The Dissertation Committee for Ina Angelova Taneva
certifies that this is the approved version of the following dissertation:

# Essays on Information and Mechanism Design

Committee:

Maxwell Stinchcombe, Co-Supervisor

Laurent Mathevet, Co-Supervisor

Thomas Wiseman

Svetlana Boyarchenko

Brian Rogers

# Essays on Information and Mechanism Design

by

# Ina Angelova Taneva, B.S.Econ.; M.Econ.; M.S.Econ.

**DISSERTATION**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**DOCTOR OF PHILOSOPHY**

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2014

*To Jutta and John Brennan*

"One looks back with appreciation to the brilliant teachers, but with gratitude to those who touched our human feelings. The curriculum is so much necessary raw material, but warmth is the vital element for the growing plant and for the soul of the child."

—Carl Jung

# Acknowledgments

First and foremost I would like to thank my main advisor, Laurent Mathevet, without whom I would not be writing this acknowledgement section. Starting to work with him in the third year of my PhD studies was the best thing that happened to me academically. He spent countless hours working with me, advising me, and preparing me for the job market. His criticism was always honest and constructive, and sometimes even harsh, but there was a lot of warmth in his well-timed words of encouragement. Laurent has been and continues to be a great mentor. His creativity and fascination with the subject of economic theory are an abounding source of inspiration for me.

Max Stinchcombe has also been a great advisor. His knowledge, expertise, and insight are truly amazing and he always knew where to direct me. His palpable passion for mathematics and theory is contagious and very motivating. I am also extremely thankful to Max for funding my last fall semester at UT, which helped tremendously during my time on the job market.

I am greatly indebted to Tom Wiseman, who spent a lot of time guiding and helping me despite not being my main advisor. His sharpness, candor, and humor made my meetings with him not only extremely productive, but also highly enjoyable. Tom is without a doubt one of the best teachers I have ever had and has been a true inspiration since my very first day at UT.

I would like to thank Svetlana Boyarchenko and Brian Rogers for serving on my committee and attending my presentations. Their questions and feedback have been very valuable. Special thanks are due to Svetlana for repeatedly choosing me as

her teaching assistant for her Microeconomic Theory class, which made it a lot easier for me to devote more time to research.

I would like to thank Jason Abrevaya who always took the time out of his busy schedule to help me with important decisions, answer my questions, and even come to campus for a last minute mock interview during the winter holiday break. A big thank you goes to the best program coordinator ever — Vivian Goldman-Leffler — for her efficiency, promptness, and help, as well as her great sense of humor and friendship.

Jutta and John Brennan, to whom this dissertation is dedicated, have been amazing teachers, senseis, and friends. Without their faith, encouragement, and unconditional support, I would not have completed this degree. In fact, during my time as an undergraduate student at Otto-von-Guericke Universität Magdeburg, Dr. Brennan was the first to suggest that I pursue a doctoral degree in economics. In a way, he has followed this long process from its very inception. I also thank Gerhard Schwödiauer, Karim Sadrieh, Ludwig von Auer, and Ronnie Schöb for their help and advice during the application process which resulted in my admission to UT.

I am grateful to my fellow graduate students and friends here in Austin. I have learned a lot from them both in the context of academic and social fellowship. I would also like to thank my family and friends back in Bulgaria, who have been cheering me on all these years. I am especially grateful to my mom, who always supported me and never stopped believing. Last but certainly not least, I would like to thank my true friend and running buddy, Penny, who taught me by example that there is always time to stop and smell the flowers.

# Essays on Information and Mechanism Design

Ina Angelova Taneva, Ph.D.
The University of Texas at Austin, 2014

Co-Supervisors:    Maxwell Stinchcombe
                   Laurent Mathevet

My dissertation studies the optimal design of institutions and information structures for different objectives of a designer or a social planner. The questions addressed are interesting both from a theoretical point of view, and in terms of their real-life applications.

The first chapter of the dissertation focuses on supermodular mechanism design in environments with arbitrary finite type spaces and interdependent valuations. In these environments, the designer may have to use Bayesian equilibrium as a solution concept, because ex post implementation may not be possible. We propose direct Bayesian mechanisms that are robust to certain forms of bounded rationality while controlling for equilibrium multiplicity. In quasi-linear environments with informational and allocative externalities, we show that any Bayesian mechanism that implements a social choice function can be converted into a supermodular mechanism that also implements the original decision rule. The proposed supermodular mechanism can be chosen in a way that minimizes the size of the equilibrium set, and we provide two sets of sufficient conditions to this effect: for general decision rules and for decision rules that satisfy a certain requirement. This is followed by conditions for supermodular implementation in unique equilibrium.

The second chapter looks at the incentives of a revenue-maximizing seller (designer) who discloses information to a number of interacting bidders (agents). In particular, the designer chooses the level of precision with which agents can infer the quality of a common-value object from their privately observed signals. We restrict attention to the second-price sealed-bid auction format. If the seller has perfect commitment power and can choose the precision level before observing the quality of the object, in the presence of any small cost to precision it is ex ante optimal for her to choose completely uninformative signals. For the case when the seller chooses the precision level after observing the quality of the object, we characterize pooling, partial pooling, and separating equilibria. We show that in this setting the cost associated with precision can be viewed as a form of commitment device: if costs are too low, the best pooling equilibrium ceases to exist as the high type seller is too tempted to separate. Thus, the seller ends up with a lower ex ante expected payoff than in the case when cost parameters are above a certain threshold.

The third chapter of this dissertation studies the optimal choice of information structure from the perspective of a designer maximizing a certain objective function. Generally speaking, there are two ways of creating incentives for interacting agents to behave in a desired way. One is by providing appropriate payoff incentives, which is the subject of mechanism design. The other is by choosing the information that agents observe, which we refer to as information design. We consider a model of symmetric information where a designer chooses and announces the information structure about a payoff relevant state. The interacting agents observe the signal realizations, update their beliefs, and take actions which affect the welfare of both the designer and the agents. We characterize the general finite approach to deriving the optimal information structure — the one that maximizes the designer's ex ante expected utility subject to agents playing a Bayes Nash equilibrium. We then apply the general

approach to a symmetric two state, two agent, and two actions environment in a parameterized underlying game and fully characterize the optimal information structure. It is never strictly optimal for the designer to use conditionally independent private signals. The optimal information structure may be a public signal, or may consist of correlated private signals. Finally, we examine how changes in the underlying game affect the designer's maximum payoff. This exercise provides a joint mechanism/information design perspective.

# Table of Contents

# Chapter 1

# Finite Supermodular Design with Interdependent Valuations

## 1.1   Introduction

This paper studies supermodular mechanism design in environments with interdependent valuations and arbitrary (in particular, multidimensional) finite type spaces.[1] This approach was introduced by Mathevet [45] in differentiable quasilinear environments with private values and one-dimensional types.[2] The main motivation is to design direct mechanisms that are robust to certain forms of bounded rationality while controlling for equilibrium multiplicity. It is important to extend supermodular mechanism design to environments with informational and allocative externalities and multidimensional types for at least two reasons. First, these environments capture many realistic situations. Second, it is often impossible to use dominant strategy or ex-post implementation in these settings (see Jehiel *et al.* [38] and Section 1.2),

---

[1]Substantial portions of this chapter have been published in an article coauthored by Laurent Mathevet and Ina Taneva: "Finite Supermodular Design with Interdependent Valuations", *Games and Economic Behavior* 82 (2013), 327-349. The part of the work conducted by the present author cuts across all aspects and portions of the paper, which may be summarized as follows. The idea of supermodular mechanism design is due to Mathevet. The current author contributed the simple motivating example, which exemplifies the impossibility of ex post implementation and the virtues of supermodular implementation in the context of interdependent valuations and multidimensional types. The authors split, more or less equally, the ideas on how to approach the mathematical analysis and proofs of the results.

[2]Chen [14] was the first to propose a supermodular mechanism (to implement the Lindahl correspondence).

and thus the designer may resort to Bayesian equilibrium as a solution concept. It becomes useful to have a simple method for improving the behavioral robustness of Bayesian mechanisms.

In this paper, we are concerned with the design of supermodular mechanisms whose equilibrium set is of minimal size. We call this minimal supermodular implementation. Supermodular mechanism design aims to induce the right incentives so that agents play a supermodular game. Supermodular games are games where players have monotone best responses, i.e. each player wants to play a "larger" strategy if others do so as well. On the theoretical front, the reasons for using supermodular mechanisms stem from Milgrom and Roberts [48], [49] and Vives [57]: supermodular games have extremal equilibria, a smallest and a largest one, that enclose all the iteratively undominated strategies and all the limit points of all adaptive and sophisticated learning dynamics. Therefore, supermodular games are robust to a wide range of behaviors, including boundedly rational behaviors. In particular, if the designer had the opportunity to use her mechanism repeatedly, then adaptive learners (Milgrom and Roberts [48]) would end up within the interval prediction, which is the interval between the extremal equilibria. Therefore, the objective of minimizing the size of the interval prediction has several virtues. It minimizes the multiple equilibrium problem, since all equilibria are contained in it.[3] It also guarantees a more accurate convergence of the learning dynamics. Ideally, this interval reduces to a single point in certain situations (see Section 1.2), thereby solving the multiplicity issue and ensuring convergence of all dynamics.

Supermodular mechanisms have other attractive theoretical properties. Not

_____

[3]If the outcome function of the mechanism is continuous and if the interval prediction is tight, then all equilibrium outcomes are close, so that the output of the mechanism must be close to the socially desirable objective.

only are their mixed strategy equilibria unstable (Echenique and Edlin [25]), which justifies ruling them out of the analysis, but many pure equilibria are stable, such as the extremal equilibria (Echenique [26]). Thus, a perturbation should not destabilize permanently a socially desirable alternative implemented via a supermodular mechanism (provided the underlying equilibrium was stable).

The robustness properties of supermodular mechanisms have been corroborated by several experiments. Chen and Gazzale [17] run experiments on a game for which they control the amount of supermodularity. They show how convergence in that game is significantly better when it is supermodular. Healy [33] tests five public goods mechanisms and he observes that subjects learn to play the equilibrium only in those mechanisms that induce a supermodular game. Other experiments (e.g. Chen and Plott [15], and Chen and Tang [16]) provide results emphasizing the importance of dynamic convergence in the context of implementation. Most of these experiments demonstrate that convergence to an equilibrium is not a trivial issue.

In this paper, we generalize supermodular mechanism design to environments with allocative externalities, interdependent valuations (i.e. informational externalities) and arbitrary (finite) type spaces. There are two important reasons for doing so.

Firstly, it allows us to cover mechanism design problems of interest. The importance of allocative externalities is well documented in the literature. Jehiel and Moldovanu [36] use patent licensing in an oligopolistic market as an example. Informational externalities are also a realistic assumption, proved to be interestingly challenging by many papers (Cremer and McLean [21], Maskin [44], Dasgupta and Maskin [23], Perry and Reny [52], Chung and Ely [19], Bergemann and Morris [4], etc). Finally, it is often natural to interpret information as a multidimensional type in many design problems. Consider, for example, oil companies bidding to obtain a

3

drilling permit. Their private information is modeled as a multidimensional signal (e.g., expected amount of oil in the oil field, proximity to other reserves, etc).

Secondly, it is difficult to use dominant strategy or ex-post implementation in those environments — with allocative externalities, interdependent valuations and multidimensional types — and thus behaviorally-robust Bayesian mechanisms become especially appealing. In quasilinear environments with interdependent valuations and multidimensional types, many impossibility results limit the set of available solution concepts. The conclusions are rather pessimistic about dominant strategy equilibrium and ex-post equilibrium. Williams and Radner [59] show that efficient dominant strategy implementation is generally not possible when agents have interdependent valuations. Jehiel *et al.* [38] prove a strong impossibility result: when types are multidimensional and valuations are interdependent, only trivial decision rules can generically be implemented in ex-post equilibrium. If the designer wants to implement a meaningful social choice function, not even necessarily efficient, she may have to use Bayesian equilibrium as a solution concept (see Section 1.2). Even then, impossibility results exist. Jehiel and Moldovanu [37] show that it is difficult to reconcile Bayesian incentive compatibility with some efficiency constraint. These negative results indicate that Bayesian equilibrium may often be a natural candidate as a solution concept. However, playing a Bayesian equilibrium requires more, in general, on the part of the agents. Agents have to be Bayesian rational, and strong knowledge assumptions about the information structure and the rationality must hold (Brandenburger and Dekel [11]). For example, Bayesian equilibrium calls for correct predictions of opponents' play to determine one's own strategy. In this context, the ability to construct supermodular Bayesian mechanisms is attractive, because eventual play of some equilibrium can be achieved by unsophisticated agents who follow simple behavioral rules.

4

Our paper provides methods for converting *any* truthful Bayesian mechanism into a (truthful) supermodular mechanism whose equilibrium set is of minimal size. The idea is to create complementarities between agents' announcements by augmenting the original transfer scheme with a function. This function vanishes in expectation and therefore preserves incentive compatibility. Although there exist many ways in which a mechanism can be transformed into a supermodular mechanism, we derive the one that most adequately addresses the multiple equilibrium problem. To this purpose, we add just enough strategic complementarities to ensure that a supermodular game is induced, but not in any excess of that level.

We present two sets of results for minimal supermodular implementation. In both instances, "best" is used to designate the mechanism with the smallest interval prediction. The first result shows that if a social choice function is implementable, then its decision rule can be implemented by the best supermodular mechanism among all the supermodular mechanisms *whose transfers are in a certain class*. No additional condition is required. In particular, this result holds for all (implementable) decision rules and all valuation functions. The result also provides an explicit transfer scheme.

The second result characterizes the *overall* best supermodular mechanism among all possible supermodular mechanisms or transfers: if a social choice function is implementable, then its decision rule can be implemented by the (overall) best supermodular mechanism if and only if some (explicitly stated) finite system of linear equations admits a solution. This result determines the existence and the numerical values of the minimally supermodular transfers. As a complementary result, we provide a sufficient condition, order reducibility, under which existence of a solution is guaranteed. Although the former result reaches a weaker conclusion than the latter, it applies under very general conditions.

Finally, we provide conditions that ensure that truthtelling is the (essentially)

unique equilibrium. For fine type spaces, this guarantees stability: all learning dynamics converge to the truthful equilibrium and the game is dominance solvable.

Beyond the generalizations of supermodular mechanism design, this paper provides new insights into the design of minimally supermodular mechanisms. The use of finite types clarifies the existence and the construction of these mechanisms. For example, the fact that minimally supermodular mechanisms are a solution to a system of linear equations allows the application of numerical methods for designing them. Further, this formulation helps clarify what the necessary conditions for the existence of such transfers might look like. In this regard, we provide a simple sufficient condition under which the system admits a solution; although this condition is not necessary, it gives valuable information about the type of mechanism design problems that might cause issues for minimally supermodular implementation.

A number of papers are related to our work. The first paper to present a supermodular mechanism was Chen [14]. Mathevet [45] developed supermodular mechanism design as a general method under incomplete information. Since his paper is the closest to ours, our contribution deserves clarification. As already said, our environment is more general, due to the interdependent values and the multidimensional types, but we use a finite setup. In Mathevet [45]'s environment, dominant strategy implementation is still a powerful tool, while it in our environment it is significantly less applicable. The present paper also clarifies the construction of minimally supermodular mechanisms, especially with our reducibility condition and the formulation as a linear system. Finally, we propose different options for minimal supermodular design when sufficient conditions fail, while Mathevet [45] does not. In particular, our first main result always applies. Cabrales and Serrano [13] study implementation with boundedly rational agents who follow adaptive better response dynamics. This learning rule excludes learning processes that our mechanisms allow for. Finally, our

6

paper is related to the literature on rationalizable implementation (e.g. Abreu and Matsushima [1], and Bergemann, Morris, and Tercieux [6]), because this solution concept has the potential to imply nice learning properties when a unique equilibrium is rationalizable. Abreu and Matsushima [1] show that any social choice function can be virtually implemented in iteratively undominated strategies. Their result is very powerful but their mechanism remains complex, as the dimension of the message space must be arbitrarily large. Experimental evidence does not support this mechanism (Sefton and Yavas [54]). Instead we look at direct mechanisms and exact implementation. In general, the concept of rationalizability is such that a strategy may not be rationalizable, because it is dominated by another dominated strategy, an argument a la Jackson [35]. For example, in Bergemann, Morris, and Tercieux [6], the best responses are not well-defined off-equilibrium, and off-equilibrium behaviors are one of our motivations.

The remainder of the paper is organized as follows. A motivating example is presented in Section 1.2. Section 1.3 defines the framework of supermodular mechanism design. Section 1.4 introduces the notion of minimal implementation and contains our two main results. Section 1.5 studies supermodular implementation in unique equilibrium. Concluding remarks appear in Section 1.7. All proofs are relegated to the Appendix.

## 1.2 Motivating Example

This section illustrates our approach in a simple example. In this example, the designer would like to make an efficient decision, but this cannot be done in ex-post equilibrium and, hence, also not in dominant strategies. Thus, the designer

may decide to work with Bayesian implementation.[4] We show how the designer can start from *any* truthful Bayesian mechanism, in particular one with poor stability properties, and turn it into a (truthful) supermodular mechanism with a unique equilibrium.

Consider a social planner who has to make a decision between two public projects, $A$: improving the roadway or $B$: building a park. There are two agents (e.g., food vendors), 1 and 2, on each side of the road who will be affected by the decision. Each agent $i$ receives a two-dimensional signal $\theta_i = (\theta_i^A, \theta_i^B) \in \{(1,2),(2,1)\}$, where each dimension represents a projected flow of traffic resulting from the implementation of the respective project. Types are drawn with equal probabilities: $\Pr(\theta_1 = (1,2)) = \Pr(\theta_2 = (1,2)) = .5$. Let $V_i^g(\theta)$ be $i$'s valuation for project $g \in \{A,B\}$ at $\theta = (\theta_1, \theta_2)$. Valuations represent the expected gain for each vendor from a project and are given in the following matrix (rows represent agent 1):

| $V(\theta)$ | $(1,2)$ | | $(2,1)$ | |
|---|---|---|---|---|
| $(1,2)$ | $V_1^A = 1$ | $V_2^A = 1$ | $V_1^A = 2$ | $V_2^A = 0$ |
| | $V_1^B = 2.5$ | $V_2^B = -1$ | $V_1^B = 1.5$ | $V_2^B = 1$ |
| $(2,1)$ | $V_1^A = 2$ | $V_2^A = 0$ | $V_1^A = 4$ | $V_2^A = -2$ |
| | $V_1^B = 1.5$ | $V_2^B = 1$ | $V_1^B = 1$ | $V_2^B = -1.$ |

Agent 1 always interprets the higher flow of traffic as good news in terms of revenue, and that view is reinforced by agent 2's signal. That is, 1's value for project $A$ $(B)$ increases in the first (second) dimension of her signal and of agent 2's signal. From 1's perspective, there are positive informational externalities coming from agent 2.

---

[4]There are many situations in which Bayesian implementation is possible, while ex-post implementation is not. So, we are not restricting attention to non-generic cases. Bayesian incentive compatibility (BIC) can be viewed as a weighted average of ex-post incentive constraints. Thus, some ex-post constraints can be violated while BIC holds, if other ex-post constraints hold with enough slack.

Agent 2, on the other hand, considers the first dimension of both signals to be bad news for project $A$. In particular, he thinks that the higher the increase in traffic due to improved roadway, the lower its profits will be, because increased traffic flow should attract other competitors. That is, 2's value for $A$ decreases in the first dimension of his signal and of 1's signal. As for project $B$, 2 thinks that it will be profitable for him only if there are medium levels of traffic, which corresponds to opposite estimates from 1 and 2, $\theta_i^B \neq \theta_j^B$. When both second dimensions are low, the park should not increase the market size and the profits; the same applies if both estimates are high, due to (expected) increased competition. Overall, from 2's perspective, the informational externalities coming from 1 are ambiguous (they are negative for project $A$ and ambiguous for $B$).

The designer's objective is to choose the efficient project, i.e., the project which maximizes the sum of valuations for each possible profile of types. Given the valuations, the efficient decision rule, which the designer would like to implement, is:

| $x(\hat{\theta})$ | $(1,2)$ | $(2,1)$ |
|---|---|---|
| $(1,2)$ | A | B |
| $(2,1)$ | B | A |

Denote agent $i$'s transfers as a function of reported types by:

| $t_i(\hat{\theta})$ | $(1,2)$ | $(2,1)$ |
|---|---|---|
| $(1,2)$ | $t_i^1$ | $t_i^2$ |
| $(2,1)$ | $t_i^3$ | $t_i^4$ |

The efficient decision rule is not ex-post incentive compatible.[5] To see why, let us consider the ex-post incentive compatibility conditions for agent 1 when agent 2's

---

[5]Ex-post incentive compatibility requires that for all $i$ and $\theta$, $u_i(x(\theta), \theta) \geq u_i(x(\theta_i', \theta_{-i}), \theta)$ for all $\theta_i'$. This means that if all other agents report truthfully, truthtelling is a best response for each agent $i$ at every possible realizations of types $\theta$.

type is $\theta_2 = (1, 2)$ and is truthfully reported. At type $\theta_1 = (1, 2)$, ex-post incentive compatibility for agent 1 requires

$$1 + t_1^1 \geq 2.5 + t_1^3.$$

At type $\theta_1 = (2, 1)$, ex-post incentive compatibility requires

$$1.5 + t_1^3 \geq 2 + t_1^1.$$

The last two inequalities cannot be jointly satisfied, which proves that the efficient decision rule is not ex-post implementable. The designer is therefore inclined to work with Bayesian implementation. We proceed to show that there exist transfers that implement the efficient decision rule in Bayesian equilibrium.

Bayesian incentive compatibility for agent 1 requires that truthtelling be weakly preferred to lying when her true type is $(1, 2)$

$$.5(1 + t_1^1) + .5(1.5 + t_1^2) \geq .5(2.5 + t_1^3) + .5(2 + t_1^4)$$

and when her true type is $(2, 1)$

$$.5(1.5 + t_1^3) + .5(4 + t_1^4) \geq .5(2 + t_1^1) + .5(1 + t_1^2).$$

Combining these two inequalities, we obtain that for any $t_1$ such that

$$2.5 \geq t_1^1 + t_1^2 - t_1^3 - t_1^4 \geq 2$$

the efficient decision rule is Bayesian incentive compatible for agent 1. Similarly, Bayesian incentive compatibility is satisfied for any $t_2$ such that

$$0 \geq t_2^1 - t_2^2 + t_2^3 - t_2^4 \geq -3.$$

In particular, the designer can choose:

10

| $t_1(\hat{\theta})$ | $(1,2)$ | $(2,1)$ |
|---|---|---|
| $(1,2)$ | -5 | 7 |
| $(2,1)$ | 4.75 | -5 |

| $t_2(\hat{\theta})$ | $(1,2)$ | $(2,1)$ |
|---|---|---|
| $(1,2)$ | 5 | -5 |
| $(2,1)$ | -5.5 | 5 |

As we will see, the magnitude of these transfers is large enough to offset any consideration about the valuations. Given these transfers, the resulting payoff matrix in the ex ante Bayesian game is:

| EU | truthtelling | constant $(1,2)$ | constant $(2,1)$ | always lie |
|---|---|---|---|---|
| truthtelling | 2.43*; .1* | 1.25; 0 | 3.5; -.5 | 2.31; -.62 |
| constant (1,2) | 2.37; .25 | -2.75; 4.75* | 8.62*; -5 | 3.5*; -.5 |
| constant (2,1) | 2.37;-.75 | 6.37*; -5.5 | -2.75; 4.75* | 1.25; 0 |
| always lie | 2.31; -.62 | 2.37; -0.75 | 2.37; 0.25* | 2.44; .12 |

where row entries and first payoffs correspond to agent 1, while column entries and second payoffs correspond to agent 2. Best responses are denoted by asterisks. The game described by this payoff matrix is not (ex ante) dominance solvable. Despite being the unique equilibrium, truthtelling is unstable; after a small perturbation, convergence to it fails under various dynamics. The intuition goes as follows. If agent 2 plays the constant announcement $(1,2)$ irrespective of his true type, then agent 1 will best-respond by announcing $(2,1)$ regardless of her type. In return, agent 2 will also announce $(2,1)$ for every type. Then agent 1 will want to play the constant announcement $(1, 2)$, followed by a constant announcement of $(1, 2)$ by agent 2. We are back to the original strategy of agent 2. These transfers give rise to cycling behaviors and the problem extends beyond best-response dynamics.

To overcome this problem, we propose converting the mechanism into a supermodular mechanism. The idea is to modify the original transfers in a way that adds complementarity between agents' reports. Start from the above transfers $t_i$. Given a collection of numbers $\{\delta_i(\theta) : \theta \in \Theta\}$, define

$$t_i^{SM}(\hat{\theta}) = \delta_i(\hat{\theta}) - E_{\theta_{-i}}[\delta_i(\hat{\theta}_i, \theta_{-i})] + E_{\theta_{-i}}[t_i(\hat{\theta}_i, \theta_{-i})] \tag{1.1}$$

for each agent $i$. Transfers $t_i^{SM}$ satisfy $E_{\theta_{-i}}[t_i^{SM}(\theta_i, \theta_{-i})] = E_{\theta_{-i}}[t_i(\theta_i, \theta_{-i})]$ for all $\theta_i$, i.e., these two transfer functions have the same expected value when agents other than $i$ report their type truthfully. Thus, if agent $i$ finds it optimal to play truthfully under $t_i$ (when others do so), then she also finds it optimal under $t_i^{SM}$. Thus, transfers $t^{SM}$ elicit truthful revelation for every collection $\{\delta_i(\theta) : \theta \in \Theta\}$. Choosing this collection is the next question. In differentiable environments, supermodularity is characterized by positive cross-partial derivatives (Milgrom and Roberts [48]). If our environment were differentiable, we would have $\partial^2 t_i^{SM}/\partial\hat{\theta}_i\partial\hat{\theta}_j = \partial^2\delta_i/\partial\hat{\theta}_i\partial\hat{\theta}_j$. That is, $\delta_i$ controls the supermodularity, which is also true in our environment. In Section 4, we provide the formula for an appropriate $\delta_i$. Since an agent's utility is $V_i + t_i^{SM}$, the formula essentially adds enough supermodularity to compensate any effects coming from $V_i$, but not in any excess. When applied to this example, the formula and (1.1) output

| $t_1^{SM}(\hat{\theta})$ | $(1,2)$ | $(2,1)$ |
|---|---|---|
| $(1,2)$ | 1 | 1 |
| $(2,1)$ | -1.625 | 1.375 |

| $t_2^{SM}(\hat{\theta})$ | $(1,2)$ | $(2,1)$ |
|---|---|---|
| $(1,2)$ | -.25 | -1 |
| $(2,1)$ | -.25 | 1 |

which translates into the ex ante payoff matrix:

| EU | truthtelling | constant (1,2) | constant (2,1) | always lie |
|---|---|---|---|---|
| truthtelling | 2.44*; .13* | 1.06; 0 | 3.69*; -.5 | 2.31; -.63 |
| constant (1,2) | 2.38; $-.38$* | 3.25*;-.5 | 2.63; -1 | 3.5*; -1.13 |
| constant (2,1) | 2.38; -.13 | 0; -.25 | 3.63; .75* | 1.25; .63 |
| always lie | 2.31; -.63 | 2.19; -.75 | 2.56; .25* | 2.44; .13 |

This payoff matrix describes a supermodular game — assuming $(2,1) > (1,2)$ — in which truthtelling is the unique equilibrium. Supermodularity implies that, for every true type, each agent wants to make larger announcements (under the assumed order) if the other agent does so as well. This mechanism has nice properties: iterative deletion of strictly dominated strategies gives a unique prediction, truthtelling, and

12

all adaptive learning dynamics converge to the truthful equilibrium (wherever they are initiated) by Milgrom and Roberts [48]. The original instability problem is solved.

In Section 1.8.1 of the Appendix, we present another version of this example where the designer starts with transfers that create multiple equilibria. In this case, our transformation technique solves the multiple equilibrium problem, as truthtelling is the unique equilibrium of our supermodular mechanism.

## 1.3   Finite Supermodular Design: The Framework

Consider $n$ agents, each endowed with quasilinear preferences over a set of alternatives. The set of agents will be denoted by $N$. An alternative is a vector $(x, t) = (x_1, \ldots, x_n, t_1, \ldots, t_n)$, where $x_i \in X_i$ and $t_i \in \mathbb{R}$ for all $i$. In this environment, $x_i$ is interpreted as agent $i$'s allocation; $x \in X \equiv \prod_{i=1}^n X_i$ is the complete allocation profile; $t_i$ is the money transfer to $i$; and $t \in \mathbb{R}^n$ is the vector of transfers.

Each agent $i$ has a finite type space $\Theta_i$ with generic element $\theta_i$. The types of agents other than $i$ are denoted by $\theta_{-i} \in \Theta_{-i} \equiv \prod_{j \neq i} \Theta_j$, and $\theta \in \Theta \equiv \prod_{i \in N} \Theta_i$ denotes a full type profile. There are no restrictions on the nature of the type spaces: each $\Theta_i$ could be, for example, a subset of $\mathbb{R}$, $\mathbb{R}^n$, or any other finite collection of elements. Information is incomplete. There is a common prior with probability mass function $\phi$ on $\Theta$ known to the mechanism designer. Types are assumed to be independently distributed, and $\phi$ has full support.

A mechanism designer wishes to implement an allocation for each realization of types. This objective is represented by a decision rule $x : \Theta \mapsto (x_i(\theta))_{i=1}^n$. To this end, the designer sets up a transfer scheme $t_i : \Theta \to \mathbb{R}$ for each $i$. A mechanism is denoted by $\Gamma = (\{\Theta_i\}, (x, t))$. Agents are asked to announce a type, and from the vector of announced types, an allocation and a transfer accrue to each agent. The

13

pair $f = (x, t)$ is called a social choice function. We adopt the conventional notation where $\hat{\theta}_i$ is agent $i$'s announced type, $\hat{\theta}_{-i}$ is the vector of announced types of all agents but $i$, and $\hat{\theta}$ denotes the announced types of all agents.

Each agent $i$'s preferences over alternatives are represented by a utility function $u_i(x, t_i, \theta) = V_i(x; \theta) + t_i$, where $V_i : X \times \Theta \to \mathbb{R}$ is referred to as $i$'s valuation. This formulation allows for allocational externalities, as $V_i$ can depend on the allocations of agents other than $i$. It also captures the case of informational externalities (interdependent valuations) since the valuations may depend on other agents' types. Agent $i$'s utility function at type $\theta$ in $\Gamma$ is $u_i^\Gamma(\hat{\theta}; \theta) = V_i(x(\hat{\theta}); \theta) + t_i(\hat{\theta})$. A pure strategy for agent $i$ under incomplete information is a function $\hat{\theta}_i : \Theta_i \to \Theta_i$ that maps true types into announced types. Strategy $\hat{\theta}_i(\cdot)$ is called a deception. Agent $i$'s (ex ante) utility function in $\Gamma$ is $U_i^\Gamma(\hat{\theta}_i(.), \hat{\theta}_{-i}(.)) = E_\theta[u_i^\Gamma(\hat{\theta}(\theta); \theta)]$.

A partial order $\geq$ on a set $X$ is a binary relation that satisfies reflexivity, antisymmetry, and transitivity (see Topkis [56]). The couple $(X, \geq)$ is referred to as a partially ordered set. For $x, y \in X$, if $y \geq x$ and $y \neq x$, then we write $y > x$. A total order on set $X$ is a binary relation that satisfies comparability, antisymmetry, and transitivity.[6] If $\geq$ is a total order on $X$, then $(X, \geq)$ is called a totally ordered set. An interval in $(X, \geq)$ is a set of the form $[x', x''] = \{x \in X : x'' \geq x \geq x'\}$.

A total order $\geq^*$ on set $X$ is called a linear extension of the partial order $\geq$ if $(i)$ $(X, \geq^*)$ is a totally ordered set and $(ii)$ for every $x, y$ in $X$, if $y \geq x$, then $y \geq^* x$. Elements that are ordered under $\geq$ remain identically ordered under $\geq^*$, but $\geq^*$ also orders all the elements that are unordered under $\geq$. By the order-extension principle (Marczewski [43]), every partial order can be extended to a total order. Therefore, every partially ordered set admits a linear extension.

---

[6]Comparability means that $x \geq y$ or $y \geq x$ for all $x, y$ in $X$. Note that comparability implies reflexivity; hence, every total order is also a partial order.

Suppose that $(X, \geq_X)$ and $(Y, \geq_Y)$ are partially ordered sets. A function $h : X \times Y \to \mathbb{R}$ has increasing (decreasing) differences in $(x, y)$ on $X \times Y$ if for all $x'' \geq_X x'$ and all $y'' \geq_Y y'$, $h(x'', y'') - h(x', y'') \geq (\leq) h(x'', y') - h(x', y')$. Increasing (decreasing) differences express the notion of strategic complementarities (substitutes) when applied to payoff functions.

Take $x, x'$ in a partially ordered set $(X, \geq)$. If $x$ and $x'$ have a least upper bound (greatest lower bound) in $X$, it is referred to as their join (meet) and denoted by $x \vee x'$ ($x \wedge x'$). A lattice is a partially ordered set that contains the join and meet of every pair of its elements. Given a lattice $X$, a function $h : X \to \mathbb{R}$ is supermodular if $h(x) + h(x') \leq h(x \vee x') + h(x \wedge x')$ for all $x$ and $x'$ in $X$.

A finite game is a tuple $(N, \{(\Sigma_i, \geq_i)\}, \{w_i\})$ where $N$ is a finite set of players; $(\Sigma_i, \geq_i)$ is a finite partially ordered strategy set for each $i$; and $w_i : \prod_{i \in N} \Sigma_i \to \mathbb{R}$ is Player $i$'s payoff function. In the following definition, the set of strategy profiles of $i$'s opponents, denoted $\Sigma_{-i} = \prod_{j \neq i} \Sigma_j$, is endowed with the product order induced by $\{\geq_j\}_{j \neq i}$, according to which $\sigma'_{-i}$ is weakly larger than $\sigma_{-i}$ iff $\sigma'_j \geq_j \sigma_j$ for all $j \neq i$.

DEFINITION 1. *A finite game* $\mathcal{G} = (N, \{(\Sigma_i, \geq_i)\}, \{w_i\})$ *is supermodular if for all* $i \in N$, *(1)* $(\Sigma_i, \geq_i)$ *is a lattice, (2)* $w_i$ *has increasing differences in* $(\sigma_i, \sigma_{-i})$ *on* $(\Sigma_i, \Sigma_{-i})$, *and (3)* $w_i$ *is supermodular in* $\sigma_i$ *on* $\Sigma_i$ *for each* $\sigma_{-i} \in \Sigma_{-i}$.

The paper focuses on totally ordered strategy sets. In this case, requirements (1) and (3) are automatically satisfied and we only need to satisfy (2) to ensure that the game is supermodular.

We endow agents' type spaces with ordering relations and use them to define supermodular implementation. For all $i$, let $(\geq_i^1, \geq_i^2)$ be a pair of orders such that $\geq_i^1$ is a total order on $\Theta_i$ and $\geq_i^2$ is a total order on $\Theta_{-i}$. Denote by $\geq_{-i}$ the product order

induced by $\{\geq_j^1\}_{j \neq i}$ on $\Theta_{-i}$. A profile of orders $\{(\geq_i^1, \geq_i^2)\}$ is said to be *consistent* if for all $i$, $\geq_i^2$ is a linear extension of $\geq_{-i}$ on $\Theta_{-i}$.

The game induced by mechanism $\Gamma$ can be formulated at three stages: Ex ante, interim, and ex-post (complete information). Among these three formulations, the paper considers supermodularity at the ex-post level, because this is the strongest requirement. If the ex-post game is supermodular for all possible realizations of types $\theta$, then the game will be supermodular in its ex ante and interim formulations.

Let $\mathcal{G}(\theta) = (N, \{(\Theta_i, \geq_i^1)\}, \{u_i^\Gamma(\cdot\,; \theta)\})$ denote the game induced ex-post by mechanism $\Gamma$. Let $\mathcal{G} = (N, \{(\Theta_i^{\Theta_i}, \geq_i)\}, \{U_i^\Gamma\})$ be the ex ante Bayesian game induced by $\Gamma$, where $\geq_i$ is the pointwise order induced by $\geq_i^1$ on $\Theta_i^{\Theta_i}$. The next definition introduces the main implementation concept in the context of direct mechanisms.

DEFINITION **2.** *A social choice function $f = (x, t)$ is (truthfully) supermodular implementable if truthtelling, i.e., $\hat{\theta}_i(\theta_i) = \theta_i$ for all $i$, is a Bayesian equilibrium of $\mathcal{G}$ and if $\mathcal{G}(\theta)$ is supermodular for each $\theta$.*

## 1.4 Minimal Supermodular Implementation

In this section, we present two results dealing with minimally supermodular mechanisms. The main issue with supermodular implementation lies in finding the appropriate amount of complementarity to add to a mechanism. While complementarities lead to good dominance and learning properties, via the monotonicity of the best responses, excessive complementarities can generate multiple equilibria. Therefore, one negative consequence might be enhancing the "learnability" of undesirable equilibria. In our model, only the truthful equilibrium outcome is known to be socially desirable. Hence, adding complementarities can make it easier for agents to learn; but if this induces multiplicity, they may learn to play an untruthful equilibrium.

This section is organized as follows. First, we present the foundational results and concepts behind minimal supermodular implementation. Then we present our results in separate sections. Our first result is that, for *any* implementable social choice function, its decision rule can be minimally supermodular implemented by transfers within a class. This result holds for *all* valuation functions. This is a strong result, conditionally on choosing transfers within the class. Our second result does not restrict attention to a class of mechanisms or transfers. For *any* valuation functions, if a social choice function is implementable, then its decision rule can be minimally supermodular implemented (among *all* transfers that induce a supermodular game) if and only if a finite system of linear equations admits a solution. This solution, if it exists, is the set of minimally supermodular transfers. As a complementary result, we provide a sufficient condition, called order reducibility, under which existence of a solution is guaranteed. Both results provide explicit expressions for the transfers.

### 1.4.1 Foundations

Mathevet [45] relates the degree of complementarities to the size of the equilibrium set via the following binary relation.

DEFINITION 3. *The binary relation $\succeq_{ID}$ on the space of transfer functions is defined such that $\tilde{t} \succeq_{ID} t$ if for all $i \in N$ and for all $\theta_i'' \geq_i^1 \theta_i'$ and $\theta_{-i}'' \geq_{-i} \theta_{-i}'$, $\tilde{t}_i(\theta_i'', \theta_{-i}'') - \tilde{t}_i(\theta_i'', \theta_{-i}') - \tilde{t}_i(\theta_i', \theta_{-i}'') + \tilde{t}_i(\theta_i', \theta_{-i}') \geq t_i(\theta_i'', \theta_{-i}'') - t_i(\theta_i'', \theta_{-i}') - t_i(\theta_i', \theta_{-i}'') + t_i(\theta_i', \theta_{-i}')$.*

This binary relation orders transfers according to the magnitude of their increasing differences. In differentiable environments, this definition is equivalent to saying that $\tilde{t} \succeq_{ID} t$, if and only if, for all $i \in N$ the cross-partial derivatives of $\tilde{t}_i$ are larger than those of $t_i$, $\partial^2 \tilde{t}_i(\theta)/\partial \theta_i \partial \theta_j \geq \partial^2 t_i(\theta)/\partial \theta_i \partial \theta_j$, for all $j \in N$ and $\theta \in \Theta$. This definition captures the amount of complementarities contained in transfers and

compares them accordingly. While relation $\succeq_{\text{ID}}$ is transitive and reflexive, it is not antisymmetric. Denote the set of $\succeq_{\text{ID}}$-equivalence classes of transfers by $\mathcal{T}$.[7]

In a supermodular game, the interval prediction is the interval between the largest and the smallest equilibrium. We compare supermodular mechanisms by the size of the interval prediction of the game that they induce. The next proposition, taken from Mathevet [45], provides the tool to do so. If transfers $t''$ generate more complementarities than $t'$, and if both transfers induce truthtelling and a supermodular game, then the interval prediction induced by $t''$ contains that induced by $t'$.

For any $t \in \mathcal{T}$ such that $f = (x, t)$ is supermodular implementable, let $\overline{\theta}_t^i : \Theta_i \to \Theta_i$ and $\underline{\theta}_t^i : \Theta_i \to \Theta_i$ denote the largest and the smallest equilibrium strategy of Player $i$ in the game induced by the mechanism. Define

$$[\underline{\theta}_t, \overline{\theta}_t] = \{(s_1, \ldots, s_n) | s_i : \Theta_i \to \Theta_i \text{ and}$$
$$\overline{\theta}_t^i(\theta_i) \geq_i^1 s_i(\theta_i) \geq_i^1 \underline{\theta}_t^i(\theta_i) \text{ for all } i \in N \text{ and } \theta_i \in \Theta_i\} \quad (1.2)$$

to be the interval of strategy profiles in between the extremal equilibria.

PROPOSITION 1. *If $(x, t'')$ and $(x, t')$ are supermodular implementable social choice functions and if $t'' \succeq_{ID} t'$, then $[\underline{\theta}_{t'}, \overline{\theta}_{t'}] \subset [\underline{\theta}_{t''}, \overline{\theta}_{t''}]$.*

This proposition is proved in Mathevet [45]. It implies that the objective of minimizing the equilibrium set coincides with the objective of minimizing the complementarities. A social choice function $f = (x, t^*)$ will be minimally supermodular implementable if the transfers $t^*$ elicit truthful revelation and induce a supermodular game with the weakest complementarities. This will give the tightest interval prediction around the truthful equilibrium.

---

[7]Each equivalence class contains transfer functions $t$ and $\tilde{t}$ such that $\tilde{t} \succeq_{\text{ID}} t$ and $t \succeq_{\text{ID}} \tilde{t}$ while $t \neq \tilde{t}$. Any quasi-order can be transformed into a partial order by using equivalence classes.

### 1.4.2 Minimal Implementation under Total Orders

This section addresses minimal supermodular implementation within a class of transfers. We explicitly show how to convert any truthful mechanism into a supermodular mechanism while controlling for the intensity of the complementarities.

Our approach takes advantage of the totality of orders $\geq_i^1$. If the strategy sets are totally ordered, then the only requirement to check to satisfy Definition 1 is the increasing differences condition. Therefore, if the transfer functions ensure that (I) for each $\theta$ and $i$, $u_i^\Gamma(\hat\theta, \theta)$ has increasing differences in $(\hat\theta_i, \hat\theta_{-i})$ on $(\Theta_i, \geq_i^1) \times (\Theta_{-i}, \geq_{-i})$, then (II) the ex-post game $\mathcal{G}(\theta)$ will be supermodular for each $\theta$, as desired. In this section, we restrict attention to the class of transfers that guarantee that (III) for each $\theta$ and $i$, $u_i^\Gamma(\hat\theta, \theta)$ has increasing differences in $(\hat\theta_i, \hat\theta_{-i})$ on $(\Theta_i, \geq_i^1) \times (\Theta_{-i}, \geq_i^2)$, where $\{(\geq_i^1, \geq_i^2)\}_i$ is a consistent profile of orders. Since $\geq_i^2$ is a linear extension of $\geq_{-i}$, (III) implies (I), and hence (II) holds. Consider the following family of transfers:

DEFINITION 4. *Family $\mathcal{F}(x, \{(\geq_i^1, \geq_i^2)\}_i) \subset \mathcal{T}$ is the set of transfers $t$ such that $(x, t)$ is truthfully implementable and $u_i^\Gamma(\hat\theta, \theta)$ has increasing differences on $(\Theta_i, \geq_i^1) \times (\Theta_{-i}, \geq_i^2)$ for each $\theta \in \Theta$ and $i \in N$, where $\{(\geq_i^1, \geq_i^2)\}_i$ is consistent.*

We now define our concept of minimal supermodular implementation.

DEFINITION 5. *A social choice function $f = (x, t^*)$ is minimally supermodular implementable over family $\mathcal{F}$ if it is supermodular implementable, $t^* \in \mathcal{F}$, and $t \succeq_{ID} t^*$ for all transfers $t \in \mathcal{F}$.*

Minimally supermodular transfers elicit truthful revelation and produce the supermodular game with the weakest complementarities. By Proposition 1, they give the tightest equilibrium set within the class of transfers $\mathcal{F}$. Here is our first main result.

THEOREM 1. *If $f = (x, t)$ is implementable, then for any consistent profile of orders $\{(\geq_i^1, \geq_i^2)\}_i$ there exist $t^*$ such that $(x, t^*)$ is minimally supermodular implementable over $\mathcal{F}(x, \{(\geq_i^1, \geq_i^2)\}_i)$.*

The theorem reaches a strong conclusion: for *any* implementable social choice function, its decision rule can be minimally supermodular implemented. There are no other restrictions on the decision rule or the valuation functions. Despite the finiteness of the type sets, there are infinitely many transfers that can supermodularly implement a decision rule for a given consistent profile of orders. Having a method for choosing the best among them is useful. To understand this, as well as our construction, start from any truth-revealing transfers $\{t_i(\theta) : \theta \in \Theta\}$. For each $i$ and collection of numbers $\{\delta_i(\theta) : \theta \in \Theta\}$, define

$$t_i^*(\theta_i, \theta_{-i}) = \delta_i(\theta_i, \theta_{-i}) - E_{\theta_{-i}}[\delta_i(\theta_i, \theta_{-i})] + E_{\theta_{-i}}[t_i(\theta_i, \theta_{-i})]. \tag{1.3}$$

Transfers $t_i^*(\theta)$ satisfy $E_{\theta_{-i}}[t_i^*(\theta_i, \theta_{-i})] = E_{\theta_{-i}}[t_i(\theta_i, \theta_{-i})]$ for all $\theta_i$, i.e., these two transfer functions have the same expected value when agents other than $i$ report their type truthfully. Since agent $i$ finds it optimal to play truthfully under $t_i(\theta)$ (when others do so), she must also find it optimal under $t_i^*(\theta)$. We conclude that for every collection $\{\delta_i(\theta) : \theta \in \Theta\}$, the transfers $t_i^*(\theta)$ also elicit truthful revelation. The problem becomes the choice of each $\delta_i(\theta)$, as there are infinitely many ways of inducing a supermodular game given a profile of orders. The proof provides an explicit formula for the collection $\{\delta_i(\theta) : \theta \in \Theta\}$ that generates the best transfers $\{t_i^*\}$ (within a family of transfers) in terms of minimizing the interval prediction.

To sum up, our method suggests totally ordering type sets and then using our formula. Can this method be useful? In Section 1.2, it delivered a supermodular mechanism with a unique equilibrium, while ex-post implementation was not an op-

tion. In the Appendix (see Section 1.8.1), it also delivers a supermodular mechanism with a unique equilibrium, while the original transfers produce multiple equilibria.

Given a choice of consistent orders, the theorem provides appropriate transfers. But there are many possible orders and the designer may want to discriminate among the many associated transfers. Suppose that the designer has a concept of distance, i.e., a metric $d$ on $\Theta$. Then Theorem 1 can be used to select the transfers that lead to the smallest equilibrium set across *all* the families. Let $\mathcal{F}^*(x)$ be the union of $\mathcal{F}(x, \{(\geq_i^1, \geq_i^2)\}_i)$ over all consistent orders $\{(\geq_i^1, \geq_i^2)\}_i$.

COROLLARY **1.** *If $f = (x, t)$ is implementable, then there exist transfers $t^{**}$ and consistent orders $\{(\geq_i^{*1}, \geq_i^{*2})\}_i$ such that $(x, t^{**})$ is minimally supermodular implementable over $\mathcal{F}(x, \{(\geq_i^{*1}, \geq_i^{*2})\}_i)$ and $t^{**}$ give the smallest interval prediction in $\mathcal{F}^*(x)$ given $d$.*

Our corollary ultimately says that for every metric, there is a choice of total orders $(\geq_i^{*1}, \geq_i^{*2})$ for each $i$ that is most adapted to $d$, since the equilibrium set resulting from the corresponding minimal transfers is minimized (under $d$) among all of $\mathcal{F}^*(x)$. The explanation is simple. For each profile of orders, the theorem provides the transfers that deliver the smallest interval prediction within the corresponding class. Since there are finitely many types, there are finitely many (consistent) profiles of orders. Therefore, there must be a profile of orders whose associated transfers deliver the smallest interval prediction under $d$ among all of $\mathcal{F}^*(x)$.

### 1.4.3 Unconditional Minimal Implementation

In this section, we study (unconditional) minimal supermodular implementation by looking for the overall best transfers. In the previous section, the supermodular transfers were minimal within a class. We required that, for every agent

21

$i$, increasing differences be satisfied on $(\Theta_i, \geq^1_i) \times (\Theta_{-i}, \geq^2_i)$. By doing so, we did not consider all the transfers that induce a supermodular game. Indeed, some transfers may induce increasing differences on $(\Theta_i, \geq^1_i) \times (\Theta_{-i}, \geq_{-i})$ but not on the above product set, yet it is sufficient for our purpose. This happens because $\geq_{-i}$ typically orders fewer elements than $\geq^2_i$, which changes the number of inequalities that have to hold to satisfy increasing differences. To summarize, our previous theorem was a conditional form of minimal supermodular implementation, while in this section, we aim for an unconditional form. For convenience, we write $V_i(x, \theta) = V_i(x_i, \theta)$ for all $i$ to emphasize the dimension of the decision rule on which $i$'s utility depends. This notation does not exclude allocative externalities, for an agent's own allocation $x_i$ could be a function of another agent's allocation.

DEFINITION 6. *A social choice function $f = (x, t^*)$ is minimally supermodular implementable if it is minimally supermodular implementable over family $\mathcal{T}$.*

Our main result shows that the problem of finding minimally supermodular transfers is equivalent to solving a system of linear equations. This insight is highly useful, as it allows the application of standard methods and algorithms from numerical linear algebra (e.g. Paige and Saunders [51], Demmel [24]). In what follows, we refer to the *supermodularity* of a function $h_i : \Theta \to \mathbb{R}$ between types $\theta''_i \geq^1_i \theta'_i$ and $\theta''_{-i} \geq_{-i} \theta'_{-i}$ as the expression

$$h_i(\theta''_i, \theta''_{-i}) - h_i(\theta''_i, \theta'_{-i}) - h_i(\theta'_i, \theta''_{-i}) + h_i(\theta'_i, \theta'_{-i}). \tag{1.4}$$

Consider (1.3) and note that the supermodularity of $t^*_i$ is equal to the supermodularity of $\delta_i$. Therefore, our objective of finding the overall best transfers is tantamount to finding a function $\delta_i$ that induces increasing differences on $(\Theta_i, \geq^1_i) \times (\Theta_{-i}, \geq_{-i})$ for each $i$, without introducing unnecessary complementarities. Before deriving the

linear system that corresponds to this objective, we define the concept of immediate successor/predecessor.

DEFINITION **7.** *For $x'$ and $x''$ in a partially ordered set $(X, \geq_X)$, $x''$ is an immediate successor of $x'$ (and $x'$ is an immediate predecessor of $x''$) if (a) $x'' >_X x'$, and (b) the set $\{x \in X | x'' >_X x >_X x'\}$ is empty.*

Consider the system of linear equations $A_i \cdot \delta_i = b_i$, where $\delta_i$ is a column vector that contains the values of $\delta_i(\theta)$ for every $\theta \in \Theta$; $A_i$ is a sparse matrix whose nonzero elements (four per row) are equal to -1 or 1, and positioned so as to produce the supermodularity of $\delta_i$ for types that are immediate successors; $b_i$ is a vector containing expressions (1.5), i.e., the minima of the differences in valuations between immediate successors. For example, a typical row of $A_i$ takes the form $(\mathbf{0}, 1, \mathbf{0}, -1, \mathbf{0}, -1, \mathbf{0}, 1, \mathbf{0})$ where $\mathbf{0}$ is a block of zeroes, so that the dot product with vector $\delta_i$ produces an expression such as (1.4). The system matches this expression with an entry of $b_i$ that involves the same types. In Section 1.8.2, we derive this system in a particular example.

The next proposition characterizes the (unconditional) minimally supermodular transfers as a solution to the above system of linear equations.

PROPOSITION **2.** *Minimally supermodular transfers exist, if and only if, the finite linear system $A_i \cdot \delta_i = b_i$ has a solution $\delta_i \geq 0$ for all player $i$.*

Before providing the intuition for the result, we make a few remarks.

Assuming $\delta_i \geq 0$ is without loss of generality, because we can always add any positive constant $c$ to any solution $\delta_i$ and obtain another solution. The reason is that any constant gets canceled out when we form the supermodularity of a function. Moreover, assuming $b_i \geq 0$ is also without loss of generality, since any equation

23

with a negative right hand side can be multipled by -1 (i.e. the relevant rows of $A_i$ and $b_i$ get multiplied by -1). We can use linear programming (LP) techniques and software to solve the feasibility problem implied by the system of equations stated in Proposition 2. By introducing a vector $z_i$ of positive auxiliary variables, we can restate the feasibility problem into a standard form LP:

$$\min_{\delta_i, z_i} e^T \cdot z_i$$
$$\text{s.t.} \ \ A_i \cdot \delta_i + z_i = b_i$$
$$\delta_i \geq 0, z_i \geq 0$$

where $e$ is a vector of 1s of appropriate dimension. This auxilary LP has optimal value 0 (i.e. $z = 0$) if and only if there exists a non-negative solution $\delta_i$ to the system of linear equations stated in the proposition.

Besides linearity, the main virtue of Proposition 2 is to only involve immediate successors. The necessity part of this result is intuitive. For all $\theta'_i$, $\theta''_i$ where $\theta''_i$ is an immediate successor of $\theta'_i$ in $\Theta_i$, and for all $\theta'_{-i}$, $\theta''_{-i}$, where $\theta''_{-i}$ is an immediate successor of $\theta'_{-i}$ in $\Theta_{-i}$, the supermodularity of $\delta_i$ (which corresponds to that of $t_i^*$) must equal

$$- \min_{\theta \in \Theta} [V_i(x_i(\theta''_i, \theta''_{-i}), \theta) - V_i(x_i(\theta''_i, \theta'_{-i}), \theta) - V_i(x_i(\theta'_i, \theta''_{-i}), \theta) + V_i(x_i(\theta'_i, \theta'_{-i}), \theta)],$$
$$(1.5)$$

for every $i$, for otherwise we could construct alternative transfers $\tilde{t}_i$ that satisfy this equality (at particular types); therefore, it could not hold that $\tilde{t} \succeq_{\text{ID}} t^*$, and hence no $t^*$ could be minimally supermodular.

While necessity seems clear, it is not obvious that it suffices to search for $\delta_i$'s whose supermodularity equals (1.5) for successive types only. This property, which

24

greatly simplifies the problem, comes from the proof of Theorem 1: the supermodularity of any function of two variables, when measured between non-successive elements, is equal to the sum of the supermodularities between all pairs of immediate successors in between. The intuition goes as follows. Take a function $h$ with two variables, where each variable is in $\mathbb{N}$. Note that

$$h(2,3) - h(2,1) - (h(1,3) - h(1,1)) \tag{1.6}$$

is equal to

$$[h(2,3) - h(2,2) - (h(1,3) - h(1,2))] + [h(2,2) - h(2,1) - (h(1,2) - h(1,1))]. \tag{1.7}$$

The differences between non-successive types (1 and 3 are not immediate successors in (1.6)) are sums of differences between successive types, (1.7). Therefore, if the supermodularity of $\delta_i$ between successive types equals (1.5), which is the minimal requirement, then our previous observation implies that the supermodularity of $\delta_i$ between non-successive types must also be minimal. In conclusion, we just need to be concerned with supermodularity between successive types.

In Section 1.8.4, we provide a sufficient condition that ensures minimal transfers exist and have a simple closed-form representation. Although this condition is not necessary, it appears to be tight. The condition, called *order reducibility*, is imposed on the set of decision rules.

## 1.5 Uniqueness

In this section, we provide sufficient conditions for supermodular implementation in unique equilibrium. In light of our current results, a natural question to ask is: when does a minimally supermodular mechanism, i.e., one with the smallest equilibrium set, actually have a unique equilibrium? If a supermodular game has a unique

equilibrium, then it is dominance-solvable, and many learning dynamics converge to the unique equilibrium (Milgrom and Roberts [48]). Supermodular implementation is, therefore, particularly appealing when truthtelling is the unique equilibrium. The study of unique supermodular implementation allows us to draw some conclusions regarding the type of environments — preferences and social choice functions — for which supermodular implementation may be most useful.

In this section, we impose more structure on the type sets. Suppose $(\Theta_i, d_i)$ is a metric space for every $i$. Our only requirement is that for any $i$, if $\theta_i' \geq_i^1 \theta_i$ and if $\theta_i'' \geq_i^1 \theta_i'$, then $d_i(\theta_i'', \theta_i) \geq d_i(\theta_i', \theta_i)$, so that each metric respects the order.

Recall that $i$'s utility at type $\theta$ is denoted by $u_i^\Gamma(\hat{\theta}; \theta) = V_i(x(\hat{\theta}); \theta) + t_i(\hat{\theta})$. For each $i$ and $\theta$, let $K_i(\theta)$ be a real number such that

$$
(u_i^\Gamma(\theta_i'', \theta_{-i}''; \theta) - u_i^\Gamma(\theta_i', \theta_{-i}''; \theta)) - (u_i^\Gamma(\theta_i'', \theta_{-i}'; \theta) - u_i^\Gamma(\theta_i', \theta_{-i}'; \theta))
$$
$$
\leq d_i(\theta_i'', \theta_i') K_i(\theta) \sum_{j \neq i} d_j(\theta_j'', \theta_j') \quad (1.8)
$$

for all $\theta_i'' \geq_i^1 \theta_i'$ and $\theta_{-i}'' \geq_{-i} \theta_{-i}'$.[8] Due to the finiteness of types, $K_i$ always exists. When types are real-valued and $d_i$ is the Euclidean metric, (1.8) can be written as

$$
\frac{u_i^\Gamma(\theta_i'', \theta_{-i}''; \theta) - u_i^\Gamma(\theta_i', \theta_{-i}''; \theta)) - (u_i^\Gamma(\theta_i'', \theta_{-i}'; \theta) - u_i^\Gamma(\theta_i', \theta_{-i}'; \theta))}{(\theta_i'' - \theta_i') \sum_{j \neq i} (\theta_j'' - \theta_j)} \leq K_i(\theta), \quad (1.9)
$$

and so if the environment is differentiable, $K_i(\theta) = \max_{j \neq i} \max_{\hat{\theta}} \partial^2 u_i^\Gamma(\hat{\theta}; \theta) / \partial \hat{\theta}_i \partial \hat{\theta}_j$.[9] The cross-partial derivatives measure the strategic complementarities between agents' reports. Therefore, $K_i(\theta)$ is an upper bound on the strategic complementarities (between agents' reports) induced by $\Gamma$ at $\theta$, and thus it is a nonnegative number. Note that $K_i$ is an *endogenous* quantity, as it depends on the transfers.

---

[8]In (1.8), we use the $L_1$-norm induced by $\{d_i\}$ to measure the distance between opponents' profiles.

[9]Let $\theta_k'' = \theta_k'$ for $k \neq i, j$. If $\theta_i'' \to \theta_i'$ and $\theta_j'' \to \theta_j'$, the lhs of (1.9) becomes $\partial^2 u_i^\Gamma(\theta'; \theta) / \partial \hat{\theta}_i \partial \hat{\theta}_j$.

Similarly, for each $i$ and $\theta_{-i}$, let $\gamma_i(\theta_{-i})$ be a real number such that

$$(V_i(x(\hat{\theta}_i'', \theta_{-i}); \theta_i'', \theta_{-i}) - V_i(x(\hat{\theta}_i', \theta_{-i}); \theta_i'', \theta_{-i})) - (V_i(x(\hat{\theta}_i'', \theta_{-i}); \theta_i', \theta_{-i})$$
$$- V_i(x(\hat{\theta}_i', \theta_{-i}); \theta_i', \theta_{-i})) \geq \gamma_i(\theta_{-i}) d_i(\hat{\theta}_i'', \hat{\theta}_i') d_i(\theta_i'', \theta_i') \quad (1.10)$$

for all $\hat{\theta}_i'' \geq_i^1 \hat{\theta}_i'$ and $\theta_i'' \geq_i^1 \theta_i'$. Due to the finiteness of types, $\gamma_i$ always exists. In differentiable environments with real-valued types, $\gamma_i(\theta_{-i}) = \min_{(\theta_i, \hat{\theta}_i)} \partial^2 V_i(x(\hat{\theta}_i, \theta_{-i}); \theta)/\partial\hat{\theta}_i\partial\theta_i$. Therefore, $\gamma_i$ is a lower bound on the complementarities between $i$'s own report and type when other agents report truthfully. Note that $\gamma_i$ is an *exogenous* quantity, because it is determined by the primitives of the problem. Examples 3 and 4 in Mathevet [45] provide numerical illustrations for the computation of the $K_i$'s and $\gamma_i$'s.

Numbers $K_i$ and $\gamma_i$ represent opposite forces in the shaping of the equilibrium set. An agent with a large $\gamma_i$ is very responsive to her own type, because her marginal utility is very sensitive to a change in $\theta_i$. Thus, small changes in type lead to large changes in report, independently of others' reports, which have relatively little importance. Therefore, when $\gamma_i$ is large, $i$'s behavior is not responsive to others. By definition, equilibrium multiplicity is caused by the mutual influence that players have on one another. Since a large $\gamma_i$ isolates $i$ from the other agents, this favors uniqueness. A large $K_i$, however, expresses strong interdependence between players' reports. If $i$'s behavior is strongly responsive to others, this favors multiplicity. These effects and their impact on the set of rationalizable strategy profiles in incomplete information games are the subject of Mathevet [46].

Our next results formalize the trade-off between these forces. Denote the truthful strategy by $\theta_i^T(\cdot)$. For each $i$ and $\theta_i$, let $\bar{K}_i(\theta_i) = \max_{\theta_{-i}} K_i(\theta)$.

PROPOSITION 3. *Let $f$ be a supermodular implementable social choice function. For every $i$, consider some deception $\theta_i^*(\cdot) \geq \theta_i^T(\cdot)$. If there exist $i$, $\theta_i$ and $\hat{\theta}_i \in [\theta_i, \theta_i^*(\theta_i))$*

*such that*

$$\bar{K}_i(\theta_i) \sum_{j \neq i} E_{\theta_j}[d_j(\theta_j^*(\theta_j), \theta_j] - E_{\theta_{-i}}[\gamma_i(\theta_{-i})]d_i(\hat{\theta}_i, \theta_i) < 0 \qquad (1.11)$$

*then $\theta^*(\cdot)$ is not a Bayesian equilibrium. The same conclusion applies to deceptions $\theta_i^*(\cdot) \leq \theta_i^T(\cdot)$ if there exist $i$, $\theta_i$ and $\hat{\theta}_i \in (\theta_i^*(\theta_i), \theta_i]$ such that (1.11) holds.*

The proof demonstrates that if agent $i$ deviates from his report at type $\theta_i$, $\theta_i^*(\theta_i)$, and announces $\hat{\theta}_i$ instead, while other agents play according to $\theta_{-i}^*(\cdot)$, then the lhs of (1.11) is an upper bound on how much $i$ *loses* from that deviation. If this upper bound is positive, then the deviation may not be profitable. However, if it is negative, then $\hat{\theta}_i$ is a profitable deviation and hence $\theta^*(\cdot)$ cannot be an equilibrium.[10]

Proposition 3 should read as follows: if there exist metrics $d_i$ and constants $K_i(\theta)$ and $\gamma_i(\theta_{-i})$ satisfying (1.8) and (1.10), and if the hypotheses of Proposition 3 are satisfied, then $\theta^*(\cdot)$ is not a Bayesian equilibrium. Of course, agents need not be aware of which metrics are used by the designer. In the same way, agents need not be aware of which order is used by the designer to induce a supermodular game.

Inequality (1.11) summarizes the trade-off between the opposite forces $\bar{K}_i$ and $E[\gamma_i(\cdot)]$. If the uniqueness effect dominates, i.e., $E[\gamma_i(\cdot)]$ is large enough, then the untruthful profile $\theta^*(\cdot)$ does not fall within the bounds of the interval prediction.[11]

The proposition is not useful for profiles for which, for every $i$ and $\theta_i$, $\theta_i^*(\theta_i)$ and $\theta_i$ are equal or successive types. In that case, $d_i(\hat{\theta}_i, \theta_i)$ would be zero. Furthermore, although the theorem is useful to determine whether a given strategy profile is not an equilibrium, it does not deliver an immediate conclusion about uniqueness. The next proposition addresses this question.

---

[10]It is clear what losing 2 means, for example. But losing -2 is equivalent to gaining 2. Thus, when (1.11) holds, the loss must be negative and so the deviation profitable.

[11]Using the smallest possible $\bar{K}_i(\theta_i)$ and the largest possible $\gamma_i(\theta_{-i})$ is a natural way for the designer to utilize this proposition.

Before proceeding, we define a measure of coarseness on agents' type spaces. For any type $\theta_i$ in $\Theta_i$, letting $\theta_i'$ and $\theta_i''$ be its immediate predecessor and immediate successor, we define

$$\varepsilon_i(\Theta_i) = \max_{\theta_i \in \Theta_i} \max\{d_i(\theta_i', \theta_i), d_i(\theta_i, \theta_i'')\}$$

to be a measure of the maximal distance between any type in $\Theta_i$ and its immediate successor or predecessor. As we get closer to the continuous case, $\varepsilon_i(\Theta_i) \to 0$. Define $\varepsilon(\Theta) = \max_i \varepsilon_i(\Theta_i)$ to be the overall measure of coarseness.

Our next result will be concerned with "eventual uniqueness," which first requires to define the neighborhood of truthtelling.

DEFINITION 8. *A profile $\theta^*(\cdot)$ is outside the neighborhood of truthtelling if $\theta^*(\cdot)$ and $\theta^T(\cdot)$ are ordered and if for all $i$, there is $\theta_i$ such that $(\theta_i \wedge \theta_i^*(\theta_i), \theta_i \vee \theta_i^*(\theta_i)) \neq \emptyset$.*

A profile is outside the neighborhood of truthtelling if it is larger or smaller than truthtelling (i.e., all agents always over- or under-report), and if all agents $i$ have the option for some $\theta_i$ to report a non-truthful type in between truth $\theta_i$ and her actual report $\theta_i^*(\theta_i)$. In order for an agent to have this option, her original deception must be far enough from truthtelling, for otherwise the only possible deviation would be to report her type truthfully.

PROPOSITION 4. *Let $f$ be a supermodular implementable social choice function (on $\Theta$). If for every agent $i$*

$$(n-1)E_{\theta_i}[\bar{K}_i(\theta_i)] < E_{\theta_{-i}}[\gamma_i(\theta_{-i})], \tag{1.12}$$

*then any profile $\theta^*(\cdot)$ outside the neighborhood of truthtelling such that for some $\theta$*

$$\min_i \left\{ \sum_{j \neq i} E_{\theta_j}[d_j(\theta_j, \theta_j^*(\theta_j))] \right\} > \varepsilon(\Theta)\xi(\theta) \tag{1.13}$$

29

*where*

$$\xi(\theta) = \frac{\max_i\{E_{\theta_{-i}}[\gamma_i(\theta_{-i})]\}}{\min_i\left\{\frac{E_{\theta_{-i}}[\gamma_i(\theta_{-i})]}{n-1} - E_{\theta_i}[K_i(\theta_i)]\right\}}, \tag{1.14}$$

*is not a Bayesian equilibrium.*

This proposition says that if (1.12) holds, then the size of the equilibrium set depends essentially on the richness of the type sets. Indeed, if function $\xi$ in (1.14) is bounded above as $\varepsilon(\Theta) \to 0$,[12] then $\varepsilon(\Theta)\xi(\theta) \to 0$. Thus, the set of profiles that include truthtelling and that might be equilibria vanishes as type sets become infinitely fine, and hence truthtelling eventually becomes unique. To be clear, condition (1.12) captures the main driving force behind uniqueness, but we must take the fineness of type sets into account. Otherwise, some untruthful profiles can become equilibria simply because some deviations are not available to an agent who would have otherwise chosen it. Note also that if (1.13) holds for some profile $\theta^*(\cdot)$, then it also holds for all finer type sets and all profiles $\theta^{**}(\cdot) \geq \theta^*(\cdot)$, assuming $\xi$ is stable. Therefore, as $\varepsilon(\Theta) \to 0$, the set of potential equilibria shrinks monotonically to zero measure.

This proposition generalizes Mathevet [45]'s uniqueness result (Proposition 3, p.418) to our environments. In continuous type spaces, richness is obviously not an issue — (1.13) holds automatically when $\varepsilon(\Theta) = 0$ — and only (1.12) matters.

This proposition seems to be mostly useful a posteriori. After the mechanism has been built, we can use it to check whether truthtelling is eventually unique. However, it would be useful to know *a priori* whether the design problem at hand is compatible with unique supermodular implementation given its primitives.

Since the minimally supermodular transfers minimize the size of the equilibrium set, they are a natural choice for unique implementation. We have a closed

---

[12]Note that $K_i$ and $\gamma_i$ depend on the type sets for every $i$. If the utility functions and the transfers are well-behaved, then $K_i$ and $\gamma_i$ exist in the limit as $\varepsilon(\Theta) \to 0$.

form expression for these transfers when the decision rule satisfies order reducibility (See Section 1.8.4 in the Appendix). Assuming order reducibility, the minimally supermodular transfers are constructed from two equations: (1.62) and (1.63) (p.52 in the Appendix). The critical observation is that these transfers, and hence the $K_i$'s they produce, depend entirely on the primitives of the model. In other words, the minimally supermodular transfers endogenize the $K_i$'s. This property is very useful, because (1.12) becomes a condition that only involves the primitives of the design problem. As such, we can check it before building the mechanism and determine whether unique supermodular implementation might be attainable based on the primitives. By doing so, we learn valuable information about the type of environments for which supermodular implementation may be most useful. We explain this below.

For any implementable social choice function $f$, let

$$K_i^*(\theta) = \max_{\{\theta_i'', \theta_i', \theta_{-i}'', \theta_{-i}'\}} \frac{V_i(\theta_i' \triangleright \theta_i'', \theta_{-i}''; \theta) - V_i(\theta_i' \triangleright \theta_i'', \theta_{-i}'; \theta) - H_i(\theta_i'', \theta_i', \theta_{-i}'', \theta_{-i}')}{d_i(\theta_i'', \theta_i') \sum_{j \neq i} d_j(\theta_j'', \theta_j')} \quad (1.15)$$

where $V_i(\theta_i' \triangleright \theta_i'', \cdot; \theta) = V_i(x_i(\theta_i'', \cdot); \theta) - V_i(x_i(\theta_i', \cdot); \theta)$ and $H_i(\theta_i'', \theta_i', \theta_{-i}'', \theta_{-i}')$ is the sum of elements

$$\min_{\theta \in \Theta} [V_i(\hat{\theta}_i' \triangleright \hat{\theta}_i'', \hat{\theta}_{-i}''; \theta) - V_i(\hat{\theta}_i' \triangleright \hat{\theta}_i'', \hat{\theta}_{-i}'; \theta)]$$

for all pairs of immediate successors $\hat{\theta}_i'' \geq_i^1 \hat{\theta}_i'$, such that $\theta_i'' \geq_i^1 \hat{\theta}_i'' \geq_i^1 \hat{\theta}_i' \geq_i^1 \theta_i'$, and all pairs of immediate successors $\hat{\theta}_{-i}'' \geq_{-i} \hat{\theta}_{-i}'$ for a chosen "path" between $\theta_{-i}'$ and $\theta_{-i}''$, such that $\theta_{-i}'' \geq_{-i} \hat{\theta}_{-i}'' \geq_{-i} \hat{\theta}_{-i}' \geq_{-i} \theta_{-i}'$. Notice that $\geq_i^1$ is a total order, which means that there is only one path of immediate successors connecting $\theta_i'$ and $\theta_i''$. In contrast, $\geq_{-i}$ is a partial order and there could be many paths of immediate successors that connect any two ordered elements $\theta_{-i}'$ and $\theta_{-i}''$. However, order reducibility ensures that irrespective of the path being chosen, the value of $H_i$ will be the same.

Under the minimally supermodular transfers, $K_i^*(\theta)$ is the smallest number that satisfies (1.8). That is, $\bar{K}_i^*(\theta_i) \equiv \max_{\theta_{-i}} K_i^*(\theta)$ bounds the strategic complementarities in the game induced by these transfers at $\theta$. When the designer uses the minimally supermodular transfers, $\bar{K}_i^*$ is the value that appears in condition (1.12). In this case, notice that (1.12) only involves the primitives of the model. If this inequality holds, then supermodular implementation is particularly well-suited for the design problem at hand, because the minimal transfers supermodularly implement the social choice function and truthtelling is eventually unique.

Expression (1.15) has a nice interpretation. It measures how much the supermodularity of the valuations varies across true types. We know that the designer must induce a supermodular game for any realization of types.[13] In this context, a large $\bar{K}_i^*$ can be caused by valuation functions that exhibit large substitutes for some (true) types (say $\theta$) and large complementarities for other (true) types (say $\theta'$). Since the designer does not know the realization of the true types, she will need to add a lot of complementarities through the transfers to ensure that the game is supermodular at $\theta$. But this may induce a game that is "too supermodular" at $\theta'$, since there are already enough complementarities at that type, thereby violating (1.12) and causing multiplicity.

## 1.6   Discussion

In our motivating example, the social choice function is not ex post implementable and yet it is Bayesian implementable. Therefore, the question regarding

---

[13]It is sufficient but not necessary that the ex-post game be supermodular for each realization in order for the ex ante Bayesian game to be supermodular. For example, if the prior is mostly concentrated on some subset $\Theta'$ of $\Theta$, it may not be necessary to make the ex-post payoffs supermodular for types in $\Theta \backslash \Theta'$. Of course, the possibility of neglecting $\Theta \backslash \Theta'$ depends on how unlikely that set is compared to how submodular the utility function may be for types in that set.

how generic this situation is, is an important one. We address it in this section.

Consider a generalized version of our motivating example. There are two players, each with two possible types, and two alternatives (X and Y). Small letters are used to denote 1's valuations (row player) and capital letters to denote 2's valuations (column player); upper rows in each cell represent valuations for alternative X and lower rows represent valuations for alternative Y. The alternative in the middle of each cell is the one that the designer would like to implement. The associated transfers are stated next to it.

| $V(\theta, \tau)$ | $(\tau_1, \tau_2)$ | | $(\tau_3, \tau_4)$ | |
|---|---|---|---|---|
| | $a$ | $A$ | $c$ | $C$ |
| $(\theta_1, \theta_2)$ | X, $t^1$ | | X, $t^2$ | |
| | $b$ | $B$ | $d$ | $D$ |
| | $e$ | $E$ | $g$ | $G$ |
| $(\theta_3, \theta_4)$ | Y, $t^3$ | | X, $t^4$ | |
| | $f$ | $F$ | $h$ | $H$ |

The smallest departure from a trivial (or constant) decision rule is the above. The designer would like to implement alternative X for all but one of the type profiles. The minimal conditions that need to hold for this decision rule to *not* be ex post implementable is that for at least one of the two agents, and for at least one type of his opponent, there are no transfers that achieve ex post incentive compatibility (EPIC). Without loss of generality, consider player 1 and opponent type $(\tau_1, \tau_2)$. There are two EPIC constraints, one for each possible type of player 1:

for type $(\theta_1, \theta_2)$

$$a + t_1^1 \geq b + t_1^3 \Leftrightarrow t_1^1 - t_1^3 \geq b - a \tag{1.16}$$

and for type $(\theta_3, \theta_4)$

$$f + t_1^3 \geq e + t_1^1 \Leftrightarrow t_1^1 - t_1^3 \leq f - e. \tag{1.17}$$

Therefore, if the valuations are such that $f - e < b - a$, there are no transfers that can satisfy (1.16) and (1.17). Furthermore, irrespective of what the distribution over opponent types is, the Bayesian incentive compatibility (BIC) constraints can never be satisfied. Assume that opponent type $(\tau_1, \tau_2)$ occurs with probability $q$. The two BIC constraints for player 1 are:

for type $(\theta_1, \theta_2)$

$$q(a + t_1^1) + (1 - q)(c + t_1^2) \geq q(b + t_1^3) + (1 - q)(c + t_1^4)$$

which is equivalent to

$$q(t_1^1 - t_1^3) + (1 - q)(t_1^2 - t_1^4) \geq q(b - a) \tag{1.18}$$

and for type $(\theta_3, \theta_4)$

$$q(f + t_1^3) + (1 - q)(g + t_1^4) \geq q(e + t_1^1) + (1 - q)(g + t_1^2)$$

which is equivalent to

$$q(t_1^1 - t_1^3) + (1 - q)(t_1^2 - t_1^4) \leq q(f - e). \tag{1.19}$$

If valuations are such that $f - e < b - a$, there are no transfers that can satisfy the two BIC constraints for player 1. In this case, the impossibility of ex post implementation implies that of Bayesian implementation.

For Bayesian implementation to be possible when ex post implementation is not, there needs to be some variability in the decision rule. Consider the following situation:

| $V(\theta, \tau)$ | $(\tau_1, \tau_2)$ | | $(\tau_3, \tau_4)$ | |
|---|---|---|---|---|
| $(\theta_1, \theta_2)$ | $a$ | $A$ | $c$ | $C$ |
| | X, $t^1$ | | Y, $t^2$ | |
| | $b$ | $B$ | $d$ | $D$ |
| $(\theta_3, \theta_4)$ | $e$ | $E$ | $g$ | $G$ |
| | Y, $t^3$ | | X, $t^4$ | |
| | $f$ | $F$ | $h$ | $H$ |

In this case, the BIC constraints for player 1 are: for type $(\theta_1, \theta_2)$

$$q(a + t_1^1) + (1-q)(d + t_1^2) \geq q(b + t_1^3) + (1-q)(c + t_1^4)$$

which is equivalent to

$$q(t_1^1 - t_1^3) + (1-q)(t_1^2 - t_1^4) \geq q(b - a) + (1-q)(c - d) \qquad (1.20)$$

and for type $(\theta_3, \theta_4)$

$$q(f + t_1^3) + (1-q)(g + t_1^4) \geq q(e + t_1^1) + (1-q)(h + t_1^2)$$

which is equivalent to

$$q(t_1^1 - t_1^3) + (1-q)(t_1^2 - t_1^4) \leq q(f - e) + (1-q)(g - h). \qquad (1.21)$$

Even if the valuations are such that $f - e < b - a$ and hence ex post implementation is impossible, (1.20) and (1.21) can hold simultaneously; that is, Bayesian implementation is feasible if $g - h$ is large enough compared to $c - d$. But note that $g - h \geq c - d$ is a necessary condition for EPIC to hold (given the opponent type is $(\tau_3, \tau_4)$). As intuition suggests, BIC is a weighted average of EPIC conditions, and thus, if some EPIC conditions are violated while others hold and compensate (in numbers or magnitude), then BIC will hold.[14] Clearly, the variability of the decision rule affects the number of EPIC conditions, and in turn, the flexibility we have to satisfy BIC when EPIC is violated.

---

[14]As a consequence, the addition of players or types makes the gap between ex post and Bayesian

35

## 1.7 Conclusion

This paper extends supermodular mechanism design to environments with interdependent valuations, informational and allocative externalities, and arbitrary finite type spaces. While realistic, these environments present a serious challenge to mechanism designers. It is typically impossible to employ dominant strategy and ex-post equilibrium. This makes Bayesian implementation particularly relevant. In this context, supermodular Bayesian mechanisms are attractive.

The main motivation behind our mechanism design approach is to facilitate convergence to a desired equilibrium. This includes two problems: the robustness to bounded rationality (especially learning) and the multiple equilibrium problem. Supermodular mechanisms have nice learning properties, and the interval between their extremal equilibria contains all the limit points of learning dynamics. To some extent, this interval "measures" the multiple equilibrium problem. Our methodology is to impose orders on type sets, and given these orders, to induce a supermodular mechanism and to minimize its interval prediction by weakening the complementarities. It is worth mentioning that agents need not be aware of the orders. While the analyst can exploit the monotonicity of agents' best-responses to derive convergence properties, agents need not know or be informed that their best-responses are monotonic.

The paper has focused on behavioral robustness and left other issues unanswered.

---

implementation grow wider. Consider our impossibility condition for ex post implementation: $(f - e < b - a)$ is both necessary and sufficient for that; if we keep this as the only violation of EPIC and increase the number of opponent types, then we increase the degrees of freedom to meet BIC. Increasing the number of player types also increases the relative number of BIC versus EPIC decision rules.

First, our mechanisms are parametric. The designer needs to know the prior beliefs to construct the mechanisms, which is demanding (Ledyard [40]). Moreover, mistakes with respect to the prior may lead to shifts in equilibrium behavior and deviations from efficiency. Along this line, the literature on robust mechanism design (Bergemann and Morris [5]) advocates the use of ex-post equilibrium. But this is not always possible in these environments.

Second, we have avoided the issue of budget balancing. Robustness to bounded rationality may well come at the price of a balanced budget, i.e. full efficiency. In both of the examples presented in Section 1.2 and Section 1.8.1, the designer could achieve dominance-solvability, uniqueness, and allocation efficiency by using the minimally supermodular transfers, but these transfers were not balanced. Reconciling budget balancing and minimal supermodularity (or, in general, dominance solvability) would be optimal, but this is an open question. If both properties were exclusive in general, the designer would be faced with a difficult choice: balancing the budget at the price of the implementation target (in case players do not learn to play the desired equilibrium), or guaranteeing the implementation target is reached at the price of a balanced budget.

## 1.8  Appendix

### 1.8.1  Another Motivating Example

Consider the motivating example of Section 1.2. The designer may choose the following transfers to implement the efficient decision rule:

| $t_1(\hat{\theta})$ | $(1,2)$ | $(2,1)$ |
|---|---|---|
| $(1,2)$ | 2 | -.10 |
| $(2,1)$ | 0 | -.25 |

| $t_2(\hat{\theta})$ | $(1,2)$ | $(2,1)$ |
|---|---|---|
| $(1,2)$ | 1.5 | .25 |
| $(2,1)$ | -1 | .75 |

Given these transfers, the resulting payoff matrix for the ex ante Bayesian game is

| EU | truthtelling | constant (1,2) | constant (2,1) | always lie |
|---|---|---|---|---|
| truthtelling | 2.41*; .63* | 2.38; .05 | 2.33*;0 | 2.29; -.13 |
| constant (1,2) | 2.33; 1.13 | 4.25*;1.25* | 1.53; .25 | 3.45*; .38 |
| constant (2,1) | 2.38;-.63 | 1.63; -1 | 2; .5* | 1.25; 0.13 |
| always lie | 2.29; -.13 | 3.5; -.25 | 1.2; .75* | 2.41; .63 |

Both truthtelling and a constant announcement of (1,2) by both players are ex ante Bayesian equilibria. If we instead use the supermodular transfers that add minimal complementarities

| $t_1^{SM}(\hat{\theta})$ | $(1,2)$ | $(2,1)$ |
|---|---|---|
| $(1,2)$ | .95 | .95 |
| $(2,1)$ | -1.625 | 1.375 |

| $t_2^{SM}(\hat{\theta})$ | $(1,2)$ | $(2,1)$ |
|---|---|---|
| $(1,2)$ | .25 | -.5 |
| $(2,1)$ | .25 | 1.5 |

we obtain the ex ante payoff matrix:

| EU | truthtelling | constant (1,2) | constant (2,1) | always lie |
|---|---|---|---|---|
| truthtelling | 2.41*; .63* | 1.04; .5 | 3.66*; 0 | 2.29; -.13 |
| constant (1,2) | 2.33; .13* | 3.2*; 0 | 2.58; -.5 | 3.45*; -.63 |
| constant (2,1) | 2.38; .38 | 0; .25 | 3.63; 1.25* | 1.25; 1.13 |
| always lie | 2.29; -.13 | 2.16; -.25 | 2.54; .75* | 2.41; .63 |

Converting the original mechanism into a minimally supermodular mechanism has solved the multiple equilibrium problem. Truthtelling is the unique Bayesian equilibrium.

### 1.8.2 An Example of Linear System for Minimal Supermodular Implementation

Consider a setting with $n = 3$ agents, and types in $\Theta_i = \{1, 2\}$ for all $i$. Assume the conventional order $2 >_i^1 1$ for all $i$. For each player $i$, in order to minimally supermodular implement the decision rule $x$, we are interested in finding a solution to the following system of linear equations:

$$
\begin{pmatrix}
1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\
1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 \\
0 & 0 & 0 & 0 & 1 & -1 & -1 & 1
\end{pmatrix}
\begin{pmatrix}
\delta_i(1, (1,1)) \\
\delta_i(2, (1,1)) \\
\delta_i(1, (1,2)) \\
\delta_i(2, (1,2)) \\
\delta_i(1, (2,1)) \\
\delta_i(2, (2,1)) \\
\delta_i(1, (2,2)) \\
\delta_i(2, (2,2))
\end{pmatrix}
=
\begin{pmatrix}
-\min_\theta X(\theta) \\
-\min_\theta Y(\theta) \\
-\min_\theta Z(\theta) \\
-\min_\theta W(\theta)
\end{pmatrix}.
$$

where

$$
\begin{aligned}
X(\theta) &= V_i(x_i(2, (1,2)); \theta) - V_i(x_i(1, (1,2)); \theta) - V_i(x_i(2, (1,1)); \theta) + V_i(x_i(1, (1,1)); \theta) \\
Y(\theta) &= V_i(x_i(2, (2,1)); \theta) - V_i(x_i(1, (2,1)); \theta) - V_i(x_i(2, (1,1)); \theta) + V_i(x_i(1, (1,1)); \theta) \\
Z(\theta) &= V_i(x_i(2, (2,2)); \theta) - V_i(x_i(1, (2,2)); \theta) - V_i(x_i(2, (1,2)); \theta) + V_i(x_i(1, (1,2)); \theta) \\
W(\theta) &= V_i(x_i(2, (2,2)); \theta) - V_i(x_i(1, (2,2)); \theta) - V_i(x_i(2, (2,1)); \theta) + V_i(x_i(1, (2,1)); \theta).
\end{aligned}
$$

Consider agent $i \in N$, whose valuations are given by:

| $V_i(\cdot; \theta)$ | $(1,1,1)$ | $(1,1,2)$ | $(1,2,1)$ | $(2,1,1)$ | $(1,2,2)$ | $(2,1,2)$ | $(2,2,1)$ | $(2,2,2)$ |
|---|---|---|---|---|---|---|---|---|
| $A$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $B$ | 3 | 1 | 2 | 1 | 2 | 2 | 2 | 2 |

Let us assume the decision rule to be implemented is:

| | $(1,1,1)$ | $(1,1,2)$ | $(1,2,1)$ | $(2,1,1)$ | $(1,2,2)$ | $(2,1,2)$ | $(2,2,1)$ | $(2,2,2)$ |
|---|---|---|---|---|---|---|---|---|
| $\tilde{x}_i(\theta)$ | B | A | A | B | B | B | B | B |

Then the rhs of the system becomes:

$$
\begin{pmatrix} -\min_\theta X \\ -\min_\theta Y \\ -\min_\theta Z \\ -\min_\theta W \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ 3 \\ 3 \end{pmatrix}
$$

One possible solution for the system is $\delta_i = (0, 1, 0, 0, 1, 1, 0, 3)^T$.

### 1.8.3  Proofs

**Proof of Theorem 1** Take a consistent profile of orders $\{(\geq_i^1, \geq_i^2)\}_i$. For every $i \in N$, each element $\theta_i \in \Theta_i$ is assigned an index $k$ that corresponds to its position in the set $\Theta_i$ under the total order $\geq_i^1$. Similarly, each element $\theta_{-i} \in \Theta_{-i}$ is assigned an index $q$ according to the total order order $\geq_i^2$ on $\Theta_{-i}$. Suppose that $f = (x, t)$ is implementable. Letting

$$
\delta_i(\hat{\theta}_i^k, \hat{\theta}_{-i}^q) \equiv -\sum_{l=1}^{k-1} \sum_{z=1}^{q-1} \min_{\theta \in \Theta} [V_i(x(\hat{\theta}_i^{l+1}, \hat{\theta}_{-i}^{z+1}); \theta) - V_i(x(\hat{\theta}_i^l, \hat{\theta}_{-i}^{z+1}); \theta)
$$

$$
- V_i(x(\hat{\theta}_i^{l+1}, \hat{\theta}_{-i}^z); \theta) + V_i(x(\hat{\theta}_i^l, \hat{\theta}_{-i}^z); \theta)]. \tag{1.22}
$$

for all $\hat{\theta}_i^k \in \Theta_i$ and $\hat{\theta}_{-i}^q \in \Theta_{-i}$, we define

$$
t_i^*(\hat{\theta}_i^k, \hat{\theta}_{-i}^q) \equiv \delta_i(\hat{\theta}_i^k, \hat{\theta}_{-i}^q) - E_{\theta_{-i}}[\delta_i(\hat{\theta}_i^k, \theta_{-i})] + E_{\theta_{-i}}[t_i(\hat{\theta}_i^k, \theta_{-i})] \tag{1.23}
$$

and show that $(x, t^*)$ is minimally supermodular implementable.

Step 1. We show that $t_i^*$ has smaller one-step supermodularity than any $t_i$ such that $(x, t)$ is supermodular implementable.

Let us define the one-step supermodularity of $V_i(x(\cdot); \theta)$ at any given announcement $(\hat{\theta}_i^k, \hat{\theta}_{-i}^q)$ as

$$
g_i(k, q; \theta) \equiv V_i(x(\hat{\theta}_i^{k+1}, \hat{\theta}_{-i}^{q+1}); \theta) - V_i(x(\hat{\theta}_i^k, \hat{\theta}_{-i}^{q+1}); \theta)
$$

$$
- V_i(x(\hat{\theta}_i^{k+1}, \hat{\theta}_{-i}^q); \theta) + V_i(x(\hat{\theta}_i^k, \hat{\theta}_{-i}^q); \theta). \tag{1.24}
$$

For notational simplicity, we define

$$
\begin{aligned}
d_i(k, q) &\equiv \min_{\theta \in \Theta} [V_i(x(\hat{\theta}_i^{k+1}, \hat{\theta}_{-i}^{q+1}); \theta) - V_i(x(\hat{\theta}_i^k, \hat{\theta}_{-i}^{q+1}); \theta) \\
&\quad - V_i(x(\hat{\theta}_i^{k+1}, \hat{\theta}_{-i}^q); \theta) + V_i(x(\hat{\theta}_i^k, \hat{\theta}_{-i}^q); \theta)] \\
&= \min_{\theta \in \Theta} g_i(k, q; \theta). \tag{1.25}
\end{aligned}
$$

Since the one-step supermodularity of $t_i^*$ is equivalent to the one-step supermodularity of $\delta_i$ we have

$$
\begin{aligned}
s_i(k,q) &\equiv \delta_i(\hat{\theta}_i^{k+1}, \hat{\theta}_{-i}^{q+1}) - \delta_i(\hat{\theta}_i^k, \hat{\theta}_{-i}^{q+1}) - \delta_i(\hat{\theta}_i^{k+1}, \hat{\theta}_{-i}^q) + \delta_i(\hat{\theta}_i^k, \hat{\theta}_{-i}^q) \\
&= -\sum_{l=1}^{k}\sum_{z=1}^{q} d_i(l,z) + \sum_{l=1}^{k-1}\sum_{z=1}^{q} d_i(l,z) + \sum_{l=1}^{k}\sum_{z=1}^{q-1} d_i(l,z) - \sum_{l=1}^{k-1}\sum_{z=1}^{q-1} d_i(l,z) \\
&= -d_i(k,q) \quad\quad\quad (1.26)
\end{aligned}
$$

as the one-step supermodularity of $t_i^*$ (and $\delta_i$).

Therefore, the one-step supermodularity of $(V_i + t_i^*)$ is given by

$$
g_i(k,q;\theta) + s_i(k,q) \geq 0 \quad\quad\quad (1.27)
$$

for all $\hat{\theta}_i^k, \hat{\theta}_{-i}^q, \theta, k, q$, and $i$.

Denote the one-step supermodularity of transfer $t_i$ as $sm_1(t_i; k, q)$, that is:

$$
sm_1(t_i; k, q) = t_i(\hat{\theta}_i^{k+1}, \hat{\theta}_{-i}^{q+1}) - t_i(\hat{\theta}_i^k, \hat{\theta}_{-i}^{q+1}) - t_i(\hat{\theta}_i^{k+1}\hat{\theta}_{-i}^q) + t_i(\hat{\theta}_i^k, \hat{\theta}_{-i}^q).
$$

For all transfers $t$ such that $(x,t)$ is supermodular implementable, it must hold that $g_i(k,q;\theta) + sm_1(t_i; k, q) \geq 0$ for all $\theta \in \Theta$, which is equivalent to:

$$
\begin{aligned}
sm_1(t_i; k, q) \geq - \min_{\theta\in\Theta}[&V_i(x(\hat{\theta}_i^{k+1}, \hat{\theta}_{-i}^{q+1}); \theta) - V_i(x(\hat{\theta}_i^k, \hat{\theta}_{-i}^{q+1}); \theta) \\
& - V_i(x(\hat{\theta}_i^{k+1}, \hat{\theta}_{-i}^q); \theta) + V_i(x(\hat{\theta}_i^k, \hat{\theta}_{-i}^q); \theta)] = s_i(k,q). \quad (1.28)
\end{aligned}
$$

The above shows that if $(x,t)$ is supermodular implementable then the one-step supermodularity of transfers $t$ is necessarily (weakly) greater than the one-step supermodularity of transfers $t^*$, which establishes Step 1.

Step 2. We show that the (multiple-step) supermodularity of any function of two variables is a sum of one-step supermodularities. Let us define the "$(\eta, \gamma)$-step supermodularity" of any function $t_i(\hat{\theta}_i^k, \hat{\theta}_{-i}^q)$ as

$$
SM_{(\eta,\gamma)}(t_i; k, q) = t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q+\gamma}) - t_i(\hat{\theta}_i^k, \hat{\theta}_{-i}^{q+\gamma}) - t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^q) + t_i(\hat{\theta}_i^k, \hat{\theta}_{-i}^q). \quad (1.29)
$$

Note that

$$
\begin{aligned}
t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q+\gamma}) = sm_1(t_i; k+\eta-1, q+\gamma-1) &+ t_i(\hat{\theta}_i^{k+\eta-1}, \hat{\theta}_{-i}^{q+\gamma}) \\
&+ t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q+\gamma-1}) - t_i(\hat{\theta}_i^{k+\eta-1}, \hat{\theta}_{-i}^{q+\gamma-1}), \quad (1.30)
\end{aligned}
$$

41

and so it follows from (1.29) that

$$SM_{(\eta,\gamma)}(t_i; k, q) = \left[ sm_1(t_i; k + \eta - 1, q + \gamma - 1) + t_i(\hat{\theta}_i^{k+\eta-1}, \hat{\theta}_{-i}^{q+\gamma}) + t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q+\gamma-1}) \right.$$
$$\left. - t_i(\hat{\theta}_i^{k+\eta-1}, \hat{\theta}_{-i}^{q+\gamma-1}) \right] - t_i(\hat{\theta}_i^{k}, \hat{\theta}_{-i}^{q+\gamma}) - t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q}) + t_i(\hat{\theta}_i^{k}, \hat{\theta}_{-i}^{q})). \quad (1.31)$$

Note that

$$t_i(\hat{\theta}_i^{k+\eta-1}, \hat{\theta}_{-i}^{q+\gamma}) = sm_1(t_i; k + \eta - 2, q + \gamma - 1) + t_i(\hat{\theta}_i^{k+\eta-2}, \hat{\theta}_{-i}^{q+\gamma})$$
$$+ t_i(\hat{\theta}_i^{k+\eta-1}, \hat{\theta}_{-i}^{q+\gamma-1}) - t_i(\hat{\theta}_i^{k+\eta-2}, \hat{\theta}_{-i}^{q+\gamma-1}), \quad (1.32)$$

and therefore it follows from (1.31) that

$$SM_{(\eta,\gamma)}(t_i; k, q) = sm_1(t_i; k + \eta - 1, q + \gamma - 1) + \left[ sm_1(t_i; k + \eta - 2, g + \gamma - 1) \right.$$

$$+ t_i(\hat{\theta}_i^{k+\eta-2}, \hat{\theta}_{-i}^{q+\gamma}) + t_i(\hat{\theta}_i^{k+\eta-1}, \hat{\theta}_{-i}^{q+\gamma-1}) - t_i(\hat{\theta}_i^{k+\eta-2}, \hat{\theta}_{-i}^{q+\gamma-1}) \bigg]$$
$$+ t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q+\gamma-1}) - t_i(\hat{\theta}_i^{k+\eta-1}, \hat{\theta}_{-i}^{q+\gamma-1}) - t_i(\hat{\theta}_i^{k}, \hat{\theta}_{-i}^{q+\gamma}) - t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q}) + t_i(\hat{\theta}_i^{k}, \hat{\theta}_{-i}^{q})$$
$$(1.33)$$

which is equal to

$$\sum_{n=1}^{2} sm_1(t_i; k + \eta - n, q + \gamma - 1) + t_i(\hat{\theta}_i^{k+\eta-2}, \hat{\theta}_{-i}^{q+\gamma}) - t_i(\hat{\theta}_i^{k+\eta-2}, \hat{\theta}_{-i}^{q+\gamma-1})$$

$$+ t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q+\gamma-1}) - t_i(\hat{\theta}_i^{k}, \hat{\theta}_{-i}^{q+\gamma}) - t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q}) + t_i(\hat{\theta}_i^{k}, \hat{\theta}_{-i}^{q}). \quad (1.34)$$

Proceeding iteratively with this process of substitution and regrouping of terms for $n = (1, \ldots, \eta)$ we obtain

$$SM_{(\eta,\gamma)}(t_i; k, q) = \sum_{n=1}^{\eta} sm_1(t_i; k + \eta - n, q + \gamma - 1) + t_i(\hat{\theta}_i^{k}, \hat{\theta}_{-i}^{q+\gamma}) - t_i(\hat{\theta}_i^{k}, \hat{\theta}_{-i}^{q+\gamma-1})$$
$$+ t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q+\gamma-1}) - t_i(\hat{\theta}_i^{k}, \hat{\theta}_{-i}^{q+\gamma}) - t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q}) + t_i(\hat{\theta}_i^{k}, \hat{\theta}_{-i}^{q})$$
$$= \sum_{n=1}^{\eta} sm_1(t_i; k + \eta - n, q + \gamma - 1)$$
$$+ t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q+\gamma-1}) - t_i(\hat{\theta}_i^{k}, \hat{\theta}_{-i}^{q+\gamma-1}) - t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q}) + t_i(\hat{\theta}_i^{k}, \hat{\theta}_{-i}^{q})$$
$$= \sum_{n=1}^{\eta} sm_1(t_i; k + \eta - n, q + \gamma - 1) + SM_{(\eta,\gamma-1)}(t_i; k, q). \quad (1.35)$$

Iterating on Equation (1.35) for $m = 1, \ldots, \gamma - 1$ we obtain:

$$SM_{(\eta,\gamma)}(t_i; k, q) = \sum_{n=1}^{\eta} sm_1(t_i; k + \eta - n, q + \gamma - 1) + SM_{(\eta,\gamma-1)}(t_i; k, q)$$

$$= \sum_{n=1}^{\eta} sm_1(t_i; k + \eta - n, q + \gamma - 1) + \sum_{n=1}^{\eta} sm_1(t_i; k + \eta - n, q + \gamma - 2) + SM_{(\eta,\gamma-2)}(t_i; k, q)$$

$$= \sum_{n=1}^{\eta} \sum_{m=1}^{\gamma-1} sm_1(t_i; k + \eta - n, q + \gamma - m) + SM_{(\eta,1)}(t_i; k, q). \tag{1.36}$$

Now, using the fact that

$$
\begin{aligned}
SM_{(\eta,1)}(k, q) &= t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q+1}) - t_i(\hat{\theta}_i^{k}, \hat{\theta}_{-i}^{q+1}) - t_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q}) + t_i(\hat{\theta}_i^{k}, \hat{\theta}_{-i}^{q}) \\
&= sm_1(t_i; k + \eta - 1, q) + SM_{(\eta-1,1)}(t_i; k, q) \\
&= \sum_{n=1}^{\eta} sm_1(t_i; k + \eta - n, q)
\end{aligned}
$$

and plugging this into Equation (1.36), we obtain

$$
\begin{aligned}
SM_{(\eta,\gamma)}(t_i; k, q) &= \sum_{n=1}^{\eta} \sum_{m=1}^{\gamma-1} sm_1(t_i; k + \eta - n, q + \gamma - m) + \sum_{n=1}^{\eta} sm_1(t_i; k + \eta - n, q) \\
&= \sum_{n=1}^{\eta} \sum_{m=1}^{\gamma} sm_1(t_i; k + \eta - n, q + \gamma - m) \\
&= \sum_{l=k}^{k+\eta-1} \sum_{z=q}^{q+\gamma-1} sm_1(t_i; l, z). \tag{1.37}
\end{aligned}
$$

Thus, the multiple-step supermodularity of any function of two ordered variables is equal to the sum of one-step supermodularities, which establishes Step 2.

Step 3. Conclusion. Note that

$$E_{\theta_{-i}}[t_i^*(\hat{\theta}_i^k, \theta_{-i})] = E_{\theta_{-i}}[\delta_i(\hat{\theta}_i^k, \theta_{-i})] - E_{\theta_{-i}}[\delta_i(\hat{\theta}_i^k, \theta_{-i})] + E_{\theta_{-i}}[t_i(\hat{\theta}_i^k, \theta_{-i})] = E_{\theta_{-i}}[t_i(\hat{\theta}_i^k, \theta_{-i})] \tag{1.38}$$

and therefore transfers $t_i$ and $t_i^*$ have the same expected value given that all other agents report their types truthfully. That is, assuming truthful reporting, the expected utility of an agent is the same under $t_i$ and $t_i^*$. Since $(x, t)$ is truthfully implementable, the above implies that $(x, t^*)$ is also truthfully implementable.

Using the result eshablished in Step 2, the $(\eta, \gamma)$-step supermodularity of $V_i(x(\cdot); \theta)$ at any given announcement $(\hat{\theta}_i^k, \hat{\theta}_{-i}^q)$ can now be written as:

$$
\begin{aligned}
G_i^{(\eta,\gamma)}(k, q; \theta) &= V_i(x(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q+\gamma}); \theta) - V_i(x(\hat{\theta}_i^k, \hat{\theta}_{-i}^{q+\gamma}); \theta) \\
&\quad - V_i(x(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^q); \theta) + V_i(x(\hat{\theta}_i^k, \hat{\theta}_{-i}^q); \theta) \\
&= \sum_{l=k}^{k+\eta-1} \sum_{z=q}^{q+\gamma-1} g_i(l, z; \theta).
\end{aligned}
\tag{1.39}
$$

and the $(\eta, \gamma)$-step supermodularity of $t_i^*$ is analogously given by

$$
\begin{aligned}
S_i^{(\eta,\gamma)}(k, q) &= \delta_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^{q+\gamma}) - \delta_i(\hat{\theta}_i^k, \hat{\theta}_{-i}^{q+\gamma}) - \delta_i(\hat{\theta}_i^{k+\eta}, \hat{\theta}_{-i}^q) + \delta_i(\hat{\theta}_i^k, \hat{\theta}_{-i}^q) \\
&= -\sum_{l=1}^{k+\eta-1} \sum_{z=1}^{q+\gamma-1} d_i(l, z) + \sum_{l=1}^{k-1} \sum_{z=1}^{q+\gamma-1} d_i(l, z) + \sum_{l=1}^{k+\eta-1} \sum_{z=1}^{q-1} d_i(l, z) - \sum_{l=1}^{k-1} \sum_{z=1}^{q-1} d_i(l, z) \\
&= -\sum_{l=k}^{k+\eta-1} \sum_{z=q}^{q+\gamma-1} d_i(l, z).
\end{aligned}
\tag{1.40}
$$

It is straightforward to check that $G_i^{(\eta,\gamma)}(k, q; \theta) + S_i^{(\eta,\gamma)}(k, q) \geq 0$ for all $\hat{\theta}_i^k, \hat{\theta}_{-i}^q, \theta, k, q, \eta, \gamma$ and $i$ and, therefore, $t^*$ is supermodular implementable.

Moreover, Step 1 says that $t^*$ has the smallest one-step supermodularity among all supermodular transfers $t$. Combined with Step 2, this establishes that $t^*$ has the smallest $(\eta, \gamma)$-step supermodularity for any $(\eta, \gamma)$ among all supermodular transfers $t$. Thus we conclude that $(x, t^*)$ is minimally supermodular implementable under the chosen order profile $\{(\geq_i^1, \geq_i^2)\}_i$. $\qquad \square$

**Proof of Corollary 1** In the proof of Theorem 1, we constructed transfers that minimally supermodular implemented the decision rule $x$ under some chosen consistent profile of orders $\{(\geq_i^1, \geq_i^2)\}_i$. Each $(\geq_i^1, \geq_i^2)$ is a pair of complete orders on finite sets. Since there are finitely many agents, for each $i$ there are finitely many complete orders, and consequently, finitely many consistent profiles. For each such profile, we can compute the distance between the largest and the smallest equilibrium in the ex ante induced game under the minimal transfers, using a metric $d$. Among all consistent profiles of orders we can thus choose the one associated with the smallest interval prediction as measured by $d$: denote this profile of orders by $\{(\geq_i^{*1}, \geq_i^{*2})\}_i$ and the corresponding minimal transfers by $t^{**}$. Therefore, $t^{**}$ give the smallest interval prediction under $d$ among all minimally supermodular transfers on consistent profiles of

orders. □

**Proof of Proposition 2**
⇐ (**sufficiency**) The structure of the system implies that if there exists a solution $\delta_i$ for all $i$, then for all $\theta'_i$, $\theta''_i$, where $\theta''_i$ is an immediate successor of $\theta'_i$ in $\Theta_i$, and for all $\theta'_{-i}$, $\theta''_{-i}$, where $\theta''_{-i}$ is an immediate successor of $\theta'_{-i}$ in $\Theta_{-i}$, the supermodularity of $\delta_i(\theta)$ is equal to

$$- \min_{\theta \in \Theta} [V_i(x_i(\theta''_i, \theta''_{-i}), \theta) - V_i(x_i(\theta''_i, \theta'_{-i}), \theta) - V_i(x_i(\theta'_i, \theta''_{-i}), \theta) + V_i(x_i(\theta'_i, \theta'_{-i}), \theta)].$$
(1.41)

We construct transfers $t_i^*(\theta)$ by transforming any set of truth-revealing transfers $t_i(\theta)$ according to (1.3). Note that transfers $t_i(\theta)$ and $t_i^*(\theta)$ have the same expected value given that all other agents report truthfully. Hence transfers $t_i^*(\theta)$ also achieve truthful implementation of the same decision rule.

Moreover, it is easy to see from (1.3) that the supermodularity of $t_i^*(\theta)$ is equivalent to that of $\delta_i(\theta)$ for any two pairs of announcements. Since the one-step supermodularity of $\delta_i(\theta)$ is equivalent to that of $t_i^*(\theta)$, and in turn equal to the above equation, we can conclude that transfers $t_i^*(\theta)$ have the smallest possible one-step supermodularity.

To argue that transfers $t_i^*(\theta)$ also have the smallest multiple-step supermodularity, we rely on the proof of Theorem 1 (Step 2). In particular, (1.37) states that the multiple-step supermodularity of any function of two ordered variables is equal to the sum of all one-step supermodularities in between. We develop our argument below.

Take any $\theta''_i \geq_i^1 \theta'_i$ and $\theta''_{-i} \geq_{-i} \theta'_{-i}$. Consider the sequence of immediate successors $(\theta_i^0, \theta_i^1, \ldots, \theta_i^M)$ such that $\theta_i^0 = \theta'_i$, $\theta_i^M = \theta''_i$, and $\theta_i^{m+1}$ is the immediate successor of $\theta_i^m$ for all $m = 1, \ldots, M - 1$. Since $\Theta_i$ is totally ordered by $\geq_i^1$, this sequence is unique. On the other hand, the set of opponent types $\Theta_{-i}$ is only partially ordered by $\geq_{-i}$. Thus, there may be several different sequences of immediate successors that connect $\theta'_{-i}$ to $\theta''_{-i}$. Take any sequence of immediate successors $(\theta_{-i}^0, \theta_{-i}^1, \ldots, \theta_{-i}^N)$ such that $\theta_{-i}^0 = \theta'_{-i}$, $\theta_{-i}^N = \theta''_{-i}$, and $\theta_{-i}^n$ is an immediate successor of $\theta_{-i}^{n-1}$ for all

45

$n = 1, \ldots, N - 1$. We can now apply (1.37) to these sequences:

$$\delta_i(\theta_i'', \theta_{-i}'') - \delta_i(\theta_i'', \theta_{-i}') - \delta_i(\theta_i', \theta_{-i}'') + \delta_i(\theta_i', \theta_{-i}') =$$
$$\delta_i(\theta_i^M, \theta_{-i}^N) - \delta_i(\theta_i^M, \theta_{-i}^0) - \delta_i(\theta_i^0, \theta_{-i}^N) + \delta_i(\theta_i^0, \theta_{-i}^0) =$$
$$\sum_{m=1}^{M} \sum_{n=1}^{N} [\delta_i(\theta_i^m, \theta_{-i}^n) - \delta_i(\theta_i^m, \theta_{-i}^{n-1}) - \delta_i(\theta_i^{m-1}, \theta_{-i}^n) + \delta_i(\theta_i^{m-1}, \theta_{-i}^{n-1})]. \quad (1.42)$$

Combining (1.41) and (1.42), we obtain:

$$\delta_i(\theta_i'', \theta_{-i}'') - \delta_i(\theta_i'', \theta_{-i}') - \delta_i(\theta_i', \theta_{-i}'') + \delta_i(\theta_i', \theta_{-i}') =$$
$$\sum_{m=1}^{M} \sum_{n=1}^{N} - \min_{\theta \in \Theta} [V_i(x_i(\theta_i^m, \theta_{-i}^n), \theta) - V_i(x_i(\theta_i^m, \theta_{-i}^{n-1}), \theta)$$
$$- V_i(x_i(\theta_i^{m-1}, \theta_{-i}^n), \theta) + V_i(x_i(\theta_i^{m-1}, \theta_{-i}^{n-1}), \theta)]. \quad (1.43)$$

Irrespective of which sequence of immediate successors connecting $\theta_{-i}'$ to $\theta_{-i}''$ we choose, the multiple-step supermodularity of $\delta_i(\theta)$, and therefore of $t_i^*(\theta)$, can always be represented as a sum of minimal one-step supermodularities. Thus, there are no other transfers that can do better on any multiple step while ensuring that all one steps are not smaller than the corresponding expression (1.41). Hence, transfers $t_i^*(\theta)$ have the smallest possible multiple-step supermodularities. We conclude $t \succeq_{\text{ID}} t^*$ for all $t \in \mathcal{T}$.

$\Rightarrow$ **(necessity)** Suppose the system does not have a solution. Then any collection of numbers $\{\delta_i(\theta) : \theta \in \Theta, i \in N\}$, and in particular any transfers $\{t_i(\theta) : \theta \in \Theta, i \in N\}$, must fall into one or both of the following cases:

<u>Case 1</u>: There exist $i$ and two pairs of immediate successors $(\theta_i', \theta_i'')$ and $(\theta_{-i}', \theta_{-i}'')$ such that

$$t_i(\theta_i'', \theta_{-i}'') - t_i(\theta_i'', \theta_{-i}') - t_i(\theta_i', \theta_{-i}'') + t_i(\theta_i', \theta_{-i}') >$$
$$- \min_{\theta \in \Theta} [V_i(x_i(\theta_i'', \theta_{-i}''), \theta) - V_i(x_i(\theta_i'', \theta_{-i}'), \theta) - V_i(x_i(\theta_i', \theta_{-i}''), \theta) + V_i(x_i(\theta_i', \theta_{-i}'), \theta)].$$
$$(1.44)$$

In this case, we can choose different numbers $\{\tilde{\delta}_i(\theta)\}$ such that, for these particular

reports, the equality is satisfied:

$$\tilde{\delta}_i(\theta_i'', \theta_{-i}'') - \tilde{\delta}_i(\theta_i'', \theta_{-i}') - \tilde{\delta}_i(\theta_i', \theta_{-i}'') + \tilde{\delta}_i(\theta_i', \theta_{-i}') =$$
$$- \min_{\theta \in \Theta}[V_i(x_i(\theta_i'', \theta_{-i}''), \theta) - V_i(x_i(\theta_i'', \theta_{-i}'), \theta) - V_i(x_i(\theta_i', \theta_{-i}''), \theta) + V_i(x_i(\theta_i', \theta_{-i}'), \theta)].$$
(1.45)

Using numbers $\{\tilde{\delta}_i(\theta)\}$, we build transfers $\{\tilde{t}_i(\theta)\}$ according to equation (1.3). Clearly, it does not hold that $\tilde{t} \succeq_{\mathrm{ID}} t$, which violates the definition of minimally supermodular implementation.

$\underline{\text{Case 2}}$: There exist $i$ and two pairs of immediate successors $(\theta_i', \theta_i'')$ and $(\theta_{-i}', \theta_{-i}'')$ such that

$$t_i(\theta_i'', \theta_{-i}'') - t_i(\theta_i'', \theta_{-i}') - t_i(\theta_i', \theta_{-i}'') + t_i(\theta_i', \theta_{-i}') <$$
$$- \min_{\theta \in \Theta}[V_i(x_i(\theta_i'', \theta_{-i}''), \theta) - V_i(x_i(\theta_i'', \theta_{-i}'), \theta) - V_i(x_i(\theta_i', \theta_{-i}''), \theta) + V_i(x_i(\theta_i', \theta_{-i}'), \theta)].$$
(1.46)

In this case, the supermodularity contained in $t_i$ is not sufficient to induce a supermodular game, which also violates the definition of minimally supermodular implementation. $\square$

**Proof of Proposition 3** By way of contradiction, suppose that profile $\theta^*(\cdot) \geq \theta^T(\cdot)$ is an equilibrium so that player $i$'s best response to $\theta_{-i}^*(\cdot)$ is $\theta_i^*(\cdot)$. Thus, for all $i$, $\theta_i$, and $\hat{\theta}_i$ such that $\theta_i^*(\theta_i) >_i^1 \hat{\theta}_i \geq_i^1 \theta_i$, the following must hold

$$E_{\theta_{-i}}[\Delta u_i(\theta_{-i}^*(\theta_{-i}); \theta)] \equiv E_{\theta_{-i}}[u_i(\theta_i^*(\theta_i), \theta_{-i}^*(\theta_{-i}); \theta)] - E_{\theta_{-i}}[u_i(\hat{\theta}_i, \theta_{-i}^*(\theta_{-i}); \theta)] \geq 0.$$
(1.47)

We will show that this condition is not satisfied if the inequality in the theorem holds, i.e. there must be a player for whom a deception closer to the truthful strategy is strictly better than $\theta_i^*(\cdot)$. For simplicity, define[15]

$$E_{\theta_{-i}}[\Delta V_i(\theta_{-i}; \hat{\theta}_i, \theta_{-i})] \equiv E_{\theta_{-i}}[V_i(x(\theta_i^*(\theta_i), \theta_{-i}); \hat{\theta}_i, \theta_{-i})] - E_{\theta_{-i}}[V_i(x(\hat{\theta}_i, \theta_{-i}); \hat{\theta}_i, \theta_{-i})].$$
(1.48)

---

[15]The notation we used in Equation (1.15) becomes cumbersome in this proof, and so we replace $V_i(\hat{\theta}_i \triangleright \theta_i^*(\theta_i), \theta_{-i}; \hat{\theta}_i, \theta_{-i})$ with $\Delta V_i(\theta_{-i}; \hat{\theta}_i, \theta_{-i})$.

It follows from (1.8) and the definition of $\bar{K}_i(\theta_i)$ that for each $i$ and $\theta_i$:

$$E_{\theta_{-i}}[\Delta u_i(\theta^*_{-i}(\theta_{-i}); \theta)] \leq E_{\theta_{-i}}[\Delta u_i(\theta_{-i}; \theta)] + d_i(\theta^*_i(\theta_i), \hat{\theta}_i)\bar{K}_i(\theta_i)\sum_{j \neq i} E_{\theta_j}[d_j(\theta^*_j(\theta_j), \theta_j)].$$
(1.49)

Since the social choice function $(x, t)$ is implementable, the transfers $\{t_i\}$ induce truthful revelation. Therefore, it must be that for all $i$ and $\theta_i$ the incentive compatibility constraint is satisfied, that is:

$$E_{\theta_{-i}}[V_i(x(\hat{\theta}_i, \theta_{-i}); \hat{\theta}_i, \theta_{-i})] - E_{\theta_{-i}}[V_i(x(\theta^*_i(\theta_i), \theta_{-i}); \hat{\theta}_i, \theta_{-i})] \geq$$
$$E_{\theta_{-i}}[t_i(\theta^*_i(\theta_i), \theta_{-i})] - E_{\theta_{-i}}[t_i(\hat{\theta}_i, \theta_{-i})]. \quad (1.50)$$

Thus, we obtain

$$\begin{aligned}
E_{\theta_{-i}}[\Delta u_i(\theta_{-i}; \theta)] &= E_{\theta_{-i}}[V_i(x(\theta^*_i(\theta_i), \theta_{-i}); \theta_i, \theta_{-i})] - E_{\theta_{-i}}[V_i(x(\hat{\theta}_i, \theta_{-i}); \theta_i, \theta_{-i})] \\
&\quad + E_{\theta_{-i}}[t_i(\theta^*_i(\theta_i), \theta_{-i})] - E_{\theta_{-i}}[t_i(\hat{\theta}_i, \theta_{-i})] \\
&\leq E_{\theta_{-i}}[\Delta V_i(\theta_{-i}; \theta_i, \theta_{-i})] - E_{\theta_{-i}}[\Delta V_i(\theta_{-i}; \hat{\theta}_i, \theta_{-i})] \\
&\leq -E_{\theta_{-i}}[\gamma_i(\theta_{-i})]d_i(\theta^*_i(\theta_i), \hat{\theta}_i)d_i(\hat{\theta}_i, \theta_i).
\end{aligned}$$
(1.51)

where the first inequality is derived after substituting in the LHS of (1.50) and the second inequality follows from (1.10). Combining (1.49) and (1.51), we arrive at

$$\frac{E_{\theta_{-i}}[\Delta u_i(\theta^*_{-i}(\theta_{-i}); \theta)]}{d_i(\theta^*_i(\theta_i), \hat{\theta}_i)} \leq \bar{K}_i(\theta_i)\sum_{j \neq i} E_{\theta_j}[d_j(\theta^*_j(\theta_j), \theta_j] - E_{\theta_{-i}}[\gamma_i(\theta_{-i})]d_i(\hat{\theta}_i, \theta_i). \quad (1.52)$$

If there exist $i$, $\theta_i$, and $\hat{\theta}_i \in [\theta_i, \theta^*_i(\theta_i))$ such that

$$\bar{K}_i(\theta_i)\sum_{j \neq i} E_{\theta_j}[d_j(\theta^*_j(\theta_j), \theta_j] - E_{\theta_{-i}}[\gamma_i(\theta_{-i})]d_i(\hat{\theta}_i, \theta_i) < 0 \quad (1.53)$$

then by (1.52) $E_{\theta_{-i}}[\Delta u_i(\theta^*_{-i}(\theta_{-i}); \theta)] < 0$, which contradicts (1.47). Therefore, $\theta^*(\cdot)$ is not a Bayesian equilibrium.

The same reasoning applies when $\theta^*(\cdot) \leq \theta^T(\cdot)$: if the condition of the theorem holds, $\theta^*(\cdot)$ cannot be a Bayesian equilibrium. $\qquad\square$

**Proof of Proposition 4** Take any profile $\theta^*(\cdot) \geq \theta^T(\cdot)$ outside the neighborhood of truthtelling. By way of contradiction, suppose that $\theta^*(\cdot)$ is an equilibrium. Then, for

all $i$ and $\theta_i$, the following must hold:

$$E_{\theta_{-i}}[\Delta u_i(\theta^*_{-i}(\theta_{-i}); \theta)] =$$

$$E_{\theta_{-i}}[u_i(\theta^*_i(\theta_i), \theta^*_{-i}(\theta_{-i}); \theta)] - E_{\theta_{-i}}[u_i(\hat{\theta}_i(\theta_i), \theta^*_{-i}(\theta_{-i}); \theta)] \geq 0 \quad (1.54)$$

for all deceptions $\hat{\theta}_i(\cdot) \in [\theta^T_i(\cdot), \theta^*_i(\cdot)]$. This in turn implies that the rhs of (1.52) must be nonnegative for all $i$, $\theta_i$ and $\hat{\theta}_i(\cdot)$, and thus its expected value (over $\theta_i$) must also be nonnegative, that is

$$E_{\theta_i}[\bar{K}_i(\theta_i)] \sum_{j \neq i} E_{\theta_j}[d_j(\theta^*_j(\theta_j), \theta_j] - E_{\theta_{-i}}[\gamma_i(\theta_{-i})] E_{\theta_i}[d_i(\hat{\theta}_i(\theta_i), \theta_i)] \geq 0. \quad (1.55)$$

We will show that there is an agent $i$ and a strategy $\hat{\theta}_i(\cdot)$ for which this inequality is violated, which is a contradiction.

Pick any agent $i$ such that $E_{\theta_i}[d_i(\theta^*_i(\theta_i), \theta_i)] \geq E_{\theta_j}[d_j(\theta^*_j(\theta_j), \theta_j)]$ for all $j$. Let us show that agent $i$ has an incentive to deviate from $\theta^*_i(\cdot)$ if the conditions of the proposition hold. By the definition of a metric,

$$E_{\theta_i}[d_i(\theta^*_i(\theta_i), \theta_i)] \leq E_{\theta_i}[d_i(\theta^*_i(\theta_i), \hat{\theta}_i(\theta_i))] + E_{\theta_i}[d_i(\hat{\theta}_i(\theta_i), \theta_i)]$$

and

$$E_{\theta_i}[d_i(\hat{\theta}_i(\theta_i), \theta_i)] \leq E_{\theta_i}[d_i(\theta^*_i(\theta_i), \hat{\theta}_i(\theta_i))] + E_{\theta_i}[d_i(\theta^*_i(\theta_i), \theta_i)],$$

and hence

$$|E_{\theta_i}[d_i(\theta^*_i(\theta_i), \theta_i)] - E_{\theta_i}[d_i(\hat{\theta}_i(\theta_i), \theta_i)]| \leq E_{\theta_i}[d_i(\theta^*_i(\theta_i), \hat{\theta}_i(\theta_i))] \quad (1.56)$$

for any $\hat{\theta}_i(\cdot)$. Note also that

$$\frac{\sum_{j \neq i} E_{\theta_j}[d_j(\theta^*_j(\theta_j), \theta_j]}{E_{\theta_i}[d_i(\theta^*_i(\theta_i), \theta_i)]} \leq (n-1) < \frac{E_{\theta_{-i}}[\gamma_i(\theta_{-i})]}{E_{\theta_i}[\bar{K}_i(\theta_i)]}, \quad (1.57)$$

where the first inequality follows from our choice of $i$, and the second inequality follows from the condition of the proposition. Hence,

$$E_{\theta_i}[\bar{K}_i(\theta_i)] \sum_{j \neq i} E_{\theta_j}[d_j(\theta^*_j(\theta_j), \theta_j)] - E_{\theta_{-i}}[\gamma_i(\theta_{-i})] E_{\theta_i}[d_i(\theta^*_i(\theta_i), \theta_i)] \quad (1.58)$$

$$\leq E_{\theta_i}[\bar{K}_i(\theta_i)] \sum_{j \neq i} E_{\theta_j}[d_j(\theta^*_j(\theta_j), \theta_j)] - E_{\theta_{-i}}[\gamma_i(\theta_{-i})] \frac{\sum_{j \neq i} E_{\theta_j}[d_j(\theta^*_j(\theta_j), \theta_j)]}{n-1}$$

$$= \sum_{j \neq i} E_{\theta_j}[d_j(\theta^*_j(\theta_j), \theta_j)] \left[ E_{\theta_i}[K_i(\theta_i)] - \frac{E_{\theta_{-i}}[\gamma_i(\theta_{-i})]}{n-1} \right] < 0. \quad (1.59)$$

49

It follows from (1.56) and the definition of $\varepsilon(\Theta)$ that we can choose a deviation $\hat{\theta}_i(\cdot)$, which is closer to truthtelling, i.e. $\theta_i^*(\cdot) >_i \hat{\theta}_i(\cdot)$, and which is close enough to $\theta_i^*(\cdot)$ so that

$$E_{\theta_i}[d_i(\theta_i^*(\theta_i), \theta_i)] - \varepsilon(\Theta) \le E_{\theta_i}[d_i(\hat{\theta}_i(\theta_i), \theta_i)]. \tag{1.60}$$

By (1.58) and (1.60), we obtain

$$E_{\theta_i}[\bar{K}_i(\theta_i)] \sum_{j \ne i} E_{\theta_j}[d_j(\theta_j^*(\theta_j), \theta_j)] - E_{\theta_{-i}}[\gamma_i(\theta_{-i})] E_{\theta_i}[d_i(\hat{\theta}_i(\theta_i), \theta_i)]$$

$$\le E_{\theta_i}[\bar{K}_i(\theta_i)] \sum_{j \ne i} E_{\theta_j}[d_j(\theta_j^*(\theta_j), \theta_j)] - E_{\theta_{-i}}[\gamma_i(\theta_{-i})] \left[ E_{\theta_i}[d_i(\theta_i^*(\theta_i), \theta_i)] - \varepsilon(\Theta) \right]$$

which is negative if

$$\sum_{j \ne i} E_{\theta_j}[d_j(\theta_j^*(\theta_j), \theta_j)] \left[ \frac{E_{\theta_{-i}}[\gamma_i(\theta_{-i})]}{n-1} - E_{\theta_i}[K_i(\theta_i)] \right] > E_{\theta_{-i}}[\gamma_i(\theta_{-i})]\varepsilon(\Theta),$$

and thus, if

$$\min_i \left\{ \sum_{j \ne i} E_{\theta_j}[d_j(\theta_j, \theta_j^*(\theta_j))] \right\} > \varepsilon(\Theta) \frac{\max_i\{E_{\theta_{-i}}[\gamma_i(\theta_{-i})]\}}{\min_i \left\{ \frac{E_{\theta_{-i}}[\gamma_i(\theta_{-i})]}{n-1} - E_{\theta_i}[K_i(\theta_i)] \right\}}. \tag{1.61}$$

Therefore, (1.55) is violated, which means that $\hat{\theta}_i(\cdot)$ is a profitable deviation from $\theta_i^*(\cdot)$. Thus, $\theta^*(\cdot)$ is not an equilibrium. An analogous argument applies to the case $\theta^*(\cdot) \le \theta^T(\cdot)$. $\qquad\qquad\square$

### 1.8.4  Unconditional Minimal Implementation and Order Reducibility

We first define a richness condition on decision rule $x$.

DEFINITION 9. *A decision rule $x(\theta)$ is order reducible if for each $i$, there are sets $\{G_p^i\}_{p=1}^P$ such that (a) $\Theta_{-i} = \cup_{p=1}^P G_p^i$, (b) for each $\theta_i$, $x_i(\theta) = x_i(\theta_i, \theta_{-i}')$ for all $\theta_{-i}, \theta_{-i}' \in G_p^i$, and (c) if $\theta_{-i} \in G_p^i$, all immediate successors of $\theta_{-i}$ must be in $G_p^i \cup G_{p+1}^i$.*

Order reducibility ensures that, through the structure of the decision rule, opponents' type profiles can be put into groups to form a linear path between the images of $x_i$. This linear path preserves the product order on $\Theta_{-i}$ and does not

impose any ordering of images between unordered types. To illustrate the definition, consider a setting with $n = 3$ agents and $\Theta_i = \{1, 2\}$ for all $i$. Assume types are ordered according to the usual order, i.e. $2 >_i^1 1$ for all $i$. Suppose the decision rule is $x_i(\theta) = x(\theta) = h(\sum \theta_i)$ where $h$ is some strictly increasing real-valued function (Mathevet [45] presents several examples where the efficient decision rule takes this form). This decision rule is order reducible: for each agent $i$, it yields partition $G_1^i = \{(1, 1)\}$, $G_2^i = \{(1, 2), (2, 1)\}$ and $G_3^i = \{(2, 2)\}$. Note that for $n = 2$, order reducibility is trivially satisfied by all decision rules. Indeed, for each $j \neq i$, let each type in $\Theta_j$ form its own group with an index that corresponds to the position of the type under $>_j^1$. Below we present an example where order reducibility is violated.

PROPOSITION 5. *Let $f = (x, t)$ be a social choice function such that $x$ is order reducible. If $f$ is implementable, then there exist $t^o$ such that $(x, t^o)$ is minimally supermodular implementable.*

This proposition establishes minimal supermodular implementability of a class of social choice functions. For any implementable social choice function, if the decision rule satisfies order reducibility, then there exist transfers $t^o$ that guarantee truthful supermodular implementation as well as the smallest equilibrium set among all supermodular transfers. There are many ways in which a mechanism can be converted into a supermodular one. It is therefore useful to describe the best way to convert it (and when it exists) given the objective of minimized equilibrium set. In the proof of the proposition, we provide an explicit formula for transfers $t^o$.

Order reducibility may seem to be a restrictive condition. Unfortunately, relaxing it just a little in a simple setting already defies existence of minimal transfers, as the following example demonstrates. Consider a three-agent two-type example. Let $\Theta_i = \{1, 2\}$ and $2 >_i^1 1$ for all $i$. Choose a decision rule $x$ such that for some $i$, the only possible grouping is $G_1^i = \{(1, 1)\}$, $G_2^i = \{(1, 2)\}$, $G_3^i = \{(2, 1), (2, 2)\}$ (actual group indexes do not matter). This decision rule is not order reducible since $(2,1)$, despite being an immediate successor of $(1,1)$, is in a group that does not immediately follow $G_i^1$. For most valuation functions, a solution to our system of linear equations does not exist in this case. Thus, the transfers $t^o$ are not minimal but no other transfers are.

**Proof of Proposition 5** Suppose $f = (x, t)$ is implementable and $x$ is order reducible. For every $i \in N$, assign to each element $\theta_i \in \Theta_i$ an index $k$ that corresponds to its position in the set $\Theta_i$ under the total order $\geq_i^1$. Since $x$ is order reducible, each element $\theta_{-i} \in \Theta_{-i}$ can be assigned an index $p$ according to the group $G_p^i$ to which

51

it belongs. Note that more than one element $\theta_{-i}$ can be assigned the same index $p$, because all the elements in group $G_p$ share the same index $p$. Letting

$$\delta_i(\theta_i^k, \theta_{-i}^p) = -\sum_{l=1}^{k-1}\sum_{z=1}^{p-1}\min_{\theta\in\Theta}[V_i(x(\theta_i^{l+1}, \theta_{-i}^{z+1}), \theta) - V_i(x(\theta_i^l, \theta_{-i}^{z+1}), \theta)$$
$$- V_i(x(\theta_i^{l+1}, \theta_{-i}^z), \theta) + V_i(x(\theta_i^l, \theta_{-i}^z), \theta)] \qquad (1.62)$$

for all $\theta_i^k \in \Theta_i$ and $\theta_{-i}^p \in \Theta_{-i}$, we define

$$t_i^o(\theta_i^k, \theta_{-i}^p) = \delta_i(\theta_i^k, \theta_{-i}^p) - E_{\theta_{-i}}[\delta_i(\theta_i^k, \theta_{-i})] + E_{\theta_{-i}}[t_i(\theta_i^k, \theta_{-i})] \qquad (1.63)$$

and show that $(x, t^o)$ is minimally supermodular implementable.

Note that $E_{\theta_{-i}}[t_i^o(\theta_i^k, \theta_{-i})] = E_{\theta_{-i}}[t_i(\theta_i^k, \theta_{-i})]$ and thus $(x, t^o)$ is truthfully implementable. Moreover, the supermodularity of $t_i^o(\theta_i^k, \theta_{-i}^p)$ is equal to the supermodularity of $\delta_i(\theta_i^k, \theta_{-i}^p)$. We proceed to show in separate steps of the proof that transfers $t^o$ achieve minimal supermodularities across immediate successors on $(\Theta_i, \geq_i^1)$ and $(\Theta_{-i}, \geq_{-i})$ (Step 1) and that the supermodularities of $t_i^o$ across (multiple-step) successive types are sums of supermodularities between immediate (one-step) successors (Step 2).

Step 1. Consider any two pairs of immediate successors $\theta_i'' \geq_i^1 \theta_i'$ and $\theta_{-i}'' \geq_{-i} \theta_{-i}'$. As they are immediate successors, we can instead write $\theta_i^{k+1} \geq_i^1 \theta_i^k$. The (one-step) supermodularity of $t_i^o$ is

$$t_i^o(\theta_i^{k+1}, \theta_{-i}'') - t_i^o(\theta_i^k, \theta_{-i}'') - t_i^o(\theta_i^{k+1}, \theta_{-i}') + t_i^o(\theta_i^k, \theta_{-i}') =$$
$$\delta_i(\theta_i^{k+1}, \theta_{-i}'') - \delta_i(\theta_i^k, \theta_{-i}'') - \delta_i(\theta_i^{k+1}, \theta_{-i}') + \delta_i(\theta_i^k, \theta_{-i}'). \quad (1.64)$$

Since $x$ is order reducible and $\theta_{-i}'' \geq_{-i} \theta_{-i}'$ are immediate successors, it must be that either $\theta_{-i}', \theta_{-i}'' \in G_p^i$ or $\theta_{-i}' \in G_p^i$ and $\theta_{-i}'' \in G_{p+1}^i$.
Case 1. If $\theta_{-i}', \theta_{-i}'' \in G_p^i$, then by order reducibility, $x(\theta_i, \theta_{-i}') = x(\theta_i, \theta_{-i}'')$ for all $\theta_i$ and we obtain

$$V_i(x(\theta_i^{k+1}, \theta_{-i}''); \theta) - V_i(x(\theta_i^k, \theta_{-i}''); \theta) - V_i(x(\theta_i^{k+1}, \theta_{-i}'); \theta) + V_i(x(\theta_i^k, \theta_{-i}'); \theta) = 0. \qquad (1.65)$$

Using equation (1.62) for $\delta_i$ we have that $\delta_i(\theta_i, \theta_{-i}') = \delta_i(\theta_i, \theta_{-i}'') = \delta_i(\theta_i, \theta_{-i}^p)$ for all $\theta_i$. The supermodularity of $t_i^o$ hence becomes:

$$t_i^o(\theta_i^{k+1}, \theta_{-i}'') - t_i^o(\theta_i^k, \theta_{-i}'') - t_i^o(\theta_i^{k+1}, \theta_{-i}') + t_i^o(\theta_i^k, \theta_{-i}') =$$
$$\delta_i(\theta_i^{k+1}, \theta_{-i}^p) - \delta_i(\theta_i^k, \theta_{-i}^p) - \delta_i(\theta_i^{k+1}, \theta_{-i}^p) + \delta_i(\theta_i^k, \theta_{-i}^p) = 0. \quad (1.66)$$

Hence, for all $t_i$ such that $(x,t)$ is supermodular implementable it must hold that:

$$t_i(\theta_i^{k+1},\theta''_{-i}) - t_i(\theta_i^k,\theta''_{-i}) - t_i(\theta_i^{k+1},\theta'_{-i}) + t_i(\theta_i^k,\theta'_{-i}) \geq$$
$$- \min_\theta [V_i(x(\theta_i^{k+1},\theta''_{-i});\theta) - V_i(x(\theta_i^k,\theta''_{-i});\theta) - V_i(x(\theta_i^{k+1},\theta'_{-i});\theta) + V_i(x(\theta_i^k,\theta'_{-i});\theta)]$$
$$= 0 = t_i^o(\theta_i^{k+1},\theta''_{-i}) - t_i^o(\theta_i^k,\theta''_{-i}) - t_i^o(\theta_i^{k+1},\theta'_{-i}) + t_i^o(\theta_i^k,\theta'_{-i}). \quad (1.67)$$

Therefore, for all $i$ and immediate successors $\theta'_{-i},\theta''_{-i} \in G_p^i$, transfers $t_i^o$ have the smallest one-step supermodularity.

<u>Case 2.</u> If $\theta'_{-i} \in G_p^i$ and $\theta''_{-i} \in G_{p+1}^i$, using equation (1.62) to obtain the supermodularity of $t_i^o$ we get

$$t_i^o(\theta_i^{k+1},\theta''_{-i}) - t_i^o(\theta_i^k,\theta''_{-i}) - t_i^o(\theta_i^{k+1},\theta'_{-i}) + t_i^o(\theta_i^k,\theta'_{-i}) =$$
$$\delta_i(\theta_i^{k+1},\theta_{-i}^{p+1}) - \delta_i(\theta_i^k,\theta_{-i}^{p+1}) - \delta_i(\theta_i^{k+1},\theta_{-i}^p) + \delta_i(\theta_i^k,\theta_{-i}^p) =$$
$$- \min_\theta [V_i(x(\theta_i^{k+1},\theta_{-i}^{p+1});\theta) - V_i(x(\theta_i^k,\theta_{-i}^{p+1});\theta) - V_i(x(\theta_i^{k+1},\theta_{-i}^p);\theta) + V_i(x(\theta_i^k,\theta_{-i}^p);\theta)] =$$
$$- \min_\theta [V_i(x(\theta_i^{k+1},\theta''_{-i});\theta) - V_i(x(\theta_i^k,\theta''_{-i});\theta) - V_i(x(\theta_i^{k+1},\theta'_{-i});\theta) + V_i(x(\theta_i^k,\theta'_{-i});\theta)].$$
$$(1.68)$$

Hence, for all $t_i$ such that $(x,t)$ is supermodular implementable it must hold that:

$$t_i(\theta_i^{k+1},\theta''_{-i}) - t_i(\theta_i^k,\theta''_{-i}) - t_i(\theta_i^{k+1},\theta'_{-i}) + t_i(\theta_i^k,\theta'_{-i}) \geq$$
$$- \min_\theta [V_i(x(\theta_i^{k+1},\theta''_{-i});\theta) - V_i(x(\theta_i^k,\theta''_{-i});\theta) - V_i(x(\theta_i^{k+1},\theta'_{-i});\theta) + V_i(x(\theta_i^k,\theta'_{-i});\theta)]$$
$$= t_i^o(\theta_i^{k+1},\theta''_{-i}) - t_i^o(\theta_i^k,\theta''_{-i}) - t_i^o(\theta_i^{k+1},\theta'_{-i}) + t_i^o(\theta_i^k,\theta'_{-i}). \quad (1.69)$$

Therefore, for all $i$ and immediate successors $\theta'_{-i} \in G_p^i$ and $\theta''_{-i} \in G_{p+1}^i$, transfers $t_i^o$ have the smallest one-step supermodularity.

Cases 1 and 2 allow us to conclude that transfers $t^o$ achieve minimal supermodularities across any pair of immediate successors on $(\Theta_i, \geq_i^1)$ and $(\Theta_{-i}, \geq_{-i})$, as long as $x$ is order reducible.

Step 2. Consider the supermodularity between successive types $\theta_i^k$, $\theta_i^{k+q}$ and $\theta'_{-i} \in G_p^i$, $\theta''_{-i} \in G_{p+m}^i$. For $q = 1$ and $m = 1$ (or $m = 0$) this would reduce to the case of supermodularities between immediate successors considered in Step 1. Using

equation (1.62), we obtain

$$t_i^o(\theta_i^{k+q}, \theta_{-i}'') - t_i^o(\theta_i^k, \theta_{-i}'') - t_i^o(\theta_i^{k+q}, \theta_{-i}') + t_i^o(\theta_i^k, \theta_{-i}') =$$

$$\delta_i(\theta_i^{k+q}, \theta_{-i}^{p+m}) - \delta_i(\theta_i^k, \theta_{-i}^{p+m}) - \delta_i(\theta_i^{k+q}, \theta_{-i}^p) + \delta_i(\theta_i^k, \theta_{-i}^p) =$$

$$- \sum_{l=k}^{k+q-1} \sum_{z=p}^{p+m-1} \min_{\theta \in \Theta} [V_i(x(\theta_i^{l+1}, \theta_{-i}^{z+1}), \theta) - V_i(x(\theta_i^l, \theta_{-i}^{z+1}), \theta)$$

$$- V_i(x(\theta_i^{l+1}, \theta_{-i}^z), \theta) + V_i(x(\theta_i^l, \theta_{-i}^z), \theta)]. \quad (1.70)$$

Hence, the $q, m$-step supermodularity of transfers $t_i^o$ is a sum of all the one-step supermodularities between the groups $G_p$ and $G_{p+m}$. We next show that this sum between the groups is equivalent to a sum of minimal one-step supermodularities on $\Theta_i \times \Theta_{-i}$, all of which need to be minimized for minimal supermodular implementation to hold.

Take a sequence $\theta_i^k, \ldots, \theta_i^{k+q}$ of immediate successors under $\geq_i^1$, and a sequence $\theta_{-i}^1, \ldots, \theta_{-i}^{1+s}$ of immediate successors under $\geq_{-i}$ such that $\theta_{-i}^1 = \theta_{-i}'$ and $\theta_{-i}^{1+s} = \theta_{-i}''$. Since $\theta_{-i}' \in G_p^i$, $\theta_{-i}'' \in G_{p+m}^i$, and $x$ is order reducible, it cannot be that $\theta_{-i}''$ is more that $s$ groups away from $\theta_{-i}'$, i.e. it must be that $s \geq m$.
Case 1. If $m = s$, then

$$t_i^o(\theta_i^{k+q}, \theta_{-i}'') - t_i^o(\theta_i^k, \theta_{-i}'') - t_i^o(\theta_i^{k+q}, \theta_{-i}') + t_i^o(\theta_i^k, \theta_{-i}') = \quad (1.71)$$

$$- \sum_{l=k}^{k+q-1} \sum_{z=p}^{p+m-1} \min_{\theta \in \Theta} [V_i(x(\theta_i^{l+1}, \theta_{-i}^{z+1}), \theta) - V_i(x(\theta_i^l, \theta_{-i}^{z+1}), \theta)$$

$$- V_i(x(\theta_i^{l+1}, \theta_{-i}^z), \theta) + V_i(x(\theta_i^l, \theta_{-i}^z), \theta)] = \quad (1.72)$$

$$- \sum_{l=k}^{k+q-1} \sum_{w=1}^{s-1} \min_{\theta \in \Theta} [V_i(x(\theta_i^{l+1}, \hat{\theta}_{-i}^{w+1}), \theta) - V_i(x(\theta_i^l, \theta_{-i}^{w+1}), \theta)$$

$$- V_i(x(\theta_i^{l+1}, \theta_{-i}^w), \theta) + V_i(x(\theta_i^l, \theta_{-i}^w), \theta)]. \quad (1.73)$$

Since the supermodularity of $V_i$ is equal to

$$\sum_{l=k}^{k+q-1} \sum_{w=1}^{s-1} [V_i(x(\theta_i^{l+1}, \theta_{-i}^{w+1}), \theta) - V_i(x(\theta_i^l, \theta_{-i}^{w+1}), \theta)$$

$$- V_i(x(\theta_i^{l+1}, \theta_{-i}^w), \theta) + V_i(x(\theta_i^l, \theta_{-i}^w), \theta)] \quad (1.74)$$

54

and all of the summands involve one-step supermodularities, it holds that

$$V_i(x((\theta_i^{k+q}, \theta''_{-i}), \theta) - V_i(x(\theta_i^k, \theta''_{-i}), \theta) - V_i(x(\theta_i^{k+q}, \theta'_{-i}), \theta) + V_i(x(\theta_i^k, \theta'_{-i}), \theta)]$$
$$+ [t_i^o(\theta_i^{k+q}, \theta''_{-i}) - t_i^o(\theta_i^k, \theta''_{-i}) - t_i^o(\theta_i^{k+q}, \theta'_{-i}) + t_i^o(\theta_i^k, \theta'_{-i})] \geq 0 \quad (1.75)$$

and the multiple-step supermodularity of $t_i^o$ is the smallest possible, so that all one-steps are minimally supermodular.

<u>Case 2.</u> If $s > m$, it means that $s - m$ immediate successors $\tilde{\theta}''_{-i}$ under $\geq_{-i}$ are in the same category as their immediate predecessors $\tilde{\theta}'_{-i}$ and are disregarded in the sum (1.72). However, note that for all of these successors, it holds that:

$$V_i(x((\theta_i^{k+1}, \tilde{\theta}''_{-i}), \theta) - V_i(x(\theta_i^k, \tilde{\theta}''_{-i}), \theta) - V_i(x(\theta_i^{k+1}, \tilde{\theta}'_{-i}), \theta) + V_i(x(\theta_i^k, \tilde{\theta}'_{-i}), \theta)] = 0$$
$$(1.76)$$

and hence equality between (1.72) and (1.73) prevails. The rest of the argument for this case follows that for case 1.

Steps 1 and 2 prove that transfers $t_i^o$ minimally supermodular implement the decision rule $x$ under the chosen profile of total orders $\{\geq_i^1\}_i$.

# Chapter 2

# Precision of Information in Second Price Common Value Auctions

## 2.1 Introduction

This papers considers the optimal information revelation policy when a privately informed seller has control over the precision of signals that potential bidders receive about the common value of a single object. The objective of the seller is to maximize his payoff, which is defined as expected revenue minus cost of precision. There is a common prior over the value of the object. Bidders form their valuations based on the publicly observed level of precision set by the seller and the privately observed signals they receive.

To our best knowledge, this is the first paper to study the information revelation incentives of the auctioneer in a common value setting where bidders observe conditionally independent signal realizations and precision of information is costly to the seller. The seminal results of Milgrom and Weber (1982) are related to our work, but they are derived in a different informational environment. In a model with affiliated values they show that it is optimal for the seller to reveal as much information as possible about the quality of her product through a public signal observable by all bidders. This holds true irrespective of whether the seller has a high or a low quality product and results in a fully revealing equilibrium. The provision of public information decreases the winner's curse and the informational rent of the winning bidder, thus increasing revenues.

In our environment the conclusions are quite different. The first distinction of our model lies in the fact that the only way the seller can control the amount of information she supplies is by determining the precision of the *private* signals that bidders observe. Therefore, if the seller can commit to her policy before observing the value of the object, the winner's curse is the lowest either for low values of precision or for high values of precision, and the highest for intermediate values of precision. In the presence of any cost to precision, it is hence ex ante optimal to choose the smallest possible level of precision. The second distinction is that in our set-up there are two different channels through which the seller can influence bidders' beliefs. The first channel is the choice and announcement of a precision level which, in the case of no commitment, is a signal of the type of seller. The second channel is stochastic in nature and represents the privately observed signals, the joint distribution of which is determined by the chosen level of precision. The most important contribution of the model is to show that pooling equilibria can be sustained in the presence of cost to precision of information when sellers send conditionally independent private signals. This is in contrast with the full separation results of Milgrom and Weber (1982) and might explain observed the lack of transparency in some markets.

Our model is closely related to the signaling literature, and in particular to the paper by Daley and Green (2012). They consider a market signaling model in which receivers observe both a costly signal as well as a stochastic "grade" that is correlated with the sender's type. The main difference, and an important one at that, is the fact that in their framework the stochastic grades are publicly observable, while in ours they are privately observed. This paper is also related to the literature on information in mechanism design, e.g. Bergemann and Pesendorfer (2007), Eso and Szentes (2007), and Gershkov (2002).

Our paper is organized as follows: Section 2 describes the model. Section 3

addresses the optimal information precision for the case of perfect commitment, while Section 4 contains the characterization of equilibria in the signaling environment. Section 5 addresses some interesting extensions of the model, and Section 6 concludes.

## 2.2 Model

A seller (sender) has a single object for sale, which she values at 0. She privately observes the quality $V$ of the object she is selling, which can be either high (1) or low (0), i.e. $V \in \{0, 1\}$. We will use $V$ to refer to both the quality of the object and the type of seller.

There is a set of $N$ bidders (receivers) each indexed by $i \in \{1, 2, \ldots, N\}$, which share a commonly known prior over the quality of the object $p \equiv \Pr(V = 1) \in (0, 1)$. We will first focus on the case when there are two bidders, i.e. $N = 2$. Each bidder privately observes the realization of a random signal $S_i$ that reveals information about $V$. The signals $S_i$'s are independently distributed conditional on $V$ and can take on the values $\{0, \frac{1}{2}, 1\}$. The seller chooses and publicly announces the precision of signals $\delta \in [0, 1]$, which determines for each $S_i$ the probability of a perfectly informative signal realization. In particular, the $S_i$'s have the following distribution: $\Pr(S_i = V) = \delta$, $\Pr(S_i = \frac{1}{2}) = 1 - \delta$.

We focus on this particular signal structure since it allows for straightforward updating following Bayes' rule. Conditional on observing a signal realization $S_i = 0$ or $S_i = 1$ bidder $i$ knows he has observed the true quality $V$. On the other hand, conditional on observing $S_i = \frac{1}{2}$ (a completely uninformative signal or a "blank page") bidder $i$ continues to assign the prior probabilities associated with each possible value of $V$.

We assume that precision is costly. Any level of precision is equally costly for

both types of sellers, i.e. a seller of a high quality product does not have any cost advantage when compared to a low quality seller. We restrict attention to affine cost functions of the form $C(\delta) = a + b \cdot \delta$ with $a \geq 0$ and $b \geq 0$ and $C(0) = 0$. The objective of the seller is to maximize her payoff, which is equal to expected revenue minus cost of precision.

Each bidder $i$ submits a bid $b_i \in \mathbb{R}_+$. The format of the auction is fixed at a standard second-price sealed-bid auction, in which the highest bidder receives the object and pays the second highest bid. In the event of a tie at the highest bid, each of the highest bidders are awarded the object with equal probability. The utility of each bidder is quasi-linear and given by $u_i = \mathbb{1}(V - b_{-i})$ where $b_{-i} = \max_{j \neq i} b_j$. The indicator function $\mathbb{1}$ takes on the value of 1 if bidder $i$ is awarded the object and 0 otherwise. In the analyses to follow we restrict attention to symmetric equilibrium strategies.

## 2.3   Perfect Commitment

In this section we consider the case when the seller chooses and publicly announces the precision $\delta$ before observing the quality of the product she is selling. We assume that once announced, $\delta$ cannot be changed and the seller is thus perfectly committed to the chosen level of precision. In this case the choice of $\delta$ cannot be interpreted as a signal regarding the quality of the product. Consequently, a bidder would update his prior belief $p$ only if he were to observe a perfectly informative signal $S_i = 0$ or $S_i = 1$. It is a weakly dominant strategy for each player to bid according to the following bidding function:

$$b_i(\delta, S_i) = \begin{cases} S_i & \text{if } S_i = 0, 1 \\ p & \text{if } S_i = \frac{1}{2}. \end{cases}$$

The expected revenue for a high quality seller who chooses a level of precision

$\delta$ is $E[R|V = 1, \delta] = \delta^2 + (1 - \delta^2)p$, while for a low quality seller the expected revenue is given by $E[R|V = 0, \delta] = (1 - \delta)^2 p$. Thus, the ex ante expected seller revenue for any chosen level of precision $\delta$ is:

$$E[R|\delta] = p[1 - 2\delta(1 - p) + 2\delta^2(1 - p)] \tag{2.1}$$

This is a convex function in $\delta$, which achieves a minimum at $\delta = 0.5$.
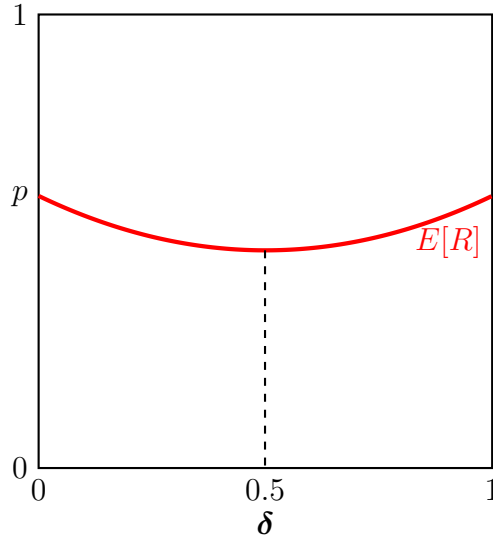


Figure 2.1: Ex Ante Expected Revenue

Therefore, in the presence of any cost to precision it is ex ante optimal for the seller to choose the lowest possible level of precision, i.e. $\delta = 0$. This is the statement of our first result.

PROPOSITION **6.** *Assume the seller can perfectly commit to a precision level $\delta$ before observing the quality of the product she is selling. In the presence of any cost to precision, i.e. $C(\delta) > 0$ for $\delta \neq 0$, the ex ante expected payoff maximizing level of precision is $\delta = 0$:*

$$\arg\max_{\delta}[E[R|\delta] - C(\delta)] = 0$$

60

*Proof* Ex ante expected revenue as given by 2.1 is maximized at $\delta = 0$ or $\delta = 1$. The cost of precision are minimized at $\delta = 0$ whenever $a > 0$ and/or $b > 0$, or in other words whenever there is some strictly positive costs associated with strictly positive levels of precision. Therefore, in the presence of any positive cost to precision, $\delta = 0$ maximizes ex ante expected revenue and minimizes cost. This implies that ex ante expected payoff defined as the difference between ex ante expected revenue and cost associated with precision is maximized at $\delta = 0$. □

It is worthwhile to interpret this result in the context of Milgrom and Weber's linkage principle. Notice that in their analysis the seller discloses information in the form of a public signal, while here the seller chooses the precision or probabilistic informativeness of privately observed signals. Therefore, in their paper, the statistical linkage between a bidder's information and the price he would pay upon winning increases with the release of more or more precise public information, which reduces the winner's curse and increases expected revenue. In contrast, in our case, this linkage is the strongest either when $\delta = 0$, i.e. completely uninformative signals, or for $\delta = 1$, in which case the signals are perfectly informative. This provides the intuition for the above results.

## 2.4   No Commitment: The Signaling Environment

In this section we focus on the informed seller case. In other words, the seller knows the quality of the object that she has for sale at the time when she makes and announces the decision regarding the level of signal precision $\delta$. The choice of $\delta$ can be therefore interpreted as a signal by the buyers.

### 2.4.1 Solution Concept and Beliefs

The solution concept that we use throughout the analysis is that of perfect bayesian equilibrium (PBE). All bidders hold identical beliefs off the equilibrium path. We also consider a refinement on the off-equilibrium-path beliefs (D1) in order to derive sharper predictions and more concrete insights. Buyers use Bayes rule to update their beliefs after any history for which it is possible to do so.

After observing the publicly announced precision $\delta$, buyers update their prior belief to an *interim belief* $\mu(\delta) \equiv \Pr(V = 1|\delta)$. Notice that this interim belief is common to all buyers. In addition, each buyer privately observes a signal $S_i(\delta, V)$ and subsequently updates the interim belief to a *final belief*

$$\pi_i[\mu(\delta), S_i(\delta, V)] = \pi_i(\delta, V) \equiv \Pr(V = 1|\delta, S_i).$$

The final beliefs are specific to the individual bidder as they are also based on the privately observed signals $S_i$. Furthermore, due to the distributional assumptions on the $S_i$'s, the final belief of bidder $i$ is either equal to 0 or 1 (in case a perfectly informative signal $S_i$ is observed) or it remains equal to the interim belief (when the signal is completely uninformative, i.e. $S_i = \frac{1}{2}$). More specifically, the updating works as follows: $\pi_i(\mu(\delta), S_i = 1) = 1$, $\pi_i(\mu(\delta), S_i = 0) = 0$, and $\pi_i(\mu(\delta), S_i = \frac{1}{2}) = \mu(\delta)$.

Since the update based on the privately observed signals is purely statistical, the interim belief is all a seller of a certain type needs in order to calculate her expected revenue. Indeed, the joint distribution of the private $S_i$ signals determines how the common interim belief $\mu(\cdot)$ is going to be updated to the individual final belief $\pi_i(\cdot)$. Therefore, the choice of precision level $\delta$ determines, on the one hand, the *equilibrium* interim belief $\mu(\cdot)$, and on the other, the *statistical* distribution over private signals $S_i$ conditional on quality $V$. The combination of these two channels

(equilibrium and statistical) determines the joint distribution over final beliefs $\pi_i$ for a given quality $V$.

We will now proceed by characterizing the set of PBE of this signaling game. There are pooling, separating, and partial pooling[1] equilibria. The set of PBE is indeed large, as there is a lot of flexibility when specifying off-equilibrium-path beliefs. In the following, we restrict attention to equilibria satisfying the intuitive criterion of Cho and Kreps (1987).

### 2.4.2  Pooling Equilibria

Pooling equilibria are characterized by both types of seller choosing the same level of precision $\tilde{\delta} \in [0, \bar{\delta}]$. The upper bound on the set of pooling equilibria $\bar{\delta}$ is determined from the combination of incentive compatibility constraints for the different types of seller.

In the case of pooling the interim beliefs are $\mu(\delta) = p$ for $\delta \geq \tilde{\delta}$ and $\mu(\delta) = 0$ for $\delta < \tilde{\delta}$. For any pooling equilibrium with $\tilde{\delta}$ it is a symmetric weakly dominant strategy to bid according to:

$$b_i(\tilde{\delta}, S_i) = \begin{cases} S_i & \text{if } S_i = 0, 1 \\ p & \text{if } S_i = \frac{1}{2}. \end{cases}$$

The incentive compatibility constraint for the low type seller in a pooling equilibrium (ICLp) addresses only downward deviations. Upward deviations to higher $\delta$'s are never profitable for this type of seller as they increase the probability of buyers realizing that the true quality is $V = 0$ which lowers expected revenue while at the same time increases the cost. The most profitable downward deviations is to $\delta = 0$:

$$(1 - \tilde{\delta})^2 p - a - b\tilde{\delta} \geq 0 \tag{ICLp}$$

---

[1]Partial pooling equilibria are characterized by the high type seller choosing a precision level $\hat{\delta}$ and the low type seller mixing between $\delta = 0$ and $\delta = \hat{\delta}$.

This condiction pins down a value $\bar{\delta}_L = 1 + \frac{b - \sqrt{b^2 + 4p(a+b)}}{2p}$, which represents the highest level of precision that a low type seller would choose in a pooling equilibrium. For $\bar{\delta}_L$ to be positive it needs to hold that $p \geq a$.

For the high type seller downward deviations are never profitable. The incentive compatibility constraint for the high type seller in any pooling equilibrium (ICHp) therefore guarantees that no upward deviations are profitable. In particular, for all $\tilde{\delta} > 0$ it has to hold that:

$$\tilde{\delta}^2 + (1 - \tilde{\delta}^2)p - a - b\tilde{\delta} \geq 1 - a - b \qquad \text{(ICHp)}$$

On the graph below, condition (ICHp) can be interpreted as the vertical distance between the green and the red line being bigger at any pooling equilibrium $\tilde{\delta} > 0$ than the distance between these two lines at $\delta = 1$. Condition (ICHp) holds for values of $\delta$ smaller than $\bar{\delta}_H = \frac{b}{1-p} - 1$ and at $\delta = 1$. Therefore, the upper bound on the range of sustainable pooling equilibria is determined as $\bar{\delta} = \max\left\{0, \min\{\bar{\delta}_L, \bar{\delta}_H\}\right\}$.

Notice that the lowest cost pooling equilibrium occurs at $\tilde{\delta} = 0$. This equilibrium is the most easily sustainable pooling equilibrium in terms of the incentive compatibility constraint for the high type seller, because the (ICHp) constraint for it is simply $p \geq 1 - a - b$. Therefore, this lowest cost pooling equilibrium is the last one that ceases to exist as we decrease the cost parameters. That is, if the sum of the fixed cost and the marginal cost parameters, $(a + b)$ falls below a certain level determined by the prior, $(1 - p)$, the lowest cost pooling equilibrium is no longer sustainable and neither are any other pooling equilibria. These conclusions are summarized in the next two propositions.

PROPOSITION **7.** *A necessary and sufficient condition for the existence of a pooling equilibrium with $\tilde{\delta} = 0$ is $p \geq 1 - a - b$.*
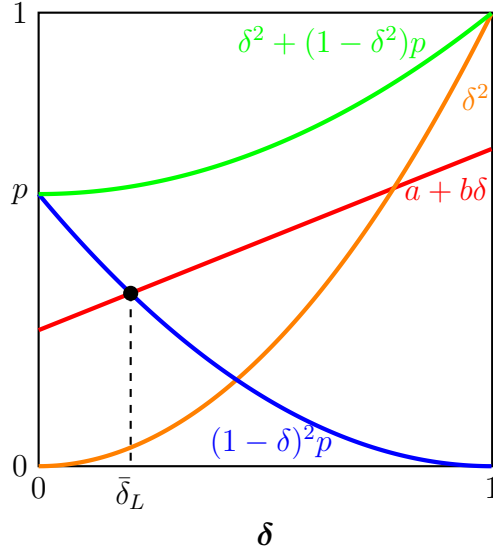
Figure 2.2: Pooling Equilibria (Constraints and Bounds)

*Proof* The result is derived from (ICHp) and (ICLp) when $\tilde{\delta} = 0$.

PROPOSITION **8.** *A necessary and sufficient condition for the existence of a strictly positive pooling equilibrium ($\tilde{\delta} > 0$) is $p \geq \max\{a, 1 - b\}$.*

*Proof* The result of the proposition follows in a straightforward way from the preceding incentive compatibility constraints. $\qquad\qquad\square$

In the presence of cost to precision, pooling equilibria can be thus sustained for a wide range of parameter values. This is in sharp contrast with the fully separating results of Milgrom and Weber (1982), according to which the high type seller always reveals all the information she has; in our environment that corresponds to the high type always choosing $\delta = 1$, and the low type choosing $\delta = 0$. Another point worth emphasizing is that our results are not driven by differences in the *explicit* costs of precision between the high and the low quality seller, as the parameters of the cost

function $a$ and $b$ are the same across types. Instead, the *implicit* costs and benefits associated with increasing precision are the driving force behind the equilibria here. The implicit costs to the low quality seller of increasing $\delta$ come with the increased probability of having bidders observe her true type, while to the high quality seller increasing $\delta$ is beneficial for the exact same reason.

It is instructive to derive the expression for the ex ante expected seller payoff in the lowest cost pooling equilibrium, so that we can later compare that to the corresponding expression in the best separating equilibrium. When $\tilde{\delta} = 0$ the expected revenue is simply equal to the prior, while the cost of precision is zero. Thus, the expected seller payoff in the lowest cost pooling equilibrium is:

$$E[P|\tilde{\delta} = 0] = p. \tag{2.2}$$

### 2.4.3   Separating Equilibria

In a separating equilibrium the high type seller chooses a level of precision $\delta^* \in [\underline{\delta}, 1]$, while the low type seller chooses $\delta = 0$. The interim beliefs in a separating equilibrium are $\mu(\delta) = 1$ for $\delta \geq \delta^*$ and $\mu(\delta) = 0$ for $\delta < \delta^*$. In any separating equilibrium characterized by $\delta^*$ it is a weakly dominant strategy to bid according to: it is a weakly dominant strategy for each player to bid according to the following bidding function:

$$b_i(\delta, S_i) = \begin{cases} S_i & \text{if } S_i = 0, 1 \\ 1 & \text{if } S_i = \frac{1}{2} \text{ and } \delta = \delta^* \\ 0 & \text{if } S_i = \frac{1}{2} \text{ and } \delta = 0. \end{cases}$$

The lower bound $\underline{\delta}$ on the set of precision levels chosen by the high type seller in any separating equilibrium is determined by the incentive compatibility constraints for the low type seller, as the ones for the high type are always satisfied for any $\delta^*$. The incentive compatibility constraint for the low type seller in any separating equiilibrium (ICLs) ensures that an upward deviation to $\delta^*$ is not profitable:

66

$$0 \geq (1 - \delta^*)^2 - a - b\delta^* \tag{ICLs}$$

On the graph below this condition can be interpreted as the purple curve being below the red cost line in any separating equilibrium. The point of intersection between these two curves determines the value of $\underline{\delta}$, which algebraically is derived as a solution to the (ICLs) condition holding with equality: $\underline{\delta} = 1 + \frac{b - \sqrt{b^2 + 4(a+b)}}{2}$.

PROPOSITION **9.** *A necessary condition for the existence of a separating equilibrium is $a < 1$.*

*Proof* For the existence of a separating equilibrium we need to have $\underline{\delta} > 0$. From the formula for $\underline{\delta}$ given above we can derive this is equivalent to $a < 1$, i.e. the magnitude of the fixed cost parameter needs to be smaller than 1. $\qquad \square$
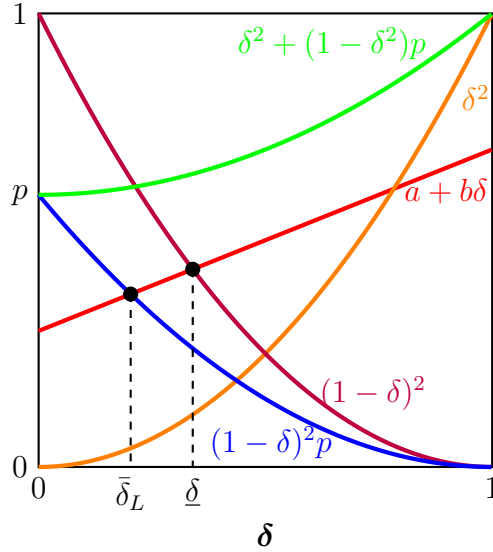


Figure 2.3: Separating Equilibria (Constraints and Bounds)

Therefore, from the perspective of the seller, the best separating equilibrium is the one associated with the lowest cost, i.e. $\delta^* = \underline{\delta}$. The ex ante expected seller

67

payoff in this lowest cost separating equilibrium is

$$E[P|\delta = \underline{\delta}] = p[1 - a - b\underline{\delta}] = p - p\left(a + b + \frac{b^2 - b\sqrt{b^2 + 4(a+b)}}{2}\right). \qquad (2.3)$$

### 2.4.4 Partial Pooling Equilibria

The set of partial pooling equilibria is characterized by the high type seller always choosing a precision level $\hat{\delta}$, while the low type seller mixes with probability $1 - \lambda$ on $\delta = 0$ and probability $\lambda$ on $\hat{\delta}$. Hence, the interim beliefs in partial pooling equilibria are given by $\mu(\delta) = \frac{p}{p + \lambda(1-p)}$ for $\delta \geq \hat{\delta}$ and $\mu(\delta) = 0$ for $\delta < \hat{\delta}$. The set of precision levels $\hat{\delta}$ which constitute partial pooling equilibria is found by solving the indifference condition for the low type seller (ICLpp):

$$0 = (1 - \hat{\delta})^2 \frac{p}{p + \lambda(1-p)} - a - b\hat{\delta} \qquad \text{(ICLpp)}$$

which also determines the mixing probability $\lambda$:

$$\lambda = \frac{p}{1 - p}\left[\frac{(1 - \hat{\delta})^2}{a + b\hat{\delta}} - 1\right]. \qquad (2.4)$$

The incentive compatibility constraint of the high type seller

$$\hat{\delta}^2 + (1 - \hat{\delta}^2)\frac{p}{p + \lambda(1-p)} - a - b\hat{\delta} \geq 0 \qquad \text{(ICHpp)}$$

is always satisfied as long as (ICLpp) holds.

The fact that $\lambda$ and $\hat{\delta}$ are inversely related becomes apparent from equation (ICLpp). An increase (decrease) in $\hat{\delta}$ leads to an unambiguous decrease (increase) of the righthand side of the equation. In order to preserve the equality, $\lambda$ has to decrease (increase). We can thus determine the bounds on the set of partial pooling equilibria by considering the extreme values that $\lambda$ can take on. For $\lambda = 1$ we obtain the lower

bound on the set of partial pooling equilibria, while for $\lambda = 0$ we obtain the upper bound on this set. These are respectively given by $\bar{\delta}_L$ and $\underline{\delta}$ in Figure 2.3. Therefore the set of partial pooling equilibria— $(\bar{\delta}_L, \underline{\delta})$ — is always in between the set of pooling equilibria and the set of separating equilibria.

### 2.4.5  Ex Ante Welfare Analysis

In this section we analyze the highest ex ante expected seller payoff in any equilibrium as a function of the cost parameters of the model. We know from the preceding sections that the lowest-cost pooling equilibrium $\tilde{\delta} = 0$ is the overall best in terms of ex ante seller payoff. Also, we observed that it is the most "robust" of the pooling equilibria, as it is the last pooling equilibrium to cease to exist as the cost parameters of the model change. If we are in the case when the fixed cost to precision is smaller than the prior $(a \leq p)$, as the sum of the fixed plus variable cost parameters, $(a + b)$ becomes smaller, the set of sustainable pooling equilibria shrinks. Once this sum falls below the value $(1 - p)$, pooling equilibria can no longer be sustained and they cease to exist. The overall best equilibrium in terms of ex ante seller payoff then becomes the lowest cost separating equilibrium $\delta^* = \underline{\delta}$, which is strictly worse than the best pooling equilibrium. In the following graph we plot ex ante expected equilibrium payoff for the seller as a function of $(a + b)$ for fixed $a \in [0, p]$ and variable marginal cost parameter $b$:

The analysis and the graph demonstrate that in fact ex ante seller welfare may be higher in the case of larger cost associated with precision. For given fixed cost $a \leq p$, if the marginal cost $b$ associated with precision are lower than $1 - p - a$, the pooling equilibria cease to exist. The reason is that in this case the cost are low enough that the temptation to separate for the high type seller is really strong and even the incentive compatibility constraints for the most robust pooling equilibrium
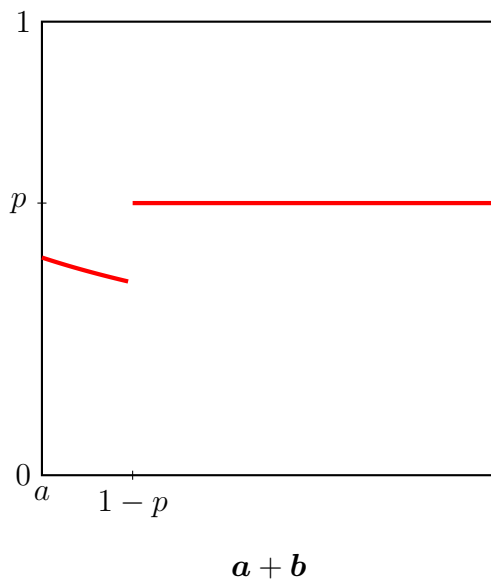
69

Figure 2.4: Ex Ante Expected Seller Payoff

$(\tilde{\delta} = 0)$ can no longer be satisfied. Therefore, if the costs associated with precision are too low, the lowest cost separating equilibrium $\delta^* = \underline{\delta}$ is the best a seller can do and the associated ex ante seller payoff is lower than $p$.

## 2.5 Extensions

There are a few extensions of the model that we are interested in pursuing. In this section we outline the basic intuition and approach to incorporating these into the basic framework presented above.

### 2.5.1 Arbitrary Number of Bidders

Allowing for an arbitrary number of bidders $N$ is an important aspect to consider. There are a couple of interacting effects that occur when we increase the number of bidders. Let us consider the pooling equilibrium constraints for the low

70

and high type seller for the general case of $N$ bidders:

$$\left(1 - \tilde{\delta}^N - N\tilde{\delta}^{N-1}(1-\tilde{\delta})\right)p - a - b\tilde{\delta} \geq 0 \qquad \text{(ICLp)}$$

$$\left((1-\tilde{\delta})^N + N\tilde{\delta}(1-\tilde{\delta})^{N-1}\right)(p-1) + 1 - a - b\tilde{\delta} \geq 1 - a - b. \qquad \text{(ICHp)}$$

Hence, for any given $\tilde{\delta}$, as $N$ increases, the pooling equilibrium incentive constraints for both types of sellers change non-monotonically. However, there is a cutoff value for $N$ beyond which any further increase in the number of bidders makes both the (ICLp) and the (ICHp) constraints easier to satisfy. Therefore, as $N$ increases beyond that cutoff value, the range of pooling equilibria expands.

On the other hand, the separating equilibrium incentive compatibility constraint of the low type seller for arbitrary $N$ is:

$$0 \geq \left(1 - \delta^{*N} - N\delta^{*N-1}(1-\delta^*)\right) - a - b\delta^* \qquad \text{(ICLs)}$$

This constraint becomes harder to satisfy as $N$ increases beyond the cutoff level mentioned above. Hence, the range of separating equilibria shrinks.

### 2.5.2 Two-Dimensional Types

Another interesting and challenging extension is allowing for the presence of both a common and a private value component to buyer valuations, in which case buyers have two-dimensional types. In particular we are interested in how robust the current results derived in the pure common value setting are to the introduction of a private value component. We consider the case when buyer valuations are a convex linear combination of a common value component $V \in \{0, 1\}$ and a private value component $t_i \in \{0, 1\}$ as given by:

$$\alpha t_i + (1 - \alpha)V \qquad (2.5)$$

where $\alpha \in (0, 1)$ is the weight assigned to the common value. We define the information structure for this environment in the following way: with probability $\delta$ each buyer observes both the his true private value $t_i$ and the true common value $V$, and with probability $(1-\delta)$ he observes nothing. There is a common prior over the private component $q \equiv \Pr(t_i = 1) \in (0, 1)$ for $i = 1, 2$ and a common prior over the common component $p \equiv \Pr(V = 1) \in (0, 1)$.

We find out that for $\alpha > \frac{1}{2}$ there exist symmetric equilibrium bidding strategies which are also truthful: informed bidders bid their valuations, while uninformed ones bid their expected valuations. For this case the analysis of the signaling game is very similar to the one presented for the pure common value environment.

However, for the case when $\alpha < \frac{1}{2}$ we run into issues of non-existence of symmetric equilibria, as pointed out by Jackson (2009). His results are actually stronger, showing non-existence of equilibria not only in symmetric, but more generally in undominated strategies, when buyers have two-dimensional types. Nonetheless, in our environment we are able to characterize an asymmetric equilibrium in undominated strategies, as our information structure is slightly different than Jackson's. The two-bidder asymmetric equilibrium is depicted in Figure 2.5 in terms of the best response correspondences. The red sets are the best responses of bidder 1, while the blue ones are bidder 2's best responses. The asymmetric equilibria are given by the overlap of the two correspondences. With these asymmetric equilibrium bidding strategies, we could theoretically proceed with the analysis of the signaling model.

## 2.6    Conlusion

Our motivation for this project was to understand how the choice of information structure affects the equilibria and the ex ante seller welfare in a pure common value auction. We use a simple model with two bidders and a parameterized infor-
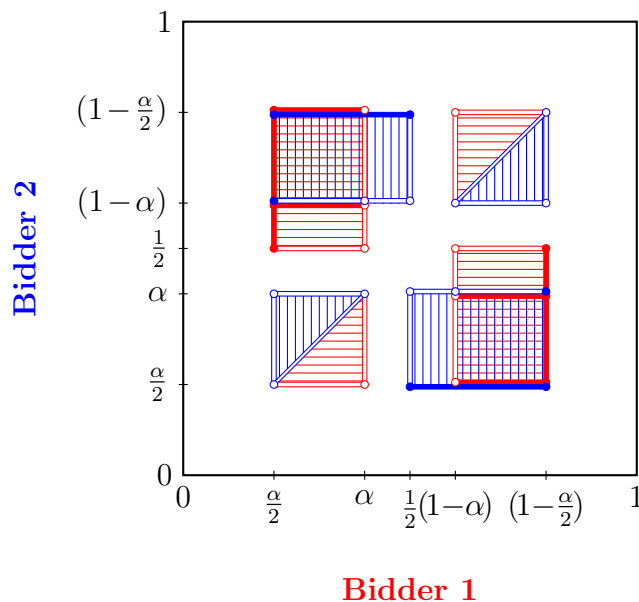
Figure 2.5: Asymmetric Equilibria in Undominated Strategies ($\alpha < \frac{1}{2}$)

mation structure that allows for easy updating, in order to derive stylized results in this environment. We believe that the resulting signaling model provides for an interesting analysis and results, in its own right.

While this project has been mainly driven by theoretical aspects and results, It is important to point out a number of limitation that are inherent to the applicability of our analysis and results to the design of auctions in practice. First of all, we are taking the format of the auction as fixed and we only look at the effects that the choice of information structure has on the equilibrium and seller welfare in a second price auction. The question of how the choice of information structure and the choice of an optimal mechanism design interact when chosen simultaneously is certainly an interesting one and constitutes a research venue we would like pursue. Another strong assumption we are making is the fact that the seller can perfectly control the precision of information. This certainly need not be the case in reality, as the seller might be

73

subject to significant constraints in the choice of information structure. Moreover, the potential bidders might have some prior information about the quality of the good, which is also assumed away here. These are all potential extensions that would make the current set-up more realistic and relevant for real-world applications.

# Chapter 3

# Information Design

## 3.1 Introduction

In many economic and social settings one person or institution communicates with multiple interacting parties. In courts, a prosecutor presents the results of her investigation to a jury consisting of several members. In advertising, a company chooses how much and what type of information to reveal about its new product to target different groups of customers through samples, demo versions, information brochures. In politics, election platforms are designed to appeal to constituents, government officials, leaders of other countries. In financial markets, a firm discloses information about its profitability that is relevant to both shareholders and competitors. In economic policy, the Fed releases information about its stimulus campaign, which affects the economic outlook of consumers, as well as domestic and foreign investors.

These are but a few settings of economic importance that provide context for our general questions: What is the optimal mode of information transmission between a self-interested sender (designer) and a group of interacting receivers (agents)? If agents are rational Bayesian players, can the designer select the information structure in a way that makes them play an equilibrium profile most beneficial to her? Is it always optimal for the sender to send a public message observed by all receivers? Or is it sometimes optimal to send privately observed signals? If so, what is the optimal degree of correlation between these private signals? Further, when is the optimal information structure symmetric, and when is it optimal to design an asymmetric

information structure? We consider this the subject of *information design*.

Consider a general environment with multiple interacting agents who choose within a set of possible actions. Their payoffs are determined by their own action, the actions of their opponents, and the realization of a payoff relevant state with a commonly known prior distribution. We refer to this as the *basic game*. In order to analyze the strategic interactions in this setting, we need to also specify what the agents believe about the payoff state, what they believe about their opponents' beliefs, and so on. This is captured by the *information structure*. Consider a designer who has preferences over the payoff state and the actions taken by the agents. Mechanism design takes the information structure as given and modifies the basic game so that the agents achieve the designer's desired objective in equilibrium. In contrast, information design takes the basic game as given and imposes the information structure which maximizes the designer's objective in equilibrium.

We study the general problem of a self-interested designer communicating with multiple agents engaged in a strategic interaction. Before observing the state of the world, the designer chooses the information structure which maximizes her objective in expectation. The designer's objective is an arbitrary function of the state and the agents' actions. The choice of information structure can be viewed as a choice of joint distributions over signals conditional on different states. Once the designer chooses the information structure, it becomes common knowledge. The agents then observe the signal realizations and formulate their beliefs about the state of the world as well as their higher order beliefs. After this, they take actions which affect their own, their opponent's, and the designer's payoffs.

The main contribution of this paper is laying out the methodology of information design in finite settings. The general problem is that of engineering the information structure which for the given basic game supports a Bayes Nash equilibrium

76

that maximizes the designer's objective in expectation. In order to do that, we need to first characterize the set of all Bayes Nash equilibria for all possible information structures. This seems like a daunting task, especially in view of the many different beliefs and higher order beliefs we would need to keep track of. Bergemann and Morris [9] provide a tool that allows us to accomplish this task. They introduce a definition of *Bayes correlated equilibrium* under which we show we can characterize the set of all Bayes Nash equilibria associated with all possible information structures for given basic game by characterizing the set of Bayes correlated equilibria when agents have no information but their prior. By using this concept of correlated equilibrium, we can characterize the set of all Bayes Nash equilibria without explicitly using information structures. Then we maximize the designer's objective function over this set to find the optimal Bayes Nash equilibrium. After that, we back out the information structure which supports it as a Bayes Nash equilibrium for the given basic game.

We apply the general methodology outlined above to a class of symmetric problems with two agents, two actions and two states, for which we are able to derive crisp results and conclusions. We work with a parameterized basic game, which is broad enough to capture many different interactions. Moreover, the parameterization allows for comparative statics with respect to degree of strategic complementarity and substitutability between agents and between each agent and the state. To the best of our knowledge, this is the first application to consider arbitrary objective functions without any a priori assumptions on the form of the information structure.

We provide a complete characterization of the optimal information structure in the symmetric binary environment. The characterization encompasses all possible designer objective functions. The optimal information structure is a function of the underlying game parameters and of the designer's objective. Not surprisingly, when the preferences of the designer and the agents are completely aligned, full information

77

revelation is optimal. However, we also find that making preferences more aligned may in fact decrease the optimal degree of information transmission. This contrasts with results from the literature on cheap talk without commitment.

For the symmetric binary setting our results demonstrate that in almost all of the cases the designer benefits from information design as opposed to revealing no information and letting the agents interact under their prior beliefs. We further show that conditionally independent private signals are never optimal, irrespective of the designer's objective function. Additionally, we ask the question of when modifying the payoffs may be beneficial to the information designer. We obtain clear-cut answers for some of the parameter values. This analysis can be viewed both as comparative statics with respect to the underlying game or as a joint mechanism-design/information-design perspective. Finally, we discuss important extensions that can be addressed in out framework.

Several assumptions are crucial to our model and analysis. The first one is that the designer chooses the information structure before observing the state of the world and is able to perfectly commit to it.[1] The second one is that once the signal realizations occur, the designer cannot change or obfuscate them. Therefore, the agents know that what they observe are undistorted signal realizations from the commonly known conditional distributions. This ensures that they can update their beliefs without considerations of the designer's incentive compatibility constraints. Third, we abstract away from any communication between the agents. In certain instances, this is a reasonable and realistic assumption. However, in other cases, it might be strategically beneficial for the agents to reveal their signals to each other. We provide some discussion regarding these issues and other possible extensions after

---

[1] In Section 1.C Kamenica and Gentzkow [39] provide an excellent discussion on why this assumption is in fact not as restrictive as it may appear at first.

presenting our main results.

The remainder of the paper is organized as follows. Section 3.2 provides an overview of the related literature, followed by the motivating example presented in Section 3.3. Section 3.4 introduces the framework and outlines the general approach to information design. In Section 3.5 we apply the general approach to a particular tractable environment. We provide a complete analysis and characterization of the optimal information structure in the symmetric binary case. Section 3.6 presents some important extensions, as well as a discussion of how these can be incorporated into the model. Section 3.7 concludes with some directions for future research. All proofs are relegated to the Appendix.

## 3.2 Literature Review

This paper is related to the literature on cheap talk communication. The cheap-talk framework analyzes the optimal information structure when the sender knows the realized state of the world and can send costless, non-verifiable messages. Alternatively, it can be viewed as the case when the designer cannot credibly commit to the ex ante chosen information structure and abide by it once the state of the world has been realized. Most related to our framework are the papers by Farrell and Gibbons [29] and Goltsman and Pavlov [31], which extend the cheap talk model of Crawford and Sobel [20] to an environment with two receivers/audiences. They study the impact of costless, non-verifiable claims on the beliefs and therefore the actions of the receivers, which in turn affect the utility of both the sender and the receivers.

There are two significant differences between these papers and ours. First, in our environment, the sender has full commitment power and credibly chooses the information structure, a collection of signal distributions conditional on the state, before the state of the world has been realized. Second, the receivers in the cheap

talk literature are independent decision makers and their payoffs depend only on their own action and the state of the world. Therefore, if the sender were to communicate via privately observed signals, the problem reduces to solving the single receiver case individually for each of the receivers. In our framework, in contrast, the receivers are involved in a strategic interaction with each other, i.e. they play a game. In this case, even if the sender were to communicate via private signals, the information structure affects the higher order beliefs of the agents, which in turn impact the equilibrium actions.

This paper is also closely related to the literature on Bayesian persuasion, which is sometimes referred to as *cheap talk with commitment*. A pivotal paper in that literature is Kamenica and Gentzkow [39], which is equivalent to information design with one agent. They characterize the optimal signal for any given set of preferences and initial beliefs with techniques from convex analysis. However, the tools used by Kamenica and Gentzkow [39] are not sufficient to address the question in an environment with multiple interacting receivers, as the authors themselves point out: *"There is an important third class of multiple-receiver models, however, where our results do not extend easily: those where the receivers care about each other's actions and Sender can send private signals to individual receivers.*[2]*"*

We use a definition of Bayes correlated equilibrium proposed by Bergemann and Morris [9] to answer this open question and show how things differ in the multiple-interacting-receivers environment. We suggest a general approach to the optimal design of information structures, which can be applied in very general environments. We also provide insights regarding the characterization of optimal information structures for different properties of the designer's objective function and of the underlying game

---

[2]See Kamenica and Gentzkow [39], p. 2609.

played by the interacting agents. While Bergemann and Morris [9] provide the tool that enables our analysis, we use it for very different purposes. They focus on characterizing the set of possible Bayes Nash equilibrium outcomes that can arise when players have observed at least a certain level of information and potentially more. They further describe a partial order on information structures under which the size of the equilibrium set varies monotonically. There is no "designer" in their paper, who chooses the information structure with her objective maximization in mind.

A paper by Wang [58] also examines the question of "Bayesian persuasion with multiple receivers". She looks at a specific voting environment, in which the sender has a state-independent utility with a preference for the same alternative. Moreover, she only allows for conditionally independent private signals or purely public signals and compares these two structures. We, on the other hand, allow for a general form of the designer's objective and impose no a priory assumptions on the types of information structures we consider. Public and conditionally independent private signals are special cases contained in our specification. More importantly, we show that restricting attention to these two special categories of persuasion mechanisms is not without loss of generality, as the optimal information structure does not always belong to one of them.

Eliaz and Forges [27] consider a specific environment in which a principal chooses what information to reveal to two symmetric agents whose actions are strategic substitutes. In their framework, the disclosure policy is restricted to verifiable evidence where the sender reports the set of possible states and must include the true state. The sender can control the precision of information by controlling the number of elements she includes in that set. In this setting the authors find that when the sender can commit to a disclosure policy before observing the state, it is optimal to reveal the state perfectly to one agent and disclose nothing to the other. However,

this result crucially relies on the specific designer objective function they look at, the strategic substitutes assumption on the players' actions and the hard evidence assumption. The main part of the analysis in Eliaz and Forges [27] however deals with the case of an informed principal who chooses the information disclosure in the absence of commitment, which is different from our framework. Moreover, their disclosure policy is always constrained to include the true state, while we impose no such restriction. Further, we allow for the agents' actions to be both strategic substitutes and strategic complements.

There is an extensive list of papers studying the comparison of information structures in strategic interactions: Bergemann and Morris [9], Gossner [32], Lehrer, Rosenberg and Shmaya [41] and [42], Peski [53], etc. Closest to ours is Lehrer, Rosenberg and Shmaya [41]. They restrict attention to symmetric games of common interest and rank information structures according to highest player payoffs they induce under different solution concepts. In contrast, we characterize the optimal information structure under Bayes Nash equilibrium and in view of the designer's welfare rather than the agents' equilibrium payoffs.

A number of papers analyze the equilibrium behavior and socially optimal use of information in a tractable class of environments with quadratic payoffs and a normally distributed state of the world (Angeletos and Pavan [2], Bergemann and Morris [10], Bergemann et al. [8]). These papers assume a specific information structure under which each player observes two normally distributed signals: a public signal common to all players and a conditionally independent signal that is privately observed. They characterize the equilibrium use of information and compare that to some efficiency benchmark. In contrast, we do not assume a particular information structure a priori. Our focus is the reverse-engineering aspect of the problem, which concerns the choice of an information structure that will decentralize the most desir-

able distribution over actions and states of the world as a Bayes Nash equilibrium. We are also interested in how this optimal choice changes with the designer's objective function, which is not necessarily socially optimal.

## 3.3  Motivating Example

Consider a prosecutor who conducts an investigation and reports the outcomes to a jury.[3] The prosecutor's objective is to convince the jury that the defendant is guilty and to achieve conviction. She chooses the investigation process and is obligated by law to fully and truthfully report the outcomes to the jury. The choice of investigation process can be viewed as the prosecutor's decisions regarding which witnesses to subpoena, what questions to ask them, which forensic and other tests to order, how to structure her arguments, etc. If the defendant is guilty, then choosing a more informative investigation will tend to help the prosecutor's case and increase the likelihood of conviction. However, if the defendant is innocent, a more informative investigation will impede the prosecutor's case. The question we focus on is whether the prosecutor can gain by choosing the investigation process optimally, in a way that maximizes the overall probability of conviction by a jury consisting of rational Bayesian agents.

To formalize the example, suppose the jury consists of two members, indexed by $i$ and $j$. There are two states of the world: the defendant is either innocent ($\theta_0$) or guilty ($\theta_1$). The prosecutor (designer) and jurors (agents) share a common prior belief, which assigns probability to the defendant being innocent 70 percent of the time, $\Pr(\theta_0) = 0.7$, and guilty 30 percent of the time, $\Pr(\theta_1) = 0.3$. The jurors get

---

[3]This is a multiple-agent version of the prosecutor-judge example of Kamenica and Gentzkow [39]. In fact, we purposefully assume the same prior distribution and prosecutor objective, which allows for direct comparisons with the single receiver case.

utility from choosing the just action: vote to acquit ($a_0$) when innocent and vote to convict ($a_1$) when guilty. Let us assume that unanimity of the jurors' votes is required for a verdict to be reached. If the votes are not unanimous, the case is declared a mistrial due to a deadlocked jury. The payoffs of the jury members are given by the following matrix:

| $\theta = \theta_0$ | $a_0$ | $a_1$ |
|---|---|---|
| $a_0$ | $2,2$ | $1,0$ |
| $a_1$ | $0,1$ | $0,0$ |

| $\theta = \theta_1$ | $a_0$ | $a_1$ |
|---|---|---|
| $a_0$ | $0,0$ | $0,1$ |
| $a_1$ | $1,0$ | $2,2$ |

Each juror receives payoff of at least 1 if he chooses the just vote irrespective of what the other juror does. If the other juror votes justly as well, the payoff is increased to 2, since then a just verdict is reached. Whenever a juror votes unjustly (to convict when innocent and to acquit when guilty) he gets a payoff of 0 irrespective of what the other juror does. The objective of the prosecutor is to achieve a conviction, irrespective of the state. Her utility function is thus given by

$$V(a^i, a^j, \theta) = \begin{cases} 1 & \text{if } a^i = a^j = a_1 \\ 0 & \text{otherwise.} \end{cases}$$

The choice of investigation can be formally represented by conditional distributions $\pi(\cdot|\theta_0)$ and $\pi(\cdot|\theta_1)$ over a set of signal relaizations. The prosecutor chooses $\pi$, which then becomes common knowledge, and the jury observes the undistorted signal realizations from the investigation. If the prosecutor chooses a completely uninformative investigation or equivalently if she chooses not to conduct one, then both jurors will vote to acquit. This is their default action profile since innocence is more likely than guilt. The prosecutor will in turn receive a certain payoff of $V(a_0, a_0) = 0$. At the other extreme, if she were to choose a completely informative investigation process, i.e. one that reveals the state perfectly, the jurors will both vote for conviction only when the defendant is indeed guilty. This happens 30 percent of the time and results in an expected payoff of 0.3 for the prosecutor.

However, the prosecutor can do better than that. The optimal investigation is in fact given by the following signal structure:

| $\theta = \theta_0$ | $t_0$ | $t_1$ |
|---|---|---|
| $t_0$ | $1/7$ | $0$ |
| $t_1$ | $0$ | $6/7$ |

| $\theta = \theta_1$ | $t_0$ | $t_1$ |
|---|---|---|
| $t_0$ | $0$ | $0$ |
| $t_1$ | $0$ | $1$ |

which is asymmetric with respect to the state. Under this information structure it is a BNE for each jury member to vote to acquit $(a_0)$ when he observes $t_0$ and to convict $(a_1)$ when he observes $t_1$. The incentive BNE compatibility constraints are as follows:

$$u(a_0|t_0) = 1 \cdot 2 = 2 > u(a_1|t_0) = 1 \cdot 0 = 0$$

and

$$u(a_1|t_1) = \frac{2}{3} \cdot 0 + \frac{1}{3} \cdot 2 = \frac{2}{3} = u(a_0|t_1) = \frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 0 = \frac{2}{3}.$$

Under this information structure and BNE, the expected value of the prosecutor's objective function is 0.9. The jury members know that 70 percent of the defendants are innocent, yet they end up convicting 90 percent of them. They are completely aware that the investigation was chosen in a way to maximize the probability of conviction; yet they react in a rational Bayesian way given the signal realizations they observe.

We observe the same two characteristics here that Kamenica and Gentzkow [39] derive for the optimal information structure in the single receiver case. First, when each juror votes to acquit, the prosecutor's least favorite option, he is certain that the defendant is innocent. In other words, we have $\pi(t_0, t_0|\theta_1) = \pi(t_0, t_1|\theta_1) = \pi(t_1, t_0|\theta_1) = 0$. If these probabilities were positive, the prosecutor could decrease them in favor of increasing $\pi(t_1, t_1|\theta_1)$. This will increase both the marginal probability of the signal realization $(t_1, t_1)$ and the willingness of each juror to convict

when observing $t_1$. Both of these effects increase the expected payoff of the prosecutor. Hence, the optimal information structure in this setting will always have $\pi(t_1, t_1 | \theta_1) = 1$.

Second, when a juror votes for conviction, he is exactly indifferent between the two votes. If he were strictly in favor of convicting, then the prosecutor could increase the probability of $\pi(t_1, t_1 | \theta_0)$ and decrease the probability of $\pi(t_0, t_0 | \theta_0)$, to the point at which the juror becomes indifferent. That will not change the juror's optimal choice given $t_1$ — he will still choose to convict — but will increase the probability of $(t_1, t_1)$ and hence, also the probability of conviction. The designer could increase $\pi(t_1, t_1 | \theta_0)$ to the point where, conditional upon receiving $t_1$, the posterior probability put on $\theta_0$ becomes so high that the juror would choose to acquit. This turning point for the posterior on $\theta_0$ is $\frac{2}{3}$ in this example.

A fundamental difference between the single receiver case of Kamenica and Genzkow [39] and the current framework concerns the posterior beliefs. The judge in their example needs to have a posterior belief (on "guilty" or $\theta_1$) of at least $\frac{1}{2}$ in order to convict. Here, in contrast, each juror convicts as long as his posterior belief on $\theta_1$ is at least $\frac{1}{3}$. This happens because of the complementarities in the strategic interaction and the choice of information structure. Since unanimity is needed for a verdict to be reached and the structure is such that both jurors always observe the same signal realization, each juror knows that voting for conviction will only really make a difference if the other juror were to vote in the same way. Therefore, receiving a signal indicative of a guilty defendant, i.e. $t_1$, will make a juror more willing to vote for conviction for two reasons. First, if the defendant is guilty, then his vote is needed for a just verdict to be reached. Second, if the defendant is innocent, then voting to acquit will not help achieve the right verdict (a payoff of 2) and will only give him the payoff from unilaterally choosing the right action (a payoff of 1). This is because,

each juror knows that conditional on receiving a signal $t_1$, the other juror has received the same signal and so, in equilibrium, is voting to convict.[4]

We can also show that in the analogous case of strategic substitutes with $c = 1$ and $d = 2$, the posterior on $\theta_1$ that is necessary for conviction increases to $\frac{2}{3}$. While the prosecutor-jury framework does not make sense with this parameterization, changing the game to one with strategic substitutes provides intuition as to how that affects the results for the same objective function. In particular, the optimal information structure now has $\pi(t_0, t_0|\theta_0) = \frac{11}{14}$ and $\pi(t_1, t_1|\theta_0) = \frac{3}{14}$, with the distribution conditional on $\theta_1$ remaining the same as before. Conditional on observing $t_1$ the posterior necessary for a juror to convict is now $\frac{2}{3}$, that is twice as high as in the case of strategic complements. This is due to the strong incentive for each juror to individually choose the just vote, irrespective of what the other juror decides. Hence, a designer who wants the agents to coordinate on the non-default action will have a harder time doing so when the underlying game is one of strategic substitutes as opposed to strategic complements.

## 3.4  The General Approach

This section describes the general approach to information design in finite environments.

---

[4]In the case when there is no benefit to choosing the just vote, i.e. when instead of 1 the payoffs to mis-coordinated votes are always 0, the posterior can be as low as the prior for an equilibrium of both jurors always convicting to be achieved. This is because under the null information structure, if the other juror were to always vote to convict, it is a best response to do the same.

### 3.4.1 Setup

There are $N$ agents engaged in a strategic interaction. The set of agents is denoted by $I$ and we index a generic player by $i = 1, \ldots, N$. Each player has a finite set of actions $A_i$ and we write $A = A_1 \times \cdots \times A_N$ for the set of action profiles and $a$ for a generic element of that set. There is a finite set of states $\Theta$ with $\theta$ denoting a generic element of that set. Each agent has a utility function $u_i : A \times \Theta \to \mathbb{R}$ that depends on the played action profile and on the ex ante unknown state of the world. The designer has a utility function $V : A \times \Theta \to \mathbb{R}$, so that her payoff is affected by the action profile that agents play and the state of the world. Designer and agents share a common full support prior $\psi \in \text{int}(\Delta(\Theta))$ and that is common knowledge. Let $G = \left((A_i, u_i)_{i=1}^N, \psi\right)$. We refer to $G$ as the *basic game*.

An *information structure* $S = \left((T_i)_{i=1}^N, \pi\right)$ consists of a finite set of signals $T_i$ for each player $i$ and a signal distribution $\pi : \Theta \to \Delta(T)$ where $T = T_1 \times \cdots \times T_N$. We denote by $t_i$ a generic element of $T_i$ and similarly by $t$, a generic element of $T$. Together, the tuple $(G, S)$ defines a game of incomplete information.[5]

Given a known basic game $G$, the designer chooses and publicly announces an information structure $S$, which becomes common knowledge. The agents then observe the choice $S$ and the subsequent signal realizations. Depending on the choice of information structure, these signal realizations may be only privately observable, they may be common to different subsets of agents, or they might be public to everyone. Upon observing his signal realization each agent formulates his first order and higher order beliefs taking into account the common knowledge of the information structure $S$. Then, each agent selects an action, which maximizes his interim expected utility.

---

[5]This division of a game of incomplete information into a basic game and an information structure has been previously used in the literature; see for example Bergemann and Morris [9] and Lehrer, Rosenberg, and Shmaya [41].

The resulting action profile is a Bayes Nash equilibrium (BNE) of the incomplete information game $(G, S)$ at the interim level. The designer's problem is to choose an information structure which induces agents to play a BNE that maximizes her ex ante expected utility. That is, the designer selects among the BNE of $(G, S)$ at the ex ante level, the one that is most beneficial to her. If there are multiple equilibria of $(G, S)$, we take a best-case approach and consider the one which yields the highest ex ante expected utility to the designer. If in turn there are multiple equilibria that maximize the designer's ex ante expected utility, we select arbitrarily among them. We follow Kamenica and Gentzkow [39] and take this best-case approach since it provides a meaningful benchmark in case of equilibrium multiplicity.

### 3.4.2   Designer's Problem

For a given basic game $G$, an information structure $S$ induces a BNE of the incomplete information game $(G, S)$, which in turn determines a distribution over action profiles and states of the world. Hence, the designer's problem can be organized as follows: 1) Characterize the set of all BNE of $G$ that could emerge under all possible information structures. We refer to this as the *constraint set* of the optimization problem. 2) Among all BNE, select (the) one which generates a distribution over actions and states that maximizes the designer's ex ante expected utility. We refer to the latter as the *objective function* of the designer's optimization. 3) Find the information structure which induces this BNE for the given basic game $G$. In this section we will show that steps 1) and 2) reduce to a linear programming problem. We will also show that without loss of generality we can focus on a particular class of information structures when approaching step 3).

### 3.4.2.1    Constraint Set

To determine the constraint set, we need to characterize the set of all BNE that could emerge under all possible information structures, of which there are infinitely many. To accomplish this task, we use a definition of correlated equilibrium introduced by Bergemann and Morris [9]. We show below that using their definition of correlated equilibrium under a special information structure, we can characterize the set of BNE that could emerge under all possible information structures. With this purpose in mind, we introduce a few definitions to establish the necessary terminology.

A (behavioral) strategy for player $i$ in $(G, S)$ is a mapping $\beta_i : T_i \to \Delta(A_i)$.

DEFINITION 10. (Bayes Nash Equilibrium)
*A strategy profile $\beta$ is a Bayes Nash equilibrium (BNE) of $(G, S)$ if for each $i \in I$, $t_i \in T_i$ and $a_i \in A_i$ with $\beta_i(a_i|t_i) > 0$, we have*

$$\sum_{a_{-i}, t_{-i}, \theta} \psi(\theta) \pi(t_i, t_{-i}|\theta) \Big( \prod_{j \neq i} \beta_j(a_j|t_j) \Big) u_i((a_i, a_{-i}), \theta)$$

$$\geq \sum_{a_{-i}, t_{-i}, \theta} \psi(\theta) \pi(t_i, t_{-i}|\theta) \Big( \prod_{j \neq i} \beta_j(a_j|t_j) \Big) u_i((a_i', a_{-i}), \theta) \quad (3.1)$$

*for all $a_i' \in A_i$.*

We next state the definition of Bayes correlated equilibrium as introduced by Bergemann and Morris [9]. Let $\sigma : T \times \Theta \to \Delta(A)$ be a distribution over action profiles conditional on type profiles and states.

DEFINITION 11. (Bayes Correlated Equilibrium)
*A distribution $\sigma$ is a Bayes correlated equilibrium (BCE) of $(G, S)$ if for each $i \in I$,*

$t_i \in T_i$ and $a_i \in A_i$, we have

$$\sum_{a_{-i},t_{-i},\theta} \psi(\theta)\pi(t_i,t_{-i}|\theta)\sigma((a_i,a_{-i})|(t_i,t_{-i}),\theta)u_i((a_i,a_{-i}),\theta)$$

$$\geq \sum_{a_{-i},t_{-i},\theta} \psi(\theta)\pi(t_i,t_{-i}|\theta)\sigma((a_i,a_{-i})|(t_i,t_{-i}),\theta)u_i((a_i',a_{-i}),\theta) \quad (3.2)$$

for all $a_i' \in A_i$.

A BCE distribution $\sigma$ reflects the assumption of common certainty of rationality and the common prior assumption in the basic game $G$ when the players have observed at least information structure $S$.

The designer is ultimately interested in what can be said about the equilibrium distributions of action profiles conditional on states of the world, as that determines the expected value of her objective function. She is not interested in the distributions conditional on the signals, as the information structure is simply a tool and not the end goal. Therefore, she would like to find the most beneficial equilibrium distribution of actions conditional on states, which maximizes the expected value of her objective function, without assuming a specific information structure to start with. Let mapping $\nu : \Theta \to \Delta(A)$ be a distribution over action profiles conditional on states.

DEFINITION **12.** *A distribution $\nu$ is a BNE of $(G,S)$ if $\beta$ is a BNE of $(G,S)$ and*

$$\sum_{t \in T} \pi(t|\theta)\Big(\prod_{j=1}^{N} \beta_j(a_j|t_j)\Big) = \nu(a|\theta) \quad (3.3)$$

*for each $a \in A$ and $\theta \in \Theta$. A distribution $\nu$ is a BCE of $(G,S)$ if $\sigma$ is a BCE of $(G,S)$ and*

$$\sum_{t \in T} \pi(t|\theta)\sigma(a|t,\theta) = \nu(a|\theta) \quad (3.4)$$

*for each $a \in A$ and $\theta \in \Theta$.*

Subsequently, for a basic game $G$ and information structure $S$, we use $BNE(G, S)$ to denote the set of BNE distributions $\nu$ and $BCE(G, S)$ to denote the set of BCE distributions $\nu$. In the designer's problem, the constraint set is the largest set of distributions $\nu$ that could emerge if agents play a BNE for a basic game $G$ under any possible information structure. To characterize this set, we show that it is easier to work with the set of BCE for $G$ under a particular information structure.

We next define the information structure, which plays an important role in the upcoming analysis. The *null* information structure $\underline{S}$ has $\underline{T}_i = \{t_i\}$ for all $i$ and $\underline{\pi}(\underline{t}|\theta) = 1$ for all $\theta \in \Theta$. Thus, the null information structure $\underline{S} = (\underline{T}, \underline{\pi})$ provides no information at all about the state of the world. The next results established the characterization of the largest set of BNE distributions through its equivalence to the set of BCE under the null information structure.

PROPOSITION **10.** *The following holds:* $BCE(G, \underline{S}) = \cup_S BNE(G, S)$.

The above result established the equivalence between the largest set of BNE distributions for a basic game $G$ and the set of BCE random choice rules for $G$ under the null information structure. It is a version of Theorem 2 by Bergemann and Morris [9]. We will work with the constraints defining the set $BCE(G, \underline{S})$ to characterize the constraint set of the designer $\cup_S BNE(G, S)$.

Intuitively, the result can be interpreted as follows. In a BCE distribution, the correlation between the actions given the state is arbitrary. In a BNE distribution, the correlation between the actions given the state can be generated only through independent probability distributions of individual actions given signals according to (3.3). To generate every possible distribution in $BCE(G, S)$ as a BNE, i.e. with behavioral strategies, the additional coordination with the state must come through the conditioning on the signals. Therefore, the information structure has to provide

the necessary correlation of the independently chosen actions and the state. Every BCE distribution can thus be replicated as a BNE distribution for an appropriately chosen information structure $S$. $S$ should provide enough information about the state to generate the required correlation in the equilibrium distribution. To summarize, a BCE distribution under $\underline{S}$ can be viewed as a stochastic device which is sophisticated in terms of how much correlation it can generate between the actions, but does not use the information structure at all. A BNE distribution under $S$, on the other hand, can be viewed as a stochastic device which generates all the correlation between the actions through $S$. Therefore more intricate information structures are required for the latter to replicate any distribution of the former.

By Proposition 10 we can characterize the set of all BNE by means of the BCE incentive constraints (3.2) under the null information structure $\underline{S}$. We need to combine these with the constraints ensuring $\nu$ is a proper probability distribution. Hence, the set $BCE(G, \underline{S})$ is the collection of $\nu(a|\theta)$ such that:

i) $\nu(a|\theta) \geq 0$ for all $a \in A$ and $\theta \in \Theta$,

ii) $\sum_{a \in A} \nu(a|\theta) = 1$ for all $\theta \in \Theta$, and

iii) $\sum_{a_{-i}, \theta} \psi(\theta)\nu((a_i, a_{-i})|\theta)u_i((a_i, a_{-i}), \theta) \geq \sum_{a_{-i}, \theta} \psi(\theta)\nu((a_i, a_{-i})|\theta)u_i((a_i', a_{-i}), \theta)$
for all $i \in I$, $a_i \in A_i$ and $a_i' \in A_i$.

The above constraints are all linear in $\nu(a|\theta)$. Therefore, the set $BCE(G, \underline{S})$ is a convex polygon. By Theorem A of Stinchcombe [55], the set of BCE is non-empty. Hence, the constraint set of the designer is a non-empty convex polygon.

### 3.4.2.2 Objective Function

The designer's utility when the agents play action profile $a$ and the state is $\theta$ is given by $V(a, \theta)$. The designer's objective is to maximize the ex ante expected value of her utility, which can be written as

$$\mathbb{E}_\nu[V] = \sum_{a,\theta} V(a, \theta)\nu(a|\theta)\psi(\theta).$$

Notice that this objective is also linear in $\nu(a|\theta)$. Hence, the designer is maximizing a linear objective function over a non-empty convex polygon and the tools of linear programming can be utilized to find the optimal solution. By the fundamental theorem of linear programming, a solution $\nu^*$ exists and is at one of the corners of the constraint set. Therefore, $\nu^* \in \cup_S BNE(G, S)$ is the BNE the designer would like to induce. We next character use the information structure $S^*$ which supports $\nu^*$ as a BNE, i.e. for which $\nu^* \in BNE(G, S^*)$.

### 3.4.2.3 Optimal Information Structure

We first simplify the problem by showing that, without loss of generality, we can restrict attention to a certain class of information structures, which we call *direct*.

DEFINITION **13.** *Given a basic game $G$, an information structure $S = (T, \pi)$ is direct if $T = A$ and there exists $\nu \in BNE(G, S)$ such that $\nu(a|\theta) = \pi(a|\theta)$ for all $a \in A$ and $\theta \in \Theta$.*

DEFINITION **14.** *Given basic game $G$, we say that an information structure $S$ has value $\tilde{V}$ if there exists a distribution $\nu \in BNE(G, S)$ such that $\mathbb{E}_\nu[V] = \tilde{V}$.*

PROPOSITION **11.** *The following are equivalent:*

*(i) There exists an information structure with value $V^*$;*

*(ii) There exists a direct information structure with value $V^*$;*

*(iii) There exists a BNE distribution $\nu$ such that $\mathbb{E}_\nu[V] = V^*$.*

The main implication of Proposition 11 is that we can work with direct information structures only. The equivalence of $(i)$ and $(ii)$ is in spirit very similar to the revelation principle (e.g., Myerson [50]). The equivalence between $(ii)$ and $(iii)$ uses Proposition 10 and a *truthful* equilibrium strategy. The intuition behind this is simple. If there is a BNE distribution $\nu$ over action profiles conditional on states, then it must be that $\nu$ is also a BCE distribution under the null information structure. Thus, if the designer uses a direct information structure with the same probability distribution $\nu$, it is Bayes incentive compatible for each agent to follow the action recommendation implied by the observed signal realization assuming that the other agents do so as well. This generates an BNE equilibrium distribution $\nu$ under a direct information structure, which in turn results in the same ex ante expected payoff for the designer.

COROLLARY **2.** *The optimal information structure is given by $S^* = (A, \pi^*)$, where $\pi^*(a|\theta) = \nu^*(a|\theta)$ and $\nu^* = \arg\max_\nu \mathbb{E}_\nu[V]$ s.t. $\nu \in BCE(G, \underline{S})$.*

This corollary establishes the equivalence between the optimal information structure and the optimal BCE distribution under the null information structure. Once we find $\nu^*$, we create a direct information structure with the same probability distribution over signal realizations conditional on states. The signal realizations are in fact the action recommendations, which agents have incentive to follow in equilibrium. This direct information structure is optimal.

In our setting, there is nothing that precludes a designer, who has chosen a partially informative information structure, from deciding to release more information

after certain "unfavorable" signal realizations. The following result establishes that regardless of the actual signal realizations and the implied equilibrium action profile, the designer would never want to deviate and send additional signals, if the initial information structure was optimally chosen to begin with and if all players have observed the resulting action profile $\underline{a}$. In this case, the designer cannot benefit from persuading the players to switch to a different action profile by gicing them more information.

PROPOSITION **12.** *If a realized action profile $\underline{a} \in A$ observed by all players was induced by an optimal signal, the designer has no incentive to release more information.*

## 3.5 Application: Symmetric Binary Environments

In this section we apply the general information design approach outlined above to a symmetric binary environment.

### 3.5.1 Setup

Consider a two-player, two-state, two-action environment with symmetric payoffs. There are $N = 2$ players, and we use $i$ as an index for the typical player, and $j$ for his opponent. The set of states of the world is $\Theta = \{\theta_0, \theta_1\}$. The set of actions is the same for both players and given by $A = \{a_0, a_1\}$. The payoffs (or utility functions) $u : A \times \Theta \to \mathbb{R}$ are also the same for both players. Further, we assume a common prior $\psi$, which is uniform on the two states, i.e. $\psi(\theta_0) = \psi(\theta_1) = \frac{1}{2}$. Hence, we have specified the basic game $G = (A^2, u, \psi)$. We will refer to this as the *symmetric $2 \times 2 \times 2$ environment.*

Consider the following parameterized framework, where the payoffs in each state are given by: with $c \geq 0$ and $d \geq 0$. The assumption that the payoff parameters

| $\theta = \theta_0$ | $a_0$ | $a_1$ |
|---|---|---|
| $a_0$ | $c, c$ | $d, 0$ |
| $a_1$ | $0, d$ | $0, 0$ |

| $\theta = \theta_1$ | $a_0$ | $a_1$ |
|---|---|---|
| $a_0$ | $0, 0$ | $0, d$ |
| $a_1$ | $d, 0$ | $c, c$ |

Table 3.1: Parameterized Basic Game

are weakly positive ensures that the participation constraints of the agents to engage in the strategic interaction are always satisfied. This two-parameter representation is rich enough to capture many different environments of interest. We will refer to the basic game with parameters $c$ and $d$ as $G_{c,d}$.

The above payoff matrices assume that players have a preference for playing different actions in the different states of the world. This is an important assumption. Notice that if the same action were preferred in both states, there would be a dominant strategy equilibrium. In this case, the information that players receive is irrelevant for their strategies, and the designer cannot use information design to achieve her desired objective. Hence, information design becomes relevant only when the players have preferences for coordinating each action with a different state. We denote by $a_k$ the action preferred in state $\theta_k$ for $k = 0, 1$. Additionally, we use superscript to signify the agent that takes the action, i.e. $a_k^i$ stands for agent $i$ taking action $a_k$.

In addition to the preference for aligning their action with the state, the players may exhibit either a preference for coordination (strategic complementarity) or mis-coordination (strategic substitutability) of their action with the action of their opponent. The strength of the preference for alignment with the state versus alignment with one's opponent depends on the relative magnitude of $c$ and $d$.

The preference of each player for coordination with the state, for any given action of the other player, is represented by $c + d$, which we will refer to as the

97

*unilateral complementarity* $(U)$. This is given by the difference:

$$U = u(a_1, a^j, \theta_1) - u(a_0, a^j, \theta_1) - u(a_1, a^j, \theta_0) + u(a_0, a^j, \theta_0) = c + d \qquad (3.5)$$

for each $a^j \in A$. Due to the symmetry, we obtain the same expression for each player and each possible opponent action. The larger (3.5), the stronger the preference for alignment between each player's own action and the state.

In each state, the preference of each player for coordination with the other player is captured by $c - d$. This is given by:

$$T = u(a_1, a_1, \theta_k) - u(a_0, a_1, \theta_k) - u(a_1, a_0, \theta_k) + u(a_0, a_0, \theta_k) = c - d \qquad (3.6)$$

for $k = 0, 1$. We will refer to this as the *strategic complementarity* $(T)$. If this difference is positive and large, there is a strong preference for coordination with one's opponent, that is, strong strategic complementarity. On the other hand, if this difference is negative and large, there is a strong preference for mis-coordination between the players and thus, strong strategic substitutability. Consequently, we say that the basic game $G_{c,d}$ exhibits *strategic complements* if $c > d$ and *strategic substitutes* if $c < d$.

This two-parameter payoff representation captures many strategic interactions of interest and different preferences for (mis)coordination. For example, $c > d > 0$ represents the beauty contest game: players want to match the state and have an added benefit if their actions match. This may correspond to a situation of two people deciding to invest in one of two projects. The profitability of the projects depends on an unknown state and on the total investment, with higher investment leading to a more profitable project. Therefore, choosing the right project is associated with a higher payoff if the opponent also invests in the same project. When $d > c > 0$, the payoffs represent the situation of two competitors trying to match the

consumer preference for a certain product. If they both match it, they split the market. However, if one of them fails to produce the product with desired features, then the other firm captures the whole market and obtains a higher payoff.

For any value $T$ of the strategic complementarity (3.6) and any value $U \geq |T|$ of the unilateral complementarity (3.5), we can choose payoff parameters $c$ and $d$ to yield these coordination preferences by setting $c = \frac{U+T}{2}$ and $d = \frac{U-T}{2}$.

### 3.5.2  Designer's Problem

### 3.5.2.1  Constraint Set

To determine the constraint set we need to characterize the set of all possible BNE for basic game $G_{c,d}$ under all possible information structures. By Proposition 10 we know that for a given basic game $G$, the largest set of distributions over actions and states of the world, which can be sustained as BNE under some information structure, is given by $BCE(G, \underline{S})$. We restrict attention to distributions which are symmetric both in terms of the players and with respect to the state. These can be fully described by two parameters — $q$ and $r$ — and denoted as $\nu(q, r)$. Hence, a symmetric distribution over action profiles conditional on state can be represented as follows

| $\theta = \theta_0$ | $a_0$ | $a_1$ |
|---|---|---|
| $a_0$ | $r$ | $q - r$ |
| $a_1$ | $q - r$ | $1 - 2q + r$ |

| $\theta = \theta_1$ | $a_0$ | $a_1$ |
|---|---|---|
| $a_0$ | $1 - 2q + r$ | $q - r$ |
| $a_1$ | $q - r$ | $r$ |

We denote a particular random choice rule as $\nu(q, r)$. The parameter $r$ represents the probability with which in each state both agents simultaneously match the state with their actions: $\Pr(a_0, a_0|\theta_0) = \Pr(a_1, a_1|\theta_1) = r$. Hence, it measures the likelihood with which the players coordinate both with each other and with the state. On the other hand, $q$ denotes the probability with which in each state each agent matches the

state with his action, irrespective of whether the other agent does so as well or not. For agent $i$ and state $\theta_0$, this probability is given by $\Pr(a_0^i, a_0^j | \theta_0) + \Pr(a_0^i, a_1^j | \theta_0) = q$.

We choose to work with symmetric distributions for a number of reasons. First, this is without loss of generality when the utility function of the designer is also symmetric in agents and in states — we will define explicitly what this means in the following subsection. However, in general, this will be a constrained optimal BNE distribution. Second, we use symmetric distributions as that allows for a two-dimensional graphical representation of the constraint set and objective function. Third, sometimes the designer is naturally constrained in her choice to symmetric information structures due to laws and regulations. In the symmetric $2 \times 2 \times 2$ environment that means that the she will be optimizing over the set of symmetric BNE, which are induced by symmetric information structures.

Our next result characterizes the set of symmetric BCE of $(G_{c,d}, \underline{S})$. We consider all possible values of the basic game parameters $c$ and $d$ which do not make the strategic interaction trivial. In other words, we consider all possible cases with $c, d \geq 0$ and for which both parameters are not simultaneously equal to zero.

PROPOSITION **13. (BCE Random Choice Rules)**
*Consider the symmetric $2 \times 2 \times 2$ environment.*

*If $c > d$ (strategic complements), the set of symmetric BCE random choice rules of $(G_{c,d}, \underline{S})$ is given by $\left\{ (q, r) \in Co\left\{ \left( \frac{d}{c+d}, \frac{d}{c+d} \right), \left( \frac{2c-d}{3c-d}, \frac{c-d}{3c-d} \right), (1, 1) \right\} \right\}$.*

*If $d > c$ (strategic substitutes), the set of symmetric BCE random choice rules of $(G_{c,d}, \underline{S})$ is given by $\left\{ (q, r) \in Co\left\{ \left( \frac{d}{c+d}, \frac{d}{c+d} \right), (1, 1), \left( \frac{d}{3d-c}, 0 \right), \left( \frac{1}{2}, 0 \right) \right\} \right\}$.*

*If $d = c > 0$, the set of symmetric BCE random choice rules of $(G_{c,d}, \underline{S})$ is given by $\left\{ (q, r) \in Co\left\{ \left( \frac{1}{2}, \frac{1}{2} \right), \left( \frac{1}{2}, 0 \right), (1, 1) \right\} \right\}$.*

The proof of the proposition shows that the set of BCE random choice rules under the null information structure for a basic game $G_{c,d}$ is the constraint set determined by four linear inequalities. Three of these inequalities ensure that the parameters of the random choice rule satisfy the consistency conditions for probability distributions. The fourth inequality represents the incentive constraints associated with BCE under the null information structure.

We make use of the following example to show the construction of the constraint set. We will return to this example throughout the rest of the section to illustrate the different steps of the information design problem.

SYMMETRIC EXAMPLE. *Consider the parameterized basic game in Table 3.1 with $c = 2$ and $d = 1$. Hence, the agents are involved in a coordination game, where they want to both match each other and the state with their actions. Suppose the designer benefits from mis-coordination between the agents' actions irrespective of the state. That is, her utility function is given by:*

$$V(a^i, a^j, \theta) = \begin{cases} 1 & \text{if } a^i \neq a^j \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

*The constraint set $BCE(G_{2,1}, \underline{S})$ is depicted in Figure 3.1. The red line represents the BCE incentive constraint. It always goes through the point $(\frac{1}{2}, \frac{1}{4})$, plotted on the graph, which represents the symmetric mixed strategy BNE when the agents have no information.*

### 3.5.2.2 Objective function

We consider a general utility function for the designer $V : A \times \Theta \to \mathbb{R}$. Hence, $V(a, \theta)$ is the designer payoff when $a$ is the action profile played by the agents and $\theta$ is
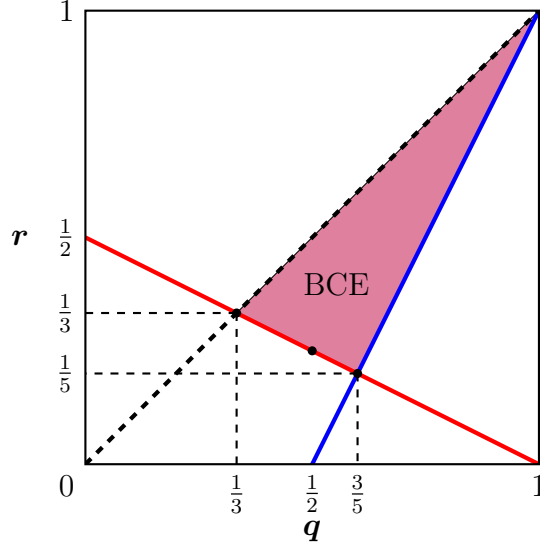
Figure 3.1: Constraint Set (Symmetric Example)

the state of the world. For any given objective function $V(a, \theta)$, where $a$ is the action profile played by the agents and $\theta$ is the realized state of the world, its expectation given a BNE distribution $\nu(q, r)$ is

$$\mathbb{E}(V) = \psi\big[rV(a_0, a_0, \theta_0) + (q-r)V(a_0, a_1, \theta_0) + (q-r)V(a_1, a_0, \theta_0) + (1-2q+r)V(a_1, a_1, \theta_0)\big]$$
$$+(1-\psi)\big[rV(a_1, a_1, \theta_1) + (q-r)V(a_0, a_1, \theta_1) + (q-r)V(a_1, a_0, \theta_1) + (1-2q+r)V(a_0, a_0, \theta_1)\big].$$
$$(3.8)$$

Reorganizing and regrouping terms gets us to:

$$\begin{aligned}
\mathbb{E}(V) = & \Big[\psi\big[V(a_1, a_1, \theta_0) - V(a_0, a_1, \theta_0) - V(a_1, a_0, \theta_0) + V(a_0, a_0, \theta_0)\big] \\
& + (1-\psi)\big[V(a_1, a_1, \theta_1) - V(a_0, a_1, \theta_1) - V(a_1, a_0, \theta_1) + V(a_0, a_0, \theta_1)\big]\Big]r \\
& + \Big[\psi\big[V(a_0, a_1, \theta_0) - V(a_1, a_1, \theta_0) + V(a_1, a_0, \theta_0) - V(a_1, a_1, \theta_0)\big] \\
& + (1-\psi)\big[V(a_0, a_1, \theta_1) - V(a_0, a_0, \theta_1) + V(a_1, a_0, \theta_1) - V(a_0, a_0, \theta_1)\big]\Big]q \\
& + \psi V(a_1, a_1, \theta_0) + (1-\psi)V(a_0, a_0, \theta_1) \\
= & \ R \cdot r + Q \cdot q + const.
\end{aligned}$$
$$(3.9)$$

102

The coefficient in front of $r$, which we denote by $R$, captures the "expected" preference for complementarity between the actions in the designer's objective function. It is indeed a weighted average of the complementarities between the actions in each state, the weights being the prior probabilities for each state. Therefore, the coefficient in front of $r$ measures the average importance of coordination of the agents' actions in the designer's objective function.

The coefficient in front of $q$, which we label $Q$, is the expected preference for *unilateral* coordination of each player's action with the state, assuming the other player mismatches the state. For example, suppose that the state is $\theta_0$. Then $V(a_0, a_1, \theta_0) - V(a_1, a_1, \theta_0)$ captures the benefit of having the first player unilaterally match the state with his action as opposed to having perfect mis-coordination between both of the actions and the state. For the second player, the relevant expression is $V(a_1, a_0, \theta_0) - V(a_1, a_1, \theta_0)$. So the sum of those two expressions represents the preference of the designer for "unilateral" coordination between the players and the state $\theta_0$. Therefore, the coefficient in front of $q$ measures the importance of unilateral coordination in the designer's utility in expectation over the two states.

Lastly, a utility function that is symmetric in both the agents' actions and the state is characterized by the following equalities: (i) $V(a_0, a_0, \theta_0) = V(a_1, a_1, \theta_1)$, (ii) $V(a_0, a_1, \theta_0) = V(a_1, a_0, \theta_0) = V(a_0, a_1, \theta_1) = V(a_1, a_0, \theta_1)$ and (iii) $V(a_1, a_1, \theta_0) = V(a_0, a_0, \theta_1)$. As mentioned above, when the designer's utility function is symmetric, restricting attention to symmetric information structures is without loss of generality.

SYMMETRIC EXAMPLE. *The utility function of the designer given by* (3.7) *is symmetric both with in the agents' actions and in the state. Substituting the values into* (3.9), *gives*

$$\mathbb{E}(V) = -2r + 2q \tag{3.10}$$

*as the ex ante expected objective function. This is represented by a level line with a slope of one, the value of which increases when shifted in the direction of the lower-right corner (see Figure 3.3).*

### 3.5.2.3 Optimal Information Structure

In the previous section, we characterized the set $BCE(G_{c,d}, \underline{S})$, which is the budget set of the designer. We can now maximize the designer's objective function (3.9) over this set. Let us denote by $\nu^*(q, r)$ the distribution which maximizes (3.9) over $BCE(G_{c,d}, \underline{S})$. Once we find $\nu^*(q, r)$, we can reverse-engineer the information structure $S^*$ which decentralizes it as a BNE. By Proposition 11 we know that there exists a direct information structure $S^*$ such that $\nu^*(q, r) \in BNE(G_{c,d}, S^*)$. And by Corollary 2 we know that $S^* = (A, \pi^*)$ with $\pi^*(a|\theta) = \nu^*(a|\theta)$ for all $a \in A$ and $\theta \in \Theta$.

Therefore, the direct information structures which support all distributions $\nu(q, r) \in BCE(G_{c,d}, \underline{S})$ as BNE, can be parameterized in an analogous way with the following conditional probabilities $\pi(\cdot|\theta)$ on signal realizations:

| $\theta = \theta_0$ | $a_0$ | $a_1$ |
|---|---|---|
| $a_0$ | $r$ | $q - r$ |
| $a_1$ | $q - r$ | $1 - 2q + r$ |

| $\theta = \theta_1$ | $a_0$ | $a_1$ |
|---|---|---|
| $a_0$ | $1 - 2q + r$ | $q - r$ |
| $a_1$ | $q - r$ | $r$ |

Table 3.2: Direct Information Structures

The information structure parameterization in Table 3.2 is very general as it represents all binary information structures which are symmetric across agents and states. The parameter $q$ is the probability with which each agent receives the action recommendation that "matches" the state, i.e. the "state-matching" action. We refer to it as the *precision* of the information structure. The parameter $r$ is the

probability with which both agents simultaneously receive the state-matching action recommendation. We refer to it as the *correlation* of the information structure.

The above parameterization also includes many important special structures. The case of conditionally independent private signals is captured by setting $r = q^2$ for $q \in (0, 1)$. In this case, each agent receives a private signal which is equal to the state-matching action with probability $q$ and is independent of the signal of his opponent. Both agents thus receive the state-matching action recommendation with probability $q \times q = r$ and receive opposite action recommendations with probability $q \times (1 - q) = q - r$. On the other hand, the case of public signals is covered by setting $r = q$. This ensures that both agents always receive the same action recommendation, where $q$ is the probability of having that action match the state. We denote public signals by $S_{q,q}$. For the general case of *private signals* with precision $q$ and correlation $r$, we write $S_{q,r}$.

Of particular importance is the *null information* structure $\underline{S}$ which provides no information about the state $\theta$. In terms of the above parameterization, the null information structure corresponds to $q = \frac{1}{2}$ and can be denoted as $S_{\frac{1}{2},r}$. In this case, the signals are completely uninformative with respect to the state. Notice also that there are infinitely many null information structures, each one associated with a different degree of correlation between the signals. On the other hand, there is only one *full information* structure $\bar{S}$ which reveals the state of the world perfectly, captured by $q = r = 1$ and written as $S_{1,1}$.

It is useful for the upcoming analysis to graphically represent the set of direct binary information structures in the $(q, r)-$space (Figure 3.2). For the conditional probabilities in Table 3.2 to be positive, we need to have $r$ smaller than $q$, greater than $2q - 1$ and greater than 0. Thus, the set of possible direct information structures is defined by three linear constraints. The first line is $r = q$, which describes the set of all
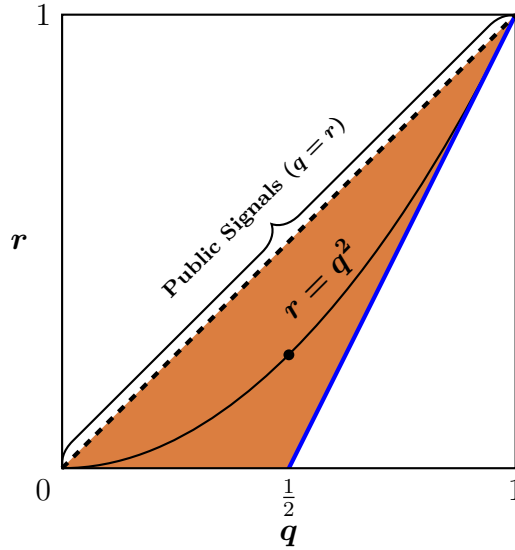
105

Figure 3.2: Set of Direct Information Structures

public signals with different levels of precision. The second line is $r = 2q - 1$, which represents all information structures with minimal levels of correlation for a given level of precision $q$. And the third line is $r = 0$, which corresponds to all information structures with zero correlation consistent with different levels of precision. That is why these three lines determine the set of direct information structures. It is easy to see that the set of conditionally independent signals, $r = q^2$ with $q \in (0, 1)$ is in the interior of the set of all possible information structures.

Before we move on to the complete characterization, let us demonstrate graphically how we obtain the optimal information structure in the symmetric example we have been using throughout this section. This is shown in Figure 3.3.

SYMMETRIC EXAMPLE. *The symmetric BCE which maximizes the expectation of the objective function is $\nu^*(\frac{3}{5}, \frac{1}{5})$. The optimal direct information structure is thus given by $S^* = (A, \nu^*)$ and is summarized in the following matrices:*
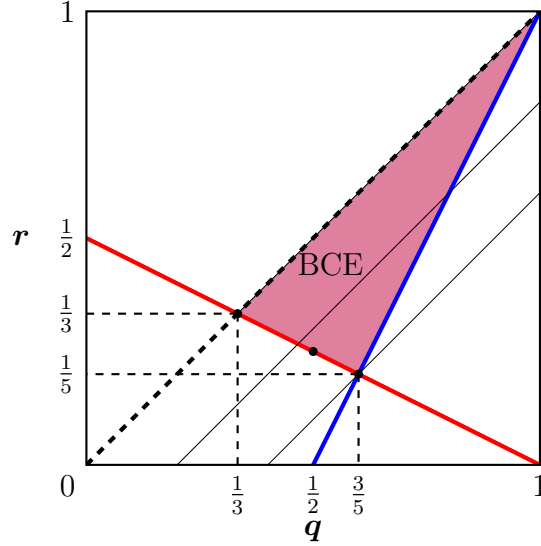
Figure 3.3: Optimal Information Structure (Symmetric Example)

| $\theta = \theta_0$ | $a_0$ | $a_1$ |
|---|---|---|
| $a_0$ | $\frac{1}{5}$ | $\frac{2}{5}$ |
| $a_1$ | $\frac{2}{5}$ | $0$ |

| $\theta = \theta_1$ | $a_0$ | $a_1$ |
|---|---|---|
| $a_0$ | $0$ | $\frac{2}{5}$ |
| $a_1$ | $\frac{2}{5}$ | $\frac{1}{5}$ |

*Under this information structure, the expected value of the prosecutor's objective function is $\frac{4}{5}$. Due to the symmetry of the binary environment and of the designer's utility function, this information structure is a global optimum. In other words, restricting attention to symmetric information structures is, in this case, without loss of generality.*

Our next result is a complete characterization of the optimal symmetric information structure for all possible designer's objective functions and basic games $G_{c,d}$. Recall that $R$ and $Q$ are defined as the average preference for coordination of the agents' actions and the average preference for unilateral coordination of each player with the state, respectively (see Section 3.5.2).

THEOREM **2.**

107

1. If $R > 0$ and $Q > 0$, the full information structure is always optimal.

2. If $R < 0$, $Q > 0$ and the basic game exhibits strategic complements, the optimal information structure is public signals with precision $\frac{d}{c+d}$ if $-\frac{Q}{R} < \frac{c-3d}{2(c-d)}$; private signals with precision $\frac{2c-d}{3c-d}$ and correlation $\frac{c-d}{3c-d}$ if $\frac{c-3d}{2(c-d)} < -\frac{Q}{R} < 2$; and the full information structure if $-\frac{Q}{R} > 2$.

3. If $R < 0$, $Q > 0$ and the basic game exhibits strategic substitutes, the optimal information structure is the null information structure if $-\frac{Q}{R} < 2$; and the full information structure if $-\frac{Q}{R} > 2$.

4. If $R > 0$, $Q < 0$ and the basic game exhibits strategic complements, the optimal information structure is the full information structure if $-\frac{Q}{R} < 1$; and private signals with precision $\frac{2c-d}{3c-d}$ and correlation $\frac{c-d}{3c-d}$ if $-\frac{Q}{R} > 1$.

5. If $R > 0$, $Q < 0$ and the basic game exhibits strategic substitutes, the optimal information structure is the full information structure if $-\frac{Q}{R} < 1$; public signals with precision $\frac{d}{c+d}$ if $1 < -\frac{Q}{R} < \frac{c-3d}{2(c-d)}$; and private signals with precision $\frac{d}{3d-c}$ and correlation $0$ if $-\frac{Q}{R} > \frac{c-3d}{2(c-d)}$.

6. If $R < 0$, $Q < 0$ and the basic game exhibits strategic complements, the optimal information structure is public signals with precision $\frac{d}{c+d}$ if $-\frac{Q}{R} < \frac{c-3d}{2(c-d)}$; and private signals with precision $\frac{2c-d}{3c-d}$ and correlation $\frac{c-d}{3c-d}$ if $-\frac{Q}{R} > \frac{c-3d}{2(c-d)}$.

7. If $R < 0$, $Q < 0$ and the basic game exhibits strategic substitutes, the optimal information structure is private signals with precision $\frac{d}{3d-c}$ and correlation $0$.

Our characterization theorem is summarized in Table 3.3 of Appendix B. The information design problem can be seen as utility maximization given the designer's

preferences over distributions of actions conditional on the states, where the budget set is $BCE(G_{c,d}, \underline{S})$. The slope of the designer's level line, $-\frac{Q}{R}$, can be viewed as a marginal rate of substitution. It represents the designer's benefit from an increase in the probability $(q)$ of state coordination relative to the benefit from an increase in the probability $(r)$ of action coordination. When this slope is negative, the designer benefits from simultaneous movements in both probabilities and so is willing to trade in an increase in one for a decrease in the other. When it is positive, however, the designer is willing to trade between simultaneous movements: an increase (decrease) in one for an increase (decrease) in the other. Thus, the slope of the designer level line represents the tradeoff that she is willing to accept between the two parameters of the equilibrium distributions.

The set of BCE under the null information structure, $BCE(G_{c,d}, \underline{S})$, is the budget set of the information designer. The slopes of its boundaries represent the tradeoffs between the parameters $q$ and $r$ that need to be maintained so that the BCE incentive compatibility constraints remain satisfied. Therefore, these slopes represent the rates at which the designer may trade changes in one parameter for changes in the other. We next explain the intuition in a few cases.

Consider the case when the designer would like the agents to both coordinate their actions with each other $(R > 0)$ and with the state $(Q > 0)$. If the game has strategic complements, then the agents have the exact same preferences as the designer. Therefore, it is best to give the agents full information, as they will use it to coordinate perfectly with each other and the state, which is the objective of the designer. Notice that for the same preferences of the designer, the optimal information structure is full information also when the game has strategic substitutes, which is somewhat counterintuitive. The reason is that once the agents have full information, it is always a dominant strategy to play the action that matches the state, because

the strategic substitutes are not strong enough. This is due to the fact that we restrict the payoff parameters to be strictly positive, which is a limitation of the model. If we were to allow for $c$ to be negative, this result would change. However, in the current setting, a designer who would like the agents to coordinate with each other and with the state achieves that by giving full information both when the game exhibits strategic complements and strategic substitutes.

Next, consider the case of a designer with preferences described by $R < 0$ and $Q > 0$. She would hence like to choose $q$ to be as high as possible and $r$ to be as low as possible. This translates into a preference for mis-coordination between the agents irrespective of the state, as it maximizes the probability of mismatched actions $\Pr(a_0, a_1) = \Pr(a_1, a_0) = q - r$. If the game has strategic substitutes, there is an underlying incentive for the agents to mismatch their actions. However, each one of them still has an incentive to match the state and the mismatched action profile is never a full information Nash equilibrium. If the preference for coordination with the state is not as strong as the disutility from coordination between the agents, then the designer would choose to reveal no information. This will maximize the probability of mis-coordinated actions. Conversely, if the preference for coordination with the state is stronger than the disutility from coordination between the actions, then it is optimal for the designer to reveal everything. The actions will never be mis-coordinated in this case; nonetheless, the designer would prefer to have perfect state coordination.

For the case of $R > 0$ and $Q < 0$, the designer wants the agents to coordinate their actions but to not coordinate with the state. If the game exhibits strategic complements, the agents would like to both coordinate with each other and with the state. Therefore, if the designer's preference for action coordination is stronger than the disutility from state coordination, she will choose the full information structure

and have them coordinate on both actions and the state. In contrast, if she really dislikes state coordination, she will choose correlated privates signals with imperfect precision. By doing this, the designer foregoes the perfect action coordination she could achieve with full information in order to achieve some degree of state mis-coordination. Thus, depending on the strength of those two preferences, the outcome is either full information or correlated private signals with imperfect precision.

Lastly, consider the case of $R < 0$ and $Q < 0$, where the designer wants the agents to mis-coordinate both on the state and the actions. If the game exhibits strategic substitutes, the optimal information structure is private signals with low precision. This ensures that the agents do not obtain enough information about the state, so that they don't coordinate too much with it. At the same time, they have enough information about what the other player has likely observed. Since the game has strategic substitutes, the agents have an incentive to mis-coordinate their actions. Hence the signals are used mainly as a device for the agents to condition their actions on, in order to achieve mis-coordination.

To contrast our results with those from the literature on cheap talk without commitment, we would like to point out that making the preferences of the designer and the agents more aligned, may in fact decrease the optimal precision of information. For example, when we have strategic substitutes, and the preferences of the designer are $R > 0, Q > 0$, full information is optimal. For strategic substitutes and $R < 0, Q > 0$, the designer also wants action mis-coordination and state coordination, just like the agents. So the preferences have become more aligned. However, the null structure is optimal in this case for certain values of the parameters.

COROLLARY 3. *Conditionally independent private signals are never optimal.*

Notice that $r = q^2$, $q \in (0, 1)$ is always in the interior of the set of information structures and is therefore never optimal.

COROLLARY **4.** *The only case when the designer may not benefit from information design is when $R < 0$, $Q > 0$, $-\frac{Q}{R} < 2$ and the basic game exhibits strategic substitutes.*

There is no benefit from information design whenever revealing no information and letting the agents operate under their prior beliefs is optimal. Thus, as long as the optimal information structure differs from the null, the designer benefits from information design. No information revelation, i.e. $q = 1/2$, is only ever strictly optimal in the case of strategic substitutes when the designer has preferences described by $R < 0$, $Q > 0$ and $-\frac{Q}{R} < 2$ . The intuition behind this case was described above. In all remaining cases, information design is beneficial.

### 3.5.2.4    Indirect Information Structures

In some cases, rather than sending direct action recommendations, the designer may prefer or be confined to using signals, which are intrinsically associated to varying degrees with the different states of the world.[6] This is the second class of information structures we consider, which we call "indirect".

When the designer is not able to send direct action recommendations to create the information structure, the interpretation of the signals becomes relevant. In a way, this can be viewed as using a predetermined "language" to create the information structure. In binary environments, the designer needs two different signals to generate the set of information structures that can support all BNE. In order to generate symmetric information structures, one of the signals has to be designated as more indicative of state $\theta_0$, and the other signal — as more indicative of state $\theta_1$. Let us denote the former by $t_0$ and the latter by $t_1$. Therefore, conditional on $\theta_0$ ($\theta_1$)

---

[6]The chairman of the Fed does not typically talk about how economic agents should be behaving. Rather, his statements include signals regarding the Fed's stimulus policy and the economic outlook.

the probability that signal $t_0$ $(t_1)$ is observed has to be at least $\frac{1}{2}$ i.e. $\Pr(t_0|\theta_0) = \Pr(t_1|\theta_1) = r + (q - r) = q$ has to be weakly greater than $\frac{1}{2}$. Otherwise the signals would not be indicative of the states.

We can use the same parameterization as in Table 3.2, only instead of the action recommendations $a_0$ and $a_1$, we use the signals $t_0$ and $t_1$ respectively. Most importantly, we now have the added restriction that the precision $q$ has to be weakly larger than $\frac{1}{2}$. Thus, the set of indirect information structures is smaller than the set of direct information structures due to this additional constraint $q \geq \frac{1}{2}$. Figure 3.4 depicts the set of indirect information structures.
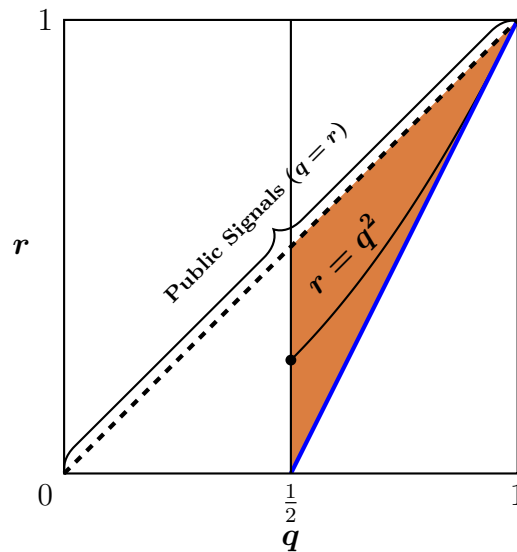


Figure 3.4: Set of Indirect Information Structures

Every $\nu^*(q, r) \in BCE(G_{c,d}, \underline{S})$ can be thus supported as a BNE by an indirect information structure $S_{q,r}$ as long as $q \geq \frac{1}{2}$. In this case, following the signal and playing $a_0$ when $t_0$ is received, and $a_1$ when $t_1$ is received, is a BNE. This can be viewed as "truthtelling" under indirect information structures.

However, for $\nu^*(q,r) \in BCE(G_{c,d}, \underline{S})$ with $q < \frac{1}{2}$ we can no longer use an indirect information structure $S_{q,r}$, as it is not defined for precision values less than a half. We need to find an indirect information structure, which will decentralize $\nu^*(q,r)$ as a BNE. The next proposition establishes the information structure that accomplishes this.

PROPOSITION 14. *For $q < \frac{1}{2}$, $\nu^*(q,r) \in BCE(G_{c,d}, \underline{S})$ only if $\nu^*(q,r) \in BNE(G_{c,d}, S_{1-q,1-2q+r})$.*

The indirect information structure which decentralizes a BCE $\nu^*(q,r)$ with $q < \frac{1}{2}$ as a BNE is given by $S_{1-q,1-2q+r}$. The intuition is that the designer creates an indirect structure that is a "mirror image" of the direct information structure she would have used. This is necessary so that the precision of the indirect structure is greater than $\frac{1}{2}$. Under $S_{1-q,1-2q+r}$, it is a BNE for both players to play the opposite action of what the signal suggests; that is, play $a_1$ if $t_0$ is received, and play $a_0$ if $t_1$ is received. This BNE results in a random choice rule which is exactly equivalent to $\nu^*(q,r)$. Since the information structure $S_{1-q,1-2q+r}$ is a mirror image of $\nu^*(q,r)$ and the BNE strategy we consider is in turn a mirror image of truthtelling, the resulting distribution over actions conditional on states of the world is exactly the desired distribution $\nu^*(q,r)$.

### 3.5.3 Mechanism and Information Design

Thus far we maintained the assumption that the payoffs of the underlying game — the parameters $c$ and $d$ — were constant. In this section we offer insights into how changes in these parameters affect the maximal payoff of the information designer. This change in parameter values may come about due to exogenous factors. Another possibility is if the designer uses state and action contingent transfers to modify the payoffs of the agents. This is feasible if the state of the world becomes observable after

114

the game has been played and if the actions of the agents are verifiable. Angeletos and Pavan [3] use taxes contingent on ex post public information about the realized state and aggregate activity. They show that such policies can improve the equilibrium use of information.

The main idea of this section is that in some cases the designer can induce an even better equilibrium outcome by combining the tools of mechanism design (payoff modification) and information design (belief modification), than when she uses only one or the other. Changes in the payoff parameters through state and payoff contingent transfers can affect the maximal utility of the information designer. An increase in $c$, while $d$ is held constant, increases both the complementarities with the state and between the agents. In contrast, an increase in $d$, while $c$ is held constant, increases the complementarities with the state, while decreasing the complementarities between the agents. These changes have different effects on the maximum utility of the designer, depending on her preferences and also on the extent of the changes.

It is not always possible to draw general conclusions from the comparative statics analysis. Frequently, if the change in the parameter is substantial enough, it may cause a shift from the case of strategic complements to the case of strategic substitutes. When this happens, the comparison is very sensitive to the size of the shift and we cannot evaluate the changes based only on its direction. In some instances, however, we are able to make clear-cut conclusions regarding the direction of the effects. The next proposition established the cases for which this is possible and for which mechanism design can improve on the outcome achieved with information design.

PROPOSITION **15.** *Holding all else equal,*

1. *and starting with $c > d$, an increase in $c$ or a decrease in $d$ is always beneficial to a utility maximizing designer with $R < 0$, $Q < 0$, who uses public signals.*

115

2. *and starting with $c < d$, a decrease in $c$ or an increase in $d$ is always beneficial to a utility maximizing designer with $R > 0$, $Q < 0$ who uses private signals and to a designer with $R < 0$, $Q < 0$.*

We next explain the intuition behind the last case: $R < 0$, $Q < 0$ and $c < d$. If the designer wants players to mis-coordinate both with the state and with each other, and the game exhibits strategic substitutes, the optimal information structure is private signals with low precision. This allows the players to use the signals in order to mis-match their actions, while obtaining very little information about the state. The signals need to have some level of precision in order for the players to pay attention to them and utilize them when formulating their strategies. This comes to the designer at the cost of the agents being able to predict the state better. When $c$ decreases, or $d$ increases, the level of strategic substitutability decreases. Therefore, the incentives of the agents for mis-coordinating their actions become stronger. They will now need less precise signals about the state to achieve this. This is beneficial to the designer as the less precise signals also result in less coordination with the state.

## 3.6   Discussion and Extensions

### 3.6.1   Communication

In certain environments it is unreasonable to assume that agents will not share the information they observe with each other, if that is beneficial to them. When the underlying game exhibits strategic complements, this issue is of particular relevance. The agents want to match the state and each other with their actions. Therefore, each agent has an incentive to disclose the private signal he observes to his opponent. The benefit from doing so is twofold. First, the agents can coordinate their actions perfectly once they have the same information. Second, by sharing the signal

116

realizations they have observed, they can improve the precision of their information and update their beliefs accordingly.

A designer who faces a situation where the agents have incentives to communicate with each other, needs to take that into account when designing the information structure. This implies making the information design robust to communication. In this case, the designer is restricted to public signal information structures, as every signal realization will be ultimately observed by both agents. Hence, she needs to include the constraint $q = r$ in her linear optimization program. This constraint imposes the restriction of public signals on the choice of information structure. From Table 3.3 it becomes clear that in some instances this constraint is binding and leads to lower optimal value of the designer's objective function. Looking at the case of strategic complements ($c > d$), whenever the optimal information structure consists of private signals, the communication constraint is binding. In these cases, the designer needs to choose a constrained optimum information structure consisting of a public signal, which will be communication robust.

### 3.6.2 Multiplicity of Equilibria and Other Solution Concepts

Information design is about finding the information structure under which the most beneficial BNE is played. However, this does not exclude the possibility of there being multiple BNE under the optimal information structure. Information design is subject to the same criticisms as mechanism design with regards to the multiple equilibrium problem. For example, consider the symmetric $2 \times 2 \times 2$ environment when there are strategic complements ($c > d$). Under the null information structure, there are three (agent) symmetric BNE: always play $a_0$, always play $a_1$, and mix with equal probabilities between the two actions. Only the last equilibrium is also symmetric in the state, which is what we focus on in our two-dimensional representation. Further

117

restrictions need to be imposed on the information structure through the incentive constraints in order to ensure the uniqueness of equilibrium.

In our framework, we exclusively focus on BNE as a solution concept. Working with different solution concepts may change the environment, by modifying the incentive constraints or adding new ones. For example, Lehrer, Rosenberg and Shmaya [41] consider solutions concepts which allow for degrees of communication and correlation between the agents that are different from Nash equilibrium. The optimal information structure, which supports the most desirable equilibrium under a different solution concept, can be derived using the approach described here with appropriately modified incentive constraints. An interesting question to investigate would be how the optimal information structures compare, in terms of their complexity, across the different solution concepts.

### 3.6.3  Exogenous Information

Our analysis and results are based on the assumption that the designer is in complete control of the informational environment. In particular, we assumed away any signals observed by the agents prior to the ones sent by the designer. In some instances, however, this assumption is unrealistic as the agents may already have some information about the state. Depending on the nature of this information, the designer's ability to achieve the highest possible objective may be impeded. In either case, the designer needs to take into account the prior signals of the agents and incorporate that as an additional constraint into her information design problem.

Consider the motivating example of Section 3.3. Let us assume that in the course of the trial, the jury already observed some evidence. In particular, she knows that the jury members have observed signal realizations from the following information structure:

| $\theta_0$ | $t_0$ | $t_1$ |
|---|---|---|
| $t_0$ | 6/20 | 7/20 |
| $t_1$ | 7/20 | 0 |

| $\theta_1$ | $t_0$ | $t_1$ |
|---|---|---|
| $t_0$ | 0 | 7/20 |
| $t_1$ | 7/20 | 6/20 |

The precision of these signals is $\frac{13}{20}$ which is slightly higher than the precision of $\frac{3}{5}$ that the optimal information structure was characterized by. Given this prior information structure observed by the jurors, the prosecutor can no longer achieve the unconstrained maximum objective value of $\frac{4}{5}$. In fact, the best she could do is choose a completely uninformative investigation and obtain a value of $\frac{7}{10}$.

Therefore, in the presence of prior information, the optimal information structure may significantly change as compared to the case of no prior information. Similar considerations apply when the agents observe additional signals beyond what the designer reveals. As long as the designer knows the structure of the exogenous information, she can incorporate that as additional constraints into her optimization problem. These constraints may not affect her ability to achieve the same maximum value of the objective as when the agents have no exogenous information. Nevertheless, she needs to modify the optimal information structure, since it is augmented by the exogenous signals.

## 3.7   Conclusion

The incentives of rational agents to behave in a certain way are determined by their payoffs and by their beliefs. Mechanism design concerns the modification of payoffs so that people have incentives to behave in desired ways. This paper lays out the methodology of information design. Information design operates on the beliefs of the agents through the choice of information structure. It thus focuses on the choice and creation of information structures under which agents achieve the most favorable outcomes.

119

As suggested in the previous section, there are many important extensions and robustness issues that can be studied with the proposed method. All of these constitute interesting directions for future research.

## Appendix A

**Proof of Proposition 10**

First we prove that $BCE(G, \underline{S}) \subseteq \cup_S BNE(G, S)$. Choose $\nu \in BCE(G, \underline{S})$. Hence, it must hold that

$$\sum_{a_{-i}, \theta} \psi(\theta)\nu(a_i, a_{-i}|\theta)u_i((a_i, a_{-i}), \theta) \geq \sum_{a_{-i}, \theta} \psi(\theta)\nu(a_i, a_{-i}|\theta)u_i((a_i', a_{-i}), \theta) \qquad (3.11)$$

for each $i \in I$, $a_i \in A_i$ and $a_i' \in A_i$. Consider the information structure $S^* = (T^*, \pi^*)$ with $T_i^* = A_i^* \times \underline{t}_i$ and

$$\pi^*((a_i, \underline{t}_i)_{i=1}^N|\theta) = \pi^*(a, \underline{t}|\theta) = \nu(a|\theta) \qquad (3.12)$$

for each $a \in A$ and $\theta \in \Theta$. In the game $(G, S^*)$ consider the "truthful" behavioral strategy $\beta_i^*$ for agent $i$ with

$$\beta_i^*(a_i|a_i', \underline{t}_i) = \begin{cases} 1, & \text{if } a_i = a_i' \\ 0, & \text{if } a_i \neq a_i' \end{cases} \qquad (3.13)$$

for all $a_i, a_i' \in A_i$. The interim payoff to agent $i$ observing signal $(a_i, \underline{t}_i)$ and choosing action $a_i'$ when his opponents follow $\beta_{-i}^*$ is

$$\sum_{a_{-i}, a_{-i}', \theta} \psi(\theta)\pi^*((a_i, a_{-i}'), \underline{t}|\theta)\Big(\prod_{j \neq i} \beta_j^*(a_j|a_j', \underline{t}_j)\Big)u_i((a_i', a_{-i}), \theta)$$

$$= \sum_{a_{-i}, \theta} \psi(\theta)\nu(a_i, a_{-i}|\theta)u_i((a_i', a_{-i}), \theta) \quad (3.14)$$

where we use (3.12) and (3.13). Therefore, the BNE interim incentive compatibility condition

$$\sum_{a_{-i}, a_{-i}', \theta} \psi(\theta)\pi^*((a_i, a_{-i}'), \underline{t}|\theta)\Big(\prod_{j \neq i} \beta_j^*(a_j|a_j', \underline{t}_j)\Big)u_i((a_i, a_{-i}), \theta)$$

$$\geq \sum_{a_{-i}, a_{-i}', \theta} \psi(\theta)\pi^*((a_i, a_{-i}'), \underline{t}|\theta)\Big(\prod_{j \neq i} \beta_j^*(a_j|a_j', \underline{t}_j)\Big)u_i((a_i', a_{-i}), \theta) \quad (3.15)$$

is equivalent to and implied by the BCE obedience constraint (3.11). Hence, $\beta^*$ is a BNE of $(G, S^*)$. The distribution over actions conditional on states generated from this equilibrium strategy is

$$\sum_{a' \in A} \pi^*(a', \underline{t}|\theta)\Big(\prod_{j=1}^N \beta_j(a_j|a_j', \underline{t}_j)\Big) = \nu(a|\theta). \qquad (3.16)$$

121

Thus, $\nu$ is a BNE of the game $(G, S^*)$, i.e. $\nu \in BNE(G, S^*)$. This implies $BCE(G, \underline{S}) \subseteq \cup_S BNE(G, S)$.

Next we prove that $BCE(G, \underline{S}) \supseteq \cup_S BNE(G, S)$. Choose $\tilde{\nu} \in \cup_S BNE(G, S)$. Hence, there exist an information structure $\tilde{S} = (\tilde{T}, \tilde{\pi})$ and a BNE behavioral strategy $\beta(a|\tilde{t})$ of $(G, \tilde{S})$ such that

$$\tilde{\nu}(a|\theta) = \sum_{\tilde{t} \in \tilde{T}} \tilde{\pi}(\tilde{t}|\theta)\Big(\prod_{j=1}^{N} \beta_j(a_j|\tilde{t}_j)\Big). \tag{3.17}$$

We write $\tilde{\pi}(\tilde{t}, \underline{t}|\theta) = \tilde{\pi}(\tilde{t}|\theta)$, $\beta(a|\tilde{t}, \underline{t}) = \beta(a|\tilde{t})$ and $\tilde{\nu}(a|\underline{t}, \theta) = \tilde{\nu}(a|\theta)$ for $\tilde{t} \in \tilde{T}$ and $\{\underline{t}\} = \underline{T}$, which trivially holds.

For each $a_i$ such that $\beta_i(a_i|\tilde{t}_i, \underline{t}_i) > 0$, by the BNE incentive compatibility condition it must hold that

$$\sum_{a_{-i}, \tilde{t}_{-i}, \theta} \psi(\theta)\tilde{\pi}((\tilde{t}_i, \tilde{t}_{-i}), \underline{t}|\theta)\Big(\prod_{j \neq i} \beta_j(a_j|\tilde{t}_j, \underline{t}_j)\Big) u_i((a_i, a_{-i}), \theta)$$

$$\geq \sum_{a_{-i}, \tilde{t}_{-i}, \theta} \psi(\theta)\tilde{\pi}((\tilde{t}_i, \tilde{t}_{-i}), \underline{t}|\theta)\Big(\prod_{j \neq i} \beta_j(a_j|\tilde{t}_j, \underline{t}_j)\Big) u_i((a_i', a_{-i}), \theta) \quad (3.18)$$

for each $i \in I$, $\tilde{t}_i \in \tilde{T}_i$, and $a_i' \in A_i$. Multiplying both sides by $\sum_{\tilde{t}_i} \beta_i(a_i|\tilde{t}_i, \underline{t}_i)$ gives

$$\sum_{a_{-i}, \tilde{t}, \theta} \psi(\theta)\tilde{\pi}((\tilde{t}_i, \tilde{t}_{-i}), \underline{t}|\theta)\Big(\prod_{j=1}^{N} \beta_j(a_j|\tilde{t}_j, \underline{t}_j)\Big) u_i((a_i, a_{-i}), \theta)$$

$$\geq \sum_{a_{-i}, \tilde{t}, \theta} \psi(\theta)\tilde{\pi}((\tilde{t}_i, \tilde{t}_{-i}), \underline{t}|\theta)\Big(\prod_{j=1}^{N} \beta_j(a_j|\tilde{t}_j, \underline{t}_j)\Big) u_i((a_i', a_{-i}), \theta) \quad (3.19)$$

which by (3.17) is equivalent to

$$\sum_{a_{-i}, \theta} \psi(\theta)\tilde{\nu}(a|\underline{t}, \theta)u_i((a_i, a_{-i}), \theta) \geq \sum_{a_{-i}, \theta} \psi(\theta)\tilde{\nu}(a|\underline{t}, \theta)u_i((a_i', a_{-i}), \theta). \tag{3.20}$$

Thus, $\tilde{\nu} \in BCE(G, \underline{S})$, which implies $BCE(G, \underline{S}) \supseteq \cup_S BNE(G, S)$. $\qquad \square$

**Proof of Proposition 11**

By definition, (ii) implies (i) and (iii). Let us first show that (i) implies (ii). Take a

basic game $G$ and an information structure $S = (T, \pi)$ with value $V^*$. Suppose that $\beta$ is the BNE of $(G, S)$ which generates that value, that is

$$\sum_{a,t,\theta} V(a, \theta) \pi(t|\theta) \Big( \prod_{i=1}^{N} \beta_i(a_i|t_i) \Big) \psi(\theta) = V^*. \tag{3.21}$$

Let $T^a = \{t|\beta_i(a_i|t_i) > 0 \ \forall \ i \in I\}$. Consider the direct information structure $S' = (A, \pi')$ with

$$\pi'(a|\theta) = \sum_{t \in T^a} \pi(t|\theta) \Big( \prod_{i=1}^{N} \beta_i(a_i|t_i) \Big). \tag{3.22}$$

We will show that the truthful strategy of playing the action implied by the signal realization is a BNE of $(G, S')$.

Since $\beta_i(a_i|t_i) > 0$ we have the BNE incentive compatibility condition

$$\sum_{a_{-i}, t_{-i}, \theta} \psi(\theta) \pi((t_i, t_{-i})|\theta) \Big( \prod_{j \neq i} \beta_j(a_j|t_j) \Big) u_i((a_i, a_{-i}), \theta)$$

$$\geq \sum_{a_{-i}, t_{-i}, \theta} \psi(\theta) \pi((t_i, t_{-i})|\theta) \Big( \prod_{j \neq i} \beta_j(a_j|t_j) \Big) u_i((a_i', a_{-i}), \theta) \tag{3.23}$$

for each $i \in I$, $t_i \in T_i$ and $a_i' \in A_i$. Multiplying both sides of the above inequality by $\beta_i(a_i|t_i)$ and summing across all $t_i$ we get

$$\sum_{t_i} \beta_i(a_i|t_i) \sum_{a_{-i}, t_{-i}, \theta} \psi(\theta) \pi((t_i, t_{-i})|\theta) \Big( \prod_{j \neq i} \beta_j(a_j|t_j) \Big) u_i((a_i, a_{-i}), \theta)$$

$$\geq \sum_{t_i} \beta_i(a_i|t_i) \sum_{a_{-i}, t_{-i}, \theta} \psi(\theta) \pi((t_i, t_{-i})|\theta) \Big( \prod_{j \neq i} \beta_j(a_j|t_j) \Big) u_i((a_i', a_{-i}), \theta) \tag{3.24}$$

which can be rewritten as

$$\sum_{a_{-i}, \theta} \psi(\theta) \sum_{t} \pi((t_i, t_{-i})|\theta) \Big( \prod_{i=1}^{N} \beta_i(a_i|t_i) \Big) u_i((a_i, a_{-i}), \theta)$$

$$\geq \sum_{a_{-i}, \theta} \psi(\theta) \sum_{t} \pi((t_i, t_{-i})|\theta) \Big( \prod_{i=1}^{N} \beta_i(a_i|t_i) \Big) u_i((a_i', a_{-i}), \theta) \tag{3.25}$$

123

for each $i \in I$ and $a_i' \in A_i$. Substituting in with (3.22) we obtain

$$\sum_{a_{-i},\theta} \psi(\theta)\pi'((a_i, a_{-i})|\theta)u_i((a_i, a_{-i}), \theta) \geq \sum_{a_{-i},\theta} \psi(\theta))\pi'((a_i, a_{-i})|\theta)u_i((a_i', a_{-i}), \theta). \quad (3.26)$$

In the game $(G, S')$ consider the behavioral strategy $\beta_i'$ for agent $i$ with

$$\beta_i'(a_i|a_i') = \begin{cases} 1, & \text{if } a_i = a_i' \\ 0, & \text{if } a_i \neq a_i' \end{cases} \quad (3.27)$$

for all $a_i, a_i' \in A_i$. The interim payoff to agent $i$ observing signal $a_i$ and choosing action $a_i'$ in $(G, S')$ when each opponent $j$ follows strategy $\beta_j'$ is

$$\sum_{a_{-i}, a_{-i}', \theta} \psi(\theta)\pi'((a_i, a_{-i}')|\theta)\Big(\prod_{j \neq i}\beta_j'(a_j|a_j')\Big)u_i((a_i', a_{-i}), \theta)$$

$$= \sum_{a_{-i},\theta} \psi(\theta)\pi'((a_i, a_{-i})|\theta)u_i((a_i', a_{-i}), \theta). \quad (3.28)$$

Hence, (3.26) implies the BNE incentive compatibility conditions for strategy profile $\beta'$. Under $\beta'$ the expected payoff to the designer is given by

$$\mathbb{E}[V] = \sum_{a,\theta} V(a, \theta)\pi'(a|\theta)\psi(\theta) = V^*. \quad (3.29)$$

where we use (3.21) and (3.22). Hence, the direct information structure $S'$ also has value $V^*$.

Next we show that (iii) implies (ii). For basic game $G$, consider a BNE distribution $\nu$ such that

$$\mathbb{E}_\nu[V] = \sum_{a,\theta} V(a, \theta)\nu(a|\theta)\psi(\theta) = V^*.$$

By Proposition 10 we know that $\nu \in BCE(G, \underline{S})$. Hence, it holds

$$\sum_{a_{-i},\theta} \psi(\theta)\nu((a_i, a_{-i})|\theta)u_i((a_i, a_{-i}), \theta) \geq \sum_{a_{-i},\theta} \psi(\theta))\nu((a_i, a_{-i})|\theta)u_i((a_i', a_{-i}), \theta). \quad (3.30)$$

Consider the direct information structure $S = (A, \pi)$ with $\pi(a|\theta) = \nu(a|\theta)$ for all $a \in A$ and $\theta \in \Theta$. In the game $(G, S)$ consider the behavioral strategy $\beta_i$ for agent $i$ with

$$\beta_i(a_i|a_i') = \begin{cases} 1, & \text{if } a_i = a_i' \\ 0, & \text{if } a_i \neq a_i' \end{cases} \quad (3.31)$$

124

for all $a_i, a_i' \in A_i$. The interim payoff to agent $i$ observing signal $a_i$ and choosing action $a_i'$ in $(G, S)$ when each opponent $j$ follows strategy $\beta_j$ is

$$\sum_{a_{-i}, a_{-i}', \theta} \psi(\theta) \pi((a_i, a_{-i}')|\theta) \Big( \prod_{j \neq i} \beta_j(a_j|a_j') \Big) u_i((a_i', a_{-i}), \theta)$$

$$= \sum_{a_{-i}, \theta} \psi(\theta) \pi((a_i, a_{-i})|\theta) u_i((a_i', a_{-i}), \theta) = \sum_{a_{-i}, \theta} \psi(\theta) \nu((a_i, a_{-i})|\theta) u_i((a_i', a_{-i}), \theta).$$
(3.32)

where the first equality follow by (3.31) and the second equality follows from $\pi(a|\theta) = \nu(a|\theta)$. Hence, (3.32) implies the BNE incentive compatibility conditions for strategy profile $\beta$. The distribution of actions conditional on states of the world under $\beta$ is

$$\sum_{a' \in A} \pi(a'|\theta) \Big( \prod_{i=1}^{N} \beta_i(a_i|a_i') \Big) = \pi(a|\theta) = \nu(a|\theta)$$
(3.33)

and thus, the expected payoff of the designer under strategy profile $\beta$ is given by

$$\mathbb{E}[V] = \sum_{a, \theta} V(a, \theta) \nu(a|\theta) \psi(\theta) = V^*.$$
(3.34)

Hence, the direct information structure $S$ has value $V^*$. $\qquad \square$

**Proof of Proposition 12**

Let $\nu^*$ be an optimal direct signal structure that generates an action profile $\underline{a}$, i.e. $\nu^*(\underline{a}|\theta) > 0$ for some $\theta \in \Theta$. The expected payoff for the designer under realization $\underline{a}$ is given by $\mathbb{E}[V(\underline{a})] = \sum_{\theta} V(\underline{a}, \theta) \nu^*(\theta|\underline{a})$. Suppose that upon observing the signal realization $\underline{a}$, the designer decides to release a new signal structure $\hat{\nu}$. She will only strictly prefer to do that if this gives her a higher expected payoff, that is if:

$$\sum_{\theta} V(\underline{a}, \theta) \nu^*(\theta|\underline{a}) < \sum_{\theta} \nu^*(\theta|\underline{a}) \sum_{a} V(a, \theta) \hat{\nu}(a|\theta).$$
(3.35)

Multiplying both sides of the above inequality by $\nu^*(\underline{a}) = \sum_{\theta \in \Theta} \nu^*(\underline{a}|\theta)$ we obtain:

$$\sum_{\theta} V(\underline{a}, \theta) \nu^*(\underline{a}, \theta) < \sum_{\theta} \nu^*(\underline{a}, \theta) \sum_{a} V(a, \theta) \hat{\nu}(a|\theta).$$
(3.36)

Suppose that the designer released the following information structure to start with:

$$\nu(a|\theta) = \begin{cases} \nu^*(\underline{a}|\theta)\hat{\nu}(a|\theta) & \text{if } \hat{\nu}(a|\theta) > 0 \text{ and } \nu^*(a|\theta) = 0 \\ \nu^*(a|\theta) + \nu^*(\underline{a}|\theta)\hat{\nu}(a|\theta) & \text{if } \hat{\nu}(a|\theta) > 0 \text{ and } \nu^*(a|\theta) > 0 \\ \nu^*(a|\theta) & \text{if } \hat{\nu}(a|\theta) = 0, \nu^*(a|\theta) > 0 \text{ and } a \neq \underline{a} \\ 0 & \text{otherwise} \end{cases}$$

(3.37)

for each $a \in A$ and $\theta \in \Theta$. It is straightforward to see that $\nu(a|\theta) \in [0,1]$ and $\sum_a \nu(a|\theta) = 1$ and hence $\nu(a|\theta)$ is a proper probability distribution. Obedience of $\nu(a|\theta)$ requires:

$$\sum_{a_{-i},\theta} \psi(\theta)\nu((a_i,a_{-i})|\theta)u_i((a_i,a_{-i}),\theta) \geq \sum_{a_{-i},\theta} \psi(\theta))\nu((a_i,a_{-i})|\theta)u_i((a_i',a_{-i}),\theta). \quad (3.38)$$

$$\sum_{a_{-i},\theta} \psi(\theta)[\nu((a_i,a_{-i})|\theta) + \nu^*(\underline{a}|\theta)\hat{\nu}(a_i,a_{-i}|\theta)] \; u_i((a_i,a_{-i}),\theta)$$

$$\geq \sum_{a_{-i},\theta} \psi(\theta)[\nu((a_i,a_{-i})|\theta) + \nu^*(\underline{a}|\theta)\hat{\nu}(a_i,a_{-i}|\theta)] \; u_i((a_i',a_{-i}),\theta). \quad (3.39)$$

This condition is satisfied due to the obedience of $\nu^*$ under the prior $\psi(\theta)$ and the obedience of $\hat{\nu}$ under the prior $\nu^*(\theta|\underline{a})$.

The above signal structure $\nu(a|\theta)$ would have resulted in an expected payoff of

$$\sum_\theta \nu^*(\underline{a}|\theta)\psi(\theta) \sum_a V(a,\theta)\hat{\nu}(a|\theta) + \sum_{a\backslash\underline{a},\theta} V(a,\theta)\nu^*(a|\theta)\psi(\theta)$$

$$= \sum_\theta \nu^*(\underline{a},\theta) \sum_a V(a,\theta)\hat{\nu}(a|\theta) + \sum_{a\backslash\underline{a},\theta} V(a,\theta)\nu^*(a,\theta) \quad (3.40)$$

By (3.36) the above expression is strictly larger than the expected payoff under $\nu^*$, which can be written as

$$\sum_\theta V(\underline{a},\theta)\nu^*(\underline{a},\theta) + \sum_{a\backslash\underline{a},\theta} V(a,\theta)\nu^*(a,\theta). \quad (3.41)$$

This is a contradiction to $\nu^*$ being optimal. $\qquad\qquad\square$

**Proof of Proposition 13**

For the null information structure $\underline{S}$ and a basic game $G_{c,d}$ the general BCE constraints given in Definition 11 become:

for $a_i = a_0$, $a'_i = a_1$:

$$\frac{1}{2}rc + \frac{1}{2}(q-r)d \geq \frac{1}{2}(q-r)c + \frac{1}{2}(1-2q+r)d$$

and

for $a_i = a_1$, $a'_i = a_0$:

$$\frac{1}{2}rc + \frac{1}{2}(q-r)d \geq \frac{1}{2}(q-r)c + \frac{1}{2}(1-2q+r)d.$$

These two constraints are equivalent and reduce to only one inequality:

$$2(c-d)r \geq d + (c-3d)q. \tag{3.42}$$

Additionally, the parameters need to satisfy:

$$r \leq q \tag{3.43}$$

$$r \geq \max\{2q-1, 0\} \tag{3.44}$$

and

$$q \in [0,1]. \tag{3.45}$$

Therefore, the set of BCE random choice rules if $(G_{c,d}, \underline{S})$ is equivalent to the set of $(q,r)$-pairs which satisfy constraints (3.42)–(3.45).

**Case 1:** Assume $c > d \geq 0$ (strategic complements). The obedience constraint (3.42) can thus be written as:

$$r \geq \frac{d}{2(c-d)} + \frac{c-3d}{2(c-d)}q \tag{3.46}$$

In this case, constraint (3.45), which essentially coincides with the $x$-axis of the graph, is never binding. The reason behind this is the following. The intercept of constraint (3.46) is always positive. When in addition $c \geq 3d$, the slope is also positive. Hence, this constraint is always more binding than (3.45), as it always lies above the $x$-axis. On the other hand, when $c < 3d$, the slope of (3.46) is negative. However, it is easy to show that (3.46) intersects (3.44) before it intersects the $x$-axis. Therefore, for the relevant range of values, (3.46) lies above the $x$-axis also in this case Hence, (3.45) is never binding.
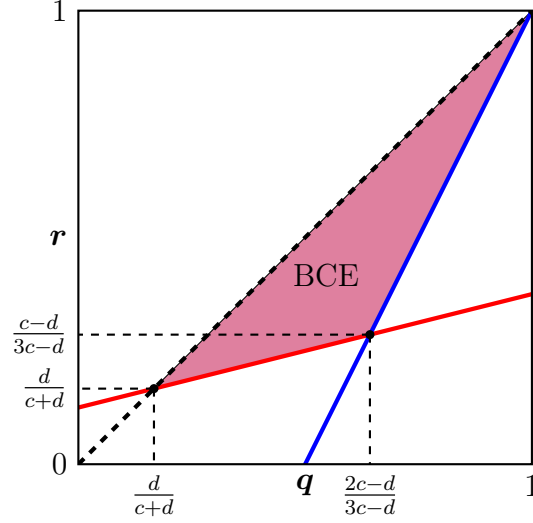
Figure 3.5: Strategic Complements $(c > d \geq 0)$

The set of random choice rules which satisfy (3.46), (3.43) and (3.44) is thus equivalent to the convex hull formed by the intersection points $(q_1, r_1) = \left(\frac{d}{c+d}, \frac{d}{c+d}\right)$ (of (3.46) and (3.43)), $(q_2, r_2) = \left(\frac{2c-d}{3c-d}, \frac{c-d}{3c-d}\right)$ (of (3.46) and (3.44))and $(q_3, r_3) = (1, 1)$ (of (3.43) and (3.44)).

**Case 2:** Assume $d > c \geq 0$ (strategic substitutes). The obedience constraint (3.42) can thus be written as:

$$r \leq \frac{d}{2(c - d)} + \frac{c - 3d}{2(c - d)}q \tag{3.47}$$

This constraint has a negative intercept and a positive slope. In fact, it always holds that the slope $\frac{c-3d}{2(c-d)} \geq \frac{3}{2}$. When $c > 0$ the slope is strictly greater than $\frac{3}{2}$ and (3.47) intersects only constraints (3.45) and (3.43). In this case, all four constraints (3.47), (3.43), (3.44) and (3.45) are binding. The set of random choice rules which satisfy all of them is equivalent to the hull formed by the intersection points $(q_1, r_1) = \left(\frac{d}{c+d}, \frac{d}{c+d}\right)$ (of (3.47) and (3.43)), $(q_3, r_3) = (1, 1)$ (of (3.43) and (3.44)), $(q_5, r_5) = \left(\frac{1}{2}, 0\right)$ (of (3.44) and (3.45)) and $(q_4, r_4) = \left(\frac{d}{3d-c}, 0\right)$ (of (3.45) and (3.47)).

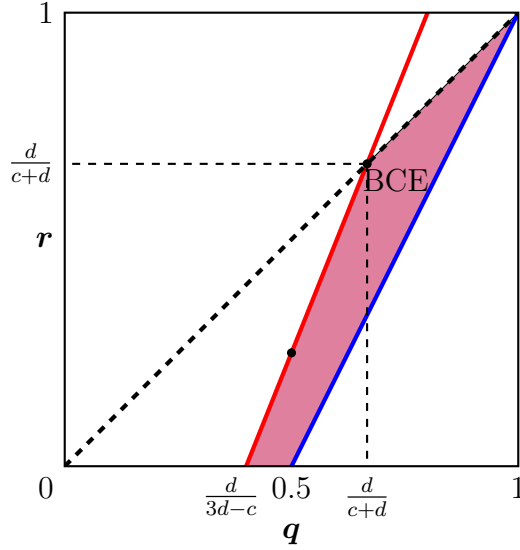When $c = 0$, the slope of (3.47) is exactly equal to $\frac{3}{2}$. In this case (3.47), (3.43)

128

Figure 3.6: Strategic Substitutes $(d > c \geq 0)$

and (3.44) all intersect at one point — $(q_3, r_3) = (1, 1)$ — and (3.43) is never binding. The set of random choice rules is equivalent to the hull formed by the intersection points $(q_3, r_3) = (1, 1)$, $(q_5, r_5) = \left(\frac{1}{2}, 0\right)$ and $(q_4, r_4) = \left(\frac{1}{3}, 0\right)$.

**Case 3:** In the special case of $c = d > 0$, the obedience constraint (3.42) becomes $q \geq \frac{1}{2}$. The set of BCE is then equivalent to the convex hull of $(q_1, r_1) = \left(\frac{1}{2}, \frac{1}{2}\right)$, $(q_2, r_2) = \left(\frac{1}{2}, 0\right)$, and $(q_3, r_3) = (1, 1)$. □
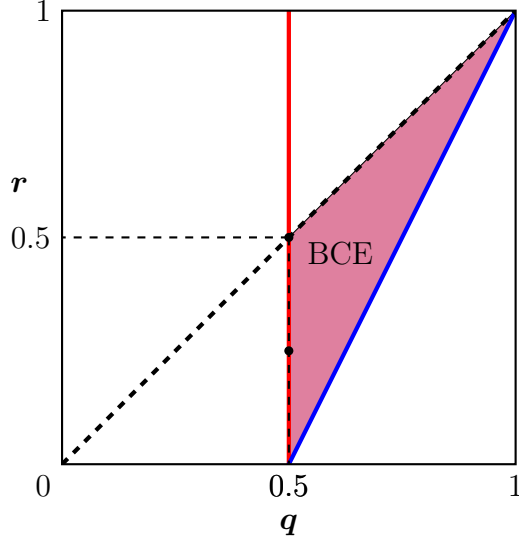
Figure 3.7: $c = d > 0$

**Proof of Proposition 14**

Since $\nu^*(q,r) \in BCE(G_{c,d}, \underline{S})$, we know that (3.42) holds. We will show that this condition implies the behavioral strategy

$$\beta_i(a_k|t_n) = \begin{cases} 1 & \text{if } k \neq n \\ 0 & \text{if } k = n \end{cases} \tag{3.48}$$

for $k, n = 0, 1$ and $i = 1, 2$ is a BNE in the incomplete information game $(G_{c,d}, S_{1-q,1-2q+r})$. For (3.48) to be an equilibrium, it needs to hold that the BNE incentive compatibility conditions given by (3.1) in Definition 10 are satisfied. Due to the symmetry of the players, we need to consider only player $i$. The incentive constraint for $\beta_i(a_1|t_0) = 1$

is given by

$$\frac{1}{2}\pi(t_0,t_0|\theta_0)\beta_j(a_1|t_0)u(a_1,a_1,\theta_0) + \frac{1}{2}\pi(t_0,t_0|\theta_1)\beta_j(a_1|t_0)u(a_1,a_1,\theta_1)$$

$$+ \frac{1}{2}\pi(t_0,t_1|\theta_0)\beta_j(a_0|t_1)u(a_1,a_0,\theta_0) + \frac{1}{2}\pi(t_0,t_1|\theta_1)\beta_j(a_0|t_1)u(a_1,a_0,\theta_1)$$

$$\geq \frac{1}{2}\pi(t_0,t_0|\theta_0)\beta_j(a_1|t_0)u(a_0,a_1,\theta_0) + \frac{1}{2}\pi(t_0,t_0|\theta_1)\beta_j(a_1|t_0)u(a_0,a_1,\theta_1)$$

$$+ \frac{1}{2}\pi(t_0,t_1|\theta_0)\beta_j(a_0|t_1)u(a_0,a_0,\theta_0) + \frac{1}{2}\pi(t_0,t_1|\theta_1)\beta_j(a_0|t_1)u(a_0,a_0,\theta_1).$$

$$(3.49)$$

When we substitute in the probabilities $\pi(\cdot|\theta)$ of the information structure $S_{1-q,1-2q+r}$, the equilibrium strategy probabilities $\beta_j$ and the basic game payoffs, the above condition reduces to

$$rc + (q-r)d \geq (1-2q+r)d + (q-r)c. \qquad (3.50)$$

The incentive constraint for $\beta_i(a_0|t_1) = 1$ is given by

$$\frac{1}{2}\pi(t_1,t_0|\theta_0)\beta_j(a_1|t_0)u(a_0,a_1,\theta_0) + \frac{1}{2}\pi(t_1,t_0|\theta_1)\beta_j(a_1|t_0)u(a_0,a_1,\theta_1)$$

$$+ \frac{1}{2}\pi(t_1,t_1|\theta_0)\beta_j(a_0|t_1)u(a_0,a_0,\theta_0) + \frac{1}{2}\pi(t_1,t_1|\theta_1)\beta_j(a_0|t_1)u(a_0,a_0,\theta_1)$$

$$\geq \frac{1}{2}\pi(t_1,t_0|\theta_0)\beta_j(a_1|t_0)u(a_1,a_1,\theta_0) + \frac{1}{2}\pi(t_1,t_0|\theta_1)\beta_j(a_1|t_0)u(a_1,a_1,\theta_1)$$

$$+ \frac{1}{2}\pi(t_1,t_1|\theta_0)\beta_j(a_0|t_1)u(a_1,a_0,\theta_0) + \frac{1}{2}\pi(t_1,t_1|\theta_1)\beta_j(a_0|t_1)u(a_1,a_0,\theta_1).$$

$$(3.51)$$

After substituting in we obtain:

$$(q-r)d + rc \geq (q-r)c + (1-2q+r)d. \qquad (3.52)$$

Notice that (3.50) and (3.52) are equivalent and, moreover, implied by the BCE condition (3.42). Hence, the behavioral strategy (3.48) is an equilibrium in the incomplete information game $(G_{c,d}, S_{1-q,1-2q+r})$.

We now need to show that this BNE strategy generates the random choice rule $\nu^*(q,r)$. We will show that for the distribution conditional on $\theta_0$, as the rest follows by analogy. The decision rule induced by strategy (3.48) is given by $\sigma(a_0,a_0|t_1,t_1,\theta_0) =$

$\sigma(a_0, a_1 | t_1, t_0, \theta_0) = \sigma(a_1, a_0 | t_0, t_1, \theta_0) = \sigma(a_1, a_1 | t_0, t_0, \theta_0) = 1$ and zero otherwise. Hence, we obtain:

$$\nu(a_0, a_0 | \theta_0) = \sigma(a_0, a_0 | t_1, t_1, \theta_0) \pi(t_1, t_1 | \theta_0) = r$$

$$\nu(a_0, a_1 | \theta_0) = \sigma(a_0, a_1 | t_1, t_0, \theta_0) \pi(t_1, t_0 | \theta_0) = q - r$$

$$\nu(a_1, a_0 | \theta_0) = \sigma(a_1, a_0 | t_0, t_1, \theta_0) \pi(t_0, t_1 | \theta_0) = q - r$$

$$\nu(a_1, a_1 | \theta_0) = \sigma(a_1, a_1 | t_0, t_0, \theta_0) \pi(t_0, t_0 | \theta_0) = 1 - 2q + r.$$

By analogy, we obtain the corresponding values for $\nu(\cdot | \theta_1)$. This is exactly equivalent to the random choice rule $\nu^*(q, r)$. Thus, (3.48) is a BNE strategy which induces $\nu^*(q, r)$ in $(G_{c,d}, S_{1-q, 1-2q+r})$. $\qquad \square$

**Proof of Proposition 15**

<u>Case 1</u>: Consider a designer with $R < 0$, $Q < 0$, and $-\frac{Q}{R} < \frac{c-3d}{2(c-d)}$ in the case of strategic complements. By Theorem 3.3 we know that the optimal information structure is a public signal with precision $q = \frac{d}{c+d}$. Notice that $\frac{c-3d}{2(c-d)}$ is increasing in $c$ and decreasing in $d$. Therefore an increase in $c$ and a decrease in $d$ will both increase its value. This implies that after the change, the optimal information structure will be a public signal with a new precision level. Therefore, when we consider such changes in the parameters, we can write the designer's utility under the optimal information structure as:

$$V^* = (R + Q) \frac{d}{c+d} + C$$

where $C$ is a constant. Notice that $V^*$ is increasing in $c$:

$$\frac{\partial V^*}{\partial c} = -(R + Q) \frac{d}{(c+d)^2} > 0$$

and decreasing in $d$:

$$\frac{\partial V^*}{\partial c} = (R + Q) \frac{c}{(c+d)^2} < 0.$$

<u>Case 2</u>: Consider a designer with $R > 0$, $Q < 0$, and $-\frac{Q}{R} > \frac{c-3d}{2(c-d)}$ in the case of strategic substitutes. Since $\frac{c-3d}{2(c-d)}$ is increasing in $c$ and decreasing in $d$, a decrease in $c$ and a increase in $d$ will both decrease its value. This implies that after the change, the optimal information structure will be a private signal with a new precision level and same correlation $r = 0$. Therefore, when we consider such changes

132

in the parameters, we can write the designer's utility under the optimal information structure as:
$$V^* = Rr + Q\frac{d}{3d - c} + C = Q\frac{d}{3d - c} + C$$
where $C$ is a constant. Notice that $V^*$ is decreasing in $c$:
$$\frac{\partial V^*}{\partial c} = Q\frac{d}{(3d - c)^2} < 0$$
and increasing in $d$:
$$\frac{\partial V^*}{\partial c} = -Q\frac{c}{(3d - c)^2} > 0$$
a decrease in $c$ or an increase in $d$ always increases the maximal utility of the designer. The same argument applies to the case of $R < 0$, $Q < 0$ and strategic substitutes. $\quad\square$

# Appendix B

Table 3.3: Characterization of Optimal Information Structure

| | complements $[c > d]$ | substitutes $[c < d]$ |
|---|---|---|
| $R > 0, Q > 0$ | full information | full information |
| $R < 0, Q > 0$ | public signal $\left(q = \frac{d}{c+d}\right)$ if $-\frac{Q}{R} < \frac{c-3d}{2(c-d)}$ | ——— |
| | private signals $\left(q = \frac{2c-d}{3c-d}, r = \frac{c-d}{3c-d}\right)$ if $\frac{c-3d}{2(c-d)} < -\frac{Q}{R} < 2$ | null information if $-\frac{Q}{R} < 2$ |
| | full information if $-\frac{Q}{R} > 2$ | full information if $-\frac{Q}{R} > 2$ |
| $R > 0, Q < 0$ | ——— | private signals $\left(q = \frac{d}{3d-c}, r = 0\right)$ if $-\frac{Q}{R} > \frac{c-3d}{2(c-d)}$ |
| | private signals $\left(q = \frac{2c-d}{3c-d}, r = \frac{c-d}{3c-d}\right)$ if $-\frac{Q}{R} > 1$ | public signal $\left(q = \frac{d}{c+d}\right)$ if $1 < -\frac{Q}{R} < \frac{c-3d}{2(c-d)}$ |
| | full information if $-\frac{Q}{R} < 1$ | full information if $-\frac{Q}{R} < 1$ |
| $R < 0, Q < 0$ | public signal $\left(q = \frac{d}{c+d}\right)$ if $-\frac{Q}{R} < \frac{c-3d}{2(c-d)}$ | private signals $\left(q = \frac{d}{3d-c}, r = 0\right)$ always |
| | private signals $\left(q = \frac{2c-d}{3c-d}, r = \frac{c-d}{3c-d}\right)$ if $-\frac{Q}{R} > \frac{c-3d}{2(c-d)}$ | ——— |

Full information: $(q = 1, r = 1)$; null information: $(q = \frac{1}{2}, r = 0)$; public signals: $q = r$.

# References

[1] Abreu, D., H. Matsushima (1992), "Virtual Implementation in Iteratively Un-dominated Strategies: Complete Information," *Econometrica*, 60, 993-1008.

[2] Angeletos, G.-M., A. Pavan(2007), "Efficient Use of Information and Social Value of Information," *Econometrica*, 75(4), 1101-1142.

[3] Angeletos, G.-M., A. Pavan(2009), "Policy with Dispersed Information," *Journal of the European Economic Association*, 7(1), 11-60.

[4] Bergemann, D., S. Morris (2003),"Robust Mechanism Design," *Econometrica*, 73, 1771-1813.

[5] Bergemann, D., S. Morris (2009), "Robust Implementation in Direct Mecha-nisms," *Review of Economic Studies*, 76, 1175-1204.

[6] Bergemann, D., S. Morris, O. Tercieux (2011),"Rationalizable Implementation," *Journal of Economic Theory*, 146, 1253-1274.

[7] Bergemann, D. and M. Pesendorfer (2007). Information Structures in Optimal Auctions. *Journal of Economic Theory* **137**: 580-609.

[8] Bergemann, D., T. Heumann, S. Morris (2013), "Information, Interdependence, and Interaction: Where Does the Volatility Come From?," *mimeo*, 1-82.

[9] Bergemann, D., S. Morris (2013), "Bayes Correlated Equilibrium and the Com-parison of Information Structures," *mimeo*, 1-90.

[10] Bergemann, D., S. Morris (2013), "Robust Predictions in Games with Incomplete Information," *Econometrica*, 81(4), 1251-1308.

[11] Brandenburger, A. and E. Dekel (1989), "The Role of Common Knowledge Assumptions in Game Theory," *in* The Economics of Missing Markets, Information Games (F. Hahn, Ed.), Clarendon Press, Oxford.

[12] Cabrales, A. (1999), "Implementation in Adaptive Better-Response Dynamics: Towards a General Theory of Bounded Rationality in Mechanisms," *Games and Economic Behavior*, 86, 159-184.

[13] Cabrales, A., R. Serrano (2011), "Adaptive dynamics and the implementation problem with complete information," *Journal of Economic Theory*, 86, 159-184.

[14] Chen, Y. (2002), "A Family of Supermodular Nash Mechanisms Implementing Lindahl Allocations," *Economic Theory*, 19, 773-790.

[15] Chen Y., C. R. Plott (1996), "The Groves-Ledyard Mechanism: An Experimental Study of Institutional Design," *Journal of Public Economics*, 59, 335-364.

[16] Chen Y., F-F. Tang (1998), "Learning and Incentive Compatible Mechanisms for Public Goods Provision: An Experimental Study," *Journal of Political Economy* 106, 633-662.

[17] Chen Y., R. Gazzale (2004), "When Does Learning in Games Generate Convergence to Nash Equilibria? The Role of Supermodularity in an Experimental Setting," *American Economic Review*, 94, 1505-1535.

[18] Cho, I-K., and D. M. Kreps (1987). Signaling Games and Stable Equilibria. *Quarterly Journal of Economics* **102**: 179-221.

[19] Chung, K-S., J. Ely (2002), "Ex-post Incentive Compatible Mechanism Design," *mimeo*, Northwestern University.

[20] Crawford, V., J. Sobel (1982), "Strategic Information Transmission," *Econometrica*, 50(6), 1431-1451.

[21] Cremer, J., R. McLean (1985), "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent," *Econometrica*, 53, 345-361.

[22] Daley, B. and B. Green (2012). Market Signaling with Grades. Mimeo.

[23] Dasgupta, P. S., E. Maskin (2000), "Efficient Auctions," *Quarterly Journal of Economics*, 115, 341-388.

[24] Demmel, J. (1997), "Applied Numerical Linear Algebra," *SIAM*, Philadelphia.

[25] Echenique F., A. Edlin (2004), "Mixed Equilibria are Unstable in Games of Strategic Complementarities," *Journal of Economic Theory*, 118, 61-79.

[26] Echenique F. (2000), "Comparative Statics by Adaptive Dynamics and The Correspondence Principle," *Econometrica*, 70(2), 833-844.

[27] Eliaz, K., F. Forges (2012), "Disclosing Information to Interacting Agents," *mimeo*, 1-15.

[28] Eso, P., and B. Szentes (2007). Optimal Information Disclosure in Auctions and the Handicap Auction. *Review of Economic Studies* **74(3)**: 705-731.

[29] Farrell, J., R. Gibbons (1989), "Cheap Talk with Two Audiences," *American Economic Review*, 79(5), 1214-1223.

[30] Gershkov, A. (2002). Optimal Auctions and Endogenous Information Structure. Technical Report, Hebrew University.

[31] Goltsman, M., G. Pavlov (2011), "How to Talk to Multiple Audiences," *Games and Economic Behavior*, 72, 100-122.

[32] Gossner, O. (2000), "Comparison of Information Structures," *Games and Economic Behavior*, 30, 44-63.

[33] Healy P. J. (2005), "Learning Dynamics in Mechanism Design: An Experimental Comparison of Public Goods Mechanisms," *Journal of Economic Theory*, in press.

[34] Jackson, M. O. (2009). Non-Existence of Equilibrium in Vickrey, Second-Price, and English Auctions. *Review of Economic Design* **13**: 137-145.

[35] Jackson M. O. (1992), "Implementation in Undominated Strategies: A Look at Bounded Mechanisms," *Review of Economic Studies*, 59(201), 757-775.

[36] Jehiel P., B. Moldovanu (1996), "Strategic Nonparticipation," *RAND Journal of Economics*, 27(1), 84-98.

[37] Jehiel P., B. Moldovanu (2001), "Efficient Design with Interdependent Valuations," *Econometrica*, 65(5), 1237-1259

[38] Jehiel P., M. Meyer-ter-Vehn, B. Moldovanu, W. R. Zame (2006), "The Limits of Ex-Post Implementation," *Econometrica*, 74(3), 585-610.

[39] Kamenica, E., M. Gentzkow (2011), "Bayesian Persuasion," *American Economic Review*, 101, 2590-2615.

[40] Ledyard, J. (1986), "The Scope of the Hypothesis of Bayesian Equilibrium," *Journal of Economic Theory*, 39, 59-82.

[41] Lehrer, E., D. Rosenberg, E. Shmaya (2010), "Signaling and Mediation in Games with Common Interest," *Games and Economic Behavior*, 68, 670-682.

[42] Lehrer, E., D. Rosenberg, E. Shmaya (2011), "Garbling of Signals and Outcome Equivalence," *Discussion Paper*, Tel Aviv University, University of Paris and Northwestern University.

[43] Marczewski, E. (1930), "Sur l'extension de l'ordre partiel," *Fundamenta Mathematicae (in French)*, 16, 386-389.

[44] Maskin, E. (1992), "Auctions and Privatization," in *Privatization*, ed. by H. Siebert, Tubingen: Mohr, 115-136.

[45] Mathevet, L. (2010), "Supermodular Mechanism Design," *Theoretical Economics*, 5, 403-443.

[46] Mathevet, L. (2011), "Beliefs and Rationalizability in Games with Complementarities," *mimeo*.

[47] Milgrom, P.R., and R.J. Weber (1982). A Theory of Auctions and Competitive Bidding. *Econometrica* **50**: 1089-1122.

[48] Milgrom P., J. Roberts (1990), "Rationalizability Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica*, 58, 1255-1277.

[49] Milgrom P., J. Roberts (1991), "Adaptive and Sophisticated Learning in Normal Form Games," *Games and Economic Behavior*, 3, 82-100.

[50] Myerson, R. (1979), "Incentive Compatibility and the Bargaining Problem," *Econometrica*, 47, 61-73.

[51] Paige C. C., M. A. Saunders (1975), "Solutions of Sparse Indefinite Systems of Linear Equations," *SIAM Journal of Numerical Analysis*, 12, 617-629.

[52] Perry, M., P. Reny (2002), "An Efficient Auction," *Econometrica*, 70, 1199-1213.

[53] Peski, M. (2008), "Comparison of Information Structures in Zero-Sum Games," *Games and Economic Behavior*, 62, 732-735.

[54] Sefton, M., A. Yavas (1996), "Abreu-Matsushima Mechanisms: Experimental Evidence," *Games and Economic Behavior*, 16, 280-302.

[55] Stinchcombe, M. B. (2011), "Correlated Equilibrium Existence for Infinite Games with Type-Dependent Strategies," *Journal of Economic Theory*, 146(2), 638-655.

[56] Topkis D.M. (1998), Supermodularity and Complementarity, Princeton, New Jersey.

[57] Vives X. (1990), "Nash Equilibrium with Strategic Complementarities," *Journal of Mathematical Economics*, 19, 305-321.

[58] Wang, Y. (2012), "Bayesian Persuasion with Multiple Receivers," *Job Market Paper*, Department of Economics, University of Pittsburgh, 1-45.

[59] Williams S., R. Radner (1988), "Informational Externalities and the Scope of Efficient Dominant Strategy Mechanisms," Discussion Paper No. 761, Northwestern University.