BELIEF EXTRACTION IN MECHANISM DESIGN

Benjamin Aaron Brooks

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF ECONOMICS

Advisors: Dilip Abreu and Stephen Morris

June 2014

Abstract

In many settings, the designer of an institution is less informed about the economy than are the agents who will ultimately participate in that institution. This dissertation explores how such an uninformed designer can learn features of the economy that are common knowledge among the agents and use the extracted information to design better institutions.

Chapters 1 and 2 study how an uninformed seller can induce potential buyers to reveal the revenue maximizing reservation price for an auction. Chapter 1 explores mechanisms in which a seller runs a sealed-bid second-price auction and simultaneously surveys the buyers' beliefs about others' valuations. The seller offers bets that incentivize truthful reporting of beliefs, and for a general class of environments, truth-telling is the unique equilibrium. Losing bidders' reports are used to set an interim optimal reserve price for the winner. As a result, these mechanisms guarantee the seller an optimal worst-case revenue-share of the efficient surplus.

Chapter 2 considers sealed-bid second-price auctions in which instead of reporting beliefs, each bidder recommends a reservation price to be used when they lose the auction. If the recommendation is used, the bidder is rewarded with a small share of revenue. Revenue sharing aligns the incentives of the seller and losing buyers, but creates an incentive to "throw" the auction when a buyer expects to win at a price close to his valuation. When the distribution of valuations satisfies a monotone hazard rate assumption, the mechanism has a symmetric and monotonic equilibrium. As the bidders' revenue shares go to zero, bid shading disappears and the equilibrium results in optimal reserve prices.

Chapter 3 explores general mechanisms that a designer can use to extract common knowledge for the purpose of building that information into a mechanism. For private-good environments such as those considered in Chapters 1 and 2, mechanisms are constructed that allow the designer to recover the agents' common knowledge at arbitrarily small cost to any ultimate mechanism design goals.

Acknowledgments

I would like to express my sincere thanks to the many individuals without whose support I would never have been able to complete this body of work. I am eternally grateful to my advisors Dilip Abreu and Stephen Morris, who have displayed a seemingly endless generosity with their time and wisdom. In addition to being excellent economists, educators, mentors, and colleagues, they are both true mentshn.

Dirk Bergemann, Sylvain Chassang, and Yuliy Sannikov have had a profound impact on my development as an economist and on my research, for which I am extremely grateful. The experience of collaborating with Karla Hoff and Priyanka Pandey has been hugely enriching and educational. I have also greatly benefited from discussing my research with Roland Benabou, Marco Battaglini, Faruk Gul, Bo Honoré, Myrto Kalouptsidi, Wolfgang Pesendorfer, Satoru Takahashi, and Juan Pablo Xandri.

I am very grateful to Kathleen DeGennaro, Ellen Graf, Laura Hedden, Geralyn McDermott, Matthew Parker, and Barbara Radvanady for always providing assistance when I have been in need.

Thank you to my colleagues Nemanja Antic, Peter Buisseret, Gonzalo Cisternas, Moshe Katzwer, Rohit Lamba, Jay Lu, Juan Ortner, Takuo Sugaya, and Juha Tolvanen, from whom I have learned a great deal and who have the basement of Fisher Hall a brighter place to work. Also thanks to my dear friends Karyne Charbonneau, Jacob Goldin, Nikolaj Harmon, Tatiana Homonoff, Yan Lau, Quynh Nguyen, Deborah Schwartz, and Nicholas and Siobhan Skizim. These past six years would not have been the same without you.

Thanks to my Labrador retriever Lola, whose warm disposition and affinity for belly rubs has often "given my heart a change of mood/and saved some part of a day I had rued."

I have been blessed with a loving and caring family: my cousins Paul, Jennifer, Naomi, David, Madeline, Amelia, Lilly, Jocelyn, and Jemma; my aunt and uncle Linda and Dennis; my sister-in-law Riva; my siblings Lee and Sarah; and my parents Risa and Chuck. And finally, thank you to Kate. You have been a constant source of support and guidance throughout all the time that I have known you.

To my parents, Risa and Charles.
They value education highly and have given much that mine should be excellent.

Contents

Chapter 1

Introduction

Many important problems of institutional design involve complex informational frictions. The value of one outcome over another can depend on many characteristics, only some of which will be known to any single individual. Information is both incomplete, in that strategic agents can only partially specify the states of the world that affect preferences, and differential, in that agents generally do not all have the same information. The mechanism design literature has studied design under such incomplete information almost since its inception, but the classical literature has generally made an important simplifying assumption: that both the mechanism designer and the agents in the economy share a common prior distribution over the variables that determine preferences, and that it is from this common prior that agents' beliefs are derived.

That the designer and the agents share the same common prior can be a natural assumption in many settings. But what if the designer is less informed, in that there is some feature of the economy that is common knowledge among the agents but is unknown to the designer? The designer might be a government agency deciding how to allocate spectrum licenses to telecom firms, though the firms know much more than the government about the value of a given block of frequencies. Or perhaps the designer is a general purpose auction house selling some rare work of art, in which case art collectors may be much better at judging an appropriate starting bid. In either case, the designer will want to learn more about the market in which the mechanism will operate in order to improve the design. If there are uninterested third-parties who do not care about the outcome but are nonetheless well-informed, then the designer could hire them as consultants and effectively restore common knowledge of the common prior. On the other hand, if the informed parties all have a stake in the decisions made by the institution, then it may not be so easy to incentivize them to share their common knowledge, since by misreporting they may be able to manipulate the design and induce a more preferred outcome.

This is the problem of belief extraction in mechanism design: to learn the common knowledge of the agents, as represented by their beliefs, for the purpose of incorporating that common knowledge into the design of a mechanism. Ideally, the designer would like to incentivize the agents to reveal as much information as possible, while maintaining the flexibility to use that information as the designer sees fit. These goals may be at odds when there is conflict of interest between the designer and the agents, since the more the designer uses extracted information to customize the mechanism, the greater are the opportunities for manipulation through misreporting. This dissertation investigates the extent to which the designer can achieve these goals. The key finding is that in many settings, there are techniques for recovering the agents' common knowledge that do not require the designer to compromise on how the information is used.

Without the classical assumption that the designer shares the same common prior as the agents, important issues arise in the modeling of the mechanism design problem. First, though the designer may have traditional preferences conditional on a given specification of the prior, how should the designer's preferences incorporate the ambiguity as to which prior is correct? One approach would be for the designer to look for mechanisms that provide a good approximation of some benchmark, regardless of which is the true prior. Alternatively, the designer may look at more than just performance relative to a benchmark and prefer mechanisms that never perform worse and sometimes perform better. These different approaches to modeling the designer's preferences may lead to very different optimal mechanisms. Second, the extent to which the mechanism can be specialized for the true prior implicitly depends on how much information the designer extracts from the agents. Greater information extraction may well necessitate more complicated mechanisms with richer message spaces. At the same time, practically useful mechanisms need to be easy for the agents to understand, and any justification for truthful reporting should be transparent. Third, though there is ambiguity about the prior, the designer may be willing to make additional assumptions that limit the scale of the ambiguity to facilitate specialized mechanisms. However, such specialization comes at the expense of wider applicability of the mechanism.

In the following chapters, I provide three complementary approaches to belief extraction in mechanism design that explore different compromises between optimality, complexity, and generality of the mechanism. Throughout, I will focus on private-good allocation problems in which there is a set of objects to be distributed, and each agent is only concerned with their own allocation and not the allocations of others. The chapters consider various ways in which this broad framework can be specialized: Chapters 1 and 2 look at single-unit allocation with private values and quasilinear preferences, whereas Chapter 3 addresses multi-unit allocation with interdependent expected utility preferences. The preferences of the designer range from the limited goal of optimal worst-case revenue guarantees in Chapters 1 and 2, to demanding optimality conditional on each possible environment in Chapter 3. For each specification of the designer's assumptions and preferences, I construct mechanisms that achieve the designer's goals. Generally speaking, less demanding preferences and more restrictive assumptions lead to simpler optimal mechanisms. Taken as a whole, the results provide a broad view of the possibilities for closing the gap in knowledge between the designer and the agents, even when the agents rationally anticipate that the information they reveal will be used to advance the designer's interest.

To start, Chapter 1 considers a seller of a single unit of a good for which the potential buyers have private valuations. The seller is ambiguity averse and would like to design a mechanism that provides favorable revenue guarantees for all possible true environments. Of course, it is possible that with probability one every buyer has a valuation of zero, and in such an environment, all mechanisms would generate a revenue of zero. Consequently, if the seller simply evaluated mechanisms by their worst-case revenue, all designs would perform equally badly. Instead, the seller considers the revenue generated by a mechanism relative to the potential for generating revenue, as measured by the surplus that would be generated if the good were allocated to the buyer who values it the most. I refer to this ratio of expected revenue to the expected efficient surplus as the *extraction ratio*, and the seller's goal is to maximize the worst-case extraction ratio. The extraction ratio has been previously studied by other authors, notably Neeman (2003), who characterized the extraction ratio of a second-price auction with a reserve price. The environment itself is described with the standard modeling device of a type space, in which each buyer has a type that contains a private value and also a belief about others' types. The seller further supposes that there are limitations on how dispersed the buyers' valuations can be, in the form of a constraint on the size of buyers' valuations relative to the average highest value. The main result is to characterize the best extraction ratio that the seller can guarantee himself as a function of the constraint on dispersion. This max min extraction ratio turns out to be substantial: if values can be at most ten times larger than the efficient surplus, then the seller is guaranteed at least 20% of the efficient surplus as revenue; if values can be a hundred times larger, the seller is guaranteed a 10% extraction ratio. The extraction ratio declines as the bound is relaxed, though it does so at a relatively slow rate. Even if values can be ten million times larger, the seller is still guaranteed an extraction ratio of 5%.

In the process of characterizing the max min extraction ratio, I construct a class of mechanisms that guarantee extraction ratios arbitrarily close to the optimal lower bound. These *belief survey auctions* have a relatively simple structure: the seller runs a sealed-bid second-price auction and also asks the bidders to report their beliefs about order statistics of others' bids. The seller adds noise bids with small probability, so that bidding one's value is uniquely optimal. Furthermore, bidders are incentivized to report their true beliefs using carefully structured bets known as scoring rules. Such techniques have also been used by Azar, Chen, and Micali (2012) for belief extraction in mechanism design. The bids and belief reports of losing bidders are then used to set an interim optimal reserve price for the high bidder. This calibrated reserve price overcomes dangers associated with using a fixed reserve that is independent of the type space: if the reserve price is set equal to zero, it might be that the high bid is positive whereas the second-highest bid is zero; and if the reserve price is greater than zero, it might have been set too high, so that no one purchases the good. Either way, the resulting extraction ratio would be zero. By having the reserve price depend on the losing bidders' reported beliefs, the seller is guaranteed that revenue should not fall too low as long as there is some efficient surplus to be extracted. The critical environment at which the worst-case extraction ratio is attained has special structure that minimizes the amount of information conveyed by losing bidders' reports, and therefore constrains the effectiveness of the choice of reserve price. For these worst-case environments, the belief survey auction reduces to the optimal posted price mechanism, which is in fact an optimal mechanism for these environments. As a result, the bound on the extraction ratio guaranteed by the belief survey auction cannot be improved.

The belief survey auction is simple in two respects: (i) the mechanism is strategically straightforward in that truth-telling is the unique equilibrium for a very general class of environments, and this equilibrium can be arrived at using two rounds of iterated deletion of strictly dominated strategies; and (ii) the mechanism only elicits private values and first-order beliefs about statistics of others' values, as opposed to say higher-order beliefs of arbitrary degree. With regard to (ii), the seller never actually learns the common prior and does not use the globally optimal mechanism or even reserve price conditional on the true type space, but nonetheless the mechanism guarantees an optimal lower bound on performance as measured by the extraction ratio. In spite of these positive attributes, the mechanism is still complicated in that a large amount of information must be elicited from the buyers, and truthful revelation is incentivized through the subtle logic of the scoring rule.

One might ask if there are mechanisms which achieve similar objectives, but elicit less information and provide more intuitive incentives. Such an auction might be desirable if, say, the bidders have proprietary information about their competitors which they would prefer not to reveal, or if there are practical limitations on how complex the mechanism can be. Chapter 2 considers mechanisms in which the seller runs a second-price sealed-bid auction and simply asks bidders what they think is an appropriate reserve price. The seller then selects one of the losing bidders at random and implements that bidder's suggested reserve price. The incentives of losing bidders are aligned with those of the seller by giving each bidder a small amount of equity in the revenue generated when using that bidder's suggestion. Thus, bidders will report the reserve price that maximizes revenue conditional on them losing the auction. While incentives for truthful reporting of reserve prices are very clean, revenue sharing distorts incentives to truthfully bid one's value. The reason is that sharing in revenue might be more attractive than winning the good at a price close to one's valuation. As a result, the buyers shade their bids in equilibrium.

Even though some shading must occur, I show that a small amount of revenue sharing creates only a small amount of shading. Thus, by giving the bidders a small amount of equity, the seller is able to recover optimal reserve prices at minimal cost. To be more specific, I consider environments in which bidders' private values are drawn from a symmetric and differentiable distribution with compact support that also satisfies a monotone hazard rate condition. This condition requires that values be positively correlated in a certain sense and plays a similar role as the affiliation assumption of Milgrom and Weber (1982). For this class of environments, I construct an intuitive equilibrium in which bid shading balances the benefits of winning the good outright versus sharing in revenue. The equilibrium is characterized by a monotonic bidding function that is defined piecewise as either being (i) equal to one's value, when suggesting a reserve that is strictly greater than the bid, or (ii) the solution of a differential equation, when the bidder is suggesting a reserve price equal to their bid. In case (ii), the trade-off between winning and revenue sharing gives rise to an interior optimum, and the first-order condition defines the differential equation. In addition, I show that the equilibrium bid is bounded below by the bidder's valuation divided by one plus the revenue share. Hence, as revenue sharing goes to zero, bids must converge to values, and suggested reserves converge to the optimal quantities with respect to the prior distribution. As a result, by making the bidders' equity sufficiently small, these mechanisms guarantee the seller revenue that is arbitrarily close to that of a second-price auction with an optimally chosen reserve price.

Relative to the belief survey auction, the revenue sharing mechanism has a simpler message space and more transparent incentives for truthful reporting. However, these benefits must be balanced against the greater complexity of equilibrium behavior, which involves bid shading according to a non-trivial formula, as opposed to the truthful bidding supported in the belief survey auction. In addition, solving for an explicit equilibrium requires additional structure on the type space, such as that the buyers do not get additional information about one another's values in addition to their own values and differentiability of the joint distribution. In contrast, the belief survey auction is characterized for much more general type spaces in which buyers can learn much more about others' values than just the information contained in their own value. Thus, there are trade-offs between these two mechanisms along the dimensions of complexity and generality.

A feature that these models have in common, however, is that they both posit a designer with a relatively coarse notion of optimality. In Chapter 1, the seller simply wishes to maximize the minimum possible extraction ratio, whereas in Chapter 2, the seller wants to guarantee himself the revenue from a second-price auction with an optimally chosen reserve price. For each objective, there is no assertion that the mechanisms constructed are uniquely optimal. In fact, there are many mechanisms that always generate as much revenue as the belief survey and revenue sharing auctions and even generate strictly greater revenue in some type spaces. All complexity concerns aside, we would expect the seller to prefer such mechanisms that dominate uniformly. Indeed, the first best outcome for the seller would be to recover the true type space and then implement the mechanism that maximizes revenue conditional on the true environment. Under weak implementation concepts, such an outcome is easy to achieve using old folk arguments from the complete information mechanism design literature. For example, each bidder announces a type space, and if the report is unanimous, the seller implements the revenue maximizing mechanism. Otherwise, no one receives the good and all bidders have to pay a large fine to the seller. This mechanism does support truthful reporting of the type space as an equilibrium, but it also supports other non-truthful equilibria in which bidders coordinate on a false report. This begs the question, are there mechanisms in which all equilibria involve truthful revelation of the type space, at minimal cost to the seller's goals?

This is precisely the subject of Chapter 3. The setting is generalized to private-good allocation problems with interdependent preferences, of which the single-unit, private-value, and quasilinear environments of Chapters 1 and 2 are special cases. The designer identifies a collection of type spaces that are considered to be possible, and for each type space in the collection specifies a mechanism that the designer would like to implement conditional on the given type space being the correct description of beliefs. Such a specification is referred to as a mechanism mapping. This formulation implicitly allows the designer to have diverse objectives that motivate the choice of mechanism for each type space. The goal is to find a single uniform mechanism which will be "close" to the desired mechanism on each given type space. Closeness is formalized with a notion of strategic equivalence of mechanisms, in which two mechanisms are equivalent for a given type space and up to a certain distance if it is possible to identify rationalizable messages under one mechanism with rationalizable messages in the other mechanism such that (i) lotteries induced by identified message profiles are within the given distance, and (ii) the difference in the lottery between two rationalizable message profiles in one mechanism is proportional to the difference in the lottery between identified message profiles in the other mechanism. This definition of equivalence allows the designer to augment the desired mechanism by adding messages and slightly perturbing outcomes in order to elicit the type space. However, the additional messages must not be rationalizable, and the perturbation should not affect preferences over rationalizable message profiles. Both the model and the solution concept draw heavily on a recent paper by Bergemann, Morris, and Takahashi (2011) which studies belief extraction, though not for the purpose of implementation. If it is possible to find a uniform mechanism which is strategically equivalent to the desired mechanism on each type space at arbitrarily small distances, I will say that the mechanism mapping is uniformly virtually implementable.

There are some intuitive and unavoidable constraints on which kinds of mechanism mappings can be implemented in this manner. For example, if two type spaces are identical except for different names for the types, then no single mechanism can guarantee that agents from the different type spaces will separate themselves and effectively play non-strategically equivalent mechanisms. The chapter shows that in order for a mechanism mapping to be uniformly virtually implementable, it must satisfy a condition which I call local preference measurability. As is shown by Bergemann et al., a type in a type space can be identified with a hierarchy of preferences, which consists of that type's unconditional preference over outcomes, the type's preference conditional on others' unconditional preferences, etc. Local preference measurability of a mechanism mapping requires that if two type spaces have subsets of types which have the same higher-order preferences, then the desired mechanisms must be strategically equivalent on the overlap. Local preference measurability roughly corresponds to a measurability condition identified by Abreu and Matsushima (1992b), which says that two types cannot be strictly incentivized to separate themselves if they have the same higher-order preferences. In addition to being necessary, local preference measurability is also a sufficient condition for uniform virtual implementability as long as it is possible to punish players with some undesirable and state-independent outcome, as in the quasilinear setting of Chapters 1 and 2. This result has implications for the revenue maximization problems discussed previously. In particular, Chapter 3 shows that it is always possible for the seller to implement the revenue maximizing mechanism conditional on the smallest belief closed subspace of the true type space at an arbitrarily small cost. However, the uniform equivalent mechanism constructed in Chapter 3 is quite complicated and requires players to report their entire hierarchy of higher-order preferences. As such, this result should be interpreted as a benchmark of what the designer can accomplish with arbitrarily complex mechanisms.

To sum up, these chapters provide complementary approaches to mechanism design when the designer does not possess information that is common knowledge among the agents. If the designer is willing to use arbitrarily complicated mechanisms, then subject to the necessary condition of local preference measurability, it is possible to extract the common knowledge of the agents and use this to implement any desired mechanism. On the other hand, if the designer uses a weaker notion of optimality in the form of optimal approximation of simple benchmarks, then simpler mechanisms are also optimal. In the context of revenue maximization in the single-good private-value allocation problem of Chapters 1 and 2, the seller can achieve substantial performance guarantees with mechanisms that only extract as much information as required to implement optimal reserve prices.

This dissertation enhances our understanding of belief extraction in mechanism design, but there remain many important open questions. The chapters all rely on the private-good structure, that an allocation can be taken away from one agent while leaving others' allocations and preferences unchanged. In many economically important settings, there is a public aspect to the good in that the allocation must be received by all of the agents or by none of them. Also, the current results rely on the existence of some state-independent punishment outcome, such as not receiving an allocation and paying a fine. It remains to be seen what kind of implementation results are possible when the private-good and state-independent punishment assumptions are relaxed. Additionally, I have highlighted the variation in complexity of the mechanisms constructed in the different chapters. Complexity has been evaluated in a purely subjective manner, and there are many different notions that one could use. For example, there is the complexity of the message space, the complexity of describing equilibrium behavior, or even the complexity of the logic behind equilibrium. Ideally, these notions should be given a formal description and explicitly incorporated into the designer's preferences. Finally, the result of Chapter 3 gives sufficient conditions under which a mechanism mapping can be virtually implemented, but it does not provide a more general characterization of which social choice functions can be virtually implemented as in Abreu and Matsushima (1992b). I hope to revisit these topics in future work.

Chapter 2

Surveying and selling: Belief and surplus extraction in auctions

2.1 Introduction

2.1.1 Why survey?

Consider a small municipality that is replacing a public school building. The replacement of schools is generally a rare event, with the average age of public schools in the U.S. being 42 years.[1] As such, it is reasonable to suppose that municipal officials do not have great expertise in assessing construction costs. On the other hand, the firms that bid for the contract are likely to have detailed knowledge of one another's costs and capabilities. Is it possible for the municipality to get the contractors to truthfully reveal what they know about one another's costs, even though the information they reveal will influence the award?

[1]According to the National Center for Education Statistics, as of 1999.

The elicitation of potential buyers' opinions by a seller is more than just a theoretical possibility. After the use of auctions for allocating radio spectrum was authorized by the U.S. Congress, the Federal Communications Commission (FCC) elicited feedback on its proposed rules from potential bidders and industry experts. The FCC received "written comments from 222 parties and reply comments from 169 parties" (FCC, 1997, p. 9). Such feedback was no doubt crucial to gauging the welfare effects of the new mechanism. The FCC does not specify when and how it incorporated this feedback into the auction design, but surely the responses of the interested parties were influenced by their strategic concerns vis-à-vis the ultimate allocation and costs of licenses.

In this chapter, I will consider such situations, in which the seller of a good is *uninformed* about demand, whereas the potential buyers are *well-informed*. By well-informed, I mean that each agent knows their own private valuation for the good. In addition, they have a belief about others' values which is derived from a common prior. The buyers' private valuations and beliefs can be thought of as being induced by informative signals, with the common prior corresponding to the ex-ante distribution over the signals. The set of signals together with the prior specify a *type space*, where a buyer's "type" is precisely the realized signal.[2] The seller could greatly benefit from knowing the type space: at the very least, such knowledge could facilitate the selection of a revenue enhancing reserve price, and in particular cases, the seller can even use variation in bidders' beliefs to extract all of the potential surplus as revenue. However, the seller in my model does not know the type space, and therefore he cannot build such detail-dependent features

[2]Aside from requiring the common prior and that the set of types is finite (for tractability), I impose no additional restrictions on the type space. The set of possible environments is therefore quite general, and in fact includes type spaces for which there is no known characterization of the revenue maximizing Bayesian mechanism. See Farinha Luz (2013) for probably the most general characterization to date.

directly into the mechanism. Moreover the seller only interacts with the buyers after their types have been realized, and thus cannot easily incentivize them to reveal the prior. The remaining option, which the seller takes advantage of, is to use a mechanism that determines the allocation and transfers while simultaneously eliciting the buyers' interim beliefs, i.e., their beliefs after they learn their types but without knowing others' types. In this way, the outcome of the mechanism can be made to depend on the true type space. The buyers are of course aware of the effects of their reports, and will take advantage of any opportunity to misreport in order to favorably influence the outcome of the mechanism.

Given such large uncertainty about the type space, it is natural for the seller to use a worst-case criterion: the seller seeks a mechanism that will perform well irrespective of the true distribution of values and beliefs. Since it is possible for the buyers' valuations to be arbitrarily small, every mechanism has zero expected revenue in the worst-case. As a result, worst-case expected revenue is not a useful criterion to distinguish between mechanisms. Instead, I posit that the seller evaluates the performance of a mechanism by its expected revenue relative to the expected surplus that could be generated by allocating the good efficiently. I term this metric the *extraction ratio*: the ratio of expected revenue to expected efficient surplus. In addition, the seller makes no presumption that the buyers will behave according to his preferred equilibrium, so he evaluates a mechanism by its lowest extraction ratio over all type spaces and over all equilibria. Similar criteria have been considered in the literature, most notably by Neeman (2003) and by the computer science literature on mechanism design, surveyed in Hartline (2012). I will revisit the connections with these and other papers below and in some detail. By using a mechanism that maximizes the minimum extraction ratio, the seller will be guaranteed at least a minimum share of the expected efficient surplus, regardless of the true distribution of buyers' values and beliefs.

2.1.2 Overview of main results

My main result is that there is a simple class of mechanisms that the seller can use to achieve the max min extraction ratio. Moreover, this max min extraction ratio is economically substantial, as I will elaborate upon shortly. These mechanisms are essentially modified second-price sealed-bid auctions, in which the buyers simultaneously submit bids as well as respond to a survey of their beliefs about the values of others. The high bidder will be "offered" the good at a price determined using others' reports and ultimately receives the good if this price is less than the high bid. Because each buyer's bid does not affect the price of the good, but only whether or not the good is received at an exogenously chosen price, truthful bidding is a weakly dominant strategy. A slight perturbation makes bidding one's value strictly dominant. Also, given that others' will bid truthfully, buyers can be incentivized to report their true beliefs about others' values using a scoring rule. The use of scoring rules to elicit beliefs in mechanism design has also been considered by Azar, Chen, and Micali (2012). To calculate the price offered to the high bidder, the seller uses one of the losing bidders' survey reports as a "consultation" about the conditional distribution of the highest value. This consult, together with the second-highest bid, is used to compute an optimal price to offer the winner. I give these mechanisms the descriptive moniker of *belief survey auctions* (BSA), since the seller uses a survey of losing bidders' beliefs to set the winner's price.

I derive the minimum extraction ratio for the BSA, and I show that no other mechanism could achieve a greater extraction ratio in the worst-case. The strict incentives to bid one's value and report beliefs truthfully can be provided at arbitrarily small cost to the extraction ratio, so the BSA *virtually* achieves the max min. It turns out that if the support of valuations is unbounded, the max min extraction ratio is zero. The reason is that there are distributions of values that have arbitrarily large expected efficient surplus but also hold the seller to finite expected revenue. However, these type spaces are extreme in that they have a lot of mass in the tail of the distribution of the highest value. A natural assumption is that the support of the highest value is bounded by a constant multiple of the expected efficient surplus, effectively limiting the dispersion of the highest value around its mean. I study how the max min extraction ratio changes with the bound on the dispersion of values. For any bound, the max min extraction ratio is strictly positive, and even for very generous bounds on values, the max min is economically substantial. As an example, if buyers' values cannot be more than 10 times the expected efficient surplus, then the seller is guaranteed an extraction ratio of at least 20%. If values can be 1,000 times larger, the seller is still guaranteed a 10% extraction ratio.

It is particularly interesting that the seller is able to achieve these bounds with such simple mechanisms. The seller never recovers the prior distribution over values, but rather sets reserve prices using bidders' interim beliefs. As I will argue below, this is actually a virtue of the mechanism; bidders' interim beliefs are weakly more informative than the prior, and thus allow the seller to set better reserve prices at the interim stage. Moreover, the seller does not even need to elicit beliefs about the entire distribution of buyers' values; it is sufficient for the seller to ask bidders the conditional distribution of the top two valuations of other bidders and the number of bidders who tie for the highest value.

In addition to guaranteeing the seller a minimum extraction ratio, the BSA has desirable revenue properties away from the worst case. Aside from the small cost of providing strict incentives, the BSA guarantees the seller at least the revenue of a second-price auction with an optimal anonymous reserve price, i.e., a uniform reserve price for all bidders that maximizes expected revenue. As such, the BSA virtually maximizes expected revenue over all Bayesian mechanisms when the distribution of values is independent, symmetric, and regular, as in Myerson (1981).

A potential concern with the max min extraction ratio is that the seller seems to be indifferent between outcomes with very different expected revenues, as long as the expected efficient surplus varies proportionally. However, no such comparisons are necessary to justify the use of the BSA. An alternative way to model the seller's preference over mechanisms is the following conditional ordering: the seller prefers greater worst-case expected revenue conditional on the level of the expected efficient surplus, but he will not compare revenue outcomes between type spaces in which the social value of the good varies. As a result, one mechanism is preferred to another only if it has greater worst-case expected revenue conditional on *every* possible level of the expected efficient surplus. Observe that this ranking is only a partial order on the set of mechanisms, because the seller does not compare mechanisms whose worst-case revenue ranking switches depending on the surplus level. It turns out that the BSA is maximal with respect to this partial ordering: if the seller wishes to maximize the minimum expected revenue conditional on a particular level of the expected efficient surplus, then he can select no better mechanism than the BSA.

2.1.3 The logic behind the BSA

Here I will give a brief summary of how my results are obtained. The BSA offers the good to the high bidder, so on average, the winner's valuation is drawn from the distribution of the highest value among the n bidders. If the seller knew the prior distribution over values, he could set the reserve price which maximizes revenue without knowing the identity of the winner, which is the optimal anonymous reserve price. By assumption, the seller does not know the prior distribution over values, but the reports of the losing bidders allow the seller to set reserve prices conditional on more detailed information. Specifically, the seller learns the distribution of the winner's value conditional on (1) the winner not being the bidder who was consulted, (2) the realization of the second-highest bid, and (3) any extra information the consulted bidder has about the distribution, as encoded in his type. Conditional on (1)-(3), the seller can always set a reserve price that generates weakly more revenue than could be achieved with an optimally chosen anonymous reserve price. Thus, the BSA performs better on average than any second-price auction with an anonymous reserve, in spite of the fact that the seller never recovers the prior nor does he know the optimal anonymous reserve price.

There is an analogy to be made with third-degree price discrimination, in which a monopolist receives information that divides a market into segments. If the monopolist can set different prices in different segments, then this information must be weakly revenue increasing, since it is always feasible to set the optimal uniform price. A similar property holds in the auction setting. More informative reports by losing bidders allows the seller to set better reserve prices, and hence the worst-case environments for the BSA are ones in which (1)-(3) are minimally informative. These type spaces exhibit the property that bidders get no information beyond their private values, which minimizes the informativeness of (3). Moreover, the worst-case type spaces are lopsided, in the sense that at any time there is only one "serious" bidder who submits the high bid, and the other bidders know that they will not win, which minimizes the learning from (1) and (2). In a sense, this reduces the seller's problem to designing a mechanism for selling to a single serious buyer. Even so, multiple bidders will participate in the auction, and their reports are used by the seller to set an optimal reserve price for that single buyer. Finally, I derive the distribution for the serious bidder's value that minimizes revenue, subject to a given level of the efficient surplus.

With additional restrictions on the environment, the seller can achieve the same goals with mechanisms that are even simpler. Throughout the analysis, careful attention is paid to the possibility of multiple bidders having the same valuation, so that the winner is determined by a tie break. The tie break induces a selection effect: conditional on winning the auction, the winner is less likely to have a valuation at which ties are likely to have occurred. For this reason, the seller must survey bidders' beliefs about the likelihood of ties. One might think of ties as being a non-generic phenomenon, for example if values are drawn from a non-atomic distribution. If attention is restricted to type spaces in which ties do not occur with positive probability, the conditional distribution of the high bidder's value can be calculated much more simply. Also, as mentioned in the previous section, the seller elicits the buyers' beliefs about the top two valuations among other buyers. By leveraging the information about the highest value contained in the second-highest value, the BSA always perform better than a second-price auction with an optimal reserve price. However, if the seller is only concerned about worst-case extraction ratio, then the same bounds can be achieved with a mechanism that only elicits beliefs about the highest value of others.

2.1.4 Related literature

The results described above have a tight connection to the work of Neeman (2003), who studies the worst-case extraction ratio of the second-price auction. Neeman considers a seller who has three different levels of sophistication with regard to reservation prices. At the most basic level, the seller cannot set any reserve price. At the next level, the seller can use a fixed reserve price that is independent of the true type space. At the highest level of sophistication, the seller knows the distribution of values and is able to set the optimal anonymous reserve price for the true type space, although the seller is not sufficiently sophisticated to design and run the optimal auction. It is this last case that is the most relevant to the present chapter. For this setting, Neeman derives bounds on the extraction ratio that are equal to my own, albeit with a slightly different parametrization of the set of type spaces. Indeed, since the BSA always generates as much revenue as a second-price auction with an anonymous reserve, and since the revenue of these two mechanisms coincides on the worst-case type spaces derived by Neeman, it is necessarily the case that both have the same minimum extraction ratio. However, I will give a direct proof of worst-case type spaces for the BSA, to better illuminate the connection with third-degree price discrimination described above.

Another paper which is closely related is that of Azar, Chen, and Micali (2012). They also consider a seller who is uninformed about the type space while the agents are well-informed, and they look for general mechanisms that achieve a favorable worst-case performance relative to the benchmark of maximum revenue in a dominant strategy ex-post individually rational mechanism. Similar to the present work, they extract buyers' beliefs using scoring rules. They consider a restricted class of environments, for which the gap between first-order beliefs and the prior distribution is relatively small.[3] By eliciting the buyers' first-order beliefs, the seller is able to recover a truncated view of the prior, and this is used as an input into a dominant strategy mechanism. In comparison, the present work is in much more general environments, in which buyers can have arbitrary conditional beliefs about the distribution of values. As such, very different arguments are required to arrive at my results. Also, I use a different benchmark which does not assume a restriction to a particular implementation concept. Nonetheless, to achieve the $\max\min$ extraction ratio, it is sufficient for the seller to use simple mechanisms that only extract first-order beliefs about statistics of others' values.

More broadly, my work is part of the large literature on robust mechanism design (Bergemann and Morris, 2012b, provide an overview). At least since the critique of Wilson (1987), the mechanism design literature has held as a desideratum that mechanisms should be *detail-free*, in the sense that the rules of the game should not vary with fine details of the environment. This is in contrast to classical auction design, e.g., Myerson (1981) and Crémer and McLean (1988), in which the mechanism can be tailored to specific and highly structured type spaces. A more recent contribution of Farinha Luz (2013) considers very general type spaces but still allows the mechanism to depend on the type space.

The robust mechanism design literature has explored various ways to operationalize the Wilson critique. Much of the literature focuses on more stringent implementation concepts. For example, Bergemann and Morris (2009a, 2011) require that a particular social choice function be implemented regardless of the beliefs of the agents. I consider auction formats that are compatible with a slightly

[3]Specifically, they consider environments in which bidders' beliefs are derived from a common prior in the following manner: each bidder is associated with a partition of others' values, and bidders learn their own value and the cell of the partition containing other bidders' values. The first-order beliefs of different types of the same bidder have disjoint supports, and beliefs are always proportional to the prior distribution on their support.

different interpretation of the detail-free criterion: the mechanisms that the uninformed seller can use are detail-free in that the distribution of values and beliefs of the agents cannot be hard-wired into the mechanism. However, the outcome of the mechanism can depend on details of the environment through equilibrium behavior, if these details are known to the agents.

Other authors have considered criteria akin to max min extraction ratio. As discussed above, the closest such related work is that of Neeman (2003). Bergemann and Schlag (2011) consider a monopolist facing unknown demand from a single buyer, and characterize the pricing rule that achieves min max regret, which is the absolute difference between expected revenue and expected efficient surplus. Chassang (2013) studies dynamic incentive contracts, and solves for contracts that achieve a target that is analogous to max min extraction ratio. Carroll (2012) also considers max min preferences over contracts in a static setting. Chung and Ely (2007) give a foundation for dominant strategy mechanisms by positing a seller with worst-case preferences and who knows the distribution of private values but not the beliefs of the agents, which may be inconsistent with a common prior.

The criterion of max min extraction ratio is similar to the competitive ratios studied by computer scientists (see Hartline (2012) for a comprehensive survey). This literature looks at worst-case revenue ratios, with a variety of benchmarks in the denominator. The benchmark is often tailored to a specific solution concept, such as maximum revenue over all dominant strategy mechanisms. The efficient surplus is in a sense a more demanding benchmark, as it does not presume a restriction to a particular class of mechanisms. An assumption throughout some of the literature is that mechanisms can only elicit one-dimensional bids, which precludes the belief extraction approach of the present model. Chawla, Hartline, and Kleinberg (2007) and Hartline and Roughgarden (2009) study worst-case competitive ratios for the second-price auction with optimal reserve prices, which presumes that the seller knows the prior. Goldberg et al. (2004) and Goldberg and Hartline (2003) look at mechanisms which do not depend on the prior, with a benchmark which is the revenue the seller could generate selling $k \geq 2$ units of the good at the kth highest price. Such a benchmark could be zero in cases where the efficient surplus is positive.

Others have considered how a seller can learn about demand. Baliga and Vohra (2003) and Segal (2003) consider a seller who forecasts the distribution of values using past realizations. In contrast, I will look at a situation where the seller asks agents for their beliefs, rather than dynamic learning based on reported values. Caillaud and Robert (2005) construct detail-free mechanisms that use agents' beliefs to partially implement the optimal auction of Myerson (1981). Choi and Kim (1999) consider belief extraction in the context a public goods problem, but assume the existence of an ex-ante stage at which the seller can extract prior beliefs, before the realization of agents' private information.

Finally, this work is part of my broader investigation into mechanisms that harness the agents' beliefs about the environment, to make up for a lack of knowledge on the part of the designer. I see the present model as a midpoint in the trade-off between the simplicity of the mechanism and the strength of the optimality criterion. In Chapter 3, I will investigate the limits of how much the seller could learn about the environment. This relates to a classic "folk argument" in the mechanism design literature, that if a common prior were known to the agents and not to the designer, then the designer could recover the prior for free (Bergemann and Morris, 2012a), in the sense that the need to recover the prior does not restrict the social choice functions that the seller can implement once the prior is known. I show that the designer can indeed extract the prior, without compromising on how the prior will be used, by using a mechanism which elicits bidders' infinite hierarchy of beliefs. While complexity is not explicitly modeled, it is safe to say that this mechanism would be much more challenging to implement than the BSA. At the other end of the spectrum, Chapter 2 looks at mechanisms in which the seller runs a second-price auction and simply asks each bidder to suggest a reserve price for the other bidders. The seller incentivizes truth-telling by sharing revenue generated through a bidder's suggestion. In more structured type spaces, this mechanism has a natural equilibrium in which bids are close to values, and bidders suggest reserve prices that are approximately optimal. I revisit the broader agenda in Section 5.

The rest of this chapter is organized as follows. In Section 2.2, I describe the model and the seller's mechanism design problem. In Section 2.3, I present a simple example that illustrates some of the main ideas of the chapter. Section 2.4 presents the main results. Section 2.5 is a discussion, and Section 2.6 concludes. Omitted proofs appear at the end of the chapter.

2.2 Model

There are n potential buyers for a single unit of a private good, indexed by $i \in N = \{1, \ldots, n\}$. I adopt the usual convention that $-i = \{j \in N | j \neq i\}$, and vectors x_S denotes the sub-vector of x containing indices in S, e.g., $t = (t_i, t_{-i}) \in T$ is a profile of types. For real vectors x, $x^{(1)}$ denotes the highest value in x, $x^{(2)}$ denotes the second-highest value, and $x^{(1,2)}$ is the ordered pair of the highest and second-highest values. If x only has a single coordinate, then $x^{(2)} = -\infty$.

The values and beliefs of the bidders are modeled with the language of type spaces. In particular, there is a finite set of types $T = \times_{i \in N} T_i$ and a joint distribution $\pi \in \Delta(T)$. The notation $\Delta(X)$ denotes the set of probability measures on X with finite support. Each type $t_i \in T_i$ is associated with a private value:

$$\phi_i(t_i) \in \mathbb{R}.$$

Together, a *type space* is a triple $\mathcal{T} = (T, \pi, \phi)$. I will write $\pi(t_{-i}|t_i)$ for the conditional distribution of types given t_i, and $\pi_i(t_i)$ for the marginal distribution on T_i. For each type space $\mathcal{T}$, the expected surplus generated if the good were allocated efficiently is:

$$S(\mathcal{T}) = \sum_t \phi^{(1)}(t)\pi(t). \tag{2.1}$$

I write $\underline{v}$ and $\overline{v}$ for smallest and largest values in the support of the measure over values induced by π under the mapping ϕ.

A type space is *symmetric* if (T_i, ϕ_i) is the same for all bidders, and π is exchangeable in the types, i.e., $\pi(t_1, \ldots, t_n) = \pi(t_{\psi(1)}, \ldots, t_{\psi(n)})$ where ψ is a permutation of N. A *payoff* type spaces has the property that $\left|\phi_i^{-1}(v_i)\right| \leq 1$ for all $v_i \in \mathbb{R}_+$. In other words, each valuation is associated with at most one type, so all bidders with a given valuation have the same conditional beliefs about other bidders' types as well as other bidders' valuations. I will say that a type space is *lopsided* if with probability one, at most one bidder has a valuation above the minimum of the support. These type spaces are lopsided in the sense that the winner's valuation tends to be much larger than the second-highest value.

The seller must design an auction for the sale of the good. A mechanism consists of a measurable space of messages M_i for each player, with $M = \times_{i \in N} M_i$, and mappings $q : M \to \mathbb{R}^n$ and $p : M \to \mathbb{R}^n$. The quantity $q_i(m)$ is the probability that agent i is allocated the good, and $p_i(m)$ is agent i's net transfer to the seller when the message profile m is sent. Naturally, $q_i(m)$ is required to be non-negative and $\sum_{i\in N} q_i(m) \leq 1$. A *mechanism* is a triple $\mathcal{M} = (M, q, p)$.

A mechanism and a type space together define a Bayesian game, in which each player's strategy set is $\Sigma_i(\mathcal{M}, \mathcal{T}) = \{\sigma_i : T_i \to \Delta(M_i)\}$. I write $\sigma_i(dm_i; t_i)$ for the probability measure over bidder i's messages m_i given type t_i. For a strategy profile $\sigma \in \Sigma(\mathcal{M}, \mathcal{T}) = \times_{i\in N}\Sigma_i(\mathcal{M}, \mathcal{T})$ and type $t_i \in T_i$, bidder i's payoff is:

$$u_i(\sigma, t_i) = \sum_{t_{-i}\in T_{-i}} \pi(t_{-i}|t_i) \int_{m\in M} [\phi_i(t_i)q_i(m) - p_i(m)]\, \sigma(dm; t),$$

where $\sigma(dm; t) = \times_{i\in n}\sigma_i(dm_i; t_i)$ is the product measure on M. A profile σ is a *Bayesian Nash equilibrium* if:

$$\sigma_i \in \arg \max_{\sigma_i'\in\Sigma_i(\mathcal{M},\mathcal{T})} \sum_{t_i\in T_i} \pi_i(t_i) u_i\left((\sigma_i', \sigma_{-i}), t_i\right).$$

I denote by BNE$(\mathcal{M}, \mathcal{T})$ the set of all Bayesian Nash equilibria. Note that this set may be empty for particular choices of $(\mathcal{M}, \mathcal{T})$. The revenue of $\mathcal{M}$ under a particular type space $\mathcal{T}$ and strategy profile σ is:

$$R(\mathcal{M}, \mathcal{T}, \sigma) = \sum_{t\in T} \pi(t) \int_{m\in M} \sum_{i\in N} p_i(m)\sigma(dm; t). \tag{2.2}$$

The corresponding *extraction ratio* is:

$$E(\mathcal{M}, \mathcal{T}, \sigma) = \frac{R(\mathcal{M}, \mathcal{T}, \sigma)}{S(\mathcal{T})}, \tag{2.3}$$

with the convention that when $S = R = 0$, $E = 1$.

The seller's goal is to find mechanisms that solve:

$$\sup_{\mathcal{M}} \inf_{\mathcal{T}} \inf_{\sigma \in \text{BNE}(\mathcal{M},\mathcal{T})} E(\mathcal{M}, \mathcal{T}, \sigma). \tag{2.4}$$

The interpretation of this problem is: the seller must select a mechanism $\mathcal{M}$, following which Nature[4] will select both the type space $\mathcal{T}$ and the equilibrium $\sigma \in \text{BNE}(\mathcal{M}, \mathcal{T})$ to make E as small as possible. If $\text{BNE}(\mathcal{M}, \mathcal{T})$ is empty, then our convention is that the infimum is zero. I will refer to the value (2.4) as the max min extraction ratio.

The formulation of (2.4) is quite demanding: if the seller chooses a mechanism for which there are multiple equilibria, Nature will select the one with the lowest extraction ratio. The exposition is simplified by initially allowing the seller to choose the equilibrium, which permits us to achieve a "partial" max min extraction ratio. The results will later be strengthened to "full" max min by allowing Nature to select the equilibrium with the lowest ratio. This terminology is modeled after the partial and full implementation concepts in mechanism design, although to be clear, full max min does not require equilibrium uniqueness, but just that the extraction ratio be maximized in the worst type space and worst equilibrium. When it is clear which equilibrium is used, I will simply write $E(\mathcal{M}, \mathcal{T})$ for the extraction ratio.

2.3 Example

Let us start by considering a simple example that will illustrate some of the main ideas. There are two potential buyers $i = 1, 2$, and each buyer could be of type

[4] Throughout, I use Nature as the personification of all minimization operations beyond the control of the designer.

L or type H. Type L thinks the good is worth v, and type H thinks the good is worth $2v$. The distribution of types is given by the following table:

	L	H
L	$1-2\rho-\psi$	ρ
H	ρ	ψ

Relative to the general model, I have assumed that the type space is a symmetric and payoff type space and that the support of values is of the form $\{v, 2v\}$. These features are common knowledge among the buyers and the seller. However, the parameters ρ, ψ, and v are known to the buyers and unknown to the seller. These assumptions are made for the simplicity of the example, and will be relaxed for the main results.

The maximum surplus that can be generated by allocating the good efficiently is:

$$
\begin{aligned}
S &= v(2(\psi+2\rho)+1-2\rho-\psi) \\
&= v(1+2\rho+\psi).
\end{aligned}
$$

The seller has to select a mechanism to sell the good which is independent of v, ρ, and ψ, and therefore independent of S as well. As discussed in the introduction, the seller is highly uncertain about S, and lacks beliefs about which parameters are likely to obtain. As a result, the seller compares mechanisms by a worst-case performance metric which is scale free in S: the minimum extraction ratio. Our seller will first pick a mechanism, and then Nature will choose the parameters to minimize the extraction ratio.

For starters, let us consider what would happen if the seller were to use a second-price auction. If there is a positive reserve price $r > 0$, then Nature could always select v such that $2v < r$. S would be positive, but $R = 0$ (since the reserve is greater than the highest value). This is an important observation: introducing a positive reserve price that is totally unresponsive to the parameters of the model leads to extremely unfavorable outcomes in the worst-case. With a reserve price of zero, revenue is:

$$\begin{aligned} R &= v(2\psi + 1 - \psi) \\ &= v(1 + \psi). \end{aligned}$$

Hence, the extraction ratio would be:

$$E = \frac{1 + \psi}{1 + 2\rho + \psi}.$$

Clearly, to minimize the ratio, Nature should make ψ as small as possible and ρ as large as possible, so $\psi = 0$ and $\rho = \frac{1}{2}$. The resulting extraction ratio is $E = \frac{1}{2}$. Note well that it does not actually matter what level v Nature chooses: The model is "scale-free" in v.

	L	H
L	0	$\frac{1}{2}$
H	$\frac{1}{2}$	0

The second-price auction with no reserve is of course just one mechanism. Let us consider a simple modification. In addition to accepting bids b_i, the seller canvasses the bidders for what they believe about the distribution of others' bids. Each bidder's response to the survey will be used to set the reserve price only if that bidder does not win. Specifically, bidders submit a quantity w_i, which is the bidder's report for the value of v, and a quantity μ_i, which is the bidder's reported probability that $b_j = 2v$. If bidder i has the high bid, or if bids are equal and i wins a uniform tie break, then bidder i will be "offered" the good at a reserve price r_j that only depends on (b_j, w_j, μ_j). Moreover, this price is always at least b_j. Thus, bidder i is facing an exogenous price which depends on the other bidder's report, and he will receive the good and pay the price as long as $b_i \geq r_j$. As in the second-price auction, truthful bidding is a weakly dominant strategy: $b_i = v_i$. In the following, I impose that bidders follow this strategy.

Bidders will receive a small side reward for their survey response (w_i, μ_i). In particular, bidder i is paid according to the scoring rule:

$$\epsilon\, b_j \left(\mu_i \mathbb{I}_{b_j = 2w_i} + (1 - \mu_i) \mathbb{I}_{b_j = w_i} - \frac{(\mu_i)^2 + (1 - \mu_i)^2}{2} \right). \tag{2.5}$$

Given a report $w_i = v$, there is a strict incentive to report $\mu_i = Pr(b_j = 2v | v_i)$, and in fact the bidder receives a positive net payment from the seller. Moreover, any report other than $w_i = v$ induces a smaller payoff: if $w_i \neq 2v$ as well, then the bidder always makes a net transfer to the seller, and if $w_i = 2v$, bidder i only gets paid when $b_j = 2v$, and not when $b_j = v$. Thus, in any equilibrium in which the buyers bid their values, they must also truthfully report $w_i = v$ and their conditional belief that $b_j = 2v$. Note that the payment is scaled by $b_j = v_j > 0$, so that in expectation, the transfer to bidder i from (2.5) is no more than $\epsilon\, \mathbb{E}[v_j] \leq \epsilon S$.

The seller offers the winner the good at the revenue maximizing price, conditional on the winner being the high bidder and winning any tie breaks, and also conditional on the loser's reported beliefs. Suppose bidder i has the high bid and wins a tie break. Clearly, if the second-highest bid is $b_j = 2w_j = 2v$, then the seller should set i's price at $2v$. If $b_j = v$, then the probability of $v_i = 2v$ is:

$$\frac{2\mu_j}{3} = \frac{2}{3}\frac{\rho}{1-\rho-\psi},$$

and the conditional probability of $v_i = v$ is:

$$\frac{1-\mu_j}{3} = \frac{1}{3}\frac{1-2\rho-\psi}{1-\rho-\psi}.$$

These formulae exhibit the selection effect of the tie: the "raw" probability of both players having valuation v is $1-\mu_j$, and one having $2v$ is μ_j. But if both have a low value, bidder i only wins half the time, thus leading to the formulae above. The optimal price in this case is $2v$ if $\rho \geq \frac{1-\psi}{4}$ and v if $\rho \leq \frac{1-\psi}{4}$.

Thus, Nature has two options. If $\rho \leq \frac{1-\psi}{4}$, then revenue is $R_1 = v(1+\psi)$, and the extraction ratio is:

$$E_1 \geq \frac{R_1}{S} = \frac{1+\psi}{1+2\rho+\psi},$$

which is minimized by making ρ as large and ψ as small as possible. Hence, E_1 is minimized at $\rho = \frac{1}{4}$ and $\psi = 0$. At these values, $E_1 = \frac{2}{3}$.

If $\rho \geq \frac{1-\psi}{4}$, then the price is always $2v$, and revenue is $R_2 = 2v(2\rho+\psi)$ and the extraction ratio (not counting transfers associated with (2.5) is:

$$E_2 \geq \frac{R_2}{S} = \frac{2(2\rho+\psi)}{1+2\rho+\psi}.$$

The ratio is decreasing in ρ and ψ. Substituting in $\rho = \frac{1-\psi}{4}$, the ratio is still decreasing in ψ, so again the optimal values are $\rho = \frac{1}{4}$ and $\psi = 0$. Hence, $E_2 = \frac{2}{3}$.

Taking into account at most ϵS in lost revenue for each bidder due to (2.5), the extraction ratio for this mechanism is therefore at least $\frac{2}{3} - 2\epsilon$.

	L	H
L	$\frac{1}{2}$	$\frac{1}{4}$
H	$\frac{1}{4}$	0

Note that this type space is lopsided, in the sense introduction in Section 2.2: with probability one, at most one bidder has a valuation greater than v, which is the bottom of the support.

The bottom line is that a simple modification of the second-price auction yields a substantial improvement in the worst-case extraction ratio from $\frac{1}{2}$ to $\frac{2}{3}$. This mechanism accepts bids and also surveys bidders' beliefs about the distribution of others' bids, with the truthful revelation of this information being incentivized with a scoring rule. Each bidder's survey response is used to set the reserve price when the other bidder wins, which protects the seller from low revenue when there is a large gap between the highest and second-highest values.

It is worth noting that this mechanism, while a significant improvement over the second-price auction with no reserve, does *not* maximize the minimum possible extraction ratio. It is easy to see that in the worst-case distribution, bidders' beliefs determine their preferences in the sense of Neeman (2004), since H puts zero probability on the other bidder being of type H. With a more complicated mechanism in which the seller elicits second-order beliefs and introduces side-bets, this property could be exploited to extract all of the efficient surplus. The only type spaces which do not have this property are those in which types are drawn independently, for which the extraction ratio is minimized when $Pr(v_i = 2v) = \sqrt{2} - 1$, and the max min extraction ratio is approximately 0.7071.

This wedge is entirely due to the assumption that the support of values is restricted to being of the form $\{v, 2v\}$. In the rest of the chapter, I will pursue a similar analysis but in the more general setting of Section 2.2, without restrictions on the number of bidders, on the support of valuations, or on the kinds of information that bidders might learn about the distribution. It will turn out that the worst-case type spaces approach a continuous distribution of values, in contrast to this discrete example. A straightforward generalization of the mechanism described above achieves the max min extraction ratio for this more general problem.

2.4 Characterizing the $\max\min$ extraction ratio

2.4.1 Preamble

I now proceed to characterize the max min extraction ratio and present simple mechanisms that achieve the max min. I begin by defining a particular mechanism that I call the *belief survey auction* (BSA). This mechanism is a modified second-price auction in which the seller accepts bids and also elicits reports of first-order beliefs. A bidder's reported belief is used to set the reserve price when one of the other bidders wins the auction. I show that truthful reporting of values and beliefs is incentive compatible, and in this truthful equilibrium, the seller is guaranteed a tight lower bound on the extraction ratio. In particular, Lemma 2.1 shows that there is a small subset of type spaces, namely symmetric and lopsided payoff type spaces, within which the extraction ratio for the BSA can be minimized. The argument proceeds by taking a given type space as input, and producing a new symmetric and lopsided payoff type space with the same efficient surplus and weakly lower revenue, and hence a lower extraction ratio. The is a simple intuition for why these type spaces minimize the extraction ratio. The reports of losing bidders allow the seller to set reserve prices, and the more informative the losers' reports are, the better reserves the seller is able to set. In these worst-case type spaces, bidders' reports are minimally informative: since they are payoff type spaces, bidders beliefs contain no information beyond the value, and symmetry implies that all bidders' reports are equally informative. Finally, lopsidedness implies that all losing bidders have the same value, equal the bottom of the support, so there is just one belief that is used to set the reserve price.

The extraction ratio of the BSA on such type spaces is completely determined by the distribution of the highest value. I show that for a given level of revenue, the efficient surplus is maximized by drawing the highest value from a particular Pareto distribution. If the support of the distribution were unbounded, the highest value would have infinite expected value, with the resulting extraction ratio being zero. As a result, I consider type spaces in which the support of the highest value is bounded as a constant multiple γ of the efficient surplus. This constant parametrizes the set of type spaces, and for each value of γ I characterize the $\max\min$ extraction ratio.

In addition, the BSA turns out to be an optimal auction on symmetric and lopsided payoff type spaces. As a result, no mechanism can have a higher extraction ratio in the worst case, and the bound on the extraction ratio is tight. This establishes the partial $\max\min$ extraction ratio result of Theorem 2.1. Finally, I show that if the seller rewards bidders for their reported beliefs using a scoring rule, truthful reporting can be made the unique strategy profile that survives iterated deletion of dominated strategies. Hence, the partial result is strengthened to full max min in Theorem 2.2.

I note that there are in fact many mechanisms which approach the solution to (2.4). The mechanisms I consider are notable for their simplicity, but in Section 2.5 I will discuss some alternatives.

2.4.2 The belief survey auction

Our foundation for constructing the BSA is the second-price auction. This auction has an important property: each bidder is facing a random price at which he could purchase the good, where the distribution of the price is completely determined by the strategies of other bidders. In the second-price auction, this price is the highest bid made by other bidders, $b_{-i}^{(1)}$. The own bid b_i is the cutoff such that bidder i would like to purchase the good if the realized price is less than b_i. Since it is optimal to buy the good at any price below the bidder's value, $b_i = v_i$ is a weakly dominant strategy.

The BSA will retain this property. Each buyer submits a bid b_i, which is the cutoff at which they accept a price which is a function of other buyers' reports. The point of departure from the second-price auction is that this price is not $b_{-i}^{(1)}$, but rather incorporates more information that is elicited from the other bidders. Specifically, in addition to a bid, each bidder will submit a report of their beliefs about the joint distribution of (1) the highest bid of others $b_{-i}^{(1)}$, (2) the second-highest bid of others $b_{-i}^{(2)}$, and (3) the number k of high bidders amongst the other players, i.e., the number of players j such that $b_j = b_{-i}^{(1)}$. Naturally, if $b_{-i}^{(1)} > b_{-i}^{(2)}$, then under a truthful report, $k = 1$ with probability 1. Assuming the report is truthful, bidder j's reported beliefs allow the seller to determine an optimal price to charge the winning bidder conditional on the winner not being bidder j, and conditional on the second-highest bid of others. The report of the number of tied bidders allows the seller to control for the selection effect induced by tie breaking.

This mechanism strikes a balance between the amount of information about the environment that the seller elicits from bidders and the range of type spaces in which the mechanism maximizes revenue. In particular, by canvassing beliefs about $b_{-i}^{(2)}$ in addition to $b_{-i}^{(1)}$, the seller is able to set reserve prices that are better on average than the optimal anonymous reserve price in the second-price auction (Proposition 2.2). If the seller collected less information, namely beliefs about $b_{-i}^{(1)}$ and the number of high bidders, he could still achieve max min extraction ratio (Theorems 2.1 and 2.2) but would no longer be guaranteed to do as well as the second-price auction.

More formally, I define a mechanism $\mathcal{M}^{BSA}$ as follows. Each message m_i consists of a bid b_i and a distribution μ_i in $\Delta\left(\mathbb{R}_+^2 \times \mathbb{N}\right)$, where $\mathbb{R}_+ = [0, \infty)$ and $\mathbb{N}$ is the set of positive integers. For a vector x, let:

$$W(x) = \left\{ i \,\middle|\, x_i = x^{(1)} \right\} \tag{2.6}$$

denote the set of maximal indices in x. The interpretation is that μ_i is bidder i's reported beliefs about the distribution of $b_{-i}^{(1,2)}$ and $|W(b_{-i})|$. Thus, $M_i = \mathbb{R}_+ \times \Delta(\mathbb{R}_+^2 \times \mathbb{N})$. A typical message will be written $m_i = (b_i, \mu_i)$. The allocation rule is specified as follows. Suppose bidder i submits the highest bid, $b_i = b^{(1)}$. If there are ties, the mechanism selects i uniformly from the set of high bidders $W(b)$. We will then pick a bidder $j \neq i$ uniformly to calibrate the price $r_j\left(m_j, b_{-j}^{(2)}\right)$ for bidder i, which will be greater than $b^{(2)} = b_{-i}^{(1)}$. If $b_i \geq r_j\left(m_j, b_{-j}^{(2)}\right)$, bidder i wins the good and pays $r_j\left(m_j, b_{-j}^{(2)}\right)$. Otherwise, the good remains unallocated.

Hence:

$$q_i^{BSA}(m) = \frac{1}{|W(b)|}\frac{1}{n-1}\sum_{j\neq i}\mathbb{I}_{b_i \geq r_j\left(m_j, b_{-j}^{(2)}\right)},$$
$$p_i^{BSA}(m) = \frac{1}{|W(b)|}\frac{1}{n-1}\sum_{j\neq i}\mathbb{I}_{b_i \geq r_j\left(m_j, b_{-j}^{(2)}\right)} r_j\left(m_j, b_{-j}^{(2)}\right),$$

where $\mathbb{I}_C$ is the indicator function, equal to one if condition C is met and zero otherwise. The price $r_j\left(m_j, b_{-j}^{(2)}\right)$ in fact does not depend on m_i when i is allocated the good, since $b_i \geq b^{(2)}$. Thus, bidding one's value is a weakly dominant strategy and for now I impose that this occurs in equilibrium.

Also, note that bidder i's report of μ_i has no effect on any price $r_j\left(m_j, b_{-j}^{(2)}\right)$ when bidder $j \neq i$ is consulted, nor does it affect whether or not i is offered the good at any price. Hence, any report of μ_i is incentive compatible. I consider the "truth-telling" equilibrium in which bidders report:

$$\mu_i\left(v_{-i}^{(1,2)}, k\right) = \sum_{\left\{t_{-i}\,\middle|\, \substack{\phi_{-i}^{(1,2)}(t_{-i}) = v_{-i}^{(1,2)}, \\ |W(\phi_{-i}(t_{-i}))| = k}\right\}} \pi(t_{-i}|t_i).$$

In plain language, bidders report the conditional joint distribution of the first two order statistics of others' bids, and the number of high bidders among the other players. We will subsequently see that this strategy can be made the unique equilibrium, for any $\mathcal{T}$, at an arbitrarily small cost to the extraction ratio.

I still have to specify the prices that bidders are offered. What I would like to implement is a "monopoly" price with respect to the conditional distribution of the winner's value when bidder j is consulted. Each bidder reports their beliefs μ_j conditional on t_j, but the seller only consults bidder j in particular situations, namely when j is not a high bidder or when j is a high bidder but loses a tie break. Hence, μ_j is *not* the distribution of $\left(v_{-j}^{(1,2)}, |W(v_{-j})|\right)$ conditional on j being consulted. Rather, the mechanism takes into account the fact that $v_{-j}^{(1)} \geq b_j$ and that j must have lost any and all tie breaks. Finally, since the price can depend on any information from losing bidders, the mechanism additionally uses the fact that $v_{-j}^{(1)} \geq b_{-j}^{(2)}$. As previously discussed, by conditioning on the second-highest bid of $-j$, the seller makes sure that the auction generates weakly greater revenue than a second-price auction with the optimal reserve price. These computations result in an upper cumulative conditional distribution of the winner's value when j is consulted, which is:

$$G_j\left(r; m_j, b_{-j}^{(2)}\right) = \sum_{\left\{v_{-j}^{(1,2)}, k \,\middle|\, \substack{v_{-j}^{(1)} \geq \max\{r, b^{(2)}\}, \\ v_{-j}^{(2)} = b_{-j}^{(2)}}\right\}} \frac{k}{\mathbb{I}_{b_j = v_{-j}^{(1)}} + k} \mu_j\left(v_{-j}^{(1,2)}, k\right). \tag{2.7}$$

This is the probability that the winner's value is at least r, conditional on bidder j being consulted and on $b_{-j}^{(2)}$. The price induced by bidder j's report is a monopoly price with respect to this distribution:

$$r_j\left(m_j, b_{-j}^{(2)}\right) \in \arg\max_r r\; G_j\left(r; m_j, b_{-j}^{(2)}\right). \tag{2.8}$$

Note that $r_j\left(m_j, b_{-j}^{(2)}\right)$ is always at least $b^{(2)}$, since $G_j\left(r; m_j, b_{-j}^{(2)}\right)$ is constant for $r \leq b^{(2)} = \max\left\{b_j, b_{-j}^{(2)}\right\}$.

2.4.3 Worst-case extraction ratio for the BSA

In this section, I characterize the minimum extraction ratio and the minimizing type spaces for the BSA under the truth-telling equilibrium. I will give an informal argument, with a rigorous proof appearing at the end of the chapter.

The reports of the losing bidders contain information that the seller uses to optimally set the winner's price. In particular, the seller conditions on:

(1) bidder j not being offered the good, i.e., $v^{(1)} \geq v_j$ and j loses any tie breaks;

(2) the winner's value being greater than $b_j = v_j$ and $b_{-j}^{(2)} = v_{-j}^{(2)}$;

(3) the consulted bidder j's realized type being t_j.

These three pieces of information are incorporated into G_j and the optimized reserve price r_j.

Note that the distribution of $v_{-j}^{(1)}$ conditioning on (1)-(3) will on average be the distribution of $v_{-j}^{(1)}$, given that j is not being offered the good, which is just (1). The fact that j is not offered the good means that $v_{-j}^{(1)} \geq v_j$ and j loses any tie breaks. This average distribution does not condition on the fact that v_j and $v_{-j}^{(2)}$ have particular realized values b_j and $b_{-j}^{(2)}$, respectively, and it does not incorporate the additional information contained in t_j.

The conditioning of the reserve price on (1)-(3) facilitates a kind of monopoly price discrimination when selling to the high bidder. We could view the distribution of the highest value as being an aggregate demand curve. Instead of having to a single price for the entire market, the seller sees demand broken up into pieces conditional on (1)-(3). Such price discrimination is always beneficial to the seller, since the seller could ignore the extra information and set uniform prices. Hence, it is weakly worse for the seller to have less information, which is when b_j, $b_{-j}^{(2)}$, and t_j are less informative about $v^{(1)}$.

Given a particular type space $\mathcal{T}$ with $S(\mathcal{T}) = S$, the seller generally learns more from (1)-(3) than from just (1). However, it is possible to find another type space $\mathcal{T}'$ in which the distribution of $v^{(1)}$ conditional on (1) is the same, but in which the seller learns nothing from (2) and (3). This alternative type space $\mathcal{T}'$

has the same efficient surplus, but revenue must be weakly lower, since the seller does not benefit by setting discriminatory reserve prices based on (2) and (3). Hence, the extraction ratio is lower as well.

How is this type space obtained? First, if bidders have more than one type t_j associated with a particular realization v_j, then $\mathcal{T}'$ can be defined so that these types are effectively merged. In other words, bidder j's types $t_j \in \phi_j^{-1}(v_j)$ are replaced with a single type, so that bidder j only learns that he has one of the types such that $\phi_j(t_j) = v_j$. As a result, $\mathcal{T}'$ is a payoff type space with the same marginal distribution over values as $\mathcal{T}$. This transformation makes bidder j's report μ_j weakly less informative. Second, the realized values b_j and $b_{-j}^{(2)}$ are also informative. We can modify the type space so that every bidder except the winner has the minimum valuation in the support $\underline{v}$, meaning the type space is lopsided. Thus, the losing bidders' values are completely uninformative as lower bounds on the winner's valuation. Finally, there may be asymmetries wherein one bidder's losing report is on average more informative than others'. We can "symmetrize" the distribution so that all bidders' losing reports are equally informative. This is formalized in the following:

Lemma 2.1 (Worst-case type spaces)**.** *For any $\mathcal{T}$, there exists a symmetric and lopsided payoff type space $\mathcal{T}'$ such that:*

$$E\left(\mathcal{M}^{BSA}, \mathcal{T}'\right) \leq E\left(\mathcal{M}^{BSA}, \mathcal{T}\right).$$

Any type space within the class described in Lemma 2.1 is of the following form: pick one bidder $i \in N$ uniformly, and set $v_j = \underline{v}$ for $j \neq i$. Bidder i's value is drawn from the distribution $F^{(1)} \in \Delta(\mathbb{R}_+)$, which is the unconditional distribution of $v^{(1)}$. Losing bidders j always report the conditional belief that

$v_{-j}^{(1)} \sim F^{(1)}$, and $b_j = b_{-j}^{(2)} = \underline{v}$. The reserve price r^* is a solution to:

$$\max_{r \geq 0} r\ G^{(1)}(r),$$

where $G^{(1)}(r) = 1 - \lim_{v \uparrow r} F^{(1)}(v)$ is the probability of the offered price of r being "accepted" by the high bidder, when the bidder buys whenever indifferent. Revenue $R(\mathcal{T})$ is simply this maximum.

The spirit of Lemma 2.1 is that holding fixed the expected efficient surplus, there is a certain class of type spaces within which revenue can be minimized. It is now instructive to reverse the question: suppose we wanted to maintain $R(\mathcal{T}) \leq R$. Which distributions $F^{(1)}$ will maximize $S(\mathcal{T})$ subject to this revenue constraint? It must be that for every $r \geq 0$, $r\, G^{(1)}(r) \leq R$, so $F^{(1)} \geq 1 - \frac{R}{r}$. On the other hand, pushing down the cumulative distribution of $v^{(1)}$ always increases $\mathbb{E}[v^{(1)}] = S(\mathcal{T})$. Thus, the supremum of $E(\mathcal{M}^{BSA}, \mathcal{T})$ will be attained when $F^{(1)}$ is precisely:

$$F^{(1)}(v) = \begin{cases} 0 & \text{if } v < R \\ 1 - \frac{R}{v} & \text{if } R \leq v < \overline{v} \\ 1 & \text{if } v \geq \overline{v} \end{cases}, \tag{2.9}$$

where $\overline{v}$ is the largest valuation in the support, which is a truncated Pareto distribution with scale R and shape of 1. An example of such a distribution is given in Figure 2.1.

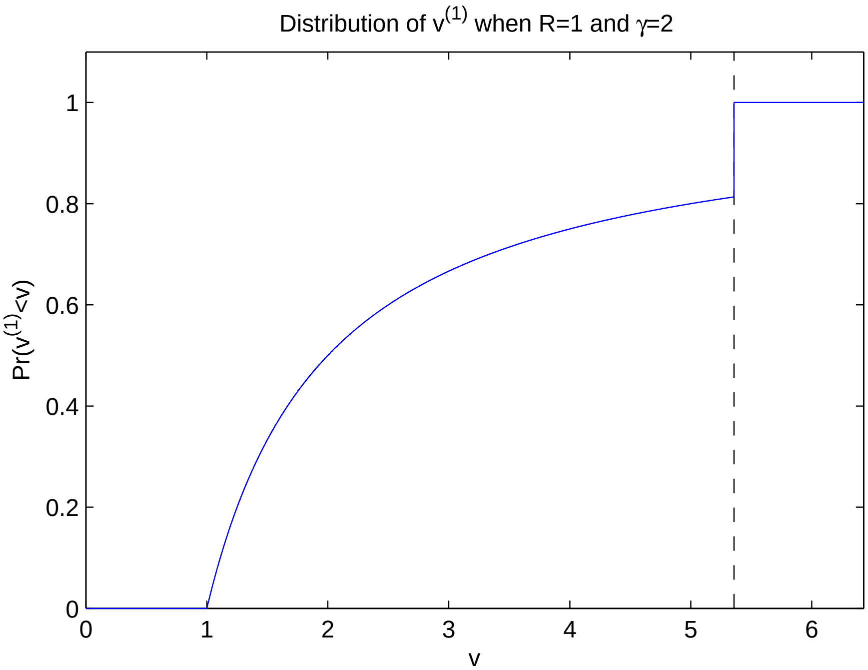

Figure 2.1: An example of the lower bound on $F^{(1)}$ for $R = 1$ and $\overline{v} = 2$. The efficient surplus in this case is approximately 2.68, and the extraction ratio is approximately 0.37.

If the distribution of $v^{(1)}$ is given by (2.9), then the efficient surplus is given by the Riemann-Stieltjes integral:

$$\begin{aligned} S &= \int_{v=R}^{\overline{v}} v \frac{R}{v^2} dv + \overline{v}\frac{R}{\overline{v}} \\ &= R\left(1 + \log(\overline{v}) - \log(R)\right). \end{aligned} \tag{2.10}$$

Note that (2.9) implies a mass point on $\overline{v}$ of size $\frac{R}{\overline{v}}$. It is evident that S is increasing without bound in $\overline{v}$. If the distribution of values can be unbounded, then a fixed level of revenue is consistent with arbitrarily large efficient surplus. However, this requires putting a lot of mass on extremely large valuations, far from the efficient surplus. It is natural to ask how the extraction ratio behaves when there are limits to how dispersed values can be. A simple way to limit dispersion is to require that values not be too much larger than S. This is formalized in Assumption 2.1:

Assumption 2.1 (Bounded support)**.** The support of values is contained in $[0, \gamma S(\mathcal{T})]^n$ for some $\gamma \geq 1$.

Let:

$$\mathbf{T}(\gamma) = \{\mathcal{T} | \operatorname{supp}(\phi_* \pi) \subset [0, \gamma S(\mathcal{T})]^n\},$$

where $\phi_* \pi$ is the pushforward measure on values, i.e., the distribution on values induced by the distribution π and the mapping ϕ. Under this assumption, γS and E can be substituted for $\overline{v}$ and $\frac{R}{S}$, so that (2.10) becomes:

$$E(1 + \log(\gamma) - \log(E)) = 1. \tag{2.11}$$

This equation represents an accounting identity. When the highest value is drawn from (2.9), the expected efficient surplus S must be equal to the Riemann-Stieltjes integral with respect to (2.9), where γS is the upper limit of the integral. As shown in the proof of the following proposition, this equation has a unique solution, denoted $E^*(\gamma)$. We have the following:

Proposition 2.1 (min extraction ratio for BSA)**.** *For any $\gamma > 0$, the worst-case extraction ratio for the BSA under the truth-telling equilibrium when type spaces are restricted to $\mathbf{T}(\gamma)$ is the unique $E^*(\gamma)$ which solves* (2.11)*. This extraction ratio is attained by type spaces of the form described in Lemma 2.1, with the distribution of the highest value approaching* (2.9)*.*

This concludes the characterization of the worst-case extraction ratio for the BSA.

2.4.4 $\max\min$ extraction ratio

In fact, for the worst-case type spaces in which at most one bidder has a positive value, $\mathcal{M}^{BSA}$ is an optimal auction. On these type spaces, the seller's problem is formally equivalent to the selling of a single unit to a single buyer. It is well known that the optimal mechanism is a posted price, and the reports of the losers allow the seller to set the optimal price (cf. Riley and Zeckhauser, 1983).

This implies that the lower bound of $E^*(\gamma)$ is tight. In general:

$$\sup_{\mathcal{M}} \inf_{\mathcal{T}\in\mathbf{T}(\gamma)} E(\mathcal{M},\mathcal{T}) \leq \inf_{\mathcal{T}\in\mathbf{T}(\gamma)} \sup_{\mathcal{M}} E(\mathcal{M},\mathcal{T}),$$

since any mechanism that the seller chooses when forced to move first could also be chosen when moving second, and therefore guarantees at least as large of a payoff. For many problems, it turns out that the inequality is in fact an equality, as in the minimax theorems of zero sum games. This is not automatically the case here since the setup does not satisfy the regularity conditions of the minimax theorems known to the author.[5]

However, a solution to the LHS is given by the seller using $\mathcal{M}^{BSA}$ and Nature choosing a type space satisfying the conditions of Lemma 2.1 with $\underline{v} = 0$. Moreover, for the RHS, Nature could always use these same type spaces, and the seller can do no better than with $\mathcal{M}^{BSA}$. Hence, the two sides are in fact equal. This observation, combined with Proposition 2.1 gives us the following:

[5]Specifically, von Neumann's minimax theorem only applies to finite domains, and Sion's minimax theorem requires the domains to be linear topological spaces. This structure is lacking on type spaces and mechanisms.

Theorem 2.1 (Partial max min extraction ratio)**.** *The solution to* (2.4) *restricted to* $\mathcal{T} \in \mathbf{T}(\gamma)$ *is no greater than* $E^*(\gamma)$. *Hence, the BSA under the truth-telling equilibrium partially solves* (2.4)*.*

2.4.5 Characterizing performance

The number $E^*(\gamma)$ gives the max min extraction ratio that the seller is guaranteed by using the BSA. But is this lower bound economically meaningful? It would be useful to know that the lower bound guarantees the seller a substantial revenue-share of the efficient surplus. On the right panel of Figure 2.2, $E^*(\gamma)$ is plotted for values of γ ranging from 1 to 50. The left hand panel gives E^* for six values of γ. We see that $E^*(\gamma)$ is monotonically decreasing, quickly for small γ, with the rate of decrease falling rapidly. For example, going from $\gamma = 1.1$ to $\gamma = 2$ entails a decreasing from $E^* = 0.67$ to $E^* = 0.37$, whereas the difference between $\gamma = 100$ and $\gamma = 1,000$ is only 3 percentage points. In the latter case, the seller is guaranteed at least a 10% revenue-share of the efficient surplus. Even for $\gamma = 10,000,000$, the seller is guaranteed approximately a 5% revenue-share.

The slow rate of decay of $E^*(\gamma)$ can be formalized as follows. Asymptotically:

$$E^*(\gamma) = O\left(\frac{1}{\log(\gamma)}\right).$$

This follows from (2.11), as:

$$\begin{aligned} 1 &= \lim_{\gamma\to\infty} \frac{\frac{1}{E^*(\gamma)} + \log(E^*(\gamma))}{1+\log(\gamma)} \\ &= \lim_{\gamma\to\infty} \frac{\frac{1}{E^*(\gamma)} + \log(E^*(\gamma))}{\log(\gamma)}, \end{aligned}$$

γ	$E^*(\gamma)$
1.1	0.6656
2	0.3734
10	0.2045
100	0.1309
1,000	0.0977
10^7	0.0497

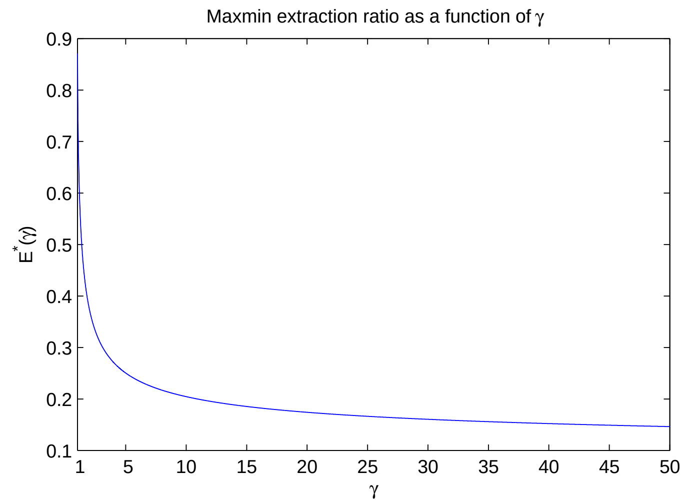

Figure 2.2: $E^*(\gamma)$ for a range of values.

since the expression inside the first limit is equal to 1 for all γ. It is easy to see that $\lim_{\gamma\to\infty} E^*(\gamma) = 0$, since if it were bounded away from 0, the left hand side of (2.11) would blow up. Hence, by L'Hôpital's rule:

$$\lim_{\gamma\to\infty} \frac{\frac{1}{E^*(\gamma)}}{\frac{1}{E^*(\gamma)} + \log(E^*(\gamma))} = \lim_{\gamma\to\infty} \frac{1}{1 + E^*(\gamma)} = 1,$$

where the derivative exists because of the implicit function theorem. Since both limits exist, the limit of the product is equal to the product of the limits, and:

$$\begin{aligned} 1 &= \lim_{\gamma\to\infty} \frac{\frac{1}{E^*(\gamma)} + \log(E^*(\gamma))}{\log(\gamma)} \frac{\frac{1}{E^*(\gamma)}}{\frac{1}{E^*(\gamma)} + \log(E^*(\gamma))} \\ &= \lim_{\gamma\to\infty} \frac{\frac{1}{\log(\gamma)}}{E^*(\gamma)}, \end{aligned}$$

which proves the result. In sum, $E^*(\gamma)$ goes to zero exponentially slower than the rate of growth of γ, so even for very generous bounds on the dispersion in values, the seller will still be guaranteed a substantial share of the efficient surplus.

Another potential concern is that the type spaces uses to achieve the lower bound are highly stylized. Symmetry of the type space is not necessary, but what is necessary is the consequence of the lopsided property, that there is a large gap between the average highest and second-highest values. In some situations, this property could be quite natural. In the school construction example from the introduction, it is possible that there is one dominant contractor that tends to have the lowest cost, and this asymmetry is common knowledge among the firms. Nonetheless, in many situations one would not expect to find such a large gap. Is the extraction ratio in the worst-case radically different from extraction ratios in the kinds of environments that are more frequently modeled?

Figure 2.3 gives examples of four different distributions over values in which the gap between highest and second-highest values is modest. Each type space has a different efficient surplus and consequently different $\gamma = \frac{\overline{v}}{S}$. For example, in the first panel values are independent and uniformly distributed on $[0, 1]$. With two bidders, the efficient surplus is 0.67, so $\gamma = 1.5$. For each of these type spaces, the extraction ratio the seller obtains with a second-price auction and the optimally chosen anonymous reserve price is compared to the lower bound guaranteed by the BSA for the same γ. In the uniform example, the seller could set the optimal reserve of 0.5 and obtain the optimal extraction ratio of 0.625. In contrast, $E^*(1.5) = 0.4569$. The point of these examples is that although the worst-case type spaces are stylized, the lower bound is not orders of magnitude different from the extraction ratio on "typical" examples with similar γ.

In fact, there are classes of environments, namely symmetric payoff type spaces with regular and independent distributions, in which the BSA will implement the optimal auction. The reason is simply that each bidder will report the independent distribution from which other bidders' values are drawn, and the seller will

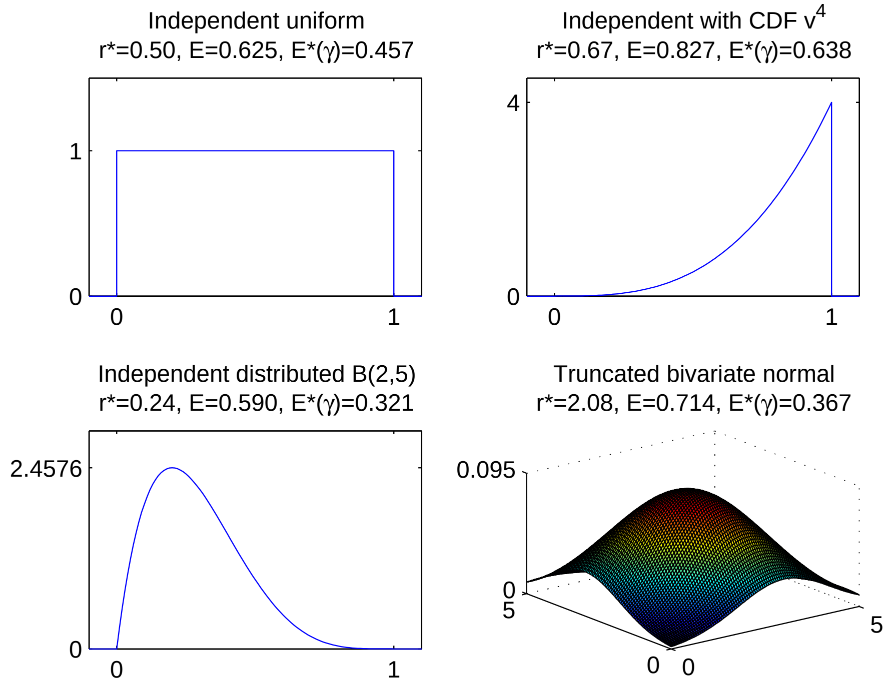

Figure 2.3: Examples of the extraction ratio on particular type spaces in $\mathbf{T}(\gamma)$ versus $E^{*}(\gamma)$. For the first three frames, values are i.i.d. and drawn from the depicted PDF. In the fourth frame, values are correlated and drawn from the depicted joint PDF. In each case, E^{*} is on the same order of magnitude as the actual extraction ratio that would obtain with the second-price auction with an optimal anonymous reserve.

set the winner's price equal to the maximum of the second-highest bid and the optimal reserve price, which is where the virtual valuation is zero. More generally, in any type space the BSA has to generate at least as much revenue as a second-price auction with an optimal uniform reserve price. The reason is that the BSA essentially breaks up the seller's reserve pricing problem into a bunch of conditional pricing problems. For each of these problems, it is always feasible for the seller to set the price equal to $\max\left\{r^{*}, b^{(2)}\right\}$, where r^{*} is an anonymous reserve price. Hence, the optimal pricing rule for each of these problems must generate weakly more revenue than the fixed reserve rule. We have proven the following:

Proposition 2.2 (Comparison with second-price auction)**.** *Expected revenue in the BSA when the seller does not know the prior is weakly greater than expected revenue of the second-price auction when the seller knows the prior and sets the optimal anonymous reserve price. If the distribution of values is independent, symmetric, and regular, then the BSA is an optimal auction.*

The bottom line is that the BSA guarantees the seller a relevant lower bound on the worst-case extraction ratio, and also does not greatly disadvantage the seller away from the worst-case. Revenue is always weakly better than in the second-price auction with anonymous reserve, which is probably the most widely used auction format in the world and is known to be an optimal auction in benchmark environments.

2.4.6 Equilibrium uniqueness

In this section, I extend the partial max min result of Theorem 2.1 to full max min. This is facilitated by simple perturbations of $\mathcal{M}^{BSA}$ that make the truth-telling equilibrium unique. Specifically, I will construct a mechanism $\mathcal{M}^{\epsilon}$ for every $\epsilon > 0$ with the message space M^{BSA}. This mechanism implements the same allocation and transfers as $\mathcal{M}^{BSA}$ with probability $1 - \epsilon$, but is perturbed in such a way that truth-telling is the unique strategy profile that survives iterated deletion of strictly dominated strategies. In particular, the message space is $M^{\epsilon} = M^{BSA}$, and the allocation and payoff rules are:

$$q^\epsilon(m) = (1-\epsilon)q^{BSA}(m) + \epsilon q^1(m),$$
$$p^\epsilon(m) = (1-\epsilon)p^{BSA}(m) + \epsilon p^1(m) + \epsilon p^2(m),$$

where q^1, p^1, and p^2 will be defined presently.

Recall that truthful bidding is a weakly dominant strategy of $\mathcal{M}^{BSA}$. The functions q^1 and p^1 make it strictly dominant, by adding a small probability event that each bidder is selected to be offered the good at a price drawn from a distribution $G(r)$ with positive density $g(r)$ and support equal to $\mathbb{R}_+$. Specifically, define:

$$q_i^1(m) = \frac{1}{n}\int_{r=0}^{b_i} g(r)dr,$$
$$p_i^1(m) = \frac{1}{n}\int_{r=0}^{b_i} r\, g(r)dr.$$

Since p^2 will not depend on b_i at all, $b_i = v_i$ is uniquely optimal. This trick is similar to one used in Bergemann and Morris (2012a).

The second new component of the transfer p^2 is a modified scoring rule that rewards bidders for correctly guessing the distribution of $\left(b_{-i}^{(1,2)}, |W(b_{-i})|\right)$. I say a modified scoring rule, as the transfer is weighted it so that the seller never has to pay too much in expectation to incentivize bidders to report their beliefs. In particular:

$$p_i^2(m) = \frac{1}{n}\left[\sum_{v_{-i}\in\mathbb{R}_+^{n-1}} v_{-i}^{(1)} \frac{\left(\mu\left(v_{-i}^{(1,2)}, |W(v_{-i})|\right)\right)^2}{2} - b_{-i}^{(1)}\mu_i\left(b_{-i}^{(1,2)}, |W(b_{-i})|\right)\right].$$

Since bidders report $b_{-i} = v_{-i}$ in equilibrium, by reporting μ_i, bidder i's expected payoff is:

$$\mathbb{E}[p_i^2(m)|t_i] = \frac{1}{n} \sum_{v_{-i} \in \mathbb{R}_+^{n-1}} v_{-i}^{(1)} \left[\frac{\left(\mu\left(v_{-i}^{(1,2)}, |W(v_{-i})|\right)\right)^2}{2} \right.$$
$$\left. - \pi\left(v_{-i}^{(1,2)}, |W(v_{-i})| \,\middle|\, t_i\right) \mu_i\left(v_{-i}^{(1,2)}, |W(v_{-i})|\right) \right],$$

so the first-order condition implies (as long as $v_{-i}^{(1)} > 0$) that the type t_i reports:

$$\mu_i\left(v_{-i}^{(1,2)}, k\right) = \pi\left(v_{-i}^{(1,2)}, \,\middle|\, t_i\right)$$

for all $v_{-i}^{(1,2)}$ and k. There is a unique v_{-i} such that $v_{-i}^{(1)} = 0$, and since bidders must report a distribution, they report the probability of this event accurately as well. Thus, $\mathcal{M}^\epsilon$ has a truth-telling as the unique equilibrium for all $\mathcal{T}$.

Finally, observe that in equilibrium, it must be that:

$$\begin{aligned}
\mathbb{E}[p_i^2(m)] &= -\frac{1}{n} \sum_{t_i} \pi_i(t_i) \sum_{v_{-i} \in \mathbb{R}_+^{n-1}} v_{-i}^{(1)} \frac{\left(\pi\left(v_{-i}^{(1,2)}, |W(v_{-i})| \,\middle|\, t_i\right)\right)^2}{2} \\
&\geq -\frac{1}{n} \sum_t \phi_{-i}^{(1)}(t)\pi(t) \\
&\geq -\frac{1}{n} \sum_t \phi^{(1)}(t)\pi(t) \\
&= -\frac{1}{n} S(\mathcal{T}).
\end{aligned}$$

On average, the seller makes such a transfer for each of the n bidders, at a total cost of at most $S(\mathcal{T})$. Since $\mathbb{E}[p^1(m)] \geq 0$, it will be true that:

$$E(\mathcal{M}^\epsilon, \mathcal{T}) \geq E(\mathcal{M}^{BSA}, \mathcal{T}) - 2\epsilon.$$

The results of Proposition 2.1 and Theorem 2.1, together with the fact that $\mathcal{M}^\epsilon$ has a unique equilibrium that always has an extraction ratio within 2ϵ of the truth-telling equilibrium of $\mathcal{M}^{BSA}$ imply the following theorem:

Theorem 2.2 (Full max min extraction ratio)**.** *$E^*(\gamma)$ is the solution to* (2.4)*, and the mechanisms $\mathcal{M}^\epsilon$ guarantee the seller an extraction ratio that is at least $E^*(\gamma) - 2\epsilon$.*

2.4.7 Conditional preferences

The max min extraction ratio criterion implicitly assumes that the seller compares expected revenues for different type spaces not one-for-one, but normalized by the respective expected efficient surpluses. I regard this as reasonable if the seller uses the expected efficient surplus as the target for revenue. However, the manner in which the seller compares revenue on type spaces with different surpluses is not essential to my results.

In this section, I consider a much weaker preference in which the seller does not compare outcomes across different levels of the efficient surplus. Define:

$$\mathbf{T}(S, \gamma) = \left\{ \mathcal{T} \,\middle|\, {}_{\mathrm{supp}(\phi_* \pi) \subseteq [0, \gamma S]^n}^{S(\mathcal{T}) = S} \right\}.$$

Consider the incomplete preference over mechanisms, where mechanism $\mathcal{M}$ is weakly preferred to mechanism $\mathcal{M}'$ (denoted $\mathcal{M} \succeq \mathcal{M}'$) if for every (S, γ):

$$\inf_{\mathcal{T} \in \mathbf{T}(S,\gamma)} \inf_{\sigma \in \mathrm{BNE}(\mathcal{M},\mathcal{T})} R(\mathcal{M}, \mathcal{T}, \sigma) \geq \inf_{\mathcal{T} \in \mathbf{T}(S,\gamma)} \inf_{\sigma \in \mathrm{BNE}(\mathcal{M}',\mathcal{T})} R(\mathcal{M}', \mathcal{T}, \sigma). \tag{2.12}$$

This preference is a partial ordering, since two mechanisms $\mathcal{M}$ and $\mathcal{M}'$ are incomparable if (2.12) holds for one (S, γ), but not for (S', γ'). However, $M \succeq M'$ indicates a strong notion of dominance in that for every efficient surplus, $\mathcal{M}$ performs better in terms of worst-case revenue. A mechanism is maximal in the ordering $\succeq$ if it solves:

$$\sup_{\mathcal{M}} \inf_{\mathcal{T} \in \mathbf{T}(S,\gamma)} \inf_{\sigma \in \mathrm{BNE}(\mathcal{M},\mathcal{T})} R(\mathcal{M}, \mathcal{T}, \sigma) \tag{2.13}$$

for every (S, γ). Denote the solution to this problem by $R^*(S, \gamma)$. My next result is that the mechanisms $\mathcal{M}^\epsilon$ are virtually maximal according to $\succeq$:

Proposition 2.3. *The solution to* (2.13) *is* $R^*(S, \gamma) = E^*(\gamma)S$. *For every* (S, γ)*:*

$$\inf_{\mathcal{T} \in \mathbf{T}(S,\gamma)} \inf_{\sigma \in \mathrm{BNE}(\mathcal{M}^\epsilon,\mathcal{T})} R(\mathcal{M}^\epsilon, \mathcal{T}, \sigma) \geq (E^*(\gamma) - 2\epsilon)S.$$

Proof. For the first part, clearly it cannot be that $R^*(S, \gamma) < E^*(\gamma)S$, since this implies that:

$$\sup_{\mathcal{M}} \inf_{\mathcal{T}} \inf_{\sigma \in \mathrm{BNE}(\mathcal{M},\mathcal{T})} \frac{R(\mathcal{M}, \mathcal{T}, \sigma)}{S} \leq \sup_{\mathcal{M}} \inf_{\mathcal{T} \in \mathbf{T}(S,\gamma)} \inf_{\sigma \in \mathrm{BNE}(\mathcal{M},\mathcal{T})} \frac{R(\mathcal{M}, \mathcal{T}, \sigma)}{S} < E^*(\gamma).$$

Moreover, if $R^*(S, \gamma) > E^*(\gamma)S$ for some (S, γ), then it would have to be strictly larger for every (S, γ), since every type space $\mathcal{T} = (T, \phi, \pi) \in \mathbf{T}(\gamma)$ with strictly positive $S(\mathcal{T})$ can be mapped to some $\mathcal{T}' = (T', \phi', \pi') \in \mathbf{T}(S, \gamma)$ by defining

$T' = T$, $\pi' = \pi$, and $\phi'(t) = \phi(t)\frac{S}{S(T)}$. Hence, if $R^*(S,\gamma) > (E^*(\gamma) + \epsilon)S$ for some S, then $\frac{R^*(S,\gamma)}{S} > E^*(\gamma)$ for all S, a contradiction.

The second part follows almost directly, since the extraction ratio is invariant to scaling of valuations as in the previous paragraph. As $\mathcal{M}^\epsilon$ achieves $E^*(\gamma) - 2\epsilon$ for some value of S, it must achieve the same extraction ratio for all S, and therefore revenue is at least $(E^*(\gamma) - 2\epsilon)S$. □

In sum, it does not matter how the seller compares revenue across environments in which the expected social value of the good is different. As long as the seller has max min preferences over revenue for a fixed expected efficient surplus, the BSA is an optimal auction.

2.5 Discussion

2.5.1 Belief extraction

This chapter has been focused on the selection of a mechanism by a seller who evaluates mechanisms by their worst-case extraction ratio. Under such preferences, the seller is willing to tolerate suboptimal extraction ratios on particular type spaces, as long as this ratio is greater than the worst-case. The BSA can result in such suboptimal extraction ratios, since it collects rather limited information about the environment and therefore will not maximize revenue on most type spaces. In particular, the seller only asks each bidder to estimate the distribution of the top two order statistics of other bidders' values and the number of ties. In principle, the seller could have collected information more ambitiously. Is there a limit to how much the seller could learn by asking the bidders more complicated questions? Could the seller, for example, collect enough information and in such a way that a revenue maximizing auction is always implemented?

This question is related to a folk argument that has long existed in the mechanism design literature: if a common prior is known to the agents, but not to the designer, then the prior could be extracted by the designer for free (Bergemann and Morris, 2012a). By "free", I mean that having to incentivize truthful revelation of the prior does not restrict the class of social choice functions that can be implemented, according to any solution concept. There is an obvious partial implementation solution to this problem: ask all of the agents to simultaneously announce the prior, and if they disagree, punish all of the agents severely. Of course, this is not entirely satisfactory because this mechanism would also enforce coordinated misreporting of the prior, which is counter to the full implementation philosophy of the present work.

However, Chapter 3 provides a stronger resolution of the folk argument. In general, it is possible for the seller to extract agents' beliefs in such a way that the common prior is revealed to the seller in *every* equilibrium. A caveat is that the general mechanism accomplishing this goal is quite complicated, and requires the seller to elicit each agent's infinite hierarchy of beliefs. Transfinite iterated deletion of strictly dominated strategies forces the agents to report their hierarchy truthfully in any equilibrium. Nonetheless, if such complex mechanisms are permitted, then it is possible for the seller to implement a revenue-maximizing mechanism for every realized type space, as long as the common prior is known to the agents.

The folk argument is also related to the works of Neeman (2003) and Azar, Chen, and Micali (2012). If the seller can extract the prior for free, then it is possible to implement any mechanism, including the second-price auction with an optimally chosen anonymous reserve price or the optimal dominant strategy and ex-post individually rational mechanism. The distinct contribution of this chapter is to show that for a particular criterion, i.e., the $\max\min$ extraction ratio, the seller need not use such an elaborate mechanism. It is sufficient for the seller to extract beliefs about simple statistics, and use these statistics to guard against the downside risk associated with a large gap between the highest and second-highest values, as in lopsided type spaces.

2.5.2 Simpler mechanisms

The previous section discussed more complicated mechanisms that the seller could use to achieve optimal performance in a wider range of environments. But what about the other direction: are there classes of environments in which the BSA can be further simplified, without greatly compromising performance?

There are at least two dimensions along which the BSA can be easily simplified. First, because of the selection effect induced by ties, the BSA needs to extract bidders beliefs about the number of bidders who will make high bids. Ties would not occur with positive probability if valuations were drawn from a non-atomic distribution, or if the finite supports of values were non-overlapping. Furthermore, ties do not occur with positive probability in the worst-case type spaces for extraction ratios. If attention is restricted to type spaces in which ties occur with zero probability, then clearly the seller can get away with just extracting bidders beliefs about $b_{-i}^{(1,2)}$. Second, extracting beliefs about $b_{-i}^{(2)}$ is necessary to make sure that the BSA does as well as the second-price auction. If the seller is purely concerned with $\max\min$ extraction ratio, then the seller could extract beliefs about the just highest bid of others, and set an optimal reserve price conditional on

$b^{(1)} = b^{(1)}_{-j} \geq b_j$. The same analogy with third-degree price discrimination applies, and the seller must obtain at least as much revenue as if the seller sold the good to the high bidder at the optimal monopoly price with respect to $F^{(1)}$.

Perhaps the most natural method of aligning the incentives of the seller and buyers would be to give buyers a direct stake in the revenue generated by their reports. The seller could for example share a small portion of revenue with bidder j whenever a sale is made with a price based on bidder j's report. However, this creates complicated incentives to influence the allocation of the good: for example, in the BSA, the marginal event affected by bidding $b_i = v_i$ is when the bidder is allocated the good at a price equal to v_i. In this case, the marginal surplus from being allocated the good is zero. The bidder may have an incentive to "throw" the auction at such marginal events, so as to instead obtain a positive share of revenue from selling to others at price v_i. Chapter 2 studies auctions of this form, and shows that in reasonably structured environments, an equilibrium exists in which the bidding strategy equates the marginal surplus from the allocation and the marginal surplus from sharing in revenue. Moreover, as revenue sharing becomes small, these distortions disappear and the seller recovers optimal reserve prices.

Finally, the BSA sets a price using the interim beliefs of a single losing bidder, combined with the relatively sparse information of the second-highest bid amongst all other bidders. It is natural to ask if there is a straightforward way to aggregate all of the losers' information, so that the seller sets an optimal price for the winning bidder i conditional on t_{-i}. This could easily be accomplished by having bidders report their entire hierarchy of beliefs as described above. Unfortunately, there is not an obvious simpler solution. One possibility would be to allow the losing bidders to "converse" about the optimal price, by iteratively reporting their conditional beliefs about the winner's value. Arguments in the vein of Geanakoplos and Polemarchakis (1982) would show that if such communication was allowed over multiple rounds, the bidders would eventually agree on a posterior. However, this posterior need not coincide with the true posterior conditional on t_{-i}. Moreover, if only losing bidders can have this conversation and receive the rewards that incentivize truth-telling, then these rents could create an incentive for bidders to throw the auction. Even so, such mechanisms are a promising direction for future research.

2.5.3 Common values

The assumption of private values is more appropriate in some settings than others. In the motivating example of a school construction project, it is reasonable to suppose there are private value components to firms' costs, such as prior commitments, worker abilities, etc. However, the firms might also have a common value in the idiosyncrasies of the project, such as the suitability of the land on which the school is to be built. Auction design with interdependent values can be challenging due to the buyers updating their preferences upon winning the auction. I briefly sketch the scope for generalizing my results to this broader setting. I have concluded that given reasonable assumptions on the interdependence, the same $\max\min$ extraction ratio obtains even if Nature is allowed to choose type spaces with interdependent preferences, although a much more complicated mechanism is required than the BSA.

As in much of the literature, I distinguish between "information" types t_i and "payoff" types $\theta_i \in \Theta_i$. A buyer's valuation is a function $\phi_i(\theta)$ of the profile of payoff types but does not depend on t. Thus, the definition of a type space is expanded to $\mathcal{T} = (\Theta, T, \pi, \phi)$. With interdependent values, the seller needs to elicit not just bidders' beliefs about θ, but also the form of the interdependence, i.e., $\phi(\theta)$.

Let us suppose for the moment that the seller knows Θ and ϕ. Many of the positive results in the literature require the assumption that Θ_i is one-dimensional, and that $\phi_i(\theta)$ is monotonically increasing.[6] This assumption, combined with a single-crossing property on ϕ, is sufficient for the existence of an efficient equilibrium of the English auction (Dasgupta and Maskin, 2000; Maskin, 1992; Krishna, 2003; Birulin and Izmalkov, 2011). Starting from this efficient equilibrium, the seller can partially implement a mechanism similar to the BSA, where bidders report θ_i and beliefs about θ_{-i}. These reported beliefs can be used to find the reserve price for when bidder i wins that maximizes:

$$r\,Pr\left(\left\{\widetilde{\theta}_i \middle| \phi_i\left(\widetilde{\theta}_i, \theta_{-i}\right) \geq r\right\} \middle| \theta_{-i}, t_{-i}\right).$$

The seller only sells the good to bidder i at this price if the realized value conditional on θ is at least r. Since the seller optimally accepts this price on behalf of the winning bidder, and the price does not depend on bidder i's report, truthful reporting is still an equilibrium. I will call this the interdependent belief survey auction (IBSA).

Under this mechanism, the efficient surplus and revenue with an interdependent value type space $\mathcal{T} = (\Theta, T, v, \pi)$ are the same as under the BSA with a particular private value type space $\mathcal{T}' = (T', \phi', \pi')$. Intuitively, I would like to find a private value type space which has the same distribution of the winner's

[6]For multidimensional Θ_i, existence of mechanisms with efficient equilibria becomes problematic. See Jehiel and Moldovanu (2001).

value conditional on losers' information. The bidders cannot simply be told their ex-post values, since they will reveal this to the seller, who might then be able to identify the winner's value from losers' reported values. However, a buyer can be told his value $\phi_i(\theta)$ on events where he wins, while losing bidders observe θ_i but receive private values of zero. Under such a private value type space, losers' beliefs about the winner's value are the same as under $\mathcal{T}$, and the distribution of the highest value is the same. This discussion implies that the IBSA achieves the same extraction ratio with $\mathcal{T}$ as the BSA achieves on the private value type space $\mathcal{T}'$. Hence, when minimizing the extraction ratio for this richer mechanism, it is sufficient to look at private value type spaces, and therefore the same lower bound $E^*(\gamma)$ obtains.

Finally, I return to the issue of extracting the form of the interdependence. The seller needs a general "detail-free" language in which to have the buyers communicate what they know. The preference hierarchies of Bergemann, Morris, and Takahashi (2011) are just such a language, in which bidders report a sequence of state-dependent preferences over Anscombe-Aumann acts. At the first level, the preference is over a state space with a single element corresponding to (θ_i, t_i). This preference is the player's willingness to pay for the good unconditional on other buyers' information. The second-order preference is over acts that depend on the first-order preferences of other buyers, and so on. At each level, types are separated by their preferences conditional on what others have revealed about their types, and these separated types can then be used to separate more types, in a manner analogous to Abreu and Matsushima (1992b). It turns out that bidders can be given strict incentives to truthfully reveal their interdependent preferences in this language, using the techniques of Subsection 2.4.6, and these reports can be used as an input to construct the IBSA. It is important to note that the seller can only provide strict incentives to recover a coarsened state space which corresponds to the distinguishable types of Bergemann, Morris, and Takahashi (2011, 2012). I revisit this connection in Chapter 3.

2.5.4 The role of the common prior

I have assumed throughout that the buyers' beliefs are derived from a common prior. A natural question to ask is whether or not my results can be extended to environments in which beliefs do not satisfy this restriction. The meaning of the common prior has been debated and critiqued in the literature (see Aumann, 1987; Morris, 1995; Gul, 1998). There are two possible interpretations: one is that there is some ex-ante stage before private information is realized, at which point there is common knowledge of the distribution over future private information. With such a temporal structure, the consistency of interim beliefs with a common prior is a consequence of ex-ante common knowledge. In the other interpretation, there is no ex-ante stage, but rather the common prior is a restriction on agents' higher-order beliefs (see Samet (1998) for a characterization of the common prior in terms of interim beliefs).

In my model, even if there is an ex-ante stage, the seller only interacts with the potential buyers *after* private information is realized. Hence, anything the seller learns about the prior must be obtained through the buyers' interim beliefs. The most natural interpretation is that there is some physical process which generates signals that the buyers see, and this process is know to the buyers but not the seller. As such, all of the bidders' interim beliefs about the profile of valuations are distributed around the average belief generated by this signal structure.

The prior distribution provides is a neutral perspective from which the seller can calculate expected revenue and efficient surplus. Also, since the buyers' reported beliefs will average to the prior, their interim reports give the seller access to a "segmentation" of the prior distribution of the highest valuation. Without a common prior, it would still be possible to elicit interim beliefs from the agents, but these reports would require a more complex interpretation to be useful to the seller. Thus, there is no immediate generalization of my result to non-common prior type spaces. However, as discussed by Azar, Chen, and Micali (2012), the result does not require each agent's entire hierarchy of beliefs to be consistent with a common prior. A similar result would obtain as long as the agents' first-order beliefs average to the same prior over values, and the seller uses this prior to calculate expected revenue and the expected efficient surplus.

2.6 Conclusion

This chapter has considered the mechanism design problem faced by an uninformed seller, who believes that agents are well-informed. The seller uses mechanisms that survey bidders' beliefs in addition to their private values. This information is used to optimize the prices offered to winning bidders. Such mechanisms achieve an optimal lower bound on the share of the efficient surplus that the seller can extract as revenue, and they also perform well away from the worst-case.

In practice, auctions are much more complicated than the stylized models studied by economists. For example, in the case of auctions for government contracts or natural resources, the auction designers conduct extensive research into demand, and surely engage in informal discussion with potential buyers about the pros and cons of different formats. This chapter has characterized particular ways in which the seller can elicit useful information from the buyers without distorting incentives to bid truthfully. But there is a more general message to be gleaned: if the seller has sufficient commitment power with regard to how information will be used, then it is indeed possible to have these informal discussions without allowing for adverse manipulation of the auction format.

2.A Proofs

Proof of Lemma 2.1. Fix a type space $\mathcal{T}$. I will show that there exists a payoff type space $\widehat{\mathcal{T}}$ such that $S(\mathcal{T}) = S(\widehat{\mathcal{T}})$ and $R(\mathcal{T}) \geq R(\widehat{\mathcal{T}})$. To that end, define $R_j(t_j)$ to be the revenue generated when bidder j is consulted. Since bidders report truthfully, this is:

$$R_j(t_j) = \frac{1}{n-1} \sum_{v_{-j}^{(2)}} \max_r r \sum_{v_{-j}^{(1)} \geq r, k} \frac{k}{\mathbb{I}_{\phi_j(t_j) = v_{-j}^{(1)}} + k} \pi\left(v_{-j}^{(1,2)}, k \middle| t_j \right),$$

where:

$$\pi\left(v_{-j}^{(1,2)}, k \middle| t_j \right) = \sum_{\left\{ t_{-j} \in T_{-j} \middle| \substack{\phi_{-j}^{(1,2)}(t_{-j}) = v_{-j}^{(1,2)}, \\ |W(v_{-j})| = k} \right\}} \pi\left(t_{-j} | t_j \right).$$

Consider the payoff type space in which $\widehat{T} = \phi(T)$, $\widehat{\phi}_i(t_i) = t_i$, and $\widehat{\pi}(t) = \sum_{t' \in \phi^{-1}(t)} \pi(t')$. The type space $\widehat{\mathcal{T}} = \left(\widehat{T}, \widehat{\pi}, \widehat{\phi} \right)$. Since the marginal distribution over values is the same between the two type spaces, clearly the same efficient surplus obtains. Also, the revenue generated by consulting bidder j when $\phi_j(t_j) =$

v_j is:

$$
\begin{aligned}
&\sum_{t_j \in \phi_j^{-1}(v_j)} \pi_j(t_j) R_j(t_j) \\
&\quad = \sum_{t_j \in \phi_j^{-1}(v_j)} \frac{\pi_j(t_j)}{n-1} \sum_{v_{-j}^{(2)}} \max_r r \sum_{v_{-j}^{(1)},k} \frac{k}{\mathbb{I}_{\phi_j(t_j)=v_{-j}^{(1)}} + k} \pi\left(v_{-j}^{(1,2)}, k \middle| t_j \right) \\
&\quad \geq \frac{1}{n-1} \sum_{v_{-j}^{(2)}} \max_r r \sum_{v_{-j}^{(1)},k} \frac{k}{\mathbb{I}_{\phi_j(t_j)=v_{-j}^{(1)}} + k} \sum_{t_j \in \phi_j^{-1}(v_j)} \pi_j(t_j) \pi\left(v_{-j}^{(1,2)}, k \middle| t_j \right) \\
&\quad = \frac{1}{n-1} \sum_{v_{-j}^{(2)}} \max_r r \sum_{v_{-j}^{(1)},k} \frac{k}{\mathbb{I}_{\phi_j(t_j)=v_{-j}^{(1)}} + k} \widehat{\pi}(v_{-j}|v_j) \\
&\quad = \widehat{R}_j(v_j).
\end{aligned}
$$

Thus, it is without loss of generality to consider payoff type spaces.

Symmetry follows. For a payoff type space $\mathcal{T}$ can always be made symmetric by uniformly randomizing over the $n!$ permutations $\xi : N \to N$ of the players' identities. Let the set of such permutations be denoted Ξ. Players types are $\widehat{T}_i = \Xi \times \cup_{i \in N} T_i$, $\widehat{\phi}_i(\xi_i, v_i) = v_i$, and:

$$
\widehat{\pi}(\xi, v) = \begin{cases} 0 & \text{if } \xi_i \neq \xi_j \text{ for some } i \text{ and } j \\ \frac{1}{n!}\pi(v^\xi) & \text{otherwise} \end{cases},
$$

where $v^\xi_{\xi(i)} = v_i$. Clearly, revenue and the efficient surplus are the same under the type space $\widehat{\mathcal{T}}$, but if the new types ξ_i are integrated out, S stays the same and R weakly decreases, and we are left with the symmetric and payoff type space $\widetilde{\mathcal{T}}$ which has the distribution over values:

$$
\widetilde{\pi}(v) = \frac{1}{n!} \sum_{\xi \in \Xi} \pi(v^\xi),
$$

types $\widetilde{T} = \cup_{i \in N} T_i$ and $\widetilde{\phi}_i(v_i) = v_i$.

Finally, a similar argument shows that revenue can always be lowered by giving every bidder except the winner have a value of $\underline{v}$. Starting with a symmetric and payoff type space $\mathcal{T}$, define $\widehat{\mathcal{T}}$ that has the same support for values T_i, but a distribution equal to:

$$\widehat{\pi}(v) = \begin{cases} f(x) & \text{if } V_i = x, v_j = \underline{v} \forall j \neq i \\ \pi(v) & \text{if } v_i = \underline{v} \ \forall i \\ 0 & \text{otherwise} \end{cases},$$

where:

$$f(x) = \sum_{\left\{\widetilde{v} \middle| v_i = \widetilde{v}^{(1)} = x\right\}} \frac{\pi(v)}{|W(v)|},$$

which is independent of i by symmetry.

I verify that $\mathcal{T}$ and $\widehat{\mathcal{T}}$ have the same efficient surplus by checking that the probability that $v^{(1)} = x$ is the same for both type spaces:

$$\begin{aligned}
\sum_{\left\{v \middle| v^{(1)} = x\right\}} \pi(v) &= \sum_{\left\{v \middle| v^{(1)} = x\right\}} \sum_i \mathbb{I}_{v_i = x} \frac{\pi(v)}{|W(v)|} \\
&= \sum_i \sum_{\left\{v \middle| v_i = v^{(1)} = x\right\}} \frac{\pi(v)}{|W(v)|} \\
&= \sum_i \sum_{\{v | v_i = x, v_j = \underline{v} \ \forall j \neq i\}} \widehat{\pi}(v) \\
&= \sum_{\left\{v \middle| v^{(1)} = x\right\}} \widehat{\pi}(v).
\end{aligned}$$

Revenue is lower, as:

$$\begin{aligned}
\sum_{v_j} \pi_j(v_j) R_j(v_j) &= \sum_{v_j} \sum_{y} \max_{r \geq 0} r \sum_{x \geq \max\{v_j, r\}} \sum_{\left\{\widetilde{v} \middle| \widetilde{v}_j = v_j, \widetilde{v}_{-j}^{(1,2)} = (x,y)\right\}} \frac{|W(\widetilde{v})| - \mathbb{I}_{\widetilde{v}_j = x}}{|W(\widetilde{v})|} \pi(\widetilde{v}) \\
&\geq \max_{r \geq 0} r \sum_{v_j} \sum_{y} \sum_{x \geq \max\{v_j, r\}} \sum_{\left\{\widetilde{v} \middle| \widetilde{v}_j = v_j, \widetilde{v}_{-j}^{(1,2)} = (x,y)\right\}} \frac{|W(\widetilde{v})| - \mathbb{I}_{v_j = x}}{|W(\widetilde{v})|} \pi(\widetilde{v}) \\
&= \max_{r \geq 0} r \sum_{x \geq \max r} \sum_{\left\{\widetilde{v} \middle| \widetilde{v}^{(1)} = x\right\}} \frac{|W(\widetilde{v})| - \mathbb{I}_{\widetilde{v}_j = x}}{|W(\widetilde{v})|} \pi(\widetilde{v}) \\
&= \max_{r \geq 0} r \sum_{x \geq r} \frac{|W(\widetilde{v})| - \mathbb{I}_{\widetilde{v}_j = x}}{|W(\widetilde{v})|} f(x) \\
&= \max_{r \geq 0} r \sum_{\left\{\widetilde{v} \middle| \widetilde{v}_j = \underline{v}, v^{(1)} \geq r\right\}} \widehat{\pi}(\widetilde{v}) = \widehat{R}_j.
\end{aligned}$$

□

Proof of Proposition 2.1. It is clear that symmetric payoff type spaces in which at most one bidder has a positive value are defined by the distribution of the highest value $F^{(1)}$. Moreover, for a given R, it must be that (2.9) is a lower bound on the distribution. Thus, it must be that:

$$\begin{aligned}
S(\mathcal{T}) &= \int_{v=0}^{\gamma S(\mathcal{T})} v \, dF^{(1)}(v) \\
&\geq R(1 + \log(\gamma) + \log(S(\mathcal{T})) - \log(R)),
\end{aligned}$$

and therefore:

$$1 \geq E(\mathcal{M}^{BSA}, \mathcal{T})(1 + \log(\gamma) - \log(E(\mathcal{M}^{BSA}, \mathcal{T})))$$

for any $\mathcal{T}$. Now let us consider the quantity:

$$h(x) = 1 - x(1 + \log(\gamma) - \log(x)).$$

It is straightforward to derive:

$$h'(x) = -(1 + \log(\gamma) - \log(x)) + 1 = \log(x) - \log(\gamma) < 0,$$

since $x \in [0, 1]$ and $\gamma > 1$, and h is strictly decreasing. Also:

$$\begin{aligned} \lim_{x \to 0} h(x) &= 1 - \lim_{x \to 0} \frac{1 + \log(\gamma) - \log(x)}{x^{-1}} \\ &= 1 - \lim_{x \to 0} \frac{-x^{-1}}{-x^{-2}} = 1 \end{aligned}$$

via L'Hôpital's rule and $h(1) = 1 - (1 + \log(\gamma)) = -\log(\gamma) < 0$. Thus, there exists a unique point x^* at which $h(x^*) = 0$, and $h(x) > 0$ iff $x < x^*$. This implies that $E(\mathcal{M}^{BSA}, \mathcal{T}) \geq E^*(\gamma)$ for every $\gamma > 1$.

All that remains to be seen is that there is a sequence of type spaces $\mathcal{T}^k$ such that $E(\mathcal{M}^{BSA}, \mathcal{T}^k) \to E^*(\gamma)$. We are careful to make sure each $\mathcal{T}^k$ has support in $[0, \gamma S(\mathcal{T}^k)]$. Take S^k any sequence converging to $\frac{R}{E^*(\gamma)}$, and $V^k = \{v_0, \ldots, v_{m_k}\}$ is the support of $\mathcal{T}^k$ with $v_l = \frac{l}{m_k} \gamma S^k$. We use the CDF of the highest value $F_k^{(1)}$ defined by $F_k^{(1)}(v) = F^{(1)}(v_{l+1})$ for all $v \in [v_l, v_{l+1})$ with $F^{(1)}$ as in (2.9). For each k, as $m_k \to \infty$:

$$\int_{v=0}^{\gamma S^k} v \, dF_k^{(1)}(v) \to_{m_k \to \infty} R(1 + \log(\gamma) + \log(S^k) - \log(R)) > S^k,$$

since $h\left(\frac{R}{S^k}\right) < 0$. Thus, m_k can be taken large enough so that $S(\mathcal{T}^k) > S^k$, and therefore satisfies Assumption 2.1. Moreover, $S(\mathcal{T}^k) \leq \frac{R}{E^*(\gamma)}$ by the argument of the previous paragraph, so $S(\mathcal{T}^k) \to \frac{R}{E^*(\gamma)}$ by the squeeze theorem. $\square$

Proof of Theorem 2.1. I show that for the sequence $\mathcal{T}^k$ constructed in the proof of Proposition 2.1, the supremum of $E(\mathcal{M}, \mathcal{T}^k, \sigma)$ over all $\mathcal{M}$ and σ converges to $E^*(\gamma)$. Since the supremum over σ is weakly greater than the infimum over σ, this will prove the lemma. The rest of the proof is standard, and follows Myerson (1981) or Börgers (2013).

Since the designer is allowed to pick σ, it is sufficient to look at direct revelation mechanisms where $M = T$. We can divide profiles of valuations into those in which bidder i has a positive value, and all other bidders have a zero valuation. Let $q_i(v)$ and $p_i(v)$ be bidder i's allocation and transfer if i has a positive value v, and other bidders have zero values. Let $\underline{q}^i_j(v)$ and $\underline{p}^i_j(v)$ be bidder j's allocation and transfer when j has a zero value, and bidder i has a positive value v. If $v > v'$, then it must be:

$$
\begin{aligned}
& v\, q_i(v) - p_i(v) \geq v\, q_i(v') - p_i(v') \\
& v'\, q_i(v') - p_i(v') \geq v' q_i(v) - p_i(v) \\
\implies & v(q_i(v) - q_i(v')) \geq p_i(v) - p_i(v') \geq v'(q_i(v) - q_i(v')),
\end{aligned}
$$

so $q_i(v) \geq q_i(v')$, and the allocation must be weakly increasing. Moreover, it must be that:

$$\begin{aligned} u_i(v) - u_i(v') &= v\, q_i(v) - p_i(v) - v'\, q_i(v') + p_i(v') \\ &\geq v\, q_i(v') - p_i(v') - v'\, q_i(v') + p_i(v') \\ &= (v - v')q_i(v'), \end{aligned}$$

and thus, if the support of values is indexed by $\{v^0 = 0, \ldots, v^L = \overline{v}\}$, then:

$$u_i(v^l) \geq u_i(0) + \sum_{m=0}^{l-1} (v^{m+1} - v^m) q_i(v^m),$$

and thus:

$$\begin{aligned} p_i(v^l) &= v^l\, q_i(v^l) - u_i(v^l) \\ &\leq v^l\, q_i(v^l) - u_i(0) - \sum_{m=0}^{l-1} (v^{m+1} - v^m) q_i(v^m). \end{aligned}$$

Finally, individual rationality tells us that $u_i(0) \geq 0$.

Also, since the utility of the low type is always $0\underline{q}^i_j(v) - \underline{p}^j_i(v) = -\underline{p}^j_i(v)$, which must be non-negative to satisfy individual rationality, it follows that $\underline{p}^j_i(v) \leq 0$. It is never beneficial for the seller to allocate the good to a bidder with valuation zero, since they will never pay a positive amount. The seller might as well leave the good unallocated.

Hence, an upper bound on the seller's revenue is:

$$\sum_{i \in N} \sum_{l=0}^{L} p_i(v^l) \pi_i(v^l),$$

which is a linear function of the $q_i(v^l)$, as shown above. The set of weakly increasing $q_i(v^l)$ is a convex set and its extreme points are those functions for which $q_i(v^l) \in \{0, 1\}$. These are precisely the allocations that are implemented by posted price rules, where there are bidder specific reservation prices r_j^*. But since each bidder's valuation has distribution proportional to $F_k^{(1)}$ when positive, an optimal reserve price is by construction $r_j^* = R$. At this price, revenue is exactly R, since the bidder with the high value always buys. This proves the result. $\square$

Proof of Proposition 2.2. The seller would achieve the same revenue as in a second-price auction with anonymous reserve r^* if the price set for winning bidders is $\max\{r^*, b^{(2)}\}$. Observe:

$$
\begin{aligned}
R_j(t_j) &= \frac{1}{n-1} \sum_{v_{-j}^{(2)}} \max_r r \sum_{v_{-j}^{(1)} \geq r, k} \frac{k}{\mathbb{I}_{\phi_j(t_j) = v_{-j}^{(1)}} + k} \pi\left(v_{-j}^{(1,2)}, k \middle| t_j \right) \\
&\geq \frac{1}{n-1} \sum_{v_{-j}^{(2)}} \max\left\{ r^*, v_j, v_{-j}^{(2)} \right\} \\
&\qquad \cdot \sum_{v_{-j}^{(1)} \geq \max\left\{ r^*, v_j, v_{-j}^{(2)} \right\}, k} \frac{k}{\mathbb{I}_{\phi_j(t_j) = v_{-j}^{(1)}} + k} \pi\left(v_{-j}^{(1,2)}, k \middle| t_j \right) \\
&= \frac{1}{n-1} \sum_{\left\{ t_{-j} \middle| \substack{\phi_{-j}^{(1)}(t_{-j}) \\ \geq \max\{r^*, \phi_j(t_j)\}} \right\}} \max\left\{ r^*, \phi^{(2)}(t) \right\} \left(1 - \frac{\mathbb{I}_{\phi_j(t_j) = \phi^{(1)}(t)}}{|W(\phi(t))|} \right) \pi(t_{-j} | t_j).
\end{aligned}
$$

Hence, the total revenue from the auction satisfies:

$$\begin{aligned}
&\sum_j \sum_{t_j \in T_j} \pi_j(t_j) R_j(t_j) \\
&\geq \sum_{j, t_j \in T_j} \frac{\pi_j(t_j)}{n-1} \sum_{\left\{t_{-j} \middle| \substack{\phi_{-j}^{(1)}(t_{-j}) \\ \geq \max\{r^*, \phi_j(t_j)\}}\right\}} \max\left\{r^*, \phi^{(2)}(t)\right\} \left(1 - \frac{\mathbb{I}_{\phi_j(t_j)=\phi^{(1)}(t)}}{|W(\phi(t))|}\right) \pi(t_{-j}|t_j) \\
&= \frac{1}{n-1} \sum_j \sum_{\left\{t \in T \middle| \phi^{(1)}(t) \geq r^*\right\}} \max\left\{r^*, \phi^{(2)}(t)\right\} \left(1 - \frac{\mathbb{I}_{\phi_j(t_j)=\phi^{(1)}(t)}}{|W(\phi(t))|}\right) \pi(t) \\
&= \frac{1}{n-1} \sum_{\left\{t \in T \middle| \phi^{(1)}(t) \geq r^*\right\}} \max\left\{r^*, \phi^{(2)}(t)\right\} \pi(t) \sum_j \left(1 - \frac{\mathbb{I}_{\phi_j(t_j)=\phi^{(1)}(t)}}{|W(\phi(t))|}\right) \\
&= \sum_{\left\{t \in T \middle| \phi^{(1)}(t) \geq r^*\right\}} \max\left\{r^*, \phi^{(2)}(t)\right\} \pi(t),
\end{aligned}$$

since:

$$\frac{|W(\phi(t))| - \mathbb{I}_{\phi_j(t_j)=\phi^{(1)}(t)}}{|W(\phi(t))|} \pi(t_{-j}|t_j)$$

is zero if $\phi_j(t_j) > \phi_{-j}^{(1)}(t_{-j})$ and also:

$$\sum_j \frac{|W(\phi(t))| - \mathbb{I}_{\phi_j(t_j)=\phi^{(1)}(t)}}{|W(\phi(t))|} = n - 1.$$

Since $\sum_{\left\{t \in T \middle| \phi^{(1)}(t) \geq r^*\right\}} \max\left\{r^*, \phi^{(2)}(t)\right\} \pi(t)$ is revenue under a second-price auction with the anonymous reserve r^*, this proves the result. $\square$

Chapter 3

Revenue sharing in second-price auctions

3.1 Introduction

The second-price auction with optimal reserve prices has many desirable properties. There is a compelling equilibrium in which bidders follow the unique weakly undominated strategy of bidding their values. In benchmark environments, the second-price auction with a judiciously chosen reserve price is an optimal auction (Myerson, 1981). Even for more general classes of environments, the second-price auction is a fair approximation of the optimal auction. Hartline and Roughgarden (2009) show that the second-price auction with an optimal anonymous reserve price attains at least 25% of the revenue of the optimal auction, as long as the distribution of values is independent and regular. With bidder specific reserves, this improves to 50% (Chawla, Hartline, and Kleinberg, 2007). Also, it is known that if the seller sets an optimal reserve price for the winner conditional on the losers' values, the auction generates at least 50% of the maximum revenue possible in a dominant strategy and ex-post individually rational mechanism (Ronen, 2001).

The list goes on. The gist is that the second-price auction is a reliable mechanism for generating revenue, with the proviso that the seller must set the correct reserve price. This is no trivial matter. A large literature in auction econometrics has explored the two step process of estimating the distribution of values from bid data, and using this as an input to calculate an optimal reserve price (Athey and Haile (2007) survey this literature). An inescapable fact is that these methods require lots of data and/or non-trivial assumptions about the distribution of values in order to calculate the optimal reserve from past bid data.

In this chapter, I explore an alternative route to the optimal reserve price: ask the bidders. If the buyers are themselves well-informed about the distribution of values, the seller could elicit this information and have the bidders set reserve prices for one another. I consider a variation on the second-price auction in which the seller asks losing bidders to suggest reservation prices for the high valuation bidder. When a sale is made using a bidder's suggestion, that bidder receives a small share of the resulting revenue. This linear revenue-sharing contract perfectly aligns the incentives of the buyer and seller, conditional on the buyer losing. As such, this seems to be the simplest and most natural method for eliciting the optimal reserve.

In order to obtain a share of revenue, however, a bidder has to lose. Since a bidder will only lose the auction when bidding b when some other participant bids more, the suggested reserve price conditional upon losing must be at least b. I call this the pricing constraint. If the pricing constraint is binding, then at the margin, shared revenue would increase if the bidder were to shade his bid. On the other hand, the pivotal allocation that is affected by shading just below one's value occurs when the price paid would also be close to one's value. In this event, the surplus from receiving the good will be zero. Thus, players cannot bid their value if the pricing constraint binds. Of course, if the buyer were to shade a large amount so as to set a very low reserve price, then price at which revenue would be shared would be much lower than v. Clearly there would not be a benefit to shading close to zero. This suggests that there could be an equilibrium level of shading that is positive but not too large.

My main result is to show that for environments satisfying a particular positive correlation condition, there is indeed a simple equilibrium in which bidders shade to balance the marginal surplus they could obtain by being allocated the good and the marginal shared revenue they obtain by losing and suggesting reserve prices for others. The positive correlation condition takes the form of a requirement that the conditional hazard rate for the distribution of the highest value of others is monotonically decreasing in one's own value. This condition is similar in spirit to other positive correlation conditions used in the literature, such as the affiliated values of Milgrom and Weber (1982) or monotone likelihood ratios of Athey (2001). The equilibrium bidding function is characterized by regions on which the pricing constraint binds, where the bidding function solves a differential equation equating the sum of marginal surplus and marginal revenue with zero, and regions on which the pricing constraint does not bind and bidding one's value is a best response.

In equilibrium, a bidder with valuation v bids no less than $\frac{v}{1+\frac{\alpha}{n-1}}$, where α is the share of revenue when the suggested reserve price is used and n is the number of participants. Thus, as α goes to zero, shading disappears, bids converge to values, and the reserve prices converge to their optimal quantities. As a practical matter, there may be a point at which α is so small that bidders are not sensitive to the revenue sharing incentive. A nice feature of my results is that for strictly positive α, I have a tight characterization of an equilibrium that yields bounds on the revenue lost from sharing and distortions. In particular, the seller is always guaranteed a revenue of $\pi^* \frac{1-\alpha}{1+\frac{\alpha}{n-1}}$, where π^* is revenue from the second-price auction with an optimally chosen anonymous reserve price.

In addition to the benchmark model described above, I also consider several extensions. First, I discuss what happens with asymmetric distributions in the context of an independent two bidder example. While a symmetric distribution gives rise to strictly increasing bidding functions, with asymmetric bidders, there may be regions where one player's bidding function is constant and the other's is decreasing. Moreover, there may be regions of the bid space on which one bidder's pricing constraint is binding and the other bidder's constraint is slack. Second, in the benchmark mechanism described above, bidders are rewarded with a share of the seller's realized revenue, but only on the event that the bidder loses. A variation on this mechanism rewards each bidder i regardless of whether they win the auction, with a share of "simulated" revenue that would have obtained if the seller had run the $n-1$ bidder auction excluding bidder i and using his suggested reserve price. Since bidders are rewarded regardless of the allocation, there is no incentive to throw the auction, and truthful bidding is an equilibrium. For regular, symmetric, and independent distributions, this mechanism does just as well, but I discuss its limitations in more general environments. Finally, the assumptions of positive correlation and a one-dimensional structure on types are needed to demonstrate the existence of an equilibrium. If existence is assumed, then I give an example of a mechanism with a slightly less intuitive structure that nonetheless provides similar revenue bounds in general type spaces.

In related work, I have explored how an auction designer can extract details of the environment from well-informed buyers. Chapter 3 will show that the designer can effectively extract “for free” all of the information that is common knowledge among the agents. The seller can use this information to design the mechanism as he sees fit, without affecting bidders’ incentives to tell the truth. The caveat is that this mechanism requires the agents to report their entire hierarchy of beliefs, a decidedly complicated object. In contrast, Chapter 1 looked at simpler mechanisms that maximize the minimum extraction ratio, which is the ratio between expected revenue and the expected surplus that would be generated by allocating the good efficiently. This problem is solved by a modified second-price auction, in which bidders report bids as well as first-order beliefs about the distribution of others’ values. The seller uses these first-order beliefs to calculate reserve prices. Both of these chapters cover general finite type spaces. Chapter 1 requires that the bidders have private values and higher-order beliefs that are consistent with a common prior, but Chapter 3 allows for interdependent preferences and non-common prior beliefs, though it does require that a “common support” assumption be satisfied.

The present chapter, in contrast, looks at more structured environments in which bidders have a single hierarchy of beliefs corresponding to each value and the joint distribution of values admits a density. In such environments, second-price auctions with revenue sharing allow the seller to elicit the optimal reserve prices with the simplest possible message space, consisting of bids and suggested prices. Granted, bidders are asked to compute an optimal reserve conditional on losing at their value, which is a non-trivial task. However, by sharing revenue, the losing bidders’ incentives are closely aligned with those of the seller. In contrast, the mechanisms explored in Chapters 1 and 3 incentivize the bidders to report their beliefs using scoring rules, and then the seller performs the computation. It may well be that computing optimal or near optimal reserves is easier for the bidders than communicating a distribution. For example, the bidders may have privacy concerns with regard to their private information about their competitors. Reporting a price allows the bidders to communicate what the seller needs without divulging any extra information that might be a liability.

There is a small but growing literature on how the seller can run an auction and simultaneously calibrate auction parameters using ancillary reports made by the buyers about the environment. Caillaud and Robert (2005) consider how a seller can partially implement the optimal auction of Myerson (1981) through a dynamic mechanism. Dasgupta and Maskin (2000) construct a mechanism that partially implements the efficient outcome in interdependent value settings, in which bidders submit a function that gives a bid for every possible valuation of the other player. The seller computes the winner and price by looking for a fixed point of the reported mappings, and as such, the seller needs no additional information about the environment beyond what is reported. Azar, Chen, and Micali (2012) also study the use of scoring rules to recover a truncated prior distribution over values, and then use this prior to design the mechanism.

The rest of this chapter is organized as follows. Section 3.2 describes the environment and defines the second-price auction with revenue sharing. Section 3.2 also gives a definition of my equilibrium concept, which imposes regularity conditions on the bidding function. Section 3.3 gives a simple example that illustrates some of the main ideas of my construction. Section 3.4 provides this general characterization for joint distributions of private values with positive correlation. Section 3.5 discusses several extensions, and Section 3.6 concludes.

3.2 Model

Some preliminary notation: for a vector x, let $x^{(1)}$ denote the highest value, $x^{(2)}$ the second-highest value, and $x^{(1,2)}$ the ordered pair of the highest and second-highest values. If x has a single coordinate, then $x^{(2)} = -\infty$. Let $W(x) = \{i | x_i = x^{(1)}\}$ denote the set of high value indices. I denote by $x \vee y$ and $x \wedge y$ the maximum and minimum of x and y, respectively. I also use the usual convention that x_S denotes the sub-vector of x with indices in S, and x_{-S} denotes the sub-vector with indices not in S.

There are n bidders, indexed by $i \in N = \{1, \ldots, n\}$. Bidders have private valuations for a single unit of a good that are jointly distributed according to the cumulative distribution $F(v_1, \ldots, v_n)$ with compact support $[\underline{v}, \overline{v}]^n$. This distribution is symmetric in v and admits a strictly positive and continuous density $f(v)$. In Section 3.5, I extend the analysis to a class of asymmetric distributions. I assume that bidders do not receive any additional information beyond their private value. This distribution, while unknown to the seller, is known to the bidders.

I will have need of several conditional densities and cumulative distributions, including but not limited to $f_{v_j|v_i}(\cdot|\cdot)$, $f_{v_{-i}^{(1)}|v_i}(\cdot|\cdot)$, $f_{v_{-ij}^{(1)}|v_i,v_j}(\cdot|\cdot,\cdot)$ where ij is shorthand for $\{i, j\}$. For reasons which will subsequently become clear, I assign compact notation to the following quantity:

$$g(x, y|v) = \frac{F_{v_{-ij}^{(1)}|v_i,v_j}(y|x, v) f_{v_j|v_i}(x|v)}{F_{v_j,v_{-ij}^{(1)}|v_i}(x, \overline{v}|v) - F_{v_j,v_{-ij}^{(1)}|v_i}(x, y|v)}.$$

This is the hazard rate of bidder j's value when values of bidders $k \neq i, j$ are less than y, conditional on bidder i's value being v. This quantity is closely related to the choice of an optimal reserve price if some bidder i with valuation v were to sell the good to the remaining $n-1$ bidders. With a reserve price r, revenue would be:

$$\int_{x=r}^{\overline{v}} \left[rF_{v_{-ij}^{(1)}|v_i,v_j}(r|v,x) + \int_{y=r}^{x} y f_{v_{-ij}^{(1)}|v_i,v_j}(y|v,x)dy \right] f_{v_j|v_i}(x|v)dx, \tag{3.1}$$

where bidder j is taken to be a "representative" bidder in $-i$ with the highest value. Such a bidder pays r if $v_{ij}^{(1)} \leq r$, and pays $v_{-ij}^{(1)}$ otherwise. The derivative with respect to r is:

$$\begin{aligned} &\int_{x=r}^{\overline{v}} F_{v_{-ij}^{(1)}|v_i,v_j}(r|v,x) f_{v_j|v_i}(x|v)dx - rF_{v_{-ij}^{(1)}|v_i,v_j}(r|v,r) f_{v_j|v_i}(r|v) \\ &\qquad = \left(F_{v_j,v_{-ij}^{(1)}|v_i}(\overline{v},r|v) - F_{v_j,v_{-ij}^{(1)}|v_i}(r,r|v) \right) (1 - r\, g(r,r|v))\,. \end{aligned} \tag{3.2}$$

I make the following two assumptions:

A1 For every x and y, the function $g(x,y|v)$ is weakly decreasing in v.

A2 There exists finitely many v at which $g(r,r|v) = \frac{1}{v}$.

A1 is a substantive restriction, analogous to the positive correlation conditions of Milgrom and Weber (1982) and Athey (2001). It essentially requires that higher values of v_i make higher values of v_j more likely, in the sense that for every interval $[x, \overline{v}]$, v_j is less likely to be at the bottom of the interval x when $v_{-ij}^{(1)} \leq y$. **A1** is trivially satisfied in the case of independent values, and for two bidders it reduces to the familiar monotone hazard rate condition, that $\frac{f(r|v)}{1-F(r|v)}$ is decreasing in v. This property will ensure that higher valuation bidders want to set higher reserve prices in equilibrium. **A2** is a technical restriction which facilitates a simple equilibrium construction. Without **A2**, I would have to address the possibility that there are regions where $g(v,v|v) = \frac{1}{v}$, which in the independent private value setting correspond to cases where the virtual valuation is zero. Also, if there are infinitely many points at which this equality holds, transfinite induction would be required for the proofs of my main results, as opposed to the finite induction currently used.

The seller of the good uses the following *revenue sharing auction* (RSA). Each bidder i submits a bid b_i and a reserve price r_j. If the profile of bids is b, then the seller picks a bidder $i \in W(b)$ uniformly, and then picks a losing bidder $j \neq i$ uniformly to consult for the price. If $b_i \geq r_j$, then bidder i is awarded the good at price $\max\left\{r_j, b_{-j}^{(2)}\right\}$, and bidder j receives $\alpha \max\left\{r_j, b_{-j}^{(2)}\right\}$ as his "share" of the revenue, with $\alpha \in (0,1]$. Otherwise, the good remains unallocated, and no transfers are made. The interpretation is that bidder j sets the reserve price in the $n-1$ bidder auction excluding j, and receives an α share of revenue generated by that auction.

I will define a class of symmetric equilibria consisting of a bidding function $\beta(v)$ and a pricing function $\rho(v)$, which are required to satisfy the following two properties:

E1 The function β is continuous and strictly increasing.

E2 For all $v \in [\underline{v}, \overline{v})$, one of the two holds:

(i) $\beta(v) = v$, or

(ii) A bidder of type v strictly prefers bidding $\beta(v)$ to bidding v.

In other words, each player bids their value unless there is a strict incentive to do otherwise. I exclude the type with $v_i = \overline{v}$ from this requirement, for reasons which will be seen shortly. Note that **E1** implies that β has a well defined inverse β^{-1} on its range. If $b > \beta(\overline{v})$, take $\beta^{-1}(b) = \overline{v}$, and similarly if $b < \beta(\underline{v})$, $\beta^{-1}(b) = \underline{v}$.

In addition, the following incentive constraint must be satisfied:

$$S(v, \beta(v)) + R(v, \beta(v), \rho(v)) \geq S(v, b) + R(v, b, r) \tag{3.3}$$

for all (v, b, r), where:

$$S(v, b) = \mathbb{E}\left[\left(v - \rho(v_j) \vee \beta\left(v_{-ij}^{(1)}\right)\right) \mathbb{I}_{\rho(v_j) \vee \beta\left(v_{ij}^{(1)}\right) < b} \middle| v_i = v\right] \tag{3.4}$$

$$R(v, b, r) = \alpha\, \mathbb{E}\left[\rho(v) \vee \beta\left(v_{-i}^{(2)}\right) \mathbb{I}_{\beta\left(v_{-i}^{(1)}\right) \geq b \vee r} \middle| v_i = v\right] \tag{3.5}$$

are respectively surplus from being allocated the good and revenue from selling to others, when other participants use the bidding function β and pricing function ρ. I will say that (β, ρ) constitute a *regular equilibrium* if they satisfy **E1**, **E2**, and (3.3).

My main result is that a regular equilibrium of the RSA exists. This equilibrium has an intuitive structure in which bidders sometimes shade in response to the revenue-sharing incentives.

3.3 A simple example

It is instructive to start with a simple example of the kind of equilibrium that I will construct. Let us suppose that there are two bidders whose values are distributed independently and uniformly between 0 and 1. As described above, each bidder submits a bid b_i and a price r_i. If $b_i > b_j$, then bidder i "wins" the auction, but only receives the good if $b_i > r_j$. In this case, bidder i pays r_j to the seller, and bidder j receives $\alpha\, r_j$.

In the undominated equilibrium of the second-price auction, bidders bid their values, i.e., $\beta(v) = v$. Let us investigate whether or not bidders could use such a strategy in the RSA with a strictly positive α. If so, the distribution of bids is uniform on $[0, 1]$, meaning that the optimal reserve price unconditional on losing is 0.5. However, bidders only share revenue when they lose, and the distribution of the other bidder's bids conditional on losing with a bid of b is uniform on $[b, 1]$. Naturally, it cannot be optimal to set a reserve price $r < b$. This observation is valid more generally: In any regular equilibrium, it must be that $\beta(v) \leq \rho(v)$. As such, it makes sense to impose the pricing constraint that $r \geq b$ and simply write $R(v, r)$ instead of $R(v, b, r)$.

Hence, in an equilibrium in which $\beta(v) = v$, it must be that $\rho(v) \in \arg\max_{r \geq v} r(1-r)$, so that $\rho(v) = 0.5$ for $v < 0.5$, and $\rho(v) = v$ for $v > 0.5$. For this to be an equilibrium, it must be that (3.3) is satisfied. Note that:

$$S(v, b) = (v - 0.5)0.5\, \mathbb{I}_{b \geq 0.5} + \int_{x=0.5}^{b} (v - x)dx,$$

$$R(v, r) = \alpha\, r(1 - r).$$

The bidder's goal is to maximize:

$$U(v, b, r) = S(v, b) + R(v, r),$$

subject to $r \geq b$. For any deviation $b \geq 0.5$, it is optimal to set $r = b$. Hence, the equilibrium bid must satisfy the following first-order condition:

$$\left. \frac{\partial S(v, b)}{\partial b} + \frac{\partial R(v, b)}{\partial b} \right|_{b=\beta(v)} \geq 0. \tag{3.6}$$

Under the assumption that $\beta(v) = v$, this evaluates to:

$$v - v + \alpha(1 - 2v) = 0,$$

which is obviously violated for $v > 0.5$. The intuition is as follows: The marginal allocation affected by shading when $b = v$ is when the other player sets a price $r = v$. In this case, the marginal surplus from the allocation is small, since $v - r \approx 0$. On the other hand, if $r > 0.5$, marginal revenue is strictly negative, since a price of 0.5 would be optimal if the constraint $r \geq b$ were not binding. At the margin, a bidder could shade a bit, and replace events on which he wins the good at a price close to v with events on which he sells at prices close to v, which leads to a strict improvement.

Indeed, there is a bidding function which does satisfy (3.6) when $r = \beta(v)$ is optimal, which is:

$$\beta(v) = \begin{cases} v & \text{if } v < v^* \\ \frac{\alpha + (1+\alpha)v}{(1+2\alpha)(1+\alpha)} & \text{if } v \geq v^* \end{cases}, \tag{3.7}$$

where $v^* = \frac{1}{2(1+\alpha)}$. This bidding function is depicted in Figure 3.1 for the case where $\alpha = \frac{1}{4}$. Note that the probability of a bid less than r is:

$$F(r) = \begin{cases} r & \text{if } r < v^* \\ (1+2\alpha)r - \frac{\alpha}{1+\alpha} & \text{if } r \geq v^* \end{cases}. \tag{3.8}$$

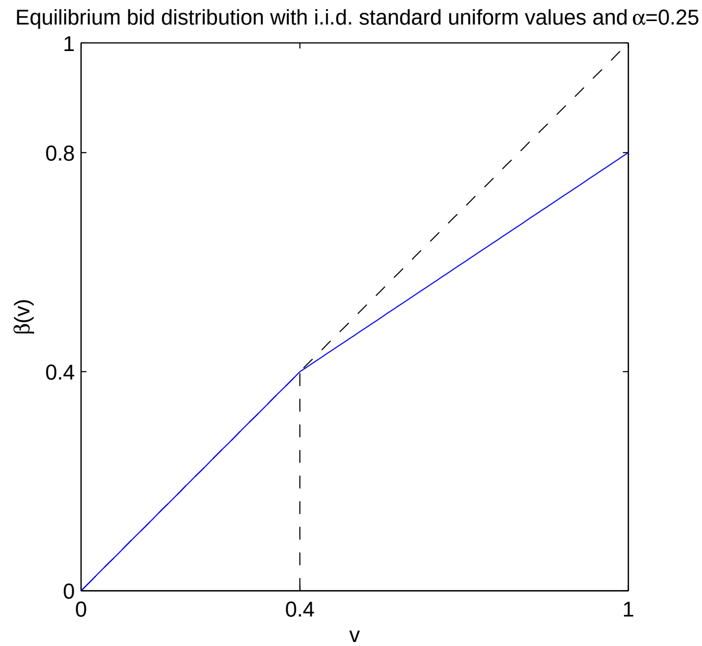

Figure 3.1: The equilibrium bidding function when values are distributed uniformly and independently on $[0,1]$, and $\alpha = \frac{1}{4}$. Note that $v^* = 0.4$ and $\beta(1) = 0.8$.

Thus, marginal revenue is:

$$1 - F(r) - rf(r) = \begin{cases} 1 - 2r & \text{if } r < v^* \\ (1+2\alpha)\left(\frac{1}{1+\alpha} - 2r\right) & \text{if } r \geq v^* \end{cases},$$

which is clearly positive if $r < v^*$, and negative otherwise, as:

$$\frac{1}{1+\alpha} - 2(1+2\alpha)r \leq -\frac{\alpha}{1+\alpha}.$$

As such, if $\beta(v)$ is an equilibrium bidding function, it must be that $\rho(v) = v^*$ if $v < v^*$ and $\rho(v) = \beta(v)$ otherwise. Consequently:

$$\begin{aligned} S(v,b) &= (v - v^*)v^* \, \mathbb{I}_{b \geq v^*} + \int_{x=v^*}^{b} (v-x)(1+2\alpha)dx, \\ R(v,r) &= \alpha \, r(1 - F(r)), \end{aligned}$$

where $F(r)$ is given by (3.8), and for $v > v^*$, (3.6) evaluates to:

$$\begin{aligned} &\left(v - \frac{\alpha + (1+\alpha)v}{(1+\alpha)(1+2\alpha)}\right)(1+2\alpha) + \alpha(1+2\alpha)\left(\frac{1}{1+\alpha} - 2\frac{\alpha+(1+\alpha)v}{(1+\alpha)(1+2\alpha)}\right) \\ &= (1+2\alpha)\left(v + \frac{\alpha}{1+\alpha} - (1+2\alpha)\frac{\alpha+(1+\alpha)v}{(1+\alpha)(1+2\alpha)}\right) = 0. \end{aligned}$$

We conclude that (β, ρ) is indeed a regular equilibrium.

The form for β was not chosen arbitrarily. Note that $\beta(1) = \frac{1}{1+\alpha}$. In a regular equilibrium, the bidder with the highest valuation must win all the time with the highest bid, $\beta(1)$. The marginal surplus lost from shading is $(1 - \beta(1))f(1)$. On the other hand, shading makes it possible to sell to the other bidder when he has the highest value, the marginal revenue from which would be $-\alpha\beta(1)f(1)$. If the bidder with valuation 1 is indifferent to shading in equilibrium, then it must be that $\beta(1) = \frac{1}{1+\alpha}$. I will show in Lemma 3.2 that this condition generalizes to a requirement that in a regular equilibrium, $\beta(\overline{v}) = \frac{\overline{v}}{1+\frac{\alpha}{n-1}}$.

Also, suppose $\beta(v) = \rho(v)$ and is differentiable on a neighborhood of v in equilibrium, and consider local deviations to a bid $\beta(w)$ made by a nearby valuation w. In that case, the first order condition:

$$\left.\frac{\partial S(v, \beta(w))}{\partial w} + \frac{\partial R(v, \beta(w))}{\partial w}\right|_{w=v} = 0$$

reduces to:

$$(v - \beta(v))\beta'(v) + \alpha\left[(1-v)\beta'(v) - \beta(v)\right] = 0,$$

which can be rearranged to:

$$\beta'(w) = \frac{\beta(w)(1+\alpha) - v}{\alpha}\frac{1}{1-v}.$$

One can guess that there is a linear solution in which $\beta'(v) = C$, and indeed there is, with $C = \frac{1}{1+2\alpha}$, which is precisely (3.7). Our equilibrium bidding function follows this differential equation until it hits v, at which point bidders are simply required to bid their values. I will show in Lemma 3.1 that this differential equation has a natural generalization to the framework of Section 3.2, and that this formula must be satisfied whenever $\beta = \rho$ and β is differentiable on some neighborhood.

It is worth pointing out some nice features of this equilibrium. Given the explicit solution for β in (3.7), it is easy to see that as $\alpha \to 0$, $\beta(v) \to v$. In other words, the distortion created by revenue sharing is continuous in the amount shared, and as this amount becomes small, bids converge to values. Moreover, for each α, bidders report the correct reserve price conditional on them losing with respect to the equilibrium bid distribution, which is $\frac{1}{2(1+\alpha)}$. Hence, as $\alpha \to 0$, the bid distribution converges to the value distribution, and the reserve price converges to $\frac{1}{2}$, and hence the seller is able to get close to revenue in the second-price auction with the optimally chosen reserve prices.

Finally, let us consider a slight variation of the linear example. Instead of uniform on $[0, 1]$, take the distribution of values to be uniform on $[\gamma - 1, \gamma]$, with

$\gamma \geq 1$. The analogous differential equation is:

$$\beta'(w) = \frac{\beta(w)(1+\alpha) - v}{\alpha} \frac{1}{\gamma - v},$$

which has the solution:

$$\beta(v) = \frac{\gamma\alpha + v(1+\alpha)}{(1+\alpha)(1+2\alpha)}.$$

For γ sufficiently large, $\beta(\gamma - 1) < \gamma - 1$, so that the bidding function never leaves the regime with the binding pricing constraint. Moreover, as $\gamma \to \infty$, the ratio:

$$\frac{\gamma - 1}{\beta(\gamma - 1)} \to 1 + \alpha,$$

so that in the limit, $\beta(v) \approx \frac{v}{1+\alpha}$. This result is intuitive: as γ becomes large, bids become large as well, but because values are compressed into the relatively small region $[\gamma - 1, \gamma]$, the bidders have to shade a large amount in order to obtain a large enough marginal surplus from winning to offset the loss in marginal revenue. Nonetheless, the marginal surplus from winning in equilibrium is $v - \beta(v)$ and the marginal revenue from selling is $\alpha(\gamma - v - \beta(v)) \geq -\alpha\beta(v)$, so shading obeys the proportional bound of $\beta(v) \geq \frac{v}{1+\alpha}$.

3.4 A general symmetric equilibrium

In this section, I will construct an equilibrium analogous to that of Section 3.3 for the general model of Section 3.2. In Section 3.4.1, I will investigate two necessary conditions of regular equilibrium, namely a boundary condition for the bid made by the highest valuation buyer, and a differential equation that must be satisfied when bids are equal to suggested prices. With these necessary conditions in hand, Section 3.4.2 describes an algorithmic construction of a bidding and pricing function. Section 3.4.3 gives a rich example that showcases features of the construction not appearing in the example of the previous section. Section 3.4.4 contains a summary of the proof that this strategy profile is indeed a regular equilibrium. Section 3.4.5 explores the revenue properties of the RSA, relative to the second-price auction with an optimal anonymous reserve price. All omitted proofs appear at the end of the chapter.

3.4.1 Necessary conditions for regular equilibrium

To begin the analysis, I will prove that there are two necessary conditions for a regular equilibrium. The first is a generalization of the first-order condition (3.6), and the second is a boundary condition for the bid made by the buyer with valuation $\overline{v}$.

For starters, the choice of $\beta(v)$ effectively pins down $\rho(v)$, and more generally, it pins down the optimal price when bidding b. Conditional on bidder i losing with a bid of b when others are using the regular strategy (β, ρ), it must be that $\beta\left(v_{-i}^{(1)}\right) \geq b$. Each of the remaining bidders is equally likely to have the highest value among $-i$, so bidder j can be taken to be a "representative" high valuation player. Revenue is:

$$\begin{aligned} R(v,b,r) = \\ \alpha \int_{x=\beta^{-1}(b\vee r)}^{\overline{v}} \Bigg[& r\, F_{v_{-ij}^{(1)}|v_i,v_j}\left(\beta^{-1}(r)\middle|\, v, x\right) \\ & + \int_{y=\beta^{-1}(r)}^{x} \beta(y)\, f_{v_{-ij}^{(1)}|v_i,v_j}\left(\beta^{-1}(r)\middle|\, v, x\right) \Bigg] f_{v_j|v_i}(x|v)dx. \end{aligned} \tag{3.9}$$

The interpretation is that bidder i makes a sale if $\beta(v_j) \geq r$ and if $\beta(v_j) \geq b$ (since i has to lose the auction), which is the outer integral. Conditional on a particular realization for v_j, bidder i makes a sale at price r if $\beta(v_{-ij}^{(1)}) \leq r$, and makes a sale at price $\beta(v_{-ij}^{(1)})$ if $\beta(v_{-ij}^{(1)}) > r$. Since the lower limit for the first integral is $\beta^{-1}(b \vee r)$, it is never optimal to set a price less than b. Let us write:

$$r^*(v, b) = \arg\max_{r \geq b} R(v, b, r).$$

Then an equilibrium condition is that $\rho(v) \in r^*(v, \beta(v))$. Note that continuity of β and f imply that $r^*(v, b)$ is non-empty and compact for all b, and upper-hemicontinuous in b. In general, the price $\rho(v)$ will fall into one of two cases: either there is an interior maximum of $R(v, \beta(v), r)$ for $r \geq b$, in which case $\rho(v)$ is locally constant in v, or the maximizer is $r = \beta(v)$. When this second case obtains, and if $r^*(v, b) = \{b\}$, it will be the case that $\rho(v) = \beta(v)$ for a neighborhood around $[b, b + \epsilon)$. In the following, I build in the fact that bidders would only choose $r \geq b$, in equilibrium or otherwise, and simply write $R(v, r)$.

Now consider the surplus that a bidder receives being allocated the good when bidding b. Note that this does not depend on the price that the bidder suggests:

$$\begin{aligned} S(v, b) = & \\ & \int_{x=\underline{v}}^{\beta^{-1}(b)} \mathbb{I}_{b \geq \rho(x)} \Bigg[(v - \rho(x)) F_{v_{-ij}^{(1)}|v_i, v_j}(\beta^{-1}(\rho(x))|v, x) \\ & \qquad + \int_{y=\beta^{-1}(\rho(x))}^{\beta^{-1}(b)} (v - \beta(y)) f_{v_{-ij}^{(1)}|v_i, v_j}(y|v, x) dy \Bigg] f_{v_j|v_i}(x|v) dx. \end{aligned} \tag{3.10}$$

Here, I am using j as the index of the representative consulted bidder amongst $-i$, when bidder i wins the auction. A winning bidder i's total surplus is:

$$U(v, b, r) = S(v, b) + R(v, r).$$

Suppose there is a neighborhood $(v-\epsilon, v+\epsilon)$ of v on which $r^*(w, \beta(w)) = \{\beta(w)\}$, so that $\rho(w) = \beta(w)$, and β and ρ are differentiable at v. Note that deviations to b near $\beta(v)$ can equivalently be thought of as deviations to $\beta(w)$ for nearby w, due to the continuity of β. Thus, the bidding function β must satisfy a first-order condition:

$$\left.\frac{dU(v, \beta(w), \beta(w))}{dw}\right|_{w=v} = \left.\frac{\partial S(v, \beta(w))}{\partial w}\right|_{w=v} + \left.\frac{\partial R(v, \beta(w))}{\partial w}\right|_{w=v} = 0. \tag{3.11}$$

This first-order condition can be translated into an equilibrium condition on β:

Lemma 3.1. *Suppose that there is a neighborhood $(v - \epsilon, v + \epsilon)$ of v on which $r^*(w, \beta(w)) = \{\beta(w)\}$, and β is differentiable at v. Then:*

$$\beta'(v) = \frac{\beta(v)\,(1 + \widehat{\alpha}) - v}{\widehat{\alpha}} g(v, v|v), \tag{FOC}$$

where $\widehat{\alpha} = \frac{\alpha}{n-1}$.

Proof of Lemma 3.1. Using the definition of S, and the fact that $\rho(\beta^{-1}(b)) = b$ for b in $(\beta(v - \epsilon), \beta(v + \epsilon))$, marginal surplus can be rewritten as:

$$\frac{\partial S(v, \beta(w))}{\partial w} = (v - \beta(w)) F_{v_{-ij}^{(1)}|v_i,v_j}(w|v, w) f_{v_j|v_i}(w|v) + (v - \beta(w)) \int_{x=\underline{v}}^{w} f_{v_{-ij}^{(1)}|v_i,v_j}(w|v, x) f_{v_j|v_i}(x|v) dx.$$

By symmetry, it must be that:

$$\int_{x=\underline{v}}^{w} f_{v_{-ij}^{(1)}|v_i,v_j}(w|v,x) f_{v_j|v_i}(x|v) dx = (n-2) F_{v_{-ij}^{(1)}|v_i,v_j}(w|v,w) f_{v_j|v_i}(w|v),$$

so:

$$\frac{\partial S(v,\beta(w))}{\partial w} = (n-1)(v-\beta(w)) F_{v_{-ij}^{(1)}|v_i,v_j}(w|v,w) f_{v_j|v_i}(w|v).$$

The interpretation is that since $\beta = \rho$ around v, the marginal allocation event affected by bid $\beta(w)$ is when $v_{-i}^{(1)} = w$, which is the probability that one of the remaining bidders has a valuation of w and the other bidders have valuations less than w. Additionally, there are $n-1$ choices for the bidder with valuation exactly w.

The second term in (3.11) is:

$$\frac{\partial R(v,\beta(w))}{\partial w} = \alpha \Bigg[\left(F_{v_j,v_{-ij}^{(1)}|v_i}(\overline{v},w|v) - F_{v_j,v_{-ij}^{(1)}|v_i}(w,w|v) \right) \beta'(w) \\ - \beta(w) F_{v_{-ij}^{(1)}|v_i,v_j}(w|v,w) f_{v_j|v_i}(w|v) \Bigg],$$

with the other terms canceling or dropping out. Combining results, the marginal payoff is:

$$\frac{\partial U(v,\beta(w),\beta(w))}{\partial w} = (n-1)\left(v - \beta(w)\left(1+\frac{\alpha}{n-1}\right)\right) F_{v_{-ij}^{(1)}|v_i,v_j}(w|v,w) f_{v_j|v_i}(w|v) \\ - \alpha \left(F_{v_j,v_{-ij}^{(1)}|v_i}(\overline{v},w|v) - F_{v_j,v_{-ij}^{(1)}|v_i}(w,w|v) \right) \beta'(w).$$

Now, evaluating at $w = v$ and rearranging yields equilibrium condition:

$$\beta'(v) = \frac{\beta(v)\,(1+\widehat{\alpha}) - v}{\widehat{\alpha}} \frac{F_{v_{-ij}^{(1)}|v_i,v_j}(v|v,v) f_{v_j|v_i}(v|v)}{F_{v_j,v_{-ij}^{(1)}|v_i}(\overline{v},v|v) - F_{v_j,v_{-ij}^{(1)}|v_i}(v,v|v)},$$

where $\widehat{\alpha} = \frac{\alpha}{n-1}$. □

There might be a concern that $g(v, v|v)$ can blow up as $v \to \overline{v}$. However, one can prove directly that a solution to (FOC) subject to the boundary condition $\beta(\overline{v}) = \frac{\overline{v}}{1+\widehat{\alpha}}$ exists. If $\int_{x=\underline{v}}^{v} g(x, x|x)dx$ diverges as $v \to \overline{v}$, the solution is:

$$\begin{aligned} \beta(v) = \frac{1}{\widehat{\alpha}} \exp\left(\frac{1+\widehat{\alpha}}{\widehat{\alpha}} \int_{x=\underline{v}}^{v} g(x,x|x)dx\right) \\ \cdot \int_{x=v}^{\overline{v}} \exp\left(-\frac{1+\widehat{\alpha}}{\widehat{\alpha}} \int_{y=\underline{v}}^{x} g(y,y|y)dy\right) x g(x,x|x)dx. \end{aligned} \tag{3.12}$$

Observe, the quantity:

$$\int_{x=v}^{w} \exp\left(-\frac{1+\widehat{\alpha}}{\widehat{\alpha}} \int_{y=\underline{v}}^{x} g(y,y|y)dy\right) x g(x,x|x)dx$$

can be integrated by parts to give:

$$\begin{aligned} -\frac{\widehat{\alpha}}{1+\widehat{\alpha}} \exp\left(-\frac{1+\widehat{\alpha}}{\widehat{\alpha}} \int_{y=\underline{v}}^{x} g(y,y|y)dy\right) x\Bigg]_{x=v}^{w} \\ + \int_{x=v}^{w} \frac{\widehat{\alpha}}{1+\widehat{\alpha}} \exp\left(-\frac{1+\widehat{\alpha}}{\widehat{\alpha}} \int_{y=\underline{v}}^{x} g(y,y|y)dy\right) dx, \end{aligned}$$

which converges to:

$$\begin{aligned} \frac{\widehat{\alpha}}{1+\widehat{\alpha}} \exp\left(-\frac{1+\widehat{\alpha}}{\widehat{\alpha}} \int_{y=\underline{v}}^{v} g(y,y|y)dy\right) v \\ + \int_{x=v}^{\overline{v}} \frac{\widehat{\alpha}}{1+\widehat{\alpha}} \exp\left(-\frac{1+\widehat{\alpha}}{\widehat{\alpha}} \int_{y=\underline{v}}^{x} g(y,y|y)dy\right) dx \end{aligned}$$

as $w \to \overline{v}$. Then taking $v \to \overline{v}$, this expression must converge to zero. We can then apply L'Hôpital's rule to (3.12) to find that:

$$\lim_{v\to\overline{v}} \beta(v) = \lim_{v\to\overline{v}} \frac{1}{\widehat{\alpha}} \frac{-\exp\left(-\frac{1+\widehat{\alpha}}{\widehat{\alpha}} \int_{x=\underline{v}}^{v} g(x,x|x)dx\right) vg(v,v|v)}{-\frac{1+\widehat{\alpha}}{\widehat{\alpha}} g(v,v|v) \exp\left(-\frac{1+\widehat{\alpha}}{\widehat{\alpha}} \int_{x=\underline{v}}^{v} g(x,x|x)dx\right)} = \lim_{v\to\overline{v}} \frac{v}{1+\widehat{\alpha}}.$$

If $\int_{x=\underline{v}}^{\overline{v}} g(x,x|x)dx$ converges, then a term $C\int_{x=\underline{v}}^{\overline{v}} g(x,x|x)dx$ can be added so that the boundary condition obtains.

Now, consider the bidder with valuation $\overline{v}$. In a regular equilibrium, this type must make the largest bid $\beta(\overline{v})$. According to the rules described above, the bidder always wins and hence sells to other bidders with probability 0. In order for this to be incentive compatible, it must be that the type $\overline{v}$ does not have an incentive to shade, and start selling to bidders of lower valuation. At the margin, the surplus lost from not receiving the good when others bid $\beta(\overline{v})$ is $(v - \beta(\overline{v}))f_{v_{-i}^{(1)}|v_i}(\overline{v}|\overline{v})$, i.e. when one of the other bidders has a valuation of $\overline{v}$ conditional on $v_i = \overline{v}$. On the other hand, the revenue gained from selling to such a type is precisely $\widehat{\alpha}\beta(\overline{v})f_{v_{-i}^{(1)}|v_i}(\overline{v}|\overline{v})$. Hence, for shading not to be attractive for the highest type, it must be that $\beta(\overline{v})$ is less than $\frac{\overline{v}}{1+\widehat{\alpha}}$. In fact, if this were a strict inequality, then there is some type with value $v_i \in (\overline{v}-\epsilon, \overline{v}]$ who would prefer to shade less. This informal argument suggests the following Lemma, whose proof is at the end of the chapter.

Lemma 3.2. *In any regular equilibrium, it must be that:*

$$\beta(\overline{v}) = \frac{\overline{v}}{1+\widehat{\alpha}}, \tag{3.13}$$

where $\widehat{\alpha} = \frac{\alpha}{n-1}$.

3.4.2 A constructive algorithm

I will now construct a regular equilibrium of the RSA. The equilibrium consists of a partition of the interval of valuations $[\underline{v}, \overline{v}]$ into a sequence of intervals with endpoints:

$$\underline{v} = \underline{w}^K \leq \overline{w}^K \leq \cdots \leq \underline{w}^0 \leq \overline{w}^0 = \overline{v}.$$

The partition, and the bidding and pricing functions β and ρ, will be defined inductively on the regions:

$$\begin{aligned} \overline{W}^k &= (\underline{w}^k, \overline{w}^k] \\ \underline{W}^k &= (\overline{w}^{k+1}, \underline{w}^k]. \end{aligned} \tag{3.14}$$

On regions $\overline{W}^k$, I set $\beta(v) = \rho(v)$ where β solves (FOC) with the initial condition $\beta(\overline{w}^0) = \frac{\overline{v}}{1+\hat{\alpha}}$ and $\beta(\overline{w}^k) = \overline{w}^k$ for $k > 1$. On regions of the form $(\overline{w}^{k-1}, \underline{w}^k]$, I set $\beta(v) = v$ and $\rho(v) = r^*(v) = \inf r^*(v, v) \leq \underline{w}^k$.

In particular, let $\beta_k(v)$ be the solution to (FOC) on $[\underline{v}, \overline{w}^k]$ with the boundary condition:

$$\beta_k(\overline{w}^k) = \begin{cases} \overline{w}^k & \text{if } k > 0 \\ \frac{\overline{v}}{1+\hat{\alpha}} & \text{if } k = 0 \end{cases}. \tag{BC}$$

Define:

$$\underline{w}^k = \sup\left(\{\underline{v}\} \cup \{v < \overline{w}^k | \beta_k(v) > v\}\right). \tag{3.15}$$

Note that this definition implies that if $\beta_k(v) \leq v$ for all v, then $\underline{w}^k = \underline{v}$. I define $\beta(v) = \beta_k(v)$ for all $v \in [\underline{w}^k, \overline{w}^k]$.

If $\underline{w}^k > \underline{v}$, let:

$$\overline{w}^{k+1} = \sup\left(\{\underline{v}\} \cup \left\{v < \underline{w}^k \,\middle|\, R(v,v) > R(v,w)\ \forall w \in (v, \underline{w}^k]\right\}\right). \tag{3.16}$$

If $\underline{w}^k \in r^*(v,v)$ for all $v < \underline{w}^k$, then set $\overline{w}^{k+1} = \underline{v}$. Otherwise, $\beta(v) = v$ and $\rho(v) = r^*(v)$ for $v \in \underline{W}^k$.

The construction starts with $\overline{w}^0 = \overline{v}$, and continues inductively alternating between defining new $\underline{w}^k$ and $\overline{w}^{k+1}$. The algorithm terminates when the next of these two suprema are $\underline{v}$. This is formalized in Algorithm 3.1, and Proposition 3.1 provides a characterization of the algorithm.

Algorithm 3.1: Constructing a regular equilibrium

---

initialize $k = 0$, $\overline{w}^0 = \overline{v}$.
initialize $\beta(v) = \rho(v) = \beta_0(v)$, which solves (FOC) and (BC)
while true
    define $\underline{w}^k$ according to (3.15).
    redefine $\beta(v) = \rho(v) = \beta_k(v)$ for $v \in [0, \underline{w}^k]$,
            where $\beta_k(v)$ solves (FOC) and (BC).
    if $\underline{w}^k = \underline{v}$,
        break.
    define $\overline{w}^{k+1}$ according to (3.15).
    redefine $\beta(v) = v$ and $\rho(v) = r^*(v)$ for $v \in [0, \underline{w}^k]$.
    if $\underline{w}^k = \underline{v}$,
        break.
    redefine k=k+1.
end while

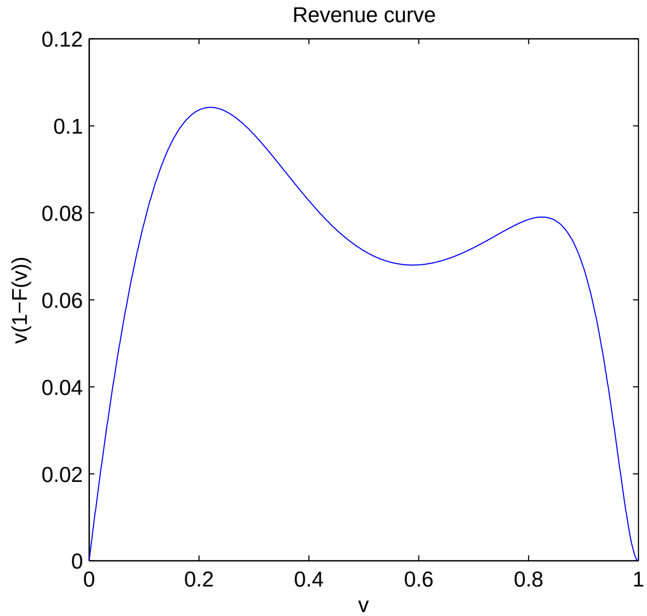

Figure 3.2: The revenue curve for the irregular example of Section 3.4.3. Note that the profit function has two peaks.

Proposition 3.1. *The inductive construction of Algorithm 3.1 terminates after finitely many steps. It defines a continuous and strictly increasing bidding function.*

Hence, the algorithm defines a bidding function. It is easy to see that this bidding function is continuous, since it either solves the differential equation (FOC) or is $\beta(v) = v$, and I have defined β at boundary points so that it is continuous. At this point, it is not known that β is strictly increasing, but this will follow from Lemma 3.3 below.

3.4.3 A more complicated example

Before showing that Algorithm 3.1 defines an equilibrium, let us look at an example that showcases the richness of the construction. The uniform example was relatively simple because Algorithm 3.1 converged after just two steps, which is

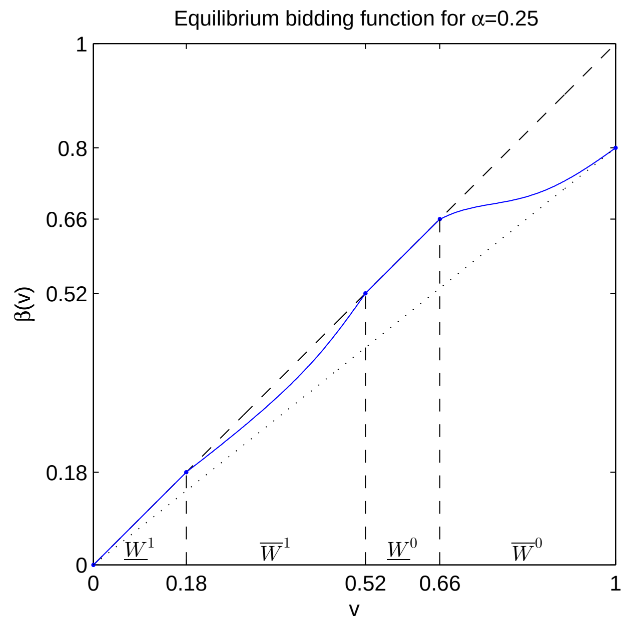

Figure 3.3: The equilibrium bid distribution for the example of Section 3.4.3. Algorithm 3.1 takes four steps to converge, with $\underline{w}^0 \approx 0.66$, $\overline{w}^1 \approx 0.52$, $\underline{w}^1 \approx 0.18$, and $\overline{w}^2 = 0$. The dotted line is $\frac{v}{1+\widehat{\alpha}}$.

a consequence of the fact that the value distribution has monotonic virtual valuation, i.e., is regular in the sense of Myerson (1981). It also turned out that the bidding function on $\overline{W}^0$ had a simple linear form. Here I present a more complicated example involving two bidders in which values are still independent, but the independent distribution is highly irregular. The cumulative distribution of values is a weighted sum of Beta distributions, and in particular each bidder's valuation is distributed $B[\alpha = 1.5, \beta = 5.5]$ with probability 0.9 and is distributed $B[\alpha = 25, \beta = 2]$ with probability 0.1. The revenue sharing parameter is $\alpha = \frac{1}{4}$. This cumulative distribution results in the revenue curve $v(1 - F(v))$ depicted in Figure 3.2.

The equilibrium bidding function is depicted in Figure 3.3. Note that the solid line, $\beta(v)$, is everywhere above the dotted line, which is $\frac{v}{1+\alpha}$. The algorithm takes four regime changes to converge. It starts with $\overline{w}^0 = 1$ and $\beta(\overline{w}^0) = \frac{1}{1+\alpha} = 0.8$. Initially, β solves the differential equation (FOC) starting at $v = 1$ and going downwards, until $\beta(v)$ hits v at $\underline{w}^0 \approx 0.66$. At this point, the regime switches to $\beta(v) = v$ and $r^*(v) = 0.66$, until $\overline{w}^1 \approx 0.52$. At this point, r^* jumps down to 0.52, and the regime switches back to solving (FOC) with the boundary condition $\beta(\overline{w}^1) = \overline{w}^1$. The bidding function again hits v at $\underline{w}^1 \approx 0.18$, and the regime switches back to the $\beta(v) = v$ with $r^*(v) = 0.18$, until v hits zero.

In general, the algorithm could require many regime changes before reaching zero, although the number of regime changes is bounded above by two times the number of zeros of $g(v, v|v) - \frac{1}{v}$, as shown in the proof of Proposition 3.1.

3.4.4 The algorithm defines an equilibrium

My main result is the following:

Theorem 3.1. *The bidding and pricing functions (β, ρ) defined by Algorithm 3.1 constitute a regular equilibrium of the revenue-sharing second-price auction.*

I will provide a general overview of the proof. To start, observe that (FOC) is zero when $\beta(v) = \frac{v}{1+\hat{\alpha}}$. Hence, it is impossible for β to fall below this level, and indeed it is impossible for it to remain at this level for an open interval. This means that the bidding function is always strictly increasing, so if it is an equilibrium, it will be regular. This is formalized in Lemma 3.3.

Lemma 3.3. $\frac{v}{1+\hat{\alpha}} \leq \beta(v) \leq v$.

An observation which greatly simplifies the proof is that there is a relatively small number of deviations which need to be checked. In particular, a deviation to (r, r) with $r < v$ dominates all deviations of the form (b, r) with $b < r < v$.

The reason is that for $b < v$, $S(v,b)$ is weakly increasing, so it is without loss of generality to take b as large as possible subject to the pricing constraint. On the other side, only deviations of the form (v,r) where $r > v$ need to be considered. The reason is the same: $S(v,b)$ is weakly decreasing when $b > v$.

The next Lemma will help rule out some downward deviations. Recall that (FOC) defines the bidding function on regions $\overline{W}^k$. The first-order condition was obtained by differentiating $U(v,\beta(w),\beta(w))$ with respect to v and setting it equal to zero for $w = v$. However, substituting the definition of $\beta'(w)$ into the derivative of U yields the expression:

$$\frac{\partial U(v,\beta(w),\beta(w))}{\partial w} = C(v,w) \cdot \Big[(\beta(w)(1+\widehat{\alpha}) - w)g(w,w|w) - (\beta(w)(1+\widehat{\alpha}) - v)g(w,w|v)\Big],$$

where $C(v,w)$ is some strictly positive number that depends on v and w. But because of Lemma 3.3, the term multiplying $g(w,w|w)$ is always non-negative, and also $\beta(w)(1+\widehat{\alpha})-v$ is greater (less) than $\beta(w)(1+\widehat{\alpha})-w$ if v is less (greater) than w. Combined with the fact that $g(w,w|v)$ is monotonically decreasing, these observations imply that the bidder's deviation payoff at a deviation of the form (r,r) is increasing if $b < \beta(v)$ and decreasing if $b > \beta(v)$.

Lemma 3.4. *For all w on the interior of $\overline{W}^k$ for some k:*

$$\frac{dU(v,\beta(w),\beta(w))}{dw} \begin{cases} \leq 0 & \text{if } w > v \\ \geq 0 & \text{if } w < v \end{cases}.$$

This Lemma tells us that a bidder's payoff is always decreasing as a deviation of the form (r, r) moves away from $(\beta(v), \rho(v))$, when the deviation bid is in a region on which the bidding function solves (FOC).

But to rule out large deviations, it must be that deviation payoffs are decreasing when crossing regions of the form $\underline{W}^k$, and also when deviating to (v, r) with $r > v$. This is facilitated by the following Lemma 3.5.

Lemma 3.5.

1. *For* $r \in \underline{W}^k$, $\frac{\partial R(v,r)}{\partial r}$ *is increasing in* v. *As a result, if* $r \geq r'$ *and* $v \geq v'$, *then:*

$$v < \overline{w}^{k+1} \implies R(v, r) \leq R(v, \overline{w}^{k+1})$$
$$v > \overline{w}^{k+1} \implies R(v, r) \leq R(v, r^*(v) \wedge \underline{w}^k).$$

2. *For* $w \in \overline{W}^k$ *and* $w \geq v$, $\frac{dR(v,\beta(w))}{dw} \leq 0$.

The Lemma makes two assertions. The first concerns regions of the form $\underline{W}^k$, and asserts that if $v < \overline{w}^{k+1}$, then $\overline{w}^{k+1}$ generates greater expected revenue than any price $r \in \underline{W}^k$. Note that the result holds trivially when $\overline{w}^{k+1} = \underline{v}$. On the other hand, if $v > \overline{w}^{k+1}$, then either (1) $v \in \underline{W}^k$, and $r^*(v)$ is better than any price in $\underline{W}^k$, or (2) $v \notin \underline{W}^k$ and $\underline{w}^k$ is a better price than any $r \in \underline{w}^k$. Note that this is trivially satisfied when $\overline{w}^k = \overline{v}$. More generally, the result is a consequence of the positive correlation assumption **A1**, which is that higher types are more optimistic about the distribution of others' values. As a result, higher valuations always want to set higher reserve prices for other bidders.

The second part of Lemma 3.5 concerns the sign of marginal revenue on regions $\overline{W}^k$ with $v < \underline{w}^k$. By substituting in the formula for $\beta'(w)$, the derivative of U is:

$$\frac{\partial U(v, \beta(w))}{\partial w} = C(v, w) \cdot \left[\frac{\beta(w) - w}{\widehat{\alpha}} g(w, w|w) + \beta(w)(g(w, w|w) - g(w, w|v)) \right],$$

where $C(v, w)$ is strictly positive. Since $\beta(w) \leq w$, and $g(w, w|w) < g(w, w|v)$, again it is the case that marginal revenue is non-positive on such regions.

The results of Lemmas 3.4 and 3.5 facilitate an inductive argument that the equilibrium strategy is optimal. For simplicity, let us consider $v \in \overline{W}^k$ for some k. The case when $v \in \underline{W}^k$ is not substantially different. The quasiconcavity of $U(v, \beta(w), \beta(w))$ means that there are no deviations of the form (r, r) that are optimal with $r \in \beta(\overline{W}^k)$. In particular, this means that $\underline{w}^k$ is not a profitable deviation. But then Lemma 3.5 implies that there is no profitable deviation on $[\overline{w}^{k+1}, \underline{w}^k] = \underline{W}^k$, since $R(v, \underline{w}^k)$ is greater than $R(v, r)$ for $r \in \underline{W}^k$ and $S(v, b)$ is weakly decreasing when $b < v$. Hence, a deviation to $(\overline{w}^{k+1}, \overline{w}^{k+1})$ is not profitable. But now the quasiconcavity kicks in again on $\overline{W}^{k+1}$, and there are no profitable deviations here either. This induction continues, and so that there are no profitable downward deviations.

With regard to upward deviations, it has already been shown that (v, v) is not a profitable deviation, since either (1) $v > \beta(\overline{v})$, in which case this is obvious, or (2) since $\beta(\overline{w}^k) \geq v$, (v, v) is a downward deviation in $\overline{W}^k$, which is not profitable because of Lemma 3.4. But now part 2 of Lemma 3.5 can be used to show that the payoff at (v, v) is weakly greater than the profit at (v, r) for all $r > v$. On regions $\overline{W}^k$, marginal revenue is non-positive because of Lemma 3.5, so $(v, \beta(\overline{w}^k))$ is not profitable. If $k > 0$, the first part of Lemma 3.5 shows that $(v, \overline{w}^k)$ is better than any deviation (v, r) with $r \in \underline{W}^k$. The induction continues, showing that no upward deviation is profitable. This concludes the proof sketch that (β, ρ) constitute a regular equilibrium of the RSA.

3.4.5 Equilibrium net revenue

Let us now turn our attention to revenue properties of the RSA, specifically with an interest in comparative statics as $\alpha \to 0$. My basis for comparison is revenue from the second-price auction if the seller knew the distribution of values and was able to set the optimal anonymous reserve price. Formally, define:

$$\pi^* = \max_{r \geq \underline{v}} \mathbb{E}\left[r\, \mathbb{I}_{v^{(2)} \leq r \leq v^{(1)}} + v^{(2)} \mathbb{I}_{r \leq v^{(2)}}\right],$$

and define r^* to be a revenue maximizing r, which is an optimal anonymous reserve price. Gross revenue from the RSA is:

$$\pi^G = \int_{v=\underline{v}}^{\overline{v}} \mathbb{E}\left[\rho(v)\, \mathbb{I}_{\beta\left(v_{-i}^{(2)}\right) \leq \rho(v) \leq \beta\left(v_{-i}^{(1)}\right)} + v_{-i}^{(2)} \mathbb{I}_{\rho(v) \leq \beta\left(v_{-i}^{(2)}\right)} \middle| v_{-i}^{(1)} \geq v\right] \cdot \left(1 - F_{v_{-i}^{(1)}|v_i}(v|v)\right) f_{v_i}(v) dv,$$

and net revenue from the RSA is $\pi = (1-\alpha)\pi^G$, since an α share of revenue is awarded to the bidder who suggests the reserve price.

I will prove the following result:

Proposition 3.2. *For any $\alpha > 0$, net revenue π from the RSA under the equilibrium defined by Algorithm 3.1 is at least $\pi^* \frac{1-\alpha}{1+\alpha}$. Hence, as $\alpha \to 0$, revenue converges to a limit weakly greater than π^*.*

Proof of Proposition 3.2. Because of Lemma 3.3, we know that $\beta(v) \geq \frac{v}{1+\widehat{\alpha}}$. This implies that:

$$\begin{aligned}
&\mathbb{E}\left[\frac{r^*}{1+\widehat{\alpha}}\,\mathbb{I}_{\beta\left(v^{(2)}\right)\leq\frac{r^*}{1+\widehat{\alpha}}\leq\beta\left(v^{(1)}\right)}+\beta\left(v^{(2)}\right)\mathbb{I}_{\frac{r^*}{1+\widehat{\alpha}}\leq\beta\left(v^{(2)}\right)}\right]\\
&\geq\mathbb{E}\left[\frac{r^*}{1+\widehat{\alpha}}\,\mathbb{I}_{v^{(2)}\leq r^*\leq v^{(1)}}+\frac{v^{(2)}}{1+\widehat{\alpha}}\,\mathbb{I}_{r^*\leq v^{(2)}}\right]\\
&=\frac{\pi^*}{1+\widehat{\alpha}}.
\end{aligned}$$

Hence, if the seller were to use the anonymous reserve price $\frac{r^*}{1+\widehat{\alpha}}$ with the equilibrium bid distribution induced by β, gross revenue would be at least $\frac{r^*}{1+\widehat{\alpha}}$.

In fact, the seller does not set the reserve price $\frac{r^*}{1+\widehat{\alpha}}$, but rather the reserve price $\rho(v)$ of a losing bidder with valuation v. However, each such bidder is setting an optimal reserve price conditional on $v^{(1)} \geq v$, and therefore is setting a reserve price which generates weakly greater expected revenue conditional on this event. Formally, gross revenue when using a particular bidder's recommendation is:

$$\max_{r\geq\underline{v}}\mathbb{E}\left[r\,\mathbb{I}_{\beta\left(v_{-i}^{(2)}\right)\leq r\leq\beta\left(v_{-i}^{(1)}\right)}+v_{-i}^{(2)}\mathbb{I}_{r\leq\beta\left(v_{-i}^{(2)}\right)}\,\middle|\,v_{-i}^{(1)}\geq v\right].$$

Hence, gross revenue is:

$$\begin{aligned}
&\int_{v=\underline{v}}^{\overline{v}}\max_{r\geq\underline{v}}\mathbb{E}\left[r\,\mathbb{I}_{\beta\left(v_{-i}^{(2)}\right)\leq r\leq\beta\left(v_{-i}^{(1)}\right)}+v_{-i}^{(2)}\mathbb{I}_{r\leq\beta\left(v_{-i}^{(2)}\right)}\,\middle|\,v_{-i}^{(1)}\geq v\right]\left(1-F_{v_{-i}^{(1)}|v_i}(v|v)\right)f_{v_i}(v)dv\\
&=\int_{v=\underline{v}}^{\overline{v}}\max_{r\geq\underline{v}}\mathbb{E}\left[r\,\mathbb{I}_{\beta\left(v^{(2)}\right)\leq r\leq\beta\left(v^{(1)}\right)}+v^{(2)}\mathbb{I}_{r\leq\beta\left(v^{(2)}\right)}\,\middle|\,v_{-i}^{(1)}\geq v\right]\left(1-F_{v_{-i}^{(1)}|v_i}(v|v)\right)f_{v_i}(v)dv\\
&\geq\max_{r\geq\underline{v}}\int_{v=\underline{v}}^{\overline{v}}\mathbb{E}\left[r\,\mathbb{I}_{\beta\left(v^{(2)}\right)\leq r\leq\beta\left(v^{(1)}\right)}+v^{(2)}\mathbb{I}_{r\leq\beta\left(v^{(2)}\right)}\,\middle|\,v_{-i}^{(1)}\geq v\right]\left(1-F_{v_{-i}^{(1)}|v_i}(v|v)\right)f_{v_i}(v)dv\\
&=\max_{r\geq\underline{v}}\mathbb{E}\left[r\,\mathbb{I}_{\beta\left(v^{(2)}\right)\leq r\leq\beta\left(v^{(1)}\right)}+v^{(2)}\mathbb{I}_{r\leq\beta\left(v^{(2)}\right)}\right].
\end{aligned}$$

The second line comes from the fact that $r \geq \beta(v)$, $v_{-i}^{(2)} \geq v^{(2)}$, and $v_{-i}^{(1)} = v^{(1)}$. The third line comes from the integral of the maximum being greater than the maximum of the integral. The final line is just the law of iterated expectations. The last line is at least $\frac{\pi^*}{1+\widehat{\alpha}}$, so gross revenue under the RSA is at least this quantity as well.

However, the seller is also making payments to the agents of α times realized revenue. Hence, net revenue is $1 - \alpha$ times gross revenue. □

Thus, the loss from revenue sharing becomes small as $\alpha \to 0$. This result is intuitive: for any α, bidders suggest optimal reserve prices conditional on them losing the auction, with respect to the equilibrium bid distribution. Since these prices are optimal conditional on more information than the prior, namely the realization of the loser's value v and the fact that $v_{-i}^{(1)} \geq v$, revenue generated with such prices is at least the revenue with an optimal ex-ante reserve prices. Moreover, because of Lemma 3.3, as $\alpha \to 0$ the equilibrium bid distribution converges weakly to the distribution of values, and since expected revenue is weakly continuous in the distribution of values, gross revenue converges to a quantity weakly greater than π^*. Lastly, for small α, the revenue lost from sharing is small relative to gross revenue.

In light of Proposition 3.2, it is fair to say that the RSA accomplishes the goal described in the introduction, which is to approximate revenue from the second-price auction with an optimal anonymous reserve, even in situations where the seller does not know the distribution of values but the buyers do.

3.5 Discussion

3.5.1 Asymmetric distributions

Throughout the analysis, I have restricted attention to symmetric case. The extension to asymmetric distribution involves a somewhat more complicated construction than Algorithm 3.1, and some new conceptual challenges. In the symmetric case, bidders all used the same bidding function and hence at any valuation v, all bidders were either in the regime determined by the first-order condition (FOC) or were bidding their values. With asymmetric bidders and asymmetric bidding functions, the bidders' regimes need not coincide, and indeed I must make allowance for one bidder to be following the asymmetric version of (FOC) and the other bidder to bid his value.

To illustrate, let us consider a two bidder example in which each bidder i's value is drawn independently from the distribution with cumulative distribution F_i. I assume that both F_i have the same support $[\underline{v}, \overline{v}]$ and both admit strictly positive and continuous densities f_i. I will first derive the asymmetric analog of (FOC). To that end, it is useful to define the functions:

$$z_i(b) = \beta_i^{-1}(b),$$

which are the inverse bid functions. In the asymmetric case, I will solve directly for the inverse functions, and then invert them to obtain bidding functions.

In that case, surplus and revenue can be written:

$$S_i(v,b) = \int_{x=\beta(\underline{v})}^{b} (v - \rho_j(z_j(x))) f_j(z_j(x)) z_j'(x) dx,$$
$$R_i(v,r) = \alpha\, r \left(1 - F_j\left(z_j(r)\right)\right).$$

Hence, the condition that bidder i's marginal surplus plus marginal revenue equal zero reduces to:

$$(v-b) f_j(z_j(b)) z_j'(b) + \alpha \left[1 - F_j(z_j(b)) - b\, f_j(z_j(b)) z_j'(b)\right] = 0,$$

which evaluated at $v = z_i(b)$ can be rewritten as:

$$z_j'(b) = \frac{\alpha}{b(1+\alpha) - z_i(b)} \frac{1 - F_j(z_j(b))}{f_j(z_j(b))}. \qquad \text{(FOC}'\text{)}$$

This formula has an important feature missing from the symmetric case: The first-order condition for bidder i to price at his bid is actually a constraint on bidder j's bidding function. The boundary condition is unchanged:

$$z_j\left(\frac{\overline{v}}{1+\alpha}\right) = \overline{v}. \qquad \text{(BC}'\text{)}$$

Our new algorithm again calls for initially solving (FOC′) subject to (BC′). The construction starts with $z_i(b) \geq b$, and the regime switches when some $z_i(b) - b$ hits 0. With symmetric distributions, both z_i would hit b at the same time (if at all). However, we must confront the possibility that $z_i(b) - b$ hits 0 at $\widehat{b}$, but $z_j(b) - b > 0$ for all $b \geq \widehat{b}$.

One idea would be to look for hybrid regimes in which one player's z_i is defined using the first-order condition, and the other player has $z_j(b) = b$. However, this cannot be part of an equilibrium. For suppose that this is the case, say with z_1 solving (FOC$'$) and $z_2(b) = b$. This would imply that bidder 1 is pricing strictly above his bid, and bidder 2 is pricing at his bid on some region. But this requires that $z_1(b) > b$, in which case bidder 1 is shading, even though his pricing constraint is not binding and there is positive probability of bidder 1 setting a price of b' between b and $z_1(b)$. As such, bidder 1 would want to increase his bid, so as to win on these events!

Therefore, a hybrid regime cannot exist when leaving a regime where both pricing constraints bind. But, it still might be the case that the solution of $z_2(b) - b$, say, hits 0 at $\widehat{b}$, while $z_1(\widehat{b}) > \widehat{b}$. What then? If z_1 is to be monotonic, the only option is to have z_1 jump down to $z_1(b) = b$, so that z_1 has a discontinuity. This corresponds to a range of valuations for player 1, between $z_1^+(\widehat{b})$ and $z_1^-(\widehat{b})$, limits from the right and left respectively, who all bid $\widehat{b}$ and set a price of $\widehat{b}$. Intuitively, these types all want to sell to a bidder 2 with value greater than $\widehat{b}$, and bidder 2 bids his value and sets a price of $\widehat{b}$, effectively selling to the mass point.

To illustrate, let us solve a simple asymmetric example. The support of values is $[0, 1]$, and $F_1(x) = x$ and $F_2(x) = x^2$. Hence, bidder 2 is the "high demand" consumer, with expected valuation of $\frac{2}{3}$, whereas bidder 1's expected value is $\frac{1}{2}$. The differential equations are:

$$z_1'(b) = \frac{\alpha}{b(1+\alpha) - z_2(b)}(1 - z_1(b)), \tag{3.17a}$$

$$z_2'(b) = \frac{\alpha}{b(1+\alpha) - z_1(b)} \frac{1 - (z_2(b))^2}{2z_2(b)}. \tag{3.17b}$$

The construction starts with $\widehat{b}^0 = \frac{1}{1+\alpha}$ and $\omega^0 = (B, B)$. It turns out that $z_2(b) - b$ hits 0 first, at around $\widehat{b} \approx 0.4430$, and thereafter set $z_1(b) = z_2(b) = b$. This is an equilibrium, since a price of $\widehat{b}$ dominates all lower bids. This can be seen from the fact that $v(1 - F_1(v))$ is concave with a maximum at $v = \frac{1}{2}$, and $v(1 - F_2(v))$ is concave with a maximum at $v = \frac{1}{\sqrt{3}} \approx 0.5774$.

What if the z_i are in the regime where $z_1(b) = z_2(b) = b$, and then at some $\widehat{b}$ bidder 1's pricing constraint binds, so that he would want to start shading in equilibrium? In order to satisfy bidder 1's indifference while maintaining $z_1(b) = b$, bidder 2 would need to start shading. However, this shading cannot be incentive compatible if bidder 1 is setting prices between b and $z_2(b)$ with positive probability. The solution is to solve (FOC′) with $z_2'(b) = 1$, so that:

$$(b(1 + \alpha) - z_1(b))f_2(b) = \alpha(1 - F_2(b)),$$

until bidder 2's pricing constraint binds, at which point the z_i solve the full system of first-order conditions.

Note that the difference between leaving the both-not-binding regime considered here and leaving the both-binding regime considered above is that the player who continues to bid his value must have a weak incentive to price above his own value. When leaving the both-binding regime, when $z_i(b) - b$ hits 0 at $\widehat{b}$, this means that bidder j now has an incentive to price at $\widehat{b}$, and hence he cannot shade to a bid below $\widehat{b}$. On the other hand, when leaving the both-not-binding regime, there is no problem having one bidder continue to bid his value as long as he prices above his value, while the other bidder starts to shade.

Thus, the general lessons for the two bidder asymmetric case are

1. When leaving a regime with both z_1 and z_2 solving the first-order condition, and when $z_i(b) - b$ hits 0 first at $\widehat{b}$, then $z_j(b)$ jumps down, so that β_j is constant at $\widehat{b}$ until $v = \widehat{b}$.

2. When leaving a regime with $z_1(b) = z_2(b) = b$, and bidder i's pricing constraint binds first, then bidder j continues to have $z_j(b) = b$ while z_i solves (FOC′) with $z_j' = 1$ and $z_j(b) = b$.

Finally, I observe that while this proposed algorithm leads to continuous and weakly increasing bidding functions, they are not strictly increasing because of the discontinuities in z_i. Hence, for asymmetric bidders, the definition of a regular equilibrium would need to be relaxed to allow for weakly increasing bids.

3.5.2 Simpler auctions

In the symmetric independent and regular case, I could have used a very simple auction to accomplish my stated goal: each bidder submits a bid b_i and price r_i, but instead of being rewarded with revenue only when losing the auction, bidder i receives a payment of $r_i \vee b_{-i}^{(2)}$ if $b_{-i}^{(1)} \geq r_i$. In effect, the seller "simulates" the revenue that the bidder would receive from setting a reserve price of r_i. In the symmetric independent and regular case with distribution $F(v)$, the optimal reserve price is independent of the number of bidders and simply solves:

$$1 - F(v) - vf(v) = 0.$$

so bidders will report a reserve price solving this first-order condition. The seller can then implement this reserve price for the remaining bidders.

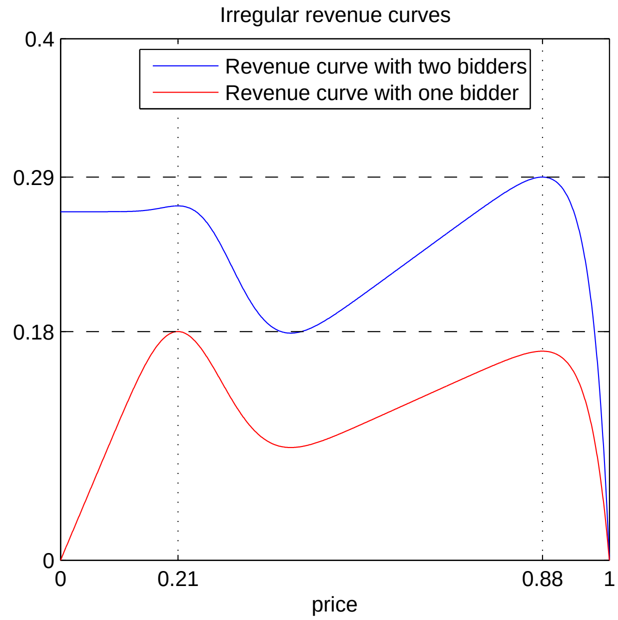

Figure 3.4: Comparison of revenue with one bidder versus two, when values are independently drawn from a $B[\alpha = 11, \beta = 30]$ distribution with probability 0.81 and from a $B[\alpha = 25, \beta = 1]$ distribution with probability 0.19. In the simulated revenue auction, each bidder suggests a price of 0.21, even though 0.88 is the optimal anonymous reserve price with two bidders.

This auction generates no incentives to shade to throw the auction, since the simulated revenue is received regardless of whether the bidder wins the good. However, with irregular, asymmetric or correlated distributions, there is no simple formula for the optimal reserve price, nor an easy way to relate it to some ex-ante reserve price that does not condition on whether or not a bidder is the loser. Indeed, for a modified version of the example from Section 3.4.3 depicted in Figure 3.4, each bidder would suggest a price of 0.21 in the simulated revenue auction, even though 0.88 is the optimal anonymous reserve with two bidders.

3.5.3 Uniqueness

In Section 3.4.1, I characterized two necessary conditions for a symmetric regular equilibrium, namely that when the bid function is differentiable, (FOC) must be satisfied, and $\beta(\overline{v}) = \frac{\overline{v}}{1+\widehat{\alpha}}$. I strongly suspect that the equilibrium of Algorithm 3.1 is unique among the class of regular equilibria, though I have not proven this result. Other authors have investigated uniqueness of auction equilibria in similar settings, notably Lizzeri and Persico (2000) and Lebrun (2006). Lizzeri and Persico (2000) in particular use a notion of regularity that is analogous to my own, though my requirement that bidders bid their values when indifferent is unnecessary for the auctions they consider. This assumption could be dispensed with by modifying the auction format by adding a small probability event that bidder i is sold the good at a randomly drawn price r if $r \leq b$, where r is drawn from the cumulative distribution $G(r)$ with support equal to $\mathbb{R}_+$. This extra incentive to bid close to one's value interacts smoothly with (FOC), and does not substantively change the structure of equilibrium. I hope to consider the question of uniqueness in the future.

3.5.4 Extension to general type spaces

At the heart of my arguments is that bidders should not shade too much in equilibrium, because of the requirement that the sum of marginal surplus and marginal revenue must be zero if bidders shade a positive amount. I showed that for the RSA, in order for the bidding function to solve (FOC), bids must be at least $\frac{v}{1+\widehat{\alpha}}$. However, in order to prove that the bound holds, I had to construct an equilibrium, which required the monotonicity property.

In more general type spaces, there is an easy way to achieve a similar bound using a first-order condition. Consider an auction in which the bidder elicits bids and prices, as in the RSA. With probability $1-\alpha$, the seller picks a bidder to consult at random and uses that bidder's suggested reserve price, also as in the RSA. With probability α, the seller simply uses the second-highest bid as the price for the winner. Crucially, the share of revenue that goes to the consulted bidder is α^2 times realized revenue.

Let us consider the marginal incentive to shade using this auction format. If the pricing constraint is not binding, then there is no incentive to shade, and bidding one's value is a weakly undominated strategy. If the pricing constraint binds, then the marginal surplus is:

$$\begin{aligned}\frac{\partial S(v,\beta(w))}{\partial w} &= \left[(1-\alpha)(v-\rho(w))\mathbb{I}_{\beta(w)\geq\rho(w)} + \alpha(v-\beta(w))\right] \\ &\quad \cdot F_{v^{(1)}_{-ij}|v_i,v_j}(w|v,w) f_{v_j|v_i}(w|v) \\ &\geq \alpha(v-\beta(w)) F_{v^{(1)}_{-ij}|v_i,v_j}(w|v,w) f_{v_j|v_i}(w|v)\end{aligned}$$

and marginal revenue is:

$$\begin{aligned}\frac{\partial R(v,\beta(v))}{\partial w} = \alpha^2 \Bigg[& \left(F_{v_j,v^{(1)}_{-ij}|v_i}(\overline{v},w|v) - F_{v_j,v^{(1)}_{-ij}|v_i}(w,w|v)\right)\beta'(w) \\ & - \beta(w) F_{v^{(1)}_{-ij}|v_i,v_j}(w|v,w) f_{v_j|v_i}(w|v)\Bigg] \\ \geq -\alpha^2 & \beta(w) F_{v^{(1)}_{-ij}|v_i,v_j}(w|v,w) f_{v_j|v_i}(w|v)\end{aligned}$$

If it is true that:

$$\frac{\partial S(v,\beta(w))}{\partial w} + \frac{\partial R(v,\beta(v))}{\partial w} = 0,$$

then:

$$\alpha^2\beta(w) \geq \alpha(v - \beta(w))$$
$$\implies \beta(w) \geq \frac{v}{1+\alpha}.$$

Thus, if an equilibrium exists for this more general mechanism, and if bidders bid their values unless they have a strict incentive to shade (as they would be if the trick referred to in Section 3.5.3 were used), then $\beta(w)$ must be at least $\frac{v}{1+\alpha}$. As a result, bounds similar to those of Proposition 3.2 would obtain. However, existence is no small order, as has been pointed out in the literature (see Reny, 1999; Athey, 2001; Reny and Zamir, 2004).

3.6 Conclusion

This chapter has considered a setting in which the buyers know the distribution of values, and therefore know the optimal reserve price, but the seller does not. The seller wishes to have the bidders communicate enough of what they know so that the seller can obtain the greater revenue associated with a well-chosen reserve price, but the seller also desires that the bidders communicate as little information in as concise a manner as possible. This leads us to a mechanism in which each bidder simply recommends a reserve price for the seller to use in the event that the bidder loses the auction. Truthful reporting of the reserve price is incentivized with revenue sharing.

This rule distorts bidders' incentives to bid their values, and therefore pushes down the equilibrium bid distribution relative to the value distribution. Nonetheless, the distortions are small when the seller only shares a small amount of revenue, and the seller is able to extract virtually all of the revenue that he would obtain if he knew the distribution and set the optimal anonymous reserve price. In that sense, this mechanism accomplishes the seller's goal.

3.A Proofs

Proof of Lemma 3.2. A type v can win for sure at price $\beta(\overline{v})$ and obtain a payoff of:

$$\mathbb{E}\left[v - \rho(v_j) \vee \beta\left(v_{-ij}^{(1)}\right)\middle| v_i = v\right].$$

On the other hand, by bidding $\beta(v)$, a bidder with valuation v can obtain:

$$\mathbb{E}\left[\left(v - \rho(v_j) \vee \beta\left(v_{-ij}^{(1)}\right)\right)\mathbb{I}_{\rho(v_j)\vee\beta(v_{-ij}^{(1)})<\beta(v)}\middle| v_i = v\right]$$
$$+ \widehat{\alpha}\,\mathbb{E}\left[\rho(v) \vee \beta\left(v_{-i}^{(2)}\right)\mathbb{I}_{\beta(v_{-i}^{(1)})>\rho(v)}\middle| v_i = v\right].$$

The difference is:

$$\mathbb{E}\left[\left(v - \rho(v_j) \vee \beta\left(v_{-ij}^{(1)}\right)\right)\mathbb{I}_{\rho(v_j)\vee\beta(v_{-ij}^{(1)})>\beta(v)} - \widehat{\alpha}\rho(v) \vee \beta\left(v_{-i}^{(2)}\right)\mathbb{I}_{\beta(v_{-i}^{(1)})>\rho(v)}\middle| v_i = v\right]$$
$$> \mathbb{E}\left[\left(v - \rho(v_j) \vee \beta\left(v_{-ij}^{(1)}\right) - \widehat{\alpha}\rho(v) \vee \beta\left(v_{-i}^{(2)}\right)\right)\mathbb{I}_{\beta(v_{-i}^{(1)})>\beta(v)}\middle| v_i = v\right],$$

since $\rho(v_j) \geq \beta(v_j)$ and $\rho(v) \geq \beta(v)$. Clearly, $\rho(v) \leq \beta(\overline{v})$, since otherwise no revenue would be generated. Hence, this quantity is at least:

$$\mathbb{E}\left[(v - \beta(\overline{v})(1 + \widehat{\alpha}))\,\mathbb{I}_{v_{-i}^{(1)}>v}\middle| v_i = v\right].$$

If $\beta(\overline{v}) < \frac{\overline{v}}{1+\widehat{\alpha}}$, then this quantity is positive for v sufficiently close to $\overline{v}$, in which case deviating to $\beta(\overline{v})$ will be attractive for such a v.

On the other side, if $\beta(\overline{v}) > \frac{\overline{v}}{1+\overline{\alpha}}$, then type $\overline{v}$'s payoff from bidding $\beta(\overline{v})$ is:

$$\mathbb{E}\left[\overline{v} - \rho(v_j) \vee \beta(v_{-ij}^{(1)})\middle| v_i = \overline{v}\right],$$

whereas the payoff from bidding $\rho(v) < \beta(\overline{v})$ and setting the same price is:

$$\mathbb{E}\left[\left(\overline{v} - \rho(v_j) \vee \beta\left(v_{-ij}^{(1)}\right)\right) \mathbb{I}_{\rho(v_j)\vee\beta(v_{-ij}^{(1)})\leq\rho(v)} \middle| v_i = \overline{v}\right] \\ + \widehat{\alpha}\mathbb{E}\left[\rho(v) \vee \beta\left(v_{-i}^{(2)}\right) \mathbb{I}_{\beta(v_{-i}^{(1)})\geq\rho(v)} \middle| v_i = \overline{v}\right],$$

so that the difference is:

$$\mathbb{E}\left[\left(\overline{v} - \rho(v_j) \vee \beta\left(v_{-ij}^{(1)}\right)\right) \mathbb{I}_{\rho(v_j)\vee\beta\left(v_{-ij}^{(1)}\right)\geq\rho(v)} \middle| v_i = \overline{v}\right] \\ - \widehat{\alpha}\mathbb{E}\left[\rho(v) \vee \beta\left(v_{-i}^{(2)}\right) \mathbb{I}_{\beta(v_{-i}^{(1)})\geq\rho(v)} \middle| v_i = \overline{v}\right].$$

Since $\rho(v_j) \geq \beta(v_j)$, I conclude:

$$\rho(v_j) \vee \beta\left(v_{-ij}^{(1)}\right) \mathbb{I}_{\rho(v_j)\vee\beta\left(v_{-ij}^{(1)}\right)\geq\rho(v)} \geq \rho(v)\mathbb{I}_{\rho(v_j)\vee\beta\left(v_{-ij}^{(1)}\right)\geq\rho(v)},$$

and also $\mathbb{I}_{\rho(v_j)\vee\beta\left(v_{-ij}^{(1)}\right)\geq\rho(v)} \leq \mathbb{I}_{\beta\left(v_{-i}^{(1)}\right)\geq\rho(v)}$. Hence, the difference is at most:

$$\mathbb{E}\left[\left(\overline{v} - \rho\left(v\right)\left(1 + \widehat{\alpha}\right)\right) \mathbb{I}_{\beta\left(v_{-i}^{(1)}\right)\geq\rho(v)} \middle| v_i = \overline{v}\right],$$

which must be negative for $\rho(v)$ close to $\beta(\overline{v})$, since $\rho(v)$ is being squeezed to $\beta(\overline{v}) > \frac{\overline{v}}{1+\widehat{\alpha}}$. □

Proof of Proposition 3.1. Consider a point $\underline{w}^k$, which is the supremum of $v < \overline{w}^k$ such that $\beta_k(v) > v$. By continuity, it must be that $\beta_k(\underline{w}^k) = \underline{w}^k$, so the derivative β' at such a point is:

$$\beta_k'(\underline{w}^k) = \underline{w}^k g(\underline{w}^k, \underline{w}^k | \underline{w}^k).$$

If $\beta_k'(\underline{w}^k) > 1$, then $\beta_k(v) \geq v$ for $v \in [\underline{w}^k, \underline{w}^k + \epsilon)$, which contradicts the definition of $\underline{w}^k$. Hence, it must be that $g(\underline{w}^k, \underline{w}^k | \underline{w}^k) \leq \frac{1}{\underline{w}^k}$.

Next, I show that $g(\overline{w}^k, \overline{w}^k | \overline{w}^k) \geq \frac{1}{\overline{w}^k}$, as long as $\underline{v} < \underline{w}^k < \overline{w}^k$. Clearly this is true at $\overline{w}^0 = \overline{v}$, since $g(\overline{w}^k, \overline{w}^k | \overline{w}^k)$ blows up at that point. For $k > 0$, according to the constructed equilibrium, $\beta(\overline{w}^k) = \overline{w}^k$ and $\overline{w}^k \in r^*(\overline{w}^k, \overline{w}^k)$. This inclusion follows from upper-hemicontinuity of r^*. Moreover, on $(\overline{w}^k - \epsilon, \overline{w}^k]$ it must be that $\beta(v) = v$. As a result, marginal revenue at the reserve price r is exactly (3.2). Clearly, marginal revenue is positive if $g(\overline{w}^k, \overline{w}^k | \overline{w}^k) < \frac{1}{\overline{w}^k}$, which contradicts $\underline{w}^k \in r^*(\underline{w}^k, \underline{w}^k)$.

To summarize, it must be that $g(\overline{w}^k, \overline{w}^k | \overline{w}^k) \geq \frac{1}{\overline{w}^k}$ and $g(\underline{w}^k, \underline{w}^k | \underline{w}^k) \leq \frac{1}{\underline{w}^k}$. By **A2**, there can be at most finitely many points at which $g(v, v|v)$ changes sign. Finally, $\underline{w}^k$ cannot coincide with $\overline{w}^{k+1}$, since if $R(v, v) > R(v, \underline{w}^k)$ for v near $\underline{w}^k$, marginal revenue must be negative so that $1 - v\, g(v, v|v) < 0$ for v near $\underline{w}^k$ (since this function has finitely many zeros), so:

$$\beta'(v) \leq \frac{\beta(v)(1 + \widehat{\alpha}) - v}{\widehat{\alpha} v} < 1,$$

since $\beta(v) \leq v$, so $\beta_k(v) < v$ for v in $(\underline{w}^k - \epsilon, \underline{w}^k]$. This contradicts the definition of $\underline{w}^k$. Hence, any sequence of decreasing v^k for which $g(v^k, v^k | v^k)$ alternates sign (weakly) must terminate after finitely many steps. □

Proof of Lemma 3.3. This is obviously true on regions $[\overline{w}^k, \underline{w}^{k-1}]$, when $\beta(v) = v$. Second, suppose that $\beta(v) < \frac{v}{1+\widehat{\alpha}}$ for some $v \in [\underline{w}^k, \overline{w}^k]$. Since $\beta(v)$ is continuous and $\beta(\overline{w}^k) \geq \frac{\overline{w}^k}{1+\widehat{\alpha}}$, the following quantity is well defined:

$$\widetilde{v} = \inf \left\{ w \geq v \,\middle|\, \beta(w) \geq \frac{w}{1 + \widehat{\alpha}} \right\}.$$

Then $\beta(w) < \frac{w}{1+\widehat{\alpha}}$ for all $w \in (v, \widetilde{v})$. By the mean value theorem, there exists $\widehat{w} \in (v, \widetilde{v})$ such that:

$$\beta'(\widehat{w}) = \frac{\frac{\widetilde{v}}{1+\widehat{\alpha}} - \beta(v)}{\widetilde{v} - v} > \frac{1}{1+\widehat{\alpha}} > 0.$$

But by (FOC), $\beta'(\widehat{w}) > 0$, a contradiction. $\square$

Proof of Lemma 3.4. The derivative of this function is:

$$\begin{aligned}
\frac{dU(v,\beta(w),\beta(w))}{dw} &= \left(\frac{\partial S(v,b)}{\partial b}\bigg|_{b=\beta(w)} + \frac{\partial R(v,r)}{\partial r}\bigg|_{r=\beta(w)} \right) \beta'(w) \\
&= (n-1)(v-\beta(w)) F_{v^{(1)}_{-ij}|v_i,v_j}(w|v,w) f_{v_j|v_i}(w|v) \\
&\quad + \alpha \Bigg[\beta'(w) \left(F_{v_j,v^{(1)}_{-ij}|v_i}(\overline{v}, w|v) - F_{v_j,v^{(1)}_{-ij}|v_i}(w,w|v) \right) \\
&\qquad - \beta(w) F_{v^{(1)}_{-ij}|v_i,v_j}(w|v,w) f_{v_j|v_i}(w|v) \Bigg] \\
&= (n-1) \left(F_{v_j,v^{(1)}_{-ij}|v_i}(\overline{v}, w|v) - F_{v_j,v^{(1)}_{-ij}|v_i}(w,w|v) \right) \\
&\qquad \cdot \Big[(v - \beta(w)(1+\widehat{\alpha})) g(w,w|v) + \widehat{\alpha}\beta'(w) \Big].
\end{aligned}$$

Substituting in (FOC) yields:

$$\begin{aligned}
\frac{dU(v,\beta(w),\beta(w))}{dw} &= (n-1) \left(F_{v_j,v^{(1)}_{-ij}|v_i}(\overline{v}, w|v) - F_{v_j,v^{(1)}_{-ij}|v_i}(w,w|v) \right) \\
&\quad \cdot \Big[(\beta(w)(1+\widehat{\alpha}) - w) g(w,w|w) - (\beta(w)(1+\widehat{\alpha}) - v) g(w,w|v) \Big].
\end{aligned}$$

Note that $\beta(w)(1+\widehat{\alpha}) \geq w$ by Lemma 3.3, so the first term is positive.

Take $w < v$. Then $g(w, w|w) \geq g(w, w|v)$. Hence, it must be that:

$$\begin{aligned}
&(\beta(w)(1+\widehat{\alpha}) - w)g(w, w|w) - (\beta(w)(1+\widehat{\alpha}) - v)g(w, w|v) \\
&\geq (\beta(w)(1+\widehat{\alpha}) - w)g(w, w|v) - (\beta(w)(1+\widehat{\alpha}) - v)g(w, w|v) \\
&= (v - w)g(w, w|w) \geq 0.
\end{aligned}$$

If $w > v$, then $g(w, w|w) < g(w, w|v)$, and:

$$\begin{aligned}
&(\beta(w)(1+\widehat{\alpha}) - w)g(w, w|w) - (\beta(w)(1+\widehat{\alpha}) - v)g(w, w|v) \\
&\leq (\beta(w)(1+\widehat{\alpha}) - w)g(w, w|v) - (\beta(w)(1+\widehat{\alpha}) - v)g(w, w|v) \\
&= (v - w)g(w, w|v) \leq 0.
\end{aligned}$$

□

Proof of Lemma 3.5. To prove the first part of the Lemma, note that:

$$\log\left(F_{v_j, v_{-ij}^{(1)}|v_i}(\overline{v}, r|v) - F_{v_j, v_{-ij}^{(1)}|v_i}(r, r|v)\right) = -\int_{x=\underline{v}}^{\overline{v}} g(x, r|v)dx,$$

which is increasing in v. Hence:

$$F_{v_j, v_{-ij}^{(1)}|v_i}(\overline{v}, r|v) - F_{v_j, v_{-ij}^{(1)}|v_i}(r, r|v)$$

is also increasing in v, as is $1 - rg(r, r|v)$. Given the expression for $\frac{\partial R(v,r)}{\partial r}$, this proves the claim.

By definition, $R(\overline{w}^k, \overline{w}^k) \geq R(\overline{w}^k, r)$ for all $w \in \underline{W}^k$. Hence, if $v < \overline{w}^k$, the difference is:

$$\begin{aligned} R(v,w) - R(v,\overline{w}^k) &= \int_{x=\overline{w}^k}^{w} \frac{\partial R(v,r)}{\partial r} \\ &\leq \int_{x=\overline{w}^k}^{w} \frac{\partial R(\overline{w}^k,r)}{\partial r} \\ &= R(\overline{w}^k, w) - R(\overline{w}^k, \overline{w}^k) \leq 0. \end{aligned}$$

The other direction is significantly more complicated. Our goal is to show that the integral:

$$R(v, \underline{w}^{k-1}) - R(v,w) = \int_{x=w}^{\underline{w}^{k-1}} \frac{\partial R(v,x)}{\partial x} dx$$

is non-negative for $v \geq \underline{w}^{k-1}$. First, I will show that $r^*(v) = \inf(r^*(v,v))$ is monotonically increasing on $\underline{W}^k$. Take $v > v'$ and $x \in r^*(v,v)$. Then $R(v,x) \geq R(v,y)$ for all $y \geq v$. This implies that $R(v',x) \geq R(v,y)$ for all $y \geq x$, by the fact that $\frac{\partial R(v,r)}{\partial r}$ is increasing in v. Hence, if $x \notin r^*(v',v')$, it means that there must be a $y \in r^*(v',v')$ such that $R(v',y) > R(v',x)$ and hence $y < x$. Thus, either (1) $r^*(v) \in r^*(v',v')$, in which case weakly increasing is obvious, or (2) $r^*(v) \notin r^*(v',v')$, in which case there must exist $y < r^*(v)$ in $r^*(v',v')$.

With this monotonicity result in hand, for any point in the image of $x = r^*(v)$ on $(\underline{w}^k, \overline{w}^{k-1}]$, x must be an optimal price for type v, and moreover must satisfy an interior first-order condition. Otherwise, if the constraint $x \geq v$ were binding, it would be the case $R(v,v) > R(v,w)$ for all $w \in (v, \underline{w}^{k-1}]$, which contradicts the definition of $\overline{w}^k$. Hence:

$$\left.\frac{\partial R(v,x)}{\partial x}\right|_{x=r^*(v)} = 0 \leq \left.\frac{\partial R(v,x)}{\partial x}\right|_{x=r^*(v)}.$$

In other words, for any x in the image of r^*, marginal revenue is non-negative for type $\underline{w}^{k-1}$. On the other hand, for any $x \in [\overline{w}^k, \underline{w}^{k-1}] \setminus r^*([\overline{w}^k, \underline{w}^{k-1}])$, it must be that x is passed over at a jump discontinuity of the monotonic function r^*. Let K be the countable collection of intervals that result from jump discontinuities, i.e., the set of $[a,b]$ such that $a = r^*(v')$ and $b = \lim_{v'' \downarrow w} r^*(v'')$ with $b > a$. Clearly, $r^*([\overline{w}^k, \underline{w}^{k-1}]) \cup \{I \in K\} = [\overline{w}^k, \underline{w}^{k-1}]$.

For each $[a,b] \in K$ which is the jump at v', it must be that $\{a,b\} \subset r^*(v',v')$, for if $R(v',a) > R(v',b)$, then this will also be true for $v'' > v'$ but nearby. Moreover, it must be that $R(v',a) \geq R(v',w)$ for any $w \in [a,b]$. Hence:

$$0 = \int_{x=w}^{b} \frac{\partial R(v,x)}{\partial x} dx \leq \int_{x=w}^{b} \frac{\partial R(\underline{w}^{k-1},x)}{\partial x} dx.$$

Finally, this shows that:

$$\int_{x=w}^{\underline{w}^k} \frac{\partial R(v,x)}{\partial x} dx = \int_{(\cup_{I \in K} I) \cap [w, \underline{w}^{k-1}]} \frac{\partial R(v,x)}{\partial x} dx + \int_{r^*([w, \underline{w}^{k-1}])} \frac{\partial R(v,x)}{\partial x} dx \geq 0,$$

which proves the other direction. It is straightforward to repeat the argument with $r^*(v)$ instead of $\underline{w}^k$. For that case, I would show that:

$$\int_{x=w}^{\underline{w}^k} \frac{\partial R(v,x)}{\partial x} dx \geq 0,$$

which is established by analogous arguments.

For the second part of the Lemma, marginal revenue has the same sign as:

$$\begin{aligned}\frac{dU(v,\beta(w))}{dw} &= \alpha\left(F_{v_j,v_{-ij}^{(1)}|v_i}(\overline{v},w|v) - F_{v_j,v_{-ij}^{(1)}|v_i}(w,w|v)\right)[\beta'(w) - \beta(w)g(w,w|v)] \\ &= \alpha\left(F_{v_j,v_{-ij}^{(1)}|v_i}(\overline{v},w|v) - F_{v_j,v_{-ij}^{(1)}|v_i}(w,w|v)\right) \\ &\quad \cdot \left[\frac{\beta(w)-w}{\widehat{\alpha}}g(w,w|w) + \beta(w)(g(w,w|w) - g(w,w|v))\right].\end{aligned}$$

Clearly $\beta(w) \leq w$, so the first term in the brackets is non-positive. Also, for $w \geq v$, **A1** implies that $g(w,w|v) \geq g(w,w|w)$, so the second term is non-positive as well. Hence, marginal revenue is non-positive at $w \in \overline{W}^k$ if $w > v$. □

Proof of Theorem 3.1. To start, fix a valuation v and consider deviations to some (b,r). If $r < v$, then it is without loss of generality to consider the deviation (r,r), since $S(v,b)$ is weakly increasing as long as $b \leq v$. On the other hand, if $r > v$, I can consider deviations of the form (v,r), since $S(v,b)$ is weakly decreasing for $b \geq v$. These are referred to as downward and upward deviations, respectively.

Downward deviations. I prove a base step and an inductive step. The base step considers the cases where $v \in \overline{W}^k$ or $v \in \underline{W}^k$.

For $v \in \overline{W}^k$, consider deviations to some $r \leq \beta(\overline{w}^k)$, since either (1) $\overline{w}^k = \overline{v}$, and it would never be profitable to set a price above the support of bids, or (2) $\beta(\overline{w}^k) = \overline{w}^k > v$. Since β is continuous, and deviations not in the support of bids would never be attractive, the deviation r is equal to $\beta(w)$ for some w. Lemma 3.4 shows that $\beta(v)$, $\beta(v)$ is weakly better than any downward deviation $\beta(w)$ for $w \in \overline{W}^k$.

Now suppose that $v \in \underline{W}^k$. Lemma 3.5 shows that $R(v,\underline{w}^k) \geq R(v,r)$ for all $r \in \underline{W}^k$. Also, any downward deviation would entail lower surplus as well, since

$S(v, b)$ is weakly increasing for $b \leq v$. As a result, there can be no profitable downward deviations to $r = \beta(r) \in \underline{W}^k$. This concludes the base step.

For the first half of the inductive step, suppose that deviating to $(\beta(\overline{w}^k), \beta(\overline{w}^k))$ is not profitable, where $\overline{w}^k \leq v$. Again, Lemma 3.4 shows that $U(v, \beta(w), \beta(w))$ is weakly decreasing for $w \in \overline{W}^k$, so any deviation to $\beta(\overline{W}^k)$ is weakly worse than $\beta(\overline{w}^k)$.

For the second half, suppose that deviating to $(\underline{w}^k, \underline{w}^k)$ is not profitable. Lemma 3.5 again shows that $R(v, \underline{w}^k) \geq R(v, r)$ for all $r \in \underline{W}^k$, and $S(v, b)$ is weakly decreasing, so there are no profitable deviations in $\underline{W}^k$.

Hence, for any $\overline{W}^k$ with $\overline{w}^k < v$ or $\underline{W}^k$ with $\underline{w}^k < v$, it cannot be that there are any profitable downward deviations to $\beta(w)$ with $w \in \overline{W}^k$ or $w \in \underline{W}^k$.

Upward deviations. I again show a base step and an inductive step, as in the downward case. If $v \in \overline{W}^k$, the downward case has shown that (v, v) is not a profitable deviation. By Lemma 3.5, marginal revenue is non-positive if $w > v$, so there cannot be profitable deviations to some $(v, \beta(w))$ with $w \in \overline{W}^k$ and $\beta(w) \geq v$, since this implies $w \geq v$ as well.

On the other hand, if $v \in \underline{W}^K$, Lemma 3.5 shows that $R(v, \underline{w}^k) \geq R(v, r)$ for all $r \in \underline{W}^k$. This concludes the base step for upward deviations.

I have already shown that marginal revenue is non-positive on $\overline{W}^k$. Hence, the deviation $(v, \underline{w}^k)$ is weakly better than $(v, \beta(w))$ for $w \in \overline{W}^k$.

For the other half of the inductive step, Lemma 3.5 shows that $R(v, \overline{w}^{k+1}) \geq R(v, r)$ for all $r \in \underline{W}^k$. So if $(v, \overline{w}^{k+1})$ is not profitable, then neither is such a deviation (v, r). This concludes the inductive step. $\square$

Chapter 4

Extracting common knowledge: Strengthening a folk argument

4.1 Introduction

An assumption underlying much of classical mechanism design is that the designer knows features of the environment which are common knowledge among the agents. Consider the simple problem of designing an auction for the sale of a single unit of a good. An objective frequently attributed to the designer is to maximize expected revenue, where the expectation is taken with respect to a prior distribution of agents' valuations or signals. Numerous results have demonstrated that the optimal mechanism may depend on fine details of this prior distribution, and even the ability to take such an expectation, let alone design the optimal mechanism, presumes that the designer knows what the prior distribution is. In practice, this may very well not be the case.

In this work, I consider what would happen if a designer did not know the prior distribution of values, signals, states, etc. More generally, even if there is no prior because of disagreement among the agents, we can ask what would happen if the designer did not know agents' possible beliefs. The purpose of my inquiry is to determine whether the designer is truly limited by this lack of knowledge, or if he is able to recover the classical results by eliciting such information that is common knowledge among the agents. For example, if bidders' values in an auction were drawn from a prior distribution which is known to the agents, could the designer incentivize the agents to tell him what the prior was, and use this information to design an auction as he pleases?

We have good reason to think that the designer might be able to do just that. An old folk argument in mechanism design, going back to the early days of complete information implementation, says that if a feature of the environment is common knowledge among the agents, then the designer can recover this information at no cost. The designer constructs a mechanism in which the agents simultaneously announce the common knowledge. If the report is unanimous, then the designer goes on to implement his desired mechanism as a function of the agents' reports. If the announcement is divided, then some harsh punishment is meted out, for example, slowly lowering the agents into a shark tank. Such a mechanism has been formally described by Choi and Kim (1999) in a public goods setting and is discussed in Bergemann and Morris (2012a). To the extent that such a grim outcome is less desirable than the designer's choice of mechanism, truthful reporting will be a Nash equilibrium. Nonetheless, this mechanism has many equilibria, most of which involve coordinated misreporting of what the agents know.

I seek a stronger resolution of the designer's problem, in which there are greater assurances that the agents will truthfully reveal their common knowledge. Ideally, the designer would like to use a mechanism in which all rationalizable messages involve truthful reporting of their common knowledge, and such that any compromise of the designer's ultimate goals is minimal. This should be true even though the agents rationally anticipate that the designer will use their reports to design a secondary mechanism, the outcome of which is valued by the agents. For a wide class of private-good environments, I show that it is possible for the designer to accomplish these goals, although there are limitations on the kind of common knowledge that the designer can elicit.

In particular, I consider environments in which outcomes consist of a vector of agent specific components, where there are joint restrictions on which outcomes can be implemented across agents. The environment is private-good-like in that each agent only cares about their own component, and moreover it is always feasible to exclude an agent by giving them a status quo outcome while not changing others' outcomes. An example of such an environment is the allocation problem previously alluded to: the designer has finitely many goods, each of which cannot be allocated to more than one agent at a time. Nevertheless, it is always possible to exclude one agent from receiving any good without changing others' allocations. Such environments give the designer flexibility to punish or reward one agent at a time.

Information and preferences of the agents are modeled using type spaces. Each agent has one of finitely many types, and this type is associated with beliefs over and preferences conditional on others' types. The types are taken to be a sufficient statistic for the distribution of any payoff relevant states of the world that influence preferences. Agents' beliefs are not required to be consistent with a common prior, though there is a "common support" assumption that if a given profile of types can be realized, each agent's type must consider the others' types to be possible, i.e., others' types lie in the support of beliefs. The designer can use type spaces to describe various kinds of common knowledge that may exist among the agents and how that common knowledge should influence the mechanism design. In particular, the designer specifies a collection of type spaces, and for each type space in that collection a mechanism which he would like to implement. Such a specification is referred to as a mechanism mapping, and represents the designer's ideal choice of mechanism if he knew the true type space. However, since the designer does not know the true type space, he must choose a single uniform mechanism to use in all events. The designer's goal is to find such a uniform mechanism in which agents in a given type space will behave similarly to how they would behave in the desired mechanism.

My inquiry is closely related to the work of Abreu and Matsushima (1992a,b, hereafter AM), who consider virtual implementation of social choice functions under the assumptions of expected utility preferences and finite type spaces and outcomes. AM discovered a "measurability" condition which a social choice function must satisfy in order to be virtually implementable. This condition requires that the social choice function prescribe the same outcome for types that have the same preferences over lotteries unconditional on others' types, the same preferences over lotteries conditional on others' unconditional preferences, etc. This measurability condition is further studied and developed in a recent paper of Bergemann, Morris, and Takahashi (2011, hereafter BMT), who construct a "universal preference space" of hierarchies of preferences to better understand when two types can be strictly incentivized to behave differently. BMT explore a solution concept which they term interim preference correlated rationalizability, according to which a message is rationalizable if it is a best response to a correlated conjecture about how others will play, when others' actions can be informative about an agent's own preferences. A key result is that two types have identical preference hierarchies if and only if they have the same rationalizable messages for every mechanism.

I use this solution concept to formalize what it means for the uniform mechanism to be strategically similar to the desired mechanism for the true type space. Fixing a type space, I will say that two mechanisms are ϵ-strategically equivalent if it is possible to identify each type's rationalizable messages in the two mechanisms in such a way that the lotteries over outcomes induced by identified message profiles are the same up to an order ϵ. Moreover, the order ϵ difference in outcomes is such that agents' preferences over rationalizable message profiles are the same between the two mechanisms. This definition captures the idea that the designer is allowed to augment the desired mechanism with additional features to elicit common knowledge, but the mechanism with extra features should still reduce to the original strategic environment. For example, adding a message that cannot be rationalized results in a strategically equivalent mechanism, as does merely relabeling the messages or affine perturbations of lotteries. In other words, the designer is also allowed to modify outcomes slightly, but not in a way that changes the relative merits of rationalizable messages. Note that the notion of rationalizability employed here is quite permissive, in that there is a large set of possible conjectures about others' behavior and one's own preferences that could justify using a particular message. This permissiveness strengthens my result, since it provides a stronger assurance that agents would never find misreporting common knowledge to be optimal.

The designer would like to find a single mechanism which, for each given type space, is ϵ-strategically equivalent to the desired mechanism. When this is possible for ϵ arbitrarily small, then the designer can recover the common knowledge among the agents which is captured in the type space at arbitrarily small cost to his original objectives. In Theorem 4.1, I show that a mechanism mapping can be strategically approximated in this manner by a uniform mechanism only if it satisfies a local preference measurability condition. In particular, if the designer specifies two type spaces that contain types with identical preference hierarchies, then it must be that the desired mechanisms are strategically equivalent on a smaller type space that contains these repeated types. Local preference measurability is the analogue of AM's measurability and BMT's strategic indistinguishability in the context of mechanism mappings. In Theorem 4.2, I demonstrate that local preference measurability is a sufficient condition for a mechanism mapping to admit a strategically equivalent mechanism that is independent of the type space at arbitrarily small cost, given sufficient flexibility to punish agents for misreporting.

Here I give a brief summary of the argument when there are at least three agents. Under the private-good assumption, I construct a mechanism in which agents have a strict incentive to reveal their higher-order preferences. This mechanism essentially rewards agents using scoring rules for accurately reporting the subjective distribution of others' types and subjective relative utilities of outcomes given others' types. Because of the common support assumption, agents must report preference hierarchies that lie in the same smallest belief-closed subset of a universal preference space. The revelation mechanism is used to construct a general uniform equivalent mechanism. Note that if we implement any of the mechanisms specified for a type space containing types with these hierarchies, then by local preference measurability, we will have implemented a mechanism which is strategically equivalent to all of the desired mechanisms. Thus, for each preference hierarchy that appears on the domain of the mechanism mapping, we will pick one such mechanism as the mechanism to be implemented for those preference hierarchies.

In the uniform mechanism, agents report their preference hierarchy, a mechanism that they "suggest" should be implemented, and a message in the suggested mechanism. With small probability, we will implement the outcome that incentivizes truthful reporting of preferences based on the reported preference hierarchies. Importantly, this is the only part of outcome function through which an agent's reported preferences influence the marginal lottery over that agent's component of the outcome. This implies that agents must truthfully report their preference hierarchies in any rationalizable message. Each agent's hierarchy implies a particular mechanism, selected from among the desired mechanisms for type spaces containing the reported hierarchies. If all agents suggest the same mechanism, which is the same as the mechanism implied by the preference reports of all agents, then the designer implements the outcome for that mechanism under the reported messages. It will turn out that any other message profile which does not suggest the correct mechanism is not rationalizable. For example, an agent is allowed to deviate in their reported preferences without changing their suggestion, and although this adversely affects others' outcomes, it will not affect the marginal outcome for the "whistle-blower". This allows an agent to deviate from a unanimous misreport of preference hierarchies. On the other hand, if reports are close to a unanimous report, except for inconsistent suggested mechanisms, then a combination of nudges and more severe punishments induce the agents to switch to suggestions that agree with others' implied mechanisms.

Thus, in any rationalizable message profile, agents report their true preference hierarchies and suggest the correct mechanism for the true hierarchies. But such message profiles can be identified with message profiles in the desired mechanism, and any remaining message is rationalizable if and only if its counterpart is rationalizable in the suggested mechanism. Moreover, in any rationalizable message profile, the parts of the outcome which incentivize truthful revelation of preferences are constant given types. The only variation in outcome lotteries comes from the high probability event in which the desired mechanism's outcome lottery is used. Hence, the uniform mechanism will satisfy the definition of strategic equivalence. This canonical uniform mechanism virtually implements the mechanism mapping by making the probability of implementing the preference revealing mechanism sufficiently small.

It is worth noting that the focus in the present work is quite different from much of the mechanism design literature, which is primarily concerned with the implementation of a social choice function or correspondence that maps states to outcomes. In contrast, I am concerned with the recovery of common knowledge, as captured in the type space, to facilitate the implementation of a mechanism. In that sense, I am agnostic about the specific solution concept that the designer would like to use for implementing social choice functions. For example, if the designer specifies mechanisms with unique Nash equilibria that implement particular social choice functions, then this feature will be preserved under the uniform equivalent mechanism that I construct. On the other hand, a designer may prefer mechanisms that have multiple Nash equilibria but have other desirable properties like low complexity or an equilibrium in weakly dominant strategies, such as a second-price auction with a reserve price. In that case, the designer may want to learn about the type space in order to calibrate the reserve price while forgoing a full-on optimization of revenue. More generally, my results characterize aspects of the agents' common knowledge with respect to which the implemented social choice can vary in an arbitrary fashion.

4.1.1 Related literature

The present work contributes to the literatures on mechanism design and on strategic distinguishability of agents with different beliefs. Many notable results in mechanism design and auction theory rely on the designer knowing a prior distribution of types in order to calibrate a mechanism (e.g., Myerson, 1981; Crémer and McLean, 1988; d'Aspremont and Gérard-Varet, 1979). The present chapter relaxes this assumption, so that the designer must choose a mechanism that is independent of the distribution of types, or more generally of the beliefs of the agents. Similar goals are pursued in the literature on robust mechanism design (Bergemann and Morris, 2005, 2012b). Much of this literature asks not only that the mechanism be independent of the type space, but that the implemented social choice function be independent as well.

In contrast, the current chapter allows the implemented outcome to depend on the type space, but restricts the designers ability to "hard wire" the agents' beliefs into the mechanism. This was also the premise in Chapters 1 and 2, as well as papers by other authors such as Azar, Chen, and Micali (2012). These works assume more structure on preferences, namely private values and quasilinearity, and more structure is placed on the designer's objective. In Chapter 1, the designer's goal is to achieve an optimal worst-case revenue-share of the efficient surplus, and in Chapter 2, the designer simply wishes to guarantee himself the revenue from a second-price auction with an optimally chosen reserve price. Such specific objectives lead to simpler optimal mechanisms, though an implication of the results here is that there are many optimal mechanisms for these objectives. For example, Theorem 4.2 below will imply that the designer can always implement the revenue maximizing reserve price in a second-price auction as a function of “minimal” prior distributions, where minimality simply means that the prior cannot be written as a randomization over priors on disjoint sets of valuations. Indeed, the present model does not assume anything about the designer’s motives conditional on the true type space, and the results provide conditions under which optimization of the mechanism can be performed type space by type space, with the results of these optimizations being approximated by a single type space-independent mechanism.

As previously mentioned, the particular notion of approximation is very much in the spirit of virtual implementation as in AM (see also Sen and Abreu, 1991; Bergemann and Morris, 2009b). Broadly speaking, some objective is achieved with arbitrarily high probability according to some solution concept. For AM, the objective is to implement a social choice function according to iterated deletion of strictly dominated strategies. Here, the objective is to implement a mechanism for a given type space according to strategic equivalence. Either way, the small probability events on which the objective is not achieved are used to provide incentives that pressure the agents to reveal information. AM and much of the subsequent literature on virtual implementation are concerned with general environments in which goods may be public, in contrast to the private-good environments studied here. AM’s constructions also make heavy use of finiteness and the fact that there is a uniform lower bound on preference differences for different types. In the present setting, the designer may wish to distinguish between infinitely many type spaces, so that there may be no such uniform lower bound, and thus the arguments end up being quite different.

Strategic equivalence of mechanisms for a given type space is closely related to strategic equivalence of types. Many authors have considered when two types from two different type spaces will exhibit the same behavior. Dekel, Fudenberg, and

Morris (2007) explore a solution concept called interim correlated rationalizability (ICR), which is stronger than the notion of rationalizability used here and in BMT, and they show that two types have the same higher-order beliefs as in Mertens and Zamir (1985) if and only if they have the same ICR actions in every game. BMT show a similar result for interim preference correlated rationalizability and finite mechanisms. Other authors have explored solution concepts, namely Bayesian Nash equilibrium, under which redundant types that repeat the same higher-order beliefs can contain information that changes equilibrium behavior (Liu, 2009; Sadzik, 2011). The present chapter uses the insights of this literature to define a notion of strategic equivalence of mechanisms that is invariant to such redundant types. In other words, if two type spaces induce the same preference hierarchies, then they also induce the same strategic equivalence classes of mechanisms.

My work makes use of rationalizability in mechanisms with infinitely many messages, and the particular mechanisms involved require transfinitely many rounds of deletion in order to arrive at a stable set of messages for each type, as in Lipman (1994). As such, care must be taken to make sure that the set of rationalizable messages is well-defined and to avoid the pathological behavior illustrated in Dufwenberg and Stegeman (2002). I adopt a definition of transfinite deletion of never-best responses which is adapted from Chen, Long, and Luo (2007, hereafter CLL) in a complete information setting, which ensures that the set of rationalizable messages is uniquely defined and does not depend on the order in which messages are eliminated. Strange behavior can still arise with this definition, whereby a non-rationalizable message can be a better response to a rationalizable strategy profile than any rationalizable message. This could in principle allow two mechanisms to be strategically equivalent even though they do not have comparable Nash equilibria. The constructions employed in the proof of my main result do not make use of such features, however, and it is without loss of generality to restrict to a class of regular mechanisms in which rationalizable messages dominate non-rationalizable messages, as long as others' strategies are rationalizable.

The rest of the chapter proceeds as follows. Section 4.2 presents a model of state dependent preferences over outcomes, and formulates the designer's problem. Section 4.3 defines notions of strategic equivalence of mechanisms and virtual implementation of a mechanism mapping by a single uniform mechanism. Section 4.4 applies this solution concept to characterize when it is possible to extract common knowledge for the purpose of implementing the designer's desired mechanism mapping. Section 4.5 concludes.

4.2 Model

A designer must select an outcome from a finite set A, and there is a finite set of agents $N = \{1, \ldots, n\}$ who have preferences over which outcome is implemented.[1] The agents' information and preferences are described by a Harsanyi type space: each agent has one of finitely many types in T_i, with the set of feasible tuples of types being denoted by $T \subseteq \times_{i \in N} T_i$. I will assume that for every $t_i \in T_i$, there is some $t' \in T$ such that $t'_i = t_i$. Each type has preferences over functions that map types in T_{-i} to lotteries on A. This preference can be represented by beliefs about others' types $p_i : T_i \to \Delta(T_{-i})$, where:

$$T_{-i} = \left\{ t_{-i} \in \times_{j \in N \setminus \{i\}} T_j | \exists t_i \text{ s.t. } (t_i, t_{-i}) \in T \right\},$$

[1]The basic setup of the model is adapted from BMT.

and type contingent utility functions $u_i : T \times A \to (0, \infty)$, so that for $f, f' : T_{-i} \to \Delta(A)$, f is preferred to f' by type t_i if and only if:

$$\sum_{\substack{t_{-i} \in T_{-i}, \\ a \in A}} p_i(t_{-i}|t_i) u_i((t_i, t_{-i}), a) f(a|t_{-i}) \geq \sum_{\substack{t_{-i} \in T_{-i}, \\ a \in A}} p_i(t_{-i}|t_i) u_i((t_i, t_{-i}), a) f'(a|t_{-i}).$$

This representation of preferences is unique up to an affine transformation. I apply the normalization that $\sum_{t_{-i}, a} p(t_{-i}|t_i) u_i(t, a) = 1$, so that preferences can be jointly represented by $\pi : T_i \to \Delta(T_{-i} \times A)$, where f is preferred to f' by type t_i if and only if:

$$\sum_{\substack{t_{-i} \in T_{-i}, \\ a \in A}} \pi(t_{-i}, a|t_i) f(a|t_{-i}) \geq \sum_{\substack{t_{-i} \in T_{-i}, \\ a \in A}} \pi_i(t_{-i}, a|t_i) f'(a|t_{-i})$$

We can map a probability/utility representation into this joint representation by setting:

$$\pi(t_{-i}, a|t_i) = \frac{p(t_{-i}|t_i) u((t_i, t_{-i}), a)}{\sum_{t'_{-i}, a'} p(t'_{-i}|t_i) u((t_i, t'_{-i}), a')}.$$

The *type space* is collectively denoted by $\mathcal{T} = (T, \{\pi_i\}_{i \in N})$.

Such a type space can be thought of as a reduced form representation for a model in which there is a payoff relevant state θ which lives in a set Θ, and agents have beliefs $p_i : T_i \to \Delta(\Theta \times T_{-i})$ and state contingent utilities $u_i : T_i \times \Theta \times A_i \to \mathbb{R}_+$. The present formulation can be obtained by integrating out the states in Θ, taking the types t_{-i} to be a sufficient statistic for the underlying payoff-relevant state.

The following assumptions will be key to the results that follow:

Assumption 4.1 (Private good)**.** $A \subseteq \times_{i\in N} A_i$ and $\pi_i(t_{-i}, a|t_i) = \pi_i(t_{-i}, a'|t_i)$ if $a_i = a'_i$. Moreover, there exists $0_i \in A_i$ such that if $a \in A$, there exists $a' \in A$ where $a'_i = 0_i$ and $a'_{-i} = a_{-i}$.

This assumption ensures that the outcome space is private-good-like, in that the outcome consists of components for each agent, and each agent is only concerned with their own component of the outcome. Even though A has components for each agent, there may be joint feasibility restrictions on which allocations can be implemented for each agent, which can be captured by A being a strict subset of $\times_{i\in N} A_i$. In light of Assumption 4.1, I will simply write $\pi_i(t_{-i}, a_i|t_i)$. The next assumption is:

Assumption 4.2 (Worst outcome and non-triviality)**.** For each $i \in N$, there exists $\underline{a}^i \in A$ such that $\pi_i(t_{-i}, \underline{a}^i_i|t_i) = 0$ for all t.

Assumption 4.2 is substantive and with some loss of generality, and directly correspond to assumptions made in BMT and Morris and Takahashi (2012). It says that for each agent, there is some outcome (not necessarily the same as the null outcome) which is worse in every state than any other outcome. Combined with the assumption that $\sum_{t_{-i},a_i} \pi_i(t_{-i}, a_i|t_i) = 1$, this implies that the agent is not indifferent between all outcomes for any types that occur with positive subjective probability. Moreover, this assumption implies that the set of preferences can be represented by the compact set $\Delta(T_{-i} \times A_i^+)$, where $A_i^+ = A_i \setminus \{\underline{a}^i_i\}$. Note that the worst-outcome assumption plays a more substantive role in the present work than it does in BMT, since the worst outcome is sometimes used to punish agents for not making unanimous reports.

The final assumption concerns agents' beliefs. Let $\pi_i(t_{-i}|t_i) = \sum_{a_i\in A_i} \pi_i(t_{-i}, a_i|t_i)$ denote type t_i's conditional marginal beliefs over T_{-i}.

Assumption 4.3 (Common support)**.** If $t \in T$, then $\pi_i(t_{-i}|t_i) > 0$.

In words, Assumption 4.3 says that if agent j has a particular type t_j, then agent i must believe it is possible for agent j to have type t_j.

A subset $T' \subseteq T$ is *belief-closed* if for all $t \in T'$, $\sum_{t_{-i} \in T'_{-i}} \pi_i(t_{-i}|t_i) = 1$. A type space $\mathcal{T}' = (T', \{\pi'_i\}_{i \in N})$ is a belief-closed subspace of $\mathcal{T}$ if $T' \subseteq T$ and T' is belief closed, and if π' is the restriction of π to T'_i. $\mathcal{T}$ is *smallest belief-closed* (SBC) if there does not exist a belief closed subspace of $\mathcal{T}$ that is not equal to $\mathcal{T}$. Assumption 4.3 implies that for all $t \in T$, the SBC subspace of $\mathcal{T}$ containing t_i is the same as the SBC subspace containing t_j for all $i, j \in N$.

A motivating example for the kind of environment captured by this formalism is an auction for one or more goods in a finite set Q. The designer can require the agents to send transfers in $\{\underline{t}, \overline{t}\}$, with lotteries generating all expected payments in $[\underline{t}, \overline{t}]$. Outcomes for agent i can be represented by $A_i = \{0,1\}^Q \times \{\underline{t}, \overline{t}\}$ with $\underline{t} \leq 0 \leq \overline{t}$. Thus, an outcome is an (f_i, t_i) such that $f_i : Q \to \{0,1\}$ and $f_i(q) = 1$ if agent i is allocated item q and t_i is a transfer in $\{\underline{t}, \overline{t}\}$. The joint feasibility restriction is that (f, t) must satisfy $\sum_{i \in N} f_i(q) \leq 1$ for all $q \in Q$, which simply says that a given item cannot be allocated to more than one agent. The worst outcome for agent i is that $f_i(q) = 0$ for all i and $t_i = \overline{t}$. This assumes a "free disposal" property, that if an agent preferred not to receive a good q, then he could always just throw it away to avoid incurring disutility. More than one outcome could count as the "default" option in this setting, but a natural status quo is $f_i(q) = 0$ for all i and an average transfer of zero, which is achieved by a lottery that puts weight $\frac{-\underline{t}}{\overline{t}-\underline{t}}$ on $\overline{t}$ and weight $\frac{\overline{t}}{\overline{t}-\underline{t}}$ on $\underline{t}$ (strictly speaking, this is not a pure outcome, but we could add a 0 transfer to the set of outcomes in order to exactly nest the model).

The agents' types are unknown to the designer, but the designer can commit to a mechanism by which agents reports will determine which outcome is implemented. A *mechanism* consists of measurable sets of messages M_i for each agent and a measurable mapping $g : M \to A$, where $M = \times_{i \in N} M_i$. Such a mechanism is denoted by $\mathcal{M} = (M, g)$. When necessary, I will write $M(\mathcal{M})$ and $g(\mathcal{M})$ for the message space and outcome functions associated with a given mechanism $\mathcal{M}$. A mechanism together with a type space comprise a game of incomplete information, in which each agent's strategies are mappings from types to probability distributions over messages. Thus, the set of strategies for agent i is $\Sigma_i = \{\sigma_i : T_i \to \Delta(M_i)\}$. I identify strategy profiles $\sigma \in \Sigma = \times_{i \in N} \Sigma_i$ with the product measure $\times_{i \in N} \sigma_i(dm_i | t_i)$. Fixing such a profile, agent i's payoff conditional on his type is:

$$\sum_{t_{-i} \in T_{-i}} \int_{m \in M} \sum_{a_i \in A_i^+} g_i(a_i | m) \sigma(dm | t_i, t_{-i}) \pi(t_{-i}, a_i | t_i).$$

In words, this is the expected utility integrated over other agents' types, over messages sent conditional on others' types, and over outcomes implemented for a given profile of messages. A Bayesian Nash equilibrium σ is a strategy profile such that each agent's $\sigma_i(t_i)$ maximizes the expected utility of type t_i, holding σ_{-i} fixed.

Rather than designing a mechanism for a single type space, the designer posits that one of a number of possible type spaces may obtain from a collection $\mathbf{T}$. The designer is able to make conditional statements of the form, "If type space $\mathcal{T} \in \mathbf{T}$ obtains, then I would like to implement the mechanism $\mathbf{M}(\mathcal{T})$." Collectively, I will refer to the objects $(\mathbf{T}, \mathbf{M})$ as a *mechanism mapping.* The purpose of the chapter is to find conditions under which there is a single mechanism $\mathcal{M}$, independent of the type space, which will implement the mechanism mapping $(\mathbf{T}, \mathbf{M})$ in the sense that agents in type space $\mathcal{T} \in \mathbf{T}$ face a similar strategic situation and induce similar outcomes in mechanism $\mathcal{M}$ as they would in $\mathbf{M}(\mathcal{T})$. The precise notion of strategic similarity will be developed in the next section.

Note that if the domain of the mechanism mapping $\mathbf{T}$ were finite, then the setup would be very similar to that of AM. In that case, one could directly apply their solution concept and construction to implement a social choice function, subject to their conditions of measurability and self-selection. However, there is no such restriction, and $\mathbf{T}$ is allowed to be an arbitrary set of type spaces. This complicates matters, since the construction of AM relies on there being only finitely many types so that there is a uniform lower bound on the differences in preferences across types and outcomes. Nonetheless, we will see that the private-good structure and other assumptions facilitate mechanisms for implementing mechanism mappings that do not require discrete differences in preferences.

4.3 Strategic equivalence

I now turn to the issue of defining a reasonable notion of strategic equivalence of mechanisms. I start with a formal description of rationalizability, adapted from BMT's interim preference correlated rationalizability to infinite mechanisms, and I show that the set of rationalizable messages is always well defined. Rationalizability is then used to define strategic equivalence of mechanisms, in which two mechanisms are ϵ-strategically equivalent for a given type space if the mechanisms restricted to rationalizable messages induce similar lotteries and similar preferences over message profiles. The remainder of the section characterizes the relationship between strategic equivalence and higher-order preferences. In particular I show that strategic equivalence classes of mechanisms only depend on higher-order preferences, and not on any other features of the type space.

4.3.1 Rationalizability

The main solution concept employed in this chapter will be the notion of interim preference correlated rationalizability due to BMT, which I will simply refer to as *rationalizability*. Let $\Xi_i : T_i \rightrightarrows M_i$ be a correspondence mapping types into measurable subsets of messages for each i, which is extended to $\Xi : T \rightrightarrows M$ by $\Xi(t) = \times_{i \in N} \Xi_i(t_i)$. A Ξ-*consistent conjecture* for t_i is a probability measure $\mu \in \Delta(\operatorname{graph} \Xi_{-i} \times A_i^+)$ such that:

$$\int_{m_{-i} \in M_{-i}} \mu(t_{-i}, dm_{-i}, a_i) = \pi(t_{-i}, a_i | t_i).$$

We say a message $m_i \in M_i$ is *rationalizable for type* t_i *given* Ξ if there exists a Ξ-consistent conjecture μ such that:

$$\begin{aligned} &\sum_{t_{-i} \in T_{-i}} \int_{m_{-i} \in M_{-i}} \sum_{a_i \in A_i^+} g_i(a_i | m_i, m_{-i}) \mu(t_{-i}, dm_{-i}, a_i) \\ \geq &\sum_{t_{-i} \in T_{-i}} \int_{m_{-i} \in M_{-i}} \sum_{a_i \in A_i^+} g_i(a_i | m_i', m_{-i}) \mu(t_{-i}, dm_{-i}, a_i) \end{aligned}$$

for all $m_i' \in M_i$. In other words, m_i is a best reply to the conjecture μ. Let $R_i(t_i, \Xi, \mathcal{T}, \mathcal{M})$ denote the set of messages in $\Xi_i(t_i)$ which are rationalizable for type t_i given Ξ. $R(\Xi, \mathcal{T}, \mathcal{M})$ is the correspondence Ξ' such that $\Xi_i'(t_i) = R_i(t_i, \Xi, \mathcal{T}, \mathcal{M})$. Note that a sufficient condition for a message m_i not to be in $R_i(t_i, \Xi, \mathcal{T}, \mathcal{M})$ is that there exists some σ_i that generates strictly higher utility for all Ξ-consistent conjectures μ, so that m_i is strictly dominated.

A *deletion sequence* is an indexed set $\{\Xi^\lambda\}_{\lambda\in\Lambda}$ where Λ is well-ordered such that (1) $\Xi_i^{\lambda_0}(t_i) = M_i$ for $\lambda_0 = \min\Lambda$, (2) $\Xi^{\lambda+1} \supseteq R(\Xi^\lambda, \mathcal{T}, \mathcal{M})$ where $\lambda + 1 = \min\{\lambda' \in \Lambda | \lambda' > \lambda\}$, (3) $\Xi^\lambda = \cap_{\{\lambda'\in\Lambda|\lambda'<\lambda\}}\Xi^{\lambda'}$ if $\lambda \neq \lambda' + 1$ for some λ', and (4) $\Xi^* = \cap_{\lambda\in\Lambda}\Xi^\lambda$ satisfies $\Xi^* = R(\Xi^*, \mathcal{T}, \mathcal{M})$. Such a Ξ^* is called a *maximal reduction.* This definition roughly corresponds to that of CLL, who explored transfinitely iterated deletion of strictly dominated strategies in a complete information setting.

Proposition 4.1 (Maximal reduction)**.** *A maximal reduction exists and is unique.*

Proof of Proposition 4.1. Fact: if $\Xi_i(t_i) \subseteq \Xi_i'(t_i)$ for all i and t_i and message m_i is not rationalizable for type t_i given Ξ' then it is not rationalizable for t_i given Ξ. If m_i is rationalizable for t_i given Ξ, there exists a Ξ-consistent conjecture for which m_i is a best reply. But a Ξ-consistent conjecture is Ξ'-consistent as well, so that m_i is rationalizable given Ξ'. Hence, we conclude that $R(\Xi, \mathcal{T}, \mathcal{M}) \subseteq R(\Xi', \mathcal{T}, \mathcal{M})$.

The rest of CLL's argument goes through, which is replicated for completeness: we can define a deletion sequence by $\Xi^{\lambda+1} = R(\Xi^\lambda, \mathcal{T}, \mathcal{M})$ and $\Xi^\lambda = \cap_{\{\lambda'\in\Lambda|\lambda'<\lambda\}}\Xi^{\lambda'}$ for limit ordinals, and Λ is an ordinal with the same cardinality as the power set of graph Ξ. This sequence is decreasing by the previous paragraph. If $R(\Xi^{\lambda+1}, \mathcal{T}, \mathcal{M}) = R(\Xi^\lambda, \mathcal{T}, \mathcal{M})$, then $\Xi^{\lambda'} = \Xi^\lambda$ for all $\lambda' > \lambda$ and Ξ^λ is a maximal reduction. If it is strictly decreasing, we can define an injective mapping from Λ into graph Ξ by associating to each λ a unique $m_i \in \Xi_i^\lambda(t_i) \setminus \Xi_i^{\lambda+1}(t_i)$, so that graph Ξ has greater cardinality than Λ and hence its own power set, a contradiction. Hence, this sequence must be constant after some point and a maximal reduction Ξ^* exists.

For uniqueness, suppose that $\widehat{\Xi}^*$ is also a maximal reduction and the limit of $\{\widehat{\Xi}^\lambda\}_{\lambda\in\widehat{\Lambda}}$, maintaining Ξ^* as the maximal reduction constructed in the previous paragraph. Note that it is without loss of generality to take $\widehat{\Lambda} = \Lambda$, since if

$\widehat{\Lambda} < \Lambda$, we can extend the deletion sequence to $\{\widehat{\Xi}^\lambda\}_{\lambda\in\Lambda}$ by keeping it constant for $\lambda > \widehat{\Lambda}$, or vice versa if $\Lambda < \widehat{\Lambda}$. By definition, $\Xi^{\lambda_0} \supseteq \Xi^* \cup \widehat{\Xi}^*$. Inductively, if $\Xi^\lambda \supseteq \Xi^* \cup \widehat{\Xi}^*$, then again by monotonicity $\Xi^{\lambda+1} \supseteq R(\Xi^\lambda, \mathcal{T}, \mathcal{M}) \supseteq R(\Xi^* \cup \widehat{\Xi}^*, \mathcal{T}, \mathcal{M}) \supset R(\widehat{\Xi}^*, \mathcal{T}, \mathcal{M}) = \widehat{\Xi}^*$. Thus, we conclude that $\widehat{\Xi}^* \subseteq \Xi^*$, and the converse argument shows that the two must be equal. □

The proof is essentially that of CLL adapted to the incomplete information setting and the different solution concept. The key property is the monotonicity of R, which is ensured by allowing any message in M_i to be a better reply than a message in $\Xi_i^\lambda(t_i)$. I will use the notation $\Xi^*(\mathcal{T}, \mathcal{M})$ for the maximal reduction of a mechanism $\mathcal{M}$ on a type space $\mathcal{T}$. This correspondence can always be identified by the fast deletion sequence $\{\Xi^\lambda(\mathcal{T}, \mathcal{M})\}_{\lambda\in\Lambda}$, in which $\Xi^{\lambda+1}(\mathcal{T}, \mathcal{M}) = R(\Xi^\lambda(\mathcal{T}, \mathcal{M}), \mathcal{T}, \mathcal{M})$ for all λ.

4.3.2 Strategic equivalence

Fixing a type space $\mathcal{T}$ and $\epsilon > 0$, we say a mechanism $\mathcal{M} = (M, g)$ is $(\mathcal{T}, \epsilon)$-*strategically equivalent* to $\mathcal{M}' = (M', g')$ if there exists bijections $\eta_i(t_i) : \Xi_i^*(t_i, \mathcal{T}, \mathcal{M}) \to \Xi_i^*(t_i, \mathcal{T}, \mathcal{M}')$ and $\alpha(t) > 0$ such that:

$$\|g(m) - g'(\eta(t, m))\| \leq \epsilon \tag{4.1a}$$

$$g(m) - g(m') = \alpha(t)\left[g'(\eta(t, m)) - g'(\eta(t, m'))\right] \tag{4.1b}$$

for all $t \in T$ and $m, m' \in M$ such that $m_i, m_i' \in \Xi_i^*(\mathcal{T}, \mathcal{M})(t_i)$ for all i. I will refer to the bijections η_i as *ϵ-outcome preserving mappings.* In words, we can identify the rationalizable messages in $\mathcal{M}$ with the rationalizable messages in $\mathcal{M}'$ in such a way that outcomes only differ by ϵ, and differences in outcomes are proportional according to positive scaling parameters $\alpha(t)$. This relation is indicated by $\mathcal{M} \sim_{(\mathcal{T},\epsilon)} \mathcal{M}'$. If $\mathcal{M} \sim_{(\mathcal{T},\epsilon)} \mathcal{M}'$ for all $\epsilon > 0$, I will simply write $\mathcal{M} \sim_{\mathcal{T}} \mathcal{M}'$.

This notion of equivalence is motivated by the following normative properties. First, mechanisms should be regarded as equivalent if they are the same except for relabeling of messages. Second, mechanisms are equivalent if one can be constructed from the other by adding a non-rationalizable message. And third, mechanism equivalence should satisfy transitivity. With $\epsilon = 0$, these properties exactly characterize the relation defined above for finite mechanisms. With positive ϵ, I allow for small perturbations of outcomes so that different types select rationalizable messages differently from otherwise redundant messages. However, these perturbations should not change the strategic calculus of the agents in the sense of affecting the relative merits of one rationalizable message profile over another. Note that proportional differences in lotteries is sufficient but not necessary for the agents' preferences over message profiles to be preserved (cf. Morris and Ui, 2004).

Lemma 4.1 (Transitivity)**.** *If* $\mathcal{M}^A \sim_{(\mathcal{T},\epsilon)} \mathcal{M}^B$*, and* $\mathcal{M}^B \sim_{(\mathcal{T},\epsilon')} \mathcal{M}^C$*, then* $\mathcal{M}^A \sim_{(\mathcal{T},\epsilon+\epsilon')} \mathcal{M}^C$*.*

Proof of Lemma 4.1. By assumption, there are bijective mappings:

$$\eta_i^{A\to B}(t_i) : \Xi^*(\mathcal{T},\mathcal{M}^A)(t_i) \to \Xi^*(\mathcal{T},\mathcal{M}^B)(t_i)$$
$$\eta_i^{B\to C}(t_i) : \Xi^*(\mathcal{T},\mathcal{M}^B)(t_i) \to \Xi^*(\mathcal{T},\mathcal{M}^C)(t_i).$$

Then clearly:

$$\eta_i^{A\to C} = \eta_i^{B\to C} \circ \eta_i^{A\to B}$$

are bijections from $\Xi^*(\mathcal{T}, \mathcal{M}^A)$ to $\Xi^*(\mathcal{T}, \mathcal{M}^C)$. From the definition of equivalence, there are scaling factors $\alpha^A(t), \alpha^B(t) > 0$ such that:

$$\begin{aligned} g^A(m) - g^A(m') &= \alpha^A(t)\left[g^B(\eta^{A\to B}(t,m)) - g^B(\eta^{A\to B}(t,m'))\right] \\ g^B(m) - g^B(m') &= \alpha^B(t)\left[g^C(\eta^{B\to C}(t,m)) - g^C(\eta^{B\to C}(t,m'))\right] \end{aligned}$$

so that:

$$g^A(m) - g^A(m') = \alpha^B(t)\alpha^C(t)\left[g^C(\eta^{A\to C}(t,m)) - g^C(\eta^{A\to C}(t,m'))\right].$$

Also, the triangle inequality implies that for all m:

$$\begin{aligned} \|g^A(m) - g^C(m')\| &\leq \|g^A(m) - g^B(m')\| + \|g^B(m) - g^C(m')\| \\ &< \epsilon + \epsilon'. \end{aligned}$$

□

4.3.3 Preference hierarchies

An important feature of strategic equivalence is that it only depends on higher-order preferences that appear in a given type space and not on any other information contained in the type space. Formally, the space of higher-order preferences is constructed as follows. Define:

$$X_i^0 = \{\emptyset\}$$
$$X_i^k = X_i^{k-1} \times \Delta(X_{-i}^{k-1} \times A_i^+)$$
$$X_i^* = \prod_{k=0}^{\infty} \Delta(X_{-i}^k \times A_i^+),$$

so that:

$$X_{-i}^k = X_{-i}^{k-1} \times \prod_{j \neq i} \Delta(X_{-j}^{k-1} \times A_j \setminus \{\underline{a}_j^j\}).$$

X_i^* is the set of hierarchies of preferences for agent i. A sequence $\{x_i^k\}_{k=1}^{\infty} \in X_i^0$ is *coherent* if the marginal distribution of x_i^k on $X_{-i}^{k-1} \times A_i^+$ is equal to x_i^{k-1}. Define T_i^0 to be the set of coherent hierarchies in X_i^*. The Kolmogorov extension theorem implies that for any coherent hierarchy $x_i \in T_i^0$, we can find a unique measure $\pi_i^X(x_i) \in \Delta(X_{-i}^* \times A_i^+)$ such that x_i^k is the marginal of $\pi_i^X(x_i)$ on $X_{-i}^k \times A_i^+)$. Now inductively define:

$$T_i^k = \{x_i \in T_i^{k-1} | \pi_i^X(T_{-i}^{k-1} \times A_i^+, x_i) = 1\},$$

and define $T_i^* = \cap_{i=1}^{\infty} T_i^k$, which is the set of hierarchies in which there is common certainty of coherency. T^* is called the *universal preference space.* The mapping π_i^X restricted to T_i^* defines a homeomorphism $\pi_i^* : T_i^* \to \Delta\left(T_{-i}^* \times A_i^+\right)$.

For a given type space $\mathcal{T}$, we can identify a type t_i with the universal preference $\phi_i(t_i, \mathcal{T}) \in T_i^*$ where:

$$\phi_i^1(\{\emptyset\}, a_i, t_i, \mathcal{T}) = \sum_{t_{-i} \in T_{-i}} \pi_i(t_{-i}, a_i | t_i)$$
$$\phi_i^k(\phi_{-i}^{k-1}, a_i, t_i, \mathcal{T})(a_i, \phi_{-i}^{k-1}) = \sum_{\{t_{-i} \in T_{-i} | \phi_{-i}^{k-1}(t_{-i}, \mathcal{T}) = \phi_{-i}^{k-1}\}} \pi_i(t_{-i}, a_i | t_i).$$

I will write $\phi(\mathcal{T})$ for the type space consisting of types $\{\phi(t, \mathcal{T}) | t \in T(\mathcal{T})\}$, and with the preferences represented by π_i^*. Thus, $\phi(\mathcal{T})$ is the image of $\mathcal{T}$ in the universal preference space.

It is a result of BMT (Theorem 3) that two types have the same rationalizable messages in every finite mechanism if and only if they have the same preference hierarchies. This result is readily extended to the infinite mechanisms used here.

Proposition 4.2 (Strategic equivalence of types)**.** $\Xi_i^*(t_i, \mathcal{T}, \mathcal{M}) = \Xi_i^*(t_i', \mathcal{T}', \mathcal{M})$ *for every mechanism* $\mathcal{M}$ *if and only if* $\phi_i(t_i, \mathcal{T}) = \phi_i(t_i', \mathcal{T}')$.

Proof of Proposition 4.2. The only if is proven by example with finite mechanisms in BMT. Here I will show that the if part extends to infinite mechanisms.

Suppose that there are correspondences $\Xi_i : T_i \rightrightarrows M_i$ and $\Xi_i' : T_i' \rightrightarrows M_i$ such that $\phi_i(t_i, \mathcal{T}) = \phi_i(t_i', \mathcal{T}')$ implies that $\Xi_i(t_i) = \Xi_i'(t_i')$. Then I claim that $R_i(t_i, \Xi, \mathcal{T}, \mathcal{M}) = R_i(t_i', \Xi', \mathcal{T}', \mathcal{M})$. Let $m_i \in R_i(t_i, \Xi, \mathcal{T}, \mathcal{M})$. Then there exists a Ξ-consistent conjecture $\mu \in \Delta(\text{graph}\, \Xi \times M_{-i} \times A_i^+$ for t_i such that m_i is a best reply for t_i to μ. We will construct a Ξ'-consistent conjecture μ' such that m_i is a best reply for t_i' to μ'. Write:

$$\begin{aligned}\widehat{\pi}_i(\phi_{-i}, a_i|t_i) &= \sum_{\{t_{-i}\in T_{-i}|\phi_{-i}(t_{-i},\mathcal{T})=\phi_{-i}\}} \pi_i(dt_{-i}, a_i|t_i),\\ &= \sum_{\{t'_{-i}\in T'_{-i}|\phi_{-i}(t'_{-i},\mathcal{T}')=\phi_{-i}\}} \pi'_i(t'_{-i}, a_i|t'_i).\end{aligned}$$

Note that these two are equal precisely because $\phi_i(t_i, \mathcal{T}) = \phi_i(t_i', \mathcal{T}')$, and therefore have the same marginal distribution over $T^*_{-i} \times A_i^+$. Also, for ϕ_{-i} such that $\phi_{-i}^{-1}(\phi_{-i}, \mathcal{T})$ intersects the support of $\pi_i'(t'_{-i}, a_i|t_i')$, define:

$$\sigma_{-i}(dm_{-i}|\phi_{-i}, a_i) = \frac{1}{\widehat{\pi}_i(\phi_{-i}, a_i|t_i)} \sum_{\{t_{-i}\in T_{-i}|\phi_{-i}(t_{-i},\mathcal{T})=\phi_{-i})\}} \mu(t_{-i}, dm_{-i}, a_i).$$

Then the distribution:

$$\mu'(t'_{-i}, dm_{-i}, a_i) = \pi_i'(t'_{-i}, da_i|t_i')\sigma_{-i}(dm_{-i}|\phi_{-i}(t'_{-i}, \mathcal{T}'), a_i)$$

is a Ξ'-consistent conjecture for t_i'. By construction, the support of messages sent by types t'_{-i} is the same as those sent by t_{-i} with the same universal preference types ϕ_{-i}, and by the inductive hypothesis these messages must be available to t'_{-i} as well. Integrating out messages, we would also arrive at π_i' being the marginal of μ' over $T'_{-i} \times A_i^+$. Moreover, integrating out the types, the marginal distribution over $M_{-i} \times A_i^+$ is the same for μ and μ' and equal to:

$$\sum_{\phi_{-i}\in \operatorname{supp}\phi_i(t_i,\mathcal{T})} \sigma_{-i}(dm_{-i}|\phi_{-i}, a_i)\widehat{\pi}_i(\phi_{-i}, a_i|t_i),$$

so that both t_i and t_i' have the same best reply, which must be m_i.

Now consider the fast deletion sequences $\{\Xi^{\lambda}(\mathcal{T},\mathcal{M})\}_{\lambda\in\Lambda}$ and $\{\Xi^{\lambda}(\mathcal{T}',\mathcal{M})\}_{\lambda\in\Lambda}$. Clearly, $\Xi_i^{\lambda_0}(t_i,\mathcal{T},\mathcal{M}) = M_i = \Xi_i^{\lambda_0}(t_i',\mathcal{T}',\mathcal{M})$ for all $t_i \in T_i$ and $t_i' \in T_i'$. Let $P(\lambda)$ be the property that for all $\lambda' < \lambda$, $\Xi_i^{\lambda'}(t_i,\mathcal{T},\mathcal{M}) = \Xi_i^{\lambda'}(t_i',\mathcal{T}',\mathcal{M})$ for all t_i, t_i' such that $\phi_i(t_i,\mathcal{T}) = \phi_i(t_i',\mathcal{T}')$, then $\Xi_i^{\lambda}(t_i,\mathcal{T},\mathcal{M}) = \Xi_i^{\lambda}(t_i',\mathcal{T}',\mathcal{M})$. We have just shown that if λ is a successor ordinal, then $P(\lambda-1)$ implies $P(\lambda)$. If λ is a limit ordinal, then $\Xi_i^{\lambda}(t_i,\mathcal{T},\mathcal{M}) = \cap_{\{\lambda'\in\Lambda|\lambda'<\lambda\}}\Xi_i^{\lambda'}(t_i,\mathcal{T},\mathcal{M})$. If t_i and t_i' have the same universal preference, then by the inductive hypothesis the sets in the intersections are the same for the two types, so obviously the intersections are the same as well, so that $P(\lambda)$ is true. Thus, $P(\lambda)$ holds for all $\lambda \in \Lambda$, and in particular $\Xi_i^*(t_i,\mathcal{T},\mathcal{M}) = \Xi_i^*(t_i',\mathcal{T}',\mathcal{M})$ if $\phi_i(t_i,\mathcal{T}) = \phi_i(t_i',\mathcal{T}')$. □

The first part of the argument, an inductive step, is essentially that of BMT. They show an even stronger result, that if t_i and t_i' have the same kth order preference, then $\Xi_i^k(t_i,\mathcal{T},\mathcal{M})$ is equal to $\Xi_i^k(t_i',\mathcal{T}',\mathcal{M})$.

This proposition implies the following important properties of equivalence:

Corollary 4.1 (Dependence on universal preferences)**.** *(a) $\mathcal{M} \sim_{(\mathcal{T},\epsilon)} \mathcal{M}'$ if and only if $\mathcal{M} \sim_{(\mathcal{T}',\epsilon)} \mathcal{M}'$ for every SBC subspace $\mathcal{T}'$ of $\mathcal{T}$. (b) If $\phi(\mathcal{T}) = \phi(\mathcal{T}')$, then $\mathcal{M} \sim_{(\mathcal{T},\epsilon)} \mathcal{M}'$ if and only if $\mathcal{M} \sim_{(\mathcal{T}',\epsilon)} \mathcal{M}'$.*

Proof of Corollary 4.1. For (a), note that for every type profile t in $\mathcal{T}$, there is a unique SBC subspace $\mathcal{T}'$ of $\mathcal{T}$ containing t such that $\phi_i(t_i,\mathcal{T}) = \phi_i(t_i,\mathcal{T}')$ for all i. Hence, Proposition 4.2 implies that $\Xi_i^*(t_i,\mathcal{T},\mathcal{M}) = \Xi_i^*(t_i,\mathcal{T}',\mathcal{M})$ and $\Xi_i^*(t_i,\mathcal{T},\mathcal{M}') = \Xi_i^*(t_i,\mathcal{T}',\mathcal{M}')$ for all i. Let $\eta_i(t_i')$ be the bijection from $\Xi_i^*(t_i',\mathcal{T}',\mathcal{M})$ to $\Xi_i^*(t_i',\mathcal{T}',\mathcal{M}')$ and $\alpha'(t)$ scaling proportions that satisfy the definition of ϵ-strategic equivalence for $\mathcal{M} \sim_{(\mathcal{T}',\epsilon)} \mathcal{M}'$. Letting $\eta_i(t_i) = \eta_i'(t_i)$ and $\alpha(t) = \alpha'(t)$, we have bijections from $\Xi_i^*(t_i,\mathcal{T},\mathcal{M})$ to $\Xi_i^*(t_i,\mathcal{T},\mathcal{M}')$ and scaling proportions that satisfies the definition of $\mathcal{M} \sim_{(\mathcal{T},\epsilon)} \mathcal{M}'$.

For part (b), let $\eta_i(t_i)$ and $\alpha(t)$ satisfy the definition of $\mathcal{M} \sim_{(\mathcal{T},\epsilon)} \mathcal{M}'$. For each $t_i' \in T_i(\mathcal{T}')$, we can find $\chi_i(t_i') = t_i \in T_i(\mathcal{T})$ with the same universal preference as t_i'. Thus, $\Xi_i^*(t_i', \mathcal{T}', \mathcal{M}) = \Xi_i^*(t_i, \mathcal{T}, \mathcal{M})$ and $\Xi_i^*(t_i', \mathcal{T}', \mathcal{M}') = \Xi_i^*(t_i, \mathcal{T}, \mathcal{M}')$. Setting $\eta_i'(t_i') = \eta_i(\chi_i(t_i'))$ and $\alpha'(t') = \alpha(\chi(t'))$, we have bijections between $\Xi_i^*(t_i', \mathcal{T}', \mathcal{M})$ and $\Xi_i^*(t_i', \mathcal{T}', \mathcal{M}')$ and a scaling proportion α' that satisfies the definition of $\mathcal{M} \sim_{(\mathcal{T}',\epsilon)} \mathcal{M}'$. □

Corollary 4.1 demonstrates that strategic equivalence of mechanisms only depends on higher-order preferences, and not on any other feature of the type space. In particular, two type spaces induce the same equivalence classes of mechanisms if they have the same higher-order preferences, and the equivalence classes for a given type space are simply the intersections of equivalence classes for belief-closed subspaces.

4.4 Extracting common knowledge

We now come to the heart of the chapter, in which the notion of strategic equivalence developed in the previous section is used to formalize when the designer can extract common knowledge via a mechanism that is similar to the desired mechanism. Formally, I will say that a mechanism $\mathcal{M}$ is *ϵ-uniformly equivalent* to the mechanism mapping $(\mathbf{T}, \mathbf{M})$ if for all $\mathcal{T} \in \mathbf{T}$, $\mathcal{M} \sim_{(\mathcal{T},\epsilon)} \mathbf{M}(\mathcal{T})$. If for every $\epsilon > 0$, a mechanism mapping $(\mathbf{T}, \mathbf{M})$ admits an ϵ-uniformly equivalent mechanism, then $(\mathbf{T}, \mathbf{M})$ is *uniformly virtually implementable.*

It will turn out that there are constraints on which mechanism mappings can be uniformly virtually implemented. In particular, a mechanism mapping must satisfy a local preference measurability condition (Theorem 4.1). In fact, local preference measurability is nearly sufficient for uniform virtual implementation, which is the subject of Theorem 4.2. The section will conclude with a discussion of the preservation of the set of Nash equilibria and other solution concepts by the uniform mechanism, which requires a natural regularity condition on the mechanism mapping.

4.4.1 Local preference measurability

The invariance of strategic equivalence to information other than higher-order preferences immediately suggests that there are limitations to the kinds of mechanism mappings that the designer can virtually implement. In particular, if $\mathbf{T}$ contains type spaces that map into the same subset of the universal preference space, then clearly any given uniform mechanism lies in the same equivalence class of mechanisms for both of these type spaces. Hence, if the uniform mechanism is equivalent to the desired mechanisms on these two type spaces, then the desired mechanisms must be equivalent as well.

Of course, restrictions on the mechanism mapping will be imposed even if the type spaces merely overlap in the universal preference space rather than coincide. I will say that a mechanism is *locally preference measurable* if the following property holds: for all $\mathcal{T}, \mathcal{T}' \in \mathbf{T}$, if there are SBC subspaces $\widehat{\mathcal{T}}$ and $\widehat{\mathcal{T}}'$ such that $\phi(\widehat{\mathcal{T}}) = \phi(\widehat{\mathcal{T}}')$, then:

$$\mathbf{M}(\mathcal{T}) \sim_{\widehat{\mathcal{T}}} \mathbf{M}(\mathcal{T}').$$

In other words, if the type spaces $\mathcal{T}$ and $\mathcal{T}'$ contain types with the same preference hierarchies, then $\mathbf{M}$ must specify mechanisms which are equivalent on the subspace of overlapping types.

The goal of this chapter is to show that a designer can implement a mechanism which is virtually equivalent to a given mechanism mapping, without having to know the type space $\mathcal{T}$.

I argue that local preference measurability is a necessary condition for uniform virtual implementation of a mechanism mapping:

Theorem 4.1 (Necessity)**.** *A mechanism mapping is uniformly virtually implementable only if it is locally preference measurable.*

Proof of Theorem 4.1. Suppose that $(\mathbf{T}, \mathbf{M})$ is uniformly virtually implementable. Then for every $\epsilon > 0$, there exists an ϵ-uniformly equivalent mechanism $\mathcal{M}$. Suppose $\mathcal{T}$ and $\mathcal{T}'$ are in $\mathbf{T}$ with SBC subspaces $\widehat{\mathcal{T}}$ and $\widehat{\mathcal{T}}'$, respectively, such that $\phi(\widehat{\mathcal{T}}) = \phi(\widehat{\mathcal{T}}')$. By Corollary 4.1 (a), $\mathcal{M} \sim_{(\mathcal{T}',\epsilon)} \mathbf{M}(\mathcal{T}')$ implies that $\mathcal{M} \sim_{(\widehat{\mathcal{T}}',\epsilon)} \mathbf{M}(\mathcal{T}')$, and (b) implies that $\mathcal{M} \sim_{(\widehat{\mathcal{T}},\epsilon)} \mathbf{M}(\mathcal{T}')$. We can similarly conclude that $\mathcal{M} \sim_{(\widehat{\mathcal{T}},\epsilon)} \mathbf{M}(\mathcal{T})$. Thus, Lemma 4.1 implies that $\mathbf{M}(\mathcal{T}) \sim_{(\widehat{\mathcal{T}},2\epsilon)} \mathbf{M}(\mathcal{T}')$. But this is true for every $\epsilon > 0$, so that $\mathbf{M}(\mathcal{T}) \sim_{\widehat{\mathcal{T}}} \mathbf{M}(\mathcal{T}')$. Hence, $(\mathbf{T}, \mathbf{M})$ is locally preference measurable. $\square$

4.4.2 Sufficiency

I now turn to the question of sufficient conditions for a mechanism mapping to be uniformly virtually implementable. It turns out that if there are at least three agents, local preference measurability is both necessary and sufficient. The constructive proof proceeds in three steps: First, I construct an explicit mechanism for eliciting preference hierarchies, in which the unique rationalizable message is truthful reporting of the preference hierarchy. Second, I show that a locally preference measurable mechanism mappings can be "reduced" to a minimal mapping for which type spaces in $\mathbf{T}$ do not overlap in the universal preference space. Such mechanism mappings are relatively easy to implement, and Lemma 4.2 shows that uniformly virtually implementation of the reduced mapping implies uniform virtual implementation of the original map as well. Finally, in Proposition 4.3, I construct a mechanism that ϵ-uniformly virtually implements a minimal mapping for $n \geq 3$.

This mechanism uses the principle that if a message profile is one deviation away from a unanimous report, there is a unique deviator who can be punished. Clearly, if there are only two agents whose reports disagree, then neither is the sole deviator. In Proposition 4.4, I introduce another assumption, joint worst-outcome feasibility, that restores sufficiency of local preference measurability in the case of two agents. The purpose is obvious: if agents' reports disagree, then both will be punished, pressuring them to agree in equilibrium. All sufficiency results are summarized in Theorem 4.2.

A mechanism for eliciting preferences

Consider an agent who has a state dependent preference over acts that map a finite observable state space X into lotteries, with the preference represented by $\pi_i \in \Delta(X \times A_i^+)$. Suppose we run the following one-player mechanism: the agent reports a preference $\widehat{\pi}_i$, and based on this report and the realized state x, we will implement the lottery:

$$\xi_i(a_i|x, \widehat{\pi}_i, X) = \frac{1}{|A_i^+|}\left(\frac{\widehat{\pi}_i(x, a_i)}{2} + \frac{1}{2} - \sum_{\substack{x' \in X, \\ a_i' \in A_i^+}} \frac{\widehat{\pi}_i(x', a_i')^2}{4}\right),$$

for $a_i \in A_i^+$, with the complementary probability on $\underline{a}_i^i$. In words, if the agent reports $\widehat{\pi}_i(x, a_i)$, then conditional on x being realized, outcome a_i will be implemented with probability $\frac{\widehat{\pi}_i(x,a_i)}{2(|A_i^+|)}$. With probability $\frac{1}{2} - \sum_{(x',a')} \frac{\widehat{\pi}_i(x',a_i')^2}{4}$, we will pick an outcome in A_i^+ to implement at random. Otherwise, $\underline{a}_i^i$ is implemented. Note that the $\widehat{\pi}_i$ must sum to one, and since they are numbers in $[0,1]$, $\sum_{x',a_i'\neq\underline{a}_i^i} \frac{\widehat{\pi}_i(x',a_i')^2}{4} \leq \frac{1}{2}$, so these probabilities sum to less than one and our lottery is always well-defined.

If the agent's preferences are represented by a true $\pi_i(x,a_i)$, then recalling that $\pi_i(x,\underline{a}_i^i) = 0$, the expected utility from a report of $\widehat{\pi}_i$ is:

$$\frac{1}{|A_i^+|} \sum_{\substack{x\in X,\\ a_i\in A_i^+}} \frac{\widehat{\pi}_i(x,a_i)\pi_i(x,a_i)}{2} + \left(\frac{1}{2} - \sum_{\substack{x\in X,\\ a_i\in A_i^+}} \frac{\widehat{\pi}_i(x,a_i)^2}{4}\right) \frac{1}{|A_i^+|} \sum_{\substack{x\in X,\\ a_i\in A_i^+}} \pi_i(x,a_i).$$

Since the $\widehat{\pi}_i(x,a_i)$ sum to 1, this expression simplifies to:

$$\frac{1}{|A_i^+|} \left(\frac{1}{2} + \frac{1}{2} \sum_{\substack{x\in X,\\ a_i\in A_i^+}} \left(\widehat{\pi}_i(x,a_i)\pi_i(x,a_i) - \frac{\widehat{\pi}_i(x,a_i)^2}{2}\right)\right).$$

Differentiating with respect to $\widehat{\pi}_i(x,a_i)$, we can see that it is strictly optimal for agent i to report the true preference $\widehat{\pi}_i = \pi_i$. Note that this result does not depend on the particular representation for agent i's preferences.

The structure of this choice over lotteries has a marked similarity to the quadratic scoring rule used in experimental economics (Brier, 1950). We can interpret $\pi_i(x,a_i)$ as the "probability" that the outcome (x,a_i) will obtain. If it does, the designer will reward the agent with a fraction $\widehat{\pi}_i(x,a_i)$ of a dollar. This is traded off against a certain loss of a fraction $\frac{\widehat{\pi}_i(x,a_i)^2}{2}$ of a dollar. In a bet so structured, it is optimal to wager precisely the the subjective probability of the outcome (x,a_i) occurring. Such scoring rules have been used in a mechanism design context by Azar, Chen, and Micali (2012) and also in Chapter 1 to elicit beliefs in quasilinear environments, so that agents could in fact be rewarded with money if a given probabilistic outcome obtained. The same technique works with non-quasilinear and subjective expected utility preferences, where instead of using money, the agent is rewarded with shares of the subjective utility of a_i if x occurs, traded off against the average expected utility across all outcomes.

Using such lotteries, we can construct a mechanism to elicit agents' higher-order preferences. Formally, agent i's message space M_i^* in this mechanism will be the set of (possibly incoherent) preference hierarchies in X_i^* that have finite support. For a message profile $m^* \in M^*$, the outcome lottery is:

$$g^*(m^*) = \frac{1}{n}\sum_{i\in N}\left(\frac{1}{2}\xi_i\left(\emptyset, m_i^{*,1}, \{\emptyset\}\right) + \sum_{k=2}^{\infty}\frac{1}{2^k}\xi_i\left(m_{-i}^{*,k-1}, m_i^{*,k}, M_{-i}^{*,k-1}\right)\right),$$

where $m_i^{*,k}$ is agent i's kth order preference report. The mechanism will be referred to as $\mathcal{M}^* = (\{M_i^*\}_{i\in N}, g^*)$.

I claim that for any type space $\mathcal{T}$, the unique rationalizable strategy profile is to always report $m_i^* = \phi_i(t_i, \mathcal{T})$. The argument is inductive. For the base case, observe that agent i's first-order report $m_i^{*,1}$ only affects the first term inside the parentheses. Conditional on agent i's report being used, the outcome is independent of other agents' reports. Thus, for any Ξ^0-consistent conjecture μ:

$$\begin{aligned}
&\sum_{\substack{t_{-i}\in T_{-i},\\ m^*_{-i}\in M^*_{-i}}} \sum_{a_i\in A_i^+} \xi_i(a_i|\emptyset, m_i^{*,1}, \{\emptyset\})\mu(t_{-i}, m^*_{-i}, a_i) \\
&\qquad = \sum_{a_i\in A_i^+} \xi_i(a_i|\emptyset, m_i^{*,1}, \{\emptyset\}) \sum_{\substack{t_{-i}\in T_{-i},\\ m^*_{-i}\in M^*_{-i}}} \mu(t_{-i}, m^*_{-i}, a_i) \\
&\qquad = \sum_{a_i\in A_i^+} \xi_i(a_i|\emptyset, m_i^{*,1}, \{\emptyset\}) \sum_{t_{-i}\in T_{-i}} \pi_i(t_{-i}, a_i|t_i) \\
&\qquad = \sum_{a_i\in A_i^+} \xi_i(a_i|\emptyset, m_i^{*,1}, \{\emptyset\})\phi_i^1(a_i, t_i, \mathcal{T}).
\end{aligned}$$

Thus, as derived above, the optimal strategy is to report $m_i^{*,1} = \phi_i^1(t_i, \mathcal{T})$. For the inductive step, suppose that for all message profiles in Ξ_i^{k-1}, agents $-i$ are reporting $m_{-i}^{*,k-1} = \phi_{-i}^{k-1}(t_{-i}, \mathcal{T})$. Again, conditional on the outcome depending on $\widehat{\pi}_i^k$, the outcome is independent of m^*_{-i} except for $m_{-i}^{*,k-1}$. But since this is reported truthfully, agent i's preferences over acts on $m_{-i}^{*,k-1}$ is represented by precisely by $\phi_i^k(t_i, \mathcal{T})$. Formally, if μ is a Ξ^{k-1}-consistent conjecture, then agent i will choose $m_i^{*,k}$ to maximize:

$$\begin{aligned}
&\sum_{\substack{t_{-i}\in T_{-i},\\ m^*_{-i}\in M^*_{-i}}} \sum_{a_i\in A_i^+} \xi_i(a_i|m_{-i}^{*,k-1}, m_i^{*,k}, M_{-i}^{*,k-1})\mu(t_{-i}, m^*_{-i}, a_i) \\
&\qquad = \sum_{a_i\in A_i^+} \sum_{m_{-i}^{*,k-1}\in M_{-i}^{*,k-1}} \xi_i(a_i|m_{-i}^{*,k-1}, m_i^{*,k}, M_{-i}^{*,k-1}) \sum_{\left\{t_{-i}\in T_{-i} \middle| \substack{\phi_{-i}^{k-1}(t_{-i})\\ =m_{-i}^{*,k-1}}\right\}} \mu(t_{-i}, m^*_{-i}, a_i) \\
&\qquad = \sum_{a_i\in A_i^+} \sum_{m_{-i}^{*,k-1}\in M_{-i}^{*,k-1}} \xi_i(a_i|m_{-i}^{*,k-1}, m_i^{*,k}, M_{-i}^{*,k-1})\phi_i^k(m_{-i}^{*,k-1}, a_i, t_i, \mathcal{T}),
\end{aligned}$$

which is maximized by reporting $m_i^{*,k} = \phi_i^k(t_i, \mathcal{T})$. Thus, after ω^2 rounds of deletion, we conclude that the only rationalizable message has $m_i^{*,k} = \phi_i^k(t_i, \mathcal{T})$ for all k, i.e., $m_i^* = \phi_i(t_i, \mathcal{T})$.

BMT also construct mechanisms for the purpose of separating types with different preference hierarchies, though there are some important differences. First, BMT consider environments more general than the present private-good setting, in which agent i might care about the outcomes which are implemented to incentivize agents $-i$ to truthfully report their preferences. This contaminates agent i's incentives to report his kth order preference, because of concern over how that report will affect the outcome to incentivize agents $-i$'s $k+1$th order preferences. In contrast, $\mathcal{M}^*$ fixes agent i's allocation at 0_i whenever agents $-i$ are being incentivized in order to avoid this contamination. In addition, BMT work with only finite mechanisms, which precludes a single "catch-all" elicitation mechanism for possibly infinitely many different preference hierarchies. BMT finesse both issues by coarsening the mechanism to separate preference hierarchies that are sufficiently far apart in a metric compatible with the product topology on X^*. They also make use of a trick from AM, of rapidly scaling down the relative probability of lotteries to incentivize revelation of higher-order preferences, so that the contamination effect is swamped by a discrete benefit from reporting kth order preferences truthfully.

Minimal mechanism mappings

A mechanism mapping may have type spaces in its domain that repeat preference hierarchies. For locally preference measurable mappings, such information is redundant in that only one such type space would be necessary to define the

[2] ω is the first infinite ordinal number, which is order isomorphic to the natural numbers.

"equivalence class" of strategically equivalent mechanisms that the designer would like to implement for those preference hierarchies. A minimal mechanism mapping is non-redundant in that the domain contains only SBC type spaces. Note that local preference measurability is automatically satisfied for such mappings.

Formally, a mechanism mapping $(\mathbf{T}, \mathbf{M})$ is *minimal* if $\mathcal{T}$ is an SBC subspace of T^* for every $\mathcal{T} \in \mathbf{T}$. I will say that $(\widetilde{\mathbf{T}}, \widetilde{\mathbf{M}})$ is a *reduced mechanism mapping* for the locally preference measurable mapping $(\mathbf{T}, \mathbf{M})$ if $\widetilde{\mathbf{T}}$ is the set of smallest belief-closed subsets of T^* that are mapped to by types in type spaces in $\mathbf{T}$:

$$\widetilde{\mathbf{T}} = \{\phi(\mathcal{T}') | \mathcal{T}' \text{ is an SBC subspace of } \mathcal{T} \in \mathbf{T}\},$$

and if $\widetilde{\mathbf{M}}(\mathcal{T})$ is equal to some $\mathbf{M}(\mathcal{T}')$ with $\mathcal{T}' \in \mathbf{T}$ for which $\mathcal{T} \subseteq \phi(\mathcal{T}')$.

To argue the sufficiency results of Propositions 4.3 and 4.4, I will construct ϵ-uniformly equivalent mechanisms for minimal mechanism mappings, which will then imply sufficiency results for general mechanism mappings. The reason is that any locally preference measurable mechanism mapping can be associated with a minimal mapping, and as Lemma 4.2 shows below, uniformly virtual implementability of the reduced mapping implies uniform virtual implementability of the original mapping.

Lemma 4.2 (Reduced mechanism mappings)**.** *If* $(\widetilde{\mathbf{T}}, \widetilde{\mathbf{M}})$ *is a reduced mechanism mapping for the locally preference measurable* $(\mathbf{T}, \mathbf{M})$ *and if* $\mathcal{M}$ *is* ϵ*-uniformly equivalent to* $(\widetilde{\mathbf{T}}, \widetilde{\mathbf{M}})$*, then* $\mathcal{M}$ *is also* ϵ*-uniformly equivalent to* $(\mathbf{T}, \mathbf{M})$*.*

Proof of Lemma 4.2. Take any $\mathcal{T} \in \mathbf{T}$ and $\mathcal{T}'$ any SBC subspace of $\mathcal{T}$. By assumption, $\mathcal{M} \sim_{(\phi(\mathcal{T}'),\epsilon)} \widetilde{\mathbf{M}}(\phi(\mathcal{T}'))$, so by Corollary 4.1 (b), $\mathcal{M} \sim_{(\mathcal{T}',\epsilon)} \widetilde{\mathbf{M}}(\phi(\mathcal{T}'))$. But the latter mechanism is equal to $\mathbf{M}(\mathcal{T}'')$ for some $\mathcal{T}'' \in \mathbf{T}$ such that $\mathcal{T}' \subseteq \phi(\mathcal{T}'')$. By local preference measurability, it must be that $\mathbf{M}(\mathcal{T}) \sim_{\mathcal{T}'} \mathbf{M}(\mathcal{T}'')$, so by Lemma 4.1, we conclude that $\mathcal{M} \sim_{(\mathcal{T}',\epsilon)} \mathbf{M}(\mathcal{T})$. But this must be true for every SBC subspace $\mathcal{T}'$ of $\mathcal{T}$, so by Corollary 4.1 (a), $\mathcal{M} \sim_{(\mathcal{T},\epsilon)} \mathbf{M}(\mathcal{T})$. □

The convenient feature of the mechanism mapping $(\widetilde{\mathbf{T}}, \widetilde{\mathbf{M}})$ is that a given preference hierarchy appears in a single type space in $\widetilde{\mathbf{T}}$, and therefore is associated with a unique mechanism. This will simplify the exposition to follow, in which mechanisms are constructed that uniformly virtually implement minimal mechanism mappings.

A uniform equivalent mechanism for $n \geq 3$

I now show that as long as there are at least three agents, every minimal mechanism mapping $(\mathbf{T}, \mathbf{M})$ is uniformly virtually implementable. For a given $\epsilon > 0$, I construct an ϵ-uniformly equivalent mechanism as follows. Agent i's message space will be:

$$M_i(\mathcal{M}) = M_i^* \times \cup_{\widehat{\mathcal{M}} \in \mathbf{M}(\mathbf{T})} \left(\left\{ \widehat{\mathcal{M}} \right\} \times M_i \left(\widehat{\mathcal{M}} \right) \right).$$

In other words, agent i's messages are triples consisting of a preference hierarchy, a "suggested" mechanism in the image of $\mathbf{M}$, and a message in the suggested mechanism. A typical message will be written as $m_i = (m_i^*, \widehat{\mathcal{M}}_i, \widehat{m}_i)$.

Let $\phi(m_i^*)$ be the smallest belief-closed subset of T^* containing m_i^*, and let $\mathbf{M}(m_i^*) = \mathbf{M}(\phi(m_i^*))$ if $\phi(m_i^*) = \phi(\mathcal{T})$ for some $\mathcal{T} \in \mathbf{T}$, and let $\mathbf{M}(m_i^*) = \emptyset$ otherwise. Note that $\mathbf{M}(m_i^*)$ is well-defined since $(\mathbf{T}, \mathbf{M})$ is minimal. We can think of this as being the mechanism which is "implied" by agent i's reported beliefs, in that those beliefs lie in a SBC type space $\mathcal{T}$ which maps to the given mechanism.

The outcome function will consist of three pieces:

$$g(m)=\frac{\epsilon}{4}g^*(m^*)+\frac{\epsilon}{4n}\sum_{i\in N}\widehat{g}^i(m)+\left(1-\frac{\epsilon}{2}\right)\widehat{g}(m)$$

where $\epsilon\in(0,1)$. Thus, $\widehat{g}(m)$ is the lion's share of the outcome, with the $g^*(m^*)$ component incentivizing truthful reporting of beliefs and the $\widehat{g}^i$ portion providing a slight nudge towards consensus, as we shall see below.

To define $\widehat{g}(m)$, I will use the following classification of message profiles. Let M^0 be the message profiles in which $\widehat{\mathcal{M}}_i=\widehat{\mathcal{M}}_j=\mathbf{M}(m_i^*)=\mathbf{M}(m_j^*)$ for all i and j. These are the profiles in which all agents are unanimous in their report of the mechanism and the mechanism "implied" by their higher-order preferences.

Let $M^{i,1}$ be the message profiles for which $\widehat{\mathcal{M}}_i=\widehat{\mathcal{M}}_j=\mathbf{M}(m_j^*)\neq\mathbf{M}(m_i^*)$ for all $j\neq i$, which is disjoint from M^0. For these message profiles, the report is unanimous as in M^0, except for the single deviator agent i who deviates in his report of the implied mechanism $\mathbf{M}(m_i^*)$. Note that a message profile can be in $M^{i,1}$ for at most one i.

Let $M^{i,2}$ be the message profiles in which $\widehat{\mathcal{M}}_i\neq\widehat{\mathcal{M}}_j=\mathbf{M}(m_j^*)=\widehat{\mathcal{M}}_k=\mathbf{M}(m_k^*)$ for all $j\neq i\neq k$. These are profiles in which, again, agent i is dissenting from an otherwise unanimous report, though here the dissent must include a different suggested mechanism, as well as possibly a different implied mechanism. Note that if $n>2$, a given profile can be in $M^{i,2}$ for at most one i since otherwise it cannot be that both $-i$ and $-j$ are unanimous in their reports, but if $n=2$, we can have profiles in $M^{1,2}\cap M^{2,2}$. Throughout the rest of this subsection, I assume that $n>2$, with the $n=2$ case being treated in the next subsection.

The lottery $\widehat{g}(m)$ is equal to $g\left(\widehat{m},\widehat{\mathcal{M}}\right)$ if $m\in M^0$ and $\widehat{\mathcal{M}}$ is the unanimously reported mechanism. If $m\in M^{i,1}$, then $\widehat{g}(m)$ implements outcome $(a_i,0_{-i})$ with probability $\sum_{a_{-i}\in A_{-i}} g\left(a_i, a_{-i}|\widehat{m}, \widehat{\mathcal{M}}\right)$; in other words, $\widehat{g}$ implements the same marginal lottery over A_i as in $g(\widehat{m}, \widehat{\mathcal{M}})$, but the outcomes for agents in $-i$ are set equal to 0_{-i}. If $m \in M^{i,2}$ for some i, then we implement the outcome $(\underline{a}_i^i, 0_{-i})$ with probability one. Finally, $\widehat{g}(m)$ puts probability one on 0 for all other message profiles.

To define $\widehat{g}^i(m)$, let $M^{i,3}$ be the set of profiles in which $\widehat{\mathcal{M}}_i = \mathbf{M}(m_j^*) = \mathbf{M}(m_k^*)$ for all $j, k \in -i$. These are message profiles in which agents $-i$ are unanimous in their implied mechanisms, which agree with agent i's suggested mechanism. For $m \in M^{i,3}$, $\widehat{g}^i(m)$ puts equal probability $\frac{1}{|A_i^+|}$ on $(a_i, 0_{-i})$ for every $a_i \in A_i^+$, and otherwise $\widehat{g}^i(m)$ puts probability one on $(\underline{a}_i^i, 0_{-i})$. Note that agent i only cares about the lottery $\widehat{g}^i$ and not $\widehat{g}^{-i}$, and that the lottery $\widehat{g}^i(m)$ with $m \in M^{i,3}$ is strictly preferred to the outcome with $m \notin M^{i,3}$. This completes the specification of $g(m)$.

Proposition 4.3 (Sufficiency for minimal mappings, $n \geq 3$)**.** *$\mathcal{M} = (M, g)$ is ϵ-uniformly equivalent to the minimal mechanism mapping $(\mathbf{T}, \mathbf{M})$.*

Proof of Proposition 4.3. For a given $\mathcal{T} \in \mathbf{T}$, I will construct a deletion sequence for $\mathcal{M}$, and show that $\mathcal{M} \sim_{(\mathcal{T},\epsilon)} \mathbf{M}(\mathcal{T})$.

To start, observe that the report of m_i^* does not affect the marginal lottery of g on A_i except through $g^*(m^*)$: $\widehat{g}^i$ only depends on $\widehat{\mathcal{M}}_i$ and m_{-i}^*, and $\widehat{g}_i$ depends on $\widehat{\mathcal{M}}$, $\widehat{m}$, and m_{-i}^*. This is because the marginal lottery of $\widehat{g}$ on A_i is the same for $m \in M^0$ and $m' \in M^{i,1}$, as long as the profiles are the same except for m_i^*. Thus, for any type $t_i \in T_i^*$, any strategy which is rationalizable must have $m_i^* = t_i$. In particular, we can construct the first ω terms of a deletion sequence $\{\Xi^\lambda\}_{\lambda\in\omega}$ in which we rule out $m_i^* \neq t_i$. Thus, $\mathbf{M}(m_i^*) = \mathbf{M}(\mathcal{T})$ for every i and for every rationalizable message m_i. Note that this relies on Assumption 4.3, which implies that all agents types must lie in the same SBC subspace of T^*.

Let m_i be some message in which $\widehat{\mathcal{M}}_i \neq \mathbf{M}(\mathcal{T})$. I claim that this message is strictly dominated by any report in which $\widehat{\mathcal{M}}_i = \mathbf{M}(\mathcal{T})$, and therefore m_i is not rationalizable given Ξ^ω. First, if $\widehat{\mathcal{M}}_j = \mathbf{M}(\mathcal{T})$ for every $j \neq i$, then $m \in M^{i,2}$ and $m \notin M^{i,3}$ and $(\underline{a}_i^i, 0_{-i})$ is implemented if either $\widehat{g}^i$ or $\widehat{g}$ is used, whereas deviating to $m_i' = (t_i, \mathbf{M}(\mathcal{T}), \widehat{m}_i)$ with $\widehat{m}_i' \in M_i(\mathbf{M}(\mathcal{T}))$ must induce a weakly better outcome when $\widehat{g}$ is used (since we are already implementing the worst outcome for agent i) and yields a strictly better outcome when $\widehat{g}^i$ is used, since $m' \in M^{i,3}$. On the other hand, if $\widehat{\mathcal{M}}_j \neq \mathbf{M}(\mathcal{T})$ for some $j \neq i$, then $\widehat{g}$ implements the outcome 0 with probability 1 and $\widehat{g}^i$ implements $(\underline{a}_i^i, 0_{-i})$ with probability 1. By switching to any m_i' as described above, the new message profile will still result in the outcome 0_i if $\widehat{g}$ is used, since $m' \notin M^0 \cup M^{i,1}$, but again results in a strictly better lottery being implemented if $\widehat{g}^i$ is used. Hence, after $\omega + 1$ rounds of deletion, we rule out any message in which $\widehat{\mathcal{M}}_i \neq \mathbf{M}(\mathcal{T})$, and any rationalizable message profile in $\Xi_i^{\omega+1}(t_i)$ must be of the form $(t_i, \mathbf{M}(\mathcal{T}), \widehat{m}_i)$ with $\widehat{m}_i \in M_i(\mathbf{M}(\mathcal{T}))$.

As a result, we can define a bijective mapping $\eta_i(t_i) : \Xi_i^{\omega+1}(t_i) \to M_i(\mathbf{M}(\mathcal{T}))$ with $\eta_i(t_i, m_i) = \widehat{m}_i$, which is just projection onto the third coordinate. Letting $\{\widehat{\Xi}\}_{\lambda \in \widehat{\Lambda}}$ be any deletion sequence for $(\mathcal{T}, \mathbf{M}(\mathcal{T}))$, we can define the next Λ elements of the deletion sequence for $(\mathcal{T}, \mathcal{M})$ as:

$$\Xi^{\omega+1+\lambda}(t_i) = \eta_i^{-1}\left(\widehat{\Xi}^\lambda(t_i), t_i\right).$$

I claim that $\{\Xi^\lambda\}_{\lambda \in \Lambda}$ so defined, where $\Lambda = \omega + 1 + \widehat{\Lambda}$, is a deletion sequence for $\mathcal{M}$. It suffices to show that for $\lambda \geq \omega + 1$, $\Xi^{\lambda+1} \supseteq R(\Xi^\lambda, \mathcal{T}, \mathcal{M})$ and $\Xi^* = R(\Xi^*, \mathcal{T}, \mathcal{M})$. This follows from the fact that any Ξ^λ-consistent conjecture μ

in $\Delta(\text{graph}\,\Xi^{\lambda}_{-i} \times A_i^+)$ can be identified with a $\Xi^{\omega+1+\lambda}$-consistent conjecture μ' in $\Delta(\text{graph}\,\Xi^{\omega+1+\lambda} \times A_i^+)$, which is the push-forward measure induced by the mappings $\eta_i(t_i)$:

$$\mu'(t_{-i}, m_{-i}, a_i) = \mu(t_{-i}, \widehat{m}_{-i}, a_{-i}).$$

Thus, a message m_i in $\Xi_i^{\omega+1+\lambda}(t_i)$ is rationalizable for agent i given $\Xi^{\omega+1+\lambda}$ if and only if $\widehat{m}_i$ is rationalizable for agent i given Ξ^{λ}. Finally, the mappings $\eta_i(t_i)$ restricted to $\Xi^*(\mathcal{T}, \mathcal{M})$ satisfy the definition of strategic equivalence, since in equilibrium:

$$g(m) - g(m') = (1-\epsilon)\left[g(\widehat{m}, \mathbf{M}(\mathcal{T})) - g(\widehat{m}', \mathbf{M}(\mathcal{T}))\right]$$

and:

$$g(m) - g(\widehat{m}, \mathbf{M}(\mathcal{T})) = \frac{\epsilon}{2}\left(\frac{1}{2}g^*(m^*) + \frac{1}{2n}\sum_{i\in N}\widehat{g}^i(m) - g(\widehat{m}, \mathbf{M}(\mathcal{T}))\right),$$

for which the norm is less than ϵ. □

Here is some intuition for why the mechanism works. Observe that an agent's own outcome does not depend on their own reported beliefs except through the reward from g^*, so agents have the same incentives to report their preferences truthfully as they do in the mechanism $\mathcal{M}^*$. But once we know that all agents will report their higher-order preferences truthfully, agent i's suggested mechanism can be compared to other agents' preference hierarchies to determine if agent i is suggesting the correct mechanism. If more than one agent has an incorrect suggestion by this comparison, $\widehat{g}$ implements a fixed outcome, but there is a "slap on the wrist" through $\widehat{g}^i$ that encourages agents to switch to suggesting the correct mechanism. If only one agent disagrees, then in addition a harsh outcome is implemented for that agent by $\widehat{g}$ so that it is better to agree, and receive the outcome that will be decided by mechanism $\mathbf{M}(\mathcal{T})$.

4.4.3 The case of $n = 2$

The mechanism of the previous subsection relies on an important property: With $n \geq 3$, if a profile m is one deviation away from unanimity, then there is a unique agent to whom this deviation belongs and who is, in a sense, the sole "deviator". This is fortunate, because it may be that it is impossible to punish more than one agent simultaneously if there is no $a \in A$ such that $a_i = \underline{a}_i^i$ and $a_j = \underline{a}_j^j$. When there are only two agents who make inconsistent reports, neither agent could be considered the sole deviator.

This suggests a simple condition on A which, together with local preference measurability, will be sufficient for uniform virtual implementation of a minimal mechanism mapping $(\mathbf{T}, \mathbf{M})$. The worst-outcome is jointly feasible if $(\underline{a}_1^1, \ldots, \underline{a}_n^n) \in A$.

Proposition 4.4 (Sufficiency with joint worst-outcome feasibility)**.** *Suppose that worst outcomes are jointly feasible. Then every minimal mechanism mapping* $(\mathbf{T}, \mathbf{M})$ *is uniformly virtually implementable.*

Proof of Proposition 4.4. The mechanism is identical to that constructed for the proof of Proposition 4.3, except that we now implement the outcome a in which $a_i = \underline{a}_i^i$ if $m \in M^{i,2}$ and $a_i = 0_i$ otherwise. The rest of the proof proceeds as is. □

Joint feasibility of the worst-outcome is a natural assumption in many situations, as in the quasilinear auction example discussed previously. In that case, it was always feasible to not allocate any of the goods to agent i and impose the largest possible transfer $\bar{t}$. However, it may very well be the case that some of the objects are undesirable for agent i but cannot be freely disposed of. For example, the worst outcome might be the assignment of an onerous task, which only one agent can be required to perform at a time.

Summing up, the results of Lemma 4.2 and Propositions 4.3 and 4.4 imply the following general sufficiency result:

Theorem 4.2 (Sufficiency)**.** *A mechanism mapping* $(\mathbf{T}, \mathbf{M})$ *is uniformly virtually implementable if it is locally preference measurable and at least one of the following holds: (a)* $n \geq 3$*, or (b) the worst outcomes are jointly feasible.*

4.4.4 Preservation of Nash equilibria

To allow for mechanism mappings with infinite domains, I have used an infinite mechanism that always exactly extracts agents' higher-order preferences. This mechanism is similar to the one specified by the designer in the sense of ϵ-strategic equivalence; rationalizable messages in the uniform mechanism can be identified with rationalizable messages of the type space-specific mechanism so that change in outcome for a change in message profile is proportional, and outcomes for identified message profiles vary by no more than ϵ. The proportional differences property seems to preserve the strategic calculus in a strong sense: the induced preferences over rationalizable message profiles are the same under strategically equivalent mechanisms.

Unfortunately, this is not a sufficient guarantee that agents will behave the same way in the two mechanisms. Strange pathologies can arise with transfinite iterated deletion of never best replies, as described by Dufwenberg and Stegeman (2002) and CLL in complete information settings. For example, there may be an infinite set of never best replies, each of which is dominated by a message which is also a never best reply, but the eliminated messages are better replies than some message that is rationalizable. Thus, the game in which messages are restricted to the rationalizable correspondence may have more Nash equilibria than the full mechanism.

The construction of the previous section is resistant to this phenomenon: given that others will report their preference hierarchies and their suggested mechanisms truthfully, reporting one's true preferences and suggesting the correct mechanism strictly dominates any other message. Thus, if a best reply exists, it must always involve $m_i^* = \phi_i(t_i, \mathcal{T})$ and $\widehat{\mathcal{M}}_i = \widetilde{\mathbf{M}}(\phi(\mathcal{T}))$, where the latter is the mechanism specified by the reduced mechanism mapping. And yet, if $\widetilde{\mathbf{M}}(\phi(\mathcal{T}))$ and $\mathbf{M}(\mathcal{T})$ are infinite mechanisms, then the set of Nash equilibria may still be different for the two mechanisms.

There is a natural restriction on mechanisms which will rule out this kind of behavior and guarantee that the set of Nash equilibria will be preserved under strategic equivalence. Let us say that a mechanism is *regular* if for every Ξ^*-consistent conjecture μ and every $m_i \notin \Xi_i^*(t_i)$, there exists an $m_i' \in \Xi_i^*(t_i)$ that is a better reply to μ than m_i. A mechanism mapping $(\mathbf{T}, \mathbf{M})$ is *regular* if $\mathbf{M}(\mathcal{T})$ is regular for every $\mathcal{T} \in \mathbf{T}$. Further, let us say that an ϵ-uniformly equivalent mechanism $\mathcal{M}$ is *Nash preserving* if for every equilibrium strategy profile σ for $(\mathcal{T}, \mathbf{M}(\mathcal{T}))$, the strategy profile σ' of $(\mathcal{T}, \mathcal{M})$ in which $\sigma_i'(t_i, m_i') = \sigma_i(t_i, \eta_i(t_i, m_i'))$ is a Nash equilibrium of $(\mathcal{T}, \mathcal{M})$, where η is the outcome preserving bijection from $\mathcal{M}$ to $\mathbf{M}(\mathcal{T})$. A mechanism mapping is *Nash preserving uniformly virtually implementable* if for every $\epsilon > 0$, there exists an ϵ-uniformly equivalent Nash preserving mechanism $\mathcal{M}$.

Corollary 4.2. *A regular mechanism mapping is Nash preserving uniformly virtually implementable if the mechanism mapping is locally preference measurable and if either (a) $n \geq 3$ or (b) the worst outcome is jointly feasible.*

Proof of Corollary 4.2. First, I show that if $\mathcal{M} \sim_{(\mathcal{T},\epsilon)} \mathcal{M}'$ where $\mathcal{M} = (M, g)$ and $\mathcal{M}' = (M', g')$ are regular, then σ is a Nash equilibrium of $(\mathcal{T}, \mathcal{M})$ only if σ' defined by $\sigma_i'(t_i, \eta_i(t_i, m_i)) = \sigma_i(t_i, m_i)$ is a Nash equilibrium of $(\mathcal{T}, \mathcal{M}')$, where η_i is an ϵ-outcome preserving bijection from $\mathcal{M}$ to $\mathcal{M}'$. Since σ is a Nash equilibrium of $\mathcal{M}$, $\sigma_i(t_i)$ must have support in $\Xi_i^*(t_i, \mathcal{T}, \mathcal{M})$ for all i and t_i. Moreover, it must be that:

$$\sum_{t_{-i} \in T_{-i}} \int_{m \in M} \sum_{a_i \in A_i^+} (g_i(a_i|m) - g_i(a_i|m_i', m_{-i}))\sigma(dm|t_i, t_{-i})\pi(t_{-i}, a_i|t_i) \leq 0$$

for all t_i and $m_i' \in M_i$. In particular, this holds for all $m_i \in \Xi_i^*(t_i, \mathcal{T}, \mathcal{M})$. By the proportionality of differences, there are $\alpha(t) > 0$ so that:

$$g_i(a_i|m) - g_i(a_i|m_i', m_{-i}) = \alpha(t)[g_i'(a_i|\eta(t, m)) - g_i'(a_i, \eta(t, m_i', m_{-i}))].$$

Thus, $\eta_i(t_i, m_i') = m_i'' \in \Xi_i^*(t_i, \mathcal{T}, \mathcal{M}')$ is never a better reply. But if there is an $\widehat{m}_i \in M_i' \setminus \Xi_i^*(t_i, \mathcal{T}, \mathcal{M}')$ which is a better reply than $\eta_i(t_i, m_i)$, by regularity we can find an $m_i'' \in \Xi_i^*(t_i, \mathcal{T}, \mathcal{M}')$ that is better than $\widehat{m}_i$, and therefore $\eta_i(t_i, m_i)$, a contradiction. Hence, m_i' is a best reply for every m_i' in the support of $\sigma_i'(t_i)$.

It only remains to show that if $\mathbf{M}(\mathcal{T})$ is regular for every $\mathcal{T}$, then the mechanisms constructed in Propositions 4.3 and 4.4 will be regular as well. But this follows almost directly from the observation that if agents $-i$ report $m^*_{-i} = t_{-i}$ and $\widehat{\mathcal{M}}_{-i} = \widetilde{\mathbf{M}}(\phi(\mathcal{T}))$, then any report which does not involve $m^*_i = t_i$ and $\widehat{\mathcal{M}}_i = \widetilde{\mathbf{M}}(\phi(\mathcal{T}))$ is strictly dominated. Remaining actions are isomorphic to $\widetilde{\mathbf{M}}(\mathcal{T})$ in that there are proportional differences, so regularity of $\widetilde{\mathbf{M}}(\mathcal{T})$ implies that any message not in $\Xi^*_i(t_i, \mathcal{T}, \mathcal{M})$ is dominated by a message in that set. □

Similar results hold for other refinements of interim preference correlated rationalizability in addition to Bayesian Nash equilibrium, such as interim correlated rationalizability (Dekel, Fudenberg, and Morris, 2007), or interim independent rationalizability á la Bernheim (1984) and Pearce (1984). The bottom line is that the uniform mechanism of Theorem 4.2 introduces extra messages which are never optimal, or even near optimal, given that others use rationalizable messages.

I will conclude this section by reconsidering the private-good auction of the Introduction and of Chapters 1 and 2. Let us suppose further that beliefs are consistent with a common prior, i.e., there exists a joint distribution $\pi \in \Delta(\times_{i \in N} A_i \times T)$, so that π_i is the conditional distribution of (t_{-i}, a_i) given t_i:

$$\pi_i(t_{-i}, a_i | t_i) = \frac{\sum_{a_{-i} \in A_{-i}} \pi(t_i, t_{-i}, a_i, a_{-i})}{\sum_{t_{-i} \in T_{-i}, a \in A} \pi(t_i, t_{-i}, a)}.$$

Let $\mathbf{T}$ be a collection of such common prior SBC type spaces, and for each $\mathcal{T} \in \mathbf{T}$, let $R(\mathcal{T})$ be the supremum expected revenue that can be achieved over all regular mechanisms with a unique Nash equilibrium, where the expectation is taken with respect to this common prior. Theorem 4.2 tells us that the seller can achieve revenue arbitrarily close to $R(\mathcal{T})$ even if he does not know anything about the type space: for any SBC subspace $\mathcal{T} \subset T^*$ with $\mathcal{T} = \phi(\mathcal{T}')$ and $\mathcal{T}' \in \mathbf{T}$, let $\mathbf{M}^\epsilon(\mathcal{T})$ be a mechanism that attains revenue of at least $(1-\epsilon)R(\mathcal{T})$, and extend this to all of $\mathbf{T}$ by $\mathbf{M}^\epsilon(\mathcal{T}) = \mathbf{M}^\epsilon(\phi(\mathcal{T}))$. By construction, this mechanism mapping is locally preference measurable, and the quasilinear private-good setting satisfies joint worst outcome feasibility. Thus, Theorem 4.2 says that we can uniformly virtually implement $(\mathbf{T}, \mathbf{M}^{\epsilon})$. In other words, the seller can attain revenue arbitrarily close to $R(\mathcal{T})$ without knowing anything about the type space.

4.5 Conclusion

The goal of this chapter has been to understand when a designer can recover features of the environment that are common knowledge among the agents for the purpose of building this common knowledge into a mechanism. The designer specifies which mechanism he would like to implement, conditional on a particular type space being the true one. The designer is willing to augment the desired mechanism to facilitate the extraction of the common knowledge, but he would like to use a mechanism that is strategically similar to the one originally intended. Moreover, the designer is willing to settle for a mechanism whose outcomes mostly coincide with what he intended, but he would like the outcomes to coincide as much as possible. The result of the analysis is a simple necessary and nearly sufficient condition on the mechanisms specified by the designer, namely local preference measurability, for there to exist a single mechanism that will extract the common knowledge and fulfill the designer's other requirements.

There remain many open questions. The results of this chapter depend heavily on the private-good structure and the existence of a state independent worst-outcome for each agent. It remains to be seen whether or not these assumptions can be relaxed. Also, the present exercise of extracting common knowledge for the purpose of designing a mechanism is a special case of the more general virtual implementation problem of AM, albeit with infinitely many possible type spaces, and therefore types. An important direction for future work is to extend the theory of virtual implementation of social choice functions to such infinite type spaces.

Bibliography

Abreu, Dilip and Hitoshi Matsushima (1992a), "Virtual implementation in iteratively undominated strategies: Complete information." *Econometrica*, 60, 993–1008.

Abreu, Dilip and Hitoshi Matsushima (1992b), "Virtual implementation in iteratively undominated strategies: Incomplete information." Princeton University and the University of Tsukuba Working Paper.

Athey, Susan (2001), "Single crossing properties and the existence of pure strategy equilibria in games of incomplete information." *Econometrica*, 69, 861–889.

Athey, Susan and Philip A Haile (2007), "Nonparametric approaches to auctions." *Handbook of Econometrics*, 6, 3847–3965.

Aumann, Robert J (1987), "Correlated equilibrium as an expression of Bayesian rationality." *Econometrica*, 55, 1–18.

Azar, Pablo, Jing Chen, and Silvio Micali (2012), "Crowdsourced Bayesian auctions." In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, 236–248, ACM.

Baliga, Sandeep and Rakesh Vohra (2003), "Market research and market design." *Advances in Theoretical Economics*, 3.

Bergemann, Dirk and Stephen Morris (2005), "Robust mechanism design." *Econometrica*, 73, 1771–1813.

Bergemann, Dirk and Stephen Morris (2009a), "Robust implementation in direct mechanisms." *The Review of Economic Studies*, 76, 1175–1204.

Bergemann, Dirk and Stephen Morris (2009b), "Robust virtual implementation." *Theoretical Economics*, 4, 45–88.

Bergemann, Dirk and Stephen Morris (2011), "Robust implementation in general mechanisms." *Games and Economic Behavior*, 71, 261–281.

Bergemann, Dirk and Stephen Morris (2012a), "An introduction to robust mechanism design." *Foundations and Trends in Microeconomics*, 8, 169–230.

Bergemann, Dirk and Stephen Morris (2012b), *Robust Mechnism Design: The Role of Private Information and Higher Order Beliefs*. World Scientific Publishing Company.

Bergemann, Dirk, Stephen Morris, and Satoru Takahashi (2011), "Interdependent preferences and strategic distinguishability." Cowles Foundation Discussion Paper.

Bergemann, Dirk, Stephen Morris, and Satoru Takahashi (2012), "Efficient auctions and interdependent types." *The American Economic Review*, 102, 319–324.

Bergemann, Dirk and Karl Schlag (2011), "Robust monopoly pricing." *Journal of Economic Theory*, 146, 2527–2543.

Bernheim, B Douglas (1984), "Rationalizable strategic behavior." *Econometrica*, 52, 1007–1028.

Birulin, Oleksii and Sergei Izmalkov (2011), "On efficiency of the English auction." *Journal of Economic Theory*, 146, 1398–1417.

Börgers, Tilman (2013), "An introduction to the theory of mechanism design." University of Michigan Book Manuscript.

Brier, Glenn W (1950), "Verification of forecasts expressed in terms of probability." *Monthly Weather Review*, 78, 1–3.

Caillaud, Bernard and Jacques Robert (2005), "Implementation of the revenue-maximizing auction by an ignorant seller." *Review of Economic Design*, 9, 127–143.

Carroll, Gabriel (2012), "Robustness and linear contracts." Stanford University Working Paper.

Chassang, Sylvain (2013), "Calibrated incentive contracts." *Econometrica*, 81, 1935–1971.

Chawla, Shuchi, Jason D Hartline, and Robert Kleinberg (2007), "Algorithmic pricing via virtual valuations." In *Proceedings of the 8th ACM Conference on Electronic Commerce*, 243–251, ACM.

Chen, Yi-Chun, Ngo Van Long, and Xiao Luo (2007), "Iterated strict dominance in general games." *Games and Economic Behavior*, 61, 299–315.

Choi, Jaewon and Taesung Kim (1999), "A nonparametric, efficient public good decision mechanism: Undominated Bayesian implementation." *Games and Economic Behavior*, 27, 64–85.

Chung, Kim-Sau and Jeffrey C Ely (2007), "Foundations of dominant-strategy mechanisms." *The Review of Economic Studies*, 74, 447–476.

Crémer, Jacques and Richard P McLean (1988), "Full extraction of the surplus in Bayesian and dominant strategy auctions." *Econometrica*, 56, 1247–57.

Dasgupta, Partha and Eric Maskin (2000), "Efficient auctions." *The Quarterly Journal of Economics*, 115, 341–388.

d'Aspremont, Claude and Louis-André Gérard-Varet (1979), "Incentives and incomplete information." *Journal of Public Economics*, 11, 25–45.

Dekel, Eddie, Drew Fudenberg, and Stephen Morris (2007), "Interim correlated rationalizability." *Theoretical Economics*, 2, 15–40.

Dufwenberg, Martin and Mark Stegeman (2002), "Existence and uniqueness of maximal reductions under iterated strict dominance." *Econometrica*, 70, 2007–2023.

Farinha Luz, Vitor (2013), "Surplus extraction with rich type spaces." *Journal of Economic Theory*, 148, 2749–2762.

Federal Communications Commission (1997), "The FCC report to Congress on spectrum auctions." *WT Docket*.

Geanakoplos, John D and Heraklis M Polemarchakis (1982), "We can't disagree forever." *Journal of Economic Theory*, 28, 192–200.

Goldberg, Andrew V and Jason D Hartline (2003), "Envy-free auctions for digital goods." In *Proceedings of the 4th ACM Conference on Electronic Commerce*, 29–35, ACM.

Goldberg, Andrew V, Jason D Hartline, Anna R Karlin, and Michael Saks (2004), "A lower bound on the competitive ratio of truthful auctions." In *STACS 2004*, 644–655, Springer.

Gul, Faruk (1998), "A comment on Aumann's Bayesian view." *Econometrica*, 66, 923–927.

Hartline, Jason D (2012), "Bayesian mechanism design." *Foundations and Trends in Theoretical Computer Science*, 8, 143–263.

Hartline, Jason D and Tim Roughgarden (2009), "Simple versus optimal mechanisms." In *Proceedings of the 10th ACM Conference on Electronic Commerce*, 225–234, ACM.

Jehiel, Philippe and Benny Moldovanu (2001), "Efficient design with interdependent valuations." *Econometrica*, 69, 1237–1259.

Krishna, Vijay (2003), "Asymmetric English auctions." *Journal of Economic Theory*, 112, 261–288.

Lebrun, Bernard (2006), "Uniqueness of the equilibrium in first-price auctions." *Games and Economic Behavior*, 55, 131–151.

Lipman, Barton L (1994), "A note on the implications of common knowledge of rationality." *Games and Economic Behavior*, 6, 114–129.

Liu, Qingmin (2009), "On redundant types and Bayesian formulation of incomplete information." *Journal of Economic Theory*, 144, 2115–2145.

Lizzeri, Alessandro and Nicola Persico (2000), "Uniqueness and existence of equilibrium in auctions with a reserve price." *Games and Economic Behavior*, 30, 83–114.

Maskin, Eric (1992), "Auctions and privatization." In *Privatization* (Horst Siebert, ed.), p. 115–136, J.C.B. Mohr.

Mertens, Jean-Francois and Shmuel Zamir (1985), "Formulation of Bayesian analysis for games with incomplete information." *International Journal of Game Theory*, 14, 1–29.

Milgrom, Paul R and Robert J Weber (1982), "A theory of auctions and competitive bidding." *Econometrica*, 50, 1089–1122.

Morris, Stephen (1995), "The common prior assumption in economic theory." *Economics and Philosophy*, 11, 227–227.

Morris, Stephen and Satoru Takahashi (2012), "Games in preference form and preference rationalizability." Princeton University Economic Theory Research Paper.

Morris, Stephen and Takashi Ui (2004), "Best response equivalence." *Games and Economic Behavior*, 49, 260–287.

Myerson, Roger B (1981), "Optimal auction design." *Mathematics of operations research*, 6, 58–73.

Neeman, Zvika (2003), "The effectiveness of English auctions." *Games and Economic Behavior*, 43, 214–238.

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

Pearce, David G (1984), "Rationalizable strategic behavior and the problem of perfection." *Econometrica*, 52, 1029–1050.

Reny, Philip J (1999), "On the existence of pure and mixed strategy Nash equilibria in discontinuous games." *Econometrica*, 67, 1029–1056.

Reny, Philip J and Shmuel Zamir (2004), "On the existence of pure strategy monotone equilibria in asymmetric first-price auctions." *Econometrica*, 72, 1105–1125.

Riley, John and Richard Zeckhauser (1983), "Optimal selling strategies: When to haggle, when to hold firm." *The Quarterly Journal of Economics*, 98, 267–289.

Ronen, Amir (2001), "On approximating optimal auctions." In *Proceedings of the 3rd ACM conference on Electronic Commerce*, 11–17, ACM.

Sadzik, Tomasz (2011), "Beliefs revealed in Bayesian-Nash equilibrium." New York University Working Paper.

Samet, Dov (1998), "Iterated expectations and common priors." *Games and economic Behavior*, 24, 131–141.

Segal, Ilya (2003), "Optimal pricing mechanisms with unknown demand." *The American Economic Review*, 93, 509–529.

Sen, Arunav and D Abreu (1991), "Virtual implementation in Nash equilibrium." *Econometrica*, 997, 1021.

Wilson, Robert (1987), "Game-theoretic analyses of trading processes." In *Advances in Economic Theory: Fifth World Congress* (Truman Bewley, ed.), chapter 2, 33–70, Cambridge University Press.