

TOPICS IN MECHANISM DESIGN

Ludvig Sinander
University of Oxford

This version: 8 November 2024

Copyright © 2024 Carl Martin Ludvig Sinander.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled ‘GNU Free Documentation License’.

This is a ‘copyleft’ licence. Visit gnu.org/licenses/copyleft to learn more.

Preface

These are the notes for a second-year graduate course that I have been teaching at Oxford each autumn since 2021. Thanks to Gregorio Curello, Péter Eső, Deniz Kattwinkel, Elliot Lipnowski, Qingmin Liu, Alex Teytelboym and to my students for comments, suggestions and insights, and to Ufuk Altunbüken, Jo Auer, Catharina Behrens, Josh le Cornu, Tanxin Gao, Justin Hadad, Brooklyn Han, Simon Handreke, Sanjari Kalantri, Marie Kaul, Nils Lager, Arvo Munoz Moran, Sami Petersen, Filip Tokarski, Gautam Vyas and Yidan Xu for reporting errors.

Contents

0	Introduction	5
1	Selling to one agent	7
1.1	The revelation principle	8
1.2	The envelope theorem	9
1.3	Characterisation of incentive-compatibility	11
1.4	Participation	12
1.5	The optimality of posting a price	13
1.6	The role of commitment	17
1.7	Selling several goods	19
1.8	Incentives via side bets	22
1.9	Addendum: preference assumptions	27
1.10	Addendum: the envelope theorem, rigorously	28
2	Allocating among several agents	32
2.1	Border's theorem	33
2.2	Incentives	37
2.3	IPV auctions	38
2.4	Allocation with costly verification	45
2.5	Allocation with evidence acquisition	52
2.6	Allocation with extra information	54
3	Allocating over time, without money	56
3.1	Quota mechanisms	57
3.2	Optimal mechanisms with commitment	61
3.3	Optimal mechanisms without commitment	63
4	Incentivising proposal	65
4.1	Deterministic mechanisms	66
4.2	Variational calculus	68

<i>CONTENTS</i>	4
4.3 Solving	69
4.4 Random mechanisms	72
4.5 Dynamics	73
A Measure and integral	75
B Convexity	78
Bibliography	80

Chapter 0

Introduction

Much of modern economic theory concerns *incentives*. The broad question is how to get some ‘agents’ (e.g. individuals or firms) to behave as we’d like them to. This is achieved by designing institutions or contracts, usually abstractly called ‘mechanisms’. We often imagine there being a ‘principal’ who is tasked with designing the mechanism; but in some applications, there isn’t literally a principal.

Incentive provision is sometimes trivial. Suppose we wish that a child refrain from cursing in the classroom. If the child’s teacher is in the classroom observing her, and is able to threaten the child with the naughty corner (and the child understands and greatly fears this horrible prospect), then the child can be expected to behave. More abstractly, incentives are easily provided when the agents can (a) be monitored accurately and (b) be rewarded generously and/or punished harshly.¹

Incentive design becomes non-trivial as soon as the principal is

- (a) imperfectly informed (i.e. monitoring is imperfect), and/or
- (b) restricted in the rewards and punishments that she can dish out.

The literature is conventionally divided according to (a): *what* is it that the agents know and the principal doesn’t? In *hidden-action* (or *moral-hazard*) models, the agents take actions that the principal doesn’t observe. In *hidden-information* (or *adverse-selection*) models, the agents have private information from the outset (usually about their own preferences, e.g. their valuations for a good). Mechanism design is the study of incentive provision

¹Such mechanisms, in which agents are kept in line using the threat of a harsh punishment, are often called ‘(Mirrlees) shoot-the-agent’ mechanisms.

in the latter type of model. (There are models that feature both moral hazard and adverse selection, of course.)

Work on mechanism design is also usefully distinguished according to (b): what *incentive tools* are available to the principal? Older work mostly focusses on incentive provision using monetary transfers, typically restricted only by agents' ability to walk away if charged too much ('individual rationality') and/or the constraint that money cannot be *taken from* agents ('limited liability'). Much of the recent literature looks at settings in which monetary transfers are highly restricted, or entirely unavailable.

A key feature of most work on mechanism design (and on incentive design more broadly) is that the principal is given the power to *commit*. In particular, she chooses a 'mechanism' (or contract, or set of institutions), which specifies payoff-relevant outcomes as a function of how the agent(s) behave. Once agent(s) have taken their actions, the principal cannot re-optimize by choosing outcomes different from those specified by the mechanism: her previous (contingent) choice binds her. There is also a large body of work asking how incentive design changes when commitment is relaxed.

This course will sample various aspects of modern mechanism design. I'll begin by teaching a classical topic from a modern perspective. The rest of the course will cover topics on (or close to) the current research frontier. Along the way, we'll encounter a number of perspectives, tools, tricks and insights; these are just as important as the details of the models and papers themselves.

Chapter 1

Selling to one agent

There is a single agent (or ‘buyer’) and one indivisible unit of a good. The agent’s valuation $t \in [0, 1]$ for the good is privately known to her. She has quasi-linear expected-utility preferences, meaning that her payoff from getting the good with probability $q \in [0, 1]$ and paying $p \in \mathbf{R}$ is $tq - p$.

(This entails an assumption of risk-neutrality over monetary lotteries, but that is not important for most results. What *is* important for several results is that preferences are strongly separable between the good and money. See §1.9 for a discussion.)

A *principal* can design any mechanism she likes: a mechanism specifies various (observable) actions the agent can take,¹ and as a function of these actions, the probability q with which the agent receives the good and the payment p that she makes. The agent chooses optimally among these actions. Let’s write $q(t)$ for the probability with which the agent of type t gets the good, given her actions, and $p(t)$ for the payment she makes. The map $q : [0, 1] \rightarrow [0, 1]$ is called the induced *allocation* of the good, while $p : [0, 1] \rightarrow \mathbf{R}$ is the induced *payment rule*.

We’d first like to know which allocations and payment rules the principal can achieve, given that only the agent knows her valuation t and that she can be relied upon to choose optimally. (We’ll also assume that the agent can be relied upon to break any indifferences she might experience in whatever way we’d like her to. That’s not a big deal when there’s just one agent.)

¹The actions are often called ‘messages’.

1.1 The revelation principle

We may drastically reduce the space of mechanisms we consider. For any allocation $q : [0, 1] \rightarrow [0, 1]$ and payment rule $p : [0, 1] \rightarrow \mathbf{R}$, we may define a simple ‘direct revelation mechanism’ (DRM): the agent is asked to make a report $r \in [0, 1]$ of her type, and is then given the good with probability $q(r)$ and pays $p(r)$. This DRM is denoted simply by (q, p) . (So ‘ (q, p) ’ can denote an allocation–payment rule pair, or a direct mechanism; mathematically the same, but psychologically distinct.)

A DRM is called *truthful*, or *incentive-compatible (IC)*, iff every type of the agent weakly prefers to report her own type. Clearly if (q, p) is an IC DRM, then it induces the allocation q and payment rule p .

Revelation principle. If the allocation and payment rule (q, p) are induced by some mechanism, then (q, p) (viewed as a DRM) is incentive-compatible (and thus induces the allocation q and the payment rule p).

The revelation principle is both deep and trivial. By the latter, I mean that fully understanding it is tantamount to considering it obvious.

Proof. Let (q, p) be induced by some mechanism. Fix a type t .

Since type t of the agent behaves optimally, her payoff from her outcome $(q(t), p(t))$ is better than the payoff she could get from any deviation.

One deviation would be for her to take whatever actions some other type $r \in [0, 1]$ takes; we call this ‘mimicking type r ’. This would obviously give type t the outcome $(q(r), p(r))$. We infer that type t likes her own outcome $(q(t), p(t))$ weakly better than the outcome $(q(r), p(r))$ of any other type $r \neq t$.

Now consider the DRM (q, p) . The *only* deviations available to t are to mimic type r , for some $r \neq t$. We just showed that each such deviation is unprofitable; thus (q, p) is IC. ■

The revelation principle allows us to restrict attention to incentive-compatible direct revelation mechanisms for analytical purposes, and that’s what we’ll do. (But once we’ve finished analysing and have found an optimal mechanism, we shall have occasion to consider whether it admits a natural *indirect implementation*: i.e. a more natural mechanism that induces the same allocation and payment rule.)

1.2 The envelope theorem

Fix a (direct revelation) mechanism (q, p) . Type t 's problem is to choose an action (her report) $r \in [0, 1]$, where her objective $f(r, t)$ is given by

$$f(r, t) = tq(r) - p(r).$$

So programmatically, the agent's problem is to choose her report r to maximise $f(r, t)$.

Suppose that our mechanism is an incentive-compatible one, meaning that the report $r = t$ yields a global maximum of $f(\cdot, t)$. One implication of incentive-compatibility is that no type t wishes to mimic a *nearby* type r ; this is called *local incentive-compatibility*. Intuitively, local IC is captured by the first-order condition

$$\left. \frac{d}{dm} f(t + m, t) \right|_{m=0} = 0, \quad \text{or more compactly,} \quad f_1(t, t) = 0.$$

Formally, we'll say that (q, p) is locally IC if this expression holds for a.e. type $t \in [0, 1]$.

Now let's re-express that. Write $V(t) := f(t, t)$ for the value of truthful reporting. Differentiating on both sides yields

$$V'(t) = f_1(t, t) + f_2(t, t).$$

The first term on the RHS is exactly the thing that's zero if our mechanism is a locally IC one. So local IC is equivalent to

$$V'(t) = f_2(t, t) \quad \text{for a.e. } t \in [0, 1],$$

or (integrating)

$$V(t) = V(0) + \int_0^t f_2(s, s) ds \quad \text{for every } t \in [0, 1].$$

This is called the *envelope formula*. Recalling that $f(r, t) = tq(r) - p(r)$ by definition, the envelope formula more explicitly reads

$$tq(t) - p(t) = -p(0) + \int_0^t q.$$

(Here ' $\int_0^t q$ ' is a shorter way of writing ' $\int_0^t q(s) ds$ '.)

The fact that IC mechanisms satisfy the envelope formula is an instance the *envelope theorem*. (The envelope theorem is more general: it applies to

any parametrised maximisation problem. Other applications are Shephard's lemma, Roy's identity and Hotelling's lemma.)

The above argument has a big hole: given our completely arbitrary mechanism (q, p) , there is no reason why the agent's reporting payoff $f(r, t) = tq(r) - p(r)$ should be a differentiable function of r , in which case the derivative

$$\left. \frac{d}{dm} f(t + m, t) \right|_{m=0} \equiv f_1(t, t)$$

simply does not exist.² But the above argument *can* be made rigorous: it really is true that the envelope formula is a fancy way of rewriting 'local optimality' (suitably defined; see Sinander (2022)). Alternatively and more traditionally, one can use a completely different (unintuitive but elegant) argument to derive the envelope formula directly from (global) IC (Milgrom & Segal, 2002)—see §1.10 at the end of this chapter. Either way, we have learned:

Mirrlees envelope theorem. Any IC mechanism (q, p) satisfies the envelope formula:

$$tq(t) - p(t) = -p(0) + \int_0^t q \quad \text{for every } t \in [0, 1].$$

Remark 1 (revenue/payoff equivalence). It follows that any two indirect mechanisms that induce the same allocation q and induce a payment of $p(0) = 0$ from type $t = 0$ must actually induce the exact same payment rule p . (Why? Prove it!) Thus any two such mechanisms provide the same payoff to every type of the agent, and also the same payment (or 'revenue' from the perspective of the principal/seller collecting the money). This insight is usually attributed to Myerson (1981), with a precedent in Vickrey (1961).

Remark 2. Why is the envelope formula so restrictive (and thus the envelope theorem so powerful)? Fix a type $t \in (0, 1)$. The IC constraint $V(r) \geq f(t, r)$ deterring a higher type $r > t$ from mimicking t may be rewritten (by subtracting $V(t)$ from both sides and dividing by $r - t$) as

$$\frac{V(r) - V(t)}{r - t} \geq \frac{f(t, r) - f(t, t)}{r - t}, \quad (1.1)$$

so that letting $r \downarrow t$ yields $V'(t) \geq f_2(t, t)$. Similarly, the IC constraint $V(r') \geq f(t, r')$ deterring a lower type $r' < t$ from mimicking t may be

²There's another issue: the step from ' $V'(t) = f_2(t, t)$ for a.e. $t \in [0, 1]$ ' to ' $V(t) = V(0) + \int_0^t f_2(s, s)ds$ for every $t \in [0, 1]$ ' is valid iff V is *absolutely continuous*—see §1.10.

rewritten

$$\frac{V(t) - V(r')}{t - r'} \leq \frac{f(t, t) - f(t, r')}{t - r'}, \quad (1.2)$$

which as $r' \uparrow t$ yields $V'(t) \leq f_2(t, t)$. Putting together our two conclusions, we have

$$f_2(t, t) \leq V'(t) \leq f_2(t, t), \quad \text{or} \quad V'(t) = f_2(t, t).$$

Economically, the fact that we obtain an equality comes from the fact that types r, r' arbitrarily close to t must be deterred from mimicking t . (Formally, because the set $[0, 1]$ of types is a connected set.) If types were spaced out discretely, then we would obtain only the more permissive ‘inequality’ envelope formula comprising (1.1)–(1.2).

The literature. The intuitive treatment in this section can be made formal: see Sinander (2022). In §1.10 below, we’ll give a rigorous treatment of the envelope theorem from a different perspective (Milgrom & Segal, 2002).

1.3 Characterisation of incentive-compatibility

Spence–Mirrlees lemma. A mechanism (q, p) is IC if and only if it satisfies the envelope formula and q is increasing.

Proof. Fix a mechanism (q, p) . Write $V(t) := tq(t) - p(t)$ for the value of truthful reporting.

Observe first that if (q, p) satisfies the envelope formula, then the payoff loss of type t from mimicking type r instead of reporting truthfully is

$$\begin{aligned} V(t) - [tq(r) - p(r)] &= V(t) - V(r) + [rq(r) - p(r)] - [tq(r) - p(r)] \\ &= \int_r^t q(s)ds - (t - r)q(r) \\ &= \int_r^t [q(s) - q(r)]ds, \end{aligned} \quad (\star)$$

where the final equality holds by the fundamental theorem of calculus.

Suppose that (q, p) is IC. We’ve seen (the Mirrlees envelope theorem above) that it satisfies the envelope formula. So by IC and (\star) , we must have

$$\int_r^t [q(s) - q(r)]ds \geq 0 \quad \text{for all } r, t \in [0, 1],$$

which is only possible if q is increasing. (Right? Convince yourself.)

Suppose that (q, p) satisfies the envelope formula and that q is increasing. Then type t 's payoff loss from mimicking r is given by (\star) as

$$\int_r^t [q(s) - q(r)] ds,$$

which is non-negative since q is increasing. Thus (q, p) is IC. ■

The literature. The Spence–Mirrlees lemma has been extended in various ways. At the highest level of generality, the ‘outcome’ q belongs to an abstract space Q ($= [0, 1]$ in the text) equipped with a partial order, and the agent’s payoff is some function $f(q, p, t)$ ($= tq - p$ in the text). Versions of the result go through whenever f satisfies the *Spence–Mirrlees* (‘single-crossing’) *condition*, defined in terms of how different types’ indifference curves in q – p space cross each other.³ See Sinander (2022, §4) for an overview of such results (plus a new, general result).

1.4 Participation

It is natural to assume that the agent can walk away. In particular, she may consume an outside option worth zero to her (no good, no payment).

It is without loss of generality to focus on mechanisms that induce every type of the agent to participate. This is because if type t were not participating, then we could invite her to participate and award her the outcome $(q(t), p(t)) = (0, 0)$ if she does, which is no better (or worse) than non-participation.

We may thus focus on IC mechanisms (q, p) that induce participation, i.e. those such that every type t 's payoff $tq(t) - p(t)$ is non-negative. Such mechanisms are called *individually rational* (IR).

Corollary 1. A mechanism (q, p) is IC and IR if and only if it satisfies the envelope formula, q is increasing, and $p(0) \leq 0$.

Proof. ‘ \implies ’ direction: if (q, p) is IR, then the payoff $0 \times q(0) - p(0) = -p(0)$ of type $t = 0$ must be at least zero (the value of the outside option). And we’ve already seen (the Spence–Mirrlees lemma) that IC requires the other properties.

³When preferences have the quasi-linear form $f(q, p, t) = g(q, t) - p$, Spence–Mirrlees requires precisely that g be *supermodular*. In general, a broader (ordinal) notion of supermodularity (or ‘complementarity’) characterises the Spence–Mirrlees condition (Milgrom & Shannon, 1994, Theorem 3).

‘ \Leftarrow ’ direction: fix a mechanism (q, p) , and write $V(t) := tq(t) - p(t)$ for the value of type t . If (q, p) satisfies the first two properties, then we’ve seen (the Spence–Mirrlees lemma) that it is IC. If it furthermore satisfies $p(0) \leq 0$, then $V(0) = -p(0) \geq 0$, and so by the Mirrlees envelope theorem

$$V(t) = V(0) + \int_0^t q \geq 0 \quad \text{for every } t \in [0, 1],$$

which is to say that (q, p) is IR. ■

1.5 The optimality of posting a price

Suppose that the good is owned by a monopolist (principal) who wishes to sell it so as to maximise expected profit. (Equivalently, re-interpret the agent as a continuum of consumers, each with a privately-known valuation; the monopolist wishes to maximise profit. This problem bears the antiquated name of *second-degree price discrimination*.)

The monopolist may use any mechanism she likes, but must take into account that only the agent knows her valuation and that she may choose to walk away.

Here’s a very simple (indirect) mechanism that the monopolist could adopt: post a price! More fully, the monopolist sets a price $\pi \in \mathbf{R}_+$ and gives the agent two options: purchase the good at price π (pay π and get the good for sure) or don’t (pay nothing and get nothing).

What allocation and payments does this induce? Agents of type $t > \pi$ will purchase the good, so $(q(t), p(t)) = (1, \pi)$ for them, while agents of type $t < \pi$ will not, so $(q(t), p(t)) = (0, 0)$.

Exercise 1. (a) This allocation–payment pair (viewed as a direct mechanism) is IC; why? (b) Verify that this mechanism satisfies the envelope formula. (c) There are other mechanisms (q, p') , i.e. mechanisms with the same allocation but different payments. Describe them.

This is about as simple a mechanism as can be devised. One thing that makes it simple is that it does not make use of the monopolist’s power to allocate the good randomly, which is in principle very powerful. Nonetheless:

Theorem 1 (Myerson, 1981). There is a posted-price mechanism that is optimal.

The rest of this section is devoted to proving this result. Myerson’s original argument is based on a duality technique called ‘(ironed) virtual valuations’.

This technique is important and useful; indeed, Myerson's proof directly extends to the case of multiple agents. (This yields Myerson's celebrated result that a second-price auction with a reserve price is optimal.) We shall instead pursue a direct, convexity-based argument that I learned from Eran Shmaya.

We begin by using our preceding results (in particular, Corollary 1) greatly to narrow down the space of mechanisms under consideration. It is obviously not optimal to subsidise type $t = 0$ (and thus by the envelope formula to lower the payments of all types), so $p(0) = 0$ is optimal. The monopolist therefore merely has to choose the allocation, which can be any increasing function $q : [0, 1] \rightarrow [0, 1]$. The payment rule p is then pinned down by the envelope formula and $p(0) = 0$ as

$$p(t) = tq(t) - \int_0^t q \quad \text{for each } t \in [0, 1].$$

This narrows our problem down to one of choosing from among the space of all increasing functions $[0, 1] \rightarrow [0, 1]$. That's still a very large (infinite-dimensional) space, though, so we aren't out of the woods yet. It isn't obvious, for example, that it won't be optimal sometimes to allocate with interior probability (selling a coin toss), as that could have incentive benefits.

The monopolist's revenue is just the agent's payment, which depends on her type according to the equation above. The monopolist views the agent's valuation as a random variable; we'll denote it by T , and write F for its CDF. Her expected revenue is thus

$$\mathbf{E}_{T \sim F}(p(T)) = \mathbf{E}_{T \sim F}\left(Tq(T) - \int_0^T q\right) =: R(q).$$

The monopolist's problem is to maximise $R(q)$ by choosing q from the space \mathcal{Q} of all increasing maps $q : [0, 1] \rightarrow [0, 1]$.

Note that \mathcal{Q} is a convex space: if q and q' are both increasing maps $[0, 1] \rightarrow [0, 1]$, then so is $\alpha q + (1 - \alpha)q'$ for any scalar $\alpha \in [0, 1]$. Note further that the objective R is linear:

$$R(\alpha q + (1 - \alpha)q') = \alpha R(q) + (1 - \alpha)R(q') \quad \text{for any } q, q' \in \mathcal{Q} \text{ and } \alpha \in [0, 1].$$

R is also continuous in a suitable sense. Finally, \mathcal{Q} is (in fact) compact in a suitable sense.

Now let's do some convex geometry (see appendix B for a little overview). An *extreme point* of \mathcal{Q} is an element of \mathcal{Q} that cannot be expressed as the convex combination of two *distinct* elements of \mathcal{Q} .

Observation 1. Any convex and suitably continuous function $\phi : \mathcal{Q} \rightarrow \mathbf{R}$ is maximised at an extreme point of \mathcal{Q} .

Proof. It is intuitive, and in fact true, that any element of the compact convex set \mathcal{Q} can be written as an (infinite) convex combination of the extreme points of \mathcal{Q} . Results like this constitute a little field called Choquet theory, and Choquet's theorem⁴ says that we may for any $q \in \mathcal{Q}$ find a probability measure μ defined on $\text{ext } \mathcal{Q}$ such that $q = \int_{\text{ext } \mathcal{Q}} q' \mu(dq')$.⁵

Now, let $q \in \mathcal{Q}$ maximise a convex and suitably continuous function ϕ on \mathcal{Q} . Then since we have $q = \int_{\text{ext } \mathcal{Q}} q' \mu(dq')$ for some probability measure μ on $\text{ext } \mathcal{Q}$, we have

$$\phi(q) \leq \int_{\text{ext } \mathcal{Q}} \phi(q') \mu(dq')$$

by Jensen's inequality, and thus $\phi(q) \leq \phi(q')$ for some $q' \in \text{ext } \mathcal{Q}$. Since q is optimal, q' must be, too. ■

Observation 1 permits us to conclude that the monopolist's problem $\max_{q \in \mathcal{Q}} R(q)$ admits a solution that is an extreme point of the space \mathcal{Q} of increasing functions $[0, 1] \rightarrow [0, 1]$. I claim that the extreme points of \mathcal{Q} are exactly those functions $q : [0, 1] \rightarrow [0, 1]$ that satisfy

$$q(t) = \begin{cases} 0 & \text{for } t < t^* \\ 1 & \text{for } t > t^* \end{cases} \quad \text{and} \quad q(t^*) \in \{0, 1\} \quad \text{for some } t^* \in [0, 1].$$

I will call such functions *impulses* (my term). It is a fact that all and only impulses are extreme points of \mathcal{Q} .

Exercise 2. Prove it! That is, (a) show that every impulse is an extreme point of \mathcal{Q} , and (b) show that every extreme point of \mathcal{Q} is an impulse.⁶

We conclude that there is a revenue-maximising allocation q that is an impulse. In other words, all types above a threshold get the good for sure, while those below do not get the good. This is a remarkably simple allocation rule; for one thing, it is deterministic!

⁴Or its generalisation, the Choquet–Bishop–de Leeuw theorem. See e.g. Phelps (2001).

⁵That's a Lebesgue integral; it's a fancy way of saying that q is a convex combination of q' 's in $\text{ext } \mathcal{Q}$, with $\mu(\{q'\}) \in [0, 1]$ being the weight placed on $q' \in \text{ext } \mathcal{Q}$.

⁶Part (b) is harder, so here's a hint. Prove the contra-positive: fix any non-impulse $q \in \mathcal{Q}$, and try to show that it isn't an extreme point, by constructing distinct $q^-, q^+ \in \mathcal{Q}$ such that $q = \alpha q^- + (1 - \alpha) q^+$ for some $\alpha \in (0, 1)$. Be careful to ensure that your q^- and q^+ are legitimately elements of \mathcal{Q} , i.e. that they take values in $[0, 1]$ and are increasing.

What are the implied payments? You can calculate the payment rule from the envelope formula and the condition $p(0) = 0$. (Try it!) Or from scratch: although each type t can make many different reports in the direct mechanism, they fall into two categories: reports $< t^*$, which yield $q = 0$, and reports $> t^*$, which yield $q = 1$. Clearly IC requires that all types $< t^*$ make the same payment; and similarly for types $> t^*$. By $p(0) = 0$, the former group pay zero; let's write π for what the latter group pay. Evidently type t^* must be indifferent; so $\pi = t^*$.

We have shown that whatever the distribution F of the agent's type, there is a posted-price mechanism that is optimal. We have not described the optimal price π ; this depends on the distribution F , and is easily characterised via a first-order condition.⁷

The literature. The theorem is due to Myerson (1981); his result is actually more general, as it covers the case of several agents. He proved it by developing a duality technique based on '(ironed) virtual valuations' that has proved useful in other environments. The direct convexity-based approach here is more intuitive to me, and has other applications: Kleiner, Moldovanu and Strack (2021) is a nice recent example. More generally, much of the structure of modern economic theory is simply convex structure, as described by convex analysis. The standard reference here is Rockafellar (1970); for Choquet theory in particular, see e.g. Phelps (2001).

Exercise 3. Suppose that the monopolist is not a Bayesian decision-maker: instead of having a single belief F about the agent's valuation, she entertains an entire set \mathcal{F} of beliefs F . For each given belief F , let us write

$$R_F(q) := \mathbf{E}_{T \sim F}(p(T)) = \mathbf{E}_{T \sim F} \left(Tq(T) - \int_0^T q \right)$$

for the monopolist's expected revenue from an increasing allocation q .

- (a) Suppose to begin with that the monopolist is a 'maxmax' decision-maker, meaning that she evaluates an allocation q according to the most optimistic of the beliefs F in the set \mathcal{F} .⁸

$$\mathcal{U}(q) = \max_{F \in \mathcal{F}} R_F(q).$$

Prove that there is a posted-price mechanism that is optimal. (Hint: \mathcal{U} is not linear. But...?)

⁷You may have derived this first-order condition when studying monopoly pricing in introductory microeconomics.

⁸This could be interpreted as 'motivated reasoning'.

- (b) Suppose instead that the monopolist has ‘maxmin’ preferences: she evaluates an increasing allocation q according to the pessimistic criterion

$$\mathcal{V}(q) = \min_{F \in \mathcal{F}} R_F(q).$$

(This captures ‘uncertainty-aversion’, as exemplified by the Ellsberg paradox.) Can our argument above be salvaged? Explain.

- (c) Maintain the maxmin assumption, and further suppose that \mathcal{F} has a least element \underline{F} , in the sense of first-order stochastic dominance. (That is, every $F \in \mathcal{F}$ first-order stochastically dominates \underline{F} .)

Reminder: F first-order stochastically dominates G if and only if $\mathbf{E}_{T \sim F}(\phi(T)) \geq \mathbf{E}_{T' \sim G}(\phi(T'))$ for every increasing $\phi : [0, 1] \rightarrow \mathbf{R}$.

- (i) Fix an increasing allocation q . Let p be the payment rule induced by q via the envelope formula and the condition $p(0) = 0$. Show that p is increasing.
- (ii) Prove that $\mathcal{V}(q) = R_{\underline{F}}(q)$ for every increasing allocation q .
- (iii) Show that there is a posted-price mechanism which is optimal.

1.6 The role of commitment

The monopolist’s ability to commit to a mechanism is the backbone of the above analysis. To see why it matters, consider an optimal posted-price mechanism with price π . Although it maximises *expected* revenue, the monopolist might get unlucky ex post: if the agent’s valuation turns out to be less than π , then the good stays with the monopolist, and she earns no revenue at all.

Having observed the agent’s failure to purchase, the monopolist may reasonably infer that the agent’s valuation is less than π , but quite possibly still positive. Were she able to, the monopolist would now very much like to offer the good for sale once more, this time at a lower price.

Were the agent to expect the monopolist to behave in this way, however, it would change the agent’s behaviour in the first place: even if her valuation exceeds π , she may decline to purchase at this price since doing so will secure her a better deal. Continuing this reasoning suggests, correctly, that the revelation principle is invalid absent commitment by the monopolist. The trouble is that there are now IC constraints not just for the agent, but also for the monopolist.

The monopolist's inability to commit not to try to sell the good again is an inherently dynamic problem, and so we need a dynamic model. (We were able to avoid this previously because the revelation principle made static mechanisms without loss.) We'll let the 'length' of a period be $\Delta > 0$, so that the periods are $n \in \{0, \Delta, 2\Delta, \dots\}$. The discount rate (for both the agent and the monopolist) is $r > 0$. (So between adjacent periods, the agent discounts by factor $\delta = e^{-r\Delta}$.) Note that we are assuming the good to be durable (non-perishable): its value does not diminish over time.

Previously, we allowed the monopolist to commit (in each period) to either (a) hand over the good (immediately) or (b) to retain the good (forever). As discussed above, it is plausible to suppose that the monopolist is unable to commit to (b): if she fails to sell the good, then she cannot bind her future selves not to try to sell it again. It remains reasonable (do you agree?) to assume that the monopolist has the power to commit to (a) hand over the good: this means that she is able to prevent her future selves from clawing back the good from the agent.⁹

This suggests the following interaction within each period: the monopolist offers the good for sale at a price, and the agent accepts or rejects. If the agent accepts, then the monopolist gives her the good (and thus commits permanently to relinquish the good—she cannot claw it back). If not, then the monopolist keeps the good until the next period.¹⁰ This is a bargaining game, and we'll be interested in its (perfect Bayesian) equilibria. (Let's not get bogged down in defining PBE formally, though.)

Let's assume that the agent's valuation T is supported on an interval $S \subseteq [0, 1]$. Recall that we normalise the monopolist's valuation of the good to zero.

The *Coase conjecture* asserts that the monopolist's bargaining power is small when offers are frequent (Δ is small): in equilibrium, she sells at a low price. More formally, the claim is roughly that the good sells a.s. in finite time, at a price p^Δ that approaches $\inf S$ as $\Delta \rightarrow 0$. The classical intuition for this conjecture is that when the period length Δ is small, the monopolist in period n engages in near-perfect competition with her period- $(n+1)$ self.

This intuition is incomplete, it seems to me, at least when $\inf S > 0$,

⁹This commitment power on the monopolist's part presumably comes from an external system of enforced property rights. Where such enforcement or rights are absent, a monopolist may not be able to prevent her future self from expropriating the agent.

¹⁰The monopolist might wish to offer a new price right away, instead of waiting until the next period. We are ruling this out as infeasible—the (possibly very short) 'period length' $\Delta > 0$ captures constraint on how long it takes to carry out one round of bargaining between buyer and monopolist.

because ‘perfect competition’ surely means selling at marginal cost (which we’ve normalised to zero), rather than at price $\inf S$! Here’s a better intuition that I learned from Francesco Nava: the seller’s lack of commitment constrains her eventually to sell the good whatever the buyer’s value, and constrains her no further than that; thus she sells the good at the highest market-clearing price, which is $\inf S$.

Is the conjecture true? If we restrict attention to stationary equilibria, then given some technical assumptions, it is indeed true (Gül, Sonnenschein & Wilson, 1986). This can be proved easily, in about a page: see the nice and short note by Qingmin Liu (2015).

What about non-stationary equilibria? (The results below involve some technical assumptions.) If $\inf S > 0$ (the ‘gap case’), there is an (essentially) unique equilibrium, and it is stationary; but if $\inf S = 0$ (the ‘no-gap case’), a folk theorem holds, meaning that there exist non-stationary ‘reputational’ equilibria which support trade at any price between zero and the full-commitment monopoly price, provided $\Delta > 0$ is small enough (Gül, Sonnenschein & Wilson, 1986; Ausubel & Deneckere, 1989).

The literature. The conjecture is from Coase (1972). The big papers are Fudenberg, Levine and Tirole (1985), Gül, Sonnenschein and Wilson (1986) and Ausubel and Deneckere (1989); see Ausubel, Cramton and Deneckere (2002) for a survey. For the intuition I attributed to Francesco Nava, see Nava and Schiraldi (2019) and the references therein. For a taste of recent work on this topic, consider Doval and Skreta (2024), Brzustowski, Georgiadis-Harris and Szentes (2023) and Lomys and Yamashita (2022).

1.7 Selling several goods

Suppose now that the monopolist has two goods to sell. We consider the simplest case: the agent’s payoff is $t_1 q_1 + t_2 q_2 - p$.

The monopolist’s revenue-maximisation problem with two goods is an open problem, despite its apparent simplicity! This is true even when t_1, t_2 are assumed to be statistically independent. In this section, we’ll try to get a feel for what the issues are, and why this problem is hard.

Mechanisms are now $(q_1, q_2, p) : [0, 1]^2 \rightarrow [0, 1]^2 \times \mathbf{R}$. It remains true that an allocation (q_1, q_2) is implementable (i.e. a payment rule p can be found such that (q_1, q_2, p) is IC) exactly if (q_1, q_2) is suitably monotone; in particular, *cyclically monotone*.¹¹

¹¹The term comes from convex analysis (see Rockafellar, 1970). This link between

Let's write \mathcal{Q}_2 for all cyclically monotone allocations. \mathcal{Q}_2 is convex (and compact), and revenue is a linear function $R : \mathcal{Q}_2 \rightarrow \mathbf{R}$, so our previous analysis tells us that there must be an optimal mechanism that is an extreme point of \mathcal{Q}_2 . But characterising these extreme points is hard, and there are many of them. More broadly: cyclic monotonicity is intractable.

So what can happen? For one thing, it can easily be that optimal mechanisms feature random assignment. Here's an example from Hart and Reny (2015b):

Proposition 1. Suppose that the agent's valuation (t_1, t_2) equals either $(1, 0)$, $(0, 2)$ or $(3, 3)$, each with equal probability. The uniquely optimal direct mechanism is

$$(q_1(t), q_2(t), p(t)) = \begin{cases} \left(\frac{1}{2}, 0, \frac{1}{2}\right) & \text{for } t = (1, 0) \\ (0, 1, 2) & \text{for } t = (0, 2) \\ (1, 1, 5) & \text{for } t = (3, 3). \end{cases}$$

There is an IC constraint for each pair of distinct types (so six IC constraints), plus an IR constraint for each type.

Exercise 4. Verify that this mechanism is IC and IR.

Some of these constraints bind, and others do not. The 'downward' IC constraints that deter type $t = (3, 3)$ from pretending to be one of the lower-valuation types both bind: this type's truthful payoff is

$$3 \times 1 + 3 \times 1 - 5 = 1,$$

while she earns

$$3 \times \frac{1}{2} + 3 \times 0 - \frac{1}{2} = 1$$

by mimicking type $t = (1, 0)$ and earns

$$3 \times 0 + 3 \times 1 - 2 = 1$$

by mimicking type $t = (0, 2)$. You can also easily verify that the IR constraints for types $(1, 0)$ and $(0, 2)$ both bind.

This is where the random outcome for type $(1, 0)$ is valuable to the monopolist. By randomising, she is able to keep *both* 'downward' IC constraints binding, while still giving type $(1, 0)$ a payoff of zero. Using deterministic

implementability and convex analysis was spotted by Rochet (1987).

outcomes, it is typically not possible to satisfy several downward IC constraints with equality without ceding ‘information rents’ to the lower types in question.

If that was too loose for you, here’s a proof. It’s a lot of algebra, *but with ‘economic’ remarks added in italics.*

Proof. For an arbitrary mechanism (q_1, q_2, p) , we may write $(\alpha_1, \beta_1, \pi_1)$, $(\alpha_2, \beta_2, \pi_2)$ and $(\alpha_3, \beta_3, \pi_3)$ for $(q_1(t), q_2(t), p(t))$ for $t = (1, 0)$, $t = (0, 2)$ and $t = (3, 3)$, respectively. Expected revenue is $\frac{1}{3}\pi_1 + \frac{1}{3}\pi_2 + \frac{1}{3}\pi_3$.

The mechanism in the proposition is easily shown to satisfy all of the IC and IR constraints. We shall show that it is uniquely optimal in a relaxed revenue-maximisation problem in which some of these constraints are ignored; that obviously implies that it is uniquely optimal in the original problem.

So consider the problem of choosing a mechanism subject only to the IR constraints for types $(1, 0)$ and $(0, 2)$ and the ‘downward’ IC constraints $(3, 3) \rightarrow (1, 0)$ and $(3, 3) \rightarrow (0, 2)$, i.e.

$$\begin{aligned}\alpha_1 - \pi_1 &\geq 0 \\ 2\beta_2 - \pi_2 &\geq 0 \\ 3\alpha_3 + 3\beta_3 - \pi_3 &\geq 3\alpha_1 + 3\beta_1 - \pi_1 \\ 3\alpha_3 + 3\beta_3 - \pi_3 &\geq 3\alpha_2 + 3\beta_2 - \pi_2.\end{aligned}$$

Rewriting yields

$$\begin{aligned}\pi_3 + 3\alpha_1 + 3\beta_1 - 3\alpha_3 - 3\beta_3 &\leq \pi_1 \leq \alpha_1 \\ \pi_3 + 3\alpha_2 + 3\beta_2 - 3\alpha_3 - 3\beta_3 &\leq \pi_2 \leq 2\beta_2.\end{aligned}$$

Clearly to maximise $\frac{1}{3}\pi_1 + \frac{1}{3}\pi_2 + \frac{1}{3}\pi_3$, we must choose $\pi_1 = \alpha_1$ and $\pi_2 = 2\beta_2$. (*IR binds for the two low-valuation types: it is not optimal to give information rents to ‘low’ types.*) The remaining constraints are

$$\begin{aligned}\pi_3 &\leq 3\alpha_3 + 3\beta_3 - 2\alpha_1 - 3\beta_1 \\ \pi_3 &\leq 3\alpha_3 + 3\beta_3 - 3\alpha_2 - \beta_2.\end{aligned}$$

It is then clearly optimal to let $\alpha_3 = \beta_3 = 1$ and $\beta_1 = \alpha_2 = 0$. (*This is intuitive: the high-valuation type gets both goods, while each of the lower-valuation types get zero of the good that they do not value.*) Clearly π_3 should be set as high as possible subject to these two constraints:

$$\pi_3 = \min\{6 - 2\alpha_1, 6 - \beta_2\}.$$

The remainder of the problem is to choose α_1, β_2 to maximise

$$\frac{1}{3}\alpha_1 + \frac{2}{3}\beta_2 + \frac{1}{3}\min\{6 - 2\alpha_1, 6 - \beta_2\}.$$

The trade-off here is that raising α_1 allows a higher payment ($\pi_1 = \alpha_1$) to be extracted from type $t = (1, 0)$, but also tightens the IC constraint $(3, 3) \rightarrow (1, 0)$, potentially requiring π_3 to be lowered. Similarly for β_2 .

Since this expression is increasing in β_2 , it is optimal to set $\beta_2 = 1$. We now merely have to choose α_1 to maximise

$$\frac{1}{3}\alpha_1 + \frac{2}{3} + \frac{1}{3}\min\{6 - 2\alpha_1, 5\}.$$

This is (uniquely) achieved by setting $\alpha_1 = 1/2$.

The economics of the last step are as follows. As we decrease α_1 from 1, we decrease the payment $\pi_1 = \alpha_1$ that we may extract from type $(1, 0)$, but the amount $\pi_3 = 6 - 2\alpha_1$ that we may charge type $(3, 3)$ rises twice as quickly since lowering α_1 slackens the IC constraint $(3, 3) \rightarrow (1, 0)$; so this is worth it. But once α_1 hits $1/2$, decreasing it further continues to be costly, but without the benefit, because now the other downward IC constraint $(3, 3) \rightarrow (0, 2)$ binds, preventing us from raising π_3 any further. Thus $\alpha_1 = 1/2$ is uniquely optimal. ■

The literature. Hart and Reny (2015b) provide a good guide to why this problem is difficult and why optimal mechanisms are stranger than one might expect; they also provide good references to the literature. Computer scientists have taken an interest in this slice of economic theory, and much recent progress has come from that corner, including the closest thing to a recent breakthrough: Daskalakis, Deckelbaum and Tzamos (2017).¹² Although ‘simple’ mechanisms (e.g. selling each good separately at a posted price, or selling all the goods in one big bundle at a posted price) are typically not optimal, they can sometimes do rather well: see Hart and Nisan (2017).

1.8 Incentives via side bets

Return to the case of a single good (with commitment), and let $\mathcal{T} \subseteq [0, 1]$ denote the set of types. We previously assumed that $\mathcal{T} = [0, 1]$, but in this section, it will sometimes be convenient to suppose that \mathcal{T} is finite.

¹²Their proof uses fancy optimal-transport techniques; see Galichon (2016) for an economist’s introduction. Kleiner and Manelli (2019) found an elementary proof.

In defining a ‘mechanism’ at the beginning of this chapter, we implicitly ruled out incentive schemes in which the outcome $(q, p) \in [0, 1] \times \mathbf{R}$ depends on factors beyond the agent’s control: only the agent’s actions in the mechanism were allowed to affect the outcome.

We now relax this assumption. There is a contractible *signal* whose realisation becomes known (i) late enough that the agent has already chosen her actions in the mechanism, but (ii) early enough that monetary payments can be made contingent on it. (It won’t matter whether the allocation of the good can depend on the realisation of the signal.) Suppose for simplicity that there are only finitely many possible signal realisations, $\mathcal{X} = \{1, \dots, |\mathcal{X}|\}$.

The signal can arise from a noisy monitoring technology. (For example, any external information about the market value of the good.) Alternatively, the signal could be a *second* agent’s type, provided (as will turn out to be the case) that this second agent can costlessly be prevailed upon truthfully to reveal her type.

Each type $t \in \mathcal{T}$ of the agent has a belief $\mu(t)$ about how likely various signal realisations are. Formally, a *belief* (about the signal) is a vector $\mu \in [0, 1]^{|\mathcal{X}|}$ whose entries sum to unity. We write Δ for the set of all beliefs.

Remark 3. This model is ‘ex interim’, which you may find unintuitive. In the (slightly more restrictive) ‘ex ante’ version of the model, the signal X and agent type T are random variables, drawn from some joint (‘prior’) distribution, and the belief $\mu(t)$ of type $t \in \mathcal{T}$ is simply the vector whose x^{th} entry (for $x \in \{1, \dots, |\mathcal{X}|\}$) is $\mu(t)_x = \mathbf{P}(X = x | T = t)$.

Insofar as X and T are not independent, the signal is informative about the agent’s type. In (only) this case, the reverse is also true: the agent’s type is informative about the signal, meaning that from observing her own type, the agent learns something about the likely realisation of the signal. In other words, $\mu(t)$ differs across types $t \in \mathcal{T}$ if (and only if) X, T are not independent.¹³

As before, an *allocation* is a map $q : \mathcal{T} \rightarrow [0, 1]$. A *contingent payment rule* is a vector-valued map $p : \mathcal{T} \rightarrow \mathbf{R}^{|\mathcal{X}|}$; the interpretation is that if the agent reports that her type is $r \in \mathcal{T}$ and the signal turns out to be $x \in \mathcal{X}$, then the agent pays $p(r)_x \in \mathbf{R}$.

A *direct revelation mechanism* is now (q, p) , where q is an allocation and p is a contingent payment rule. A direct revelation mechanism (q, p) is *incentive-compatible* iff each type $t \in \mathcal{T}$ finds it optimal (in expectation,

¹³A simple example: if X and T are ‘positively correlated’ in the sense of affiliation, then higher types have ‘higher’ beliefs in the sense of monotone likelihood ratio.

given her belief) to report truthfully:

$$q(t)t - p(t) \cdot \mu(t) \geq q(r)t - p(r) \cdot \mu(t) \quad \text{for all } r, t \in \mathcal{T}.$$

As before, a revelation principle permits us without loss to restrict attention to incentive-compatible direct revelation mechanisms, henceforth called simply ‘IC mechanisms’. An allocation q is called *implementable* iff there is a contingent payment rule p such that (q, p) is incentive-compatible.

Suppose first that all types of the agent have the same belief about the signal ($\mu(r) = \mu(t)$ for all $r, t \in \mathcal{T}$). (In the ‘ex ante’ model of Remark 3, this means precisely that the signal is uninformative about the agent’s type—they are independent random variables.) In this case, contingent payments cannot help with incentive provision: it remains true that (all and) only increasing allocations q are implementable.

Exercise 5. Prove it!

As soon as different types of the agent have different beliefs about the signal (i.e. the signal is informative about type), contingent payments can be used to provide additional incentives. The reason is that different types then have different preferences over ‘side bets’ (monetary gambles on the realisation of the signal), which can be exploited to induce self-selection.

The power of side bets is most clearly illustrated in the case in which *all* types have ‘separate’ beliefs. The appropriate notion of ‘separate’ is slightly stronger than ‘distinct’ ($\mu(r) \neq \mu(t)$ for all $r, t \in \mathcal{T}$): it also requires that different types’ beliefs not be linearly dependent in a weak (‘convex’) sense. Precisely:

Definition 1. When the set \mathcal{T} of types is finite, the beliefs $\mu : \mathcal{T} \rightarrow \Delta$ are said to satisfy *convex independence* iff no type’s belief may be written as a convex combination of *other* types’ beliefs: for each $t \in \mathcal{T}$, $\mu(t)$ lies outside the convex hull of $\{\mu(r) : r \in \mathcal{T} \setminus \{t\}\}$.

More explicitly, convex independence requires that for no $t \in \mathcal{T}$ can we write $\mu(t) = \sum_{r \in \mathcal{T} \setminus \{t\}} \lambda_r \mu(r)$ for a collection $(\lambda_r)_{r \in \mathcal{T} \setminus \{t\}} \subseteq [0, 1]$ of weights satisfying $\sum_{r \in \mathcal{T} \setminus \{t\}} \lambda_r = 1$. This is weaker than *linear independence* of the vectors $\{\mu(t) : t \in \mathcal{T}\}$ because only weights $(\lambda_r)_{r \in \mathcal{T} \setminus \{t\}}$ that sum to unity are contemplated.

Exercise 6. Let \mathcal{T} be finite, fix beliefs $M := \{\mu(t) : t \in \mathcal{T}\} \subseteq \Delta$, and let $\text{co } M$ denote the convex hull of M . If beliefs M satisfy convex independence, must $\mu(t)$ be an extreme point of $\text{co } M$, for every t ? If $\mu(t)$ is an extreme point of $\text{co } M$ for every t , must M satisfy convex independence?

If $|\mathcal{X}| > |\mathcal{T}|$, then convex independence is ‘generic’, in the formal sense that if we draw each type’s belief independently from the Lebesgue measure on Δ , then the probability that convex independence fails is zero.

Theorem 2. If the set \mathcal{T} of types is finite and the beliefs $\mu : \mathcal{T} \rightarrow \Delta$ satisfy convex independence, then for any mechanism (q, p) , there is a contingent payment rule $p' : \mathcal{T} \rightarrow \mathbf{R}^{|\mathcal{X}|}$ such that (q, p') is incentive-compatible and $p(t) \cdot \mu(t) = p'(t) \cdot \mu(t)$ for every type $t \in \mathcal{T}$.

In other words, any mechanism can be made incentive-compatible without changing either the allocation or the expected payment of any type $t \in \mathcal{T}$.

Corollary 2. Under the hypotheses of Theorem 2, every allocation $q : \mathcal{T} \rightarrow [0, 1]$ is implementable.

A much-emphasised consequence of Theorem 2 is the following:

Corollary 3. Under the hypotheses of Theorem 2, a revenue-maximising seller can design a ‘full-extraction’ mechanism that induces every type t to pay, in expectation, her full valuation t for the good.

Proof. Start with the mechanism (q, p) given by $(q(t), p(t)) := (1, t)$ for every $t \in \mathcal{T}$. Theorem 2 delivers a payment rule p' under which each type t (still) pays $p'(t) \cdot \mu(t) = p(t) \cdot \mu(t) = t$ in expectation, and such that (q, p') is incentive-compatible. Finally, since (q, p) is individually rational by inspection, the same is true of (q, p') . ■

Proof of Theorem 2. Fix an arbitrary mechanism (q, p) . For each type $t \in \mathcal{T}$, by convex independence and the (strict) separating hyperplane theorem, there is a (‘side bet’) $b(t) \in \mathbf{R}^{|\mathcal{X}|}$ that satisfies $b(t) \cdot \mu(t) = 0 < b(t) \cdot \mu(r)$ for every $r \in \mathcal{T} \setminus \{t\}$.

Define a new contingent payment rule $p' : \mathcal{T} \rightarrow \mathbf{R}^{|\mathcal{X}|}$ by $p'(t) := p(t) + \alpha b(t)$ for each $t \in \mathcal{T}$, where $\alpha > 0$ is a constant large enough that

$$\alpha > \max_{r \neq t \text{ in } \mathcal{T}} \frac{-[q(t) - q(r)]t + [p(t) - p(r)] \cdot \mu(t)}{b(r) \cdot \mu(t)}.$$

Clearly $p'(t) \cdot \mu(t) = p(t) \cdot \mu(t)$ for every $t \in \mathcal{T}$. And for all $r \neq t$ in \mathcal{T} , we have

$$\begin{aligned} q(t)t - p'(t) \cdot \mu(t) &= q(t)t - p(t) \cdot \mu(t) \\ &> q(r)t - [p(r) + \alpha b(r)] \cdot \mu(t) = q(r)t - p'(r) \cdot \mu(t), \end{aligned}$$

where the inequality holds by our choice of α . ■

In the proof, the ‘side bet’ $b(t)$ is a monetary gamble on the signal; it has zero expected value for type t , but strictly positive expected value for all the other types. By tacking a sufficiently blown-up version of this gamble onto the payment rule p , we ensure that the agent’s incentive to choose the right *gamble* overwhelms all other considerations: she reports truthfully regardless.

The usual interpretation of these results is that (essentially) *any* correlation between the signal and type sharply alters the model’s predictions: incentive-provision becomes trivial. This is widely felt to be very disturbing. There are several reasons for this; one of them is large side bets are rarely seen in the wild.

I encourage you to reflect on the significance and implications of these results. I also encourage you think about what sorts of modifications of the model would break the proof.

Exercise 7. Suppose that the agent is risk-averse: type t ’s payoff when her (ex-post) payment is $p \in \mathbf{R}$ is $t - \psi(p)$ in case she gets the good and $-\psi(p)$ otherwise, where $\psi : \mathbf{R} \rightarrow \mathbf{R}$ is convex and strictly increasing. (We have relaxed risk-neutrality, but *not* strong separability between the good and money—see the discussion in §1.9.) Prove that Corollary 2 remains true. (Hint: Theorem 2 is not true as stated... but a version of it is.) Do you think that Corollary 3 remains true?

The literature. The extraordinary power of side bets was glimpsed by Myerson (1981, §7), who gave an example, and then demonstrated at increasing levels of generality by Crémer and McLean (1985, 1988) and McAfee and Reny (1992). Lopomo, Rigotti and Shannon (2022) provide a careful and insightful synthesis of such results, paying special attention to the case (neglected in this section) in which the set \mathcal{T} of types is infinite (e.g. $\mathcal{T} = [0, 1]$).

There is a cottage industry devoted to ‘debunking’ these results. For example, the logic fails if payments cannot be unboundedly large, e.g. due to an ex-post IR or ‘limited liability’ constraint (Robert, 1991), or if the good and money are not strongly separable in the agent’s preferences (Robert, 1991; Eső, 2005). More subtly, Neeman (2004) argues that convex independence is not so ‘generic’ after all.¹⁴ Others have highlighted that large side bets have problematic incentive side effects not captured by the model: for example, they provide strong incentives to collude when the signal is a second agent’s

¹⁴This nice paper sparked an arcane literature about whether convex independence is ‘generic’ in the universal type space. The answer: it depends (greatly) on what is meant by ‘generic’. (*Universal what?! See my lecture notes from Eddie Dekel’s course ‘Economic theory and methods’, available at ludvigsinander.net/lecture_notes.*)

type (Laffont & Martimort, 2000), and also to acquire information (at a cost) about the realisation of the signal (Laohakunakorn, 2023).

1.9 Addendum: preference assumptions

In discussions of the model of this chapter (and other models like it), there is often much confusion about what ‘risk-neutrality’ means and what role it plays. In this section, I clarify.

In the model of this chapter, the payoff-relevant ex-post outcomes are $(g, p) \in \{0, 1\} \times \mathbf{R}$, where $g = 1$ iff the agent gets the good. A *lottery* (over outcomes) is a probability measure on $\{0, 1\} \times \mathbf{R}$. Each type of the agent has a preference over lotteries. *Expected-utility (EU) preferences* over lotteries ℓ are those that admit a *linear* representation: $U(\ell) = \int_{\{0,1\} \times \mathbf{R}} u d\ell$ for some (‘utility index’) $u : \{0, 1\} \times \mathbf{R} \rightarrow \mathbf{R}$. (I have omitted the measure-theoretic details.) You will recall that u is unique up to affine transformations: if (and only if) $U(\ell) = \int_{\{0,1\} \times \mathbf{R}} u d\ell$ and $V(\ell) = \int_{\{0,1\} \times \mathbf{R}} v d\ell$ represent the same EU preference over lotteries, then $u = \alpha + \beta v$ for some $\alpha \in \mathbf{R}$ and $\beta > 0$.

Risk attitude is a property of preferences over *monetary* lotteries, not over joint good–money lotteries ℓ . We may contemplate the ‘risk attitude’ of the agent’s *marginal* preferences, however: that is, her preferences among lotteries ℓ that all assign the good for sure ($\ell(\{1\} \times \mathbf{R}) = 1$), and also among those that certainly do not ($\ell(\{0\} \times \mathbf{R}) = 1$). The agent’s marginal preferences are risk-neutral (risk-averse) iff $u(0, \cdot)$ and $u(1, \cdot)$ are both affine (concave). In this case, we shall say that ‘the agent is risk-neutral (risk-averse)’.

EU preferences are *monotone* iff $u(g, p)$ is strictly increasing in g and strictly decreasing in p , and *strongly separable* iff u is additively separable: $u(g, p) = \phi(g) - \psi(p)$ for some $\phi : \{0, 1\} \rightarrow \mathbf{R}$ and $\psi : \mathbf{R} \rightarrow \mathbf{R}$. By inspection, for strongly separable EU preferences, (i) given monotonicity, we may choose $\phi(g) = gt$ for any $t > 0$, and (ii) under risk-neutrality (risk-aversion), ψ is affine (convex).

Our assumption in this chapter was that each type $t \in (0, 1]$ has *monotone and quasi-linear EU preferences*, meaning an index of the form $u(g, p) = gt - p$; in other words, strongly separable, monotone and risk-neutral EU preferences. If risk-neutrality were weakened to risk-aversion, then we would instead have $u(g, p) = gt - \psi(p)$ for some convex and strictly increasing $\psi : \mathbf{R} \rightarrow \mathbf{R}$.

Most of the analysis in this chapter remains valid in the latter (risk-averse) case. Simply ‘change currency’ from dollars to utils via $p \mapsto \psi(p)$: in this currency, the agent is risk-neutral, so preferences are quasi-linear again, and thus everything that we said about incentive-compatibility remains true.

Differences arise only when maximising the seller's revenue, since the seller cares about revenue denominated in dollars, not utils. (In the language of cooperative game theory: when both agent and principal were risk-neutral, utility was perfectly transferable between them; now it is not.)

The analysis in this chapter is less robust to dropping strong separability (e.g. $u(g, p) = \chi(gt - p)$ for some concave and strictly increasing $\chi : \mathbf{R} \rightarrow \mathbf{R}$). Some results remain true (e.g. the Spence–Mirrlees lemma). But others do not, for example the results in §1.8 about the extraordinary incentive power of side bets (see Robert, 1991).

1.10 Addendum: the envelope theorem, rigorously

In §1.2, we gave a heuristic treatment of the envelope theorem, arguing that the envelope formula is intuitively a rewritten form of a first-order condition capturing ‘local optimality’; thus optimality implies (local optimality and thus) the envelope formula. In this section, we provide a rigorous treatment.¹⁵

Formalities. We will be working with the unit interval $[0, 1]$. We equip it with the Lebesgue σ -algebra and the Lebesgue measure. All integrals will be Lebesgue integrals (with respect to the Lebesgue measure).

Recall that for a constant $K \geq 0$, a function $\phi : [0, 1] \rightarrow \mathbf{R}$ is called *K-Lipschitz continuous* iff $|\phi(t) - \phi(r)| \leq K|t - r|$ for all $r, t \in [0, 1]$. (If ϕ is continuously differentiable, then it is *K-Lipschitz* iff $|\phi'| \leq K$. [Prove it!]) A function is *Lipschitz continuous* iff it is *K-Lipschitz* for some $K \geq 0$.

Absolute continuity is a continuity concept that is weaker than Lipschitz continuity, but still stronger than ordinary continuity.¹⁶ The fundamental theorem of calculus for Lebesgue integrals (see e.g. Folland, 1999, §3.5, p. 106) states that a function $\phi : [0, 1] \rightarrow \mathbf{R}$ is absolutely continuous iff it is an integral¹⁷ iff it is the integral of its own derivative.¹⁸

The setting is a (smoothly) parametrised maximisation problem: there are *actions* \mathcal{X} , *parameters* $[0, 1]$, and an *objective function* $f : \mathcal{X} \times [0, 1] \rightarrow \mathbf{R}$. The interpretation is that $f(\cdot, t)$ is the objective to be maximised, when the parameter is $t \in [0, 1]$.

¹⁵The text of this section is partly drawn from Sinander (2022).

¹⁶The definition: $\phi : [0, 1] \rightarrow \mathbf{R}$ is absolutely continuous iff for each $\varepsilon > 0$, there is a $\delta > 0$ such that for any finite collection $\{(r_n, t_n)\}_{n=1}^N$ of disjoint intervals of $[0, 1]$, $\sum_{n=1}^N (t_n - r_n) < \delta$ implies $\sum_{n=1}^N |\phi(t_n) - \phi(r_n)| < \varepsilon$.

¹⁷There is an integrable $\psi : [0, 1] \rightarrow \mathbf{R}$ such that $\phi(t) = \phi(0) + \int_0^t \psi$ for every $t \in [0, 1]$.

¹⁸ ϕ is differentiable a.e., its (a.e.-defined) derivative $\phi' : [0, 1] \rightarrow \mathbf{R}$ is integrable, and $\phi(t) = \phi(0) + \int_0^t \phi'$ for every $t \in [0, 1]$.

The action set \mathcal{X} is completely arbitrary—it need not have any convex, topological or measurable structure, for example. The assumption that the parameter $t \in [0, 1]$ is one-dimensional is less restrictive than it may appear; see Remark 4 at the end of this section.

As for the objective function, our only assumptions will be that it varies smoothly with the parameter and that this derivative is not ‘very crazy’. Specifically:

Basic assumptions. $f(x, \cdot)$ is differentiable for every $x \in \mathcal{X}$, and the family $\{f(x, \cdot)\}_{x \in \mathcal{X}}$ of functions $[0, 1] \rightarrow \mathbf{R}$ is an *absolutely equi-continuous* family.

See Sinander (2022) for the definition of absolute equi-continuity. A simple sufficient condition is that the derivative f_2 be bounded (there is a constant $K \geq 0$ such that $|f_2(x, t)| \leq K$ for all $x \in \mathcal{X}$ and $t \in [0, 1]$).

Example 1. In the model considered throughout this chapter, the ‘actions’ are outcomes $x = (q, p) \in [0, 1] \times \mathbf{R} =: \mathcal{X}$, and the (agent’s) objective is $f(x, t) = f((q, p), t) := qt - p$. Clearly $f(x, \cdot)$ is differentiable for each $x \in \mathcal{X}$; furthermore, the derivative is bounded (why?).

A *decision rule* is a map $X : [0, 1] \rightarrow \mathcal{X}$ that prescribes an action for each parameter. The *value function* of a decision rule X is $V_X(t) := f(X(t), t)$. A decision rule X is *optimal* iff $V_X(t) \geq f(x, t)$ for all $x \in \mathcal{X}$ and $t \in [0, 1]$.

A decision rule X satisfies the *envelope formula* iff

$$V_X(t) = V_X(0) + \int_0^t f_2(X(s), s) ds \quad \text{for every } t \in [0, 1].$$

Equivalently, by Lebesgue’s fundamental theorem of calculus, X satisfies the envelope formula iff V_X is *absolutely continuous* (hence a.e. differentiable) and satisfies $V_X'(t) = f_2(X(t), t)$ for a.e. $t \in [0, 1]$.

Milgrom–Segal envelope theorem. If f satisfies the basic assumptions, then any optimal decision rule satisfies the envelope formula.

This is really a slight refinement of Theorem 2 in Milgrom and Segal (2002), because the assumptions are weaker. You can find this exact version in Sinander (2022).

Example 1 (continued). In the model considered throughout this chapter, ‘decision rules’ are mechanisms $X = (q, p) : [0, 1] \rightarrow \mathcal{X}$. Fix a mechanism X , and assume without loss that it is onto \mathcal{X} ;¹⁹ then X is incentive-compatible

¹⁹This is without loss because we may re-define \mathcal{X} to be the image of X . (This has no effect on incentives, since it removes from consideration those actions/outcomes x that never arise under the mechanism X .)

exactly if it is ‘optimal’ in the sense defined above. Thus the Milgrom–Segal envelope theorem implies the Mirrlees envelope theorem from §1.2: every incentive-compatible mechanism satisfies the envelope formula.

The Milgrom–Segal envelope theorem may be decomposed into two lemmata: the former asserts that the derivatives $V'_X(t)$ and $f_2(X(t), t)$ are equal whenever they (both) exist, and the latter says that V_X may be written as the integral of its derivative.

Lemma 1. If X is optimal, then for any $t \in (0, 1)$ at which $V'_X(t)$ and $f_2(X(t), t)$ exist, we have $V'_X(t) = f_2(X(t), t)$.

Proof. For any $r < t$, we have

$$V_X(t) - V_X(r) = f(X(t), t) - f(X(r), r) \leq f(X(t), t) - f(X(t), r),$$

so that dividing both sides by $t - r$ and sending $r \uparrow t$ yields $V'_X(t) \leq f_2(X(t), t)$. Similarly, for any $r > t$ we have $V_X(r) - V_X(t) \geq f(X(t), r) - f(X(t), t)$, so that dividing by $r - t$ and sending $r \downarrow t$ delivers $V'_X(t) \geq f_2(X(t), t)$. ■

Lemma 2. If f satisfies the basic assumptions, then for any optimal decision rule X , the value V_X is absolutely continuous.

We’ll prove a special case. (For a proof of the general case, see Sinander (2022)—the relevant result is labelled ‘the necessity lemma’.)

Proof of a special case. Assume a stronger version of the basic assumptions: f_2 exists and is *bounded*. Write $K \geq 0$ for the constant that bounds f_2 . Fix an optimal X ; we will show that V_X is *K-Lipschitz continuous* (hence absolutely continuous).

So fix any $r < t$ in $[0, 1]$, and let $B := \sup_{x \in \mathcal{X}} |f(x, t) - f(x, r)|$. Then

$$V_X(t) - V_X(r) \leq f(X(t), t) - f(X(t), r) \leq \sup_{x \in \mathcal{X}} [f(x, t) - f(x, r)] \leq B$$

and (similarly)

$$V_X(t) - V_X(r) \geq f(X(r), t) - f(X(r), r) \geq \inf_{x \in \mathcal{X}} [f(x, t) - f(x, r)] \geq -B,$$

so

$$|V_X(t) - V_X(r)| \leq B = \sup_{x \in \mathcal{X}} \left| \int_r^t f_2(x, s) ds \right| \leq K(t - r)$$

by the fundamental theorem of calculus and the bound $|f_2| \leq K$. ■

Proof of the Milgrom–Segal envelope theorem. Let f satisfy the basic assumptions, and let $X : [0, 1] \rightarrow \mathcal{X}$ be optimal. Then V_X is absolutely continuous by Lemma 2, so (by Lebesgue’s fundamental theorem of calculus) V'_X exists a.e. and $V_X(t) = V_X(0) + \int_0^t V'_X$ for every $t \in [0, 1]$. For every $t \in [0, 1]$ at which $V'_X(t)$ exists, $f_2(X(t), t)$ also exists by the basic assumptions, so $V'_X(t) = f_2(X(t), t)$ by Lemma 1. ■

The literature. The envelope theorem originated in the theories of the consumer and firm (Hotelling, 1932; Roy, 1947; Shephard, 1953), and was then considered more abstractly by Samuelson (1947). The perspective on the envelope theorem presented in this section, and the Milgrom–Segal envelope theorem in which it culminates, are due to Milgrom and Segal (2002).

The *textbook* perspective on the envelope theorem (e.g. Mas-Colell, Whinston & Green, 1995, §M.L) is quite different. The argument is typically that optimality implies *local optimality* in the form of a first-order condition, and that this first-order condition may be re-expressed as the envelope formula. (We saw a version of this reasoning in §1.2.) A nice feature of this perspective is that it tells you what the envelope formula *means* economically: namely, ‘local optimality’. The downside is that the textbook argument requires very strong assumptions—far too strong for mechanism-design applications.

But it is in fact possible to obtain the general Milgrom–Segal envelope theorem from the textbook argument (with some frills), as long as one uses a suitable notion of ‘local optimality’. It is therefore quite generally legitimate to interpret the envelope formula as ‘local optimality’. The details are in Sinander (2022).

Remark 4. Suppose that instead of a one-dimensional parameter $t \in [0, 1]$, there is a parameter θ drawn from a convex subset Θ of a normed vector space (e.g. \mathbf{R}^n , or a higher-dimensional space). Write $F : \mathcal{X} \times \Theta \rightarrow \mathbf{R}$ for the objective function. Then along any smooth path $[0, 1] \ni t \mapsto \theta_t \in \Theta$ through the parameter space Θ , we may define $f(x, t) := F(x, \theta_t)$; the envelope theorem applies along this path, and that is true for any such smooth path $t \mapsto \theta_t$.

Chapter 2

Allocating among several agents

‘Who gets what?’ is perhaps the most fundamental question in economics. This chapter concerns research on that question.

The setting will be the same throughout: there are $I \geq 2$ agents, each of whom has a ‘type’ T_i that is randomly drawn from some set \mathcal{T}_i according to a probability measure μ_i (defined on some σ -algebra). Draws are independent across agents.¹ In most (but not all) models studied in this chapter, each agent i privately knows her type $t_i \in \mathcal{T}_i$ from the outset.

There is a single indivisible good to be allocated. When convenient, we think of this allocation as performed by a ‘principal’. The good may be either assigned to an agent or kept by the principal. The agents’ preferences and the principal’s objective will vary from model to model.

The standard IPV (independent private values) auction model fits this framework. Here agent i ’s type t_i is her privately-known valuation of the good. In an *auction*, the agents take some actions (often called ‘bids’), and as a function of these, the good is allocated and monetary payments are made by the agents. Agents’ payoffs are quasi-linear in money: $q_i t_i - p_i$, where $q_i \in [0, 1]$ is i ’s probability of getting the good, and $p_i \in \mathbf{R}$ is her payment.

In many models considered in this chapter, monetary transfers are not available. The design of incentives without the use of monetary transfers is currently a fairly hot topic.

Let’s fix some terminology and notation. An *allocation* is a (measurable)

¹This is important. When types are correlated, it is ‘typically’ possible to incentivise agents to do essentially anything by offering them monetary bets on the other agents’ types—see §1.8.

function $q : \mathcal{T}_1 \times \cdots \times \mathcal{T}_I \rightarrow [0, 1]^I$ that satisfies $\sum_{i=1}^I q_i \leq 1$; the interpretation is that when types are $\mathbf{t} \in \mathcal{T}_1 \times \cdots \times \mathcal{T}_I$, agent i gets the good with probability $q_i(\mathbf{t})$. We'll use the notation

$$(t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_I) =: \mathbf{t}_{-i} \in \mathcal{T}_{-i} := \mathcal{T}_1 \times \cdots \times \mathcal{T}_{i-1} \times \mathcal{T}_{i+1} \times \cdots \times \mathcal{T}_I.$$

In many models, each agent i privately knows her type t_i . We can imagine three stages: an *ex-ante* stage at which agents do not know anyone's type, an *interim* stage at which agents know their own types but not those of others, and an *ex-post* stage at which all agents' types are common knowledge.

At the interim stage, given allocation q , agent i expects to receive the good with probability

$$Q_i(t_i) := \mathbf{E}_{\mathbf{T}_{-i} \sim \mu_{-i}}[q_i(t_i, \mathbf{T}_{-i})].$$

This family $(Q_i)_{i=1}^I$ of (measurable) functions, where $Q_i : \mathcal{T}_i \rightarrow [0, 1]$, is called the *interim allocation* induced by the (ex-post) allocation q .

In all models we consider, agents have expected-utility preferences, so that their payoffs are linear in q ; then if agents know their types when they take decisions, it only is the interim allocation that matters for their incentives. Similarly, if the principal has an expected-utility objective that is additively separable across agents, then only the interim allocation matters to her.

2.1 Border's theorem

Whereas an allocation q is a high-dimensional object, its induced interim allocation $(Q_i)_{i=1}^I$ is not. For this reason, it is frequently easier to work directly with interim allocations. To adopt this approach, we must do some housekeeping first: we must determine which families $(Q_i)_{i=1}^I$ of functions are legitimately the interim allocation induced by some allocation q .

Suppose you give me a *candidate interim allocation*, meaning a collection $(Q_i)_{i=1}^I$ of (measurable) functions, where Q_i maps \mathcal{T}_i into $[0, 1]$. Can I always construct an allocation q that realises this collection as its interim allocation? (No: consider $\mathcal{T}_i = [0, 1]$, μ_i with full support, and $Q_i(t_i) = 0$ if $t_i < 1/2$ and $Q_i(t_i) = 1$ otherwise, for all agents i .)

How, then, can I tell whether a given collection $(Q_i)_{i=1}^I$ is realisable by some q ? Border's theorem answers this question.

Asymmetric Border's theorem. For a candidate interim allocation $(Q_i)_{i=1}^I$, the following are equivalent:

- (1) $(Q_i)_{i=1}^I$ is realised by some allocation.

- (2) The Border inequality

$$\sum_{i=1}^I \int_{T_i} Q_i d\mu_i \leq 1 - \prod_{i=1}^I (1 - \mu_i(T_i))$$

holds for all (measurable) sets $T_1 \subseteq \mathcal{T}_1, \dots, T_I \subseteq \mathcal{T}_I$ of types.

- (3) For any $\alpha_1, \dots, \alpha_I \in [0, 1]$, the Border inequality holds for the sets $T_i = \{t_i \in \mathcal{T}_i : Q_i(t_i) \geq \alpha_i\}$.

Item (3) is valuable because it reduces the number of inequalities that must be checked to verify realisability (which could be very large if we rely on (2)) to a finite-dimensional family of inequalities.

We won't prove this; instead we'll introduce and prove a symmetric version. Call the environment *symmetric* iff all agents' types are drawn from the same type space, call it \mathcal{T} , according to the same distribution, call it μ . A *symmetric candidate interim allocation* is a (measurable) map $Q : \mathcal{T} \rightarrow [0, 1]$; we interpret this as the candidate interim allocation $(Q_i)_{i=1}^I$ in which $Q_i = Q$ for every i . An allocation q is *symmetric* iff agents' treatment does not depend on their names/labels: for any type profiles \mathbf{t} and \mathbf{t}' that differ only in that i 's and j 's types are reversed, $q(\mathbf{t})$ and $q(\mathbf{t}')$ differ only in that i 's and j 's probabilities are reversed.²

Symmetric Border's theorem. Let the environment be symmetric. For a symmetric candidate interim allocation $Q : \mathcal{T} \rightarrow [0, 1]$, the following are equivalent:

- (1) Q is realised by some symmetric allocation.
 (2) The Border inequality

$$I \int_{\mathcal{T}} Q d\mu \leq 1 - (1 - \mu(\mathcal{T}))^I$$

holds for all (measurable) sets $T \subseteq \mathcal{T}$ of types.

- (3) For any $\alpha \in [0, 1]$, the Border inequality holds for the set $T = \{t \in \mathcal{T} : Q(t) \geq \alpha\}$.

We'll prove the equivalence of (1) and (2). The necessity of the Border inequalities for realisability is straightforward:

²That is: for any $\mathbf{t} = (t_1, \dots, t_I)$ and $\mathbf{t}' = (t'_1, \dots, t'_I)$ such that $t_i = t'_j$, $t_j = t'_i$ and $t_k = t'_k$ for every $k \in \{1, \dots, I\} \setminus \{i, j\}$, we have $q_i(\mathbf{t}) = q_j(\mathbf{t}')$, $q_j(\mathbf{t}) = q_i(\mathbf{t}')$, and $q_k(\mathbf{t}) = q_k(\mathbf{t}')$ for every $k \in \{1, \dots, I\} \setminus \{i, j\}$.

Proof that (1) implies (2). If Q is realised by some symmetric q , then for any measurable $T \subseteq \mathcal{T}$,

$$\begin{aligned}
\int_T Q d\mu &= \mathbf{P}_{\mu,q}(\text{Ms. 1 has type in } T \text{ and gets the good}) \\
&= \mathbf{P}_{\mu,q}(\text{someone has type in } T) \\
&\quad \times \mathbf{P}_{\mu,q}\left(\begin{array}{c} \text{someone with type} \\ \text{in } T \text{ gets the good} \end{array} \middle| \text{someone has type in } T\right) \\
&\quad \times \mathbf{P}_{\mu,q}\left(\begin{array}{c} \text{Ms. 1 has type in } T \\ \text{and gets the good} \end{array} \middle| \begin{array}{c} \text{someone has type in } T \\ \text{and gets the good} \end{array}\right) \\
&\leq \mathbf{P}_{\mu,q}(\text{someone has type in } T) \\
&\quad \times \mathbf{P}_{\mu,q}\left(\begin{array}{c} \text{Ms. 1 has type in } T \\ \text{and gets the good} \end{array} \middle| \begin{array}{c} \text{someone has type in } T \\ \text{and gets the good} \end{array}\right) \\
&= \left[1 - (1 - \mu(T))^I\right] \times \frac{1}{I}
\end{aligned}$$

by symmetry. ■

For sufficiency, call a symmetric allocation q *hierarchical* iff for some pairwise disjoint and non-empty (measurable) sets $R_1, \dots, R_K \subseteq \mathcal{T}$,

- If there are agents with type in R_1 , then the good is allocated among these agents by a fair lottery.
- If for some $k \in \{1, \dots, K\}$, there are no agents with type in $R_1 \cup \dots \cup R_{k-1}$, but there are agents with type in R_k , then the good is allocated among these agents by a fair lottery.

Hierarchical allocations are special because they make the Border inequalities tight for certain sets:

Lemma 3. If a symmetric allocation q is hierarchical with type sets R_1, \dots, R_K , then for each $k \in \{1, \dots, K\}$, it satisfies the Border inequality for the set $T = R_1 \cup \dots \cup R_k$ with equality.

Proof. Fix a $k \in \{1, \dots, K\}$, and write $T := R_1 \cup \dots \cup R_k$. We have

$$\begin{aligned}
 \int_T Q d\mu &= \mathbf{P}_{\mu,q}(\text{Ms. 1 has type in } T \text{ and gets the good}) \\
 &= \mathbf{P}_{\mu,q}(\text{someone has type in } T) \\
 &\quad \times \mathbf{P}_{\mu,q}\left(\text{someone with type in } T \text{ gets the good} \mid \text{someone has type in } T\right) \\
 &\quad \times \mathbf{P}_{\mu,q}\left(\text{Ms. 1 has type in } T \text{ and gets the good} \mid \text{someone has type in } T \text{ and gets the good}\right) \\
 &= \left[1 - (1 - \mu(T))^I\right] \times 1 \times \frac{1}{I}. \quad \blacksquare
 \end{aligned}$$

Border's (1991) sufficiency proof uses hierarchical allocations and a separating-hyperplane argument. Since separating hyperplanes are simpler in finite-dimensional spaces, we'll restrict our attention to the case of finitely many types.³ (To learn how to extend this reasoning to infinitely many types, see Border (1991).)

Proof that (2) implies (1), finite case. Assume that \mathcal{T} is finite, and (wlog) that μ is strictly positive. We shall prove the contra-positive: fix a candidate interim allocation Q that is not realisable; we shall find a set $T \subseteq \mathcal{T}$ on which the relevant Border inequality fails.

The set of all realisable allocations is convex (why?), and is easily seen (how?) to be closed. Thus by the strict separating hyperplane theorem, there are weights $(\gamma_t)_{t \in \mathcal{T}}$, not all zero, such that

$$\sum_{t \in \mathcal{T}} \gamma_t \mu(t) Q(t) > \sum_{t \in \mathcal{T}} \gamma_t \mu(t) Q'(t) \quad \text{for any realisable } Q'.^4$$

We may perturb the γ s to make them all distinct. Note that $\mathcal{T}^+ := \{t \in \mathcal{T} : \gamma_t > 0\}$ is non-empty since 0 is realisable.

Let's re-name the types in \mathcal{T}^+ as $\mathcal{T}^+ \equiv \{1, \dots, |\mathcal{T}^+|\}$, where $\gamma_1 > \gamma_2 > \dots > \gamma_{|\mathcal{T}^+|} > 0$. Consider the hierarchical allocation q^* with type sets $\{1\}, \{2\}, \dots, \{|\mathcal{T}^+|\}$. Observe that

$$\sum_{t \in \mathcal{T}^+} \gamma_t \mu(t) Q(t) \geq \sum_{t \in \mathcal{T}} \gamma_t \mu(t) Q(t) > \sum_{t \in \mathcal{T}} \gamma_t \mu(t) Q^*(t) = \sum_{t \in \mathcal{T}^+} \gamma_t \mu(t) Q^*(t)$$

³I learned this finite-types version of Border's argument from Vohra (2011, §6.2.3).

⁴More pedantically, the separating hyperplane theorem yields weights $(\Gamma_t)_{t \in \mathcal{T}}$ such that $\sum_{t \in \mathcal{T}} \Gamma_t Q(t) > \sum_{t \in \mathcal{T}} \Gamma_t Q'(t)$ for any realisable Q' , and we may define $\gamma_t := \Gamma_t / \mu(t)$ for each $t \in \mathcal{T}$.

since $\gamma_t \leq 0$ and $Q^*(t) = 0$ for $t \in \mathcal{T} \setminus \mathcal{T}^+$. Telescoping on both sides yields

$$\sum_{t'=1}^{|\mathcal{T}^+|} \beta_{t'} \left(\sum_{t=1}^{t'} \mu(t) Q(t) \right) > \sum_{t'=1}^{|\mathcal{T}^+|} \beta_{t'} \left(\sum_{t=1}^{t'} \mu(t) Q^*(t) \right),$$

where $\beta_{t'} := \gamma_{t'} - \gamma_{t'+1}$ for $t' < |\mathcal{T}^+|$ and $\beta_{|\mathcal{T}^+|} := \gamma_{|\mathcal{T}^+|}$.

Since the weights $\beta_{t'}$ are positive, there must be a $t' \in \{1, \dots, |\mathcal{T}^+|\}$ at which

$$\sum_{t=1}^{t'} \mu(t) Q(t) > \sum_{t=1}^{t'} \mu(t) Q^*(t) = \left[1 - (1 - \mu(\{1, \dots, t'\}))^I \right] \frac{1}{I},$$

where the equality holds by Lemma 3. The left-hand side is $\int_{\{1, \dots, t'\}} Q d\mu$; thus Q violates the Border inequality for the set $T = \{1, \dots, t'\}$. ■

Exercise 8. Complete the proof of the symmetric Border theorem by showing that (2) is equivalent to (3).

The literature. The symmetric realisability problem was studied first. Maskin and Riley (1984) took the first steps. Matthews (1984) proved the necessity of the Border inequalities, and conjectured sufficiency. The conjecture was proved by Border (1991), using the above argument above based on hierarchical allocations and a separating hyperplane. Border (2007) obtained the asymmetric version for finitely many types, and Mierendorff (2011) extended this to general type spaces. Hart and Reny (2015a) provide an alternative characterisation of realisability in terms of second-order stochastic dominance. Some (not many) extensions to richer environments have been considered; for example, Che, Kim and Mierendorff (2013) allow for certain sorts of constraints.

2.2 Incentives

We shall always assume that every type of every agent likes the good (and is selfish). If the principal cannot directly monitor the agents' private types, and if she controls no payoff-relevant decisions apart from the allocation of the good, then screening (separation of different types) is impossible: it is possible to implement all and only allocations that do not depend on agents' types. (Why?)

In the sequel, we consider a variety of settings in which the principal *can* (to some extent) screen the agents. Some of these models grant the principal

control of a payoff-relevant action besides the allocation of the good, such as monetary payments by the agent (§2.3) or allocation decisions in future periods (chapter 3). With such an ‘incentive instrument’, the principal can screen an agent by forcing her to trade off her probability of getting the good against another outcome about which she cares. (For example, the principal can let the agent increase her likelihood of obtaining the good in exchange for making a higher payment, or for getting the good with a lower probability tomorrow.)

In other models, the principal has no incentive instrument besides her control of the allocation, but she has access to a monitoring technology. For example, the principal may be able perfectly to verify an agent’s type at a cost (§2.4), or she may possess a free but noisy monitoring technology (§2.6).

There are other possibilities that we will *not* consider in these notes. For example, the principal may be able perfectly to verify agents’ types ex post, but have recourse only to limited punishments for liars who get caught (see Mylovanov & Zapechelnyuk, 2017).

2.3 IPV auctions

The *IPV (independent private values) model* is the following. Agents’ types are drawn from $\mathcal{T}_i = [0, 1]$, and are called *valuations*. (As before, they are statistically independent, and each agent learns only her own type.) Payoffs are as in chapter 1: if agent i has type $t_i \in [0, 1]$ and pays $p_i \in \mathbf{R}$, then her payoff is $t_i - p_i$ in case she gets the good, and $-p_i$ otherwise.

The IPV model is used to study *auctions*, meaning mechanisms in which agents’ actions are non-negative numbers, called ‘bids’. We shall continue to allow for arbitrary mechanisms, though it will turn out that all the mechanisms of interest can be implemented using various familiar auctions formats.

2.3.1 Equivalence of dominant-strategy and Bayes–Nash IC

As before, an *allocation* is a map $q : [0, 1]^I \rightarrow [0, 1]^I$ such that $\sum_{i=1}^I q_i \leq 1$. A remarkable fact about the IPV setting is that frequently, an allocation which can be implemented as the Bayes–Nash equilibrium of some auction can in fact be implemented as a *dominant-strategy* equilibrium of some (other) auction.

Note well. In mechanism design, a ‘dominant strategy’ of a player in a game is one that yields weakly better payoff than any other strategy, whatever

the strategies of the other players. This is weaker than the game-theoretic notion of weak dominance! (Spot the difference!)

For example, an *efficient* allocation is one that allocates the good always to one of those agents whose valuation is highest. The first-price auction implements any efficient allocation as a Bayes–Nash equilibrium. (Convince/remind yourself.) But consider the second-price auction: it also has an equilibrium implementing any efficient allocation, and in this equilibrium (in which each agent always bids her true value), agents’ strategies are in fact weakly dominant.

This was until recently a somewhat baffling phenomenon, but no longer. A *payment rule* is a map $p : [0, 1]^I \rightarrow \mathbf{R}^I$. A *direct revelation mechanism* (DRM) is a pair (q, p) , where q is an allocation and p is a payment rule. A direct revelation mechanism is called *Bayes–Nash (dominant-strategy) incentive-compatible* iff truthful reporting is a Bayes–Nash equilibrium (a weakly dominant strategy profile). By a revelation principle, if an allocation–payment rule pair (q, p) is induced by some Bayes–Nash equilibrium (by some profile of weakly dominant strategies) of some mechanism, then (q, p) (viewed as a DRM) is Bayes–Nash IC (dominant-strategy IC).

An allocation q is *Bayes–Nash (dominant-strategy) implementable* iff there is a payment rule p such that the direct mechanism (q, p) is interim (dominant-strategy) IC. Implementability is standardly characterised as follows:

Lemma 4. In the IPV model, an allocation q is Bayes–Nash implementable iff its interim allocation $(Q_i)_{i=1}^I$ has Q_i increasing for every agent i , and is dominant-strategy implementable iff $q_i(\cdot, \mathbf{t}_{-i})$ is increasing for every agent i and each profile $\mathbf{t}_{-i} \in [0, 1]^{I-1}$.

Proof. For the dominant-strategies part, fix an agent i and a profile \mathbf{t}_{-i} of types of the agents apart from i . In a direct revelation mechanism (q, p) , i faces the same reporting problem as if she were the only agent (as in the previous chapter), facing the single-agent mechanism $(q_i(\cdot, \mathbf{t}_{-i}), p_i(\cdot, \mathbf{t}_{-i}))$. Thus by the Spence–Mirrlees lemma (p. 11), she is willing to report truthfully iff $q_i(\cdot, \mathbf{t}_{-i})$ is increasing and $p_i(\cdot, \mathbf{t}_{-i})$ is given by the envelope formula.

The Bayes–Nash part is similar: agent i ’s reporting problem in a direct revelation mechanism (q, p) is exactly the same as if she were the only agent (as in the previous chapter), facing the allocation Q_i and the payment rule $P_i : [0, 1] \rightarrow \mathbf{R}$ defined by

$$P_i(t_i) := \mathbf{E}_{\mathbf{T}_{-i} \sim \mu_{-i}}[p_i(t_i, \mathbf{T}_{-i})]. \quad \blacksquare$$

Our question is thus under what circumstances an interim allocation $(Q_i)_{i=1}^I$ with Q_i increasing for every agent i is realised (in the sense of the previous section) by some allocation q that has $q_i(\cdot, \mathbf{t}_{-i})$ increasing for every agent i and profile \mathbf{t}_{-i} . The answer is in fact ‘always’!

Theorem 3 (Manelli and Vincent, 2010). In the IPV model, if $(Q_i)_{i=1}^I$ is realised by some allocation and has Q_i increasing for every i , then it is realised by an allocation q such that $q_i(\cdot, \mathbf{t}_{-i})$ is increasing for every i and every $\mathbf{t}_{-i} \in [0, 1]^{I-1}$.

Thus any allocation that can be implemented by a Bayes–Nash equilibrium of some auction can in fact be implemented in dominant strategies. Furthermore, by payoff/revenue equivalence, agents’ interim payoffs and revenue are unaffected. In short, we can get dominant-strategy IC for free.

We shall sketch a proof for the *symmetric IPV model*: the case in which all agents’ types are drawn from the same probability distribution.

Corollary 4. In the symmetric IPV model, if $Q : [0, 1] \rightarrow [0, 1]$ is increasing and is realised by some symmetric allocation, then it is realised by a symmetric allocation q such that $q_1(\cdot, \mathbf{t}_{-1})$ is increasing for every $\mathbf{t}_{-1} \in [0, 1]^{I-1}$.

Sketch proof of Corollary 4. By Border’s theorem, the Q s of interest are exactly the increasing maps $Q : [0, 1] \rightarrow [0, 1]$ that satisfy the Border inequalities; let’s write \mathcal{Q} for the set of all such Q s. Observe that \mathcal{Q} is convex; it is also (in fact) closed in a suitable sense.

Claim. The extreme points of \mathcal{Q} are 0 and the interim allocations induced by symmetric hierarchical allocations q such that $q_1(\cdot, \mathbf{t}_{-1})$ is increasing for every $\mathbf{t}_{-1} \in [0, 1]^{I-1}$.⁵

Loosely, interim allocations induced by increasing symmetric hierarchical allocations are extreme because they make some Border inequalities (in fact, a maximal set of Border inequalities) hold with equality (Lemma 3 above). Conversely, nothing else is extreme because only hierarchical allocations make a maximal set of Border inequalities bind. This is just intuition, but the claim can be proved using the types of arguments found in e.g. Kleiner, Moldovanu and Strack (2021).

It is intuitive, and in fact true, that any member of the convex set \mathcal{Q} can be written as an (infinite) convex combination of the extreme points of \mathcal{Q} . Results like this constitute a little field called Choquet theory; and Choquet’s

⁵Here we allow for hierarchical allocations with infinitely many sets R_k ; the technical details of this might require some head-scratching.

theorem says that we may for any $Q \in \mathcal{Q}$ find a probability measure ν defined on $\text{ext } \mathcal{Q}$ such that $Q = \int_{\text{ext } \mathcal{Q}} Q' \nu(dQ')$.

So take any $Q \in \mathcal{Q}$; by Choquet's theorem, $Q = \int_{\text{ext } \mathcal{Q}} Q' \nu(dQ')$ for some probability measure ν on the $\text{ext } \mathcal{Q}$. Each $Q' \in \text{ext } \mathcal{Q}$ is realised by a $q^{Q'}$ that is either 0 or a symmetric hierarchical allocation with $q_1^{Q'}(\cdot, t_{-1})$ increasing. Thus Q is realised by the allocation

$$q := \int_{\text{ext } \mathcal{Q}} q^{Q'} \nu(dQ'),$$

which is evidently symmetric and has $q_1(\cdot, t_{-1})$ increasing. \blacksquare

The literature. The theorem is due to Manelli and Vincent (2010); their proof is more involved.⁶ See Gershkov, Goeree, Kushnir, Moldovanu and Shi (2013) for a generalisation using a different technique, and Kleiner, Moldovanu and Strack (2021) for yet another generalisation, using yet another technique. Equivalence generally fails if agents' types are multi-dimensional (Jéhiel, Moldovanu & Stacchetti, 1999), as is the case when there are multiple goods (recall §1.7). Jeong and Pycia (2023) use reasoning similar to that of this section to show that every extreme point of the space of all symmetric IC mechanisms is induced by a (slightly generalised) first-price auction.

2.3.2 Optimal auctions

We henceforth consider the *smooth IPV model*, in which the CDF F_i of each agent i 's random type T_i possesses a density f_i , which is continuous and strictly positive.

We seek the optimal mechanism for a seller whose objective is expected revenue, who expects agents to play a Bayes–Nash equilibrium of whatever mechanism she adopts, and who recognises that each type of each agent has access to an outside option worth zero.⁷

Applying the reasoning of chapter 1 agent-by-agent (in a similar way to how we reasoned agent-by-agent in the proof of Lemma 4 above) yields that any Bayesian IC mechanism (q, p) must satisfy the envelope formula

$$P_i(t_i) = P_i(0) + t_i Q_i(t_i) - \int_0^{t_i} Q_i \quad \text{for every } t_i \in [0, 1] \text{ and } i,$$

⁶Rather than using Choquet's theorem, they rely on the weaker Krein–Milman theorem; this necessitates an additional limit argument.

⁷By Theorem 3, nothing would change if the seller feels confident only that weakly dominated strategies will not be played. We shall see this explicitly at the end: the optimal mechanism will be dominant-strategy IC.

that (interim) IR requires exactly that $P_i(0) \leq 0$ for each i , and thus that revenue-maximising mechanisms have $P_i(0) = 0$ for every i . This reduces the revenue-maximisation problem to choosing an allocation q to maximise

$$R(q) := \sum_{i=1}^I R_i(Q_i), \quad \text{where} \quad R_i(Q_i) := \mathbf{E} \left(T_i Q_i(T_i) - \int_0^{T_i} Q_i \right),$$

subject to the constraint that Q_i be increasing for each agent i .

To rewrite the second term in $R_i(Q_i)$, let $\phi_i(t_i) := \int_0^{t_i} Q_i$; then integrating by parts yields

$$\begin{aligned} \mathbf{E} \left(\int_0^{T_i} Q_i \right) &= \int_0^1 \phi_i f_i = [\phi_i F_i]_0^1 - \int_0^1 \phi_i' F_i = \phi_i(1) - \int_0^1 Q_i F_i \\ &= \int_0^1 Q_i (1 - F_i). \end{aligned}$$

Let $\mathbf{I} : [0, 1] \rightarrow [0, 1]$ denote the identity ($\mathbf{I}(t) := t$ for every $t \in [0, 1]$), and let $V_i := \mathbf{I} - \frac{1-F_i}{f_i}$; this is called agent i 's *virtual value*. We have

$$R_i(Q_i) = \int_0^1 \mathbf{I} Q_i f_i - \int_0^1 \frac{1-F_i}{f_i} Q_i f_i = \mathbf{E}[V_i(T_i) Q_i(T_i)].$$

The right-hand side is equal to

$$\mathbf{E} \left[V_i(T_i) \mathbf{E} \left(q_i(t_i, \mathbf{T}_{-i}) \mid T_i = t_i \right) \right] = \mathbf{E}[V_i(T_i) q_i(T_i, \mathbf{T}_{-i})]$$

by definition of Q_i and the law of iterated expectations. (Recall that agents' types $(T_i)_{i=1}^I$ are independent.) Thus expected revenue reads

$$R(q) = \mathbf{E} \left[\sum_{i=1}^I V_i(T_i) q_i(\mathbf{T}) \right].$$

The *regular smooth IPV model* is the case in which each agent i 's virtual valuation V_i is increasing. A sufficient condition is that the hazard rate $f_i/(1 - F_i)$ be increasing.

A *second-price auction with reserve price* $r \in [0, 1]$ is just like a second-price auction, except that there is also a fictional bidder who always bids r ; if the fictional bidder wins, then the seller keeps the good and no-one pays anything. It is well-known (and easily verified) that it is weakly dominant for

every type of every agent to bid her valuation; thus the induced allocation and payment rule are

$$\begin{aligned} & (q_i(t_1, \dots, t_I), p_i(t_1, \dots, t_I)) \\ &= \begin{cases} (1, \max\{r, \max_{j \neq i} t_j\}) & \text{if } t_i > \max\{r, \max_{j \neq i} t_j\} \\ (0, 0) & \text{otherwise.} \end{cases} \end{aligned}$$

(We haven't specified what happens in case of a tie, but that doesn't matter since ties occur with probability zero in the smooth IPV model.)

Theorem 4 (Myerson, 1981). In the symmetric regular smooth IPV model, there is a second-price auction with reserve price that is optimal.

Proof. We have already reduced the seller's problem to choosing an allocation q to maximise

$$R(q) = \mathbf{E} \left[\sum_{i=1}^I V_i(T_i) q_i(\mathbf{T}) \right]$$

subject to monotonicity of Q_i for every agent i and $\sum_{i=1}^I q_i \leq 1$. In the relaxed problem in which the monotonicity constraint is absent, it is obviously optimal to assign the good to whichever agent has the highest virtual valuation if any has positive virtual valuation, and to keep the good otherwise:

$$q_i(t_i, \mathbf{t}_{-i}) = \begin{cases} 1 & \text{if } V_i(t_i) > \max\{0, \max_{j \neq i} V_j(t_j)\} \\ 0 & \text{otherwise.} \end{cases}$$

And this allocation satisfies the monotonicity constraint by regularity: since V_i is increasing, $q_i(\cdot, \mathbf{t}_{-i})$ is increasing for any $\mathbf{t}_{-i} \in [0, 1]^{I-1}$, and so certainly Q_i is increasing.

Under symmetry, $V_i = V_j$ for all $i \neq j$, so $V_i(t_i) > V_j(t_j)$ holds iff $t_i > t_j$, and $V_i(t_i) > 0$ holds iff $t_i > r$, where r is some number in $[0, 1]$. Thus optimal mechanisms allocate the good to agent i iff her valuation exceeds r and all other agents' types t_j (modulo ties, which we may neglect). By inspection, this is the allocation implemented by the second-price auction with reserve price r . ■

Remark 5. Absent symmetry, the first part of the proof still delivers a characterisation of the optimal allocation, and the implementing payments may be backed out from the envelope formula. If we also drop regularity, then the proof must be modified, because the monotonicity constraint can no longer

be neglected without incident. Myerson (1981) developed a technique called *ironing* to deal with this case. Ironing has recently been extended and applied in many ways (e.g. Toikka, 2011; Condorelli, 2012, 2013; Kleiner, Moldovanu & Strack, 2021; Dworczak, Kominers & Akbarpour, 2021; Loertscher & Muir, 2022; Onuchic & Ray, 2023; Jewitt & Quigley, 2023).

The literature. Everything in this section is due to Myerson (1981). The ‘Myersonian approach’ to the regular case has proved fruitful also in richer models in which more than just a single good is up for sale, e.g. dynamic problems of selling goods over many periods (Pavan, Segal & Toikka, 2014), problems of selling information (Yang, 2022), and the problem of selling to a group (Haghpanah, Kuvalekar & Lipnowski, 2024). To understand what’s really going on in the ‘Myersonian approach’ (in particular, how to interpret the virtual valuations), I recommend Hartline (2016, section 3.3).

2.3.3 The Bulow–Klemperer theorem

We have shown that the choice of reserve price is important for revenue. Something else that obviously matters is *competition*: the number of bidders. (‘Bidder’ is just another term for ‘agent’.) The following delineates their relative importance:

Bulow–Klemperer theorem. In the symmetric regular smooth IPV model, for any $I \in \mathbf{N}$, the expected revenue of the revenue-maximising mechanism with I bidders is weakly lower than the expected revenue of the second-price auction with $I + 1$ bidders.

In short, competition matters far more for revenue than the details of the auction design: given the choice between setting an optimal reserve price and having just *one* extra bidder, a revenue-maximising seller should always opt for the latter. Insofar as the model assumptions are valid in a given practical situation (e.g. there is just one good, and values are private), this result provides rather sharp advice to sellers.

The proof relies on a lemma. Recall that a mechanism (q, p) may keep the good: it may be that $\sum_{i=1}^I q_i(t_i) < 1$ for some type profiles $(t_1, \dots, t_I) \in [0, 1]^I$. Indeed, the second-price auction with reserve price keeps the good whenever all agents’ valuations are lower than the reserve price.

Lemma 5. In the symmetric regular smooth IPV model, among mechanisms that never keep the good $((q, p)$ such that $\sum_{i=1}^I q_i = 1$), the second-price auction is optimal.

Proof. Return to the proof of Theorem 4. Neglecting the monotonicity constraint, it is clearly optimal to allocate the good to the agent with the highest virtual valuation, which by symmetry means the highest-valuation agent; this allocation satisfies the monotonicity constraint by regularity, and is (by inspection) implemented by the second-price auction. ■

Proof of the Bulow–Klemperer theorem. Fix an $I \in \mathbf{N}$. With I agents, by Theorem 4, the second-price auction with reserve price r is optimal, for some $r \in [0, 1]$; write R^* for its expected revenue.

Now suppose that there is one additional agent. Consider the same second-price auction with reserve price r and bidders $\{1, \dots, I\}$ (note that agent $I + 1$ is not allowed to bid), except that whenever all bids are below the reserve price r , the good is awarded to agent $I + 1$ instead of to nobody. This mechanism obviously raises expected revenue R^* , and never keeps the good. Thus by Lemma 5, the $(I + 1)$ -bidder second-price auction raises expected revenue at least R^* . ■

The literature. The theorem is from Bulow and Klemperer (1996). The proof given here is due to Kirkegaard (2006).

2.4 Allocation with costly verification

A principal can allocate an indivisible good to one of $I \geq 2$ agents. Agents like the good and are expected-utility maximisers, so an agent's payoff is simply the probability with which she expects to receive the good.

The principal's payoff from allocating to agent i is t_i , while her payoff from not assigning the good is zero. Agent i 's type is a random draw T_i from a CDF F_i on \mathbf{R}_+ whose density f_i is strictly positive on a compact interval $\mathcal{T}_i := [\underline{t}_i, \bar{t}_i]$ and zero off it. Agents' types are independent.

Each agent i knows her type t_i ; the principal and other agents do not. The principal can *check* agent i 's type at a cost $c_i > 0$. She cannot check more than one agent.

Remark 6. The analysis changes only a little if types can be negative, so that the principal sometimes strictly prefers to keep the good.

What mechanisms are optimal? To fix ideas, assume symmetric checking costs ($c_i = c$ for all i). Here's a simple mechanism: elicit reports of all agents' types, then check the type of the highest reporter, and award her the good provided she told the truth. (It is necessary to check her with probability one, since otherwise truthful reporting would not be IC.)

This mechanism can be improved upon. When all types report lower than the checking cost c , it isn't worthwhile to check anyone. The principal is therefore better-off instituting a non-negative threshold; if anyone reports above the threshold, then the highest-reporting type is checked and awarded the good provided her report was true, but if all reports are below the threshold, then the principal keeps the good and checks no-one.

This mechanism can also be improved upon. When all agents report below the threshold, the principal wastes the good by keeping it. She can improve by allocating it incentive-compatibly also in this case. Since she isn't checking, it is not incentive-compatible to allocate contingent on reported types in this scenario; but it *is* incentive-compatible simply to hand the good to a pre-designated 'favoured agent'. It is of course optimal to pick as favoured agent someone whose expected type is highest. This is called a *favoured-agent mechanism*.

When costs $(c_i)_{i=1}^I$ differ across agents, favoured-agent mechanisms are defined as follows. Agents report their 'net types' $t_i - c_i$, and there is a net-type threshold τ . If at least one agent's reported net type exceeds τ , then the agent with the highest-reported net type gets the good; otherwise, the favoured agent gets the good.

Ben-Porath, Dekel and Lipman (2014) show that such a favoured-agent mechanism is optimal, and uniquely so (all optimal mechanisms are essentially randomisations over favoured-agent mechanisms). Their proof is really long, despite the simplicity of the model and the result! Fortunately, it can be shortened substantially by using Border's theorem.

2.4.1 The principal's problem

The standard revelation principle does not apply since there is hard evidence. Nonetheless, the same kind of logic reveals that we may restrict attention to mechanisms of the following sort. First, agents are asked to report their types, and truthful reporting is a Bayes–Nash equilibrium. As a function of these reports, it is (perhaps randomly) determined who (if anyone) will be checked. Finally, the good is allocated as a function of the reports and (if there was one) the check and its outcome.

It is clearly without loss to consider only mechanisms in which, if agent i is checked and is found to have lied, then she certainly does not get the good. That's because if we take any mechanism in which truth-telling is a Bayes–Nash equilibrium and modify it to never give the good to a verified liar, then our new mechanism allocates the good just like the old one (since agents do not lie in equilibrium), and truth-telling remains a best response for every

agent since lying is punished (even) more harshly in the new mechanism than in the old.

We can restrict one step further: without loss, if agent i is checked and (as happens in equilibrium) is found to have reported truthfully, then she gets the good. For suppose we had a mechanism in which when i is checked and is found to have told the truth, the good sometimes goes to j . We could instead first (randomly) determine who will get the good conditional on i being checked and found to have told the truth, and if it's j , then we can check her instead; this disturbs neither i 's nor j 's willingness to report truthfully.

For a mechanism of this simple form, an agent i 's incentives depend on two things: for each profile $\mathbf{t} \in \mathcal{T}_1 \times \cdots \times \mathcal{T}_I$, (a) the probability $q_i(\mathbf{t})$ with which she gets the good, and (b) the probability $e_i(\mathbf{t})$ with which she is checked. Formally, a mechanism is a pair of maps (q, e) , each of which carries $\mathcal{T}_1 \times \cdots \times \mathcal{T}_I$ into $[0, 1]^I$, which jointly satisfy feasibility: $\sum_{i=1}^I q_i(\mathbf{t}) \leq 1$ for every \mathbf{t} , and $e \leq q$.

The principal wishes to maximise

$$\mathbf{E} \left(\sum_{i=1}^I [q_i(\mathbf{T})T_i - e_i(\mathbf{T})c_i] \right)$$

among those mechanisms (q, e) that are incentive-compatible, meaning that truthful reporting is a Bayes–Nash equilibrium. (As usual, IC can be expressed as some inequalities; we'll do that in §2.4.3 below).

2.4.2 Interim reformulation

Both the agents and principal care only about the interim allocation and checking, defined by

$$Q_i(t_i) := \mathbf{E}[q_i(t_i, \mathbf{T}_{-i})] \quad \text{and} \quad E_i(t_i) := \mathbf{E}[e_i(t_i, \mathbf{T}_{-i})].$$

But which families $(Q_i, E_i)_{i=1}^I$ are realised by some mechanism (q, e) ?

Lemma 6. $(Q_i, E_i)_{i=1}^I$ is realised by some mechanism exactly if $E_i \leq Q_i$ for every i and $(Q_i)_{i=1}^I$ satisfies the Border inequalities:

$$\sum_{i=1}^I \int_{T_i} Q_i f_i \leq 1 - \prod_{i=1}^I \left(1 - \int_{T_i} f_i \right)$$

for all (measurable) sets $T_1 \subseteq \mathcal{T}_1, \dots, T_I \subseteq \mathcal{T}_I$ of types.

Proof. If $(Q_i, E_i)_{i=1}^I$ is realised by some mechanism (q, e) , then $E_i \leq Q_i$ since $e \leq q$, and $(Q_i)_{i=1}^I$ satisfies the Border inequalities by Border's theorem. Conversely, if $(Q_i)_{i=1}^I$ satisfies the Border inequalities, then by Border's theorem, there is an allocation q that realises $(Q_i)_{i=1}^I$. If in addition $E_i \leq Q_i$ for each i , then the checking rule e defined by

$$e_i(t_i, \mathbf{t}_{-i}) = \begin{cases} q_i(t_i, \mathbf{t}_{-i}) \times \frac{E_i(t_i)}{Q_i(t_i)} & \text{if } Q_i(t_i) > 0 \\ 0 & \text{otherwise} \end{cases}$$

satisfies $e \leq q$ and $\mathbf{E}[e_i(t_i, \mathbf{T}_{-i})] = E_i(t_i)$ for every i and $t_i \in \mathcal{T}_i$. \blacksquare

The principal's problem is then to choose $(E_i, Q_i)_{i=1}^I$ to maximise

$$\sum_{i=1}^I \mathbf{E}[Q_i(T_i)T_i - E_i(T_i)c_i]$$

subject to incentive-compatibility, $E_i \leq Q_i$ for each i , and the Border inequalities.

2.4.3 Equivalence of interim and ex-post IC

Recall that an incentive-compatible mechanism (q, e) is one in which truthful reporting is a Bayes–Nash equilibrium. In other words, (q, e) is IC exactly if its interim allocation and checking $(Q_i, E_i)_{i=1}^I$ satisfy

$$Q_i(t_i) \geq Q_i(t'_i) - E_i(t'_i) \quad \text{for every } i \text{ and all } t_i, t'_i \in \mathcal{T}_i.$$

The right-hand side here is the payoff of a type $t_i \neq t'_i$ from claiming to have type t'_i : she gets allocated the good as if she really did have type t'_i (viz. with probability $Q_i(t'_i)$), *except* if she gets checked (which happens with probability $E_i(t'_i)$); in that case, she does not get the good. If you don't find that explanation of the right-hand side convincing, then have a look at the addendum to this chapter (p. 51), which contains a painstaking derivation of the right-hand side.

A stronger notion of IC would require that agents would prefer truthful reporting even if they knew the other agents' types:

$$q_i(t_i, \mathbf{t}_{-i}) \geq q_i(t'_i, \mathbf{t}_{-i}) - e_i(t'_i, \mathbf{t}_{-i}) \quad \text{for every } i \text{ and all } t_i, t'_i \in \mathcal{T}_i, \mathbf{t}_{-i} \in \mathcal{T}_{-i}.$$

This is called *ex-post IC*.

Lemma 7. (q, e) is ex-post IC iff

$$e_i(\mathbf{t}) \geq q_i(\mathbf{t}) - \inf_{t'_i \in \mathcal{T}_i} q_i(t'_i, \mathbf{t}_{-i}) \quad \text{for every } i \text{ and all } \mathbf{t} \in \mathcal{T}_1 \times \cdots \times \mathcal{T}_I,$$

and is interim IC iff its interim allocation and checking $(Q_i, E_i)_{i=1}^I$ satisfy

$$E_i(t_i) \geq Q_i(t_i) - \inf_{\mathcal{T}_i} Q_i \quad \text{for every } i \text{ and all } t_i \in \mathcal{T}_i.$$

Proof. Ex-post of (q, e) requires precisely that for every agent i and profile $\mathbf{t} = (t_i, \mathbf{t}_{-i}) \in \mathcal{T}_i \times \mathcal{T}_{-i}$,

$$q_i(t'_i, \mathbf{t}_{-i}) \geq q_i(\mathbf{t}) - e_i(\mathbf{t}) \quad \text{for every type } t'_i \in \mathcal{T}_i.$$

This is equivalent to requiring for every agent i and profile \mathbf{t} that

$$\inf_{t'_i \in \mathcal{T}_i} q_i(t'_i, \mathbf{t}_{-i}) \geq q_i(\mathbf{t}) - e_i(\mathbf{t}).$$

Similarly for interim IC. ■

Ex-post IC appears much stronger than interim IC. And yet:

Theorem 5. For any (interim) IC mechanism (q, e) , there is an ex-post IC mechanism that realises the same interim allocation and checking.

(This result can actually be further strengthened to obtain *dominant-strategy IC*, but we won't go into that here.)

Proof. Let (q, e) be interim IC, and write $(Q_i, E_i)_{i=1}^I$ for its interim allocation and checking. Wlog re-label each agent i 's types so that Q_i is increasing.⁸ Then by the Manelli–Vincent theorem (p. 40), there is an allocation q' realising $(Q_i)_{i=1}^I$ such that $q'_i(\cdot, \mathbf{t}_{-i})$ is increasing (in the new order on the types) for every i and \mathbf{t}_{-i} . Thus

$$\inf_{t'_i \in \mathcal{T}_i} \mathbf{E}[q'_i(t'_i, \mathbf{T}_{-i})] = \mathbf{E}\left[\inf_{t'_i \in \mathcal{T}_i} q'_i(t'_i, \mathbf{T}_{-i})\right].$$

Define

$$e'_i(t_i, \mathbf{t}_{-i}) := q'_i(t_i, \mathbf{t}_{-i}) - \inf_{t'_i \in \mathcal{T}_i} q'_i(t'_i, \mathbf{t}_{-i}).$$

Evidently $e' \leq q'$, so (q', e') is a mechanism. This mechanism satisfies the reformulated ex-post IC constraints in Lemma 7 with equality, so it is IC.

⁸More formally, consider the alternative type spaces $\mathcal{T}'_i := \{Q_i(t_i)\}_{t_i \in \mathcal{T}_i}$.

And its interim checking costs are lower than those of (q, e) : for any i and $t_i \in \mathcal{T}_i$, we have

$$\begin{aligned} \mathbf{E}[e'_i(t_i, \mathbf{T}_{-i})] &= \mathbf{E}[q'_i(t_i, \mathbf{t}_{-i})] - \inf_{t'_i \in \mathcal{T}_i} \mathbf{E}[q'_i(t'_i, \mathbf{t}_{-i})] \\ &= Q_i(t_i) - \inf_{\mathcal{T}_i} Q_i \leq E_i(t_i), \end{aligned}$$

where the last inequality holds by Lemma 7 since (q, e) is interim IC. By adding probability to the checking rule e' , we may obtain a (clearly still ex-post IC) mechanism (q', e') with the same interim checking as (q, e) . ■

2.4.4 Solving

Fixing Q_i , by Lemma 7, IC requires precisely that $E_i(t_i) \geq Q_i(t_i) - \inf_{\mathcal{T}_i} Q_i$ for every $t_i \in \mathcal{T}_i$, and it is obviously optimal to choose E_i to make this hold with equality. (This also ensures that $E_i \leq Q_i$ for every agent i .) Then remainder of the principal's problem is to choose $(Q_i)_{i=1}^I$ to maximise

$$\sum_{i=1}^I \mathbf{E} \left[Q_i(T_i)[T_i - c_i] + c_i \inf_{\mathcal{T}_i} Q_i \right],$$

subject to the Border inequalities.

This is not a linear programme, because Q_i also enters non-linearly. But we can convert it into a parametrised linear programme: given weights $\varphi_1, \dots, \varphi_I \in [0, 1]$ that sum to ≤ 1 , consider maximising

$$\sum_{i=1}^I \mathbf{E}[Q_i(T_i)[T_i - c_i] + c_i \varphi_i]$$

subject $Q_i \geq \varphi_i$ and the Border inequalities. It can be shown by an elementary argument (see Erlanson & Kleiner, 2019) that whatever the values of $(\varphi_i)_{i=1}^I$, any solution of this linear programme is a *threshold mechanism*. Thus optimal mechanisms must be threshold mechanisms.

What's a threshold mechanism? Call an interim allocation $(Q_i)_{i=1}^I$ τ -*threshold* iff

$$Q_i(t_i) = \begin{cases} \prod_{j \neq i} F_j(t_i - c_i + c_j) & \text{if } t_i - c_i > \tau \\ \varphi_i & \text{if } t_i - c_i \leq \tau. \end{cases}$$

This is implemented by an allocation q such that if anyone reports a net type $t_i - c_i$ above τ then (one of) the highest reporters gets the good, while if

no-one does then the good is allocated so that i gets the good with probability $\varphi_i / \prod_{j \neq i} F_j(\tau - c_i + c_j)$. The cheapest checking rule e that makes this ex-post IC (i.e. the one that makes the reformulated ex-post IC in Lemma 7 hold with equality) is the one that checks the highest-reporting agent in the first case, and checks no-one in the second.

In our parametrised linear programme, among threshold interim allocations, the principal obviously prefers a lower threshold. But the constraints limit how low the threshold can be set. The optimal mechanisms are those threshold mechanisms with the lowest feasible threshold τ^* .

To get to favoured-agent mechanisms, it remains to choose the values of $\varphi_1, \dots, \varphi_I$. It turns out that not all choices are feasible. Ben-Porath, Dekel and Lipman (2014) show that the set of all feasible choices is convex, however. Since the principal's objective is linear, that means that extreme points are optimal choices. Some of the extreme points are, unsurprisingly, those with $\varphi_i = 0$ for all but one 'favoured' agent i^* , whose φ_{i^*} is instead 'as large as possible'; these correspond to favoured-agent mechanisms.

The literature. The model and main result (the optimality of favoured-agent mechanisms) are due to Ben-Porath, Dekel and Lipman (2014). Erlanson and Kleiner (2019) used the interim perspective (§2.4.2 above) to establish the equivalence of ex-post and interim IC (Theorem 5 above) and to obtain a simpler proof of the optimality of threshold mechanisms.⁹ There's been some subsequent work looking at variations of the original model (e.g. Kattwinkel & Knoepfle, 2023; Erlanson & Kleiner, 2020, 2022).

Addendum: derivation of the IC constraints

In writing down the (interim) IC constraints in §2.4.3, we had to express the payoff of a type $t_i \neq t'_i$ from claiming to be t'_i , in a given mechanism (q, e) with interim allocation and checking $(Q_i, E_i)_{i=1}^I$. The expression is $Q_i(t'_i) - E_i(t'_i)$, and I gave a simple explanation of why.

But if you weren't satisfied with my explanation, here's another one.

⁹There are at least two other shortenings of the threshold-mechanisms argument, one by Bart Lipman (in an unpublished note), and another by Ricky Vohra (in a lecture on Border's theorem and its applications: youtu.be/DbzOqBoTmm0).

(Tedium alert!) Observe first that

$$\begin{aligned}
Q_i(t_i) &= \mathbf{P}(i \text{ gets the good} | i \text{ truthfully reports } t_i) \\
&= \mathbf{P}(i \text{ gets the good} | i \text{ truthfully reports } t_i \text{ and is checked}) \\
&\quad \times \mathbf{P}(i \text{ is checked} | i \text{ truthfully reports } t_i) \\
&\quad + \mathbf{P}(i \text{ gets the good} | i \text{ truthfully reports } t_i \text{ and is not checked}) \\
&\quad \times \mathbf{P}(i \text{ is not checked} | i \text{ truthfully reports } t_i) \\
&= 1 \times E_i(t_i) \\
&\quad + \mathbf{P}(i \text{ gets the good} | i \text{ reports } t_i \text{ and is not checked}) \\
&\quad \times [1 - E_i(t_i)],
\end{aligned}$$

which rearranges to

$$\begin{aligned}
\mathbf{P}(i \text{ gets the good} | i \text{ reports } t_i \text{ and is not checked}) \times [1 - E_i(t_i)] \\
= Q_i(t_i) - E_i(t_i). \quad (2.1)
\end{aligned}$$

Now, the expected payoff of a type $t_i \neq t'_i$ from claiming to have type t'_i is

$$\begin{aligned}
&\mathbf{P}(i \text{ gets the good} | i \text{ falsely reports } t_i) \\
&= \mathbf{P}(i \text{ gets the good} | i \text{ falsely reports } t_i \text{ and is checked}) \\
&\quad \times \mathbf{P}(i \text{ is checked} | i \text{ falsely reports } t_i) \\
&\quad + \mathbf{P}(i \text{ gets the good} | i \text{ falsely reports } t_i \text{ and is not checked}) \\
&\quad \times \mathbf{P}(i \text{ is not checked} | i \text{ falsely reports } t_i) \\
&= 0 \times \mathbf{P}(i \text{ is checked} | i \text{ reports } t_i) \\
&\quad + \mathbf{P}(i \text{ gets the good} | i \text{ reports } t_i \text{ and is not checked}) \\
&\quad \times \mathbf{P}(i \text{ is not checked} | i \text{ reports } t_i) \\
&= \mathbf{P}(i \text{ gets the good} | i \text{ reports } t_i \text{ and is not checked}) \times [1 - E_i(t_i)] \\
&= Q_i(t_i) - E_i(t_i),
\end{aligned}$$

where the last equality holds by (2.1).

2.5 Allocation with evidence acquisition

Consider the same setting as in the previous section, except that agents do not know their own types t_i : instead, the principal and agents are symmetrically uninformed about the types (t_1, \dots, t_I) . By paying a cost of $c_i \geq 0$, an agent

i can learn her type. If she does, then she can verifiably disclose her type to the principal (that is, she can prove that her type is such-and-such).

A revelation argument shows that we may restrict attention to mechanisms in which

- Each agent has at most one information set, at which she is asked to acquire and deliver evidence.
- If she ever gets the good after delivering evidence, then she never gets the good *without* delivering evidence.
- Each agent is willing to acquire evidence when asked.
- (An agent asked for evidence who does not deliver it is never given the good; thus an agent who is asked for evidence is willing not to just to acquire it, but also to deliver it.)

An ex-post allocation is (q, e) , where $q : \mathcal{T}_1 \times \cdots \times \mathcal{T}_I \rightarrow [0, 1]^I$ satisfies $\sum_{i=1}^I q_i \leq 1$, and $e = (e_1, \dots, e_I)$ where e_i maps $\mathcal{T}_{-i} \rightarrow [0, 1]$.¹⁰ Which allocations (q, e) are feasible? It isn't obvious; for example, with $I = 2$ agents, we cannot ask each for evidence iff the other's type is $< 1/2$.¹¹ This is a model in which it is difficult to characterise feasibility even of *ex-post* allocations.

It turns out, however, to be *more* tractable to characterise feasible *interim* allocations! The interim allocation induced by (q, e) is $(Q_i, E_i)_{i=1}^I$, where $Q_i(t_i) = \mathbf{E}[q_i(t_i, \mathbf{T}_{-i})]$ and $E_i = \mathbf{E}[e_i(\mathbf{T}_{-i})]$.

Clearly a necessary condition for (q, e) to be feasible is for q to be a feasible allocation by itself. By Border's theorem, this is equivalent to $(Q_i)_{i=1}^I$ satisfying the Border inequalities.

Another necessary condition is that for each i , either Q_i is constant, or else $Q_i \leq E_i$. This is because if i is never asked for evidence, then her allocation cannot depend on her type, and so Q_i must be constant. And if i is sometimes asked for evidence, then she never gets the good without giving evidence, so $Q_i(t_i) \leq E_i$ for every type $t_i \in \mathcal{T}_i$.

Remarkably, these conditions are also *sufficient*:

Theorem 6. For an interim allocation $(Q_i, E_i)_{i=1}^I$, the following are equivalent:

¹⁰In this section, I'm using the term 'allocation' to encompass not only how the good is distributed, but also who acquires evidence.

¹¹If we are to ask agent 2 for evidence only contingent on the realised value of agent 1's type, then we must first ask agent 1 for evidence no matter what. But then 1 is not being asked for evidence contingent on the realised value of 2's type!

- (1) $(Q_i)_{i=1}^I$ satisfies the Border inequalities, and $Q_i(t_i) \leq E_i$ for every $t_i \in \mathcal{T}_i$ and i .
- (2) There is a game form and a strategy profile in it that realises $(Q_i, E_i)_{i=1}^I$ as its interim allocation.

The proof is based on Border's theorem and hierarchical allocations. With this characterisation of feasibility, one may go on to characterise optimal mechanisms.

The literature. The model and results are from work in progress by Elchanan Ben-Porath, Eddie Dekel and Bart Lipman titled 'Sequential mechanisms for evidence acquisition'. You can find a video recording of Eddie presenting the paper at youtu.be/B70fYgKc760, and his slides are also online.¹²

2.6 Allocation with extra information

Return to the baseline assumption of this chapter that agents are privately informed about their types (and that they always want the good). In section 2.4, the principal had the power to perfectly monitor an agent's type, at a cost. In this section, we instead consider free but imperfect monitoring.

Suppose, in particular, that after agents make their reports, the principal will observe some signals that are correlated with their types, and may condition her allocation decision on these. A revelation principle lets us focus on mechanisms in which the good is allocated (possibly randomly) contingent on agents' self-reported types and the realised signals, and in which agents are willing to report truthfully.

Here's a simple model. Each agent has an intrinsic 'quality' $\theta_i \in \{0, 1\}$. We assume that the principal must select one of the agents, i.e. she cannot keep the good. (That is a good assumption in some applications, and not in others.) The principal earns a higher payoff from allocating to a high-quality agent than to a low-quality agent. Each agent observes a private signal $t_i \in \{0, 1\}$ of her quality, and the principal also observes a signal $s_i \in \{0, 1\}$ of each agent's quality.

We assume that the agents' random qualities $(\Theta_i)_{i=1}^I$ are independent, that the random signals $(T_i, S_i)_{i=1}^I$ are independent conditional on $(\Theta_i)_{i=1}^I$,

¹²https://drive.google.com/file/d/1n2lo0NP5vIk_P1TVIRbUi4WtNvYNk8xk

and that for each i and profile $\boldsymbol{\theta}_{-i} \in \{0, 1\}^{I-1}$ we have

$$\mathbf{P}(T_i = 1 | \Theta_i = \theta_i, \boldsymbol{\Theta}_{-i} = \boldsymbol{\theta}_{-i}) = \mathbf{P}(T_i = 1 | \Theta_i = \theta_i) = \begin{cases} 1 - p & \text{for } \theta_i = 0 \\ p & \text{for } \theta_i = 1. \end{cases}$$

for some precision $p \in (1/2, 1)$ and

$$\mathbf{P}(S_i = 1 | \Theta_i = \theta_i, \boldsymbol{\Theta}_{-i} = \boldsymbol{\theta}_{-i}) = \mathbf{P}(S_i = 1 | \Theta_i = \theta_i) = \begin{cases} 1 - q & \text{for } \theta_i = 0 \\ q & \text{for } \theta_i = 1 \end{cases}$$

for some lower precision $q \in (1/2, p]$.

The literature. Bloch, Dutta and Dziubiński (2023) study this model, and obtain a ‘lexicographic’ characterisation of optimal mechanisms. The broader insight here, first articulated by Kattwinkel (2019), is that an agent can be screened by conditioning allocation decisions on an exogenous signal that is correlated with her type. This is analogous to (but less powerful than) screening using contingent monetary payments (recall §1.8).

Chapter 3

Allocating over time, without money

This chapter concerns repeated allocation problems. In each period, the principal has one good, which she can either allocate to the agent or keep, and the agent has a privately-known valuation t for the good. The principal wishes to allocate only to high types of the agent. Every type of the agent wants the good, but high types want it more.

Here's an example: the principal is the 'dean' of a university, and the agent is a department head. In each year, the department has one (best) job candidate, who may be either decent or excellent. The dean only wants to hire excellent researchers. The department (which does not internalise the cost to the university of hiring) always wants to hire, but is more keen when its candidate is excellent.

If the agent gets the good, then the agent and principal earn payoffs of t and $t - c$, respectively; otherwise they both get zero.¹ Both parties are expected-utility-maximisers who discount their future payoffs with discount factor $\delta \in (0, 1)$. Monetary transfers are unavailable.

Let's simplify by having only two types $t \in \{\ell, h\}$. We assume that $0 < \ell < c < h$, so that the principal wishes to allocate only to high types, while all types of the agent want the good (though high types want it more).

Time will be indexed by $n \in \{0, 1, \dots\}$. The agent's random type process $(T_n)_{n=0}^\infty$ is a stationary Markov chain; that just means that there are two

¹One interpretation is that the principal shares the agent's preferences, except that she also cares about the cost c of producing/allocating the good.

fixed ‘transition probabilities’ $u, d \in (0, 1)$ (‘up’ and ‘down’) such that

$$\begin{aligned} \mathbf{P}(T_{n+1} = h | T_n = t_n, \dots, T_0 = t_0) &= \mathbf{P}(T_{n+1} = h | T_n = t_n) \\ &= \begin{cases} u & \text{if } t_n = \ell \\ 1 - d & \text{if } t_n = h. \end{cases} \end{aligned}$$

The initial type T_0 is drawn from the stationary distribution of the Markov chain: this is the distribution assigning probability $p := u/(u + d)$ to $t = h$. We write μ for its mean: $\mu = (1 - p)\ell + ph$.

An important special case is when types are IID over time: $u = 1 - d = p$. More generally, $u \leq 1 - d$ means that the agent’s type is persistent (non-negatively auto-correlated). (*In §3.2, we shall assume persistence; in §3.3 we shall assume IID.*)

In principle, μ can be on either side of c . We shall focus on the more interesting case in which $\mu < c$, so that the principal wishes to allocate only if she manages to screen the agent.²

We shall also assume that patience δ is ‘high enough’ to make screening possible: $\delta > \ell/\mu$.³ This condition is easily interpreted in the IID case ($u = 1 - d = p$): it ensures that a low type is willing to forego consumption today if promised that she’ll be allowed to consume tomorrow no matter what ($\ell < \delta\mu$).

The analysis differs depending on when the agent learns her type. In §3.1, we shall assume that the agent learns the realisation of her entire type sequence $(T_n)_{n=0}^\infty$ at the outset (period $n = 0$). In §3.2–§3.3, we shall instead assume that the agent learns the realisation of T_n only at the dawn of period n , for each $n \in \{0, 1, \dots\}$.

3.1 Quota mechanisms

Assume that the agent learns the realisation of her entire type sequence $(T_n)_{n=0}^\infty$ at the outset.⁴

Here’s a mechanism: the agent freely chooses the probability $q_n \in [0, 1]$ with which she will consume the good in each period n , subject to the lifetime budget constraint

$$(1 - \delta) \sum_{n=0}^{\infty} \delta^n q_n \leq K,$$

²However, §3.1 does not require this assumption.

³This assumption is also unnecessary in §3.1.

⁴My own intuition suggests that this assumption shouldn’t matter *too* much, but I may be wrong. There is no definitive answer in the literature, to my knowledge.

where $K \in [0, 1]$ is a constant. We could call this a *discounted quota*: it caps the discounted total quantity that the agent may consume.

We could just as well replace this lifetime budget constraint with a dynamic budget constraint: the agent starts out with $M_0 := \delta(1 - \delta)^{-1}K$ ‘tokens’, which may be spent to obtain the good and which accrue interest at rate $\delta^{-1} - 1$ if saved:

$$q_n + M_{n+1} \leq \delta^{-1}M_n \quad \text{in each period } n.$$

We impose the no-Ponzi-scheme condition

$$\lim_{n \rightarrow \infty} \delta^n M_n = 0.$$

Then we have

$$\begin{aligned} M_0 &\geq \delta M_1 + \delta q_0 \\ &\geq \delta(\delta M_2 + \delta q_1) + \delta q_0 = \delta^2 M_2 + \delta(q_0 + \delta q_1) \\ &\geq \delta^2(\delta M_3 + \delta q_2) + \delta(q_0 + \delta q_1) = \delta^3 M_3 + \delta(q_0 + \delta q_1 + \delta^2 q_2) \\ &\vdots \\ &\geq \lim_{n \rightarrow \infty} \delta^n M_n + \delta \sum_{n=0}^{\infty} \delta^n q_n = \delta \sum_{n=0}^{\infty} \delta^n q_n, \end{aligned}$$

or

$$(1 - \delta) \sum_{n=0}^{\infty} \delta^n q_n \leq K.$$

The principal’s first-best allocation is the one in which the agent consumes when and only when her type is high. The first-best discounted total quantity is therefore

$$Q_\delta := (1 - \delta) \sum_{n=0}^{\infty} \delta^n \mathbf{1}_{\{h\}}(T_n).$$

Its expected value is

$$\mathbf{E}(Q_\delta) = (1 - \delta) \sum_{n=0}^{\infty} \delta^n \mathbf{E}(\mathbf{1}_{\{h\}}(T_n)) = (1 - \delta) \sum_{n=0}^{\infty} \delta^n p = p.$$

Given this, a natural choice of cap K is $K = p$.⁵

⁵I am *not* claiming that among discounted-quota mechanisms, the one with $K = p$ gives the principal the highest payoff. That’s not true in general.

The incentive properties of discounted quotas are simple: the mechanism leaves the agent no scope to obtain a larger discounted total quantity. The only question is thus in *which* periods the agent consumes, and this is a dimension along which incentives are aligned: both parties prefer for the agent to consume when her type is high.

The mechanism isn't perfect, however, since the principal's first-best discounted total quantity

$$Q_\delta = (1 - \delta) \sum_{n=0}^{\infty} \delta^n \mathbf{1}_{\{h\}}(T_n)$$

is random, not always equal to its expectation p . If the agent's type is high for several period in a row early on, then the first-best discounted total quantity is greater than p , but the agent lacks the tokens needed to consume more than p . Conversely, if her type is initially low for several periods, then she accrues enough saved tokens to be able to purchase the good in some periods in which her type is low.

This problematic variability of Q_δ is less severe the greater is the discount factor δ . To see why, assume for simplicity that the agent's type is IID over time ($u = 1 - d = p$), and calculate

$$\begin{aligned} \text{Var}(Q_\delta) &= (1 - \delta)^2 \sum_{n=0}^{\infty} \delta^{2n} \text{Var}(\mathbf{1}_{\{h\}}(T_n)) \\ &= (1 - \delta)^2 \sum_{n=0}^{\infty} \delta^{2n} p(1 - p) \\ &= \frac{(1 - \delta)^2}{1 - \delta^2} p(1 - p) = \frac{(1 - \delta)^2}{(1 - \delta)(1 + \delta)} p(1 - p) = \frac{1 - \delta}{1 + \delta} p(1 - p). \end{aligned}$$

By inspection, $\text{Var}(Q_\delta)$ vanishes as δ approaches 1. This suggests that we may expect Q_δ to be close to p with high probability if δ is large. That is indeed the case: by invoking Chebyshev's inequality, we obtain

$$\mathbf{P}(|Q_\delta - p| \geq \varepsilon) \leq \frac{\text{Var}(Q_\delta)}{\varepsilon^2} \quad \text{for any } \varepsilon > 0,$$

so that

$$\lim_{\delta \uparrow 1} \mathbf{P}(|Q_\delta - p| \geq \varepsilon) = 0 \quad \text{for any } \varepsilon > 0.$$

(In other words, Q_δ converges in probability to p as $\delta \uparrow 1$. Such a result is called a *weak law of large numbers*.)

The discounted-quota mechanism with cap $K = p$ therefore provides close to the first-best discounted total quantity with high probability when δ is high. This suggests (but does not prove!) that if δ is large, then this discounted-quota mechanism gives the principal a *payoff* close to her first-best payoff (which is $(h - c)Q_\delta$, by the way—why?). In other words, this mechanism is *approximately optimal* (for large δ). Formally:

Proposition 2. Assume that the agent's type is IID over time ($u = 1 - d = p$) and that the agent learns the realisation of her entire type sequence $(T_n)_{n=0}^\infty$ at the outset. Let U_δ^K denote the principal's (ex-post, random) payoff under the discounted-quota mechanism with cap $K \in [0, 1]$ when the discount factor is $\delta \in (0, 1)$. Then $(h - c)Q_\delta - U_\delta^K \rightarrow 0$ a.s. as $\delta \uparrow 1$.⁶

The proof relies on a simple *strong law of large numbers*. The IID assumption can be relaxed by using a fancier strong law of large numbers.

Proof. Fix a cap $K \in [0, 1]$ and a discount factor $\delta \in (0, 1)$. Clearly the agent always exhausts the budget. If $Q_\delta \geq K$, then she consumes only when $T_n = h$, so her payoff is hK . If $Q_\delta < K$, then the agent consumes whenever $T_n = h$, and spends the rest of the budget when $T_n = \ell$; thus her payoff is $hQ_\delta + \ell(K - Q_\delta)$. The principal's payoff is the agent's payoff minus cK , so

$$\begin{aligned} (h - c)Q_\delta - U_\delta^K &= (h - c)Q_\delta - \left[h \min\{K, Q_\delta\} + \ell \max\{K - Q_\delta, 0\} - cK \right] \\ &= (h - c)Q_\delta - \left[-h \max\{-K, -Q_\delta\} + \ell \max\{K - Q_\delta, 0\} - cK \right] \\ &= (h - c)Q_\delta - \left[-(h - \ell) \max\{K - Q_\delta, 0\} + (h - c)K \right] \\ &= -(h - c)(K - Q_\delta) + (h - \ell) \max\{K - Q_\delta, 0\}. \end{aligned}$$

For $K = p$, as $\delta \uparrow 1$, the right-hand side vanishes a.s. since $Q_\delta \rightarrow p$ a.s. by a strong law of large numbers for discounted sums (see Lai, 1974). ■

Remark 7. Zooming out, the model is one in which several decisions must be taken; we may re-interpret these as distinct, simultaneously-chosen actions. The principal and agent are still assumed to have payoffs additive across decisions, but the weights assigned to different decisions need not have the geometric form $1, \delta, \delta^2, \dots$. As above, the result is that ‘linking’ many different decisions via a weighted quota can yield good outcome.

⁶Hence by the bounded convergence theorem, the *expected* payoff $\mathbf{E}(U_\delta^K)$ converges as $\delta \uparrow 1$ to the first-best expected payoff $(h - c)p$.

The literature. Jackson and Sonnenschein (2007) showed that the principal’s first-best can be approximated by ‘linking’ the allocations across different periods via something like a discounted quota. (The theorem in this paper is true as stated, but the proof is wrong: see Ball, Jackson and Kattwinkel (2022) for the correction.) Ball and Kattwinkel (2024) go further, characterising the full set of allocations that are implementable via quota mechanisms,⁷ and proving a number of robustness results. (I highly recommend this paper.) Frankel (2016b) is a clear-eyed and general formulation of the quotas idea, with good references to related work. Frankel (2016a) provides a similarly insightful and general treatment of the idea in Remark 7. (None of these papers considers precisely the model above, and the mechanisms in Jackson–Sonnenschein and Ball–Kattwinkel are not quite discounted-quota mechanisms in the sense of this chapter.)

3.2 Optimal mechanisms with commitment

In this section, we’ll assume that the agent’s valuation is persistent ($u \leq 1 - d$), and that the agent learns her type in real time (she observes the realisation of T_n only at the dawn of period n).

Discounted quota mechanisms ensure IC precisely by capping the discounted total quantity. This inflexibility causes inefficiency, since the first-best discounted total quantity Q_δ is random. Is it possible to improve by supplying a greater discounted total quantity when the agent’s type turns out to be high in several early periods, while preserving IC?

Consider the case of two periods, $n \in \{0, 1\}$, with IID types ($u = 1 - d = p$), and let’s neglect discounting ($\delta = 1$). In the mechanism with cap $K = 1$, the agent consumes in period $n = 0$ if her type is high ($T_0 = h$), and otherwise waits and consumes in period $n = 1$. (Why?) The principal’s payoff is therefore

$$ph + (1 - p)\mu - c.$$

(We’re interested in parameters for which this is the best quota mechanism.⁸) Phrased differently, the $n = 0$ allocation is efficient, while at $n = 1$ an agent who was high at $n = 0$ gets nothing and an agent who was low at $n = 0$

⁷The result: all and only those that could be implemented using monetary transfers! This result refines an earlier characterisation due to Matsushima, Miyazaki and Yagi (2010) and Ishii (2016), which concerns *virtual* implementability. (There is a mistake in the former paper; see Ball and Kattwinkel (2023) for the correction.)

⁸By $\mu < c$, the principal strictly prefers $K = 0$ to $K = 2$. She strictly prefers $K = 1$ to $K = 0$ iff $ph + (1 - p)\mu > c$.

gets the good for sure. The principal doesn't like handing out the good in period 1 without screening (since $\mu < c$). So let's lower the probability e with which this occurs as much as possible, subject to preserving the period-0 low type's incentive to forego consumption; this yields $e = \ell/\mu$. This mechanism does not cap the total quantity supplied; rather, it provides a smaller total quantity when the agent's period-0 type is low.

To extend this idea to the full infinite-horizon model with discounting, we introduce *entitlement mechanisms*: the agent has an *entitlement* to claim the good for a certain number of consecutive periods 'no questions asked'. Whenever she foregoes the opportunity to obtain the good, her entitlement is augmented.

A little more fully, the agent begins each period with some entitlement $E \in [0, \infty]$. If $E > 0$, then she may claim the good, in which case her entitlement drops to $E - 1$. (If $E \in (0, 1)$, then she gets the good with probability E , and her entitlement drops to zero.) If she does not claim the good, then her entitlement rises by some history-dependent amount, possibly to ∞ .⁹

By appropriately choosing the amount by which the agent's entitlement is augmented when she does not claim the good, it is possible to ensure that as long as $E \in (0, \infty)$, the agent is (only just) willing to forego consumption when her type is low. (This clearly implies that she will claim the good when her type is high.) In other words, allocation is exactly efficient until E hits 0 or ∞ . (So inefficiencies are back-loaded.)

The agent's entitlement will eventually hit either 0 or ∞ .¹⁰ We might call the former 'termination' and the latter 'tenure'. The discounted total quantity supplied is evidently not deterministic (by contrast with discounted-quota mechanisms).

The literature. Guo and Hörner (2020) study this problem, and prove that entitlement mechanisms are optimal.¹¹ You'll learn a lot by reading it. The technique they use to solve for an optimal mechanism is that of Spear and Srivastava (1987), which is well-worth learning if you're interested in dynamic mechanism design with full commitment and no transfers. It offers a general recursive way of formulating dynamic contracting problems (with the agent's promised utility as the state variable). (It's a standard enough topic that you can find lecture notes on it; I don't know of a textbook treatment,

⁹In a period in which the entitlement rises to ∞ , the agent may in fact also receive the good with positive probability.

¹⁰In finite time, almost surely; this follows from the Borel–Cantelli lemma.

¹¹They prove this under an additional 'enough-patience' assumption: $\delta > 1/2$.

however. It is most commonly applied to dynamic models of *moral hazard*, rather than adverse selection.)

3.3 Optimal mechanisms without commitment

In this section, we'll assume that the agent's valuation is IID over time ($u = 1 - d = p$), and that the agent learns her type in real time (she observes the realisation of T_n only at the dawn of period n). We'll also add an extra 'enough patience' assumption:

$$(1 - \delta)(c - \mu) \leq \delta \left(1 - \frac{c}{\ell}\right)(\mu - \ell).$$

Suppose that the principal lacks the ability to commit over time. She can take certain actions (or 'commitments', if you prefer) within a given period, but cannot bind her future self. What can she achieve?

This is analytically a very different sort of question. The reason is that the principal now also has IC constraints: in each period, the principal must (given her belief) find it optimal to take the action specified by the mechanism.

A different way of thinking about limited commitment is that absent commitment, the agent and principal simply play a game: in particular, a repeated game of incomplete information, with the private information on one side only. Each equilibrium of this game induces some allocation,¹² and optimal incentive-provision from the principal's perspective amounts to playing a principal-preferred equilibrium.

It remains to specify what the principal can commit to (within a period). We'll assume that she may commit to a within-period delegation mechanism: she either offers the good to the agent or not, and if she offers it then the agent may accept or refuse. This gives the principal the ability to commit not to allocate (within a given period), which is potentially valuable.

The first thing to check is whether relaxing commitment matters: if the entitlement mechanism that was optimal under full commitment is an equilibrium outcome, then we'd be done! But it's not an equilibrium outcome. To see why, consider a history at which the agent's entitlement E has hit ∞ . (This occurs with positive probability in finite time.) The agent is now permitted to consume the good in every subsequent period, despite the fact that the principal prefers *not* to allocate absent screening ($\mu < c$). Clearly the principal strictly prefers to deviate to never allocate again.

¹²Note that equilibria can differ off the path of play, but still induce the same allocation.

So what are equilibrium outcomes? And which is the principal's favourite?

The literature. Lipnowski and Ramos (2020) characterise principal-optimal equilibria (and more generally, Pareto-optimal equilibria).¹³ Their analysis relies on the 'APS' technique (Abreu, Pearce & Stacchetti, 1990), which provides a recursive formulation of repeated games with imperfect monitoring. This is a standard (indeed, essential!) tool for analysing repeated interactions without commitment. The original APS paper is rather beautiful, but you may prefer to learn it from the textbook treatment of Mailath and Samuelson (2006), or from whatever lecture notes you can find around the internet. Another nice paper on dynamic allocation without commitment is de Clippel, Eliaz, Fershtman and Rozen (2021), who study a model with *two* agents.

¹³In their model, the agent learns T_n only if and when she is offered the good in period n . This does not make a difference, I believe. For suppose the principal-preferred equilibrium were to feature the agent conditioning her behaviour in period $n + 1$ on what T_n if she was not offered the good in period n . Since T_n has no statistical effect on T_{n+1} (that's our IID assumption), this just amounts to the agent randomising in period $n + 1$ (when the principal did not offer the good in period n); so we can replicate this behaviour by having the agent mix (i.e. she can use her own set of dice to randomise, instead of conditioning her behaviour on the 'natural dice' T_n).

Chapter 4

Incentivising proposal

There is an agent and a principal. A number of *allocations* are potentially available (these could also be called e.g. ‘projects’). We identify an allocation with its payoff consequences, writing it as (u, v) , where u is the agent’s payoff and v is the principal’s.

There is a *status quo* allocation that is always available; we normalise its payoffs to $(0, 0)$. It is ex-ante uncertain which other allocations are available: the available allocations are a random finite set $A \subseteq \mathbf{R}^2$. The agent observes which allocations are available; the principal does not.

The agent can *propose* a finite set $P \subseteq \mathbf{R}^2$ of allocations. The principal ultimately decides which allocation is implemented. The two key assumptions (which distinguish this model from vanilla mechanism design) are that

- (a) the agent can propose only available allocations (P must be a subset of A), and that
- (b) the principal must implement one of the proposed allocations, or else the status quo (she chooses from $P \cup \{(0, 0)\}$).

Assumption (a) says that the agent’s messages (proposals) to the principal constitute hard evidence, not cheap talk: proposing an allocation (u, v) *proves* the availability of this allocation. Property (b) grants the agent the power to constrain the principal: only with the agent’s permission/co-operation can the principal implement an allocation other than the status quo.

The principal can always guarantee herself a payoff of zero by selecting the status quo, and given (b), the agent can guarantee herself a payoff of zero by permitting only the status quo (i.e. proposing $P = \emptyset$). We need thus only consider allocations (u, v) belonging to \mathbf{R}_+^2 , and we’ll assume without loss that the availability set A is a random subset of \mathbf{R}_+^2 .

For the rest of this chapter, we make the further assumption that the agent can propose at most one allocation (besides the status quo); that is, P must be either null or a singleton. This is a reasonable assumption in some applications, and not in others.

The last assumption, together with the fact that proposals are hard evidence, turns out to imply that there is no very sharp revelation principle available.¹ We shall therefore rely on a *taxation principle* instead.

4.1 Deterministic mechanisms

A *delegation mechanism* consists of a *delegation set* $D \subseteq \mathbf{R}_+^2$ with $D \ni (0, 0)$ from which the agent freely chooses.² (That's it!) This is clearly equivalent to a *deterministic approval mechanism*, in which the agent proposes an allocation (or proposes nothing) to the principal, and the principal approves all and only allocations in D .

A (*deterministic*) *choice rule* is a map a that carries finite subsets of \mathbf{R}_+^2 into \mathbf{R}_+^2 in such a way that $a(A) \in A \cup \{(0, 0)\}$ for every finite $A \subseteq \mathbf{R}_+^2$.

Taxation principle. If a deterministic choice rule a is induced by some mechanism, then it is induced by the delegation mechanism with delegation set

$$D = \{a(A) : A \subseteq \mathbf{R}_+^2 \text{ finite}\}.$$

Proof. Fix a mechanism that induces some choice rule a . Fix a type $A \subseteq \mathbf{R}_+^2$ of the agent. Clearly she can mimic all and only types $B \subsetneq A$ (since she can propose whatever they can), and thus she is able to reach all allocations in $(D \cap A) \cup \{(0, 0)\}$; and she prefers $a(A)$. In the delegation mechanism with

¹Whatever we assume about how many allocations the agent can propose, we may invoke Myerson's (1982) general revelation principle: this tells us that we may limit our attention to 'IC direct mechanisms', where 'direct mechanism' means that the agent is asked for a cheap-talk report about which allocations are available and then told (as a function of her report) what allocation to propose, and where 'IC' means that the agent is willing to be truthful *and obedient* (i.e. to propose the allocation that she is told to propose). But this revelation principle is not very sharp, because it places no restrictions on what proposals a mechanism will instruct the agent to make.

As Guo and Shmaya (2023) point out, we get a sharper revelation principle if we assume instead that the agent can propose as many allocations as she likes: in that case, we may further limit our attention to those IC direct mechanisms that instruct the agent to propose *all* of the allocations that she claimed (in her report) to be available. This is an instance of Bull and Watson's (2007) revelation principle for settings with hard evidence.

²In other contexts, the term 'menu' is often used for a delegation set.

delegation set D , she can reach all *and only* allocations in $(D \cap A) \cup \{(0, 0)\}$; so she can be relied upon to select $a(A)$. ■

It remains to characterise optimal delegation sets D . For any set $S \subseteq \mathbf{R}_+^2$, we'll call the set

$$\{(u, v) \in \mathbf{R}_+^2 : (u, v') \in S \text{ for some } v' \leq v\}$$

the *upward closure* of S . A set that coincides with its upward closure is called *upward closed*.

Observation 2. If D is an optimal delegation set, then so is its upward closure. (It is therefore without loss to restrict attention to upward closed delegation sets.)

Proof. Exercise! ■

To any upward closed delegation set D , we associate a function $d : [0, \infty) \rightarrow [0, \infty]$ defined by

$$d(u) := \begin{cases} \inf\{v \in \mathbf{R}_+ : (u, v) \in D\} & \text{if } \{v \in \mathbf{R}_+ : (u, v) \in D\} \neq \emptyset \\ \infty & \text{otherwise.} \end{cases}$$

d is the lower boundary of the delegation set D , and fully describes D .³ Our remaining task is to characterise the optimal choice d of boundary.

To understand the trade-off involved, consider two allocations (u, v) and (u', v') , the former preferred by the principal ($v > v' > 0$) and the latter by the agent ($u < u'$). If the principal permits both, then the agent will choose the tempting allocation (u', v') over the 'good' allocation (u, v) when both are available. The principal can prevent this by banning the tempting allocation (u', v') . But this has a cost when the tempting allocation (u', v') is the only one that's available: the status quo is then implemented, rather than the Pareto-superior (and available) allocation (u', v') .

The literature. The model is due to Armstrong and Vickers (2010). The perspective here (which is quite different from that of the original paper) is one that I learned from Yingni Guo (see Guo & Shmaya, 2023).

³Okay, it really only determines D up to (topological) closure. In particular, given d , we may recover the *closure* $\text{cl } D$ of D as $\text{cl } D = \{(u, v) : d(u) \leq v\}$.

4.2 Variational calculus

In general, such a boundary may be characterised by variational-calculus methods that generalise the familiar first-order-condition reasoning to functional spaces. (We are choosing a function $d : [0, \infty) \rightarrow [0, \infty]$ here, not just a vector (x_1, \dots, x_n) .) Let's write $\Phi(d) \in \mathbf{R}$ for the principal's payoff from a boundary d .

Suppose that d is an optimal boundary. Consider perturbing d to d_ε , where ε is a real number $\neq 0$. By 'perturb', I mean that the gap $d_\varepsilon - d$ is 'small' when ε is, in the sense that d_ε converges to $d_0 = d$ as $\varepsilon \rightarrow 0$.⁴ Since d is optimal, we have

$$\frac{\Phi(d_\varepsilon) - \Phi(d)}{\varepsilon} \leq 0 \leq \frac{\Phi(d_{\varepsilon'}) - \Phi(d)}{\varepsilon'} \quad \text{for all } \varepsilon > 0 > \varepsilon'.$$

Thus letting $\varepsilon, \varepsilon' \rightarrow 0$ (and supposing, heroically, that the ratio converges!) yields a first-order condition:

$$\left. \frac{d}{d\varepsilon} \Phi(d_\varepsilon) \right|_{\varepsilon=0} \equiv \lim_{\varepsilon \rightarrow 0} \frac{\Phi(d_\varepsilon) - \Phi(d)}{\varepsilon} = 0.$$

One typically considers a large *family* of perturbations, yielding a *family* of such first-order conditions. Such a family of first-order conditions is called an *Euler equation*.

The art of variational calculus is to identify the 'right' perturbations d_ε . This should be guided by economic intuition; the formalism should (and will) follow. There are many candidates that might seem natural. For example, we may add ε on a small interval $[u', u' + \delta)$:

$$d_\varepsilon^{u', \delta}(u) := \begin{cases} d(u) + \varepsilon & \text{for } u \in [u', u' + \delta) \\ d(u) & \text{for } u \notin [u', u' + \delta). \end{cases}$$

By letting $\varepsilon \rightarrow 0$, we get a first-order condition for each u' and δ . To obtain a more parsimonious family of first-order conditions, let δ vanish, too:

$$\left. \frac{d}{d\delta} \frac{d}{d\varepsilon} \Phi(d_\varepsilon^{u', \delta}) \right|_{\varepsilon=0} \Big|_{\delta=0} = 0 \quad \text{for each } u' \in \mathbf{R}_+.$$

This is a pretty common Euler equation. Another natural (but rare) perturbation is $d_\varepsilon(u) = d(u + \varepsilon)$. These are just two examples of perturbations; both are useful,⁵ and there are many others.

⁴The formal meaning of this 'convergence' of functions must of course be specified; there are various concepts available. (In formal language, we must specify the *topology*.)

⁵E.g. the former yields the Euler equation in Curello and Sinander (2024), and the latter constitutes the main definition in Sinander (2022).

The theory of ‘optimal control’ (the Pontryagin maximum principle) is one way of thinking about a particular class of perturbations.

The literature. If you want to learn about variational calculus, and more generally about optimisation in functional spaces (aka ‘infinite-dimensional’ spaces), then Luenberger (1969) is a fantastic place to start. Optimal control is covered well by Seierstad and Sydsæter (1987).

4.3 Solving

Returning to the model, it turns out (remarkably!) that with a bit of extra structure, the optimal delegation set (or rather, its boundary d) admits a clean characterisation. This can be established using variational calculus (or more precisely, optimal control).

Write q for the probability mass function governing the random number N of available allocations. We assume that conditional on $N = n$, each allocation is an independent draw (U, V) from a smooth distribution on \mathbf{R}_+^2 ; we write f for the (differentiable) density of the marginal distribution of U , and $g(\cdot|u)$ for the (differentiable) conditional distribution of V .

For a given boundary d , write

$$p(u) := \int_{d(u)}^{\infty} g(\cdot|u)$$

for the proportion of type- u allocations that are permitted. The proportion of type- u -or-higher allocations that are permitted is then $\int_u^{\infty} pf$. Write

$$x(u) := 1 - \int_u^{\infty} pf$$

for the complementary probability that a given allocation is either banned or yields payoff less than u (see Figure 1). In particular, $x(0)$ is the proportion of banned allocations. x is evidently increasing, hence differentiable a.e. with derivative

$$x' = pf.$$

In addition, x must satisfy $\lim_{u \uparrow \infty} x(u) = 1$.

Conditional on there being $N = n$ allocations, then the probability that all allocations that yield at least u are banned is $x(u)^n$. It follows that the (unconditional) probability that all allocations that yield at least u are

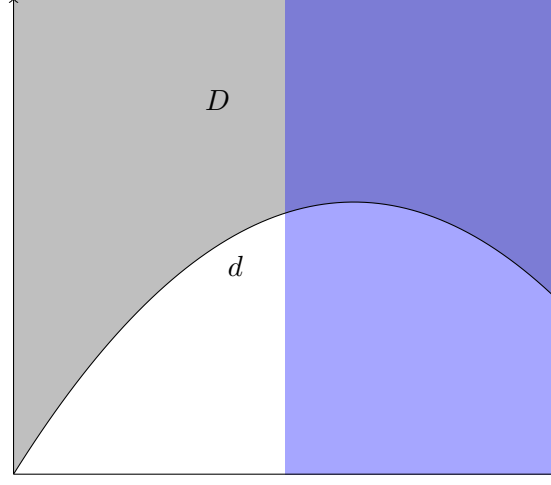


Figure 1: A delegation set D (grey) and the set of allocations that yield agent utility at least u (blue). $1 - x(u)$ is the measure of the intersection.

banned is $\phi(x(u))$, where

$$\phi(\xi) := \sum_{n=0}^{\infty} \xi^n q(n).^6$$

We can now write down the principal's problem formally. Write (U, V) for an arbitrary random allocation. For a given boundary d , the principal's payoff is

$$\begin{aligned} \Phi(d) &= \int_0^{\infty} \mathbf{E}(V | d(u) \leq V) d\phi(x(u)) \\ &= \int_0^{\infty} \frac{\int_{d(u)}^{\infty} v g(v|u) dv}{p(u)} \phi'(x(u)) p(u) f(u) du \\ &= \int_0^{\infty} V(d, u) \phi'(x(u)) p(u) f(u) du, \end{aligned}$$

where

$$V(d, u) := \mathbf{E}(V | U = u, d(u) \leq V) = \frac{1}{p(u)} \int_{d(u)}^{\infty} v g(v|u) dv.$$

⁶ ϕ is called the *probability generating function* (pgf) of the distribution q . Pgfs have certain useful general properties, such as continuity, monotonicity and convexity.

Her problem is therefore

$$\begin{aligned} & \max_{d,x,p} \int_0^\infty V(d,u) \phi'(x(u)) p(u) f(u) du \\ \text{s.t. } & x' = pf \quad \text{a.e.} \\ & \lim_{u \uparrow \infty} x(u) = 1 \\ & p(u) = \int_{d(u)}^\infty g(\cdot|u) \quad \text{for every } u \in \mathbf{R}_+. \end{aligned}$$

Letting $G(v|u) := \int_0^v g(\cdot|u)$, we have

$$p(u) = 1 - G(d(u)|u),$$

so that the principal's problem reads

$$\begin{aligned} & \max_{d,x} \int_0^\infty V(d,u) [1 - G(d(u)|u)] \phi'(x(u)) f(u) du \\ \text{s.t. } & x'(u) = [1 - G(d(u)|u)] f(u) \quad \text{for a.e. } u \in \mathbf{R}_+ \\ & \lim_{u \uparrow \infty} x(u) = 1. \end{aligned}$$

This is an optimal control problem, with d the control and x the state. The Pontryagin maximum principle gives necessary conditions for optimality. Some conditions on f and g are enough to guarantee that the program is concave, so that the necessary conditions are also sufficient.

One general feature of solutions is that $d(0) = 0$ ('no distortion at the bottom'). This is pretty obvious: the only reason to have $d(u) > 0$ is to deter the agent from choosing high- u -low- v allocations, but there are no allocations with u lower than zero.

Unsurprisingly, for the special case $q(0) + q(1) = 1$ in which the agent never has the choice between multiple allocations, it is optimal to use the 'naïve' delegation policy $d = 0$ (always permit everything). A few more qualitative things can be said about the solution. When q is the Poisson pmf, a solution is available in closed form.

The literature. Armstrong and Vickers (2010) solve for the optimal d using optimal control. A note: they parametrise their model so that the principal's payoff is $v + \alpha u$ for some $\alpha \geq 0$, so their results are stated somewhat differently. They derive comparative statics with respect to α (an 'ally principle': more aligned preferences lead to more permissive delegation), consider an extension in which allocations are generated by the agent via

costly and unobservable effort (moral hazard), and show that monetary transfers to the agent may not be used even if available. They also provide an example of how the principal can do better if the agent is able to propose more than one allocation; this issue is studied further by Guo and Shmaya (2023).

4.4 Random mechanisms

In the model of the previous section, suppose that there are two allocations (besides the status quo) for sure (i.e. $q(2) = 1$), independently drawn from a binary distribution that assigns equal probability to $(2, 4)$ and to $(3, 1)$. Optimal mechanisms must obviously permit the principal's favourite allocation $(2, 4)$; the only question is whether or not to permit the 'tempting' allocation $(3, 1)$. Banning it gives the principal an expected payoff of $\frac{3}{4} \times 4 = 3$, while allowing it gives her only $\frac{1}{4} \times 4 + \frac{3}{4} \times 1 = 1.75$. So it is best to ban $(3, 1)$ altogether.

But suppose the principal can commit to approve $(3, 1)$ with a small but positive probability. That is, the principal commits to a rule which specifies for each allocation the probability with which it is approved if proposed.

In particular, suppose she commits to approve $(2, 4)$ for sure, and to approve $(3, 1)$ with probability $p \in (0, 1)$. When $(2, 4)$ is available, the agent will still prefer to propose it rather than $(3, 1)$ provided $2 \geq 3p$, or $p \leq 2/3$. The random approval mechanism with $p = 2/3$ is clearly better: the principal still gets 4 whenever $(2, 4)$ is available (whether or not $(3, 1)$ is available, the agent will propose $(2, 4)$), and additionally, when $(3, 1)$ is the only available allocation, she now earns 1 with probability $2/3$ rather than zero for sure.

Remark 8. If the agent behaves as expected, by proposing $(3, 1)$ only when $(2, 4)$ is unavailable, then when the principal observes $(3, 1)$ being proposed, she will be tempted ex post to approve $(3, 1)$ for sure, not just with probability $2/3$. When considering random mechanisms, we are assuming that the principal can commit ex ante not to give in to such ex-post temptations: she is able *credibly* to promise ex ante that she will approve $(3, 1)$ with (exactly) probability $p = 2/3$. That is arguably a very strong assumption. Can you think of situations in which it might be reasonable?

A taxation principle for random mechanisms lets us restrict attention to 'random approval mechanisms', defined by a function $\delta : \mathbf{R}_+^2 \rightarrow [0, 1]$: the agent proposes an allocation $(u, v) \in \mathbf{R}_+^2$, and it is approved with probability $\delta(u, v)$. (If the proposal is not approved, then the status quo is implemented.)

By analogy with our ‘upward closed’ result, it is easily seen that we may restrict attention to mechanisms δ such that $\delta(u, \cdot)$ is weakly increasing.

Solving for *optimal* random approval mechanisms appears difficult; no-one has managed it so far, anyway.

The literature. Armstrong and Vickers (2010) noted the benefits of randomisation, but did not study random mechanisms systematically. Guo and Shmaya (2023) make progress by replacing our ‘Bayesian’ objective (the principal has some fixed belief about how A is distributed) with an uncertainty-averse ‘minmax regret’ objective. This allows them to obtain a clean solution, and (more importantly?) some nice qualitative insights about how the principal should optimally utilise her power to randomise.

4.5 Dynamics

There are two dynamic papers in this literature: Bird and Frug (2019) and Curello and Sinander (2024).

4.5.1 IID availability, simple payoffs, budget constraint

In Bird and Frug (2019) there is a single allocation (u, v) . The principal prefers this allocation to the status quo, while the agent does not ($u < 0 < v$). We assume that $u + v > 0$, so that the allocation (u, v) is (Kaldor–Hicks) efficient.

The principal can ‘reward’ the agent at linear cost to herself; in other words, she can increase the agent’s utility by $r \geq 0$ at a cost of cr to herself, for some fixed $c > 0$ that we’ll normalise to $c = 1$. A natural interpretation is that $r \geq 0$ reflects a monetary transfer to the agent. (The authors have a more complicated interpretation in terms of randomly-arriving ‘reward opportunities’.)

The model is dynamic, and the players have the same discount factor $\delta \in (0, 1)$. In each period, the allocation (u, v) may or may not be available; the agent knows whether it is, and the principal does not. Availability is IID over time. The principal can commit to future rewards (and approval).

As stated, the principal’s problem is straightforward: she can incentivise disclosure by committing to pay the agent $-u$ every time she discloses. Disclosure clearly cannot be incentivised more cheaply, and the principal prefers this mechanism to not incentivising disclosure at all because we assumed that $u + v > 0$. So this is what’s optimal.

The authors augment this setting with a (per-period) budget constraint for rewards: the principal only has $b \in (0, -u)$ units of reward resources (e.g. money) in each period. (If we had $b \geq -u$, then the constraint would never bind.) The authors allow the available resources b to vary randomly between periods; this doesn't materially affect the results.

The constraint means that with present resources alone, the agent cannot be rewarded sufficiently to incentivise her to disclose. So if she is to be incentivised, then she must be promised *future* rewards as well. This sounds precarious, and it is: if the allocation appears frequently enough, then the principal will eventually be forced to promise to pay the agent b in every future period, i.e. to max out her promise-making ability. Once that has happened, the principal cannot incentivise the agent any longer: having pledged all of her future resources, she resigns herself to the status quo allocation being implemented forevermore.

The authors show that this 'maxing-out' will happen a.s. in finite time. (If you know some probability theory, then you'll see that this follows from the Borel–Cantelli lemma.)

4.5.2 Single breakthrough, general payoffs

In Curello and Sinander (2024), there is initially available some set A_0 of allocations. At an uncertain time τ , a *breakthrough* occurs, expanding the set of available allocations to $A_1 \supseteq A_0$. The two sets of allocations are arbitrary, so that payoffs are entirely general. This allows for any kind of 'rewarding the agent' you like, including the linear specification of the previous paper. No exogenous constraints (e.g. budget) are imposed. The agent can disclose the availability of the new allocations (i.e. 'propose' them) at any time after the breakthrough, but not before.

We characterise optimal mechanisms in this setting. The main finding is that optimal mechanisms have a deadline structure.

Appendix A

Measure and integral

The theory of measure and (Lebesgue) integration are the foundation of modern real analysis, which is in turn the backbone of economic theory. You do not need to know it to take this course. But to understand what I'm saying, it is necessary at least to know some of the basic *language* of measure theory; that's what I'll cover here.

(If you'd like to pursue research in economic theory, then I would advise you to learn basic measure theory. I taught myself from Rosenthal (2006); this book is very accessible, except that chapter 2 is harder than it needs to be, so don't get stuck there! I now prefer the first few chapters of Folland (1999), a very beautiful book for first-year graduate students in maths. There are lots of other standard books. Efe Ok has a 'measure and probability' manuscript on his website that is specifically aimed at economists.)

Let X be a non-empty set. A σ -algebra on X is a collection of subsets of X satisfying certain properties.¹ If \mathcal{X} is a σ -algebra, we call (X, \mathcal{X}) a *measurable space*, and call the elements of \mathcal{X} the *measurable subsets* of X . Often the σ -algebra \mathcal{X} is left partly or entirely implicit, by the way.

A *measure* on a measurable space (X, \mathcal{X}) is a map $\mu : \mathcal{X} \rightarrow [0, \infty]$ that is countably additive: for any countable collection A_1, A_2, \dots of pairwise disjoint measurable subsets of X , we have $\mu(\bigcup_{n \in \mathbf{N}} A_n) = \sum_{n \in \mathbf{N}} \mu(A_n)$. The triple (X, \mathcal{X}, μ) is called a *measure space*.

A measurable set $A \subseteq X$ is called μ -null iff $\mu(A) = 0$. If a property holds at every $x \in X$, except possibly for x belonging to a null set A , then that property is said to hold *(μ -)almost everywhere*, or ' $(\mu$ -)a.e.'.

Example 2. The most commonly-used measure on $X = \mathbf{R}$ is the *Lebesgue*

¹The properties are closedness under complement and closedness under countable union.

measure, which is the unique measure λ with the property that $\lambda([a, b]) = b - a$ for all $a < b$. That is, it captures the common-sense notion of *length*.

Analogously, there's Lebesgue measure on \mathbf{R}^2 , which captures *area*, and Lebesgue measure on \mathbf{R}^3 , which captures *volume*, and so on.

The Lebesgue measure is conventionally defined on the *Lebesgue σ -algebra* (whose elements are called *Lebesgue sets*). It is often easier to work with the coarser *Borel σ -algebra* (or *Borel sets*); this is the smallest σ -algebra containing every open subset of \mathbf{R} .

If μ has the further property that $\mu(X) = 1$, then it is a *probability measure*, and (X, \mathcal{X}, μ) is a *probability space*. In the probability context, 'almost everywhere' is usually replaced with 'almost sure(ly)' or 'a.s.'.

Now consider a function $f : X \rightarrow Y$, where both X and Y are measurable spaces. For any measurable set $B \subseteq Y$, a measure on Y can tell us how 'large' the set B is. How large, then, is the set $A = \{x \in X : f(x) \in B\}$ of points x in X that lead to a value $f(x)$ that lives in B ? That question only has an answer if A is a measurable set! We call a function *measurable* if this question has an answer for every set B . Symbolically, f is measurable iff for every measurable subset B of Y , $\{x \in X : f(x) \in B\}$ is a measurable subset of X .

In modern analysis, the standard integral is the Lebesgue integral. This is the integral used almost exclusively in economic theory, including these notes. The integral of a measurable function $f : X \rightarrow \mathbf{R}$ on a measure space (X, \mathcal{X}, μ) is written $\int_X f d\mu$. The integral over a measurable subset $A \subseteq X$ is defined

$$\int_A f d\mu := \int_X f \mathbf{1}_A d\mu,$$

where $\mathbf{1}_A(x) := 1$ for $x \in A$ and $:= 0$ for $x \notin A$.

On probability spaces, measurable functions are conventionally called *random variables*, and integrals are called *expectations*. One writes

$$\mathbf{E}(f) := \int_X f d\mu.$$

If you have taken introductory calculus, then you may be more familiar with the *Riemann integral*.² The Riemann integral has the problem that the integral of some functions simply fail to exist. The Lebesgue integral extends the Riemann integral: it allows more functions to be integrated, while still giving the same result as the Riemann integral whenever the latter exists.

²The Riemann integral of a function $\mathbf{R} \rightarrow \mathbf{R}$ is defined as the limit of the areas of increasingly fine rectangles approximating the area under the graph of the function.

The Lebesgue integral is defined only for functions $f : X \rightarrow \mathbf{R}$ that are measurable. That is a necessary condition, but it is not sufficient: the existence of the integral requires a further condition. (If the further condition fails, the definition of the Lebesgue integral yields the expression $\infty - \infty$, which has no meaning; therefore we agree to say that the integral does not exist in such cases.) The integral of a function may be infinite (equal to ∞ or $-\infty$). A function is called *integrable* iff both (a) its integral exists, and (b) its integral is finite.

Whereas the Riemann integral is defined only for functions $f : \mathbf{R} \rightarrow \mathbf{R}$, the Lebesgue integral makes sense for functions $f : X \rightarrow \mathbf{R}$ defined on any space X you like, provided it has measurable structure (i.e. is equipped with a σ -algebra). This is very useful.

Appendix B

Convexity

A *vector space* (or *linear space*) is a set V for which it makes sense to (a) add two elements of V together, and (b) to multiply an element of V by a scalar. (That is, in both cases, the operation yields another element of V .) These ‘addition’ and ‘scalar multiplication’ operations have to satisfy a number of axioms that you can find on Wikipedia.

That’s rather abstract, but vector spaces are all around. The familiar example is \mathbf{R}^n , where $n \in \mathbf{N}$. There is a standard definition of what it means to ‘add’ two vectors together: it means adding up each component separately, i.e. for vectors $x, y \in \mathbf{R}^n$, $z = x + y$ is defined by

$$z_i := x_i + y_i \quad \text{for each component } i \in \{1, \dots, n\}.$$

Similarly, scalar multiplication has a standard definition: for a vector $x \in \mathbf{R}^n$ and scalar $\alpha \in \mathbf{R}$, $y = \alpha x$ is the vector whose i th entry is $y_i := \alpha x_i$. This makes \mathbf{R}^n a vector space. It is finite-dimensional (its dimension is n).

Much of the familiar structure of the finite-dimensional vector space \mathbf{R}^n extends to fancier vector spaces. Here’s a useful vector space: the space of all functions $f : [0, 1] \rightarrow \mathbf{R}$, where addition is defined ‘pointwise’ ($h = f + g$ is the function such that $h(x) := f(x) + g(x)$ for every ‘point’ $x \in [0, 1]$), and scalar multiplication is also defined ‘pointwise’ ($g = \alpha f$ for $\alpha \in \mathbf{R}$ is the function such that $g(x) := \alpha f(x)$ for every $x \in [0, 1]$). This vector space is infinite-dimensional; indeed we can think of a function $f : [0, 1] \rightarrow \mathbf{R}$ as a vector with uncountably many entries, one for each $x \in [0, 1]$, with value $f(x)$. (An n -vector (x_1, \dots, x_n) may similarly be viewed as the function $f : \{1/n, 2/n, \dots, (n-1)/n, 1\} \rightarrow \mathbf{R}$ that has $f(i/n) := x_i$ for each i .)

A subset C of a vector space is *convex* iff for any two of its members $x, y \in C$, every convex combination $\alpha x + (1 - \alpha)y$ for $\alpha \in (0, 1)$ is also a

member of C . In \mathbf{R}^n , you'll already be familiar with what convex sets look like.

Example 2. Consider the vector space of functions $[0, 1] \rightarrow \mathbf{R}$. Let C be the set of all functions $h : [0, 1] \rightarrow \mathbf{R}$ which have the form $f(x) = k + (2 - k)x$ for some $k \in [0, 2]$. (Try drawing a picture to visualise C .) This is a convex set. (Prove it!)

An element x of a convex set C is called an *extreme point* iff it cannot be written as a convex combination of two *distinct* points in C . (Any point is the convex combination of itself with itself, of course!) That is, $x \in C$ is an extreme point of C iff we cannot find $y \neq z$ in C and $\alpha \in (0, 1)$ such that $x = \alpha y + (1 - \alpha)z$.

Example 2 (continued). Return to the previous example. I claim that the function $\bar{f} \equiv 2$ (constant, always equal to 2) is an element of C , and is in fact an extreme point of C . (Convince yourself! A drawing will help.) I further claim that the function \underline{f} given by $\underline{f}(x) = 2x$ is an extreme point of C . (Convince yourself!) Finally, these are the only extreme points of C .

If C is a convex subset of a vector space, a function $\phi : C \rightarrow \mathbf{R}$ is called *convex* exactly if

$$\phi(\alpha x + (1 - \alpha)y) \leq \alpha\phi(x) + (1 - \alpha)\phi(y) \quad \text{for all } x, y \in C \text{ and } \alpha \in (0, 1).$$

If the inequality is an equality (for all x, y and α) then ϕ is called *linear*.

It is a fact that if C is a convex set and is suitably compact, and if $\phi : C \rightarrow \mathbf{R}$ is a convex function and is suitably continuous, then there is a maximiser of ϕ on C that is an extreme point of C . (Some economists call this 'Bauer's maximum principle'.) This is easy to see in \mathbf{R}^2 (visualise it!). If ϕ is linear (not merely convex), then you don't have to worry about the 'suitably continuous' bit, as linearity gives it to you for free. (You *do* have to worry about the 'suitably compact' bit. In principle.)

Bibliography

- Abreu, D., Pearce, D., & Stacchetti, E. (1990). Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica*, 58(5), 1041–1063. <https://doi.org/10.2307/2938299>
- Armstrong, M., & Vickers, J. (2010). A model of delegated project choice. *Econometrica*, 78(1), 213–244. <https://doi.org/10.3982/ECTA7965>
- Ausubel, L. M., Cramton, P., & Deneckere, R. J. (2002). Bargaining with incomplete information. In R. Aumann & S. Hart (Eds.), *Handbook of game theory with economic applications* (pp. 1897–1945, Vol. 3). North-Holland.
- Ausubel, L. M., & Deneckere, R. J. (1989). Reputation in bargaining and durable goods monopoly. *Econometrica*, 57(3), 511–531. <https://doi.org/10.2307/1911050>
- Ball, I., Jackson, M. O., & Kattwinkel, D. (2022). Comment on Jackson and Sonnenschein (2007) “Overcoming incentive constraints by linking decisions”. *Econometrica*, 90(4), 3–7. <https://doi.org/10.3982/ECTA20516>
- Ball, I., & Kattwinkel, D. (2023). Corrigendum to “Role of linking mechanisms in multitask agency with hidden information” [J. Econ. Theory 145 (2010) 2241–2259]. *Journal of Economic Theory*, 210. <https://doi.org/10.1016/j.jet.2023.105666>
- Ball, I., & Kattwinkel, D. (2024). *Linking mechanisms: Limits and robustness* [working paper, 23 Apr 2024]. <https://doi.org/10.48550/arXiv.2309.07363>
- Ben-Porath, E., Dekel, E., & Lipman, B. L. (2014). Optimal allocation with costly verification. *American Economic Review*, 104(12), 3779–3813. <https://doi.org/10.1257/aer.104.12.3779>
- Bird, D., & Frug, A. (2019). Dynamic non-monetary incentives. *American Economic Journal: Microeconomics*, 11(4), 111–150. <https://doi.org/10.1257/mic.20170025>

- Bloch, F., Dutta, B., & Dziubiński, M. (2023). Selecting a winner with external referees. *Journal of Economic Theory*, 211. <https://doi.org/10.1016/j.jet.2023.105687>
- Border, K. C. (1991). Implementation of reduced form auctions: A geometric approach. *Econometrica*, 59(4), 1175–1187. <https://doi.org/10.2307/2938181>
- Border, K. C. (2007). Reduced form auctions revisited. *Economic Theory*, 31(1), 167–181. <https://doi.org/10.1007/s00199-006-0080-z>
- Brzustowski, T., Georgiadis-Harris, A., & Szentes, B. (2023). Smart contracts and the Coase conjecture. *American Economic Review*, 113(5), 1334–59. <https://doi.org/10.1257/aer.20220357>
- Bull, J., & Watson, J. (2007). Hard evidence and mechanism design. *Games and Economic Behavior*, 58(1), 75–93. <https://doi.org/10.1016/j.geb.2006.03.003>
- Bulow, J., & Klemperer, P. (1996). Auctions versus negotiations. *American Economic Review*, 86(1), 180–194.
- Che, Y.-K., Kim, J., & Mierendorff, K. (2013). Generalized reduced-form auctions: A network-flow approach. *Econometrica*, 81(6), 2487–2520. <https://doi.org/10.3982/ECTA11405>
- Coase, R. H. (1972). Durability and monopoly. *Journal of Law and Economics*, 15(1), 143–149. <https://doi.org/10.1086/466731>
- Condorelli, D. (2012). What money can't buy: Efficient mechanism design with costly signals. *Games and Economic Behavior*, 75(2), 613–624. <https://doi.org/10.1016/j.geb.2012.02.018>
- Condorelli, D. (2013). Market and non-market mechanisms for the optimal allocation of scarce resources. *Games and Economic Behavior*, 82, 582–591. <https://doi.org/10.1016/j.geb.2013.08.008>
- Crémer, J., & McLean, R. P. (1985). Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. *Econometrica*, 53(2), 345–361. <https://doi.org/10.2307/1911240>
- Crémer, J., & McLean, R. P. (1988). Full extraction of the surplus in Bayesian and dominant strategy auctions. *Econometrica*, 56(6), 1247–1257. <https://doi.org/10.2307/1913096>
- Curello, G., & Sinander, L. (2024). *Screening for breakthroughs* [working paper, 21 Feb 2024]. <https://doi.org/10.48550/arXiv.2011.10090>
- Daskalakis, C., Deckelbaum, A., & Tzamos, C. (2017). Strong duality for a multiple-good monopolist. *Econometrica*, 85(3), 735–767. <https://doi.org/10.3982/ECTA12618>

- de Clippel, G., Eliaz, K., Fershtman, D., & Rozen, K. (2021). On selecting the right agent. *Theoretical Economics*, 16(2), 381–402. <https://doi.org/10.3982/TE4027>
- Doval, L., & Skreta, V. (2024). Optimal mechanism for the sale of a durable good. *Theoretical Economics*, 19(2), 865–915. <https://doi.org/10.3982/TE4485>
- Dworczak, P., Kominers, S. D., & Akbarpour, M. (2021). Redistribution through markets. *Econometrica*, 89(4). <https://doi.org/10.3982/ECTA16671>
- Erlanson, A., & Kleiner, A. (2019). A note on optimal allocation with costly verification. *Journal of Mathematical Economics*, 84, 56–62. <https://doi.org/10.1016/j.jmateco.2019.06.002>
- Erlanson, A., & Kleiner, A. (2020). Costly verification in collective decisions. *Theoretical Economics*, 15(3), 923–954. <https://doi.org/10.3982/TE3101>
- Erlanson, A., & Kleiner, A. (2022). *Optimal allocations with capacity constrained verification* [work in progress].
- Eső, P. (2005). An optimal auction with correlated values and risk aversion. *Journal of Economic Theory*, 125(1), 78–89. <https://doi.org/10.1016/j.jet.2004.06.005>
- Folland, G. B. (1999). *Real analysis: Modern techniques and their applications* (2nd). Wiley.
- Frankel, A. (2016a). Delegating multiple decisions. *American Economic Journal: Microeconomics*, 8(4), 16–53. <https://doi.org/10.1257/mic.20130270>
- Frankel, A. (2016b). Discounted quotas. *Journal of Economic Theory*, 166, 396–444. <https://doi.org/10.1016/j.jet.2016.08.001>
- Fudenberg, D., Levine, D., & Tirole, J. (1985). Infinite-horizon models of bargaining with one-sided incomplete information. In A. E. Roth (Ed.), *Game-theoretic models of bargaining* (pp. 73–98). Cambridge University Press.
- Galichon, A. (2016). *Optimal transport methods in economics*. Princeton University Press.
- Gershkov, A., Goeree, J. K., Kushnir, A., Moldovanu, B., & Shi, X. (2013). On the equivalence of Bayesian and dominant strategy implementation. *Econometrica*, 81(1), 197–220. <https://doi.org/10.3982/ECTA10592>
- Gül, F., Sonnenschein, H., & Wilson, R. (1986). Foundations of dynamic monopoly and the Coase conjecture. *Journal of Economic Theory*, 39(1), 155–190. [https://doi.org/10.1016/0022-0531\(86\)90024-4](https://doi.org/10.1016/0022-0531(86)90024-4)

- Guo, Y., & Hörner, J. (2020). *Dynamic allocation without money* [working paper, 26 Aug 2020].
- Guo, Y., & Shmaya, E. (2023). Regret-minimizing project choice. *Econometrica*, 91(5), 1567–1593. <https://doi.org/10.3982/ECTA20157>
- Haghpanah, N., Kuvalekar, A., & Lipnowski, E. (2024). Buying from a group. *American Economic Review*, 114(8), 2596–2632. <https://doi.org/10.1257/aer.20230914>
- Hart, S., & Nisan, N. (2017). Approximate revenue maximization with multiple items. *Journal of Economic Theory*, 172, 313–347. <https://doi.org/10.1016/j.jet.2017.09.001>
- Hart, S., & Reny, P. J. (2015a). Implementation of reduced form mechanisms: A simple approach and a new characterization. *Economic Theory Bulletin*, 3, 1–8. <https://doi.org/10.1007/s40505-014-0055-3>
- Hart, S., & Reny, P. J. (2015b). Maximal revenue with multiple goods: Nonmonotonicity and other observations. *Theoretical Economics*, 10(3), 893–922. <https://doi.org/10.3982/TE1517>
- Hartline, J. D. (2016). *Mechanism design and approximation*. <http://jasonhartline.com/MDnA>
- Hotelling, H. (1932). Edgeworth’s taxation paradox and the nature of demand and supply functions. *Journal of Political Economy*, 40(5), 577–616. <https://doi.org/10.1086/254387>
- Ishii, Y. (2016). *Implementation with quota mechanisms* [work in progress].
- Jackson, M. O., & Sonnenschein, H. F. (2007). Overcoming incentive constraints by linking decisions. *Econometrica*, 75(1), 241–257. <https://doi.org/10.1111/j.1468-0262.2007.00737.x>
- Jéhiel, P., Moldovanu, B., & Stacchetti, E. (1999). Multidimensional mechanism design for auctions with externalities. *Journal of Economic Theory*, 85(2), 258–293. <https://doi.org/10.1006/jeth.1998.2501>
- Jeong, B.-h., & Pycia, M. (2023). *The first-price principle of maximizing economic objectives* [working paper, Nov 2023].
- Jewitt, I., & Quigley, D. (2023). *Conjugate persuasion* [working paper, Aug 2023].
- Kattwinkel, D. (2019). *Allocation with correlated information: Too good to be true* [working paper, 11 Dec 2019]. <https://doi.org/10.2139/ssrn.3615398>
- Kattwinkel, D., & Knoepfle, J. (2023). Costless information and costly verification: A case for transparency. *Journal of Political Economy*, 131(2), 504–548. <https://doi.org/10.1086/721618>

- Kirkegaard, R. (2006). A short proof of the Bulow–Klemperer auctions vs. negotiations result. *Economic Theory*, 28, 449–452. <https://doi.org/10.1007/s00199-004-0593-2>
- Kleiner, A., & Manelli, A. (2019). Strong duality in monopoly pricing. *Econometrica*, 87(4), 1391–1396. <https://doi.org/10.3982/ECTA15735>
- Kleiner, A., Moldovanu, B., & Strack, P. (2021). Extreme points and majorization: Economic applications. *Econometrica*, 89(4), 1557–1593. <https://doi.org/10.3982/ECTA18312>
- Laffont, J.-J., & Martimort, D. (2000). Mechanism design with collusion and correlation. *Econometrica*, 68(2), 309–342. <https://doi.org/10.1111/1468-0262.00111>
- Lai, T. L. (1974). Summability methods for independent identically distributed random variables. *Proceedings of the American Mathematical Society*, 45(2), 253–261. <https://doi.org/10.1090/S0002-9939-1974-0356194-4>
- Laohakunakorn, K. (2023). *Rent extraction with information acquisition* [working paper, 7 Sep 2023].
- Lipnowski, E., & Ramos, J. (2020). Repeated delegation. *Journal of Economic Theory*, 188. <https://doi.org/10.1016/j.jet.2020.105040>
- Liu, Q. (2015). *A simple proof for the Coase conjecture* [working paper, 17 Sep 2015].
- Loertscher, S., & Muir, E. V. (2022). Monopoly pricing, optimal randomization, and resale. *Journal of Political Economy*, 130(3). <https://doi.org/10.1086/717894>
- Lomys, N., & Yamashita, T. (2022). *A mediator approach to mechanism design with limited commitment* [working paper, 15 Aug 2022]. <https://doi.org/10.2139/ssrn.4116543>
- Lopomo, G., Rigotti, L., & Shannon, C. (2022). Detectability, duality, and surplus extraction. *Journal of Economic Theory*, 204. <https://doi.org/10.1016/j.jet.2022.105465>
- Luenberger, D. G. (1969). *Optimization by vector space methods*. Wiley.
- Mailath, G. J., & Samuelson, L. (2006). *Repeated games and reputations: Long-run relationships*. Oxford University Press.
- Manelli, A. M., & Vincent, D. R. (2010). Bayesian and dominant-strategy implementation in the independent private-values model. *Econometrica*, 78(6), 1905–1938. <https://doi.org/10.3982/ECTA8025>
- Mas-Colell, A., Whinston, M., & Green, J. R. (1995). *Microeconomic theory*. Oxford University Press.
- Maskin, E., & Riley, J. (1984). Optimal auctions with risk averse buyers. *Econometrica*, 52(6), 1473–1518. <https://doi.org/10.2307/1913516>

- Matsushima, H., Miyazaki, K., & Yagi, N. (2010). Role of linking mechanisms in multitask agency with hidden information. *Journal of Economic Theory*, 145(6), 2241–2259. <https://doi.org/10.1016/j.jet.2010.03.014>
- Matthews, S. A. (1984). On the implementability of reduced form auctions. *Econometrica*, 52(6), 1519–1522. <https://doi.org/10.2307/1913517>
- McAfee, R. P., & Reny, P. J. (1992). Correlated information and mechanism design. *Econometrica*, 60(2), 395–421. <https://doi.org/10.2307/2951601>
- Mierendorff, K. (2011). Asymmetric reduced form auctions. *Economics Letters*, 110(1), 41–44. <https://doi.org/10.1016/j.econlet.2010.09.019>
- Milgrom, P., & Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2), 583–601. <https://doi.org/10.1111/1468-0262.00296>
- Milgrom, P., & Shannon, C. (1994). Monotone comparative statics. *Econometrica*, 62(1), 157–180. <https://doi.org/10.2307/2951479>
- Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research*, 6(1), 58–73. <https://doi.org/10.1287/moor.6.1.58>
- Myerson, R. B. (1982). Optimal coordination mechanisms in generalized principal–agent problems. *Journal of Mathematical Economics*, 10(1), 67–81. [https://doi.org/10.1016/0304-4068\(82\)90006-4](https://doi.org/10.1016/0304-4068(82)90006-4)
- Mylovanov, T., & Zapechelnyuk, A. (2017). Optimal allocation with ex post verification and limited penalties. *American Economic Review*, 107(9), 2666–2694. <https://doi.org/10.1257/aer.20140494>
- Nava, F., & Schiraldi, P. (2019). Differentiated durable goods monopoly: A robust Coase conjecture. *American Economic Review*, 109(5), 1930–68. <https://doi.org/10.1257/aer.20160404>
- Neeman, Z. (2004). The relevance of private information in mechanism design. *Journal of Economic Theory*, 117(1), 55–77. <https://doi.org/10.1016/j.jet.2003.09.001>
- Onuchic, P., & Ray, D. (2023). Conveying value via categories. *Theoretical Economics*, 18(4), 1407–1439. <https://doi.org/10.3982/TE5026>
- Pavan, A., Segal, I., & Toikka, J. (2014). Dynamic mechanism design: A Myersonian approach. *Econometrica*, 82(2), 601–653. <https://doi.org/10.3982/ECTA10269>
- Phelps, R. R. (2001). *Lectures on Choquet’s theorem* (2nd). Springer.
- Robert, J. (1991). Continuity in auction design. *Journal of Economic Theory*, 55(1), 169–179. [https://doi.org/10.1016/0022-0531\(91\)90064-B](https://doi.org/10.1016/0022-0531(91)90064-B)
- Rochet, J.-C. (1987). A necessary and sufficient condition for rationalizability in the quasilinear context. *Journal of Mathematical Economics*, 16(2), 191–200. [https://doi.org/10.1016/0304-4068\(87\)90007-3](https://doi.org/10.1016/0304-4068(87)90007-3)

- Rockafellar, R. T. (1970). *Convex analysis*. Princeton University Press.
- Rosenthal, J. S. (2006). *A first look at rigorous probability theory* (2nd). World Scientific.
- Roy, R. (1947). La distribution du revenu entre les divers biens. *Econometrica*, 15(3), 205–225. <https://doi.org/10.2307/1905479>
- Samuelson, P. A. (1947). *Foundations of economic analysis*. Harvard University Press.
- Seierstad, A., & Sydsæter, K. (1987). *Optimal control theory with economic applications*. North-Holland.
- Shephard, R. W. (1953). *Cost and production functions*. Princeton University Press.
- Sinander, L. (2022). The converse envelope theorem. *Econometrica*, 90(6), 2795–2819. <https://doi.org/10.3982/ECTA18119>
- Spear, S. E., & Srivastava, S. (1987). On repeated moral hazard with discounting. *Review of Economic Studies*, 54(4), 599–617. <https://doi.org/10.2307/2297484>
- Toikka, J. (2011). Ironing without control. *Journal of Economic Theory*, 146(6), 2510–2526. <https://doi.org/10.1016/j.jet.2011.06.003>
- Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance*, 16(1), 8–37. <https://doi.org/10.1111/j.1540-6261.1961.tb02789.x>
- Vohra, R. V. (2011). *Mechanism design: A linear programming approach*. Cambridge University Press.
- Yang, K. H. (2022). Selling consumer data for profit: Optimal market-segmentation design and its consequences. *American Economic Review*, 112(4), 1364–1393. <https://doi.org/10.1257/aer.20210616>