with incomplete information and ask what predictions can be made with the knowledge of a common prior over the payoff relevant states, and importantly in the absence of any additional information about the private information, the type space, of the agents. Thus, we consider a common prior about the relevant state, but are agnostic with respect to the beliefs and higher-order beliefs of the agents. In Bergemann and Morris (2011b), we use the structure of quadratic payoffs, and hence linear best response to analyze the set of possible equilibrium distributions in terms of moment restrictions. In Bergemann and Morris (2012a), we develop the associated equilibrium concept, which we refer to as Bayes correlated equilibrium for general finite action, finite agent games with incomplete information and establish how the equilibrium set depends on and changes with the private information of the agents. Bergemann et al. (2013a) consider the classic problem of price discrimination by a monopolist, and establish the exact set of payoffs that can be achieved when the monopolist may have access to additional information that leads him to segment the markets. In related work, Bergemann et al. (2013b) characterize the set

of equilibrium payoffs in the first price auction when the bidders may have access to strategic information, in the form of beliefs and higher order beliefs about the other bidders. In these papers, the impact of private information on the equilibrium payoffs is analyzed by means of the notion of Bayes correlated equilibrium as defined in Bergemann and Morris (2012a).

A similar approach would seem to have promise in the realm of mechanism design as well. For example, in a first price auction, one might attempt to find the set of possible equilibrium bid distributions that are generated by all information structures consistent with a given common prior over valuations. Likewise, one might investigate, the nature of the optimal auction when the principal has only limited information about the nature of the private information of the agents.

References

Abreu, D. and H. Matsushima (1992), 'Virtual implementation in iteratively undominated strategies: Incomplete information'. Discussion paper, Princeton University and University of Tokyo.

Abreu, D. and A. Sen (1991), 'Virtual implementation in Nash equilibrium'. *Econometrica* **59**, 997–1021.

Aghion, P., D. Fudenberg, R. Holden, T. Kunimoto, and O. Tercieux (2012), 'Subgame-perfect implementation under information perturbations'. *Quarterly Journal of Economics*. forthcoming.

Artemov, G., T. Kunimoto, and R. Serrano (2010), 'Robust virtual implementation with incomplete information: Towards a reinterpretation of the Wilson doctrine'. Discussion paper, University of Melbourne, McGill University and Brown University.

Ausubel, L. M. and P. Milgrom (2005), 'The lovely but lonely vickrey auction'. In: R. S. P. Cramton and Y. Shoham (eds.): *Combinatorial Auctions*.

Barelli, P. (2009), 'On the genericity of full surplus extraction in mechanism design'. *Journal of Economic Theory* **144**, 1320–1332.

Battigalli, P. and M. Siniscalchi (2003), 'Rationalization and incomplete information'. *Advances in Theoretical Economics* **3**. Article 3.

Bergemann, D., B. Brooks, and S. Morris (2013a), *Extremal Information Structures in the First Price Auction*. Yale University and Princeton University.

Bergemann, D., B. Brooks, and S. Morris (2013b), *The Limits of Price Discrimination*. Yale University and Princeton University.

Bergemann, D. and S. Morris (2001), 'Robust mechanism design'. Discussion Paper, Yale University http://www.princeton.edu/∼smorris/pdfs/robustmechanism2001.pdf.

Bergemann, D. and S. Morris (2005), 'Robust mechanism design'. *Econometrica* **73**, 1771–1813.

Bergemann, D. and S. Morris (2007), 'An ascending auction for interdependent values: Uniqueness and robustness to strategic uncertainty'. *American Economic Review Papers and Proceedings* **97**, 125–130.

Bergemann, D. and S. Morris (2008a), 'Ex post implementation'. *Games and Economic Behavior* **63**, 527–566.

Bergemann, D. and S. Morris (2008b), 'The role of the common prior in robust implementation'. *Journal of the European Economic Association Papers and Proceedings* **6**, 551–559.

Bergemann, D. and S. Morris (2009a), 'Robust implementation in direct mechanisms'. *Review of Economic Studies* **76**, 1175–1206.

Bergemann, D. and S. Morris (2009b), 'Robust virtual implementation'. *Theoretical Economics* **4**, 45–88.

Bergemann, D. and S. Morris (2011a), 'Robust implementation in general mechanisms'. *Games and Economic Behavior* **71**(1666), 261–281.

Bergemann, D. and S. Morris (2011b), 'Robust predictions in games of incomplete information'. Discussion paper, Cowles Foundation for Research in Economics, Yale University.

Bergemann, D. and S. Morris (2012a), 'Bayes correlated equilibrium and the comparison of information structures'. Discussion paper, Cowles Foundation for Research in Economics, Yale University and Princeton University.

Bergemann, D. and S. Morris (2012b), *Robust Mechanism Design*. Singapore: World Scientific Publishing.

Bergemann, D., S. Morris, and S. Takahashi (2010), 'Interdependent preferences and strategic distinguishability'. Discussion Paper 1772, Cowles Foundation for Research in Economics, Yale University.

Bergemann, D., S. Morris, and S. Takahashi (2012), 'Efficient auctions and interdependent types'. *American Economic Review: Papers and Proceedings* **102**, 319–324.

Bergemann, D., S. Morris, and O. Tercieux (2011), 'Rationalizable implementation'. *Journal of Economic Theory* **146**, 1253–1274.

Bergemann, D. and K. Schlag (2008), 'Pricing without priors'. *Journal of the European Economic Association Papers and Proceedings* **6**, 560–569.

Bergemann, D. and K. Schlag (2011), 'Robust monopoly pricing'. *Journal of Economic Theory* **146**, 2527–2543.

Bierbrauer, F. and M. Hellwig (2011), 'Mechanism design and voting for public good provision'. Discussion paper, Max Planck Institute for Research on Collective Goods.

Bikhchandani, S. (2006), 'Ex post implementation in environments with private goods'. *Theoretical Economics* **1**, 369–393.

Birulin, O. (2003), 'Inefficient ex post equilibria in efficient auctions'. *Economic Theory* **22**, 675–683.

Blume, A. and P. Heidhues (2004), 'All equilibria of the Vickrey auction'. *Journal of Economic Theory* **114**, 170–177.

Börgers, T. and D. Smith (2012), 'Robust mechanism design and dominant strategy voting rules'. *Theoretical Economics*. forthcoming.

Brandenburger, A. and E. Dekel (1987), 'Rationalizability and correlated equilibria'. *Econometrica* **55**, 1391–1402.

Chen, Y. and S. Xiong (2010), 'The genericity of belief-determine-preferences models revisited'. *Journal of Economic Theory* **146**, 751–761.

Chen, Y. and S. Xiong (2013), 'Genericity and robustness of full surplus extraction'. *Econometrica*. forthcoming.

Choi, J. and T. Kim (1999), 'A nonparametric, efficient public good decision mechanism: Undominated bayesian implementation'. *Games and Economic Behavior* **27**, 64–85.

Chung, K.-S. and J. Ely (2001), 'Efficient and dominance solvable auctions with interdependent valuations'. Discussion paper, Northwestern University.

Chung, K.-S. and J. Ely (2003), 'Implementation with Near-Complete Information'. *Econometrica* **71**, 857–871.

Chung, K.-S. and J. Ely (2007), 'Foundations of dominant strategy mechanisms'. *Review of Economic Studies* **74**, 447–476.

Cremer, J. and R. McLean (1988), 'Full extraction of the surplus in Bayesian and dominant strategy auctions'. *Econometrica* **56**, 1247–1258.

Dasgupta, P., P. Hammond, and E. Maskin (1979), 'The implementation of social choice rules. Some general results on incentive compatibility'. *Review of Economic Studies* **66**, 185–216.

Dasgupta, P. and E. Maskin (2000), 'Efficient auctions'. *Quarterly Journal of Economics* **115**, 341–388.

Dekel, E., D. Fudenberg, and S. Morris (2006), 'Topologies on types'. *Theoretical Economics* **1**, 275–309.

Eso, P. and E. Maskin (2002), 'Multi-good efficient auctions with multidimensional information'. Discussion paper, Northwestern University and Institute for Advanced Studies.

Fang, H. and S. Morris (2006), 'Multidimensional private value auctions'. *Journal of Economic Theory* **126**, 1–30.

Gizatulina, A. and M. Hellwig (2010), 'Informational smallness and the scope for limiting information rents'. *Journal of Economic Theory* **145**, 2260–2281.

Gizatulina, A. and M. Hellwig (2011), 'Beliefs, payoffs, information: On the robustness of the BDP property in models with endogeneous beliefs'. Discussion paper, Max Planck Institute for Research on Collective Goods.

Harsanyi, J. (1967–68), 'Games with incomplete information played by 'Bayesian' players'. *Management Science* **14**, 159–189, 320–334, 485–502.

Heifetz, A. and Z. Neeman (2006), 'On the generic (im)possibility of full surplus extraction in mechanism design'. *Econometrica* **74**, 213–233.

Hurwicz, L. (1972), 'On informationally decentralized systems'. In: C. McGuire and R. Radner (eds.): *Decisions and Organizations*. Amsterdam: North-Holland, pp. 297–336.

Jackson, M. (1991), 'Bayesian implementation'. *Econometrica* **59**, 461–477.

Jehiel, P., M. Meyer-Ter-Vehn, and B. Moldovanu (2008), 'Ex-post implementation and preference aggregation via potentials'. *Economic Theory* **37**, 469–490.

Jehiel, P., M. Meyer-Ter-Vehn, and B. Moldovanu (2010), 'Locally robust implementation and its limits'. *Journal of Economic Theory*. forthcoming.

Jehiel, P. and B. Moldovanu (2001), 'Efficient design with interdependent valuations'. *Econometrica* **69**, 1237–1259.

Jehiel, P., B. Moldovanu, M. Meyer-Ter-Vehn, and B. Zame (2006), 'The limits of ex post implementation'. *Econometrica* **74**, 585–610.

Kunimoto, T. and R. Serrano (2010), 'Evaluating the conditions for robust mechanism design'. Discussion paper, McGill University and Brown University.

Laffont, J. and D. Martimort (2000), 'Mechanism design with collusion and correlation'. *Econometrica* **65**, 309–342.

Ledyard, J. (1979), 'Dominant strategy mechanisms and incomplete information'. In: C. . J.-J. Laffont (ed.): *Aggregation and Revelation of Preferences*. Amsterdam: North-Holland, pp. 309–319.

Ledyard, J. O. (1978), 'Incentive compatibility and incomplete information'. *Journal of Economic Theory* **18**, 171–189.

Lopomo, G., L. Rigotti, and C. Shannon (2009), 'Uncertainty in mechanism design'. Discussion paper.

Maskin, E. (1992), 'Auctions and privatization'. In: H. Siebert (ed.): *Privatization: Symposium in Honor of Herbert Giersch*. J.C.B. Mohr: Tuebingen, pp. 115–136.

Maskin, E. (1999), 'Nash equilibrium and welfare optimality'. *Review of Economic Studies* **66**, 23–38.

Matsushima, H. (1988), 'A new approach to the implementation problem'. *Journal of Economic Theory* **45**, 128–144.

McAfee, P. and P. Reny (1992), 'Correlated information and mechanism design'. *Econometrica* **60**, 395–421.

McLean, R. and A. Postlewaite (2002), 'Informational size and incentive compatibility'. *Econometrica* **70**, 2421–2453.

Mertens, J. and S. Zamir (1985), 'Formalization of Bayesian analysis for games with incomplete information'. *International Journal of Game Theory* **14**, 1–29.

Meyer-Ter-Vehn, M. and S. Morris (2011), 'The robustness of robust implementation'. *Journal of Economic Theory* **146**, 2093–2104.

Milgrom, P. and J. Roberts (1990), 'Rationalizability, learning and equilibrium in games with strategic complementarities'. *Econometrica* **58**, 1255–1277.

Moore, J. and R. Repullo (1988), 'Subgame perfect implementation'. *Econometrica* **56**, 1191–1220.

Mueller, C. (2009), 'Robust virtual implementation under common strong belief in rationality'. Discussion paper, University of Minnesota.

Myerson, R. (1981), 'Optimal auction design'. *Mathematics of Operations Research* **6**, 58–73.

Neeman, Z. (2004), 'The relevance of private information in mechanism design'. *Journal of Economic Theory* **117**, 55–77.

Oury, M. and O. Tercieux (2012), 'Continuous implementation'. *Econometrica* **80**, 1605–1637.

Penta, A. (2011), 'Robust dynamic mechanism design'. Discussion paper, University of Wisconsin.

Perry, M. and P. Reny (2002), 'An ex post efficient auction'. *Econometrica* **70**, 1199–1212.

Peters, M. (2001), 'Surplus extraction and competition'. *Review of Economic Studies* **68**, 613–631.

Plum, M. (1992), 'Characterization and computation of Nash equilibria for auctions with incomplete information'. *International Journal of Game Theory* **20**, 393–418.

Postlewaite, A. and D. Schmeidler (1986), 'Implementation in differential information economies'. *Journal of Economic Theory* **39**, 14–33.

Robert, J. (1991), 'Continuity in auction design'. *Journal of Economic Theory* **55**, 169–179.

Smith, D. (2010), 'A prior free efficiency comparison of mechanisms for the public good problem'. Discussion paper, University of Michigan.

Wilson, R. (1985), 'Incentive efficiency of double auctions'. *Econometrica* **53**, 1101–1116.

Wilson, R. (1987), 'Game-theoretic analyses of trading processes'. In: T. Bewley (ed.): *Advances in Economic Theory: Fifth World Congress*. Cambridge: Cambridge University Press, pp. 33–70.

Yamashita, T. (2011), 'Robust welfare guarantees in bilateral trading mechanisms'. Discussion paper, Stanford University.

Yokoo, M., Y. Sakurai, and S. Matsubara (2004), 'The effect of false-name bids in combinatorial auctions: New fraud in internet auctions'. *Games and Economic Behavior* **46**, 174–188.