Essays in Microeconomic Theory

Inaugural-Dissertation

zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Deniz Kattwinkel

aus Buchen (Odenwald)

2021

Tag der Promotion: 22. Juni 2021

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Stephan Lauermann
Zweitreferent:	Prof. Dr. Benny Moldovanu
Tag der mündlichen Prüfung:	22. Juni 2021

Acknowledgements

First of all, I would like to thank my supervisors Stephan Lauermann and Benny Moldovanu for their guidance, encouragement and enduring support. I am also very grateful to Dirk Bergemann, Francesc Dilme, Marina Halac, Andreas Kleiner, Daniel Kraehmer, Sven Rady, Dezsö Szalay and Tymon Tatur.

I was fortunate to have had two fantastic co-authors for two of the three chapters of this thesis. I thank Ian Ball and Jan Knoepfle for the opportunity to work with them and the inspiring discussions we had. Furthermore, I thank my fellow students and colleagues — especially Nina Bobkova, Si Chen, Carl Heese, Claudia Herresthal, Helene Mass, Ludmila Matyskova, Andre Speit and Paul Voss — for many valuable discussions and the nice time.

I profited from the infrastructure and the financial support of both the Bonn Graduate School of Economics and the Collaborative Research Center Transregio 224 and want to thank all those involved in setting up and running these outstanding institutions, in particular the director of the BGSE, Benny Moldovanu, and the spokesperson of the CRC, Sven Rady. Special thanks for administrative support go to Silke Kinzig, Britta Altenburg and Andrea Reykers from the Bonn Graduate School of Economics and Sabine Burch from the Department of Economics.

Last but not least, I want to thank my family and friends who have always supported me during my studies.

Contents

List of Figures

List of Tables

Introduction

In the standard mechanism design paradigm, the designer commits to a mechanism, which induces a game among the agents. There are no explicit constraints on the mechanism, but there is an implicit assumption that the outcome of the induced game does not directly depend on the agent's type. Thus, each type can freely mimic every other type in the game that is induced by the mechanism. As a consequence, the principal's only channel of learning an agent's type is through the mechanism: the mechanism must make it optimal for the agent to reveal his type.

In practice, however, we observe mechanisms that go beyond this paradigm: insurers can costly verify claimants' statements before the payout. Employees receive confidential letters of recommendation about an applicant's quality.

In this thesis which consists of three self-contained papers, I address different deviations from this standard paradigm and its implications for the shape of optimal institutions.

In my first paper (Chapter 3), *Costless Information and Costly Verification: A Case for Transparency* (joint with Jan Knoepfle), we study the optimal mechanism for a principal who has to take a binary decision. She relies on information that is held by a completely biased agent. Monetary transfers between the agent and the principal are not feasible. In contrast to the standard paradigm, we assume the following: (i) The principal observes a noisy signal that is correlated with the agent's type. (ii) The principal can perfectly learn the type of the agent at a cost. In the optimal mechanism, the principal bases her decision on the signal and allows agent to appeal. If he does so, the principal costly verifies the agent's type.

We find that without money, the principal cannot complement the screening of the agent with her private information. She only screens via the costly verification after an appeal. This mechanism can be transparently implemented; the principal does not have to keep her signal secret from the agent. We then use our findings to argue that there is no value in keeping the charges secret from a defendant during a pretrial.

In my second paper (Chapter 2), *Probabilistic Verification in Mechanism Design* (joint with Ian Ball), we introduce a model of probabilistic verification in a mechanism design setting. The principal verifies the agent's claims with statistical tests. The agent's probability of passing each test depends on his type. We introduce a tractable reduced form for the testing technology. Because verification is noisy, our framework is amenable to the local first-order approach. This was not possible with previous models of verification, wherein the probability of detecting a lie discontinuously jumps from 0 to 1. Under our probabilistic verification, this probability can depend continuously on the agent's report; therefore each local constraint is loosened but not eliminated. We apply our framework to several classical revenue-maximization problems.

As the precision of the verification technology varies, our setting continuously interpolates between private information and complete information. This is reflected in the form of the optimal mechanisms we derive. They contrast with the discontinuous mechanisms that were derived under previous frameworks of verification.

In my job market paper (Chapter 1), *Allocation with Correlated Information: Too good to be true*, I study a principal–agent allocation problem. The agent privately learns the value of an indivisible good that the principal can allocate to him. The principal also has private information: she learns the cost of the allocation. Cost and value are correlated, and monetary transfers are not feasible. The principal wants to maximize welfare; whereas the agent wants to maximize the probability of receiving the good. This is not an informed principal problem, because the principal can design the mechanism contingent on the cost realization. I study how the principal utilizes her information in the optimal mechanism: when the correlation is negative, she bases her decision only on the costs. When the correlation is positive, she screens the agent. To this end, she forgoes her best allocation opportunities: when the agent reports high valuations but her costs are low. Under positive correlation, these realizations are unlikely; the principal will find them too good to be true. In contrast to standard results, this optimal mechanism may not allocate to a higher value agent with higher probability. I discuss the application to intra-firm allocations, task-delegation, and industry self-regulation.

The optimal mechanism utilizes positive correlation for screening. The possibility of using positive correlation for screening opens interesting new directions of future research. My characterization is the first step toward analyzing a richer setting in which the principal endogenously chooses the degree of positive correlation.

I introduce a novel regularity assumption on the distribution of costs and values. In my model, it guarantees the interval structure of the optimal mechanism. I believe that this assumption will be useful in other models too. In the paper I give an intuition that translates the heterogeneity of the agent's belief into heterogeneity of risk preferences. The assumption then guarantees that these risk preferences are ordered.

Chapter 1

Allocation with Correlated Information: Too good to be true *

Job Market Paper

1.1 Introduction

A principal (she) can allocate an indivisible good to an agent (he). She privately learns her costs c of allocating the good to him. He privately learns his valuation v for the good. She is benevolent and wants to allocate whenever v exceeds c. He does not incorporate the costs and wants the good in any case. In many applications of this setting, monetary transfers between the principal and the agent are not allowed. However, their information is often correlated. Examples include the following: the federal government decides whether to allocate a task to the local government, the management of a company decides whether to buy a capital good and allocate it to one of its departments, a regulator decides whether to approve a new product after she has investigated one of its features, and the producer of the good reports his assessment about the safety of the other features.

To my knowledge, this is the first study that analyzes the optimal mechanism in this bilateral trade setting with correlated information and without transfers.

As a benchmark, consider this setting when the designer could use monetary transfers to screen the agent. In this case, the efficient allocation could be implemented. The principal could charge her cost as the price for the good. Then, the agent would buy if and only if his valuation exceeds the cost.

* I thank my advisers, Stephan Lauermann and Benny Moldovanu, for their continual guidance. For helpful discussions, I thank: Ian Ball, Nina Bobkova, Gregorio Curello, Francesc Dilmé, Yingni Guo, Carl Heese, Matthew Jackson, Jan Knöpfle, Daniel Krähmer, Annie Liang, Helene Mass, Andre Speit, Rani Spiegler, Dezsö Szalay and Asher Wolinsky. I am grateful for funding from the German Research Foundation (DFG) through CRC TR 224 (Project B04). An extended abstract for this paper appeared in EC '20.

In many situations, however, monetary transfers are not feasible. This can be for organizational reasons (companies and state agencies do not want to introduce internal budgets), for technological reasons (a website and its user might not have a payment channel), or for moral and legal reasons.

Moreover, it is natural that valuation and costs are not independent but correlated:

If the federal government delegates a task to the local government, her opportunity costs are given by the value she could have created in the task herself. In this example, the values which the agent (local government) and the principal (federal government) could create in the same task are positively correlated.

If the management buys the capital good from a third party, her costs are given by the price. This price could be a quality signal. Then, the price and the valuation for the good are positively correlated. If the regulator investigates one aspect of a new product and the producer investigates another, these aspects are likely positively correlated.

In my setting, the principal designs a mechanism that can be contingent on the realization of her cost. Her objective is to maximize the ex-ante expected efficiency of the allocation. When she screens the agent to allocate efficiently, she has to discourage an agent with a value below her costs to pretend to have a higher value. Without money, she can only use the allocation itself as an incentive. Thus, she must offer a low-value agent sufficiently high expected allocation probability to prevent him from mimicking an agent with a higher value. However, allocating to low-value agents distorts the expected efficiency. In the optimal mechanism, she has to square the inefficiencies that screening entails with the efficiency gains that the information about the agent's value permits.

Results. The revelation principle applies in this setting. Therefore, it is sufficient to study direct mechanisms. In a direct mechanism, the principal commits to a menu of allocation schedules, with one allocation schedule for each valuation. An allocation schedule specifies an allocation probability for any cost realization. The agent learns his valuation, forms expectations about the cost, and then reports his valuation to the principal. The principal learns the cost and allocates according to the schedule for the reported valuation. Because cost and valuation can be correlated, agents of different valuations form different expectations about the cost and evaluate allocation schedules differently. The agent must find it optimal to report truthfully. Therefore, he must expect the allocation schedule that corresponds to his true valuation to yield him the highest probability of allocation.

When cost and valuation are independent or negatively correlated, the optimal mechanism ignores the agent's report and bases the allocation decision only on the cost: the principal commits to implement the same allocation schedule after all valuation reports. She allocates if and only if the cost falls below a cutoff. The reason is that with negative correlation, low costs tend to occur with high valuations and vice versa. Therefore, basing the decision only on the costs is already quite efficient. Screening the agent would only slightly increase the expected efficiency.

When costs and valuations are positively correlated, the optimal mechanism might screen the agent. To illustrate this screening mechanism, suppose there are two possible valuations $v_L < v_H$. If the agent reports the low valuation v_L, the principal sticks to a cutoff allocation schedule: she only allocates if the cost remains below a cutoff, $c < \bar{c}_L$. If the agent claims to have high valuation v_H, the allocation schedule is characterized by two cost cutoffs $\underline{c}_H \leq \bar{c}_H$ that form an interval. If the cost is either too low or too high, the principal does not allocate the good. She allocates only after intermediate cost realization between the two cutoffs, $c \in (\underline{c}_H, \bar{c}_H)$. This means that the principal does not allocate when the gains from allocation $v - c$ are the highest. The form of this allocation schedule exploits the difference in beliefs of the two valuation types. Under positive correlation, an agent with v_L finds low costs likelier than an agent with v_H. When the principal chooses to not allocate after a high valuation report and low costs, she makes misreporting for the low type unattractive. She forgoes the most efficient allocation opportunity; however this allows her to allocate overall with a higher probability to v_H without giving the agent with v_L incentives to misreport.

With more than two valuations, the structure from the above example extends. I introduce a novel regularity assumption on the joint distribution. Under this assumption, the allocation schedule for all valuation reports has interval form. Furthermore, the intervals are ordered. For a higher valuation, both the upper and lower support of the allocation schedule exceed the upper and lower support of the allocation schedule of a lower valuation.

With correlated information, agents of different valuations hold different beliefs about the costs. Therefore, the interim expected allocation probabilities are insufficient to describe the mechanism and the standard approaches fail. In particular, the set of incentive compatible mechanisms cannot be easily characterized. Therefore, I directly characterize the optimal mechanism. First, I use the new regularity assumption to argue that the allocation schedules of the optimal mechanism are in interval form (Proposition 1). This assumption is similar (but weaker) to an assumption Jewitt (1988) uses in a moral hazard setting. I illustrate the role of this assumption by introducing a related problem. In this related problem, agents of different valuations share the same belief about the cost but differ in their risk preferences. The regularity assumption then guarantees that in the related problem, agents with higher valuation are more risk-averse. This leads to the interval form of the allocation schedules.

Next, I show that the allocation schedules are ordered (Proposition 2). When an agent evaluates an allocation schedule, he weighs the probability of receiving the good at a certain cost realization with his belief that this cost realization occurs. This evaluation is as if the agent would have a Bernoulli utility function with respect to costs that is equal to his belief and would evaluate a lottery that is given by the allocation schedule. I establish that these utility functions have single-crossing expectational differences. Finally, I use the resulting monotone comparative statics to show that the interval allocation schedules in the optimal mechanism are ordered. In contrast to standard results, this interval monotonicity does not imply that the interim allocation probabilities are monotone in the agent's valuation.

Applications. The characterization of the optimal mechanism has many interesting consequences for applications: if the principal buys the good from a third party to allocate it to an agent inside her organization, her demand schedule is no longer monotone. When the agent reports a high valuation, she will buy for intermediate prices but will not buy for high or low prices.

The optimal mechanism under positive correlation gives a rationale for inefficient governmental allocation of resources and tasks. This demands a high degree of commitment from the principal. It might be difficult for a public official to defend the decision to not allocate a task to a subordinate agency if (i) the subordinate agency predicts to be very successful in the task whereas (ii) the principal agency expects to perform poorly. This advocates for an intransparent allocation procedure, where the performance predictions do not become public record. In contrast, under negative correlation, the optimal mechanism can be transparently implemented: the principal does not collect any information from the agent. Given only the realization of the costs, her decision is efficient.

The possibility of using positive correlation for screening has interesting effects on the value of correlation. I demonstrate this with a numerical example, wherein the principal compares two joint distributions with equal marginal expectations for costs and valuation. The second distribution exhibits a higher degree of positive correlation than the first. As a benchmark, I consider the case without information asymmetry, where the principal can observe the agent's valuation. Here, she prefers the first distribution because the higher positive correlation lowers her expected value of the efficient allocation. However, if the valuation is the agent's private information, the efficient allocation is not implementable. Then, the principal prefers the second distribution. The higher degree of positive correlation allows her to screen the agent and to allocate more efficiently in the second-best solution.

Finally, I investigate whether a regulator can delegate parts of the certification process of a new product to the producer. I model this long-standing practice of industry self-regulation as follows. The regulator analyzes one aspect of the safety of the product and learns its value a_1; a second aspect a_2 is analyzed by the producer. The two aspects are positively correlated and jointly determine the safety of the product. To fit this setting into my model of valuation and cost $v - c$, I define the principal's cost as $c = -a_1$ and the agent's valuation as $v = a_2$. With this definition, cost and valuation are negatively correlated. In the optimal mechanism, the regulator bases her decision only on the aspect that she herself analyzes. In this model of industry self-regulation, delegation of an aspect implies ignoring it for the assessment of the safety of the product.

Related literature. Myerson and Satterthwaite (1983) introduce the bilateral trade setting and show that with balanced transfers, the efficient allocation cannot be implemented. I deviate from their setting in three dimensions: (i) I let the principal (the seller in their terminology) design the mechanism. She can commit to make her decision contingent on her information. In contrast to Myerson and Satterthwaite, the optimal mechanism does not have to consider her incentive constraints. This also sets my model apart from informed principal problems (Myerson (1983) and Maskin and Tirole (1990)). (ii) I allow cost and valuation to be correlated. (iii) I do not allow for monetary transfers. If monetary transfers are allowed and the information of the players is correlated, Cremer and McLean (1988), Riordan and Sappington (1988), Johnson, Pratt, and Zeckhauser (1990), McAfee and Reny (1992), and Neeman (2004) have demonstrated how and when a designer can exploit the differences in the beliefs of the players to implement any social choice function. I revisit some of their results in section 1.3.1.

My model also relates to the literature on delegation started by Holmstrom (1977). To my knowledge, it is the first that studies delegation with two-sided information asymmetry. The interval structure of the optimal mechanism is similar to the interval delegation in Alonso and Matouschek (2008).

In settings without monetary transfers, the optimal mechanism often invents artificial money to screen agents. Hylland and Zeckhauser (1979) have agents trade allocation probabilities for indivisible goods; Jackson and Sonnenschein (2007) have voters trade the weights of their ballot in elections on different matters.

Bhargava, Majumdar, and Sen (2015) show how positively correlated beliefs among voters allow overcoming the impossibility of nondictatorial voting rules established by Gibbard (1973) and Satterthwaite (1975).

Guo and Hörner (2018) study a dynamic allocation problem between a benevolent principal and an agent. They use promises of future allocations as a means to incentivizes the agent to report truthfully.

Another strand in the literature on mechanism design studies allocation problems without money but allows for the costly verification of hidden information: Ben-Porath, Dekel, and Lipman (2014), Li (2017), Mylovanov and Zapechelnyuk (2017), Epitropou and Vohra (2019), and Kattwinkel and Knoepfle (2019). Chakravarty and Kaplan (2013) study an allocation with costly signaling.

Manelli and Vincent (2010) and Gershkov, Goeree, Kushnir, Moldovanu, and Shi (2013) analyze the equivalence between Bayesian optimal and ex-post incentive compatible implementation when monetary transfers are feasible. Ben-Porath, Dekel, and Lipman (2014) and Erlanson and Kleiner (2019) and Kattwinkel and Knoepfle (2019) show that with costly verification and without monetary transfers—similar to the results in this paper—the optimal mechanism is ex-post incentive compatible if the values are independent or the correlation is negative. The screening mechanism I derive shows that in a setting without transfers and with positively correlated information, this equivalence can fail.

Fieseler, Kittsteiner, and Moldovanu (2003) study the problem of efficient trade with interdependent values. They find that only under negative interdependent values can balanced transfers implement efficiency. Like in my setting, in this case, the expected conflict is small: small values of one agent occur with high values of the others. Therefore, they can be mediated with money. In my setting, under negative correlation, screening offers only little efficiency improvement and the principal abstains from screening.

The regularity assumption that I introduce is weaker than a similar assumption that Jewitt (1988) uses to justify the first-order approach in a moral hazard problem. The intuition for why this assumption guarantees interval allocation schedules argues with Diamond and Stiglitz's (1974) notion of utility preserving spreads. More generally, this intuition sheds light on a connection between belief heterogeneity and the heterogeneity in risk preferences.

To prove that the allocation schedules are ordered, I use that the utility functions of the players have single crossing-differences for ordered lotteries. The difference from Kartik, Lee, and Rappoport (2019)'s result is that I need this property to hold only for ordered lotteries. The result I use is a consequence of the variation diminishing property of totally positive kernels (Schoenberg (1930), Motzkin (1936), and Karlin (1968)).

1.2 Model

1.2.1 Setting

A principal (she) can allocate an indivisible good to an agent (he). Her allocation decision is denoted by $x \in \{0,1\}$. If she mixes, $x \in [0,1]$ denotes the probability of allocation. The allocation is costly for the principal. These costs are a random variable C taking values $c \in \mathbb{R}$. Of course, these costs can also be interpreted as opportunity costs. Then, C denotes the principal's valuation for the good. The realization of the costs is the principal's private information. The agent has a valuation for the good V, which takes values $v \in \mathbb{R}_{++}$. The realization of V is the agent's private information. Valuation V and cost C are jointly distributed according to a cdf $F(v,c)$. I assume that the supports of the valuation and the costs are both finite and denote them by $\mathcal{V}$ and $\mathcal{C}$, respectively. Let $f(v,c)$ denote the pdf. The assumption on the distribution of the costs is generalized in section 1.4.2.

The principal's objective is to maximize social welfare; her Bernoulli utility function is given by,

$$w(v,c,x) \; = \; x \cdot (v - c).$$

The agent does not bear the costs; his utility reads as follows:

$$u(v,c,x) \;=\; x\cdot v.$$

because $v > 0$, he always prefers receiving the good. The agent's outside option is zero so that he is always willing to participate in the mechanism.

The correlation between the agent's information V and the principal's information C is captured by the following likelihood ratios. For all $v' < v'' \in \mathscr{V}$:

$$\frac{f(v',c)}{f(v'',c)}. \tag{1.1}$$

I distinguish three cases.

(1) Negative affiliation: the likelihood ratios (1.1) are strictly increasing in c. The higher the costs, the more likely is a low valuation by the agent.

(2) Independence: the likelihood ratios (1.1) are constant in c. In this case, the costs are not informative about the valuation.

(3) Positive affiliation: the likelihood ratios (1.1) are strictly decreasing in c. The higher the costs, the more likely is a high valuation.

In the remainder I will use the term positive correlation for positive affiliation and negative correlation for negative affiliation.

1.2.2 Mechanism

The principal announces and commits to a mechanism before she learns the costs. The realization of the costs is contractible. The agent learns his valuation (but not the costs) and then plays a Bayesian best response.

Formally, a mechanism is given by a message set M and an allocation function $x : M \times \mathscr{C} \to [0,1]$ that specifies an allocation probability for any pair of message and cost realization. The revelation principle applies so that it suffices to have valuations as messages, $M = \mathscr{V}$ and to study mechanisms where the incentives are such that the agent reports his type truthfully.

Given a valuation report v, a direct mechanism specifies an allocation probability for all cost realization $c \in \mathscr{C}$. Denote this vector

$$x(v) \;=\; (x(v,c))_{c\in\mathscr{C}}$$

and term it as v's allocation schedule.

1.2.3 Agent's problem

The agent takes a mechanism x as given. He does not know the costs of the good, but forms expectations about it based on the realization of his valuation v. If he reports truthfully, he faces the random allocation lottery $x(v,C)$. If he reports $\hat{v}$, he faces a different lottery $x(\hat{v},C)$. Therefore, the Bayesian incentive constraints read as follows:

$$\forall v,\hat{v} \in \mathscr{V}: \qquad v \cdot \mathbb{E}[x(v,C) \mid V = v] \quad \geq \quad v \cdot \mathbb{E}[x(\hat{v},C) \mid V = v]. \tag{1.2}$$

Every type derives strictly positive utility from the good ($v > 0$ for all $v \in \mathscr{V}$); it follows that the intensity of type v's preferences can be eliminated from the incentive constraint: the agent maximizes his expected allocation probability.

$$IC(v,\hat{v}) = \sum_{c \in \mathscr{C}} f(v,c)\,[x(v,c) - x(\hat{v},c)] \quad \geq \quad 0. \tag{1.3}$$

This shows that the setting is equivalent to a setting where the valuation v does not enter the utility of the agent: he has utility 1 if he receives the good and 0 otherwise. In this setting, the valuation v is the value of the principal. Formally, in this equivalent setting, the principal's utility function stays the same and the agent's utility is given by

$$\overline{u}(x,v,c) = x.$$

Moreover, note that the expected allocation probability at a certain misreport is not independent of the true valuation type, as different valuation types have different conditional beliefs over the distribution of C. The interim expectations are therefore insufficient to describe the mechanism.

1.2.4 Principal's problem

The principal designs a mechanism that maximizes social welfare. If she could observe the valuation of the agent, she would allocate the good efficiently: that is, if and only if $v - c \geq 0$. However, because she only observes the costs, she must incentivize the agent to report his valuation. Her problem can be stated as the following linear program:

$$\max_{0 \leq x \leq 1} \mathbb{E}\,[\; x(V,C) \cdot (V - C) \;] \qquad \text{s.t. } \forall (v,\hat{v}) \in \mathscr{V} \times \mathscr{V}: \; IC(v,\hat{v}) \geq 0. \tag{1.4}$$

1.3 Optimal mechanism

1.3.1 If monetary transfers were feasible

If the principal could, in addition to the allocation decision $x \in [0,1]$, set a monetary transfer $t \in \mathbb{R}$ that would enter the agent's utility additively,

$$\hat{u}(v,c,x,t) \;=\; u(v,c,x) - t \;=\; x \cdot v - t,$$

the efficient allocation could be achieved: the principal can offer the good at price c to the agent and completely align the agent's interest in the mechanism with hers. To achieve efficiency, it is necessary that the agent's valuation and the principal's valuation for the good coincide. In (1.3), it is shown that without transfers, it is not relevant whether v enters the agent's utility. This is not the case if transfers are feasible. If the agent's utility is not affected by v and is just 1 if he receives the good and 0 otherwise, selling the good at price c does not induce the efficient allocation.

However, with transfers and correlation, there is another distinct way to achieve efficiency. This method exploits the differences of the beliefs that agents of different valuations v hold about the cost realization. It requires that the beliefs identify the corresponding valuations. Therefore, it must not be the case that there exists a type v' and $\lambda(v) \geq 0$ for all $v \in \mathscr{V} - \{v'\}$ such that,

$$f(c|v') \;=\; \sum_{v \in \mathscr{V} - \{v'\}} \lambda(v) \cdot f(c|v) \qquad \text{for all } c \;\in\; \mathscr{C}. \tag{1.5}$$

This rank condition[1] ensures that the belief that an agent holds is a sufficient statistic for the valuation v. If costs and valuation are independent, this condition is not fulfilled.

Theorem 1 (Cremer and McLean (1988))
If monetary transfers are feasible and the spanning condition is met, for any $x(v,c)$ there exists $t(v,c)$ implementing x with $\mathbb{E}[t(v,C)] = 0$ for all $v \in \mathscr{V}$.

The proof is an adaption of Cremer and McLean (1988)'s proof for the setting studied here: if the rank condition is met, the principal offers the agent a menu of lotteries whose payments depend on the realizations of the costs. For each valuation, there is one lottery. An agent with this valuation expects the corresponding lottery to pay out 0 and expects all other lotteries' payout to be lower than some bound $b < 0$. This bound b can be chosen uniformly for all valuations and can be set arbitrarily low. By setting b arbitrarily low, the mechanism gives the agent an arbitrarily high incentive to choose the lottery that corresponds to his valuation and thereby reveal it, irrespective of the allocation rule that the principal implements with this information. This mechanism also works if v does not enter into the agent's utility.

1. An equivalent formulation of this condition is that the set of beliefs for all types are the extreme points of their convex hull.

1.3.2 Independence and negative correlation

In the case of independence ((1.1), case (2)), the costs do not convey any information about the agent's value. In turn, agents of all valuations share the same belief about the costs. The principal cannot use the difference in the agent's belief to distinguish between them. As a result, optimally, there is no meaningful communication between the principal and the agent.

Ignorant mechanism. A mechanism that ignores the agent's report,

$$x(c,v') = x(c,v'') \text{ for all } v',v'' \in \mathscr{V},$$

is ex-post incentive compatible in the sense that an agent who would learn the cost realization before reporting to the mechanism is still incentivized to report truthfully. In practice, such a mechanism can be implemented by a procedure that does not ask the agent for his valuation at all. The optimal ignorant mechanism is given[2] by

$$x(c,v) = 1 \text{ if and only if } \mathbb{E}[V\,|\,C = c] \geq c.$$

More surprisingly, there is also no ground for communication when the correlation is negative.

Theorem 2
If costs and valuation are negatively correlated ((1.1), case (1)) or independent ((1.1), case (2)), then it is optimal for the principal to offer an ignorant mechanism. In these cases, the optimal (ignorant) mechanism is given by a simple cutoff rule:

$$\bar{c} = \min\{c \in \mathscr{C} \mid \mathbb{E}[V\,|\,C = c] \geq c\}.$$

and

$$x(c,v) = 1 \textit{ if and only if } c \leq \bar{c}.$$

The proof (in the appendix) introduces a relaxed problem: it disregards all incentive constraints for the agent to understate his valuation. For a fixed valuation report v' and two cost realizations $c' < c''$, there always exists a modification of the allocation schedule for v', $x(v') \rightarrow \tilde{x}(v')$, that decreases the allocation probability after the high cost realization c'', $x(v,c'') \downarrow$, and in turn increases the probability after the lower cost realization c', $x(v,c') \uparrow$, such that

$$\mathbb{E}[\tilde{x}(v',C)\,|\,V = v'] = \mathbb{E}[x(v',C)\,|\,V = v']$$

2. The optimal ignorant mechanism is unique except for the allocation probability after cost realizations c with $\mathbb{E}[V\,|\,C=c]=c$. I assume that in this case, the principal allocates the good.

and for all $v < v'$

$$\mathbb{E}[\tilde{x}(v', C) \,|\, V = v] \leq \mathbb{E}[x(v', C) \,|\, V = v].$$

The existence of this modification follows from the negative correlation (negative affiliation): an agent with lower valuation v puts relatively more likelihood on the realizations of high cost than on the realizations of low compared with an agent with a higher valuation v'. By shifting mass from high costs to low costs, one can keep the high valuation type indifferent while harming a lower type. The principal strictly prefers this modification because on average she has to bear lower costs while the overall probability of allocation to this valuation type remains constant.

Under negative correlation, the interest of the principal and the higher valuation type v' are aligned (put mass on low costs) and are distinct from the interest of a lower type v (put mass on high cost realization).

In the optimal mechanism, any allocation schedule must be in cutoff form. Otherwise it could be improved by a modification of the above form. Finally, the proof shows that these cutoff allocation schedule are optimally identical for all valuations. Hence, the optimal (Bayesian) mechanism is ignorant.

1.3.3 Positive correlation

Under positive correlation, the optimal mechanism offers different allocation schedules and the agents of different valuations sort themselves by their choice.

Regularity Assumption. To study the optimal mechanisms under positive correlation, I introduce a new regularity assumption on the joint distribution of cost and valuation. Under positive correlation (positive affiliation) the likelihood ratios (1.1) are strictly decreasing in c. The regularity assumption demands that these ratios are convex decreasing. The next definition formalizes the convexity of a function on a discrete set.

Definition 1. A function l on $\mathscr{C}$ is strictly convex if for all $c' < c''$ and for all $\alpha \in (0,1)$ with $\alpha \cdot c' + (1-\alpha) \cdot c'' \in \mathscr{C}$, $l(\alpha \cdot c' + (1-\alpha) \cdot c'') < \alpha \cdot l(c') + (1-\alpha) \cdot l(c'')$.

If the set is convex,this definition coincides with the usual definition of convexity. The regularity assumption reads as follows.

$$\text{For all } v' < v'' \in \mathscr{V}, \frac{f(v', c)}{f(v'', c)} \text{ is strictly convex in } c. \tag{1.6}$$

Jewitt (1988) introduces a similar condition in a moral hazard setting. He assumes that the increasing function that maps c to

$$\frac{f(v'', c)}{f(v', c)} = \frac{1}{\frac{f(v', c)}{f(v'', c)}}$$

is concave. These conditions are very similar; in general, they demand that the extent of the interference about the valuation decreases with increasing observed costs. However, the condition used here is weaker than the condition in Jewitt (1988).

Corollary 1
If a function $g : \mathbb{R} \to \mathbb{R}_{++}$ is increasing and concave, then the function $\frac{1}{g}$ is decreasing and convex. The reverse implication might not be true.

The proof is in the appendix. The regularity assumption is met by many distributions. I list some of them in section 1.4.2.

Next, the optimal mechanism is characterized.

Relaxation. Again, it is sufficient to consider a relaxed problem, where all incentive constraints that prevent the agent from understating his type are disregarded and only the upward incentive constraints,

$$IC(v, \hat{v}) \geq 0 \text{ for } v < \hat{v} \in \mathscr{V},$$

are considered.

1.3.3.1 Plateau mechanisms

The first characteristic of the optimal mechanism concerns the form of the allocation schedules that the principal offers to the agents with different values.

Definition 2 (Single plateau)**.** Let x be a mechanism and $v \in \mathscr{V}$ be a valuation. The allocation schedule $x(v)$ has a single plateau if for all $c' < c'' < c''' \in \mathscr{C}$ it holds,

$$x(v, c') > 0 \text{ and } x(v, c''') > 0 \Rightarrow x(v, c'') = 1.$$

A mechanism exhibits single plateaus if for all $v \in \mathscr{V}$, the allocation schedule $x(v)$ has a single plateau.

The support of an allocation schedule with a single plateau is always an interval in $\mathscr{C}$: it is given by $\mathscr{C} \cap [\underline{c}, \bar{c}]$ for some $\underline{c}, \bar{c} \in \mathscr{C}$. The allocation probability in the interior of this interval is always 1 and outside the interval is 0. Allocation probabilities different from $\{0, 1\}$ are only possible on the boundary of its support. In section 1.4.2, the case of continuously distributed costs is analyzed. In this case the mechanism is deterministic: the allocation probability is either 0 or 1.

The allocation schedules studied in section 1.3.2 allocate whenever the cost is below a cutoff. These cutoff schedules also have a single plateau. But the plateau is not interior. Intuitively, an allocation schedule with an interior single plateau allocates the good for intermediate costs with certainty and for extreme costs with zero probability.

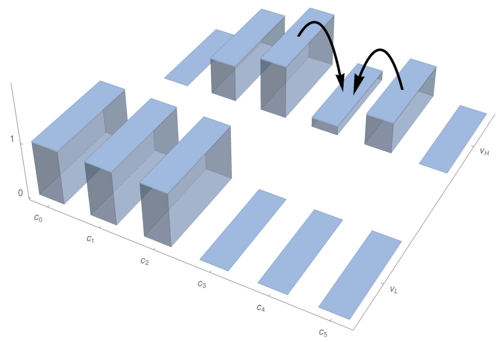


Figure 1.1. The allocation schedule for report v_H is not in single plateau form.

Proposition 1
If costs and valuations are positively correlated ((1.1), case (3)) and the distribution meets the regularity assumption (1.6), then the optimal solution to the relaxed problem has single plateaus.

If an allocation schedule has a single plateau, then the following modification is not feasible: for some triple $c' < c'' < c''' \in \mathscr{C}$, decrease $x(v, c') \downarrow$ and $x(v, c''') \downarrow$ while increasing $x(v, c'') \uparrow$. To show that the optimal mechanism has single plateaus, it sufficies to show that there always exists such a modification that respects the relaxed incentive constraints while increasing the principal's expected payoff. These shifts resemble the shifts of the payoffs that are used to study risk aversion (Diamond and Stiglitz (1974)). The formal proof of the Lemma uses a version of Farkas' Lemma and is relegated to the appendix.

Instead, I present a related problem which makes the connection to risk preferences transparent and illustrates how the regularity assumption leads to single plateau allocation schedules.

Intuition: a related problem with homogeneous beliefs. Suppose there are two valuations $\{v_L, v_H\}$ with $v_H - c \geq 0$ for all $c \in \mathscr{C}$. Let $f(v, c)$ be a twice differentiable interpolation of f for all $c \notin \mathscr{C}$. Before describing the related problem, I want to emphasize three important features of the setting that is studied in this paper. (i) Agents of different valuations hold different beliefs about the costs. (ii) The agent does not bear the costs of the allocation. (iii) The problem is equivalent to a setting where all the agents of different valuations have the same utility function.

The related problem contrasts the original setting in these three respects: (i) Agents of different valuations and the principal share a common belief about the costs. (ii) The agent bears the costs of the allocation. (iii) Agents with different valuations and the principal differ in their risk preferences with respect to the costs.

Suppose that the following related setting exists: costs are distributed according to a pdf g. Again, $g(c)$ denotes a twice differenteiable interpolation of g for all $c \notin \mathscr{C}$. There exist twice differentiable and increasing utility functions $\tilde{u}_L(-c), \tilde{u}_H(-c), \tilde{w}(-c)$ for the agent with low valuation v_L, the agent with high valuation v_H and the principal, respectively. $u_H(-c)$ describes the utility an agent with valuation v_H derives from receiving the good if the cost realization is c. The utilities are such that the preferences over allocation schedules for v_H coincide with the preferences in the original problem, formally:

$$\tilde{u}_H(-c) \cdot g(c) \; = \; f(v_H, c), \tag{1.7}$$

$$\tilde{u}_L(-c) \cdot g(c) \; = \; f(v_L, c), \tag{1.8}$$

$$\tilde{w}(-c) \cdot g(c) \; = \; f(v_H, c) \cdot (v_H - c). \tag{1.9}$$

It follows from this equation, that the optimal mechanism in the original setting and in the related problem coincide. Take any allocation schedule for v_H as given. For $c' < c'' < c''' \in \mathscr{C}$, consider the following modification of the mechanism:

$$\text{decreases } x(v_H, c') \downarrow \text{ and } x(v_H, c''') \downarrow, \text{ and increase } x(v_H, c'') \uparrow, \tag{1.10}$$

such that the utility for an agent with valuation v_H remains constant. When does such a modification simultaneously increase the principal's expected utility and decrease the expected utility of a low valuation agent?

Diamond and Stiglitz (1974) show that this is the case for all such modifications if a player with utility $\tilde{w}$ is more risk averse than a player with $\tilde{u}_H$ and a player with $\tilde{u}_H$ is more risk averse than a player with $\tilde{u}_L$, each in the sense of Arrow-Pratt:

$$\tilde{w} \; \geq_{AP} \; \tilde{u}_H \; \geq_{AP} \; \tilde{u}_L.$$

This is equivalent (see for example Jewitt (1989)) to,

$$\frac{\tilde{u}_H'(-c)}{\tilde{w}'(-c)} \text{ and } \frac{\tilde{u}_L'(-c)}{\tilde{u}_H'(-c)} \text{ are decreasing in } c. \tag{1.11}$$

In the appendix, I show that (1.7)–(1.9) and (1.11) can only hold together if

$$\frac{\partial^2}{\partial^2 c}\left(\frac{f(v_L, c)}{f(v_H, c)}\right) \; \geq \; 0.$$

In the related problem, the players' heterogeneity of beliefs is transformed into heterogeneity of their risk preferences. The regularity assumption ensures that these risk preferences can be ordered in a way that makes the modification of the form (1.10) simultaneously incentive compatible and worthwhile for the principal. In the optimal mechanism this modifications must not be feasible. Therefore, it must have single plateaus.

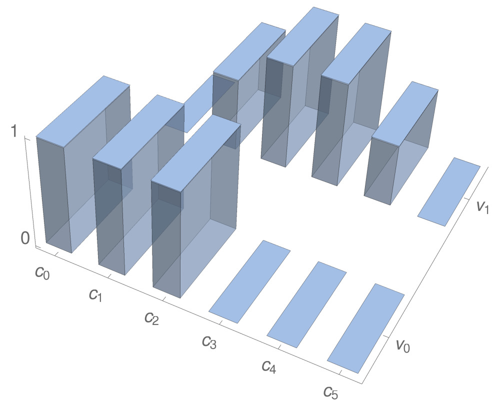


Figure 1.2. A plateau monotone mechanism.

1.3.3.2 Plateau-monotone mechanisms

Because $\mathscr{V}$ is ordered, one can number its elements.

$$\mathscr{V} = \{v_0, v_1, \dots, v_{m-1}\}.$$

To state the second characteristic of the optimal mechanism—a form of monotonicity—one needs to introduce a partial order on the set of allocation schedules $\{x(v) \,|\, v \in \mathscr{V}\}$.

Definition 3. Let x be a mechanism. Define the partial order $\succeq$ on $\{x(v) \,|\, v \in \mathscr{V}\}$ such that for $v', v'' \in \mathscr{V}$

$$x(v'') \succeq x(v') :\Longleftrightarrow \quad \forall c' < c'' \in \mathscr{C} : x(v'', c') \geq x(v', c') \Rightarrow x(v'', c'') \geq x(v', c''),$$
$$x(v', c'') \geq x(v'', c'') \Rightarrow x(v', c') \geq x(v'', c').$$

Or equivalently,

$$x(v'') \succeq x(v') :\Longleftrightarrow \quad c \mapsto x(v'', c) - x(v', c) \text{ is single-crossing from below.}$$

Define the strict order $x(v'') \succ x(v')$ if $x(v'') \succeq x(v')$ and $x(v'') \neq x(v')$.
A mechanism is monotone if

$$v'' > v' \Rightarrow x(v'') \succeq x(v')$$

or equivalently, a mechanism is monotone if $x(\cdot, \cdot)$ has single-crossing differences from below.

Proposition 2
If costs and valuations are positively correlated ((1.1), case (3)) and the distribution meets the regularity assumption (1.6), then the optimal mechanism in the relaxed problem is plateau-monotone.

Fix two valuations $v'' > v'$ and consider a plateau-monotone mechanism. Denote the upper and lower bounds of the support of the allocation schedule for v'' by $\bar{c}''$ and $\underline{c}''$, respectively. Denote the bounds for v' respectively as $\bar{c}'$ and $\underline{c}'$. Then,

$$\bar{c}'' \geq \bar{c}' \text{ and } \underline{c}'' \geq \underline{c}'.$$

The plateau of the higher valuation type is shifted to the right.

Proof sketch. The formal proof is in the appendix.
Step 1: It is established, that if x is an optimal mechanism in the relaxed problem, then the partial order $\succeq$ is total on $\{x(v) \,|\, v \in \mathcal{V}\}$. For any pair of $v', v'' \in \mathcal{V}$ either $x(v') \succeq x(v'')$ or $x(v'') \succeq x(v')$.
Step 2: The following result about the variation diminishing property of totally positive functions is used:

Lemma (Schoenberg (1930) and Karlin (1968))
If a real function function $K\colon \mathcal{V} \times \mathcal{C} \to \mathbb{R}$ is strictly totally positive of order 2 (STP_2) and there are allocation schedules: $x(\tilde{v}) \succ x(v)$, then for any $v' < v'' \in \mathcal{V}$ it holds

$$\sum_{c\in\mathcal{C}} K(v',c)\cdot(x(\tilde{v},c)-x(v,c)) \geq 0 \quad\Rightarrow\quad \sum_{c\in\mathcal{C}} K(v'',c)\cdot(x(\tilde{v},c)-x(v,c)) > 0,$$
$$\sum_{c\in\mathcal{C}} K(v'',c)\cdot(x(\tilde{v},c)-x(v,c)) \leq 0 \quad\Rightarrow\quad \sum_{c\in\mathcal{C}} K(v',c)\cdot(x(\tilde{v},c)-x(v,c)) < 0.$$

Or equivalently, $v' \mapsto \sum_{c\in\mathcal{C}} K(v',c)\cdot(x(\tilde{v},c)-x(v,c))$ crosses zero at most once and then from below.[3]

When the players compare two allocation schedules $x(\tilde{v}) \succeq x(v)$, they form expectations about the costs C and then evaluate the difference between these two schedules $x(\tilde{v},C)-x(v,C)$. The above variation diminishing results ensure that the ranking over the schedules is monotone in the valuation of the agents: setting $K(v,c) = f(v,c)$ yields, for example, that if an agent of valuation v' ranks $x(\tilde{v})$ over $x(v)$ then an agent with a higher valuation $v'' > v'$ ranks these allocation schedules in the same way.

An analogue monotone comparative static result follows for the principal. Again, suppose that $x(\tilde{v}) \succeq x(v)$. Setting $K(v,c) = f(v,c)\cdot(v-c)$ yields that if assigning the allocation $x(\tilde{v})$ to an agent with valuation v' yields her a higher expected payoff than $x(v)$ the same must hold true for an agent with a higher valuation $v'' > v'$.

3. A touching of zero is counted as a crossing.

A key step,is to establish that the assumptions of Karlin's lemma is met. The pdf $f(v,c)$ is strictly affiliated and strictly positive and therefore strictly totally positive (STP_2). Also, restricted on $\{(v,c) \in \mathcal{V} \times C \,|\, v > c\}$, the function $(v,c) \mapsto v - c$ is STP_2 and so is the product$f(v,c) \cdot (v-c)$. The proof in the appendix shows that all relevant comparisons between allocations schedules take place in this restricted set.
Step 3: Suppose for this proof sketch, that the only relevant incentive constraints are the local-upward incentive constraints of the form: $IC(v_{k-1}, v_k) \geq 0$.

If an optimal mechanism was not plateau-monotonic, there would be some k such that $x(v_{k-1}) \succ x(v_k)$. Consider the following modification: offer v_k the allocation schedule $x(v_{k-1})$. An agent with v_{k-1} ranks schedule $x(v_{k-1})$ over $x(v_k)$. Because $x(v_{k-1}) \succeq x(v_k)$, an agent with type v_k ranks them in the same way (step 2). The modification would therefore not violate any of the local-upward incentive constraints. Hence, this modification cannot be optimal for the principal:

$$\mathbb{E}[(v_k - C) \cdot x(v_{k-1}, C) \,|\, V = v_k] \leq E[(v_k - C) \cdot x(v_k, C) \,|\, V = v_k].$$

This implies (step 2) that

$$\mathbb{E}[(v_k - C) \cdot x(v_{k-1}, C) \,|\, V = v_{k-1}] < E[(v_k - C) \cdot x(v_k, C) \,|\, V = v_{k-1}].$$

However, then the mechanism can not have been optimal in the first place: modifying it such that it offers v_{k-1} the allocation schedule $x(v_k)$ would be a strict improvement for the principal and would not violate the incentive constraints. □

Theorem 3
If costs and valuations are positively correlated ((1.1), case (3)) and the distribution meets the regularity assumption (1.6), then the optimal mechanism has single plateaus and is plateau-monotone. It fulfills all local-upward incentive constraints with equality, for all $0 < k < m$*:*

$$\sum_{c \in \mathcal{C}} f(v_{k-1}, c) \cdot [x(v_{k-1}, c) - x(v_k, c)] = 0.$$

In the relaxed problem, the principal is not restrained by the incentive constraints that prevent the agent from misrepresenting his valuation as the lowest v_0. Therefore, it cannot be the case that the principal allocates the good to the lowest type at some cost c'' but does not allocate with certainty at some lower costs $c' < c''$. If this were the case, she could profitably shift the allocation probability from c'' to c', keeping the lowest type indifferent. This rules out any interior single plateau as the allocation schedule for the lowest valuation.

Corollary 2
Under the assumptions of Theorem 3, the allocation schedule for the lowest valuation v_0 *is always in cutoff form.*

$$x(v_0, c) = \begin{cases} 1, & c < \tilde{c} \\ 0, & c > \tilde{c} \end{cases}.$$

1.3.4 Monotonicity

Proposition 2 establishes that the optimal allocation schedules are plateau ordered. Of course, the ignorant cutoff mechanisms that are optimal under negative correlation are also ordered in this sense. However, in general, this monotonicity does not imply that a higher valuation type receives the good with higher probability. Again, there is a difference between positive and negative correlation. Under negative correlation, the expected allocation probabilities in the optimal ignorant mechanism are increasing in the agent's valuation. That is because, the function $x(c) = \begin{cases} 1, & c \leq \bar{c} \\ 0, & \text{otherwise} \end{cases}$ is decreasing on $\mathscr{C}$. By negative affiliation, $\mathbb{E}[x(C) \mid V = v]$ is therefore increasing in v. If the optimal mechanism under positive correlation is also ignorant, it must have the cutoff from. [4] However, in this case, the expected allocation probabilities are decreasing in the agent's valuation because under positive correlation $\mathbb{E}[x(C) \mid V = v]$ is decreasing in v.

example 4 shows that the interim expected allocation probabilities can be decreasing under positive affiliation, even if the optimal mechanism screens.

This difference with respect to standard results steams from the fact that the agents of different valuations hold different beliefs. For $v'' > v'$ by incentive compatibility,

$$\begin{aligned} \mathbb{E}[x(v'', C) \mid V = v''] &\geq \mathbb{E}[x(v', C) \mid V = v''] \\ \mathbb{E}[x(v', C) \mid V = v'] &\geq \mathbb{E}[x(v'', C) \mid V = v'], \end{aligned}$$

but because v' and v'' do not share the same belief, one cannot compare $\mathbb{E}[x(v'', C) \mid V = v'']$ and $\mathbb{E}[x(v', C) \mid V = v']$ with the above equations.

1.4 Extensions

1.4.1 Discrete costs

In section 1.2.1, I assumed that the support of the costs is finite. However, the proofs in the appendix allow the support to be countably infinite. Therefore, the case of many discrete cost distributions is covered.

Example 1. Suppose that valuations are finite: $\mathscr{V} = \{v_0, v_1, \ldots, v_{m-1}\}$.

(1) Suppose that conditional on the valuation $V = v_k$, the cost C is binomally distributed with equal number-of-trials parameter n and success probabilities p_k. Formally, $f(C = i | V = v_k) = \binom{n}{i} \cdot p_k^i \cdot (1 - p_k)^{n-i}$. The distribution is strictly positively affiliated if and only if $p_0 < p_1 < \cdots < p_{m-1}$. Then, for all $j < k$ the decreasing likelihood ratio, $\frac{f(C=i|V=v_j)}{f(C=i|V=v_k)}$ is also convex in i. Therefore, the regularity assumption (1.6) is fulfilled.

4. Otherwise, if the ignorant allocation schedule was interior, it could be shifted to lower costs, keeping the expected allocation value constant but decreasing the expected allocation costs

(2) Suppose that conditional on the valuation $V = v_k$, the cost C is geometrically distributed with failure probabilities p_k. Formally, $f(C = i|V = v_k) = (1 - p_k)^i \cdot p_k$. The joint distribution is strictly positively affiliated if and only if $p_0 > p_1 > \cdots > p_{m-1}$. Then, the regularity assumption is also fulfilled.

(3) Suppose that conditional on the valuation $V = v_k$, the cost C has a Poison distribution with parameter λ_k. Formally, $f(C = i|V = v_k) = \frac{\lambda_k^i e^{-\lambda_k}}{i!}$. The joint distribution is strictly positively affiliated if and only if $\lambda_0 < \lambda_1 < \cdots < \lambda_{m-1}$. Then, the regularity assumption is also fulfilled.

1.4.2 Continuous costs

Here, I assume that the costs are continuously distributed on an interval while the valuations remain finitely distributed. Formally, there is a measurable positive function $f(v, c)$ such that for all $B \subset \mathbb{R}$,

$$\mathbb{P}(V = v, C \in B) = \int_B f(v, c)\,\mathrm{d}c.$$

I assume that the cost distribution has the same interval as the support for all v, i.e. there exists an interval $\mathscr{C} \subset \mathbb{R}$ such that for all $v \in \mathscr{V}$:

$$\mathscr{C} = \{c \in \mathbb{R} | f(c, v) > 0\}.$$

Furthermore $c \mapsto f(v, c)$ is assumed to be continuous on $\mathscr{C}$ for all $v \in \mathscr{V}$.

An allocation schedule for a valuation report v is given by a measurable function $x(v) : \mathscr{C} \to [0, 1]$, and a mechanism specifies an allocation schedule for all valuations $v \in \mathscr{V}$. I distinguish the same three cases for the likelihood ratios with $v'' > v'$

$$\frac{f(c, v')}{f(c, v'')}.$$

Of course, manipulations on a set $B \subset \mathscr{C}$ of measure zero, neither affect the incentive constraints nor the principal's expected payoff. Mechanisms that differ only on a zero measure subset of $\mathscr{C}$ are equivalent. The regularity assumption (1.6) remains the same.

Theorem 4*(1) If costs and valuations are independent or negatively correlated, the optimal mechanism is equivalent to an ignorant cutoff mechanism of the form*

$$x(v, c) = \begin{cases} 1, & c \le c^* \\ 0, & c > c^* \end{cases},$$

for some $\bar{c} \in \mathscr{C}$.

(2) *If costs and valuations are positively correlated and the regularity assumption (1.6) is met, the optimal mechanism is equivalent to a plateau-monotone mechanism of the form*

$$x(v,c) = \begin{cases} 1, & \underline{c}(v) \le c \le \bar{c}(v) \\ 0, & \textit{otherwise} \end{cases},$$

with increasing functions $\underline{c}(v) \le \bar{c}(v)$.

If the costs are continuous, there is always an optimal mechanism that is deterministic.

Example 2. If the distribution of cost and valuation is of an exponential family, i.e., $f(c|v) = h(c) \cdot e^{\eta(v) \cdot T(c) - A(v)}$, then it is strictly positive affiliated if $T'(c) \cdot (\eta(v') - \eta(v'')) < 0$ for all c and $v' < v''$. The regularity assumption (1.6) is met if $(T'(c) \cdot (\eta(v') - \eta(v''))^2 > T''(c) \cdot (\eta(v') - \eta(v''))$ for all c and $v' < v''$. This yields that the regularity assumption is met if cost and valuation are jointly normally distributed and have positive correlation.[5] Also, if conditionally on $V = v_k$ the cost is exponentially distributed with parameter λ_k, then the distribution is positively affiliated if and only if $\lambda_0 > \lambda_1 > \cdots > \lambda_{m-1}$. In this case, the regularity assumption is also met.

Optimal ignorant mechanism. Under negative correlation or independence, the optimal ignorant mechanism was a simple cutoff rule. Despite the regularity assumption, the optimal ignorant mechanisms under positive affiliation can be quite irregular. For sets $C', C'' \subset \mathscr{C}$ define $C' \le C''$ if for all $c' \in C'$ and $c'' \in C''$: $c' \le c''$. An ignorant mechanism $x(v,c) = x(c)$ has a hole if there exist sets $C' \le C'' \le C''' \subset \mathscr{C}$ of positive measure such that $x(c) > 0$ for all $c \in \mathscr{C}' \cup C'''$ and $x(c) = 0$ for all $c \in C''$. The holes of a mechanism are counted in the following way:

$$\sup\{k \in \mathbb{N} \mid \exists C_1 \le B_1 \le C_2 \le B_2 \le \cdots \le C_{k+1}: \quad c \in C_i : x(c) = 0 \\ \& c \in B_i : x(c) > 0\}.$$

Lemma 1
For any $k \in \mathbb{N} \cup \{\infty\}$, there exists a positively correlated (item (3)) joint distribution that meets the regularity assumption such that the optimal ignorant mechanism $x(c)$ has k holes.

Under negative correlation, $\mathbb{E}[V|C = c]$ is decreasing in c. Therefore, $\mathbb{E}[V|C = c] - c$ is also decreasing, and the optimal ignorant mechanism is always in the cutoff form. Under positive correlation, $\mathbb{E}[V|C = c]$ is increasing and therefore $\mathbb{E}[V|C =$

5. Of course this paper does not cover this case, since I assume that valuation types are finite

$c] - c$ can cross zero infinitely many times. The proof (in the appendix) of Lemma 1 constructs such an example.

An optimal mechanism is in plateau form and therefore has no holes. The regularity assumption separately guarantees the plateau form of all single allocation schedules. However, the ignorant mechanism consists of one allocation schedule for all v. The separate regularity of the single optimal schedules does not aggregate to the regularity of the single optimal schedule for the ignorant mechanism. Finally if the optimal ignorant mechanism is irregular, it can be strictly improved, by an non-ignorant mechanism.

Corollary 3
If the optimal ignorant mechanism under positive affiliation has a hole, then the optimal mechanism is not ignorant.

1.4.3 Binary values

When $\mathscr{V} = \{v_L < v_H\}$ and the costs are continuously distributed on an interval $\mathscr{C} = [a, b)$ with $a \in \mathbb{R}$ and $b \in \mathbb{R} \cup \{\infty\}$, the optimal mechanism can be pinned down. With binary values, there is only one likelihood ratio of interest:

$$l(c) = \frac{f(v_L, c)}{f(v_H, c)}.$$

As a convex function, l is almost surely differentiable.

Proposition 3
When costs and valuations are positively affiliated and $[v_L, v_H] \subset \mathscr{C}$, then either:

(1) The optimal mechanism is ignorant and of a cutoff form:

$$x(v, c) = x(c) = \begin{cases} 1, & c \leq \bar{c} \\ 0, & \textit{otherwise} \end{cases}.$$

In this case, $\bar{c} \in \mathscr{C}$ is the unique solution to

$$c \in [v_L, v_H]: \quad \frac{f(v_L, c)}{f(v_H, c)} = \frac{v_H - c}{c - v_L}. \tag{1.12}$$

(2) The optimal mechanism is not ignorant and of the following form:

$$x(v_L, c) = \begin{cases} 1, & c \leq \bar{c}_L \\ 0, & \textit{otherwise} \end{cases}, \qquad x(v_H, c) = \begin{cases} 1, & \underline{c}_H \leq c \leq \bar{c}_H \\ 0, & \textit{otherwise} \end{cases}.$$

In this case, $a < \underline{c}_H \leq \bar{c}_H$ and $\bar{c}_L < \bar{c}_H$ are the unique solution to:

$$l(\underline{c}_H) = \frac{v_H - \underline{c}_H}{\bar{c}_L - v_L} \tag{1.13}$$

$$l(\bar{c}_H) = \frac{v_H - \bar{c}_H}{\bar{c}_L - v_L} \tag{1.14}$$

$$F(\bar{c}_L | V = v_L) = F(\bar{c}_H | V = v_L) - F_L(\underline{c}_H | V = v_L). \tag{1.15}$$

Remark 5. The system (1.12) always has a solution. If it is not unique then the optimal ignorant mechanism has holes and by corollary 3, the optimal mechanism is not ignorant. Then case 2 applies. Only if the solution is unique the optimal mechanism can be ignorant.

The system (1.13)–(1.15) has at most one solution. If it has no solution then the optimal mechanism is ignorant. If both (1.12) and (1.13)–(1.15) have a unique solution, then they are the two candidates for the optimal mechanism. It cannot be the case that simultaneously (1.12) is not unique and (1.13)–(1.15) do not have a solution.

When monetary transfers are feasible, any form of correlation can be exploited. Without transfers, negative correlation is not used to screen the agent. The next result shows that under positive correlation, the degree of correlation must exceed a certain threshold at the lowest cost.

Lemma 2
Suppose that l' is continuous at a. If

$$\frac{-l'(a)}{l(a)} < \frac{1}{v_H},$$

then the optimal mechanism is ignorant.

Example 3. If $f(c|V=v_\omega) = \lambda_\omega e^{-\lambda_\omega \cdot c}$ for $\omega \in \{L,H\}$ and $\lambda_L > \lambda_H$, the condition from Lemma 2 translates to $\lambda_L - \lambda_H < \frac{1}{v_H}$.

The characterizations of this section can be used to efficiently calculate numerical examples:

Example 4. $f(c, V=v_\omega) = \frac{1}{2} \cdot \lambda_\omega e^{-\lambda_\omega \cdot c}$ for $\omega \in \{L,H\}$ with $\lambda_L = 1, \lambda_H = 1/5$ and $v_H = 5, v_L = 0$. The optimal mechanism screens. It is given by $\bar{c}_L = 1.22$ and $(\underline{c}_H, \bar{c}_H) = (0.34, 4.88)$. The optimal ignorant mechanism has: $\bar{c} = 4.32$. It is worth noticing that the lower type has a higher expected allocation probability: $\mathbb{E}[x(v_L, C)\,|\,V=v_L] = .71 > \mathbb{E}[x(v_H, C)\,|\,V=v_H] = 0.56$.

1.4.4 General allocation values

The principal's value from the allocation can be generalized from $v - c$ to

$$w(x,v,c) \;=\; x \cdot z(v,c)$$

with z increasing in v and decreasing in c. Under negative correlation or independence the same results go through.

Theorem 6
If costs and valuations are negatively correlated ((1.1), case (1)) or independent ((1.1), case (2)), then it is optimal for the principal to offer an ignorant mechanism. In these cases, the optimal (ignorant) mechanism is given by a simple cutoff rule:

$$\bar{c} = \min\{c \in \mathscr{C} \mid \mathbb{E}[z(V,c)\,|\,C = c] \geq c\}.$$

and

$$x(c,v) = 1 \text{ if and only if } c \leq \bar{c}.$$

For the case of positive correlation, the regularity assumption generalizes to the following: for all $v' < v'' \in \mathscr{V}$ and all $c' < c'' < c''' \in \mathscr{C}$,

$$\frac{\frac{f(v',c'')}{f(v'',c')} - \frac{f(v',c')}{f(v'',c'')}}{\frac{f(v',c'')}{f(v'',c'')} - \frac{f(v',c''')}{f(v'',c''')}} \geq \frac{z(v'',c') - z(v'',c'')}{z(v'',c'') - z(v'',c''')}. \tag{1.16}$$

It demands that the decreasing likelihood ratio (on the left hand side) must be more convex than the allocation value of the principal (on the right hand side). This regularity assumption ensures the plateau form of the optimal mechanism in the relaxed problem. To establish that the optimal mechanism is plateau-monotone, it must be the case that z is log-supermodular on $\{(v,c)\,|\,z(v,c) > 0\}$, formally for all $v'' > v'$ and $c'' > c'$,

$$\text{if } z(v',c'') > 0: \qquad z(v'',c'') \cdot z(v',c') > z(v'',c') \cdot z(v',c'').$$

Theorem 7
Under positive correlation ((1.1), case (3)), if the generalized regularity assumption is fulfilled and z is log-supermodular where it is positive, then the optimal mechanism has single plateaus and is monotone. It fulfills all local-upward incentive constraints with equality.

1.5 Applications

1.5.1 Intra-firm allocation

A new computer model is about to enter the market. The management of a company has to decide whether it should buy the new computer. Before the market opens, the research department privately learns the value that the computer could generate for the company. The management wants to buy whenever the value exceeds the purchasing price; the research department wants to have the computer in any case. Before the competitive market price realizes, the management can communicate with the research department and commit to a demand schedule in the upcoming competitive market. Can the management use future market information to counter the internal information asymmetry?

Suppose that the correlation between market price and valuation stems from an unobservable quality factor $q \in [0,1]$ with strictly positive density $h(q)$. At $t = 0$, the quality realizes and the research department learns the value for the firm. The value is either low or high: $v \in \{v_L < v_H\}$. Let $f(v|q)$ denote the probability that valuation is $v \in \{v_L, v_H\}$ if the quality level is q. Suppose that $l(q) = \frac{f(v_L|q)}{f(v_H|q)}$ is twice differentiable and strictly decreasing. High quality indicates higher values. At $t = 1$, a competitive market for the computer opens. The management learns the rational expectation equilibrium price and can buy the computer at this price. The market is abstractly modeled as an equilibrium price function, which defines for any quality level q a price $p(q)$. Suppose that the price is strictly increasing in the quality. Defining the costs as $c = p(q)$ puts this in the positive correlation case. The regularity assumption then translates to $l(p^{-1}(p))$ being convex; this is the case if

$$-\frac{l''}{l'} > -\frac{p''}{p'}. \tag{1.17}$$

By assumption, $l' < 0$ and $p' > 0$. This condition is most easily fulfilled when the price is a convex increasing function of the quality or if the price is linear and l is convex. The price must react faster to an increase in quality than the value for the firm.

If the condition is met and the optimal mechanism involves screening, the management's demand schedule in the market is not monotone. Suppose the research department reports that the value of the computer is high. Then, the management will buy for intermediate prices and will not buy if the price is too low or too high.

1.5.2 Task delegation

The state wants to open a new hospital in a city. She has to decide if she wants to operate the hospital on her own or wants to delegate the operation to the city. Both the state and the city privately know what benefit for the society they would create as operators of the hospital. Denote the state's value by v_P and the city's value by v_A. These values are positively correlated: whenever the state can create a high value, it is more likely that the city could also create a high value. The state wants to delegate the operation to whoever creates the highest social value. The city wants to have control over the hospital in any case. If the state decides to delegate to the city, she forgoes the social value that she could create. This opportunity cost $c = -v_P$ represents her costs of allocation, whereas the social value that can the city can create is the value of the allocation: $v = v_A$. v and c are positively correlated, therefore the optimal mechanism might screen. If there is any screening then $x(\sup \mathcal{V}, \inf \mathcal{C}) < 1$. The state has to commit that she will not delegate when the city reports the highest possible value while she predicts the lowest possible value. Such a decision might be difficult to publicly defend. This advocates for an intransparent procedure, where the reported values are not part of the public record.

The next numerical example illustrates the effect of screening on the value of correlation.

Numerical example: value of correlation. Suppose that the state can build the hospital in one of two cities. After she builds the hospital, her and the city's operation value realize and she has to decide who operates it.

Both cities are with probability 1/2 good operators of the hospital and would create a social value of $v_H^1 = v_H^2 = 5$ and with probability 1/2 bad operators creating zero value. The states ability differs in the two cities. In the first city, her value is exponentially distributed with parameter $\lambda_H^1 = 1/4$ if the city itself is a good operator and with parameter $\lambda_L^1 = 1/2$ when the city is a bad operator. In the second city, the respective parameters are $\lambda_H^1 = 1/5$ and $\lambda_L^1 = 1$. The unconditional expectation of the state's operation value is the same in both cities. However, the degree of positive correlation is higher in city 2.

Without asymmetric information, the principal would prefer to build the hospital in city 1. If she could observe the realization of the city's operation value, she would implement the efficient allocation. This would yield her 2.146 for city 1 and 1.839 for city 2. With more positive correlation, if one of the two values is low, it is likelier that the other value is also low. In this sense, negative correlation serves as an insurance against low values.

The state's choice changes when she takes the asymmetric information into account. Now she prefers to build the hospital in city 2. The optimal mechanism in city 1 is ignorant. It allocates whenever $v_P > 2.376$. This yields the state 0.546. The extent of the positive correlation is not sufficient to screen out the city's type. In city 2, the positive correlation is sufficiently strong. The optimal mechanism screens the agent: When the city reports v_L^2, the state lets her operate the hospital whenever $v_P < 1.222$. when the city reports v_H^2, the principal lets her operate the hospital if $v_p \in (0.338261, 4.87646)$. This yields her $0.5887 > 0.546$.

1.5.3 Self-regulation

A firm seeks a regulator's approval for a new product. There are two aspects, $a_1 and a_2$, that are positively correlated and jointly determine the probability that the product is not faulty $p(a_1, a_2)$. The probability increases in both aspects. An example could be the approval of a new airplane model. If the regulator approves and the product turns out to not be faulty a social benefit $B > 0$ is created. If the regulator approves and the product is faulty, a social loss $L > 0$ occurs. The expected social benefit from approving reads as follows.

$$p(a_1, a_2) \cdot B - (1 - p(a_1, a_2)) \cdot L.$$

If the product is not approved, the effect on social welfare is independent of whether the product is faulty or not and normalized to zero.

It is a long-standing practice that regulators delegate parts of the certification process to the producers of the product. In the case of airplanes, the Search Results Web result with site links Federal Aviation Administration (FAA) fostered this development with their "Organizational Designation Authorization" program in 2005.[6] The extend of the delegation is substantial; the Transportation Department's Inspector General (2005) reported, "One aircraft manufacturer approved about 90 percent of the design decisions for all of its own aircraft."

The regulators argue that the delegation increases the efficiency of the process whilst ensuring its safety.[7]

To analyze this claim, suppose that the regulator delegates the certification of the first aspect to the producer and bases her decision on the reports from the producer and her own investigation of the second aspect. Setting

$$w(x, v, c) \;=\; x \cdot (p(v, -c) \cdot B - (1 - p(v, -c)) \cdot L)$$

as the principal's objective translates the positive correlation between the aspects, a_1 and a_2, into negative correlation between valuation $v = a_1$ and costs $c = -a_1$. The optimal mechanism therefore ignores the producer's report about the first aspect and bases the decision solely on the regulator's own findings. Delegation of an aspect implies that it is ignored it for the assessment of the safety of the product.

1.6 Conclusion

I study the bilateral trade setting with correlated information when monetary transfers are not feasible and characterize the welfare maximizing mechanism. This mechanism uses positive correlation to screen the agent; whereas under negative correlation, the optimal mechanism does not elicit the agent's valuation. My characterization of the optimal mechanism has interesting consequences for applications. Screening makes it necessary to forgo the highest gains from allocation. Also, the optimal mechanism may not allocate to higher valuation-types with higher probability. I introduce a novel regularity assumption that ensures an interval form of the optimal mechanism. The possibility of using positive correlation for screening opens interesting new directions of future research. My characterization is the first step toward analyzing a richer setting in which the principal endogenously chooses the degree of positive correlation.

6. see Federal Aviation Administration (2005).

7. See, for example, the testimony of the FAA's acting administrator Daniel K. Elwell (2019) in a senate hearing on the Boeing 737 Max crashes.

Appendix 1.A Proofs

All the proofs in this section are for the more general case of allocation values of the form $w(x, v, c) = x \cdot z(v, c)$ with

- $z(v, c)$ is increasing in v and decreasing in c.
- $z(v, c)$ is log-supermodular on $\{(v, c) \mid z(v, c) > 0\}$.

This generalization is introduced in section 1.4.4. The allocation value $z(v, c) = v - c$ is of this form, since

$$\partial_v \partial_c \log(v - c) = \frac{1}{(v-c)^2} \geq 0.$$

As $\mathscr{V}, \mathscr{C}$ are assumed to be discrete and ordered one can number its elements:

$$\mathscr{V} = \{v_0, v_1, \dots, v_{m-1}\}, \qquad \mathscr{C} = \{c_0, c_1, \dots, c_{n-1}\}.$$

The support of $\mathscr{C}$ can be countably infinite, then $n = \infty$.

Notation. For two vectors $a, b \in \mathbb{R}^n$ let $a \cdot b$ denote their standard inner product.[8] Let $a \circ b$ denote the vector of their element-wise products.

If there were monetary transfers

Independence and negative correlation

Proof of Theorem 2.
Take any incentive compatible mechanism $\boldsymbol{x}$. If it is not ignorant and in cutoff form I construct a modification of this mechanism that yields the principal a higher expected utility while keeping the incentive constraints satisfied.
step 0: (Relaxation)
In a relaxation of the problem the designer maximizes the same objective but disregards the incentive constraints for the agent not to report a lower type. The only incentive constraints which a solution of the relaxed problem has to respect are:

$$\forall v < \hat{v} \colon IC(v, \hat{v})$$

step 1: For all v there exists a cutoff $c(v)$ such that $x(v, c) = \begin{cases} 1, & c < c(v) \\ 0, & c > c(v) \end{cases}$.
Suppose there exists v'' and $c' < c''$ with $x(v'', c') < 1$ and $x(v'', c'') > 0$. Consider the following modification of the mechanism:

- allocate with probability $x(v'', c') + \mathrm{d}x(v'', c')$ after (v'', c')
- and with $x(v'', c'') + \mathrm{d}x(v'', c'')$ after (v'', c''),

8. If $n = 1$ than this denotes the standard product of two real numbers.

with

$$\mathrm{d}x(v'',c'') \;=\; -\frac{f(v'',c')}{f(v'',c'')}\cdot \mathrm{d}x(v'',c')$$

and $\mathrm{d}x(v'',c') > 0$ small enough such that: $x(v'',c') + \mathrm{d}x(v'',c') \le 1$ and $x(v'',c'') + \mathrm{d}x(v'',c') \ge 0$. It follows that

$$f(v'',c')\cdot \mathrm{d}x(v'',c') + f(v'',c'')\cdot \mathrm{d}x(v'',c'') \;=\; 0,$$

i.e. the probability of allocation for type v' (if he reports truthfully) remains the same in the modified mechanism. Since the allocation probability for all reports $v''' > v''$ were not modified, v'' has no new incentives to report a higher type. The principal's expected value increases:

$$f(v'',c')\cdot \mathrm{d}x(v'',c')\cdot(-c') + f(v'',c'')\cdot \mathrm{d}x(v'',c'')\cdot(-c'') \;=\; f(v'',c')\cdot\left(c''-c'\right) \;>\; 0.$$

Furthermore for any $v' \le v''$,

$$\begin{aligned} & f(v',c')\cdot \mathrm{d}x(v'',c') + f(v',c'')\cdot \mathrm{d}x(v'',c'') \\ =\; & \mathrm{d}x(v'',c')\cdot f(v',c'')\cdot\left(\frac{f(v',c')}{f(v',c'')} - \frac{f(v'',c')}{f(v'',c'')}\right) \;\le\; 0, \end{aligned}$$

i.e. for lower types misreporting their type as v'' becomes less attractive. Since the original mechanism was assumed to be incentive compatible in the relaxed problem, it follows that the modified mechanism remains compatible to the relaxed problem. We set $c(v)$ such that[9]

$$x(v,c) \;=\; \begin{cases} 1, & c \;<\; \lfloor c(v)\rfloor \\ c(v)-\lfloor c(v)\rfloor, & c \;=\; \lfloor c(v)\rfloor\,. \\ 0, & c \;>\; c(v) \end{cases}$$

step 2: There exists $x : \mathscr{C} \to [0,1]$ with $x(v,c) = x(c)$ for all v.
First, it can be ruled out that there exists $v' < v''$ such that $c(v') < c(v'')$ since this would violated $IC(v',v'')$. Suppose next, there exists $v' < v''$ such that $c(v') > c(v'')$. Define

$$v' \;=\; \sup\{v \in \mathscr{V} \,|\, \exists v'' > c : c(v) > c(v'')\}, \tag{1.A.1}$$

$$v'' \;=\; \inf\{v \in \mathscr{V} \,|\, v > v', c(v) < c(v')\}. \tag{1.A.2}$$

Suppose there was $v \in \mathscr{V}$ with $v' < v < v''$. From $v < v''$ it follows that $c(v) \ge c(v')$ and from $v > v'$ that $c(v) \le c(v'')$. Together this yields a contradiction to the assumption that $c(v') > c(v'')$. Therefore v' and v'' must be successors. This entails that all of

9. If $c(v) = 3.7$ the corresponding allocation is given by $x(v,\cdot) = (1,1,1,0.7,0,\cdots)$. If $c(v) = 3$ the allocation is $x(v,\cdot) = (1,1,0,0,\cdots)$

v' incentive constraints in the relaxed problem must slack: $c(v') > c(v'') \geq c(v''')$ for all $v''' \geq v''$. If the mechanism is optimal in the relaxed problem, it must be the case that $w(v', \lfloor c(v') \rfloor) \geq 0$. As $w(\cdot,\cdot)$ is increasing in the first component and decreasing in the second, it follows that $w(v'', c) \geq 0$ for all $c \geq \lfloor c(v) \rfloor$.
step 3: By step 2 we can assume that the optimal mechanism is ignorant. Since any ignorant mechanism is incentive compatible, the optimal ignorant mechanism is a solution to the relaxed and to the original problem. □

Positive correlation

Proof of corollary 1.
Suppose $g(c)$ is concave. For all $c' < c''$ and for all rational $\alpha \in (0,1) \cap \mathbb{Q}$ we have

$$\frac{1}{g(\alpha \cdot c' + (1-\alpha) \cdot c'')} \leq \frac{1}{\alpha \cdot g(c') + (1-\alpha) \cdot g(c'')}$$

Since α was assumed to .be rational there are natural numbers $j < n \in \mathbb{N}$ such that $\alpha = \frac{j}{n}$. But then it follows that

$$\frac{n}{j \cdot g(c') + (n-j) \cdot g(c'')} \leq \frac{j}{n} \cdot \frac{1}{g(c')} + \frac{n-j}{n} \cdot \frac{1}{g(c'')},$$

since the arithmetic mean exceeds the harmonic mean. For $\alpha \in (0,1)$ not rational, it can be expressed as limit of rational αs, and the result follows as a limit. Note that as g is assumed to be concave it must be continuous on all interior points of its support. Since $\alpha \in (0,1)$ the point $\alpha \cdot c' + (1-\alpha)c''$ is always in the interior of the support.
To see that the reverse might not be true consider the function $c \mapsto 2 - \sqrt{c}$ on $(0,1)$ it is strictly convex but $c \mapsto \frac{1}{2-\sqrt{c}}$ is not concave on $(0,1)$. □

The next result gives an equivalent characterization fo convexity for a decreasing function. It both applies when the function is defined on a interval and when the function is defined on a discrete set and convexity is discrete convexity. (See definition 1)

Corollary 4
A real decreasing function g defined on some set $I \subset \mathbb{R}$ is strictly convex iff $\forall c' < c'' < c''' \in I$:

$$(g(c') - g(c'')) \cdot (c'' - c') > (g(c'') - g(c''')) \cdot (c''' - c'').$$

Proof. For $\alpha = \frac{c'''-c''}{c'''-c'}$ it holds that $c'' = \alpha \cdot c' + (1-\alpha) \cdot c'''$. Plugging this into the definitions yields the result. □

This corollary shows that if $z(v,c) = v - c$ as in the main body, the generalized formulation of the regularity assumption (1.16) coincides with convexity of the liklihood ratios. Now, we are all set for the proof of Theorem 3. The proof proceeds in several steps. First I show that all the allocation schedules of the optimal mechanism in the relaxed problems have single plateaus.

Proof of Proposition 1.
Suppose the allocation schedule for v_j, $x(v_j)$, was not having a single plateau. This means that there exists $c' < c'' < c'''$ with $x(v_j,c') > 0, x(v_j,c'') = 0$ and $x(v_j,c''') > 0$. The remainder shows that under the regularity assumption there exists a modification of the mechanism, given by

$$\tilde{x}(v,c) \;=\; \begin{cases} x(v,c) + dx(v,c), & \text{if } v \;=\; v_j \text{ and } c \;\in\; \{c',c'',c'''\} \\ x(v,c), & \text{otherwise} \end{cases}, \tag{1.A.3}$$

with $dx(v_j,c') < 0, dx(v_j,c''') < 0$ and $dx(v_j,c''') > 0$. This proof uses a form of Farkas' Lemma (Farkas' Alternative) which I restate here:

Lemma (Farkas')
Suppose that $A \in \mathbb{R}^{k,l}, b \in \mathbb{R}^l$ then exactly one of these two alternatives is true:

(1) $\exists p \in \mathbb{R}^l$ with

- $Ap \geq 0.$
- $p \cdot b < 0.$
- $p > 0 :\Longleftrightarrow p \geq 0 \wedge p \neq 0.$

(2) $\exists y \in \mathbb{R}^k$ with

- $y'A \leq b.$
- $y \geq 0.$

A proof for this version of Farkas' Lemma can be found in Gyula Farkas' original paper, Farkas (1902). The above formulation is taken from Border (2013), Corollary 11. Setting

$$A = \begin{pmatrix} f(v_0,c') & -f(v_0,c'') & f(v_0,c''') \\ \vdots & -\vdots & \vdots \\ f(v_{j-1},c') & -f(v_{j-1},c'') & f(v_{j-1},c''') \\ -f(v_j,c') & f(v_j,c'') & -f(v_j,c''') \end{pmatrix} \in \mathbb{R}^{j+1\times 3}, \;\; b = \begin{pmatrix} z(v',c') \cdot f(v',c') \\ -z(v',c'') \cdot f(v',c'') \\ z(v',c''') \cdot f(v',c''') \end{pmatrix}$$

establishes equivalence between alternative 1 and the existence of a feasible, strictly profitable deviation of the form (1.A.3), which is then given by

$$\begin{pmatrix} dx(v_j,c') \\ dx(v_j,c') \\ dx(v_j,c') \end{pmatrix} := \begin{pmatrix} -p_1 \\ p_2 \\ -p_3 \end{pmatrix}.$$

The last entry of $Ap \geq 0$ guaranties that all incentive constraints, $IC(v_j, v'') \geq 0$ for $v' \geq v_j$ are fulfilled in the modified mechanism. The other entries ensure that $IC(v', v_j) \geq 0$ for all $v' \leq v_j$. No other incentive constraints in the relaxed are affected by the modification. Therefore the modified mechanism is also incentive compatible. Furthermore, $0 > b \cdot p =$ ensures strict profitability.

To conclude the proof one needs only to rule out alternative 2 under the regularity assumption. Suppose alternative 2 was true, then $\exists y \in \mathbb{R}_+^{j+1}$ such that $y'A \leq b$. This implies, that

$$y_j + z(v_j, c') \geq \sum_{i=0}^{j-1} \frac{f(v_i, c')}{f(v_j, c')} y_i \tag{1.A.4}$$

$$y_j + z(v_j, c'') \leq \sum_{i=0}^{j-1} \frac{f(v_i, c'')}{f(v_j, c'')} y_i \tag{1.A.5}$$

$$y_j + z(v_j, c''') \geq \sum_{i=0}^{j-1} \frac{f(v_i, c''')}{f(v_j, c''')} y_i \tag{1.A.6}$$

$(1.A.4) - (1.A.5)$ yields:

$$z(v_j, c') - z(v_j, c'') \geq \sum_{i=0}^{j-1} y_i \cdot \left(\frac{f(v_i, c'')}{f(v_j, c')} - \frac{f(v_i, c')}{f(v_j, c'')} \right) \tag{1.A.7}$$

$(1.A.6) - (1.A.5)$ yields:

$$z(v_j, c''') - z(v_j, c'') \geq \sum_{i=0}^{j-1} y_i \cdot \left(\frac{f(v_i, c''')}{f(v_j, c''')} - \frac{f(v_i, c'')}{f(v_j, c'')} \right) \tag{1.A.8}$$

$$\Leftrightarrow z(v_j, c'') - z(v_j, c''') \leq \sum_{i=0}^{j-1} y_i \cdot \left(\frac{f(v_i, c'')}{f(v_j, c'')} - \frac{f(v_i, c''')}{f(v_j, c''')} \right) \tag{1.A.9}$$

Note that by affiliation of the distribution and since $z(v_j, c)$ is decreasing c, all terms of the sum in $(1.A.7)$ and $(1.A.9)$ are positive.
Therefore $(1.A.7)$ / $(1.A.9)$ yields:

$$\frac{z(v_j, c') - z(v_j, c'')}{z(v_j, c'') - z(v_j, c''')} \geq \frac{\sum_{i=0}^{j-1} y_i \cdot \left(\frac{f(v_i, c'')}{f(v_j, c')} - \frac{f(v_i, c')}{f(v_j, c'')} \right)}{\sum_{i=0}^{j-1} y_i \cdot \left(\frac{f(v_i, c'')}{f(v_j, c'')} - \frac{f(v_i, c''')}{f(v_j, c''')} \right)} \tag{1.A.10}$$

$$\Leftrightarrow \sum_{i=0}^{j-1} y_i \cdot \left(z(v_j, c') - z(v_j, c'') \right) \cdot \left(\frac{f(v_i, c'')}{f(v_j, c'')} - \frac{f(v_i, c''')}{f(v_j, c''')} \right) \tag{1.A.11}$$

$$\geq \sum_{i=0}^{j-1} y_i \cdot \left(z(v_j, c'') - z(v_j, c''') \right) \cdot \left(\frac{f(v_i, c'')}{f(v_j, c')} - \frac{f(v_i, c')}{f(v_j, c'')} \right) \tag{1.A.12}$$

But by the regularity assumption and since $y_i \geq 0$ it holds $\forall i \in \{1,\ldots,j-1\}$:

$$y_i \cdot \left(z(v_j,c') - z(v_j,c'')\right) \cdot \left(\frac{f(v_i,c'')}{f(v_j,c'')} - \frac{f(v_i,c''')}{f(v_j,c''')}\right) \tag{1.A.13}$$

$$\leq\; y_i \cdot \left(z(v_j,c'') - z(v_j,c''')\right) \cdot \left(\frac{f(v_i,c'')}{f(v_j,c')} - \frac{f(v_i,c')}{f(v_j,c'')}\right) \tag{1.A.14}$$

Finally suppose that $\forall i \in \{1,\ldots j-1\} : y_i = 0$. Then (1.*A*.9) would imply that, $z(v_j,c'') - z(v_j,c''') \leq 0$. Contradiction since $z(v_j,c)$ is assumed to be strictly decreasing!
So we can assume that the above equation (1.*A*.14) holds strictly for at most one $i \in \{1,\ldots j-1\}$ contradicting therefore (1.*A*.12). Thus, alternative 2 can be ruled out.

□

To proof monotonicity the following Lemma is used:

Lemma 3. *Suppose plateau allocation rules $x'',x' : \mathscr{C} \to [0,1]$ are strictly ordered: $x'' \succ x'$. If there is $v \in \mathscr{V}$ with*

$$f(v) \cdot (x'' - x') \;=\; 0$$

then it must hold that

$$f(v) \circ w(v,c) \cdot (x'' - x') \;<\; 0$$

Proof. Define

$$C'' \;=\; \{c \;\in\; \mathscr{C}\,|\,x''(c) - x'(c) \;>\; 0\}, \qquad C' \;=\; \{c \;\in\; \mathscr{C}\,|\,x''(c) - x'(c) \;<\; 0.$$

As the allocation rules are in plateau form an ordered there exists $c^* \in \mathscr{C}$ with

$$\forall c'' \;\in\; C''\,\forall c' \;\in\; \mathbf{C}' : c' \;\leq\; c^* \;\leq\; c''.$$

As $w(v,\cdot)$ is strictly increasing it follows that

$$\sum_{cinC''} f(v,c) \cdot w(v,c) \cdot (x''(c) - x'(c)) \;\leq\; w(v,c^*) \cdot \left(\sum_{c \in C''} f(v,c) \cdot (x''(c) - x'(c))\right)$$

$$=\; w(v,c^*) \cdot \left(\sum_{c \in C'} f(v,c) \cdot (x''(c) - x'(c))\right) \;\leq\; \sum_{c \in C'} f(v,c) \cdot w(v,c) \cdot (x''(c) - x'(c)).$$

As $x'' \succ x'$ at least one of the two inequalities must be strict. □

I also use the following result about totally positive kernels and adapt its formulation to my setting:

Lemma 4 (Schoenberg (1930) and Karlin (1968))**.** If a real function function $K\colon \mathscr{V} \times \mathscr{C} \to \mathbb{R}$ is strictly totally positive of order 2 (STP_2) and there are allocation schedules: $x(\tilde{v}) \succ x(v)$ then for any $v' < v'' \in \mathscr{V}$ it holds

$$\sum_{c\in\mathscr{C}} K(v',c)\cdot(x(\tilde{v},c)-x(v,c)) \;\geq\; 0 \quad\Rightarrow\quad \sum_{c\in\mathscr{C}} K(v'',c)\cdot(x(\tilde{v},c)-x(v,c)) \;>\; 0,$$
$$\sum_{c\in\mathscr{C}} K(v'',c)\cdot(x(\tilde{v},c)-x(v,c)) \;\leq\; 0 \quad\Rightarrow\quad \sum_{c\in\mathscr{C}} K(v',c)\cdot(x(\tilde{v},c)-x(v,c)) \;<\; 0.$$

Or equivalently, $v' \mapsto \sum_{c\in\mathscr{C}} K(v',c)\cdot(x(\tilde{v},c)-x(v,c))$ crosses zero at most once and then from below.[10]

A proof can be found in Karlin (1968). Setting $K(v',c) = f(v',c)$ or $K(v',c' = f(v',c)\cdot z(v,c)$

Proof of Proposition 2.
Let x be a solution to the relaxed problem. With *Proposition* 1 we can assume that the mechanism is in plateau form. Let

$$k \;=\; \min\{0 \;<\; i \;<\; m \,|\, x(v_i) \;\succeq\; x(v_{i-1})\}$$

If $k = 1$, the mechanism is plateau monotone. Otherwise we have $\neg[x(v_k) \succeq x(v_{k-1})]$. This means

$$\neg[\underline{c}(v_k) \;\geq\; \underline{c}(v_{k-1}) \text{ and } \bar{c}(v_k) \;\geq\; \bar{c}(v_{k-1})].$$

One has to to distinguish three cases:

(1) $\underline{c}(v_k) \leq \underline{c}(v_{k-1})$ and $\bar{c}(v_k) \geq \bar{c}(v_{k-1})$ with at least one inequality strict. This case can be ruled out immediately, since if this were the case, then

$$f(v_{k-1})\cdot x(v_{k-1}) \;<\; f(v_{k-1})\cdot x(v_k),$$

which would be a violation of $IC(v_{k-1}, v_k)$.

(2) $\underline{c}(v_k) > \underline{c}(v_{k-1})$ and $\bar{c}(v_k) \leq \bar{c}(v_{k-1})$ with at least one inequality strict.
For all $j < k$ it holds,

$$f(v_j)\cdot x(v_j) \;\geq\; f(v_j)\cdot x(_{k-1}) \;>\; f(v_j)\cdot x(v_k).$$

Since $j \leq k-1$, the first inequality follows from incentive compatibility. The second strict inequality follows from the initial assumption about x_k and x_{k-1}. As a consequence, all incentive constraints $IC(v_j, v_{,k})$ for $j < k$ slack. This could have only been optimal if for all $c \leq \lfloor \bar{c}_{k-1} \rfloor$ it was the case that $w(v_k, c) \leq 0$. But then

10. A touching of zero is counted as a crossing.

as w is strictly increasing in v it follows that $w(v_{k-1}, v) < 0$ for all $c \leq \lfloor \bar{c}_{k-1} \rfloor$. But this is a contradiction since, for all $l \geq k$ it holds:

$$f(v_{k-1}) \cdot x(v_k - 1) \; > \; f(v_{k-1}) \cdot x(v_k) \; \geq \; f(v_{k-1}) \cdot x(v_l).$$

The first inequality is again by the initial assumption and the second follows since $x_l \succeq x_k$. This means that all incentive constraints $IC(v_{k-1}, v_l)$ for $l > k-1$ slack. So the choice of x_{k-1} could not have been optimal in the first place.

(3) $\underline{c}(v_k) \leq \underline{c}(v_{k-1})$ and $\bar{c}(v_k) \geq \bar{c}(v_{k-1})$ with at least one inequality strict.
step 1: *Define the upper end of the support of* $x(v_{k-1})$ *as* $\bar{c}$, *i.e.* $\bar{c} = \lceil \bar{c}(v_{k-1}) - 1 \rceil$. *In this step I show that:* $w(v_{k-1}, \bar{c}) > 0$.

First suppose that $w(v_{k-1},) \leq 0$. If

$$f(v_{k-1}) \cdot x(v_{k-1}) \; = \; f(v_{k-1}) \cdot x(v_k)$$

then allocating the good after v_{k-1} according to allocation rule $x(v_k)$ instead of $x(v_{k-1})$ does not generate new profitable deviations. But since $x(v_k) \succ x(v_{k-1}$ it would strictly improve the principals expected utility (Lemma 3). We can therefore assume that $IC(v_{k-1}, v_k) > 0$. But since for all $l \geq k$

$$f(v_k) \cdot x(v_k) \; \geq \; f(v_k) \cdot x(v_l) \; \Rightarrow \; f(v_{k-1}) \cdot x(v_k) \; \geq \; f(v_{k-1}) \cdot x(v_l),$$

it follows that $IC(v_{k-1}, v_j) > 0$. Since there are only finitely many types there is $\varepsilon > 0$ such that $IC(v_{k-1}, v_l) \geq \varepsilon$,i.e. v_{k-1} incentive constraints slack uniformly. This result directly in a contradiction if $w(v_{k-1}, \bar{c}) < 0$ since then $x(v_{k-1}, \lceil \bar{c}(v_{k-1}) - 1 \rceil)$ could be lowered to increase the principals expected payoff. If $w(v_{k-1}, \bar{c}) = 0$ then $x(v_{k-1}, \bar{c})$ could still be lowered until either $IC(v_{k-1}, v_k)$ binds or if the incentives still slack with $x(v_{k-1}, \bar{c}) = 0$ one could proceed and lower $x(v, \bar{c}-)$. In the former case, one could now improve the principals payoff by allocating the good after v_{k-1} according to allocation rule $x(v_k)$ instead of the modified $x(v_{k-1})$.

step 2: Consider the following modification of the mechanism: after v_k allocate according to $x(v_{k-1})$ instead of $x(v_k)$. This change does not introduce new profitable deviations since,

$$f(v_{k-1}) \cdot x(v_{k-1}) \; \geq \; f(v_{k-1}) \cdot x(v_k) \; \Rightarrow \; f(v_k) \cdot x(v_{k-1}) \; \geq \; f(v_k) \cdot x(v_k).$$

Therefore, this modification cannot yield a higher expected value for the principal:

$$f(v_k) \circ w(v_k) \cdot x(v_{k-1}) \; \leq \; f(v_k) \circ w(v_k) \cdot x(v_k)$$

It follows by Lemma 4 and since $w(v_{k-1}, \bar{c}) > 0$ that

$$f(v_{k-1}) \circ z(v_{k-1}) \cdot x(v_{k-1}) \; < \; f(v_{k-1}) \circ z(v_{k-1}) \cdot x(v_k)$$

For any $\alpha \in (0,1)$ it would be strictly profitable to modify the mechanism in the following way: after v_{k-1} allocate according to $\alpha x(v_k) + (1-\alpha)x(v_{k-1})$ instead of $x(v_{k-1})$. It must therefore be the case that any such modification would violate the incentive constraints. This can only be the case if there is a binding incentive constraint for v_{k-1}, i.e. there exists $l \geq k$ such that

$$f(v_{k-1}) \cdot x(v_{k-1}) \;=\; f(v_{k-1}) \cdot x(v_l).$$

and $x(v_l) \succ x(v_k)$. Otherwise — if $x(v_l) = x(v_k)$ — there would be no incentive violation in the modified mechanism.

As $IC(v_{k-1}, v_k) \geq 0$, it follows that

$$f(v_{k-1}) \cdot x(v_l) \;\geq\; f(v_{k-1}) \cdot x(v_k).$$

But since $x(v_l) \succ x(v_k)$ it follows that $l > k$ and

$$f(v_k) \cdot x(v_l) \;>\; f(v_k) \cdot x(v_k),$$

a contradiction to $IC(v_k, v_l) \geq 0$.

□

Lemma 5. In the optimal solution to the relaxed problem all local incentive constraints are binding. That is for all $0 < k < m$ $IC(v_{k-1}, v_k) = 0$, i.e.

$$f(v_{k-1}) \cdot x(v_{k-1}) \;=\; f(v_{k-1}) \cdot x(v_k - 1).$$

Proof. Suppose there was some k with $IC(v_{k-1}, v_k) > 0$. It cannot be the case that $x_k = x_{k-1}$. Therefore $x_k \succ x_{k-1}$. By Lemma 4 it follows for all $j < k$ that

$$f(v_j) \cdot x(v_{k-1}) \;>\; f(v_j) \cdot x(v_k)$$

For all $j < k$ the relaxed incentive compatibility implies

$$f(v_j) \cdot x(v_j) \;\geq\; f(v_j) \cdot x(v_k),$$

it follows that for all $j < k$ the incentive constraints slack, $IC(v_j, v_k) > 0$. As $x(v_k) \succ x(v_j)$ and $f(v_{k-1}) \cdot x(v_{k-1}) > f(v_{k-1}) \cdot x(v_k)$, it must be the case that $x(v_k, c_0) < 1$. Choose $u = \max\{i : 0 < i < n, x(v_k, c_i) > 0\}$. Since all incentive constraints slack one can freely shift mass from $x(v_k, c_u) \downarrow$ to $x(v_k, c_0) \uparrow$ at a rate that keeps v_k's expected allocation probability constant,

$$dx(v_k, c_0) \cdot f(v_k, c_0) \;=\; -dx(v_k, c_u) \cdot f(v_k, c_u),$$

by Lemma 3 this modification would strictly improve the principal's expected utility. Contradiction.

□

Lemma 6. The plateau-monotonic solution to the relaxed problem is a solution to the original problem.

Proof. Let x be the plateau-monotonic solution to the relaxed problem. Suppose that x was not a solution to the original problem. This means that a high type has incentive to misrepresent himself as a lower type. Set

$$k = \min\{l \in \mathbb{N} \,|\, 0 < l < m, \exists j < l \text{ with } f(v_l) \cdot x(v_l) < f(v_l) \cdot x(v_j)\}.$$

It follows that $x(v_k) \succ x(v_j)$. By Lemma 5 it holds that,

$$f(v_{k-1}) \cdot x(v_k) = x(v_{k-1}).$$

By Lemma 4 an since $x(v_k) \succeq x(v_{k-1})$ (monotonicity) it follows that

$$f(v_{k-1}) \cdot x(v_k) = x(v_{k-1}).$$

One can deduce that $k-1 > j$. Therefore, since x is monotone: $x(v_{k-1}) \succeq x(v_j)$.
k was chosen minimally, therefore:

$$f(v_{k-1}) \cdot x(v_{k-1}) \geq f(v_{k-1}) \cdot x(v_j).$$

But then again by Lemma 4 it follows that,

$$f(v_k) \cdot x(v_{k-1}) \geq f(v_k) \cdot x(v_j).$$

Contradicting, the initial assumption that $IC(V_K, v_j) < 0$. □

Taking all these results together proofs Theorem 3.

Continuous Costs

Lemma 7 (Lemma 1). For any $k \in \mathbb{N} \cup \{\infty\}$ there exists a positively affiliated joint distribution that fulfills the regularity assumption (R) such that the optimal ignorant mechanism $x(c)$ has k holes.

Proof. Suppose that there are only two valuations $v_L < v_H$. The optimal ignorant mechanism reads:

$$x(c) = \begin{cases} 1, & \mathbb{E}[V \,|\, C = c] > c \\ 0, & \mathbb{E}[V \,|\, C = c] < c \end{cases}.$$

For $c \in (v_L, v_H)$:

$$\mathbb{E}[V \,|\, C = c] > c \Leftrightarrow \frac{f(v_L, c)}{f(v_H, c)} < \frac{v_H - c}{c - v_L}.$$

Setting $v_H = 3, v_L = 0$ and

$$f(v_H, c) = \begin{cases} K, & c \in (1,2) \\ 0, & \text{otherwise} \end{cases},$$

$$f(v_L, c) = \begin{cases} K \cdot \left(\frac{v_H - c}{c - v_L} + \varepsilon \cdot \sin\left(\frac{1}{c-1}\right)(c-1)^4\right), & c \in (1,2) \\ 0, & \text{otherwise} \end{cases}$$

$\frac{v_H - c}{c - v_L}$ is a strictly convex decreasing function. For $v_H = 3, v_L = 0$ $\frac{v_H - c}{c - v_L}$, $-\left(\frac{v_H - c}{c - v_L}\right)'$, and $\left(\frac{v_H - c}{c - v_L}\right)''$ are all bounded from below by a strictly positive bound on the whole support $(1,2)$. Since $|\sin\left(\frac{1}{c-1}\right)(c-1)^4|$, $|\left(\sin\left(\frac{1}{c-1}\right)(c-1)^4\right)'|$ and $|\left(\sin\left(\frac{1}{c-1}\right)(c-1)^4\right)''|$ are bounded from above on $(1,2)$ there exists $\varepsilon > 0$ such that for all $K > 0$ $f(v_L, c) > 0$ and $\frac{f(v_L,c)}{f(v_H,c)}$ is decreasing convex. Since $\int_1^2 \sin\left(\frac{1}{c-1}\right)(c-1)^4 \, \mathrm{d}c$ exists and is bounded $K > 0$ can be chosen such that $f(v,c)$ is a density. □

Proof of Theorem 4.
The proof essentially replicates all the arguments from the finite realization case. First we show that it is in plateau form, which corresponds to intervals in this new setting:

Let λ denote the Lebesgue measure. Let C', C'', C''' be measurable subset of $\mathscr{C}$ with positive measure that are ordered in the following sense: $\forall (c', c'', c''') \in C' \times C'' \times C'''$: $c' \leq c'' \leq c'''$. The sets are assumed to be distinct on a positive measure: $\lambda(C' \Delta C'') > 0, \lambda(C'' \Delta C''') > 0$ and $\lambda(C' \Delta C''') > 0$, where Δ denotes the symmetric differences of sets.

Suppose that x is an optimal mechanism and there exists $v' \in \mathscr{V}$ with:

$$\int_{C'} x(v,c) \, \mathrm{d}c > 0, \int_{C''} (1 - x(v,c)) \, \mathrm{d}c > 0, \int_{C'''} x(v,c) \, \mathrm{d}c > 0$$

It is without loss to assume that $x(v', c) > 0$ for all $c \in C' \cup C'''$ and $x(v', c) < 1$ for all $c \in C''$ and that $c' < c'' < c'''$. Otherwise use $C' - C'' - C''' - \{c \in \mathscr{C} \,|\, x(v', c) = 0\}$ instead of C' and corresponding subsets of C'' and C'''.

Since for any measurable $C \subset \mathscr{C}$ the function $b \mapsto \lambda(C \cap [-b, b])$, is continuous, there exists $\tilde{C}' \subset C'$, $\tilde{C}'' \subset C''$, and $\tilde{C}''' \subset C'''$ with $\lambda(\tilde{C}') = \lambda(\tilde{C}'') = \lambda(\tilde{C}''') > 0$. There exist measure preserving bijections $\phi \colon \tilde{C}' \to \tilde{C}''$ and $\psi \colon \tilde{C}' \to \tilde{C}''$ (see for example Alós-Ferrer (1999) , Lemma 3). For all $c' \in \tilde{C}'$ we can apply **res:..** to $c' < \phi(c') < \psi(c'')$. This gives for any c' that the set of feasible, incentive compatible and strictly profitable deviations $dx(v', c')$, $d(v', c'')$, $dx(v', c'')$ is not empty. One needs to argue why this correspondence exhibits a measurable selection. The usual measurable selection theorems are formulated for closed correspondences. But since the deviations here must be strictly profitable the definition of the correspondence includes a strict inequality. To get around this, consider the correspondence that maps c to the set of deviations which are feasible, incentive compatible and profitable or neutral for the principal. This does not involve strict inequalities therefore it is compact valued. It is also weakly measurable, denote it by Γ.

From above we know that the most profitable deviation in $\Gamma(c')$ is a strictly profitable deviation. Also, the function which evaluates the profitability of a deviation $(c', x_1, x_2, x_3) \mapsto f(v', c') \cdot x_1 + f(v', \phi(c')) \cdot x_2 + f(v', \psi(c')) \cdot x_3$ is measurable in c' and continuous in (x_1, x_2, x_3). Hence, it is a Caratheodory function.

With this we can apply a measurable maximum theorem (Aliprantis and Border (2006), Theorem 18.19) and get a measurable selection of the most profitable deviation, which is strictly profitable. Denote it by $\gamma(c') = (\gamma_1(c'), \gamma_2(c'), \gamma_3(c'))$. With this we can construct a new measurable mechanism $\tilde{x}$, which has $\tilde{x}(v', c') = x(v', c) + \gamma_1(c')$, $x(v', \phi(c')) = x(v', \phi(c')) + \gamma_2(c')$, and $x(v', \phi(c')) = x(v', \phi(c')) + \gamma_3(c')$ for all $c' \in \tilde{C}'$ and equals x elsewhere. This new mechanism has a strictly higher payoff for the principal on a set with positive measure. Therefore the original mechanism can not have been optimal.

It follows that the optimal mechanism to the relaxed problem can be characterized by two functions: $\underline{c}, \bar{c} : \mathscr{V} \to \mathscr{C}$ with $\forall v \in \mathscr{V}$: $\underline{c}(v) \leq \bar{c}(v)$. Define an analogous partial order on the set of allocation schedules in plateau form:

$$x(v'') \succeq x(v') \iff \underline{c}(v'') \geq \underline{c}(v') \text{ and } \bar{c}(v'') \geq \bar{c}(v').$$

The variation diminishing property also holds for kernels that are continuous in c:

Lemma (Schoenberg (1930) and Karlin (1968))
If a real function function $K : \mathscr{V} \times \mathscr{C} \to \mathbb{R}$ is strictly totally positive of order 2 (STP_2) and there are allocation schedules: $x(\tilde{v}) \succ x(v)$ then for any $v' < v'' \in \mathscr{V}$ it holds

$$\int_{\mathscr{C}} K(v', c) \cdot (x(\tilde{v}, c) - x(v, c))\, \mathrm{d}c \geq 0 \quad \Rightarrow \quad \int_{\mathscr{C}} K(v'', c) \cdot (x(\tilde{v}, c) - x(v, c))\, \mathrm{d}c > 0,$$

$$\int_{\mathscr{C}} K(v'', c) \cdot (x(\tilde{v}, c) - x(v, c))\, \mathrm{d}c \leq 0 \quad \Rightarrow \quad \int_{\mathscr{C}} K(v', c) \cdot (x(\tilde{v}, c) - x(v, c))\, \mathrm{d}c < 0.$$

The cost-continuous analog of Lemma 3 directly follows, if in its proof the sums are replaced by integrals.

With these results in hand, all the other steps of the proof with finite realizations can be replicated analogously. □

Binary valuation

Proof of Proposition 3.
First, since $v_H \in (a, b)$, there always exists $d > a$ such that the ignorant mechanism $x(c) = \begin{cases} 1, & c \in [a, d) \\ 0, & c \in [d, b) \end{cases}$ dominates the mechanism that never allocates. Similarly, the ignorant mechanism $x(c) = \begin{cases} 1, & c \in [a, v_H) \\ 0, & c \in [v_H, b) \end{cases}$ dominates the mechanism that always allocates.

Suppose first, that the optimal mechanism x screens, i.e. $\bar{c}_L < \bar{c}_H$ and $\underline{c}_H > a$. I consider two feasible infinitesimal deviations that keep the incentives intact. Then it must be the case for each of them that they do not increase the principal's expected utility:

(1) Change $\bar{c}_L$ and $\underline{c}_H$ simultaneously such that

$$f(v_L, \bar{c}_L) \cdot d\bar{c}_L \;=\; f(v_L, \underline{c}_H) \cdot d\underline{c}_H \tag{1.A.15}$$

The principal's expected utility changes by

$$f(v_H, \underline{c}) \cdot (v_H - \underline{c}_H) \cdot d\underline{c}_H + f(v_L, \bar{c}_L) \cdot (v_L - \bar{c}_L) d\bar{c}_L.$$

Plugging in (1.A.15) yields:

$$f(v_L, \bar{c}_L) \cdot \left(\frac{f(v_H, \underline{c}_H)}{f(v_L, \underline{c}_H)} \cdot (v_H - \underline{c}_H) + (v_L - \bar{c}_H) \right) \cdot d\bar{c}_L$$

To rule out any strictly profitable modification of the mechanism it must be the case:

$$\frac{f(v_H, \underline{c}_H)}{f(v_L, \underline{c}_H)} \cdot (v_H - \underline{c}_H) + (v_L - \bar{c}_H) \;=\; 0 \;\Leftrightarrow\; \frac{v_H - \underline{c}_H}{\bar{c}_H - v_L} \;=\; \frac{f(v_L, \underline{c}_H)}{f(v_H, \underline{c}_H)} \;=\; l(\underline{c}_H). \tag{1.A.16}$$

(2) Change $\bar{c}_L$ and $\bar{c}_H$ simultaneously such that

$$f(v_L, \bar{c}_L) \cdot d\bar{c}_L \;=\; f(v_L, \bar{c}_H) \cdot d\bar{c}_H. \tag{1.A.17}$$

Again, the principal's expected utility cannot be strictly improved if,

$$\frac{v_H - \bar{c}_H}{\bar{c}_L - v_L} \;=\; \frac{f(v_L, \bar{c}_H)}{f(v_H, \bar{c}_H)}. \tag{1.A.18}$$

From Theorem 3 we know that $IC(v_L, v_H) = 0$. This gives a third equation

$$F(\bar{c}_L | v_L) \;=\; F(\bar{c}_h | v_L) - F(\underline{c}_H | v_L). \tag{1.A.19}$$

These three equations ((1.A.16),(1.A.18),(1.A.19)) uniquely pin down the optimal mechanism with screening, since the optimum must fulfill them and I will show that they have at most one solution.

For a a fix $\bar{c}_L$ consider the function $c \mapsto \frac{v_H - c}{\bar{c}_L - v_L}$ on $c \in [v_L, v_H]$. This linear function can be described as a straight line that hits zero at $c = v_H$. For higher $\bar{c}_L$ its slope gets less negative. This linear function intersects at most twice with the convex decreasing function $l(c)$. denote—if existent—the lower intersection point by $\underline{c}_H(\bar{c}_L)$

and the higher intersection point by $\bar{c}_H(\bar{c}_L)$. By the convexity of l it follows that for $\bar{c}_H(\bar{c}_L) - \underline{c}_H(\bar{c}_L)$ is decreasing in $\bar{c}_L$. Therefore

$$F(\bar{c}_L|v_L) - (F(\bar{c}_H(\bar{c}_L)|v_L) - F(\underline{c}_H(\bar{c}_L)|v_L))$$

is increasing in $\bar{c}_L$ It follows that there is at most one $\bar{c}_L$ solving the equations (1.A.16), (1.A.18), (1.A.19) simultaneously.

Suppose now that the optimal mechanism is ignorant. Then the optimal cutoff $\bar{c} \in [v_l, v_H]$ must be locally optimal:

$$0 = f(v_H, \bar{c}) \cdot (v_H - \bar{c}) \cdot d\bar{c} + f(v_L, \bar{c}) \cdot (v_L - \bar{c}) \cdot d\bar{c} \iff \frac{v_H - \bar{c}}{\bar{c} - v_L} = l(\bar{c})$$

If this equation is not unique then we know by corollary 3 that the optimal mechanism is not ignorant. Then it must be uniquely characterized by the equations of the screening case. □

Proof of Lemma 2.
Under which circumstances can we rule out that the system (1.A.16), (1.A.18) and (1.A.19) has a solution. One of this instances is given, if the tangent at a of l crosses zero before v_H. Since then no linear function crossing zero at v_H can twice intersect with l. Formally this is the case if:

$$l(a) + l'(a) \cdot v_H < 0 \iff v_H >$$

□

Appendix 1.B Related Problem

$$\begin{aligned} \tilde{u}_H(y) \cdot g(y) &= f(v', -y) = f(v', c) \\ \tilde{u}_L(y) \cdot g(y) &= f(v, -y) = f(v, c) \\ 0 \leq \left(\frac{\tilde{u}'_L(y)}{u'_H(y)}\right)' = \left(-l'(-y) \cdot \xi(x)\tilde{u}_L(y)\right) \cdot g(y) &= f(v, -y) \\ \tilde{w}(y) \cdot g(y) &= f(v', -y) \cdot (v' + y) \end{aligned}$$

It follows:

$$\begin{aligned} \frac{\tilde{u}_L(x)}{\tilde{u}_H(x)} = \frac{f(v, -x)}{f(v', -x)} &=: l(x) \\ \frac{\tilde{w}(x)}{\tilde{u}_H(x)} &= x. \end{aligned}$$

From that we get,

$$\tilde{u}_L'(y) \;=\; -l'(y)\cdot\tilde{u}_L(y) + l(y)\cdot\tilde{u}_L(y) \tag{1.B.1}$$

$$\tilde{w}'(y) \;=\; \tilde{u}_H(y) + y\cdot\tilde{u}_H'(y) \tag{1.B.2}$$

To have the principal with utility $\tilde{w}$ to be more risk avers than an agent with utility $\tilde{u}_H$ we need to have,

$$0 \;\geq\; \left(\frac{\tilde{w}'(y)}{\tilde{u}_H'(y)}\right)' \;=\; \left(\underbrace{\frac{\tilde{u}_H(y)}{\tilde{u}_H'(y)}}_{=:\xi(y)} + y\right)' \;=\; \xi'(y)+1$$

To also have an agent with utility function $\tilde{u}_H$ be more risk averse than an agent with utility function $\tilde{u}_L$ it must be the case, that

$$0 \;\leq\; \left(\frac{\tilde{u}_L'(y)}{u_H'(y)}\right)' \;=\; \big(-l'(-y)\cdot\xi(y) + l(-y)\big)' \;=\; l''(-y)\cdot\xi(y) - l'(-y)\cdot\xi'(y) - l'(y)$$

putting both conditions together yields:

$$l''(y)\cdot\xi(y) \;\geq\; l'(-x)\cdot(\xi'(y)+1) \;\leq\; 0$$

Assuming that $\tilde{u}_H, u_H' \geq 0$ and $l' \leq 0$ we conclude,

$$l''(-y) \;\geq\; 0.$$

References

Aliprantis, Charalambos D., and Kim Border. 2006. *Infinite dimensional analysis*. Springer. [40]

Alonso, Ricardo, and Niko Matouschek. 2008. "Optimal delegation." *Review of Economic Studies* 75 (1): 259–93. [7]

Alós-Ferrer, Carlos. 1999. "Dynamical systems with a continuum of randomly matched agents." *Journal of Economic Theory* 86 (2): 245–67. [39]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L Lipman. 2014. "Optimal allocation with costly verification." *American Economic Review* 104 (12): 3779–813. [7]

Bhargava, Mohit, Dipjyoti Majumdar, and Arunava Sen. 2015. "Incentive-compatible voting rules with positively correlated beliefs." *Theoretical Economics* 10 (3): 867–85. [7]

Border, Kim C. 2013. "Alternative linear inequalities." *Cal Tech Lecture Notes*, [32]

Chakravarty, Surajeet, and Todd R Kaplan. 2013. "Optimal allocation without transfer payments." *Games and Economic Behavior* 77 (1): 1–20. [7]

Cremer, Jacques, and Richard P McLean. 1988. "Full extraction of the surplus in Bayesian and dominant strategy auctions." *Econometrica: Journal of the Econometric Society*, 1247–57. [7, 11]

Daniel K. Elwell. 2019. "Statement before the Senate Committee on Commerce, Science, and Transportation, Subcommittee on Aviation and Space on the State of Airline Safety: Federal Oversight of Commercial Aviation." https://www.faa.gov/news/testimony/news_story.cfm?newsId=23514. [28]

Diamond, Peter A, and Joseph E Stiglitz. 1974. "Increases in risk and in risk aversion." *Journal of Economic Theory* 8 (3): 337–60. [8, 15, 16]

Epitropou, Markos, and Rakesh Vohra. 2019. "Optimal On-Line Allocation Rules with Verification." In *International Symposium on Algorithmic Game Theory*. Springer, 3–17. [7]

Erlanson, Albin, and Andreas Kleiner. 2019. "A note on optimal allocation with costly verification." *Journal of Mathematical Economics* 84: 56–62. [7]

Farkas, J. 1902. "Über die Theorie der einfachen Ungleichungen." *J. Reine Angew. Math* 124: 1–24. [32]

Federal Aviation Administration. 2005. "Establishment of Organization Designation Authorization Program." https://www.federalregister.gov/documents/2005/10/13/05-20470/establishment-of-organization-designation-authorization-program. [28]

Fieseler, Karsten, Thomas Kittsteiner, and Benny Moldovanu. 2003. "Partnerships, lemons, and efficient trade." *Journal of Economic Theory* 113 (2): 223–34. [8]

Gershkov, Alex, Jacob K Goeree, Alexey Kushnir, Benny Moldovanu, and Xianwen Shi. 2013. "On the equivalence of Bayesian and dominant strategy implementation." *Econometrica* 81 (1): 197–220. [7]

Gibbard, Allan. 1973. "Manipulation of voting schemes: a general result." *Econometrica* 41 (4): 587–601. [7]

Guo, Yingni, and Johannes Hörner. 2018. "Dynamic allocation without money." Working paper. Working Paper. [7]

Holmstrom, Bengt. 1977. "On incentives and control in organizations Ph. D." Doctoral dissertation. thesis, Stanford University. [7]

Hylland, Aanund, and Richard Zeckhauser. 1979. "The efficient allocation of individuals to positions." *Journal of Political economy* 87 (2): 293–314. [7]

Jackson, Matthew O, and Hugo F Sonnenschein. 2007. "Overcoming Incentive Constraints by Linking Decisions 1." *Econometrica* 75 (1): 241–57. [7]

Jewitt, Ian. 1988. "Justifying the first-order approach to principal-agent problems." *Econometrica: Journal of the Econometric Society*, 1177–90. [5, 8, 13, 14]

Jewitt, Ian. 1989. "Choosing between risky prospects: the characterization of comparative statics results, and location independent risk." *Management Science* 35 (1): 60–70. [16]

Johnson, Scott, John W Pratt, and Richard J Zeckhauser. 1990. "Efficiency despite mutually payoff-relevant private information: The finite case." *Econometrica*, 873–900. [7]

Karlin, Samuel. 1968. *Total positivity*. Vol. 1, Stanford University Press. [8, 18, 19, 35, 40]

Kartik, Navin, SangMok Lee, and Daniel Rappoport. 2019. "Single-crossing differences on distributions." Working paper. Working paper. [8]

Kattwinkel, Deniz, and Jan Knoepfle. 2019. "Costless Information and Costly Verification: A Case for Transparency." *Available at SSRN 3426817*, [7]

Li, Yunan. 2017. "Mechanism Design with Financially Constrained Agents and Costly Verification." [7]

Manelli, Alejandro M, and Daniel R Vincent. 2010. "Bayesian and Dominant-Strategy Implementation in the Independent Private-Values Model." *Econometrica* 78 (6): 1905–38. [7]

Maskin, Eric, and Jean Tirole. 1990. "The principal-agent relationship with an informed principal: The case of private values." *Econometrica: Journal of the Econometric Society*, 379–409. [7]

McAfee, R Preston, and Philip J Reny. 1992. "Correlated information and mecanism design." *Econometrica: Journal of the Econometric Society*, 395–421. [7]

Motzkin, Theodore Samuel. 1936. *Beiträge zur Theorie der Linearen Ungleichungen: Inaugural-Dissertation, Basel, Jerusalem*. [8]

Myerson, Roger B. 1983. "Mechanism design by an informed principal." *Econometrica: Journal of the Econometric Society*, 1767–97. [7]

Myerson, Roger B, and Mark A Satterthwaite. 1983. "Efficient mechanisms for bilateral trading." *Journal of economic theory* 29 (2): 265–81. [7]

Mylovanov, Tymofiy, and Andriy Zapechelnyuk. 2017. "Optimal allocation with ex post verification and limited penalties." *American Economic Review* 107 (9): 2666–94. [7]

Neeman, Zvika. 2004. "The relevance of private information in mechanism design." *Journal of Economic Theory* 117 (1): 55–77. [7]

Riordan, Michael H, and David EM Sappington. 1988. "Optimal contracts with public ex post information." *Journal of Economic Theory* 45 (1): 189–99. [7]

Satterthwaite, Mark Allen. 1975. "Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions." *Journal of economic theory* 10 (2): 187–217. [7]

Schoenberg, Isac. 1930. "Über variationsvermindernde lineare Transformationen." *Mathematische Zeitschrift* 32 (1): 321–28. [8, 18, 35, 40]

Transportation Department's Inspector General. 2005. "Office of Inspector GeneralAudit Report." https://www.oig.dot.gov/sites/default/files/FAA%20Oversight%20of%20ODA%20Final%20Report%5E10-15-15.pdf. [28]

Chapter 2

Probabilistic Verification in Mechanism Design*

Joint with Ian Ball

2.1 Introduction

In the standard mechanism design paradigm, the principal can commit to an arbitrary mechanism, which induces a game among the agents. There are no explicit constraints on the mechanism, but there is an implicit assumption that the outcome of the induced game does not depend directly on the agents' types. Thus, each type can freely mimic every other type. The principal learns an agent's type only if the mechanism makes it optimal for that agent to reveal it.

In practice, claims about private information are often verified. If an employee applies for disability benefits, the provider performs a medical exam to assess the employee's condition. If a driver makes an insurance claim after a car accident, the insurer checks the claim against a police report. If a consumer reports his income to a lender, the lender requests a monthly pay stub for confirmation. In these examples, verification is noisy—medical tests are imperfect, witnesses fallible, and pay stubs incomplete.

The seminal paper of Green and Laffont (1986) first incorporates *partial verification* into mechanism design. In a principal–agent setting, they illustrate how verification relaxes incentive compatibility and hence makes more social choice functions

* We thank our advisors, Dirk Bergemann and Stephan Lauermann, for continual guidance. For helpful discussions, we thank Tilman Börgers, Marina Halac, Johannes Hörner, Navin Kartik, Andreas Kleiner, Daniel Krähmer, Bart Lipman, Benny Moldovanu, Jacopo Perego, Larry Samuelson, Sebastian Schweighofer-Kodritsch, Philipp Strack, Roland Strausz, and Juuso Välimäki. We are grateful for funding from the German Research Foundation (DFG) through CRC TR 224 (Project B04). An extended abstract for this paper appeared in EC '19.

implementable. In their model, mechanisms are direct and each type faces an exogenous restriction on which reports he can send to the principal. The following interpretation is suggested. As a function of the agent's type, certain reports are always detected as false while other reports never are. The punishment for detection is prohibitively costly, so the agent chooses among the reports he can send without being detected.

Partial verification cannot capture the *noisiness* of real verification. Under partial verification, the agent knows with certainty which reports will be detected as false. But in many applications, each report is detected with an associated probability. Thus, agents trade off the benefits of successful misreporting against the risk of detection.

In this paper, we model *probabilistic verification* by endowing a principal with a stochastic testing technology. We consider a principal–agent setting. The agent has a private type. The principal has full commitment power and controls decisions (which may or may not include transfers). To this standard setting, we add a family of *pass–fail tests*. The principal elicits a message from the agent and then conducts a test. The agent sees the test and privately chooses whether to exert effort, which is costless. If he exerts effort, then his passage probability depends on his type and on the test; if he does not exert effort, then he fails with certainty. The principal observes the result of the test—but not the agent's effort—and then takes a decision.

We analyze which social choice functions can be implemented with a given testing technology. We reduce the class of mechanisms in two steps.

First, we simplify communication. Since testing does not intrude into the communication stage, we get a version of the revelation principle (Theorem 8): There is no loss in restricting attention to direct mechanisms that induce the agent to report truthfully and to exert effort on every test. We contrast our result with the failure of the revelation principle in the setting of Green and Laffont (1986).

Second, we simplify the choice of tests. In general, which test is best for verifying a particular type report depends on which types the principal would like to screen away. We introduce for each type θ an associated order over tests. One test is more *θ-discerning* than another if it can better screen all other types away from type θ. This family of orders is the appropriate analogue of Blackwell's (1953) informativeness order for our testing setting. These discernment orders provide a unified generalization of various conditions imposed in the deterministic verification literature, such as nested range (Green and Laffont, 1986), full reports (Lipman and Seppi, 1995), and normality (Bull and Watson, 2007).

A function assigning to each type θ a most θ-discerning test is a *most-discerning* testing function. The sufficiency part of our main implementation result (Theorem 9) states: If there exists a most-discerning testing function, then every implementable social choice function can be implemented using that testing function. In this case, the testing technology induces an *authentication rate*, which specifies the probabilities with which each type can pass the test assigned to each other type. The principal's problem reduces to an optimization over decision rules, subject to incentive constraints involving the authentication probabilities.

The reduction from a testing technology to an authentication rate can be inverted. We provide a necessary and sufficient condition for an authentication rate to be induced by a most-discerning testing function. In applications, we can directly specify an authentication rate that satisfies our condition—testing need not be modeled explicitly. If the authentication rate takes values 0 and 1 only, then our condition reduces to the nested range condition that Green and Laffont (1986) use to recover the revelation principle.

We are the first to analyze verification with the first-order approach. . Partial verification is not amenable to this approach because the authentication probability jumps discontinuously from 0 to 1. Under probabilistic verification, the authentication rate can depend continuously on the agent's report, so each local constraint is loosened but not eliminated. In a quasilinear environment, we aggregate these loosened local constraints to derive a *virtual value* that encodes the testing technology.

We use this virtual value to solve for revenue-maximizing mechanisms in three classical settings: nonlinear pricing, selling a single good, and auctions. With verification, the revenue-maximizing allocations have their usual expressions, except that our virtual value appears in place of the classical virtual value. The associated transfers are higher in the presence of verification. If the tests are completely uninformative, then our virtual value equals the classical virtual value. As the tests become more precise, our virtual value increases toward the agent's true value, and the revenue-maximizing allocation becomes more efficient. When selling a single good, a posted price is not optimal. Instead, the price depends on the agent's report. To study auction settings, we extend the model to allow for competing agents who submit reports and are tested separately. A virtual value is defined for each agent. The revenue-maximizing auction allocates the good to the agent whose virtual value is highest.

Finally, we consider tests with more than two results. Our discernment orders naturally extend. As in the baseline model, if there exists a most-discerning testing function, then there is no loss in using that testing function only.

The rest of the paper is organized as follows. section 2.2 presents our model of testing. section 2.3 establishes the revelation principle. section 2.4 introduces the discernment orders and states the most-discerning implementation result. section 2.5 reduces the verification technology to an authentication rate. section 2.6 considers applications to revenue-maximization. We extend the model to allow for multiple agents in section 2.7 and nonbinary tests in section 2.8. section 2.9 connects our model to previous models of verification in economics and computer science; other relevant literature is referenced throughout the text. section 2.10 concludes. Measure-theoretic definitions are in section 2.A. Proofs are in sections 2.B and 2.C.

2.2 Model

2.2.1 Setting

There are two players: a principal (she) and an agent (he). The agent draws a private type $\theta \in \Theta$ from a commonly known distribution. The principal takes a decision $x \in X$.[1] Preferences depend on the decision x and on the agent's type θ. The Bernoulli utility functions for the agent and the principal are

$$u: X \times \Theta \to \boldsymbol{R} \quad \text{and} \quad v: X \times \Theta \to \boldsymbol{R}.$$

A social choice function, denoted

$$f: \Theta \to \Delta(X),$$

assigns a decision lottery to each type.[2]

2.2.2 Verification technology

To the principal–agent setting we add a verification technology. There is a set T of *tests*, with generic element τ. Each test generates a binary result—pass or fail, denoted 1 or 0.[3] The distribution of test results is determined by the *passage rate*

$$\pi: T \times \Theta \to [0, 1],$$

which assigns to each pair (τ, θ) the probability with which type θ can pass test τ. The passage rate π is common knowledge, as is the rest of the setting, except the agent's private type.

The principal can conduct one test from the set T. The procedure is as follows. First the principal selects a test τ. Next the agent observes τ and chooses to exert effort or not, denoted $\bar{e}$ or $\underline{e}$. Effort is costless. If the agent exerts effort, nature draws the test result 1 with probability $\pi(\tau|\theta)$ and the test result 0 otherwise. If the agent does not exert effort, the test result is 0 with certainty. The principal observes the test result, but not the agent's effort choice.[4]

1. Decisions are completely abstract; they may or may not include transfers.

2. We make the following standing technical assumptions. Each set is a Polish space endowed with its Borel σ-algebra. The space of Borel probability measures on a Polish space Z is denoted $\Delta(Z)$. All primitive functions are measurable. Social choice functions and other measure-valued maps satisfy a weaker measurability condition called universal measurability. The details are in section 2.A.

3. We consider nonbinary tests in section 2.8; our main results go through.

4. The same procedure is used in Deb and Stewart (2018). They allow the principal to conduct tests (termed *tasks*) sequentially before making a binary classification. In DeMarzo, Kremer, and Skrzypacz (forthcoming), a seller of an asset can conduct a test of the asset's quality. Each test has a *null result*, which the seller can always claim to have received. If there is only one non-null result, then this technology is equivalent to ours.

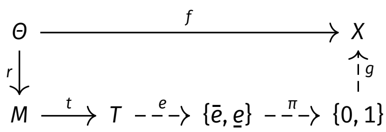


Figure 2.1. Social choice function induced by a profile

Effort is an inalienable choice of the agent, as in models of evidence in which the agent chooses whether to produce the evidence he possesses (Bull and Watson, 2007). Indeed, if the passage rate takes values 0 and 1 only, then each test can be interpreted as a request for a particular piece of hard evidence; see example 6. In general, the passage rate takes interior values, so the agent is unsure whether he will pass each test. The agent can, however, intentionally fail each test by exerting no effort. The active role played by the agent and the resulting asymmetry between passage and failure are what distinguish tests from statistical experiments.

2.2.3 Mechanisms and strategies

The principal commits to a mechanism, which induces an extensive-form game that proceeds as follows. The agent sends a message to the principal. Based on the message, the principal selects a test. The agent observes the test and chooses whether to exert effort. Nature draws the test result, as prescribed by the passage rate and the agent's effort. The principal observes this result and takes a decision.

In view of this timing, we define mechanisms and strategies.

Definition 4 (Mechanism)**.** A *mechanism* (M, t, g) consists of a message space M together with a *testing rule* $t\colon M \to \Delta(T)$ and a *decision rule*

$$g\colon M \times T \times \{0,1\} \;\to\; \Delta(X).$$

Once the principal commits to a mechanism, the agent faces a dynamic decision problem. First he sends a message. Then, after observing the selected test, he chooses effort.

Definition 5 (Strategy)**.** A *strategy* (r, e) for the agent consists of a *reporting strategy* $r\colon \Theta \to \Delta(M)$ and an *effort strategy*

$$e\colon \Theta \times M \times T \;\to\; [0,1],$$

specifying the probability of exerting effort.

A mechanism (M, t, g) and a strategy (r, e) together constitute a *profile*, which induces a social choice function by composition, as indicated in fig. 2.1. The functions e, π, and g are represented by dashed arrows as a reminder that these functions depend

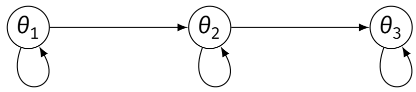


Figure 2.2. Feasible reports

on histories, not just their source sets in the diagram. Using the notation for composition from Markov processes—think of a probability row vector right-multiplied by stochastic matrices—the induced social choice function f is $(r \times t \times (e\pi))g$. The map $r \times t \times (e\pi)$ from Θ to $\Delta(M \times T \times \{0,1\})$ is applied before the map g. For measure-theoretic definitions of these operations, see section 2.A.1.

The players use expected utility to evaluate lotteries over decisions. A profile $(M, t, g; r, e)$ is *incentive compatible* if the strategy (r, e) is a *best response* to the mechanism (M, t, g), i.e., the strategy (r, e) maximizes the agent's utility over all strategies in the game induced by the mechanism (M, t, g). A profile $(M, t, g; r, e)$ *implements* a social choice function f if $(M, t, g; r, e)$ is incentive compatible and $f = (r \times t \times (e\pi))g$. A social choice function is *implementable* if there exists a profile that implements it.

2.3 Revelation principle

We now begin our analysis of which social choice functions can be implemented, given a testing technology. The space of incentive-compatible profiles is large. We reduce this space by establishing a version of the revelation principle. First we revisit the failure of the revelation principle in Green and Laffont's (1986) model of partial verification.

Example 5 (Partial verification, Green and Laffont, 1986)**.** The agent has one of three types, labeled θ_1, θ_2, and θ_3. Type θ_1 can report θ_1 or θ_2; type θ_2 can report θ_2 or θ_3; and type θ_3 can report only θ_3. This correspondence is represented as a directed graph in fig. 2.2. Each type is a node, and edges connect each type to each of his feasible reports.

The principal chooses whether to allocate a single good to the agent, who prefers to receive it, no matter his type. Consider the social choice function that allocates the good if and only if the agent's type is θ_2 or θ_3. To truthfully implement this allocation, the principal must allocate the good if the agent reports θ_2. But then type θ_1 can report θ_2 in order to get the good. Therefore, this allocation cannot be implemented truthfully. It can, however, be implemented *untruthfully*: The principal allocates the good if and only if the agent reports θ_3. Types θ_2 and θ_3 both report θ_3, but type θ_1 cannot.

What goes wrong in this example? In the partial verification model, reports are not cheap-talk messages. Instead, each report serves as a test that certain types can pass. Truthful implementation implicitly assigns to each type θ the test "report θ", regardless of which other types can pass that test. By contrast, our testing framework separates communication from verification. Reports retain their usual meaning and thus the revelation principle holds.

In the standard mechanism design setting, the revelation principle states that there is no loss in restricting to direct and truthful profiles. In our testing framework, we demand also that the agent exert effort on every test.

Definition 6 (Canonical)**.** A profile $(M, t, g; r, e)$ is *canonical* if (i) $M = \Theta$, (ii) $r = \mathrm{id}$, and (iii) $e(\theta, \theta, \tau) = 1$ for all $\theta \in \Theta$ and $\tau \in T$.

Here the identity, id, maps each type θ to the point mass δ_θ in $\Delta(\Theta)$.[5] Condition (i) says that the mechanism is direct. Condition (ii) says that the agent reports truthfully. Condition (iii), which is specific to the verification setting, says that the agent exerts effort on every test.[6] A social choice function is *canonically implementable* if it can be implemented by a canonical profile. Our revelation principle states that there is no loss in restricting to canonical implementation.

Theorem 8 (Revelation principle)
Every implementable social choice function is canonically implementable.

The structure of the proof is similar to that of the standard revelation principle. Given an arbitrary profile $(M, t, g; r, e)$ that implements a social choice function f, we construct a canonical profile that also implements f. Play proceeds as follows.[7] The agent truthfully reports his type θ. The principal feeds this report θ into the reporting strategy r to draw a message m, which is then passed to the testing rule t to draw a test τ. The agent exerts effort, so nature draws the test result 1 with probability $\pi(\tau|\theta)$ and the test result 0 otherwise. If the test result is 0, the principal feeds 0 into the decision rule g. If the test result is 1, the principal feeds into g the result 1 with probability $e(\theta, m, \tau)$ and the result 0 otherwise. Therefore, the input to g is 1 with probability $\pi(\tau|\theta)e(\theta, m, \tau)$.[8]

This canonical profile induces f. To check that this profile is incentive compatible, we show that for any deviation in the canonical mechanism, there is a corresponding deviation in the original mechanism that induces the same (stochastic) decision. The original profile is incentive compatible, so this deviation cannot be profitable.

5. We sometimes identify a map $y \mapsto f(y)$ from Y to Z with the map $y \mapsto \delta_{f(y)}$ from Y to $\Delta(Z)$.

6. The agent's off-path behavior can be ignored because the best-response condition is in strategic form. Alternatively, condition (iii) could be strengthened to require that $e(\theta, \theta', \tau) = 1$ for all $\theta, \theta' \in \Theta$ and $\tau \in T$. The results would not change.

7. In the proof, we use a more general procedure that works also for nonbinary tests.

8. Technically, this is not a mechanism but a *generalized* mechanism (Mertens, Sorin, and Zamir, 2015, Exercise 10, p. 70) because the distribution of the randomization device—the message the principal draws—depends on the agent's type report. In the proof, we eliminate this device by applying the disintegration theorem.

Suppose that the agent reports θ' and then follows the strategy of exerting effort with probability $\tilde{e}(\tau)$ on each test τ. The agent can get the same decision in the original mechanism by playing as follows. Feed θ' into the strategy r to draw a message m to send to the principal. If test τ is conducted, exert effort with probability $\tilde{e}(\tau)e(\theta', m, \tau)$.

By separating communication from testing, we recovered the revelation principle. Conceptually, our testing framework elucidates the role of verification in eliciting private information. Computationally, our progress is less clear. The complexity of untruthful reporting has been replaced by the complexity of testing rules. Reducing this class of testing rules is what we turn to next.

2.4 Ordering tests

In this section, we identify a smaller class of testing rules that suffices for implementation. To define this class, we introduce a family of orders over tests.

2.4.1 Discernment orders

The basic question is whether one test can *always* be used in place of another. More precisely, fix a type θ and tests τ and ψ. Suppose that the principal conducts test ψ on type θ. Is it *always* possible to replace test ψ with test τ and then adjust the decision rule so that (i) the decision for type θ is preserved, and (ii) no new deviations are introduced? The key is to *convert* each score[9] on test τ into an equivalent score on test ψ. The principal feeds this converted score into the original decision rule.

A score conversion is a Markov transition k on $\{0, 1\}$, which associates to each score s in $\{0, 1\}$ a measure k_s in $\Delta(\{0, 1\})$. A transition k on $\{0, 1\}$ is *monotone* if k_1 first-order stochastically dominates k_0, denoted $k_1 \geq_{\mathrm{SD}} k_0$. For each test τ and type θ, denote by $\pi_{\tau|\theta}$ the measure on $\{0, 1\}$ that puts probability $\pi(\tau|\theta)$ on 1. When a test result is drawn from the measure $\pi_{\tau|\theta}$ and a Markov transition k is applied, the resulting distribution is denoted $\pi_{\tau|\theta}k$, which can be viewed as the product of a row vector and a stochastic matrix.

Definition 7 (Discernment order)**.** Fix a type θ. A test τ is *more θ-discerning* than a test ψ, denoted $\tau \succeq_\theta \psi$, if there exists a monotone Markov transition k on $\{0, 1\}$ satisfying:

(i) $\pi_{\tau|\theta}k = \pi_{\psi|\theta}$;
(ii) $\pi_{\tau|\theta'}k \leq_{\mathrm{SD}} \pi_{\psi|\theta'}$ for all types θ' with $\theta' \neq \theta$.

9. We use *score* synonymously with *result*. Tests are pass–fail, but the language of scoring is more intuitive and our constructions immediately extend to the nonbinary case.

Conditions (i) and (ii) correspond to parts (i) and (ii) of the motivating question above. Each condition compares two testing procedures. On the left side, the agent exerts effort on test τ and his score is converted by k into a score on test ψ. On the right side, the agent exerts effort on test ψ and his score is drawn. Condition (i) says that for type θ these two procedures give the same score distribution. Condition (ii) says that for all other types the converted score distribution is first-order stochastically dominated by the unconverted score distribution on test ψ. The conversion k is required to be monotone so that effort weakly improves the distribution of the converted score. In short, the definition ensures that the conversion k from τ-scores to ψ-scores is fair for type θ but (weakly) unfavorable for all other types.

The relation $\succeq_\theta$ is reflexive and transitive. For reflexivity, take k to be the identity transition. For transitivity, compose the score conversions and note that monotone transitions preserve first-order stochastic dominance and are closed under composition; see section 2.A.2. Two tests can be incomparable under $\succeq_\theta$. We use the notation $\sim_\theta$ for equivalence under $\succeq_\theta$, and $\succ_\theta$ for the strict part of $\succeq_\theta$.

The θ-discernment order resembles Blackwell's (1953) informativeness order between experiments. Given a state space Ω and a signal space S, an *experiment* is a map from Ω to $\Delta(S)$. In an experiment, the signal realizations are drawn exogenously by nature. A *garbling* is a Markov transition on S. An experiment τ is *more Blackwell informative* than an experiment ψ if there exists a garbling g such that

$$\tau g = \psi. \tag{2.1}$$

To bring out the connection with the discernment orders, set $\Omega = \Theta$, and denote by $\pi_{\tau|\theta'}$ the distribution of signals from experiment τ in state θ'. Then (2.1) can be expressed as

$$\pi_{\tau|\theta'} g = \pi_{\psi|\theta'} \quad \text{for all } \theta' \in \Theta. \tag{2.2}$$

The Blackwell order is concerned with information, not incentives. The garbled signal from experiment τ must have the same distribution as the ungarbled signal from experiment ψ, in every state of the world. No state is privileged, and no structure on the signal space is required. In contrast, the discernment orders reflect the agent's incentives to report truthfully and to exert effort. There is a family of discernment orders, one associated with each type θ. For the distinguished type θ, the converted score on test τ must have the same distribution as the unconverted score on test ψ. For all other types—the potential deviators—the converted score on test τ need only be stochastically dominated by the score on test ψ. A conversion, unlike a garbling, is required to be monotone so that effort weakly improves the distribution of the converted score. For dominance and monotonicity to make sense, scores must be totally ordered.

2.4.2 Implementation with most-discerning testing functions

We are interested in maximum tests with respect to the discernment orders.

Definition 8 (Most discerning)**.** A test τ is *most θ-discerning* if $\tau \succeq_\theta \psi$ for all $\psi \in T$. A function $t\colon \Theta \to T$ is *most discerning* if, for each type θ, the test $t(\theta)$ is most θ-discerning.

To state the implementation result, we need a few definitions. Given a type space Θ, a *testing environment* consists of a test set T and a passage rate $\pi\colon T \times \Theta \to [0,1]$. A *decision environment* consists of a decision set X and a utility function $u\colon X \times \Theta \to \mathbf{R}$ for the agent. Given a testing rule $\hat{t}\colon \Theta \to \Delta(T)$, a social choice function f is *canonically implementable with* $\hat{t}$ if there exists a decision rule g such that the mechanism $(\hat{t}, g)$ canonically implements f.

Theorem 9 (Most-discerning implementation)
Fix a type space Θ and a testing environment (T,π). For a measurable testing function $\hat{t}\colon \Theta \to T$, the following are equivalent.

(1) $\hat{t}$ is most discerning.

(2) In every decision environment (X,u), every implementable social choice function is canonically implementable with $\hat{t}$.

The implication from condition (1) to condition (2) means that a single most-discerning testing function suffices for implementation. The proof formalizes the test replacement that motivated our definition of the discernment orders. By the revelation principle (Theorem 8), there is no loss in considering only canonical profiles. Suppose that in a canonical profile, some report θ is assigned a test ψ with $\psi \neq \hat{t}(\theta)$. Since $\hat{t}(\theta) \succeq_\theta \psi$, the principal can use a score conversion to replace test ψ with test $\hat{t}(\theta)$, without introducing any new deviations. We perform this replacement simultaneously for every type to construct an incentive-compatible canonical profile with testing rule $\hat{t}$.

The implication from condition (2) to condition (1) means that the most-discerning property is necessary. If a testing function $\hat{t}$ is not most discerning, then in some decision environment there is an implementable social choice function that cannot be canonically implemented with testing rule $\hat{t}$. The construction is as follows. Since $\hat{t}$ is not most-discerning, there exists some type θ and some test τ such that $\hat{t}(\theta) \not\succeq_\theta \tau$. We start with a social choice function that can be implemented by assigning test τ to type θ. To replace test τ with test $\hat{t}(\theta)$, we would need a score conversion from τ to $\hat{t}(\theta)$ that preserves the decision for type θ. But any such conversion is either nonmonotone or violates the dominance condition (ii). If the conversion is nonmonotone, then type θ can improve his passage probability by not exerting effort. If (ii) is violated, then there is some other type θ' whose score distribution after reporting θ is improved by the conversion. In the proof, we construct a decision environment in which these deviations are profitable.

The most-discerning property takes a simple form when tests are deterministic.

Example 6 (Deterministic tests and evidence). If the passage rate π is $\{0,1\}$-valued, then our testing framework reduces to Bull and Watson's (2007) model of hard evidence. Each test can be interpreted as a request for a piece of evidence. Define a correspondence $E\colon \Theta \twoheadrightarrow T$ by

$$E(\theta) = \{\tau \in T : \pi(\tau|\theta) = 1\}.$$

Type θ can provide the evidence requested by test τ if and only if τ is in $E(\theta)$. It can be shown that[10] a test τ in $E(\theta)$ is most θ-discerning if and only if, for every test ψ in $E(\theta)$,

$$E^{-1}(\tau) \subseteq E^{-1}(\psi).$$

That is, test τ is the hardest test that type θ can pass. The existence of a most-discerning testing function is equivalent to Bull and Watson (2007) evidentiary normality condition, which in turn is equivalent to Lipman and Seppi (1995) full reports condition.

If a most-discerning testing function does not exist, we can still reduce the class of testing rules that we need to consider.

Definition 9 (Most-discerning correspondence). A subset T_0 of T is *most θ-discerning* if for each test $\psi \in T$ there exists a test $\tau \in T_0$ such that $\tau \succeq_\theta \psi$. A correspondence $\hat{T}\colon \Theta \twoheadrightarrow T$ is *most discerning* if for each type θ the set $\hat{T}(\theta)$ is most θ-discerning.

A testing function $\hat{t}\colon \Theta \to T$ is a *selection* from a correspondence $\hat{T}\colon \Theta \twoheadrightarrow T$ if $\hat{t}(\theta) \in \hat{T}(\theta)$ for each $\theta \in \Theta$. We extend this notion to stochastic testing rules. A testing rule $\hat{t}\colon \Theta \to \Delta(T)$ is *supported on* a correspondence $\hat{T}\colon \Theta \twoheadrightarrow T$ if $\operatorname{supp} \hat{t}_\theta \subset \hat{T}(\theta)$ for each $\theta \in \Theta$. The next result says that if a correspondence is most discerning, then we can restrict attention to the testing rules supported on that correspondence. To avoid measurability problems, we impose additional regularity conditions.

Theorem 10 (Implementation with a most-discerning correspondence)
Suppose that the passage rate π is continuous. Let $\hat{T}$ be a correspondence from Θ to T with closed values and a measurable graph. If $\hat{T}$ is most discerning, then for every implementable social choice function f, there exists a testing rule $\hat{t}$ supported on $\hat{T}$ such that f is canonically implementable with $\hat{t}$.

10. Argue directly from the definition using deterministic score conversions. Alternatively, apply Proposition 5, stated below.

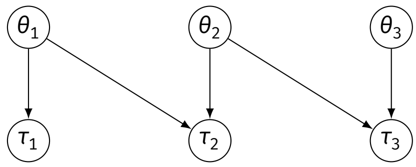


Figure 2.3. Passage correspondence

If $\hat{T}$ is singleton-valued, then exactly one testing rule is supported on $\hat{T}$, and Theorem 10 reduces to the sufficiency part of Theorem 9.[11] In general, Theorem 10 allow us to restrict attention to testing rules supported on a fixed correspondence $\hat{T}$. For example, suppose each type θ in some subset Θ_0 of Θ has a most θ-discerning test $\hat{t}(\theta)$. Take $\hat{T}(\theta) = \{\hat{t}(\theta)\}$ for $\theta \in \Theta_0$ and $\hat{T}(\theta) = T$ for $\theta \notin \Theta_0$.

We apply Theorem 10 to the the Green–Laffont example, which we reformulate with tests.

Example 7 (Green–Laffont with tests)**.** Set $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and $T = \{\tau_1, \tau_2, \tau_3\}$. Type θ_i can pass test τ_j if and only if, in example 5, type θ_i can report θ_j. This $\{0,1\}$-valued passage rate is represented in fig. 2.3 as a directed graph on $\Theta \cup T$.[12] Edges connect each type to each of the tests he can pass. The discernment relations are given by

$$\tau_1 \succ_{\theta_1} \tau_2 \succ_{\theta_1} \tau_3, \qquad \tau_2, \tau_3 \succ_{\theta_2} \tau_1, \qquad \tau_3 \succ_{\theta_3} \tau_2 \sim_{\theta_3} \tau_1.$$

Tests τ_2 and τ_3 are incomparable under $\succeq_{\theta_2}$ because τ_2 screens away type θ_3 (but not θ_1) and τ_3 screens away θ_1 (but not θ_3). The following correspondence $\hat{T}$ is most-discerning:

$$\hat{T}(\theta_1) \;=\; \{\tau_1\}, \qquad \hat{T}(\theta_2) \;=\; \{\tau_2, \tau_3\}, \qquad \hat{T}(\theta_3) \;=\; \{\tau_3\}.$$

By Theorem 10, there is no loss in restricting to testing rules supported on $\hat{T}$. But we cannot assume that type θ_2 is assigned test τ_2—this is the crux of the original counterexample.

11. In this case, the continuity assumption on π is unnecessary. We use the regularity assumptions only to apply a measurable selection theorem, which ensures that there is a measurable way to assign each pair (θ, ψ) a test τ in $\hat{T}(\theta)$ satisfying $\tau \succeq_\theta \psi$. If it can be shown independently that such an assignment exists, then the conclusion of Theorem 10 follows without any assumptions on π or $\hat{T}$. When $\hat{T}$ is singleton-valued, there is at most one such assignment.

12. Compare this graph on $\Theta \cup T$ with the graph on Θ in fig. 2.2.

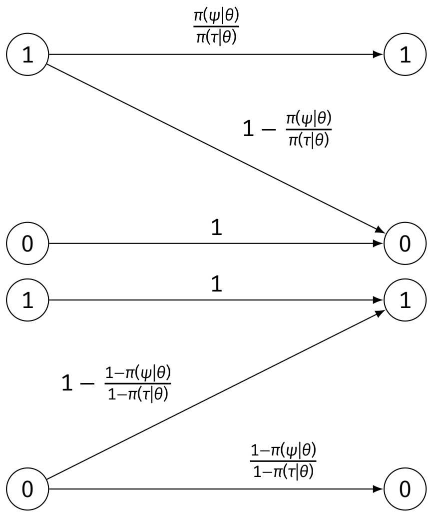


Figure 2.4. Conversion from test τ to test ψ for type θ

2.4.3 Characterizing the discernment orders

We provide a more practical characterization of each discernment order. Fix a type θ and tests τ and ψ. To determine whether τ is more θ-discerning than ψ, we first characterize the monotone Markov transitions k satisfying $\pi_{\tau|\theta}k = \pi_{\psi|\theta}$. We parameterize these transitions as convex combinations of two extreme points.

The first extreme point is obtained by matching quantiles, much like scores are converted between the SAT and the ACT. fig. 2.4 shows this Markov transition, separated into two cases. In the left panel, $\pi(\tau|\theta) \geq \pi(\psi|\theta)$, so a score of 0 on τ is never converted to a 1 on ψ. In the right panel, $\pi(\tau|\theta) < \pi(\psi|\theta)$, so a score of 1 on τ is never converted to a 0 on ψ.

To formally define this transition, we construct Markov transitions that are analogues of distribution and quantile functions. Given a type θ and a test τ, the associated *cumulative distribution transition*

$$\tilde{F}_{\tau|\theta} : \{0, 1\} \rightarrow \Delta([0, 1])$$

maps 0 and 1 to the uniform distributions over $[0, 1 - \pi(\tau|\theta)]$ and $[1 - \pi(\tau|\theta), 1]$, respectively. The associated *quantile transition*

$$\tilde{Q}_{\tau|\theta} : [0, 1] \rightarrow \Delta(\{0, 1\})$$

maps points in $[0, 1 - \pi(\tau|\theta)]$ to δ_0 and points in $(1 - \pi(\tau|\theta), 1]$ to δ_1.[13] The quantile-matching transition is the composition $\tilde{F}_{\tau|\theta}\tilde{Q}_{\psi|\theta}$.

The second extreme point is the constant transition that maps both scores 0 and 1 to the measure $\pi_{\psi|\theta}$. This transition is denoted $\pi_{\psi|\theta}$ as well.

13. To handle an edge case in Proposition 4, we redefine $\tilde{Q}_{\tau|\theta}$ to map 1 to δ_1 even if $\pi(\tau|\theta) = 0$. For more general definitions of these transitions and for some of their properties, see section 2.A.2.

Proposition 4 (Score conversion characterization)
Fix a type θ and tests τ and ψ. For a Markov transition k on $\{0,1\}$, the following are equivalent.

(1) *k is monotone and $\pi_{\tau|\theta}k = \pi_{\psi|\theta}$.*
(2) *$k = \lambda \tilde{F}_{\tau|\theta}\tilde{Q}_{\psi|\theta} + (1-\lambda)\pi_{\psi|\theta}$ for some $\lambda \in [0,1]$.*

We characterize the discernment order in terms of this parameter $\lambda \in [0,1]$. For any passage rate π, define the associated failure rate $\bar{\pi}$ by $\bar{\pi} = 1 - \pi$.

Proposition 5 (Discernment order characterization)
Fix a type θ and tests τ and ψ.

(1) *Suppose $\pi(\tau|\theta) \geq \pi(\psi|\theta)$. We have $\tau \succeq_\theta \psi$ if and only if there exists $\lambda \in [0,1]$ such that, for all types θ',*

$$[\lambda\pi(\tau|\theta') + (1-\lambda)\pi(\tau|\theta)]\pi(\psi|\theta) \;\leq\; \pi(\psi|\theta')\pi(\tau|\theta). \tag{2.3}$$

(2) *Suppose $\pi(\tau|\theta) < \pi(\psi|\theta)$. We have $\tau \succeq_\theta \psi$ if and only if there exists $\lambda \in [0,1]$ such that, for all types θ',*

$$[\lambda\bar{\pi}(\tau|\theta') + (1-\lambda)\bar{\pi}(\tau|\theta)]\bar{\pi}(\psi|\theta) \;\geq\; \bar{\pi}(\psi|\theta')\bar{\pi}(\tau|\theta). \tag{2.4}$$

Remark. If $\pi(\tau|\theta) \geq \pi(\tau|\theta')$ for all θ', then (2.3) and (2.4) are each weakest when $\lambda = 1$, so we can equivalently require $\lambda = 1$ in the statement of Proposition 5. If, in addition, the passage rates are interior, then (2.3) and (2.4) can be expressed as

$$\frac{\pi(\tau|\theta)}{\pi(\tau|\theta')} \geq \frac{\pi(\psi|\theta)}{\pi(\psi|\theta')} \quad \text{and} \quad \frac{\bar{\pi}(\tau|\theta)}{\bar{\pi}(\tau|\theta')} \leq \frac{\bar{\pi}(\psi|\theta)}{\bar{\pi}(\psi|\theta')}.$$

For the relation $\tau \succeq_\theta \psi$, the passage (failure) rate ratio is what matters if type θ is more likely to pass (fail) test τ than test ψ.

Example 8 (More $\boldsymbol{\theta}$-discerning with $\boldsymbol{\lambda \neq 1}$)**.** For simplicity, we consider a type θ and two tests τ and ψ such that $\pi(\tau|\theta)$ and $\pi(\psi|\theta)$ are equal and nonzero. In this case, test τ is more θ-discerning than test ψ if and only if there exists $\lambda \in [0,1]$ such that

$$\lambda\pi(\tau|\theta') + (1-\lambda)\pi(\tau|\theta) \;\leq\; \pi(\psi|\theta') \quad \text{for all } \theta' \in \Theta. \tag{2.5}$$

The passage rates for these tests are plotted in fig. 2.5.[14] The type space is an interval, plotted on the horizontal axis. For test τ, the passage rate is an increasing affine function; for test ψ, the passage rate is increasing and convex. The dotted line takes the constant value $\pi(\tau|\theta)$, and the dashed line is the average of the passage rate $\pi(\tau|\cdot)$ and the constant $\pi(\tau|\theta)$. From the graph we see that (2.5) is satisfied with

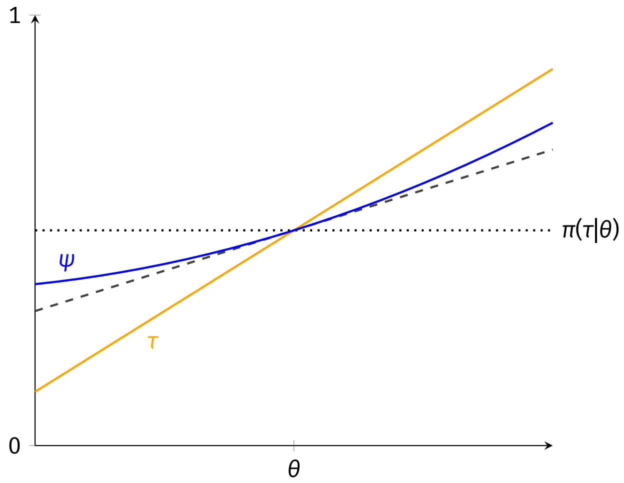


Figure 2.5. More θ-discerning with $\lambda \neq 1$

$\lambda = 1/2$, so $\tau \succeq_\theta \psi$. Moreover, the tangency at the point $(\theta, \pi(\tau|\theta))$ shows that 1/2 is the only value of λ for which (2.5) holds.

Example 9 (Relative performance and scaling tests)**.** Consider a type θ and tests τ, ψ_1, and ψ_2 whose passage rates are plotted in fig. 2.6.[15] Test τ is the test that type θ is least likely to pass, but τ is more θ-discerning than ψ_1 and ψ_2. This is possible because test τ is difficult for every type. The performance of type θ *relative to the other types* is better on test τ than on tests ψ_1 and ψ_2.

This example illustrates also the subtle effect of scaling the passage rate. The passage rate on ψ_2 is a scaling of the passage rate on ψ_1: $\pi(\psi_2|\cdot) = 0.75\pi(\psi_1|\cdot)$. Since $\pi(\psi_1|\theta) > \pi(\psi_2|\theta)$, the relation $\psi_1 \succeq_\theta \psi_2$ depends on the relative passage rates; it is satisfied. The reverse relation depends on the relative failure rates; it is not satisfied. Thus, $\psi_1 \succ_\theta \psi_2$.

Lastly, we study equivalence with respect to each discernment order. Two tests are *θ-equivalent* if they are equivalent with respect to $\succeq_\theta$. Two tests are *equal* if their passage rates are equal. A type θ is *minimal* on a test τ if $\pi(\tau|\theta) \leq \pi(\tau|\theta')$ for all $\theta' \in \Theta$.

Proposition 6 (**θ**-discernment equivalence)
Fix a type θ. Tests τ_1 and τ_2 are θ-equivalent if and only if (i) τ_1 and τ_2 are equal, or (ii) θ is minimal on τ_1 and τ_2.

14. Algebraically, $\Theta = [1/4, 3/4]$ and $\theta = 1/2$. The passage rates are $\pi(\tau|\theta') = 1/2 + (3/4)(\theta' - 1/2)$ and $\pi(\psi|\theta') = \theta'(\theta' - 1/4) + 3/8$.

15. Again, $\Theta = [1/4, 3/4]$ and $\theta = 1/2$. The passage rates are $\pi(\psi_1|\theta') = 0.75 - 4(\theta' - 0.5)^2$; $\pi(\psi_2|\theta') = 0.75\big(0.75 - 4(\theta' - 0.5)^2\big)$; and $\pi(\tau|\theta') = 0.375 - 5(\theta' - 0.5)^2$.

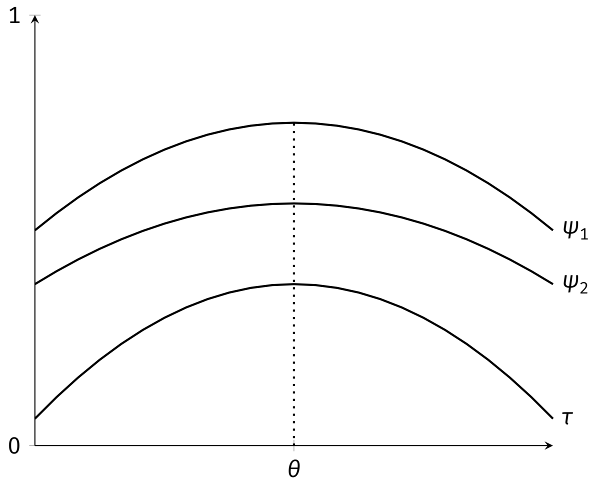


Figure 2.6. Relative performance and scaling

If two unequal tests are θ-equivalent, then these tests have no power to screen other types away from type θ. Whichever of the two tests is used, every decision that is feasible for type θ is also feasible for all other types. Proposition 6 considers θ-equivalence for a fixed type θ. If two tests are θ-equivalent for every type θ, then their passage rates are constant, but the constants are not necessarily equal.

2.5 Testing in reduced form

If there exists a most-discerning testing function, then our framework takes a reduced form in which tests do not appear explicitly.

2.5.1 Induced authentication rates

When the agent makes a type report, his passage probability depends on the test that the principal conducts after that report. If there is a most-discerning testing function $\hat{t}$, then by Theorem 9 we can restrict attention to mechanisms in which each report θ' is assigned test $\hat{t}(\theta')$. In this case, the passage probabilities for each report are pinned down for each type.

Definition 10 (Induced authentication rate)**.** Given a most-discerning testing function $\hat{t}\colon \Theta \to T$, the *authentication rate induced by* $\hat{t}$ is the function $\alpha\colon \Theta \times \Theta \to [0, 1]$ given by

$$\alpha(\theta'|\theta) \;=\; \pi(\hat{t}(\theta')|\theta).$$

There can be multiple most-discerning testing functions. Each induces a different authentication rate. But Theorem 9 guarantees that every most-discerning testing function can be used to implement the same set of social choice functions. There is a corresponding equivalence between the authentication rates induced by different most-discerning testing functions. We need a few definitions. A testing environment is *most discerning* if it admits a most-discerning testing function. A type θ is *minimal* for an authentication rate α if $\alpha(\theta|\theta) \leq \alpha(\theta|\theta')$ for all $\theta' \in \Theta$. Authentication rates α_1 and α_2 are *essentially equal* if (i) α_1 and α_2 have the same minimal types, and (ii) $\alpha_1(\theta|\cdot) = \alpha_2(\theta|\cdot)$ for all types θ that are not minimal for α_1 and α_2. Essential equality is an equivalence relation.

Proposition 7 (Essential uniqueness)
In a most-discerning testing environment, the authentication rates induced by most-discerning testing functions are all essentially equal.

Hereafter, we identify authentication rates that are essentially equal, so we speak of *the* authentication rate induced by a most-discerning testing environment.

2.5.2 Incentive compatibility

Given a most-discerning testing environment (T, π), we reformulate the principal's problem in terms of the authentication rate α induced by (T, π). When the agent reports θ', he is *authenticated* if he passes the associated most θ'-discerning test. A *reduced-form mechanism* consists of functions g_0 and g_1 from Θ to $\Delta(X)$. When the agent reports θ', the principal takes the decision $g_1(\theta')$ if the agent is authenticated and the decision $g_0(\theta')$ otherwise. If type θ reports θ' and exerts effort on the associated test, his interim utility $u(\theta'|\theta)$ is given by

$$u(\theta'|\theta) \;=\; \alpha(\theta'|\theta)u(g_1(\theta'), \theta) + (1 - \alpha(\theta'|\theta))u(g_0(\theta'), \theta).$$

On the right side, the function u is extended linearly from X to $\Delta(X)$. Even in this reduced form, the cost of lying is determined endogenously by the mechanism g, in contrast to models of lying costs.[16]

The incentive-compatibility constraint becomes

$$u(\theta|\theta) \;\geq\; u(\theta'|\theta) \vee u(g_0(\theta'), \theta) \quad \text{for all } \theta, \theta' \;\in\; \Theta. \tag{IC}$$

The right side is the interim utility for type θ if he reports θ' and then chooses effort to maximize his utility. The principal selects a reduced-form mechanism to solve

$$\begin{array}{ll} \text{maximize} & \boldsymbol{E}[\alpha(\theta|\theta)v(g_1(\theta), \theta) + (1 - \alpha(\theta|\theta))v(g_0(\theta), \theta)] \\ \text{subject to} & (IC). \end{array}$$

In the applications below, we also impose participation constraints.

16. In models of lying costs, the agent's utility is the difference between his consumption utility and an exogenous lying cost, which depends on the agent's true type and the agent's report. Lying costs relax the incentive constraints. See, for example, Lacker and Weinberg (1989), Maggi and Rodriguéz-Clare (1995), Crocker and Morgan (1998), Kartik, Ottaviani, and Squintani (2007), Kartik (2009), and Deneckere and Severinov (2017). In computer science, Kephart and Conitzer (2016) show that, with lying costs, the revelation principle holds if the lying cost function satisfies the triangle inequality.

2.5.3 Primitive authentication rates

We showed that a most-discerning testing environment has a simpler representation as an authentication rate. Can we start with the authentication rate as a primitive? To retain the testing interpretation, a primitive authentication rate must be induced by a most-discerning testing environment. We characterize when this is the case.

An authentication rate α implicitly associates to each report θ' a test $\hat{t}(\theta')$ with passage rate $\pi(\hat{t}(\theta')|\cdot) = \alpha(\theta'|\cdot)$. To check whether this testing function $\hat{t}$ is most discerning, we must specify which other tests are in the test set. We claim that there is no less in choosing the minimal test set $\hat{t}(\Theta) = \{\hat{t}(\theta') : \theta' \in \Theta\}$. If $\hat{t}$ is not most-discerning with this test set, then it cannot be most-discerning with any larger test set because adding tests adds constraints to definition 7. We translate this condition—that the testing function $\hat{t}$ is most discerning with the test set $\hat{t}(\Theta)$—into a condition on the authentication rate α. For an authentication rate α, let $\bar{\alpha} = 1 - \alpha$.

Definition 11 (Most-discerning authentication)**.** An authentication rate α is *most discerning* if the following hold for all types θ_2 and θ_3.

(1) If $\alpha(\theta_2|\theta_2) \geq \alpha(\theta_2|\theta_3)$, then there exists $\lambda \in [0,1]$ such that, for all types θ_1,

$$[\lambda\alpha(\theta_2|\theta_1) + (1-\lambda)\alpha(\theta_2|\theta_2)]\alpha(\theta_3|\theta_2) \;\leq\; \alpha(\theta_3|\theta_1)\alpha(\theta_2|\theta_2).$$

(2) If $\alpha(\theta_2|\theta_2) < \alpha(\theta_2|\theta_3)$, then there exists $\lambda \in [0,1]$ such that, for all types θ_1,

$$[\lambda\bar{\alpha}(\theta_2|\theta_1) + (1-\lambda)\bar{\alpha}(\theta_2|\theta_2)]\bar{\alpha}(\theta_3|\theta_2) \;\geq\; \bar{\alpha}(\theta_3|\theta_1)\bar{\alpha}(\theta_2|\theta_2).$$

From our characterization of the discernment order (Proposition 5), we get the following characterization for authentication rates.

Theorem 11 (Authentication rate characterization)
An authentication rate α is induced by some most-discerning testing environment if and only if α is most discerning.

Remark. If $\alpha(\theta|\theta) \geq \alpha(\theta|\theta')$ for all types θ and θ', then α is most discerning if and only if

$$\alpha(\theta_3|\theta_2)\alpha(\theta_2|\theta_1) \;\leq\; \alpha(\theta_3|\theta_1)\alpha(\theta_2|\theta_2), \tag{2.6}$$

for all $\theta_1, \theta_2, \theta_3 \in \Theta$.

If α is $\{0,1\}$-valued and $\alpha(\theta|\theta) = 1$ for all θ, then α induces a message correspondence $M\colon \Theta \twoheadrightarrow \Theta$ defined by

$$M(\theta) \;=\; \{\theta' \;:\; \alpha(\theta'|\theta) \;=\; 1\}.$$

This correspondence M satisfies $\theta \in M(\theta)$ for each θ, as in Green and Laffont (1986). In terms of M, (2.6) becomes

$$\theta_3 \;\in\; M(\theta_2) \quad \& \quad \theta_2 \;\in\; M(\theta_1) \quad\Longrightarrow\quad \theta_3 \;\in\; M(\theta_1),$$

which is exactly Green and Laffont's (1986) nested range condition.

2.6 Applications to revenue-maximization

We solve for revenue-maximizing mechanisms with the local first-order approach. The solutions use a new expression for the virtual value.

2.6.1 Quasilinear setting with verification

Consider the nonlinear pricing setting from Mussa and Rosen (1978). The agent's type $\theta \in \Theta = [\underline{\theta}, \bar{\theta}]$ is drawn from a distribution function F with strictly positive density f. The principal allocates a quantity $q \in \boldsymbol{R}_+$ and receives a transfer $t \in \boldsymbol{R}$.[17] Utilities for the agent and the principal are

$$u(q,t,\theta) \;=\; \theta q - t \quad \text{and} \quad v(q,t) \;=\; t - c(q).$$

Here, c is the cost of production. Assume that $c(0) = c'(0) = 0$ and that the marginal cost c' is strictly increasing and unbounded.

The verification technology is represented by a measurable most-discerning authentication rate $\alpha\colon \Theta \times \Theta \to [0,1]$ that satisfies the following conditions.

(i) $\alpha(\theta|\theta) = 1$ for all types θ.

(ii) For each $\theta' \in \Theta$, the function $\theta \mapsto \alpha(\theta'|\theta)$ is absolutely continuous.

(iii) For each $\theta \in \Theta$, the right and left partial derivatives $D_{2+}\alpha(\theta|\theta)$ and $D_{2-}\alpha(\theta|\theta)$ exist, and the functions $\theta \mapsto D_{2+}\alpha(\theta|\theta)$ and $\theta \mapsto D_{2-}\alpha(\theta|\theta)$ are integrable.[18]

Condition (i) ensures that the agent is authenticated if he reports truthfully. The regularity conditions (ii) and (iii) allow us to apply the envelope theorem. Since α is most discerning, (i) implies that $\alpha(\theta_3|\theta_2)\alpha(\theta_2|\theta_1) \le \alpha(\theta_3|\theta_1)$ for all $\theta_1, \theta_2, \theta_3 \in \Theta$. fig. 2.7 plots an authentication rate that satisfies our assumptions. The agent's type is on the horizontal axis. Each curve corresponds to a fixed report. In this example, the authentication probability decays exponentially in the absolute difference between the agent's type and the agent's report.

We assume that the agent is free to walk away at any time, so we impose an ex post participation constraint. Whether or not the agent is authenticated, his utility must be nonnegative.[19] Without these constraints, the principal could apply severe punishments to effectively prohibit the agent from making any report that is not

17. The pair (q,t) is the decision x in the general model. In applications, t always denotes transfers, *not* testing. Since we work directly with authentication rates, we make no reference to tests.

18. These derivatives are defined by

$$D_{2+}\alpha(\theta'|\theta) = \lim_{h\downarrow 0} \frac{\alpha(\theta'|\theta+h) - \alpha(\theta'|\theta)}{h}, \qquad D_{2-}\alpha(\theta'|\theta) = \lim_{h\downarrow 0} \frac{\alpha(\theta'|\theta) - \alpha(\theta'|\theta-h)}{h}.$$

19. In particular, we rule out upfront payments like those used in Border and Sobel (1987).

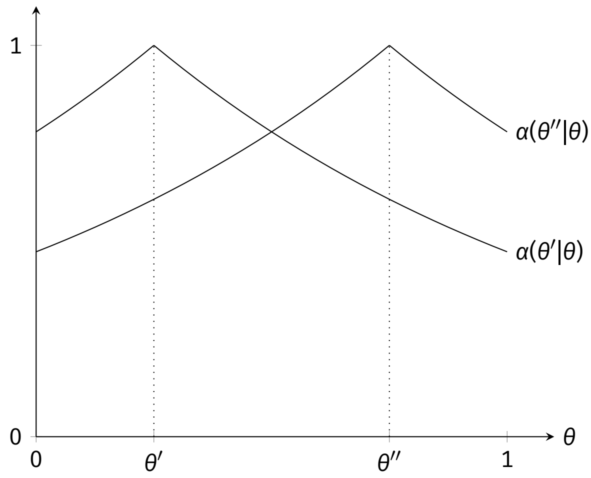


Figure 2.7. Exponential authentication rate

authenticated with certainty. In that case, the model reduces to partial verification, as in Caragiannis, Elkind, Szegedy, and Yu (2012).

We work with reduced-form mechanisms. Since the agent is always authenticated on-path, there is no loss in holding him to his outside option if he is not authenticated. We take $g_0(\theta) = (0, 0)$ for all θ, and we optimize over the decision rule g_1. Without loss, we restrict g_1 to be deterministic. The component functions of g_1 are denoted q and t.

The principal selects a quantity function $q\colon \Theta \to \mathbf{R}_+$ and a transfer function $t\colon \Theta \to \mathbf{R}$ to solve

$$\begin{aligned} &\text{maximize} && \int_{\underline{\theta}}^{\bar{\theta}} [t(\theta) - c(q(\theta))]f(\theta)\,\mathrm{d}\theta \\ &\text{subject to} && \theta q(\theta) - t(\theta) \;\geq\; \alpha(\theta'|\theta)[\theta q(\theta') - t(\theta')], \quad \theta, \theta' \in \Theta \\ &&& \theta q(\theta) - t(\theta) \;\geq\; 0, \quad \theta \in \Theta. \end{aligned} \tag{2.7}$$

The first constraint is incentive compatibility. The second is ex post participation, conditional upon being authenticated. The utility $u(\theta'|\theta)$ takes a simple form because the agent gets zero utility if he is not authenticated. The maximum operation from (IC) is dropped because it is subsumed by the participation constraint.

2.6.2 Virtual value

To motivate our new expression for the virtual value, we use the envelope theorem to compute the agent's equilibrium utility function U, defined by

$$U(\theta) \;=\; u(\theta|\theta) \;=\; \max_{\theta' \in \Theta} u(\theta'|\theta).$$

In the classical setting without verification, $u(\theta'|\theta) = \theta q(\theta') - t(\theta')$. By the envelope theorem, $U'(\theta) = D_2 u(\theta|\theta) = q(\theta)$. Integrating gives

$$U(\theta) \;=\; \int_{\underline{\theta}}^{\theta} q(z)\,\mathrm{d}z. \tag{2.8}$$

With verification, the equilibrium utility U takes a different form. We sketch the derivation here. From (2.7), the interim utility is given by

$$u(\theta'|\theta) \;=\; \alpha(\theta'|\theta)[\theta q(\theta') - t(\theta)].$$

The envelope theorem gives the bounds

$$q(\theta) + D_{2+}\alpha(\theta|\theta)U(\theta) \;\le\; U'(\theta) \;\le\; q(\theta) + D_{2-}\alpha(\theta|\theta)U(\theta).$$

Since $\alpha(\theta|\theta) = 1$, we have $D_{2+}\alpha(\theta|\theta) \le 0 \le D_{2-}\alpha(\theta|\theta)$. If α has a cusp, as in the example in fig. 2.7, these inequalities are strict and the derivative of U is not pinned down. To maximize the principal's revenue, we set $U'(\theta)$ equal to the lower bound. Define the precision function $\lambda \colon \Theta \to \boldsymbol{R}_+$ by

$$\lambda(\theta) \;=\; -D_{2+}\alpha(\theta|\theta).$$

The larger is $\lambda(\theta)$, the steeper is the function $\alpha(\theta|\cdot)$ to the right of θ. For $\theta' \le \theta$, let

$$\Lambda(\theta'|\theta) \;=\; \exp\left(\int_{\theta'}^{\theta} -\lambda(w)\,\mathrm{d}w\right).$$

The minimum equilibrium utility U is given by

$$U(\theta) \;=\; \int_{\underline{\theta}}^{\theta} \Lambda(z|\theta)q(z)\,\mathrm{d}z. \tag{2.9}$$

With this expression for equilibrium utility, we define the virtual value. Recall Myerson's (1981) *virtual value*

$$\varphi^M(\theta) \;=\; \theta - \frac{1-F(\theta)}{f(\theta)} \;=\; \theta - \frac{1}{f(\theta)}\int_{\theta}^{\bar{\theta}} f(z)\,\mathrm{d}z. \tag{2.10}$$

The virtual value of type θ is the marginal expected revenue with respect to the quantity $q(\theta)$. There are two parts. First, the principal can extract the additional consumption utility from type θ, so the marginal revenue from type θ equals θ. Second, the quantity $q(\theta)$ pushes up the equilibrium utility according to (2.8), so the marginal revenue from each type z above θ is -1; this effect is integrated against the relative density $f(z)/f(\theta)$. Verification does not change the marginal revenue

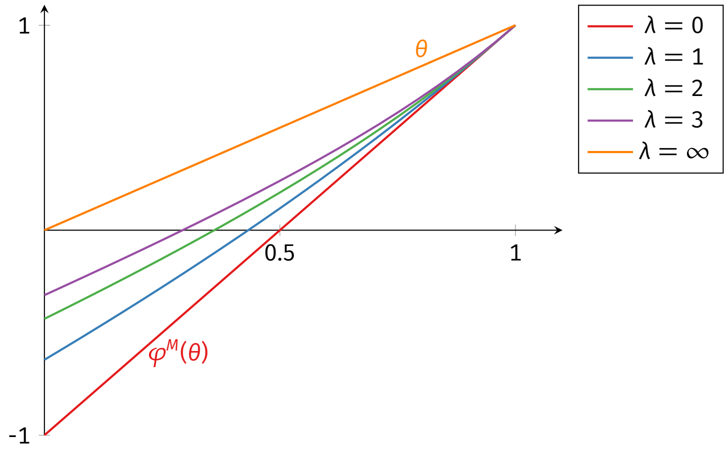


Figure 2.8. Virtual value for different precision functions λ

from type θ, but the marginal revenue from each higher type z becomes $-\Lambda(\theta|z)$, by (2.9). We define the *virtual value* by

$$\varphi(\theta) \;=\; \theta - \frac{1}{f(\theta)} \int_{\theta}^{\bar{\theta}} \Lambda(\theta|z) f(z)\,\mathrm{d}z. \tag{2.11}$$

Comparing (2.10) and (2.11) gives the inequality

$$\varphi^M(\theta) \;\le\; \varphi(\theta) \;\le\; \theta.$$

The virtual value $\varphi(\theta)$ tends towards these bounds in limiting cases.

Proposition 8 (Testing precision)
As λ converges to 0 pointwise, $\varphi(\theta)$ converges to $\varphi^M(\theta)$ for each type θ. As λ converges to ∞ pointwise, $\varphi(\theta)$ converges to θ for each type θ.

fig. 2.8 illustrates these limits in a simple example, where the agent's type is uniformly distributed on the unit interval and the precision function λ is constant.

Remark. If $\lambda(\theta) = 0$ for all θ, then $\Lambda(\theta|z) = 1$ for $\theta \le z$. Therefore, our virtual value coincides with the classical virtual value, and by the results below, the revenue-maximizing mechanism is unaffected by verification. This holds in particular if α has no kink on the diagonal, e.g., if $\alpha(\theta'|\theta) = 1 - |\theta' - \theta|^{\sigma}$ with $\sigma > 1$.

2.6.3 Nonlinear pricing

The virtual value is derived from the envelope representation of the equilibrium utility, which uses only local incentive constraints. We assume that the virtual value is increasing. But because the interim utility is not linear in the agent's type (due to the authentication rate α), we need further assumptions to ensure that the local incentive constraints imply the global incentive constraints. This implication holds in particular for the *exponential* authentication rates, given by

$$\alpha(\theta'|\theta) \;=\; \exp\left(-\int_{\theta'}^{\theta} \lambda(z)\,\mathrm{d}z\right),$$

for integrable functions $\lambda\colon \Theta \to \boldsymbol{R}_+$. We permit a larger class of authentication rates.

We impose a global condition on the relative values of α and Λ. The function Λ is determined by the behavior of α in a neighborhood of the diagonal in $\Theta \times \Theta$. Because α is most discerning, it follows that Λ is a global lower bound for α.

Proposition 9 (Lower bound)
For all types θ' and θ with $\theta' \le \theta$, we have $\alpha(\theta'|\theta) \ge \Lambda(\theta'|\theta)$.

For the exponential authentication rates, this inequality holds with equality. We require that α not be much greater than Λ. The precise condition depends on the optimal quantity function $q^\star$, which will be defined in the theorem statement. The *global upper bound* states that, for all types θ' and θ with $\theta' \le \theta$,

$$\alpha(\theta'|\theta) \;\le\; \Lambda(\theta'|\theta)A(\theta'|\theta),$$

where

$$A(\theta'|\theta) \;=\; \int_{\underline{\theta}}^{\theta} \Lambda(z|\bar{\theta})q^\star(z)\,\mathrm{d}z \Big/ \int_{\underline{\theta}}^{\theta} \Lambda(z \wedge \theta'|\bar{\theta})q^\star(z \wedge \theta')\,\mathrm{d}z.$$

For the quantity function $q^\star$ in the theorem statement, $A(\theta'|\theta) \ge 1$ for $\theta' \le \theta$.

Proposition 10 (Optimal nonlinear pricing)
Suppose that the virtual value φ is increasing. The optimal quantity function $q^\star$ and transfer function $t^\star$ are unique and given by

$$c'(q^\star(\theta)) \;=\; \varphi(\theta)_+, \qquad t^\star(\theta) \;=\; \theta q^\star(\theta) - \int_{\underline{\theta}}^{\theta} \Lambda(z|\theta)q^\star(z)\,\mathrm{d}z,$$

provided that the global upper bound is satisfied.

In the optimal mechanism, type θ receives the quantity that is efficient for type $\varphi(\theta)_+$, just as in Mussa and Rosen (1978), except that φ is our new virtual value. Transfers are pinned down by the equilibrium utility U from (2.9). The faster the virtual value φ increases, the faster the optimal quantity function $q^\star$ increases and the more permissive is the global upper bound.

2.6.4 Selling a single good

Suppose that the principal is selling a single indivisible good, which she does not value. The agent's type is his valuation for the good. The principal allocates the good with probability $q \in [0, 1]$ and receives a transfer $t \in \boldsymbol{R}$. Utilities are

$$u(q, t, \theta) = \theta q - t \quad \text{and} \quad v(q, t) = t.$$

Without verification, the revenue-maximizing mechanisms is a posted price (Riley and Zeckhauser, 1983). With verification, the price my depend on the agent's report.

Proposition 11 (Optimal sale of a single good)
Suppose that the virtual value φ is increasing. The optimal quantity and transfer functions are unique and given as follows, provided that the global upper bound is satisfied. Let $\theta^\star = \inf\{\theta : \varphi(\theta) \geq 0\}$. Each type below $\theta^\star$ receives nothing and pays nothing. Each type θ above $\theta^\star$ receives the good and pays

$$t^\star(\theta) = \theta - \int_{\theta^\star}^{\theta} \Lambda(z|\theta)\, \mathrm{d}z.$$

As in the no-verification solution, there is a cutoff type $\theta^\star$ who receives the good and pays his valuation. Each type below the cutoff is excluded; each type above the cutoff receives the good and pays less than his valuation. The allocation probability takes values 0 and 1 only—there is no randomization. Verification increases the virtual value relative to the classical virtual value, so the cutoff type is lower and more types receive the good. The price is (weakly) increasing in the agent's report, and strictly increasing if λ is strictly positive. Nevertheless, types above the cutoff cannot profit by misreporting downward—the benefit of a lower price is outweighed by the risk of not being authenticated.

2.7 Testing multiple agents

2.7.1 Testing and implementation

We extend our model to allow for n agents, labeled $i = 1, \ldots, n$. Each agent i independently draws his type $\theta_i \in \Theta_i$ from a commonly known distribution $\mu_i \in \Delta(\Theta_i)$. Set $\Theta = \prod_{i=1}^n \Theta_i$. The decision set is denoted by X, as before. Each agent i has utility function $u_i : X \times \Theta \to \boldsymbol{R}$; the principal has utility function $v : X \times \Theta \to \boldsymbol{R}$.

For each agent i, there is a set T_i of tests and a passage rate

$$\pi_i : T_i \times \Theta_i \to [0, 1],$$

where $\pi_i(\tau_i|\theta_i)$ is the probability with which type θ_i can pass test τ_i. Set $T = \prod_{i=1}^n T_i$. Each agent sees his own test—but not the tests of the other agents—and then chooses whether to exert effort. Nature draws the test result for each agent independently. A mechanism specifies a message set M_i for each agent i. Set $M = \prod_{i=1}^n M_i$. The rest of the mechanism consists of a testing rule $t \colon M \to \Delta(T)$ and a decision rule $g \colon M \times T \times \{0,1\}^n \to \Delta(X)$. The test conducted on each agent can depend on the messages sent by other agents. For each agent i, a strategy consists of a reporting strategy $r_i \colon \Theta_i \to \Delta(M_i)$ and an effort strategy $e_i \colon \Theta_i \times M_i \times T_i \to [0,1]$. The equilibrium concept is Bayes–Nash equilibrium.

In this multi-agent setting, the revelation principle (Theorem 8) holds. The discernment orders extend. For each agent i and each type θ_i in Θ_i, the θ_i-discernment order $\succeq_{\theta_i}$ over T_i is defined as in the baseline model with π_i in place of π. Given testing functions $t_i \colon \Theta_i \to T_i$ for each i, define the product testing function $\otimes_i t_i \colon \Theta \to T$ by $t(\theta_1, \dots, \theta_n) = (t_1(\theta_1), \dots, t_n(\theta_n))$. If for each agent i there is a most-discerning testing function, then the product of these testing functions suffices for implementation. In particular, the test for agent i depends only on agent i's report.

Theorem 12 (Most-discerning implementation with multiple players)
Fix a type space Θ and a testing environment (T, π). For a testing function $\hat{t} = \otimes_i \hat{t}_i$, the following are equivalent.

(1) $\hat{t}_i$ is most discerning for all i.

(2) In every decision environment (X, u), every implementable social choice function is canonically implementable with $\hat{t}$.

For applications, we make the same assumptions as in the single-agent case. Each agent is free to walk away, so we impose ex post participation constraints. For each agent i, the testing environment is represented by a measurable most-discerning authentication rate $\alpha_i \colon \Theta_i \times \Theta_i \to [0,1]$ that satisfies assumptions (i)–(iii). For each i, define λ_i and Λ_i by putting α_i in place of α in the definitions of λ and Λ.

2.7.2 Auctions

Consider an auction for a single indivisible good. Each agent i independently draws his type $\theta_i \in \Theta_i = [\underline{\theta}_i, \bar{\theta}_i]$ from a distribution function F_i with positive density f_i. The principal allocates the good to each agent i with probability $q_i \in [0,1]$, where $q_1 + \dots + q_n \le 1$; the principal receives transfers $t_1, \dots, t_n \in \boldsymbol{R}$. Set $q = (q_1, \dots, q_n)$ and $t = (t_1, \dots, t_n)$. For simplicity, we assume that the principal does not value the good. Utilities are given by

$$u_i(q, t, \theta) \; = \; \theta_i q_i - t_i \quad \text{and} \quad v(q, t) \; = \; \sum_{i=1}^n t_i.$$

Let $f_{-i}(\theta_{-i})$ denote $\prod_{j \ne i} f_j(\theta_j)$. For quantity functions $q_i \colon \Theta \to \boldsymbol{R}$, interim expectations are denoted with capital letters:

$$Q_i(\theta_i) \; = \; \int_{\Theta_{-i}} q_i(\theta_i, \theta_{-i}) f_{-i}(\theta_{-i}) \, \mathrm{d}\theta_{-i}.$$

As in the single-agent case, we impose a condition that depends on the optimal quantity function $q^\star$, which is defined in Proposition 12. For each agent i, the *global upper bound* states that, for all $\theta_i, \theta_i' \in \Theta_i$ with $\theta_i' \leq \theta_i$, we have

$$\alpha_i(\theta_i'|\theta_i) \;\leq\; A_i(\theta_i'|\theta_i)\Lambda_i(\theta_i'|\theta_i),$$

where

$$A_i(\theta_i'|\theta_i) \;=\; \int_{\underline{\theta}_i}^{\theta} \Lambda_i(z_i|\bar{\theta}_i)Q_i^\star(z_i)\,\mathrm{d}z_i \Big/ \int_{\underline{\theta}_i}^{\theta} \Lambda_i(z_i \wedge \theta_i'|\bar{\theta}_i)Q_i^\star(z_i \wedge \theta_i')\,\mathrm{d}z_i.$$

For each i and all types $\theta_i, \theta_i' \in \Theta_i$ with $\theta_i' \leq \theta_i$, we have $A_i(\theta_i'|\theta_i) \geq 1$.

Proposition 12 (Optimal auctions)
Suppose that each virtual value φ_i is increasing. The seller's maximum revenue is achieved by the allocation function $q^\star$ and transfer function $t^\star$ given by

$$q_i^\star(\theta) \;=\; \begin{cases} 1 & \text{if } \varphi_i(\theta_i) \;>\; 0 \vee \max_{j\neq i} \varphi_j(\theta_j), \\ 0 & \text{otherwise,} \end{cases}$$

and

$$t_i^\star(\theta) \;=\; q_i^\star(\theta)\left[\theta_i - \int_{\underline{\theta}_i}^{\theta_i} \Lambda_i(z_i|\theta_i)\frac{Q_i^\star(z_i)}{Q_i^\star(\theta_i)}\,\mathrm{d}z_i\right],$$

provided that the global upper bound is satisfied for each agent i.

The transfers are chosen so that each agent pays only if he receives the good, thus ensuring that the the ex post participation constraints are satisfied. The allocation coincides with Myerson's (1981) solution, except φ is our new virtual value. In Myerson's (1981) solution, the allocation rule disadvantages bidders whose valuation distributions are greater in the sense of hazard rate dominance. In our solution, the allocation rule also advantages bidders who can be verified more precisely in the sense that their local precision functions are pointwise greater.

2.8 Nonbinary tests

In the main model, we consider pass–fail tests because of their natural connection with partial verification and evidence. Here we extend the baseline principal–agent model to allow for tests with more than two results. There is a family T of tests. Each test generates scores in a finite subset S of $\boldsymbol{R}$. The passage rate

$$\pi\colon T \times \Theta \;\to\; \Delta(S)$$

assigns to each pair (τ, θ) the score distribution $\pi_{\tau|\theta}$ in $\Delta(S)$ for type θ on test τ.

We generalize the agent's effort choice. On a pass–fail test τ, when type θ exerts effort with probability e, he passes with probability $e \cdot \pi(\tau|\theta)$. Thus, there is

Table 2.1. Taxonomy of verification models

	Reduced form	**Microfoundation**
partial	message correspondence $M: \Theta \twoheadrightarrow \Theta$ Green and Laffont (1986)	evidence correspondence $E: \Theta \twoheadrightarrow \mathscr{E}$ Bull and Watson (2007)
probabilistic	authentication rate $\alpha: \Theta \times \Theta \rightarrow [0,1]$ Caragiannis et al. (2012)	passage rate $\pi: T \times \Theta \rightarrow [0,1]$ our paper

a correspondence between effort probabilities in $[0,1]$ and passage probabilities in $[0, \pi(\tau|\theta)]$. We can equivalently model the agent as choosing the passage probability directly, subject to a stochastic dominance constraint. This alternative definition of a strategy extends immediately to nonbinary tests. A *performance strategy* is a map

$$p: \Theta \times M \times T \;\rightarrow\; \Delta(S),$$

satisfying $p_{\theta,m,\tau} \leq_{\mathrm{SD}} \pi_{\tau|\theta}$ for all $(\theta, m, \tau) \in \Theta \times M \times T$. The other components of a profile are defined as before with S in place of $\{0,1\}$.

The revelation principle (Theorem 8) is proved with performance strategies, so it applies to nonbinary tests. To define the discernment orders, say that a Markov transition k on S is *monotone* if $k_r \geq_{\mathrm{SD}} k_s$ whenever $r > s$. Put S in place of $\{0,1\}$ in the definition of most-discerning. As before, a single most-discerning testing function suffices for implementation.

Theorem 13 (Most-discerning implementation with nonbinary tests)
If a testing function $\hat{t}: \Theta \rightarrow T$ is most discerning, then every implementable social choice function is canonically implementable with $\hat{t}$.

2.9 Review of verification models

Verification has been modeled in many ways, in both economics and computer science. We organize our discussion around the taxonomy in table 2.1, which focuses on the primitives in each model.

Green and Laffont (1986) introduce **partial verification**. They restrict their analysis to direct mechanisms. Verification is represented as a correspondence $M: \Theta \twoheadrightarrow \Theta$ satisfying $\theta \in M(\theta)$ for all $\theta \in \Theta$. Each type θ can report any type θ' in $M(\theta)$. In particular, each type can report truthfully. In this framework the revelation principle does not hold,[20] as Green and Laffont (1986) illustrate with a three-type counterexample, which we adapt in example 5. The revelation principle does hold, however,

20. Strausz (2016) recovers the revelation principle by modeling verification as a component of the outcome.

if the correspondence M satisfies the *nested range condition*, which requires that the relation associated to M is transitive. Without the revelation principle, it is generally difficult to determine whether a particular social choice function is implementable (Rochet, 1987; Nisan and Ronen, 2001; Singh and Wittman, 2001; Auletta, Penna, Persiano, and Ventre, 2011; Vohra, 2011; Yu, 2011; Fotakis and Zampetakis, 2015).

Bull and Watson (2004, 2007) and Lipman and Seppi (1995) model verification with **hard evidence**.[21] They introduce an evidence set $\mathscr{E}$ and an evidence correspondence $E : \Theta \twoheadrightarrow \mathscr{E}$. Type θ possesses the evidence in $E(\theta)$; he can present one piece of evidence from $E(\theta)$ to the principal. The evidence environment is *normal* if each type θ has a piece of evidence $e(\theta)$ in $E(\theta)$ that is maximal in the following sense: Every other type θ' who has $e(\theta)$ also has every other piece of evidence in $E(\theta)$. Therefore, type θ' can mimic type θ if and only if $E(\theta')$ contains $e(\theta)$. A normal evidence environment induces an abstract mimicking correspondence that satisfies the nested range condition. Normality is a special case of our most-discerning condition; see example 6.

In computer science, Caragiannis et al. (2012) and Ferraioli and Ventre (2018) study a reduced-form model of **probabilistic verification** in mechanism design. They restrict their analysis to direct mechanisms, and they specify the probabilities with which each type can successfully mimic each other type. Dziuda and Salas (2018) and Balbuzanov (2019) study a setting without commitment in which these probabilities are constant.[22] Our testing framework microfounds these models of partial verification, provided that the primitive authentication rate is most discerning. If the primitive authentication rate is not most-discerning, then it cannot be microfounded by our testing framework.

The computer science literature on probabilistic verification has a different focus than us. Caragiannis et al. (2012) allow the principal to use arbitrarily severe punishments to deter the agent from misreporting.[23] If one type cannot mimic another type perfectly, then he risks being detected and facing a prohibitive fine. Therefore, this setting reduces to partial verification. In our model, the agent can walk away at any time, so punishment is limited.

If the environment is most discerning, tests can also be interpreted as **stochastic evidence**. Each test τ in T corresponds to a request for a particular piece of evidence. The agent is asked to send a cheap talk message to the principal after he learns his payoff type but *before* he learns which evidence is available to him. With probability

21. Evidence was introduced in games (without commitment) by Milgrom (1981) and Grossman (1981); for recent work on evidence games, see Hart, Kremer, and Perry (2017), Ben-Porath, Dekel, and Lipman (2018), and Koessler and Perez-Richet (2017).

22. For some fixed $p \in (0, 1)$, the authentication probability $\alpha(\theta'|\theta)$ equals p if $\theta' \neq \theta$ and 1 if $\theta' = \theta$. This authentication satisfies (2.6) and hence is most discerning.

23. Another difference is that Caragiannis et al. (2012) investigate which allocation rules can be supported by some transfer rule. We view transfers as part of the outcome and we study revenue-maximization.

$\pi(\tau|\theta)$, type θ will have the evidence requested by test τ. Deneckere and Severinov (2008) study a different kind of stochastic evidence. In their model the agent simultaneously learns his payoff type and his set of feasible evidence messages.

In economics, “verification" traditionally means that the principal can learn the agent’s type perfectly by taking some action, e.g., paying a fee or allocating a good. This literature began with Townsend (1979) who studied **costly verification** in debt contracts. Ben-Porath, Dekel, and Lipman (2019) connect costly verification and evidence. When monetary transfers are infeasible, costly verification is often used as a substitute; see Ben-Porath, Dekel, and Lipman (2014), Erlanson and Kleiner (2015), Halac and Yared (2017), Li (2017), and Mylovanov and Zapechelnyuk (2017).

2.10 Conclusion

We model probabilistic verification as a family of stochastic tests available to the principal. Our testing framework provides a unified generalization of previous verification models. Because verification is noisy, our framework is amenable to the local first-order approach. We illustrate this approach in a few classical revenue-maximization problems. We believe this local approach will make verification tractable in other settings as well.

As the precision of the verification technology varies, our setting continuously interpolates between private information and complete information. Thus we can quantify the value to the principal of a particular verification technology. This is the first step toward analyzing a richer setting in which the principal decides how much to invest in verification.

Appendix 2.A Measure theory

2.A.1 Markov transitions

This section introduces Markov transitions, which are continuous generalizations of stochastic matrices. For further details, see Kallenberg (2017, Chapter 1).

Definition 12 (Markov transition)**.** Let $(X, \mathscr{X})$ and $(Y, \mathscr{Y})$ be measurable spaces. A *Markov transition*[24] from $(X, \mathscr{X})$ to $(Y, \mathscr{Y})$ is a function $k \colon X \times \mathscr{Y} \to [0,1]$ satisfying:

(i) for each $x \in X$, the map $B \mapsto k(x,B)$ is a probability measure on $(Y, \mathscr{Y})$;
(ii) for each $B \in \mathscr{Y}$, the map $x \mapsto k(x,B)$ is a measurable function on $(X, \mathscr{X})$.

A Markov transition k from $(X, \mathscr{X})$ to $(Y, \mathscr{Y})$ is sometimes written as $k \colon X \to \Delta(Y)$. Each measure $k(x, \cdot)$ on Y is denoted k_x, and we write $k_x(B)$ for $k(x,B)$. A

24. Markov transitions are also called kernels or Markov/stochastic/probability/transition kernels.

Markov transition from $(X, \mathscr{X})$ to $(X, \mathscr{X})$ is called a *Markov transition on* $(X, \mathscr{X})$. When the σ-algebras are clear, we will speak of Markov transitions betweens sets.

We introduce three operations between Markov transitions—composition, products, and outer products.

$$(\mu k)(B) \;=\; \int_X \mu(\mathrm{d}x) k(x, B),$$

for all $B \in \mathscr{Z}$.

First we define composition. Let k be a Markov transition from $(X, \mathscr{X})$ to $(Y, \mathscr{Y})$ and ℓ a Markov transition from $(Y, \mathscr{Y})$ to $(Z, \mathscr{Z})$. The *composition* of k and ℓ, denoted $k\ell$, is the Markov transition from $(X, \mathscr{X})$ to $(Z, \mathscr{Z})$ defined by

$$(k\ell)(x, C) \;=\; \int_Y k(x, \mathrm{d}y)\, \ell(y, C),$$

for all $x \in X$ and $C \in \mathscr{Z}$. Here the function $y \mapsto \ell(y, C)$ is integrated over Y with respect to the measure k_x. Inside the integral, it is standard to place the measure before the integrand so that the sequencing of the variables mirrors the timing of the process.

Next we define products. Let k be a Markov transition from $(X, \mathscr{X})$ to $(Y, \mathscr{Y})$ as before, and let m be a Markov transition from $(X \times Y, \mathscr{X} \otimes \mathscr{Y})$ to $(Z, \mathscr{Z})$. Here $\mathscr{X} \otimes \mathscr{Y}$ is the product σ-algebra generated by the measurable rectangles $A \times B$ for $A \in \mathscr{X}$ and $B \in \mathscr{Y}$. The *product* of k and m, denoted $k \times m$, is the unique Markov transition from $(X, \mathscr{X})$ to $(Y \times Z, \mathscr{Y} \otimes \mathscr{Z})$ satisfying

$$(k \otimes m)(x, B \times C) \;=\; \int_B k(x, \mathrm{d}y)\, m((x, y), C),$$

for all $x \in X$, $B \in \mathscr{Y}$, and $C \in \mathscr{Z}$. Here the function $y \mapsto m((x, y), C)$ is integrated over the set B with respect to the measure k_x. If m is a Markov transition from $(Y, \mathscr{Y})$ to $(Z, \mathscr{Z})$, we use the same notation, with the understanding that the integrand is $m(y, C)$ rather than $m((x, y), C)$.

Finally, we define outer products. Let k_1 be a Markov transition from $(X_1, \mathscr{X}_1)$ to $(Y, \mathscr{Y}_1)$ and k_2 a Markov transition from $(X_2, \mathscr{X}_2)$ to $(Y_2, \mathscr{Y}_2)$. The *outer product* of k_1 and k_2, denoted $k_1 \otimes k_2$, is the unique Markov transition from $(X_1 \times X_2, \mathscr{X}_1 \otimes \mathscr{X}_2)$ to $(Y_1 \times Y_2, \mathscr{Y}_1 \otimes \mathscr{Y}_2)$ satisfying

$$(k_1 \otimes k_2)((x_1, x_2), B_1 \times B_2) \;=\; k_1(x_1, B_1) \cdot k_2(x_2, B_2),$$

for all $x_1 \in X_1$, $x_2 \in X_2$, $B_1 \in \mathscr{Y}_1$, and $B_2 \in \mathscr{Y}_2$. All three operations are associative. This holds trivially for outer products; for composition and products, see Kallenberg (2017, Lemma 1.17, p. 33). We drop parentheses when there is no ambiguity.

2.A.2 Markov transitions on the real line

The real line $\boldsymbol{R}$ is endowed with its usual Borel σ-algebra.

First, we define Markov transitions that are analogues of the cumulative distribution and quantile functions. Let μ be a measure on $\boldsymbol{R}$, and let $F_\mu : \boldsymbol{R} \to [0,1]$ be the associated right-continuous cumulative distribution function. Suppose μ has compact support S. Define the left-continuous quantile function $Q_\mu : [0,1] \to S$ by

$$Q_\mu(p) = \inf\{s \in S : F_\mu(s) \geq p\}.$$

The *cumulative distribution transition* associated to μ, denoted $\tilde{F}_\mu$, is the Markov transition from $\boldsymbol{R}$ to $[0,1]$ that assigns to each point s in $\boldsymbol{R}$ the uniform measure on $[F_\mu(s-), F_\mu(s)]$, where $F_\mu(s-)$ is the left-limit of F_μ at s. In particular, if F_μ is continuous at s, then $F_\mu(s-) = F_\mu(s)$ and this uniform measure is the Dirac measure $\delta_{F_\mu(s)}$.

The *quantile transition* associated to μ, denoted $\tilde{Q}_\mu$, is the Markov transition from $[0,1]$ to $\boldsymbol{R}$ that assigns to each number p in $[0,1]$ the Dirac measure $\delta_{Q_\mu(p)}$.

These Markov transitions extend the usual properties of distribution and quantile functions to nonatomic distributions. Let $U_{[0,1]}$ denote the uniform measure on $[0,1]$.

Lemma 8 (Distribution and quantile transitions)**.** For measures μ and ν on $\boldsymbol{R}$ with compact support, the following hold:

(i) $\mu\tilde{F}_\mu = U_{[0,1]}$;
(ii) $U_{[0,1]}\tilde{Q}_\nu = \nu$;
(iii) $\mu\tilde{F}_\mu\tilde{Q}_\nu = \nu$.

For measures μ and ν on the real line, μ *first-order stochastically dominates* ν, denoted $\mu \geq_{\mathrm{SD}} \nu$, if $F_\mu(x) \leq F_\nu(x)$ for all real x. In particular, first-order stochastic dominance is reflexive.

To state the next results, we make the standing assumption that S is a compact subset of $\boldsymbol{R}$, endowed with the restriction of the Borel σ-algebra. A Markov transition d on S is *downward* if $d(s, (-\infty, s] \cap S) = 1$ for all $s \in S$.

Lemma 9 (Downward transitions)**.** For measures μ and ν on S, the following are equivalent:

(i) $\mu \geq_{\mathrm{SD}} \nu$;
(ii) $\tilde{F}_\mu\tilde{Q}_\nu$ is downward;
(iii) $\mu d = \nu$ for some downward transition d.

A Markov transition m on S is *monotone* if $s > t$ implies $m_s \geq_{\mathrm{SD}} m_t$, for all $s, t \in S$.

Lemma 10 (Monotone transitions)**.**

(i) A Markov transition m on S is monotone if and only if $\mu m \geq_{\mathrm{SD}} \nu m$ for all measures μ and ν on S satisfying $\mu \geq_{\mathrm{SD}} \nu$.
(ii) The composition of monotone Markov transitions is monotone.

2.A.3 Measurability and universal completions

Our definition of θ-discernment includes an inequality for every type. To ensure that we can select score conversions in a measurable way, we enlarge the Borel σ-algebra to its universal completion, which we introduce here.

Let $(X, \mathscr{X})$ and $(Y, \mathscr{Y})$ be measurable spaces. A function f from X to Y is $\mathscr{X}/\mathscr{Y}$-measurable if $f^{-1}(B)$ is in $\mathscr{X}$ for all B in $\mathscr{Y}$. This condition is written more compactly as $f^{-1}(\mathscr{Y}) \subset \mathscr{X}$. If the σ-algebra $\mathscr{Y}$ is understood, we say f is $\mathscr{X}$-measurable, and if both σ-algebras are understood, we say that f is measurable.

Let $(X, \mathscr{X}, \mu)$ be a probability space. A set A in $\mathscr{X}$ is a *μ-null set* if $\mu(A) = 0$. The *μ-completion* of σ-algebra $\mathscr{X}$, denoted $\overline{\mathscr{X}}_\mu$, is the smallest σ-algebra that contains every set in $\mathscr{X}$ and every subset of every μ-null set. It is straightforward to check that a subset A of X is a member of $\overline{\mathscr{X}}_\mu$ if and only if there are sets A_1 and A_2 in $\mathscr{X}$ such that $A_1 \subset A \subset A_2$ and $\mu(A_2 \setminus A_1) = 0$. The *universal completion* $\overline{\mathscr{X}}$ of $\mathscr{X}$ is the σ-algebra on X defined by

$$\overline{\mathscr{X}} = \cap_\mu \overline{\mathscr{X}}_\mu,$$

where the intersection is taken over all probability measures on $(X, \mathscr{X})$.

It is convenient to work with the universal completion because of the following measurable projection theorem (Cohn, 2013, Proposition 8.4.4, p. 264).

Theorem 14 (Measurable projection)
Let $(X, \mathscr{X})$ be a measurable space, Y a Polish space, and C a set in the product σ-algebra $\mathscr{X} \otimes \mathscr{B}(Y)$. Then the projection of C on X belongs to $\overline{\mathscr{X}}$.

By taking universal completions, we do not lose any Markov transitions.

Lemma 11 (Completing transitions). A Markov transition k from $(X, \mathscr{X})$ to $(Y, \mathscr{Y})$ can be uniquely extended to a Markov transition $\bar{k}$ from $(X, \overline{\mathscr{X}})$ to $(Y, \overline{\mathscr{Y}})$.

Next we consider the universal completion of a product σ-algebra.

Lemma 12 (Product spaces). For measurable spaces $(X, \mathscr{X})$ and $(Y, \mathscr{Y})$,

$$\overline{\mathscr{X}} \otimes \overline{\mathscr{Y}} \subset \overline{\mathscr{X} \otimes \mathscr{Y}} = \overline{\overline{\mathscr{X}} \otimes \overline{\mathscr{Y}}}.$$

We adopt the convention that when a Markov transition is defined between Polish spaces, both the domain and codomain are endowed with the universal completions of their Borel σ-algebras. In particular, when we take a product of Markov transitions, we extend this transition. By Lemma 12, it does not matter whether the component transitions are extended first.

We conclude with one note of caution about universal completions. Let k be a Markov transition k from $(X \times Y, \overline{\mathscr{X} \otimes \mathscr{Y}}$ to $(Z, \mathscr{Z})$. Since $\overline{\mathscr{X} \otimes \mathscr{Y}}$ is generally larger than $\overline{\mathscr{X}} \otimes \overline{\mathscr{Y}}$, the section k_x may not be a Markov transition. For fixed $x \in X$ and $B \in \mathscr{Z}$, the map $y \mapsto k(x, y, B)$ is defined, but may not be $\mathscr{Y}$-measurable. But if Y is countable, and $\mathscr{Y} = 2^Y$, then measurability is automatic.

2.A.4 Defining mechanisms and strategies

In the model, we make the following technical assumptions. The sets Θ, T, and X are all Polish spaces. The finite signal space S is endowed with the discrete topology. The Markov transition π is from $(\Theta \times T, \mathscr{B}(\Theta \times T))$ to $(S, \mathscr{B}(S))$. Recall that for Polish spaces Y and Z, we have $\mathscr{B}(Y \times Z) = \mathscr{B}(Y) \otimes \mathscr{B}(Z)$. In a mechanism, the message space M is Polish. Testing rules, decision rules, reporting strategies, and passage strategies are all Markov transitions, with the domain and codomain endowed with the universal completions of their Borel (product) σ-algebras.

Appendix 2.B Proofs

2.B.1 Proof of Theorem 8

We begin with new notation. To avoid confusion in the commutative diagrams below, we denote the message set in a direct mechanism by Θ'. This set is a copy of Θ, but it is helpful to keep the sets Θ and Θ' distinct.

We work with performance strategies, as defined in section 2.8. Let f be an implementable social choice function. Select a profile $(M, t, g; r, p)$ that implements f. We construct a direct mechanism $(\hat{t}, \hat{g})$ that canonically implements f.

Let $\hat{t}$ be the composition rt, which is a Markov transition from Θ' to T. By disintegration of Markov transitions (Kallenberg, 2017, Theorem 1.25, p. 39), there is a Markov transition h from $\Theta' \times T$ to M such that $r \times t = \hat{t} \otimes h$.[25]

Define a Markov transition d from $\Theta' \times M \times T \times S$ to S as follows. For each $(\theta', m, \tau) \in \Theta' \times M \times T$, set

$$d_{\theta',m,\tau} = \tilde{F}_{\tau|\theta'} \tilde{Q}_{\theta',m,\tau},$$

where $\tilde{F}_{\tau|\theta'}$ is the distribution Markov transition corresponding to $\pi_{\tau|\theta'}$ and $\tilde{Q}_{\theta',m,\tau}$ is the quantile Markov transition corresponding to $p_{\theta',m,\tau}$; see section 2.A.2 for the definitions. By Lemma 9,

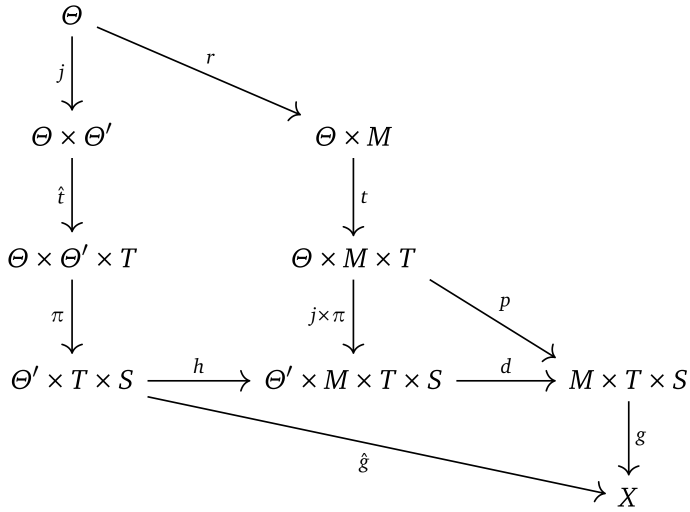
$$\pi_{\tau|\theta'} d_{\theta',m,\tau} = p_{\theta',m,\tau}, \tag{2.B.1}$$

and $d_{\theta',m,\tau}$ is downward because $p_{\theta',m,\tau} \leq_{\mathrm{SD}} \pi_{\tau|\theta'}$.

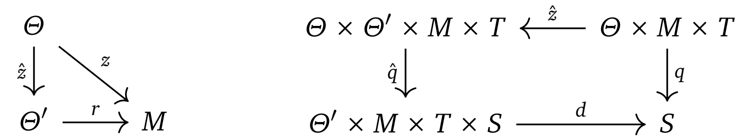
Define the direct decision rule $\hat{g}$ as the composition shown in the following commutative diagram, where j denotes the identity transition from Θ to Θ'.[26]

25. We cannot directly apply the result to the universal completions of the Borel σ-algebras. Argue as follows. First, restrict $r \times t$ to a Markov transition from $(\Theta, \overline{\mathscr{B}(\Theta')})$ to $(M \times T, \mathscr{B}(M) \otimes \mathscr{B}(T))$ and $\hat{t}$ to a Markov transition from $(\Theta', \overline{\mathscr{B}(\Theta')})$ to $(T, \mathscr{B}(T))$. By Kallenberg (2017, Theorem 1.25, p. 39), there exists a Markov transition h from $(\Theta' \times T, \overline{\mathscr{B}(\Theta') \otimes \mathscr{B}(T)})$ to $(M, \mathscr{B}(M))$ that satisfies the desired equality for the restricted transitions from $(\Theta', \mathscr{B}(\Theta'))$ to $(M \times T, \mathscr{B}(M) \otimes \mathscr{B}(T))$. By Lemma 11, we can extend h to a Markov transition from $(\Theta \times T, \overline{\mathscr{B}(\Theta) \otimes \mathscr{B}(T)})$ to $(M, \overline{\mathscr{B}(M)})$.

26. To keep the diagrams uncluttered, we adopt the following conventions. The labeled Markov transition maps a subproduct of the source space into a subproduct of the target space. The other sets

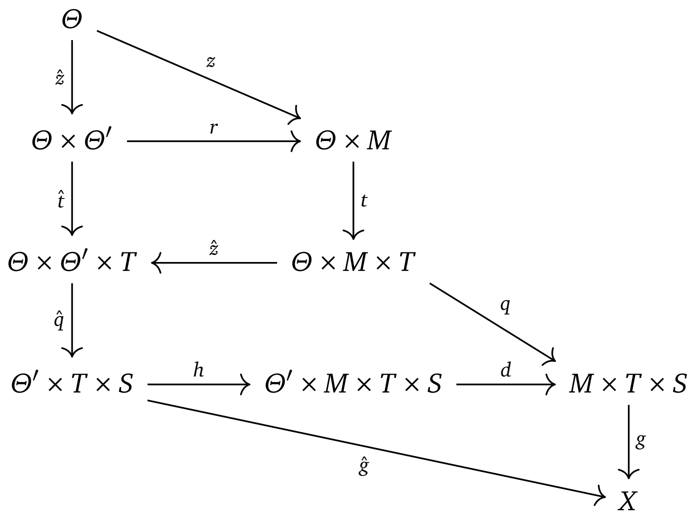


By construction, this diagram commutes, so the direct mechanism $(\hat{t}, \hat{g})$ induces f. To prove incentive compatibility, we show that for any strategy $(\hat{z}, \hat{q})$ in the direct mechanism, there is a strategy (z, q) in the original mechanism that induces the same social choice function. Given $(\hat{z}, \hat{q})$, define (z, q) by the following commutative diagrams.



(2.B.2)

Since each Markov transition $d_{\theta',m,\tau}$ is downward, it follows that q is feasible. These deviations induce the same social choice function because the following diagram commutes.



in the target space are carried along by the identity. Mechanisms act on the report set Θ' not the type space Θ; π acts on Θ.

Extension to multiple agents. Following the notation in section 2.7, let $\Theta = \prod_i \Theta_i$, $T = \prod_i T_i$, and $S = \prod_i S_i$. Set $\pi = \otimes_i \pi$. With this notation, π is a Markov transition from $T \times \Theta$ to S, as in the single-agent case. After setting $M = \prod_i M_i$, a mechanism in the multi-agent setting maps between the same sets. Similarly for the agent's strategies, set $r = \otimes_i r_i$ and $p = \otimes_i p_i$. With these new definitions, construct h as before. Define d_i separately for each agent i and then set $d = \otimes_i d_i$. The proof is exactly as before. To consider a deviation by a fixed player j, take $\hat{z} = \hat{z}_j \otimes r_{-j}$ and $\hat{q} = \hat{q}_j \otimes p_{-j}$. Then the deviation (z, q) constructed in (2.B.2) will have the form $z = z_j \otimes r_{-j}$ and $q = q_j \otimes p_{-j}$.

2.B.2 Proof of Theorem 9

We begin with some notation. As in the proof of Theorem 8, denote by Θ' the message set in a direct mechanism. Thus, Θ' is a copy of Θ. Similarly, denote by T' a copy of T that will be the codomain of a most-discerning testing function. Keeping these copies separate will be helpful for the commutative diagrams below.

The proof is organized as follows. First, we select score conversions in a measurable way. Next, we prove sufficiency and then necessity.

Selecting score conversions. Let $\mathscr{K}$ denote the space $\Delta(S)^S$ of Markov transitions on S, viewed as a subset of $\boldsymbol{R}^{S \times S}$, with the usual Euclidean topology and inner product $\langle \cdot, \cdot \rangle$. For $k \in \mathscr{K}$, denote by $k(s, s')$ the transition probability from s to s'.

Define the domain

$$D = \{(\theta, \tau, \psi) \in \Theta' \times T' \times T : \tau \succeq_\theta \psi\}.$$

Define the correspondence $K\colon D \twoheadrightarrow \mathscr{K}$ by putting $K(\theta, \tau, \psi)$ equal to the set of monotone Markov transitions k in $\mathscr{K}$ satisfying (i) $\pi_{\tau|\theta} k = \pi_{\psi|\theta}$, and (ii) $\pi_{\tau|\theta'} k \leq_{\mathrm{SD}} \pi_{\psi|\theta'}$ for all types θ'. By the choice of domain D, the correspondence K is nonempty-valued.

Endow D with the restriction of the σ-algebra $\overline{\mathscr{B}(\Theta' \times T' \times T)}$. To prove that there exists a measurable selection $\hat{k}$ from K, we apply the Kuratowski–Ryll-Nardzewski selection theorem (Aliprantis and Border, 2006, 18.13, p. 600). The correspondence K has compact convex values, so it suffices to check that associated support functions for K are measurable (Aliprantis and Border, 2006, 18.31, p. 611).

Fix $\ell \in \boldsymbol{R}^{S \times S}$. Define the map $C\colon D \to \boldsymbol{R}$ by

$$C(\theta, \tau, \psi) = \max_{k \in K(\theta, \tau, \psi)} \langle k, \ell \rangle.$$

It suffices to show that C is $\overline{\mathscr{B}(\Theta' \times T' \times T)}$-measurable. Define a sequence of auxiliary functions $C_m\colon D \times (\Theta')^m \to \boldsymbol{R}$ as follows. Let $C_m(\theta, \tau, \psi, \theta_1', \dots, \theta_m')$ be the value of the program

$$\begin{aligned} \text{maximize} \quad & \langle k, \ell \rangle \\ \text{subject to} \quad & k \in \mathscr{K} \\ & k \text{ is monotone} \\ & \pi_{\tau|\theta} k = \pi_{\psi|\theta} \\ & \pi_{\tau|\theta'_j} k \leq_{\text{SD}} \pi_{\psi|\theta'_j}, \quad j = 1, \ldots, m. \end{aligned}$$

This is a standard linear programming problem with a compact feasible set. By Berge's theorem (Aliprantis and Border, 2006, 17.30, p. 569), the value of the linear program is upper semicontinuous (and hence Borel) as a function of the coefficients appearing in the constraints. Since π is Borel, so is each function C_m. By the measurable projection theorem (Theorem 14), each map

$$(\theta, \tau, \psi) \mapsto \inf_{\boldsymbol{\theta}' \in (\Theta')^m} C_m(\theta, \tau, \psi, \boldsymbol{\theta}')$$

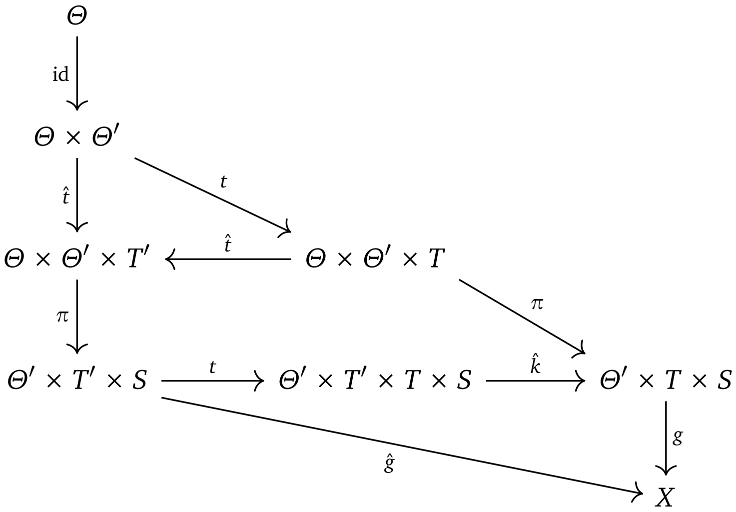
is $\overline{\mathscr{B}(\Theta' \times T' \times T)}$-measurable. A compactness argument shows that[27]

$$C(\theta, \tau, \psi) = \inf_m \inf_{\boldsymbol{\theta}' \in (\Theta')^m} C_m(\theta, \tau, \psi, \boldsymbol{\theta}'),$$

so C is also $\overline{\mathscr{B}(\Theta' \times T' \times T)}$-measurable.

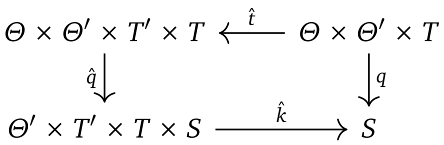
Sufficiency. Fix a decision environment (X, u) and let f be an implementable social choice function. By the revelation principle (Theorem 8), there is a direct mechanism (t, g) that canonically implements f. We now construct a decision rule $\hat{g}$ such that the direct mechanism $(\hat{t}, \hat{g})$ canonically implements f. Define $\hat{g}$ by the following commutative diagram, where j denotes the identity transition from Θ to Θ'.

27. We claim that for each positive ε there exists a natural number m and a vector $\boldsymbol{\theta}' \in (\Theta')^m$ such that $C_m(\theta, \tau, \psi, \boldsymbol{\theta}') < C(\theta, \tau, \psi) + \varepsilon$. Suppose not. For each $\theta' \in \Theta'$, let $K_{\theta'}$ be the compact set of monotone Markov transitions $k \in \mathscr{K}$ satisfying (i) $\pi_{\tau|\theta} k = \psi_{\psi|\theta}$, (ii) $\pi_{\tau|\theta'} k \leq_{\text{SD}} \pi_{\psi|\theta'}$, and (iii) $\langle k, \ell \rangle \geq C(\theta, \tau, \psi) + \varepsilon$. This family has the finite intersection property, but the intersection over all $\theta' \in \Theta'$ is empty, which is a contradiction.



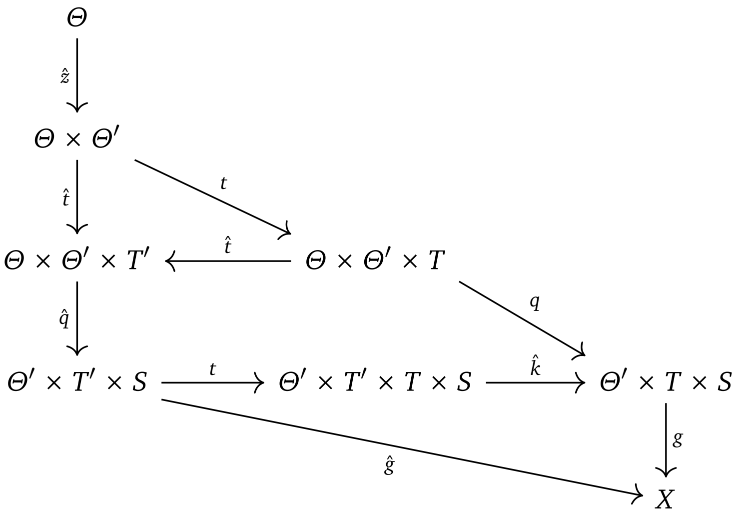
(2.B.3)

Since the diagram commutes, the direct mechanism $(\hat{t}, \hat{g})$ induces f. For incentive compatibility, we show that for any strategy $(\hat{z}, \hat{q})$ in the new mechanism, there is a strategy (z, q) in the old mechanism inducing the same social choice function. Given $(\hat{z}, \hat{q})$, set $z = \hat{z}$ and define q by the following commutative diagram.



(2.B.4)

Since each transition $\hat{k}_{\theta,\tau,\psi}$ is downward, it follows that q is feasible. These deviations induce the same social choice function because the following diagram commutes.



(2.B.5)

Necessity. If $\hat{t}$ is not most-discerning, then there exists a fixed type θ such that $\hat{t}(\theta)$ is not most θ-discerning. Set $\tau = \hat{t}(\theta)$. Select a test ψ such that $\tau \not\succeq_\theta \psi$.

Define the decision set

$$X = \{x_{\theta'} : \theta' \in \Theta\} \cup \{\underline{x}\},$$

where $\underline{x}$, $x_{\theta'}$, and $x_{\theta''}$ are distinct for all distinct types θ' and θ''.

Define utilities as follows. Type θ gets utility 1 from x_θ and utility 0 from every other decision in X. For $\theta' \neq \theta$, type θ' gets utility 1 from decision x_θ, utility $\pi(\psi|\theta')$ from $x_{\theta'}$, and utility 0 from all other decisions.

Let f be the social choice function that assigns to each type θ' with $\theta' \neq \theta$ the decision $x_{\theta'}$ with certainty, and assigns to type θ decision x_θ with probability $\pi(\psi|\theta)$ and decision $\underline{x}$ with probability $1 - \pi(\psi|\theta)$. Then f can be canonically implemented by the mechanism $(t, \hat{g})$, where t is any function satisfying $t(\theta) = \psi$ and $\hat{g}$ is the decision rule specified as follows. If the agent reports $\theta' \neq \theta$, assign $x_{\theta'}$ no matter the test result; if the agent reports θ, select x_θ if the agent passes test ψ and $\underline{x}$ if the agent fails test ψ.

We claim that f cannot be canonically implemented with the testing function $\hat{t}$. Suppose for a contradiction that there is a decision rule $\hat{g}$ such that f is canonically implemented by the mechanism $(\hat{t}, \hat{g})$. We separate into two cases.

First suppose $\pi(\tau|\theta) = 0$. Then every type can get the good with probability $\pi(\psi|\theta)$. Since $\tau \not\succeq_\theta \psi$, there is some type $\theta' \neq \theta$ such that $\pi(\psi|\theta') < \pi(\psi|\theta)$, so type θ' has a profitable deivation.

Next suppose $\pi(\tau|\theta) > 0$. Define a Markov transition k on S, represented by a vector in $[0,1]^2$ by letting $k(s)$ be the probability that the measure $g_{\theta,\tau,s}$ places on decision x_θ. Then $\pi_{\tau|\theta} k = \pi_{\psi|\theta}$. Since $\tau \not\succeq_\theta \psi$, either k is not monotone, in which case type θ can profitably deviate by reporting type θ and failing the test, or there is some type $\theta' \neq \theta$ such that $\pi_{\tau|\theta'} k >_{\text{SD}} \pi_{\psi|\theta'}$. In this case, type θ' can profitably deviate by reporting θ' and exerting effort.

2.B.3 Proof of Theorem 10

First we use the regularity assumptions to prove that there exists a measurable test selection. Then we follow the proof of the sufficiency part of Theorem 9.

Measurable test selection. We prove that there exists a measurable function $\bar{t}$ from $(\Theta' \times T, \overline{\mathscr{B}(\Theta' \times T)})$ to $(T', \mathscr{B}(T'))$ such that the test $\bar{t}(\theta, \psi)$ is in $\hat{T}(\theta)$ and satisfies $\bar{t}(\theta, \psi) \succeq_\theta \psi$, for each $\theta \in \Theta$ and $\psi \in T$. Define a correspondence $H \colon \Theta' \times T \twoheadrightarrow T'$ by

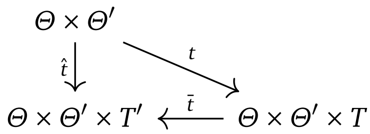
$$H(\theta, \psi) = \{\tau \in T : \tau \succeq_\theta \psi\}.$$

Since π is continuous, the graph of H is closed in $\Theta' \times T \times T'$.[28] Since the graph of $\hat{T}$ is Borel, so is the set $\{(\theta, \psi, \tau) : \tau \in \hat{T}(\theta)\}$ and also the intersection

$$\{(\theta, \psi, \tau) \ : \ \tau \ \succeq_\theta \ \psi \text{ and } \tau \ \in \ \hat{T}(\theta)\}.$$

By the measurable projection theorem (Theorem 14), the associated correspondence from $(\Theta' \times T, \overline{\mathscr{B}(\Theta' \times T)})$ to T' is measurable. Moreover, this correspondence has closed values, so we can apply the Kuratowski–Ryll-Nardzewksi selection theorem Aliprantis and Border (2006, 18.13, p. 600) to obtain the desired function $\bar{t}$.

Sufficiency. With $\bar{t}$ in hand, the proof is almost the same as the proof of Theorem 9 in section 2.B.2, but we are not given $\hat{t}$. Instead, $\hat{t}$ is defined by the commutative diagram. The second and third rows of (2.B.3) and (2.B.5) become



In (2.B.4), put $\bar{t}$ in place of $\hat{t}$. The rest of the proof is completed as above.

2.B.4 Proof of Proposition 4

A Markov transition k on $\{0, 1\}$ can be represented as a vector

$$(k(0), k(1)) \ \in \ [0, 1]^2,$$

where $k(s)$ is the probability of transitioning from $s \in \{0, 1\}$ to 1. A Markov transition k is monotone and satisfies $\pi_{\tau|\theta} k = \pi_{\psi|\theta}$ if and only if the vector $(k(0), k(1))$ satisfies

$$k(1) \ \geq \ k(0) \quad \text{and} \quad \pi(\tau|\theta)k(1) + (1 - \pi(\tau|\theta))k(0) \ = \ \pi(\psi|\theta).$$

We separate the solution into cases.

(1) If $\pi(\tau|\theta) \geq \pi(\psi|\theta)$, then the solutions are given by

$$\begin{bmatrix} k(0) \\ k(1) \end{bmatrix} \ = \ \lambda \begin{bmatrix} 0 \\ \pi(\psi|\theta)/\pi(\tau|\theta) \end{bmatrix} + (1 - \lambda) \begin{bmatrix} \pi(\psi|\theta) \\ \pi(\psi|\theta) \end{bmatrix},$$

for $\lambda \in [0, 1]$, provided that we use the convention that $0/0$ equals 1.

28. Take a sequence $(\theta_n, \psi_n, \tau_n)$ in $\operatorname{gr} H$ converging to a limit (θ, ψ, τ) in $\Theta' \times T \times T'$. For each n, there is a monotone Markov transition k_n on S such that (i) $\pi_{\tau_n|\theta_n} k_n = \pi_{\psi_n|\theta_n}$, and (ii) $\pi_{\tau_n|\theta'} k_n \leq_{\mathrm{SD}} \pi_{\psi_n|\theta'}$ for all $\theta' \in \Theta$. The space of Markov transitions on S is compact, so after passing to a subsequence, we may assume that k_n converges to a limit k, which must be monotone. Since π is continuous, taking limits gives (i) $\pi_{\tau|\theta} k = \pi_{\psi|\theta}$, and (ii) $\pi_{\tau|\theta'} k \leq_{\mathrm{SD}} \pi_{\psi|\theta'}$ for each $\theta' \in \Theta$. Therefore, (θ, ψ, τ) is in $\operatorname{gr} H$.

(2) If $\pi(\tau|\theta) < \pi(\psi|\theta)$, then the solutions are given by

$$\begin{bmatrix} k(0) \\ k(1) \end{bmatrix} = \lambda \begin{bmatrix} \bar{\pi}(\psi|\theta)/\bar{\pi}(\tau|\theta) \\ 1 \end{bmatrix} + (1-\lambda) \begin{bmatrix} \pi(\psi|\theta) \\ \pi(\psi|\theta) \end{bmatrix},$$

for $\lambda \in [0,1]$.

In each case, the left term is the vector representation of $\tilde{F}_{\tau|\theta}\tilde{Q}_{\tau|\theta}$ and the right term is the vector representation of the constant Markov transition $\pi_{\psi|\theta}$.

2.B.5 Proof of Proposition 5

Fix a type θ and tests τ and ψ. If $k = \lambda \tilde{F}_{\tau|\theta}\tilde{Q}_{\tau|\theta} + (1-\lambda)\pi_{\psi|\theta}$, then by Lemma 8,

$$\pi_{\tau|\theta'}k = \lambda \pi_{\tau|\theta'}\tilde{F}_{\tau|\theta}\tilde{Q}_{\psi|\theta} + (1-\lambda)\pi_{\psi|\theta} = \big(\lambda \pi_{\tau|\theta'} + (1-\lambda)\pi_{\tau|\theta}\big)\tilde{F}_{\tau|\theta}\tilde{Q}_{\psi|\theta}.$$

Therefore, by Proposition 4, we have $\tau \succeq_\theta \psi$ if and only if there exists $\lambda \in [0,1]$ such that

$$\big(\lambda \pi_{\tau|\theta'} + (1-\lambda)\pi_{\tau|\theta}\big)\tilde{F}_{\tau|\theta}\tilde{Q}_{\psi|\theta} \le_{\mathrm{SD}} \pi_{\psi|\theta'}, \tag{2.B.6}$$

for all $\theta' \in \Theta$. Now work with cases.

(1) If $\pi(\tau|\theta) = \pi(\psi|\theta) = 0$, the result is clear.

(2) If $\pi(\tau|\theta) \ge \pi(\psi|\theta)$ and $\pi(\tau|\theta) > 0$, we compare the probability of passage. The inequality holds if and only if

$$[\lambda \pi(\tau|\theta') + (1-\lambda)\pi(\tau|\theta)]\frac{\pi(\psi|\theta)}{\pi(\tau|\theta)} \le \pi(\psi|\theta'),$$

This reduces to the desired inequality.

(3) If $\pi(\tau|\theta) < \pi(\psi|\theta)$, we compare the probability of failure.The inequality holds if and only if

$$[\lambda \bar{\pi}(\tau|\theta') + (1-\lambda)\bar{\pi}(\tau|\theta)]\frac{\bar{\pi}(\psi|\theta)}{\bar{\pi}(\tau|\theta)} \ge \bar{\pi}(\psi|\theta').$$

2.B.6 Proof of Proposition 6

Fix a type θ and tests τ_1 and τ_2.

One direction is clear. If τ_1 and τ_2 are equal, then τ_1 and τ_2 are clearly θ-equivalent. If type θ is minimal on τ_1 and τ_2, take $\lambda = 1$ in Proposition 5 to see that τ_1 and τ_2 are θ-equivalent.

For the other direction, suppose τ_1 and τ_2 are θ-equivalent. Without loss, we may assume $\pi(\tau|\theta) \geq \pi(\psi|\theta)$. We separate into cases according to whether θ is minimal on τ.

First suppose θ is minimal on τ_1. Since $\tau_1 \succeq_\theta \tau_2$, there are probabilities $k(0)$ and $k(1)$ such that for all types θ',

$$\begin{aligned}\pi(\tau_2|\theta) &= k(0) + (k(1) - k(0))\pi(\tau_1|\theta) \\ &\leq k(0) + (k(1) - k(0))\pi(\tau_1|\theta') \\ &\leq \pi(\tau_2|\theta').\end{aligned}$$

Now suppose θ is not minimal for τ, so there exists some type θ'' such that $\pi(\tau|\theta'') < \pi(\tau|\theta)$. In particular, $\pi(\tau|\theta) > 0$.

Claim. $\pi(\tau|\theta) = \pi(\psi|\theta)$.

Suppose for a contradiction that $\pi(\tau_1|\theta) > \pi(\tau_2|\theta)$. By Proposition 5, there are constants λ_1 and λ_2 in $[0, 1]$ such that for all types θ',

$$\begin{aligned}[\lambda_1\pi(\tau_1|\theta') + (1 - \lambda_1)\pi(\tau_1|\theta)]\pi(\tau_2|\theta) &\leq \pi(\tau_2|\theta')\pi(\tau_1|\theta), \\ [\lambda_2\bar{\pi}(\tau_2|\theta') + (1 - \lambda_2)\bar{\pi}(\tau_2|\theta)]\bar{\pi}(\tau_1|\theta) &\geq \bar{\pi}(\tau_1|\theta')\bar{\pi}(\tau_2|\theta).\end{aligned}$$

With $\theta' = \theta''$ the first inequality is weakest when $\lambda_1 = 1$, so

$$\pi(\tau_1|\theta'')\pi(\tau_2|\theta) \leq \pi(\tau_2|\theta'')\pi(\tau_1|\theta). \tag{2.B.7}$$

Taking $\lambda_2 = 0$ in the second inequality, and noting that $\bar{\pi}(\tau_2|\theta) = 1 - \pi(\tau_2|\theta) > 0$, yields the contradiction $\bar{\pi}(\tau_1|\theta) \geq \bar{\pi}(\tau_1|\theta'')$. Therefore, the inequality must hold with $\lambda_2 = 1$, so

$$\bar{\pi}(\tau_2|\theta'')\bar{\pi}(\tau_2|\theta) \geq \bar{\pi}(\tau_2|\theta)\bar{\pi}(\tau_1|\theta''). \tag{2.B.8}$$

We show that (2.B.7) and (2.B.8) are incompatible. In (2.B.8), subtract $\bar{\pi}(\tau_1|\theta)\bar{\pi}(\tau_2|\theta)$ from both sides to get

$$[\bar{\pi}(\tau_2|\theta'') - \bar{\pi}(\tau_2|\theta)]\bar{\pi}(\tau_1|\theta) \geq [\bar{\pi}(\tau_1|\theta'') - \bar{\pi}(\tau_1|\theta)]\bar{\pi}(\tau_2|\theta),$$

which is equivalently,

$$[\pi(\tau_2|\theta) - \pi(\tau_2|\theta'')]\bar{\pi}(\tau_1|\theta) \geq [\pi(\tau_1|\theta) - \pi(\tau_1|\theta'')]\bar{\pi}(\tau_2|\theta).$$

The right side is strictly positive and $\bar{\pi}(\tau_2|\theta) < \bar{\pi}(\tau_1|\theta)$, so

$$\pi(\tau_2|\theta) - \pi(\tau_2|\theta'') \;>\; \pi(\tau_1|\theta) - \pi(\tau_1|\theta''). \tag{2.B.9}$$

Now negate (2.B.7) and add $\pi(\tau_1|\theta)\pi(\tau_2|\theta)$ to both sides to obtain

$$[\pi(\tau_1|\theta) - \pi(\tau_1|\theta'')]\pi(\tau_2|\theta) \;\geq\; [\pi(\tau_2|\theta) - \pi(\tau_2|\theta'')]\pi(\tau_1|\theta).$$

But $\pi(\tau_2|\theta) < \pi(\tau_1|\theta)$, so (2.B.9) gives the opposite inequality.

With the claim established, we now complete the proof. By Proposition 5 there are constants λ_1 and λ_2 in $[0, 1]$ such that for all types θ',

$$\begin{aligned} [\lambda_1\pi(\tau_1|\theta') + (1-\lambda_1)\pi(\tau_1|\theta)]\pi(\tau_2|\theta) \;&\leq\; \pi(\tau_2|\theta')\pi(\tau_1|\theta), \\ [\lambda_2\pi(\tau_2|\theta') + (1-\lambda_2)\pi(\tau_2|\theta)]\pi(\tau_1|\theta) \;&\leq\; \pi(\tau_1|\theta')\pi(\tau_2|\theta). \end{aligned}$$

After cancelling the common value of $\pi(\tau_1|\theta)$ and $\pi(\tau_2|\theta)$, which is nonzero by assumption, we have

$$\lambda_1\pi(\tau_1|\theta') + (1-\lambda_1)\pi(\tau_1|\theta) \;\leq\; \pi(\tau_2|\theta'), \tag{2.B.10}$$
$$\lambda_2\pi(\tau_2|\theta') + (1-\lambda_2)\pi(\tau_2|\theta) \;\leq\; \pi(\tau_1|\theta'). \tag{2.B.11}$$

It suffices to show that $\lambda_1 = \lambda_2 = 1$, for then $\pi(\tau_1|\theta') = \pi(\tau_2|\theta')$ for all types θ'. Take $\theta' = \theta''$ in both inequalities. We have $\pi(\tau_2|\theta) = \pi(\tau_1|\theta) > \pi(\tau_1|\theta'')$, so (2.B.11) implies that $\pi(\tau_2|\theta'') \leq \pi(\tau_1|\theta'')$. Substituting this into (2.B.10), we get $\lambda_1 = 1$. Then $\pi(\tau_1|\theta'') \leq \pi(\tau_2|\theta'')$, so (2.B.11) implies $\lambda_2 = 1$.

2.B.7 Proof of Proposition 7

We simply translate Proposition 6 into the language of authentication rates. Suppose $\hat{t}_1$ and $\hat{t}_2$ are most-discerning testing functions, and let α_1 and α_2 be the induced authentication rates. For each type θ, we know $\hat{t}_1(\theta)$ and $\hat{t}_2(\theta)$ are most θ-discerning tests. Apply Proposition 6 and translate the conclusion into the language of authentication rates.

2.B.8 Proof of Theorem 11

Let α be an authentication rate. First, suppose that α is most discerning. Let $T = \{\tau_{\theta'} : \theta' \in \Theta\}$, and define the passage rate π by $\pi(\tau_{\theta'}|\theta) = \alpha(\theta'|\theta)$ for all types θ and θ'. Combining definition 11 and Proposition 5, we see that the testing function $\theta \mapsto \tau_\theta$ is most discerning. By construction, this testing function induces α.

Now suppose α is induced by a most-discerning testing function $\hat{t}$ in a testing environment (T, π). Substitute the equality $\alpha(\theta'|\theta) = \pi(\hat{t}(\theta'), \theta)$ into Proposition 5 to conclude that α is most discerning.

2.B.9 Proof of Proposition 8

We apply the dominated convergence theorem. As λ converges to 0 pointwise, $\Lambda(z|\theta)$ converges to 1 for all z and θ with $z \leq \theta$. Hence $\varphi(\theta)$ converges to $\varphi^M(\theta)$, for each θ. Likewise, as λ converges to ∞ pointwise, $\Lambda(z|\theta)$ converges to 0 for all z and θ with $z < \theta$. Hence $\varphi(\theta)$ converges to θ, for each θ.

2.B.10 Proof of Proposition 9

We will prove a stronger result that we use below. Define functions λ_+ and λ_- from Θ to $[0, \infty)$ by

$$\lambda_+(\theta) \;=\; -D_{2+}\alpha(\theta|\theta), \qquad \lambda_-(\theta) \;=\; D_{2-}\alpha(\theta|\theta).$$

In the main text, we only work with λ_+, which is denoted λ. Extend the function Λ to $\Lambda\colon \Theta \times \Theta \to [0,1]$ by

$$\Lambda(\theta'|\theta) \;=\; \begin{cases} \exp\left(-\int_{\theta'}^{\theta} \lambda_+(s)\,\mathrm{d}s\right) & \text{if } \theta \;\geq\; \theta', \\ \exp\left(-\int_{\theta}^{\theta'} \lambda_-(s)\,\mathrm{d}s\right) & \text{if } \theta \;<\; \theta'. \end{cases}$$

With these definitions, we now prove that $\alpha(\theta'|\theta) \geq \Lambda(\theta'|\theta)$ all types θ' and θ. Fix θ and θ'. For each h, transitivity gives

$$\alpha(\theta'|\theta + h) \;\geq\; \alpha(\theta'|\theta)\alpha(\theta|\theta + h).$$

Subtract $\alpha(\theta'|\theta)$ from each side to get

$$\begin{aligned} \alpha(\theta'|\theta + h) - \alpha(\theta'|\theta) \;&\geq\; \alpha(\theta'|\theta)(\alpha(\theta|\theta + h) - 1)) \\ &=\; \alpha(\theta'|\theta)[\alpha(\theta + h, \theta) - \alpha(\theta|\theta)]. \end{aligned}$$

Dividing by h and passing to the limit as $h \downarrow 0$ and $h \uparrow 0$ gives

$$D_{2+}\alpha(\theta'|\theta) \;\geq\; -\lambda_+(\theta)\alpha(\theta'|\theta) \quad \text{and} \quad D_{2-}\alpha(\theta'|\theta) \;\leq\; \lambda_-(\theta)\alpha(\theta'|\theta).$$

Now we use absolute continuity to convert these local bounds into global bounds. Fix a report θ'. Define the function Δ on $[\underline{\theta}, \bar{\theta}]$ by

$$\Delta(\theta) \;=\; \frac{\alpha(\theta'|\theta)}{\Lambda(\theta'|\theta)}.$$

By construction, $\Delta(\theta') = 1$. We will argue that $\Delta(\theta) \geq 1$ for all θ. For $\theta' < \theta$, if $D_{2+}\Lambda(\theta'|\theta)$ exists, then

$$D_+\Delta(\theta) \;=\; \frac{1}{\Lambda(\theta'|\theta)}\big(D_+\alpha(\theta'|\theta) + \lambda_+(\theta)\alpha(\theta'|\theta)\big) \;\geq\; 0.$$

For $\theta' > \theta$, if $D_{2-}\Lambda(\theta'|\theta)$ exists, then

$$D_{-}\Delta(\theta) \;=\; \frac{1}{\Lambda(\theta'|\theta)}\big(D_{-}\alpha(\theta'|\theta) - \lambda_{-}(\theta)\alpha(\theta'|\theta)\big) \;\leq\; 0,$$

Since $\Lambda(\theta'|\cdot)$ is absolutely continuous, these inequalities hold almost surely. Moreover, the product of absolutely continuous functions on a compact set is absolutely continuous, so Δ is absolutely continuous, and hence the fundamental theorem of calculus gives $\Delta(\theta) \geq 1$, as desired.

2.B.11 Proof of Proposition 10

First we introduce notation. For a given quantity function q, there is a one-to-one correspondence between the transfer function t and the utility function U, given by $U(\theta) = \theta q(\theta) - t(\theta)$. We will interchangeably refer to such a mechanism as (q, t) or (q, U). Let

$$u(\theta'|\theta) \;=\; \alpha(\theta'|\theta)[\theta q(\theta') - t(\theta')], \qquad U(\theta) \;=\; u(\theta|\theta) \;=\; \max_{\theta'\in\Theta} u(\theta'|\theta).$$

Lemma 13 (Utility bound)**.** Let q be a bounded quantity function, and let U be a utility function. If (q, U) is incentive compatible, then for every type θ, we have

$$U(\theta) \;\geq\; \int_{\underline{\theta}}^{\theta} \Lambda(z|\theta) q(z)\,\mathrm{d}z. \tag{2.B.12}$$

If the function $\theta \mapsto \Lambda(\theta|\bar{\theta})q(\theta)$ is increasing and the global upper bound is satisfied, it is incentive compatible for (2.B.12) to hold with equality.

Lemma 13 is proved in section 2.B.12. Here we prove the theorem, taking Lemma 13 as given. There is no loss in restricting attention to bounded quantity functions.[29] Pick a bounded quantity function $q\colon \Theta \to \boldsymbol{R}_{+}$. The principal's objective function can be decomposed as the difference between the total surplus and the agent's rents:

$$\int_{\underline{\theta}}^{\bar{\theta}} [\theta q(\theta) - c(q(\theta))] f(\theta)\,\mathrm{d}\theta - \int_{\underline{\theta}}^{\bar{\theta}} U(\theta)\,\mathrm{d}\theta.$$

Plug in the bound from Lemma 13 and switch the order of integration to obtain the following upper bound on the principal's objective:

$$V(q) \;=\; \int_{\underline{\theta}}^{\bar{\theta}} [\varphi(\theta) q(\theta) - c(q(\theta))] f(\theta)\,\mathrm{d}\theta.$$

29. Pick a quantity $\bar{q}$ such that $\bar{\theta}\bar{q} = c(\bar{q})$. Offering more than $\bar{q}$ will always result in weakly negative profits, so we can remove those offerings from the menu and increase the principal's revenue. Therefore, there is no loss in focusing on quantity functions that are bounded above by $\bar{q}$.

The quantity function $q^\star$ from the theorem statement maximizes the expression in brackets pointwise, and $t^\star$ is the corresponding transfer that achieves the utility bound. Since φ is increasing and the global upper bound is satisfied, this mechanism is incentive compatible. Since c' is strictly increasing, the pointwise maximizer is unique, so this quantity function $q^\star$ is unique almost everywhere.

2.B.12 Proof of Lemma 13

Let (q, t) be a bounded incentive compatible mechanism. The first step is showing that the equilibrium utility function U is absolutely continuous. Fix types θ and θ'. We have

$$\begin{aligned} U(\theta') &\geq \alpha(\theta'|\theta)\big(\theta' q(\theta) - t(\theta)\big)_+ \\ &\geq \Lambda(\theta'|\theta)\big(\theta' q(\theta) - t(\theta)\big)_+ \\ &\geq \Lambda(\theta'|\theta)\big(\theta' q(\theta) - t(\theta)\big), \end{aligned}$$

where the first inequality uses individual rationality and incentive compatibility, and the second uses the inequality between α and Λ established in section 2.B.10. Therefore,

$$\begin{aligned} U(\theta) - U(\theta') &\leq \theta q(\theta) - t(\theta) - \Lambda(\theta'|\theta)\big(\theta' q(\theta) - t(\theta)\big) \\ &= \theta q(\theta) - t(\theta) - \Lambda(\theta'|\theta)\big((\theta' - \theta) q(\theta) + \theta q(\theta) - t(\theta)\big) \\ &= (1 - \Lambda(\theta'|\theta))(\theta q(\theta) - t(\theta)) - \Lambda(\theta'|\theta)(\theta' - \theta) q(\theta) \\ &\leq (1 - \Lambda(\theta'|\theta))(\bar{\theta}\|q\|_\infty + \|t\|_\infty) + |\theta' - \theta|\|q\|_\infty. \end{aligned}$$

To avoid separating into cases according to the relative sizes of θ and θ', set $\hat{\lambda} = \lambda_+ \vee \lambda -$. Since $\hat{\lambda} \leq \lambda_+ + \lambda_-$, we know $\hat{\lambda}$ is integrable. Using the inequality $e^x \geq 1 + x$, we have

$$1 - \Lambda(\theta'|\theta) \leq 1 - \exp\Big(-\int_{\theta' \wedge \theta}^{\theta' \vee \theta} \hat{\lambda}(z)\, dz\Big) \leq \int_{\theta' \wedge \theta}^{\theta' \vee \theta} \hat{\lambda}(z)\, dz.$$

Plug this into the inequality above to get

$$U(\theta) - U(\theta') \leq C \int_{\theta' \wedge \theta}^{\theta' \vee \theta} (\hat{\lambda}(z) + 1)\, dz,$$

where $C = \max\{1, \bar{\theta}\}\|q\|_\infty + \|t\|_\infty$. Switching the roles of θ and θ' gives the same inequality, so we conclude that

$$|U(\theta) - U(\theta')| \leq C \int_{\theta' \wedge \theta}^{\theta' \vee \theta} (\hat{\lambda}(z) + 1)\, dz,$$

which proves the desired absolute continuity.

Now we use the absolute continuity of U to establish the bound. Define the auxiliary function Δ on $[\underline{\theta}, \bar{\theta}]$ by

$$\Delta(\theta) = \Lambda(\theta|\bar{\theta})\left(U(\theta) - \int_{\underline{\theta}}^{\theta} \Lambda(z|\theta)q(z)\,\mathrm{d}z\right) = \Lambda(\theta|\bar{\theta})U(\theta) - \int_{\underline{\theta}}^{\theta} \Lambda(z|\bar{\theta})q(z)\,\mathrm{d}z.$$

The function Δ is absolutely continuous since it is the product of absolutely continuous functions on a compact set. By Theorem 1 in Milgrom and Segal (2002), whenever U is differentiable, we have

$$q(s) - \lambda_+(s)U(s) = D_{2+}u(\theta|\theta) \leq U'(s) \leq D_{2-}u(\theta,\theta) = q(s) + \lambda_-(s)U(s).$$

At each point θ where all the functions involved are differentiable, which holds almost surely, we have

$$\begin{aligned} \Delta'(\theta) &= \lambda_+(\theta)\Lambda(\theta|\bar{\theta})U(\theta) + \Lambda(\theta|\bar{\theta})U'(\theta) - \Lambda(\theta|\bar{\theta})q(\theta) \\ &= \Lambda(\bar{\theta},\theta)\left[U'(\theta) - (q(\theta) - \lambda_+(\theta)U(\theta))\right] \\ &\geq 0. \end{aligned}$$

Since $\Delta(\underline{\theta}) = 0$, the fundamental theorem of calculus implies that $\Delta(\theta) \geq 0$ for all θ, as desired.

It remains to check that the global incentive constraints are satisfied, provided that q is monotone. Expressing incentive-compatibility in terms of U, we need to show that for all types θ and θ',

$$U(\theta) \geq \alpha(\theta'|\theta)(U(\theta') + (\theta - \theta')q(\theta')). \tag{2.B.13}$$

We consider upward and downward deviations separately. First suppose $\theta' > \theta$. Write out (2.B.13) as

$$U(\theta) + \alpha(\theta'|\theta)(\theta' - \theta)q(\theta') \geq \alpha(\theta'|\theta)U(\theta'),$$

or equivalently,

$$\begin{aligned} &\int_{\underline{\theta}}^{\theta} \Lambda(z|\theta)q(z)\,\mathrm{d}z + \alpha(\theta'|\theta)\int_{\theta}^{\theta'} q(\theta')\,\mathrm{d}z \\ &\qquad \geq \alpha(\theta'|\theta)\int_{\underline{\theta}}^{\theta} \Lambda(z|\theta')q(z)\,\mathrm{d}z + \alpha(\theta'|\theta)\int_{\theta}^{\theta'} \Lambda(z|\theta')q(z)\,\mathrm{d}z. \end{aligned}$$

Compare the corresponding terms on each side. Since $\Lambda(z|\theta) \geq \Lambda(z|\theta')$ for $z \leq \theta \leq \theta'$, we get the inequality between the first terms. For the inequality between the second terms, multiply by $\Lambda(\theta'|\bar{\theta})/\alpha(\theta'|\theta)$ and use the fact that $\Lambda(z|\bar{\theta})q(z)$ is increasing in z.

Now suppose $\theta' < \theta$. Express (2.B.13) as

$$\int_{\underline{\theta}}^{\theta'} \Lambda(z|\theta)q(z)\,\mathrm{d}z + \int_{\theta'}^{\theta} \Lambda(z|\theta)q(z)\,\mathrm{d}z$$
$$\geq \alpha(\theta'|\theta)\left[\int_{\underline{\theta}}^{\theta'} \Lambda(z|\theta')q(z)\,\mathrm{d}z + \int_{\theta'}^{\theta} q(\theta')\,\mathrm{d}z\right]$$
$$= \frac{\alpha(\theta'|\theta)}{\Lambda(\theta'|\theta)}\left[\int_{\underline{\theta}}^{\theta'} \Lambda(z|\theta)q(z)\,\mathrm{d}z + \int_{\theta'}^{\theta} \Lambda(\theta'|\theta)q(\theta')\,\mathrm{d}z\right].$$

Multiply both sides by $\Lambda(\theta|\bar{\theta})$ and then rearrange to get the equivalent inequality

$$\alpha(\theta'|\theta) \;\leq\; \Lambda(\theta'|\theta)\frac{\int_{\underline{\theta}}^{\theta'} \Lambda(z|\bar{\theta})q(z)\,\mathrm{d}z + \int_{\theta'}^{\theta} \Lambda(z|\bar{\theta})q(z)\,\mathrm{d}z}{\int_{\underline{\theta}}^{\theta'} \Lambda(z|\bar{\theta})q(z)\,\mathrm{d}z + \int_{\theta'}^{\theta} \Lambda(\theta'|\bar{\theta})q(\theta')\,\mathrm{d}z}$$
$$= \Lambda(\theta'|\theta)\int_{\underline{\theta}}^{\theta} \Lambda(z|\bar{\theta})q(z)\,\mathrm{d}z\Big/\int_{\underline{\theta}}^{\theta} \Lambda(z\wedge\theta'|\bar{\theta})q(z\wedge\theta')\,\mathrm{d}z.$$

2.B.13 Proof of Proposition 11

The same argument as in the proof of Proposition 10 (section 2.B.11) shows that the principal's value as a function of q can be written as

$$V(q) \;=\; \int_{\underline{\theta}}^{\bar{\theta}} \varphi(\theta)q(\theta)\,\mathrm{d}\theta.$$

The quantity function $q^\star$ from the theorem statement maximizes this quantity pointwise, and $t^\star$ is the corresponding transfer function. Since $q^\star$ is monotone, this mechanism is incentive compatible and hence optimal. Except at points θ where $\varphi(\theta) = 0$, the pointwise maximizer is unique, and hence the mechanism is unique almost everywhere outside the set $\varphi^{-1}(0)$.

2.B.14 Proof of Theorem 12

The proof of Theorem 9 (section 2.B.2) can be extended to allow for multiple agents by taking products like in the proof of Theorem 8 (section 2.B.1). For the sufficiency, define $\hat{k}_i$ for each agent i and set $\hat{k} = \otimes_i \hat{k}_i$. For necessity, if there is some j such that $\hat{t}_j$ is not most-discerning, apply the construction above on agent j, assuming every type of every other agent is indifferent over all decisions.

2.B.15 Proof of Proposition 12

Applying the same argument player by player gives

$$V(Q) \;=\; \int_{\Theta} \Big(\sum_{i=1}^{n} \varphi_i(\theta_i) q_i(\theta_i) \Big) f(\theta)\, \mathrm{d}\theta.$$

This is maximized by the interim quantity functions $q^\star$ in the theorem statement, which induces monotone interim quantity functions. For each agent i, the transfer function is pinned down by the envelope expression for U_i. We then choose a transfer function $t^\star$ consistent with these interim transfer functions that also satisfies the ex post participation constraints.

Appendix 2.C Supplementary proofs

2.C.1 Proof of Lemma 8

(i) Fix $p \in [0,1]$. It suffices to show that $(\mu\tilde{F}_\mu)[0,p] = p$. For each $s \in \boldsymbol{R}$, let $F_\mu(s-)$ denote the left limit of F_μ at s. With this notation, we have

$$\tilde{F}_\mu(s,[0,p]) \;=\; \begin{cases} 1 & \text{if } F_\mu(s) \;\le\; p, \\ \frac{p - F_\mu(s-)}{F_\mu(s) - F_\mu(s-)} & \text{if } F_\mu(s-) \;\le\; p \;<\; F_\mu(s), \\ 0 & \text{if } F_\mu(s-) \;>\; p. \end{cases}$$

Set $F_\mu^+(p) = \sup\{t \in \boldsymbol{R} : F_\mu(t-) \le p\}$.[30] By the left-continuity of the map $t \mapsto F_\mu(t-)$, we have $F_\mu(F_\mu^+(p)) \le p$, with equality if μ is continuous at $F_\mu^+(p)$. If μ is continuous at $F_\mu^+(p)$, then

$$(\mu\tilde{F}_\mu)[0,p] \;=\; \mu(-\infty, F_\mu^+(p)] \;=\; F_\mu(F_\mu^+(p)) \;=\; p.$$

If μ is discontinuous at $F_\mu^+(p)$, then

$$(\mu\tilde{F}_\mu)[0,p] \;=\; F_\mu(F_\mu^+(p)-) + \mu(\{F_\mu^+(p)\}) \frac{p - F_\mu(F_\mu^+(p)-)}{F_\mu(F_\mu^+(p)) - F_\mu(F_\mu^+(p)-)},$$

and the right side simplifies to p.

(ii) Fix $s \in \boldsymbol{R}$. For each $p \in [0,1]$, recall that for all $s \in \operatorname{supp} \nu$, we have $Q_\nu(p) \le s$ if and only if $p \le F_\nu(s)$. Therefore,

$$\tilde{Q}_\nu(p,(-\infty,s]) \;=\; \begin{cases} 1 & \text{if } p \;\le\; F_\nu(s), \\ 0 & \text{if } p \;>\; F_\nu(s). \end{cases}$$

30. This is the right-continuous inverse of F_μ and is more commonly defined as $\inf\{t \in \boldsymbol{R} : F_\mu(t) > p\}$.

We conclude that

$$(U_{[0,1]}\tilde{Q}_\nu)(-\infty, s] = U_{[0,1]}[0, F_\nu(s)] = F_\nu(s) = \nu(-\infty, s].$$

(iii) This follows immediately from (i) and (ii).

2.C.2 Proof of Lemma 9

By Lemma 8 (iii), we have (ii) $\implies$ (iii). We prove that (i) $\implies$ (ii) $\implies$ (i).

(i) $\implies$ (ii). Suppose $\mu \succeq_{\text{SD}} \nu$. Set $k = \tilde{F}_\mu \tilde{Q}_\nu$. Fix $s \in S$ and let $S_0 = (-\infty, s] \cap S$. Recall that the left-continuous quantile function satisfies the Galois inequality

$$Q_\nu(p) \leq s \iff p \leq F_\nu(s).$$

Thus,

$$k(s, S_0) = \int_0^1 \tilde{F}_\mu(s, \mathrm{d}p)\tilde{Q}_\nu(p, S_0) = \int_0^{F_\nu(s)} \tilde{F}_\mu(s, \mathrm{d}p).$$

By first-order stochastic dominance, $F_\nu(s) \geq F_\mu(s)$. Thus, the right side is at least $\tilde{F}_\mu(s, [0, F_\mu(s)])$, which equals 1.

(iii) $\implies$ (i). Let k be a downward transition. Fix s in S, and let $S_0 = S \cap (-\infty, s]$. We have

$$(\mu k)(S_0) = \int_{\boldsymbol{R}} \mu(\mathrm{d}t)k(t, S_0) \geq \int_{(-\infty, s]} \mu(\mathrm{d}t)k(t, S_0) = \mu(S_0).$$

2.C.3 Proof of Lemma 10

(i) Suppose m is monotone on S. If $\mu \geq_{\text{SD}} \nu$, then by Lemma 9 (iii) there is a downward transition k on S such that $\mu k = \nu$. Fix $s_0 \in S$, and set $S_0 = (-\infty, s_0] \cap S$. Then

$$\begin{aligned}
(\nu m)(S_0) - (\mu m)(S_0) &= (\mu k m)(S_0) - (\mu m)(S_0) \\
&= \int_S \mu(\mathrm{d}s)[(km)(s, S_0) - m(s, S_0)] \\
&= \int_S \mu(\mathrm{d}s) \int_{-\infty}^{s} k(s, \mathrm{d}t)[m(t, S_0) - m(s, S_0)] \\
&\geq 0,
\end{aligned}$$

where we get the second equality because k is downward, and the inequality because m is monotone.

For the other direction, take $\mu = \delta_s$ and $\nu = \delta_t$ for $s > t$. Then

$$m_s = \delta_s m \geq_{\text{SD}} \delta_t m = m_t.$$

(ii) Let m and m' be monotone transitions on S. Suppose μ and ν are measures on S satisfying $\mu \geq_{\text{SD}} \nu$. Applying (i) twice, we have $\mu m \geq_{\text{SD}} \nu m$ and hence

$$\mu(mm') = (\mu m)m' \geq_{\text{SD}} (\nu m)m' = \mu(mm').$$

By (i), mm' is monotone.

2.C.4 Proof of Lemma 11

Consider a Markov transition $k\colon X \times \mathscr{Y} \to [0,1]$. Define $\bar{k}\colon X \times \overline{\mathscr{Y}} \to [0,1]$ by setting $\bar{k}_x$ equal to the extension of k_x to $\overline{\mathscr{Y}}$, for each x in X. For each B in $\mathscr{Y}$, we have $\bar{k}(x,B) = k(x,B)$, so the map $x \mapsto \bar{k}(x,B)$ is measurable and hence universally measurable. We need to check universal measurability for sets B in $\overline{\mathscr{Y}}$. Let μ be an arbitrary probability measure on $(X, \mathscr{X})$. It suffices to show that $\bar{k}(\cdot, B)$ is $\overline{\mathscr{X}}_\mu$-measurable for all B in $\overline{\mathscr{Y}}_{(\mu k)}$. If B is in $\overline{\mathscr{Y}}_{(\mu k)}$, then we can sandwich B between B_1 and B_2 satisfying

$$0 = (\mu k)(B_2 \setminus B_1) = \mu(k(\cdot, B_2)) - \mu(k(\cdot, B_1)).$$

So the function $\bar{k}(\cdot, B)$ is sandwiched between the $\mathscr{X}$-measurable functions $k(\cdot, B_1)$ and $k(\cdot, B_2)$, which agree μ–almost surely. Hence $\bar{k}(\cdot, B)$ is $\overline{\mathscr{X}}_\mu$-measurable.

2.C.5 Proof of Lemma 12

For the inclusion, it suffices to show that every probability measure μ on $(X \times Y, \mathscr{X} \otimes \mathscr{Y})$, we have

$$\overline{\mathscr{X}} \otimes \overline{\mathscr{Y}} \subset \overline{(\mathscr{X} \otimes \mathscr{Y})}_\mu.$$

Fix such a probability measure μ. Define measures μ_1 and μ_2 on $\mathscr{X}$ and $\mathscr{Y}$ by

$$\mu_1(A) = \mu(A \times Y) \quad \text{and} \quad \mu_2(B) = \mu(X \times B).$$

If A is in $\overline{\mathscr{X}}_{\mu_1}$, then there exist A_1 and A_2 in $\mathscr{X}$ sandwiching A such that

$$0 = \mu_1(A_2 \setminus A_1) = \mu((A_2 \setminus A_1) \times Y) = \mu((A_2 \times Y) \setminus (A_1 \times Y)).$$

Similarly, if B in $\overline{\mathscr{Y}}_{\mu_2}$, then $X \times B$ is in $\overline{(\mathscr{X} \otimes \mathscr{Y})}_\mu$. Taking intersections we conclude that $A \times B$ is in $\overline{(\mathscr{X} \otimes \mathscr{Y})}_\mu$. Therefore,

$$\overline{\mathscr{X}} \otimes \overline{\mathscr{Y}} \subset \overline{\mathscr{X}}_{\mu_1} \otimes \overline{\mathscr{Y}}_{\mu_2} \subset \overline{(\mathscr{X} \otimes \mathscr{Y})}_\mu.$$

Now we turn to the last equality. For each probability measure μ on $\overline{\mathscr{X}} \otimes \overline{\mathscr{Y}}$, let μ_0 be the restriction of μ to $\mathscr{X} \otimes \mathscr{Y}$. Then

$$\overline{\mathscr{X} \otimes \mathscr{Y}} \subset \overline{(\mathscr{X} \otimes \mathscr{Y})}_{\mu_0} \subset \overline{(\overline{\mathscr{X}} \otimes \overline{\mathscr{Y}})}_\mu.$$

Taking the intersection over all such μ gives $\overline{\mathscr{X} \otimes \mathscr{Y}} \subset \overline{\overline{\mathscr{X}} \otimes \overline{\mathscr{Y}}}$.

Now we prove the reverse inclusion. Each probability measure ν on $\mathscr{X} \otimes \mathscr{Y}$ has a complete extension $\bar{\nu}$ to $\overline{(\mathscr{X} \otimes \mathscr{Y})}_\nu$. Let ν_0 be the restriction of $\bar{\nu}$ to $\overline{\mathscr{X}} \otimes \overline{\mathscr{Y}}$. Then

$$\overline{(\mathscr{X} \otimes \mathscr{Y})}_\nu = \overline{(\overline{\mathscr{X}} \otimes \overline{\mathscr{Y}})}_{\nu_0} \supset \overline{\overline{\mathscr{X}} \otimes \overline{\mathscr{Y}}}.$$

Taking the intersection over all such ν gives $\overline{\mathscr{X} \otimes \mathscr{Y}} \supset \overline{\overline{\mathscr{X}} \otimes \overline{\mathscr{Y}}}$.

References

Aliprantis, Charalambos D., and Kim C. Border. 2006. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. 3rd edition. Springer. [81, 82, 85]

Auletta, Vincenzo, Paolo Penna, Giuseppe Persiano, and Carmine Ventre. 2011. "Alternatives to Truthfulness are Hard to Recognize." *Autonomous Agents and Multi-Agent Systems* 22: 200–16. [74]

Balbuzanov, Ivan. 2019. "Lies and consequences." *International Journal of Game Theory* 48 (4): 1203–40. [74]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L Lipman. 2018. "Disclosure and choice." *Review of Economic Studies* 85 (3): 1471–501. [74]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L Lipman. 2019. "Mechanisms with evidence: Commitment and robustness." *Econometrica* 87 (2): 529–66. [75]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman. 2014. "Optimal Allocation with Costly Verification." *American Economic Review* 104 (12): 3779–813. [75]

Blackwell, David. 1953. "Equivalent comparisons of experiments." *annals of mathematical statistics*, 265–72. [48, 55]

Border, Kim C, and Joel Sobel. 1987. "Samurai Accountant: A Theory of Auditing and Plunder." *Review of Economic Studies* 54: 525–40. [65]

Bull, Jesse, and Joel Watson. 2004. "Evidence Disclosure and Verifiability." *Journal of Economic Theory* 118: 1–31. [74]

Bull, Jesse, and Joel Watson. 2007. "Hard Evidence and Mechanism Design." *Games and Economic Behavior* 58: 75–93. [48, 51, 57, 73, 74]

Caragiannis, Ioannis, Edith Elkind, Mario Szegedy, and Lan Yu. 2012. "Mechanism Design: From Partial to Probabilistic Verification." In *Proceedings of the 13th ACM Conference on Electronic Commerce*. EC '12. Valencia, Spain: ACM, 266–83. DOI: 10.1145/2229012.2229035. [66, 73, 74]

Cohn, Donald L. 2013. *Measure Theory*. 2nd edition. Birkhäuser Advanced Texts Basler Lehrbücher. Birkhäuser Basel. [78]

Crocker, Keither J., and John Morgan. 1998. "Is Honesty the Best Policy? Curtailing Insurance Fraud through Optimal Incentive Contracts." *Journal of Political Economy* 106 (2): 355–75. [63]

Deb, Rahul, and Colin Stewart. 2018. "Optimal Adaptive Testing: Informativeness and Incentives." *Theoretical Economics* 13 (3): 1233–74. DOI: 10.3982/TE2914. [50]

DeMarzo, Peter M, Ilan Kremer, and Andrzej Skrzypacz. Forthcoming. "Test Design and Minimum Standards." *American Economic Review*, [50]

Deneckere, Raymond, and Sergei Severinov. 2008. "Mechanism Design with Partial State Verifiability." *Games and Economic Behavior* 64 (2): 487–513. [75]

Deneckere, Raymond, and Sergei Severinov. 2017. "Screening, Signalling and Costly Misrepresentation." Working paper. [63]

Dziuda, Wioletta, and Christian Salas. 2018. "Communication with detectable deceit." *Available at SSRN 3234695*, [74]

Erlanson, Albin, and Andreas Kleiner. 2015. "Costly Verification in Collective Decisions." Working paper. [75]

Ferraioli, Diodato, and Carmine Ventre. 2018. "Probabilistic Verification for Obviously Strategyproof Mechanisms." In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. AAMAS '18. Stockholm, Sweden: International Foundation for Autonomous Agents, and Multiagent Systems, 1930–32. URL: http://dl.acm.org/citation.cfm?id=3237383.3238027. [74]

Fotakis, Dimitris, and Emmanouil Zampetakis. 2015. “Truthfulness Flooded Domains and the Power of Verification for Mechanism Design.” *ACM Transactions on Economics and Computation* 3 (4): 20:1–29. [74]

Green, Jerry R., and Jean-Jacques Laffont. 1986. “Partially Verifiable Information and Mechanism Design.” *Review of Economic Studies* 53 (3): 447–56. [47–49, 52, 64, 73]

Grossman, Sanford J. 1981. “The informational role of warranties and private disclosure about product quality.” *Journal of Law and Economics* 24 (3): 461–83. [74]

Halac, Marina, and Pierre Yared. 2017. “Commitment vs. Flexibility with Costly Verification.” Working paper. [75]

Hart, Sergiu, Ilan Kremer, and Motty Perry. 2017. “Evidence Games: Truth and Commitment.” *American Economic Review* 107 (3): 690–713. [74]

Kallenberg, Olav. 2017. *Random Measures, Theory and Applications*. Vol. 77, Probability Theory and Stochastic Modelling. Springer. [75, 76, 79]

Kartik, Navin. 2009. “Strategic Communication with Lying Costs.” *Review of Economic Studies* 76: (4), 1359–95. [63]

Kartik, Navin, Marco Ottaviani, and Francesco Squintani. 2007. “Credulity, Lies, and Costly Talk.” *Journal of Economic Theory* 134: (1), 93–116. [63]

Kephart, Andrew, and Vincent Conitzer. 2016. “The Revelation Principle for Mechanism Design with Reporting Costs.” In *Proceedings of the 2016 ACM Conference on Economics and Computation*, 85–102. [63]

Koessler, Frédéric, and Eduardo Perez-Richet. 2017. “Evidence Reading Mechanisms.” Working paper. [74]

Lacker, Jeffrey M., and John A. Weinberg. 1989. “Optimal Contracts under Costly State Falsification.” *Journal of Political Economy* 97 (6): 1345–63. [63]

Li, Yunan. 2017. “Mechanism Design with Costly Verification and Limited Punishments.” Working paper. [75]

Lipman, Barton L., and Duane J. Seppi. 1995. “Robust Inference in Communication Games with Partial Provability.” *Journal of Economic Theory* 66: 370–405. [48, 57, 74]

Maggi, Giovanni, and Andrés Rodriguéz-Clare. 1995. “Costly Distortion of Information in Agency Problems.” *RAND Journal of Economics* 26 (4): 675–89. [63]

Mertens, Jean-Francois, Sylvain Sorin, and Shmuel Zamir. 2015. *Repeated Games*. Vol. 55, Econometric Society Monographs. Cambridge University Press. [53]

Milgrom, Paul R. 1981. “Good news and bad news: Representation theorems and applications.” *Bell Journal of Economics*, 380–91. [74]

Milgrom, Paul, and Ilya Segal. 2002. “Envelope Theorems for Arbitrary Choice Sets.” *Econometrica* 70 (2): 583–601. DOI: 10.1111/1468-0262.00296. [92]

Mussa, Michael, and Sherwin Rosen. 1978. “Monopoly and product quality.” *Journal of Economic Theory* 18: 301–17. DOI: 10.1016/0022-0531(78)90085-6. [65, 69]

Myerson, Roger B. 1981. “Optimal Auction Design.” *Mathematics of Operations Research* 6 (1): 58–73. DOI: 10.1287/moor.6.1.58. [67, 72]

Mylovanov, Tymofiy, and Andriy Zapechelnyuk. 2017. “Optimal Allocation with Ex Post Verification and Limited Penalties.” *American Economic Review* 107 (9): 2666–94. [75]

Nisan, Noam, and Amir Ronen. 2001. “Algorithmic Mechanism Design.” *Games and Economic Behavior* 35: 166–96. [74]

Riley, John, and Richard Zeckhauser. 1983. “Optimal selling strategies: When to haggle, when to hold firm.” *Quarterly Journal of Economics* 98 (2): 267–89. [70]

Rochet, Jean-Charles. 1987. "A Necessary and Sufficient Condition for Rationalizability in a Quasi-Linear Context." *Journal of Mathematical Economics* 16 (2): 191–200. [74]

Singh, Nirvikar, and Donald Wittman. 2001. "Implementation with Partial Verification." *Review of Economic Design* 6: 63–84. [74]

Strausz, Roland. 2016. "Mechanism Design with Partially Verifiable Information." Cowles Foundation Discussion Paper No. 2040. [73]

Townsend, Robert M. 1979. "Optimal Contracts and Competitive Markets with Costly State Verification." *Journal of Economic Theory* 21 (2): 265–93. [75]

Vohra, Rakesh V. 2011. *Mechanism Design: A Linear Programming Approach*. Econometric Society Monographs. Cambridge University Press. DOI: 10.1017/CBO9780511835216. [74]

Yu, Lan. 2011. "Mechanism Design with Partial Verification and Revelation Principle." *Autonomous Agents and Multi-Agent Systems* 22: 217–23. [74]

Chapter 3

Costless Information and Costly Verification: A Case for Transparency*

Joint with Jan Knoepfle

3.1 Introduction

A principal has to take a binary decision for which she relies on an agent's private information. The agent prefers one of the two actions independent of his information. Prior to the decision, the principal privately observes a signal about the agent's information. She cannot incentivize the agent through monetary transfers but has the opportunity to reveal his information at a cost.

Examples for this setting include: a human resource department decides whether to hire a candidate, a judge decides whether to acquit or convict a defendant, or a competition authority decides whether to grant or deny a company permission to merge with or acquire another firm.

While one party —the agent— has a clear preference toward one action (the candidate wants to be hired, the defendant wants to be acquitted, and the company wants to merge), the preferences of the other party —the principal— depend on information that is privately held by the agent. Here, one may think of the candidate's ability, the defendant's guilt, or the company's competitive position in the market.

* We thank Marina Halac, Stephan Lauermann, Benny Moldovanu, and Sven Rady for continual guidance and support. We likewise thank Ian Ball, Andreas Kleiner, Daniel Krähmer, Martin Pollrich, Andre Speit, Dezsö Szalay, Tymon Tatur, and participants at the 11th BiGSEM Workshop on Economic Theory (Bielefeld), the BGSE Micro Workshop (Bonn), the Jerusalem Summer School in Economic Theory 2017, the Yale Theory Breakfast, the Stony Brook Conference on Game Theory 2018, and the EEA-ESEM 2018 (Cologne) for insightful discussions. Both authors thank the Yale Economics Department for their hospitality during the respective research visits, when parts of this paper were written. This work was supported by the German Research Foundation (DFG) through CRC TR 224 (Project B04).

Often, monetary transfers to elicit the agent's private information are not feasible (for practical or moral reasons[1]), but the principal can learn the information at a cost, for example, by conducting an assessment center, a trial, or a market analysis. However, verification is costly, so the principal has an incentive to economize on it.

Typically, costly information acquisition is not the only way to learn the agent's private information. The potential employer receives references or recommendation letters from previous supervisors, the judge can inspect the outcome of pretrial investigations, and the competition authority has sector-specific knowledge derived from its supervisory function. That is, the principal privately observes factors that are correlated with the agent's type.

This paper investigates how the principal can use this private, costless information in conjunction with the costly verification to maximize her expected payoff from the decision if monetary transfers are not feasible.

Preview. We show that the optimal Bayesian incentive compatible (BIC) mechanism takes a simple cut-off structure: if the principal observes a signal that makes her sufficiently certain that the agent's preferred action is also better for her, she takes it, independent of the type report. If the signal falls below the cut-off, she takes the nonpreferred action by default but gives the agent the possibility to appeal. An appeal is always verified and induces the agent-preferred action whenever his type exceeds a threshold. This appeal threshold is set such that the agents who appeal are only those whose type makes it worthwhile for the principal to implement the agent-preferred action and pay the verification cost.

This mechanism is ex-post incentive compatible (EPIC). It would also be incentive compatible if the principal's signal was known to the agent. This result implies that the principal does not benefit from the privacy of her signal. It advocates for transparent procedures.

We extend our result to settings where the information that the principal privately observes has a positive direct effect on her utility from the agent's preferred choice. The structure of the optimal mechanism remains, and, again, transparency comes without loss for the principal. If, in contrast, the direct effect is negative, the principal benefits from hiding his information. The equivalence between EPIC and BIC optimal mechanisms breaks down; we show that the simple EPIC mechanism is no longer optimal in the larger class of BIC mechanisms.

Literature. In settings where monetary transfers are feasible, the principal can design a lottery rewarding the agent for guessing the value of her privately observed signal correctly. Different agent types hold different beliefs over the signal distribution and, therefore, reveal their type by guessing the signal they deem most likely.

1. The assumption is that payments cannot depend on the agent's report. Even though a public sector job entails payments, if the payment is fixed, it cannot be used to incentivize truthful reports of the candidate's ability.

If the agent's liability is not limited, the principal can increase reward and loss in the lottery to such an extent that the incentives to win the lottery exceed any incentives regarding the allocation decision. In doing so, she can learn the agent's type at arbitrarily small costs. Mechanisms with monetary transfers and correlated information have been discussed by Crémer and McLean (1988), Riordan and Sappington (1988), Johnson, Pratt, and Zeckhauser (1990), and McAfee and Reny (1992), who all establish conditions on the information structure that ensure full surplus extraction by the principal. As all surplus can be extracted, revenue maximization leads to ex-post efficient allocations. Neeman (2004) discusses the genericity of the above-mentioned conditions and shows that full surplus extraction is possible only if every preference type is "determined" by his belief over the correlated characteristics. Even though this condition is fulfilled, in our setting with costly verification instead of monetary transfers, full surplus extraction is not feasible and implementing the ex-post efficient allocation is not optimal for the principal.

The full surplus-extracting lotteries require potentially unbounded transfers. For the case of bounded transfers or limited liability, Demougin and Garvie (1991) show that the qualitative results, the application of rewards as bets on the signal, still apply. Different from our setting, the principal gains by maintaining her signal private.

In the absence of monetary transfers, Bhargava, Majumdar, and Sen (2015) show how positively correlated beliefs among voters allow overcoming the impossibility of nondictatorial voting rules established by Gibbard (1973) and Satterthwaite (1975).

Our result is in line with the findings in other settings where monetary transfers are not feasible but correlated information is absent. The literature (Glazer and Rubinstein, 2004; Ben-Porath, Dekel, and Lipman, 2017; Hart, Kremer, and Perry, 2017; Erlanson and Kleiner, 2019; Halac and Yared, 2019) has found optimal mechanisms to take a simple cut-off structure and to be EPIC in the sense that the agents would also report truthfully if they were informed about the other agents' type realizations before their report.

The possibility for the mechanism designer to verify an agent's private information at a cost was first introduced by Townsend (1979) considering a principal-agent model for debt contracts, which was extended to a two-period model by Gale and Hellwig (1985). These early models of state verification feature both, monetary transfers and verification. Glazer and Rubinstein (2004) introduce a setting where the principal has to take a binary decision depending on the multidimensional private information of the agent. Here, the principal cannot use monetary transfers, but she can learn about one dimension before making her decision.

Our model is most closely related to that of Ben-Porath, Dekel, and Lipman (2014), who model costly verification and consider the case of allocating a good among finitely many agents whose types are independently distributed; see also the discussion section. Erlanson and Kleiner (2019) study a collective decision problem with costly verification and show that the optimal mechanism is EPIC and can be implemented by a simple weighted majority voting rule. Mylovanov and Zapechelnyuk (2017) consider an allocation problem without monetary transfers in which the principal learns the agents' types without cost but only posterior to the allocation decision and has the ability to punish untruthful reports up to a limit. Halac and Yared (2019) consider a delegation problem and specify conditions on the verification cost that ensure optimality of a threshold mechanism with an escape clause.

Erlanson and Kleiner (2019) show further that the equivalence between BIC and EPIC mechanisms holds more generally rather than only for optimal mechanisms. This relates to Gershkov, Goeree, Kushnir, Moldovanu, and Shi (2013) and Manelli and Vincent (2010), who show equivalence between BIC and DIC mechanisms in settings with monetary transfers. All these results have been derived under the assumption that the private information of players is independently distributed; we deviate from this assumption by introducing correlation between the agent's type and the principal's signal.

As the principal has private information, our model is also related to the informed principal problem cf. Myerson (1983) and Maskin and Tirole (1990). With monetary transfers, Severinov (2008) and Cella (2008) show that correlated information allows for an efficient solution to the informed principal problem. We assume that the principal designs mechanisms with full commitment over allocation procedures before observing the signal. A priori, there is no informed principal problem in our model, but in the discussion section, we show that the optimal mechanism we derive also solves the informed principal problem.

Roadmap. After an example highlighting the difficulties that correlation adds to the canonical verification setup, and showing how our findings advocate for transparency in pretrial investigations, section 2 sets up the model. The characterization of optimal mechanisms starts in section 3 for the class of transparent mechanisms, and section 4 shows that this mechanism is optimal in the broader class of BIC mechanisms. We then extend the analysis to a specification in which the principal's valuation may be affected by the signal realization, and discuss the relation to favored-agent mechanisms and to the informed principal problem.

3.2 Example

With the following numerical example, we illustrate (1) how the principal can exploit correlation to lower the minimal verification cost required for implementing an allocation and (2) why the optimal allocation does not leave scope for such an improvement. Consider a defendant in court (the agent) who privately knows whether he is of the guilty or innocent type $t \in \{G, I\}$. The judge (principal) privately observes signal realization $s \in \{g, i\}$ as a result of pretrial investigations and asks the agent to plead either guilty or innocent (to report his type). Following the signal and plea, the judge decides whether to conduct a costly trial (to reveal the agent's type) and whether the defendant should be acquitted (as a function the trial's outcome, in case it was conducted). A trial requires verification cost $c > 0$. The defendant's utility is 1 from acquittal and 0 if he is convicted, irrespective of his type. Type and signal are jointly distributed according to

$$\begin{pmatrix} f_{G,g} & f_{G,i} \\ f_{I,g} & f_{I,i} \end{pmatrix} = \begin{pmatrix} \frac{2}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{6} \end{pmatrix}.$$

For illustration, we fix acquittal probabilities and consider the optimal verification schedule for the case when the defendant observes the outcome of the pretrial investigation and for the case when he does not observe this signal. Let the guilty type be acquitted with probability $1/2$, at both signals g and i, and the innocent type with probability 1, independent of the signal. Denote by $z_{t,s} \in [0,1]$ the probability with which the type-signal combination (t,s) is verified. After verification, the judge acquits the innocent and convicts the guilty type.[2] In a transparent mechanism, the agent observes the principal's signal before making a report. The cost-minimal verification probabilities that ensure truthful reporting are[3]

$$\begin{pmatrix} z_{G,g} & z_{G,i} \\ z_{I,g} & z_{I,i} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1/2 & 1/2 \end{pmatrix}.$$

This induces a verification cost of $\frac{1}{6} \cdot 1/2\, c + \frac{2}{6} \cdot 1/2\, c = \frac{1}{4}\, c$.

If, instead, the signal realization is not known to the agent when he makes his type report, the principal can save verification costs. The above mechanism fulfills type G's incentive constraint by verifying report I with equal probability after both signal realizations. The principal can exploit the fact that type G's subjective belief puts more weight on signal g, and shift verification probability from the type-signal combination (I,i) to (I,g). The verification probabilities

$$\begin{pmatrix} z_{G,g} & z_{G,i} \\ z_{I,g} & z_{I,i} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 3/4 & 0 \end{pmatrix}$$

produce verification costs of only $\frac{1}{6} \cdot 3/4\, c = \frac{1}{8}\, c$ but ensure truthful reporting. To see this, consider the following (Bayesian) IC constraints:

$$\begin{aligned} \frac{2}{3} \cdot 1/2 \quad + \quad \frac{1}{3} \cdot 1/2 \quad &\geq \quad \frac{2}{3} \cdot \left(1 - z_{I,g}\right) \quad + \quad \frac{1}{3} \cdot \left(1 - z_{I,i}\right) \qquad \text{and} \\ \frac{1}{3} \cdot 1 \quad + \quad \frac{2}{3} \cdot 1 \quad &\geq \quad \frac{1}{3} \cdot \left(1/2 - z_{G,g}\right) \quad + \quad \frac{2}{3} \cdot \left(1/2 - z_{G,i}\right) \quad . \end{aligned}$$

2. This assumption is inessential and made for simplicity here. As will be shown in the main body of the paper, verification optimally affects only misreporting types, with the harshest possible punishment. That is, whenever verification reveals a misreport, the defendant is not acquitted.

3. Type $t = I$ is acquitted for sure and, therefore, never has an incentive to misreport type $t = G$. Type G, who knows signal s, is willing to report his type truthfully if $1/2 \geq 1 - z_{I,s}$.

The first constraint hinders type G from reporting I. Note that $\frac{2}{3}$ is the agent's posterior belief that the signal matches his type. As before, the second constraint, hindering type I from reporting G, is satisfied independent of the verification. However, the two constraints illustrate the complication that correlation adds to the analysis: different types assign distinct probabilities to the signal realizations. Therefore, the expected utility of a given type report depends on the agent's real type. This complicates the characterization of incentive compatibility in comparison to other costly verification models.[4]

With this second verification schedule, the above-mentioned allocation,

$$\begin{pmatrix} 1/2 & 1/2 \\ 1 & 1 \end{pmatrix},$$

is not transparently implementable. If type G knows that the signal is i, he can be acquitted with probability 1 by misreporting I. This illustrates how a nontransparent procedure potentially allows the lowering of verification costs by exploiting correlation. The idea parallels the design of transfer lotteries used in Crémer and McLean (1988) and others to extract the agent's surplus.

However, conditional on the defendant's guilt, the outcome of the pretrial investigation should not affect the judge's preferences over acquittal or conviction. Therefore, we can achieve the same ex-ante expected allocation value but adjust how the probability to acquit the guilty type is distributed over the signal realizations: consider ex-post acquittal probabilities

$$\begin{pmatrix} 1/4 & 1 \\ 1 & 1 \end{pmatrix},$$

and note that the ex-ante probability for both types and, therefore, the principal's expected allocation value remain unchanged. With the above-mentioned verification probability of $z_{I,g} = \frac{3}{4}$, this allocation can be transparently implemented. The guilty type is indifferent between truth-telling or misreporting after observing either signal.[5] This exemplifies how the allocation can be re-arranged over signal realizations

4. In Ben-Porath, Dekel, and Lipman (2014) and other models of costly state verification, the absence of correlated information allows to fully characterize incentive compatibility by focusing on the type with the lowest expected utility. This technique cannot be applied to our setting. To make this explicit, consider an allocation rule that acquits the agent whenever type and signal do not match (without verification). Both types would strictly prefer to misreport. Without correlated information, such a situation cannot arise.

5. The observant reader may notice how the fact that the principal's value remains unchanged, hinges on the assumption that the signal does not have a direct effect on the allocation value for given types. This is justified in most applications, in which the signal is simply informative about the underlying fundamental type, as the outcome of pretrial investigations gives information about the defendant's guilt but should not itself affect the value of a guilty verdict. In the extensions, we show how our main result carries over with a direct effect that goes in the same direction as the informational effect, for example, when confirming the outcome of the pretrial investigations carries some value on its own.

so that the suggested improvement in verification does not impede transparency. In the main body of the paper, we show how this can be done for the optimal allocation without violating incentives with arbitrary numbers of types and signals.

The acquittal probability in this example was fixed somewhat arbitrarily to allow for informative illustration. To consider the original problem of choosing allocation and verification probabilities jointly, the principal needs to trade off verification costs and allocation value. The paper proves that this trade-off is optimally solved by a simple cut-off mechanism. In this court example, for a range of parameter values, the optimal cut-off mechanism would feature allocation

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix},$$

implemented through verification schedule

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

This mechanism resembles the proceedings of a pretrial: The case is dismissed if the signal for the defendant's innocence is strong enough, i.e., the charge is weak. If the signal for innocence is below this cut-off, the agent can plead guilty and is convicted, or he can request a trial by pleading not guilty, after which he is acquitted if he is indeed found to be not guilty and convicted otherwise.

An important implication of our result is that as in this simple example, the justice system cannot gain from keeping the charge secret during a pretrial. This is established practice in modern codes of procedures but was not always the case. Compare, for example, today's Austrian criminal code of procedure[6] with the code of 1803[7]. While the modern code grants the defendant the right to learn about all potential charges, the version from 1803 gives the court of inquiry much more discretion in the extent of information release to the defendant, stating that he has to be informed only as far as necessary to notify him that he is accused.
Kittler (2003) argues that this observation is in line with the broader development of continental European criminal procedure from the medieval inquisitorial proceedings, which exhibited secret charges, to modern forms of criminal law proceedings.[8]

6. §6 (2) StPO: www.jusline.at/gesetz/stpo/paragraf/6
7. II.3 Von Untersuchung des Beschuldigten und dem Verhöre §331
8. Maybe the most famous defendant who is not informed about the charges he faces is Josef K., the protagonist in Franz Kafka's novel *The Trial*. In fact, Kittler (2003) suggests that Kafka, who took multiple courses in legal history before completing his law degree at the University of Prague, might have based this *process* not on the proceeding standards during his time but on medieval ones.

3.3 Model

For concreteness, in the remainder of the paper, we take the binary decision to be the allocation of a single indivisible good. The principal (she) decides whether to allocate the good to the agent (he).

Types. The agent is characterized by his type T, which takes values $t \in \mathbb{R}$. We assume that the type is the agent's private information and that the set of possible types $\mathcal{T}$ is finite.

The principal privately observes a signal S with realizations $s \in \mathcal{S} \subset \mathbb{R}$, finite and ordered. This signal contains information about the agent's type: T and S are jointly distributed according to distribution $\left(f_{t,s}\right)_{t\in\mathcal{T},s\in\mathcal{S}} >> 0$. The signal satisfies the *Monotone Likelihood Ratio Property*:

$$(\textbf{MLRP}) \;\; \forall t, t' \in T \text{ with } t < t' : \frac{f_{t',s}}{f_{t,s}} \text{ is nondecreasing in } s.$$

This is equivalent to requiring that T and S be *affiliated*. It implies that a higher signal is more indicative of a higher type.

Preferences. The principal derives valuation $v(t)$ when allocating the good to an agent of type t. We normalize the value she derives from not allocating to 0. Therefore, v represents the *net* value for the principal. We assume that $v(t)$ is nondecreasing and that there are $t', t'' \in \mathcal{T}$ with $v(t') < 0 < v(t'')$ (otherwise, the principal could implement the efficient allocation decision without the agent's private information).

The signal S provides costless information about the agent's type to the principal, but does not affect her payoff from allocating the good directly. In an extension, we investigate the case in which the signal also directly affects the value of the allocation.

An agent of type t receives utility $u(t) > 0$ from the good. The agent's payoff from not receiving the good is zero.

Verification. The principal has the option to learn the realization of T after paying a cost $c > 0$. Verification is perfect; the principal always learns the exact type.[9]

Solution Concept. The principal can announce and commit to a verification and allocation mechanism before she learns her private signal. The realization of the signal is contractible. The agent learns his type (but not the signal) and then plays a Bayesian best response in the game that is induced by the mechanism. We are interested in characterizing mechanisms that maximize the principal's ex-ante expected payoff in equilibrium.

9. It turns out that with commitment over subsequent allocation decisions, whether the principal learns the agent's real type or just learns whether the information he provided is wrong does not alter the results, as long as she is certain of what she learned. Erlanson and Kleiner (2019) consider the extension to the case with imperfect verification.

3.3.1 Mechanisms

The principal can design and commit to any mechanism before her own signal or the agent's type realize. The signal is contractible, therefore, the mechanism can be contingent on the signal. A key question in this setting is: Can the principal use her private information — the signal— to elicit the agent's information? To answer this question we need to study mechanisms in which the principal could potentially reveal some of her private information to the agent. In the appendix, we, therefore, define a broad class of dynamic mechanisms. As is shown in the appendix, we can restrict our attention to a simpler class of mechanisms when we search for the optimal mechanism in this setting:

We call a mechanism of the following form **direct**: It asks the agent to report his type. Based on this report t and the signal realization s one of these three distinct events occurs:

(1) with probability $x(t,s)$ the good is allocated to the agent

(2) with probability $z(t,s)$ the agent is verified. Then, the good is allocated to him if and only if he is found to have reported truthfully

(3) with probability $1-x(t,s)-z(t,s)$ the good is not allocated to the agent and he is also not verified.

Such a mechanism is called **truthful** if it incentives the agent to report truthfully.

Theorem 3.1. *There is a direct truthful mechanism which maximizes the principal's ex ante welfare.*

Note that this theorem combines a revelation principle with optimality considerations. In the appendix we first derive a suitable revelation principle that is reminiscent of the revelation principles in Ben-Porath, Dekel, and Lipman (2014) and Akbarpour and Li (2019). Then, we show that the optimal mechanism needs to satisfy two intuitive properties: *Maximal Punishment:* if an agent is revealed to have reported $\hat{t}$ different from his actual type t, he is awarded the good with probability 0. *Minimal Verification*: Following (t,s), the agent is verified only if, after his report is verified to be true, he receives the good for sure.

It is worth noticing that these truthful direct mechanisms allow the principal to strategically release information to the agent. To see this, take a mechanism that is not in direct form: Suppose the principal commits to reveal the realization of a garbling of her signal to the agent and then let him send some message to her. Different realizations of the garbling induce different beliefs about the signal of the agent when he sends his message. By designing the garbling in this mechanism the principal can design the additional information of the agent.

The revelation principle that we prove in the appendix shows that we can replicate the outcome of this information design mechanism with a direct mechanism. The principal commits to internally simulate the original mechanism and the agent's response when the agent reports his type to her in the direct mechanism. Although the agent's belief in the direct mechanism is not changed before he reports, the simulated agent's belief is.

We can encode every direct mechanism by two matrixes (x, z) specifying the respective probabilities for all combinations of type reports and signal realizations. Feasibility requires that the total allocation probability $x_{t,s} + z_{t,s} \leq 1$ for all $(t, s) \in \mathcal{T} \times \mathcal{S}$.

3.3.2 The agent's Problem

The agent's preferences are such that he only cares about the probability of receiving the good. Consider the incentive problem of an agent of type t. He does not know the signal realization. If he reports truthfully, he faces the random allocation probability $x_{t,S} + z_{t,S}$. Whether his report is verified is irrelevant for him. If, however, t reports $\hat{t} \neq t$, he receives the good with random probability $x_{\hat{t},S}$ only if he is not verified. Therefore, the Bayesian incentive constraint (BIC) reads as follows:

$$\forall t, \hat{t} \in \mathcal{T} : u(t) \cdot \mathbb{E}_S\left[x_{t,S} + z_{t,S} \mid T = t\right] \geq u(t) \cdot \mathbb{E}_S\left[x_{\hat{t},S} \mid T = t\right].$$

As we assume that every type derives strictly positive utility from the good ($u(t) > 0$), it follows that the intensity of type t's preferences can be eliminated from the IC constraint: the agent simply maximizes his expected allocation probability. The utility he derives from the good is the same irrespective of whether he reported truthfully or not:

$$(BIC_{t,\hat{t}}) : \mathbb{E}_S\left[(x_{t,S} + z_{t,S} - x_{\hat{t},S}) \ \mathbb{1}_{\{T=t\}}\right] = \sum_{s \in \mathcal{S}} f_{t,s}\left[x_{t,s} + z_{t,s} - x_{\hat{t},s}\right] \geq 0.$$

The expected allocation probability at a certain misreport is not independent of the true type, as different types have different conditional beliefs over the distribution of S. The interim expectations are therefore insufficient to describe the mechanism.[10] In most settings without transfers and independence (Ben-Porath, Dekel, and Lipman, 2014; Erlanson and Kleiner, 2019) all types have the same interim expectations about utilities from deviations. In checking the incentive compatibility of a mechanism, it is then sufficient to restrict attention to deviations of the type with the lowest expected utility. This approach does not work in our setting.

10. This is a common feature of mechanism design with correlation: the different expected utility stemming from different beliefs is precisely how mechanisms with money exploit correlation to extract surplus.

3.3.3 The Principal's Problem

The principal designs a mechanism that maximizes her expected utility from the allocation net of the costs of verification. If the good is assigned without verification, she gains $v(T)$. In the case of allocation with prior verification, she additionally pays cost c. Hence, the principal's problem can be stated as the following linear program:

$$(LP) \qquad \max_{(x,z)\geq 0} \mathbb{E}\left[\; x_{T,S}\; v(T) + z_{T,S}\;(v(T)-c)\;\right]$$
$$\text{s.t. } \forall t,\hat{t} \in \mathcal{T} : \qquad (BIC_{t,\hat{t}}) \qquad \text{and}$$
$$\forall (t,s) \in \mathcal{T}\times\mathcal{S} : \quad x_{t,s} + z_{t,s} \leq 1.$$

3.4 Optimal Transparent Mechanisms

We call a direct mechanism transparent if the mechanism is such that the agent would report his type truthfully even if he had learned the realization of S before reporting. In our setting, transparency coincides with ex-post incentive compatibility (EPIC). The ex-post incentive constraints read as follows:

$$\forall s \in \mathcal{S},\ \forall t,\hat{t} \in \mathcal{T} : \qquad (EPIC(s)_{t,\hat{t}}) : \qquad x_{t,s} + z_{t,s} - x_{\hat{t},s} \geq 0.$$

The Bayesian incentive constraint $(BIC_{t,\hat{t}})$ is a weighted sum of the corresponding EPIC constraints over all signal realizations. A transparent mechanism is therefore necessarily BIC. In this section, we solve for the optimal transparent mechanisms. In the remainder of the paper, we show that this mechanism will also be optimal in the wider set of Bayesian incentive compatible mechanism.

Lemma 3.2. *The optimal transparent mechanism is as follows: For all $s \in \mathcal{S}$,*

$$\begin{cases} x_{t,s} = 1,\ z_{t,s} = 0 & \text{if } \mathbb{E}_T[\;(v(T) - (v(T)-c\;)^+ \mid S = s\;] > 0 \\ x_{t,s} = 0,\ z_{t,s} = \mathbb{1}_{\{v(t)>c\}} & \text{otherwise.} \end{cases}$$

The good is allocated without verification whenever the signal alone (without considering the type report) makes the principal sufficiently optimistic about the allocation value. If she is not convinced by the signal, she will only allocate after the successful verification of the agent's report. This happens if the reported allocation value exceeds the costs of verification. The induced allocation rule is not ex-post efficient. At low signals and high types, the allocation value may be positive but smaller than the verification cost so that the good is not allocated. At high signals and low types, the good may be allocated even though $v(t) < 0$.

Implementation. The optimal transparent mechanism can be implemented as a provisional decision: the principal bases her initial decision only on the signal but gives the agent the option to appeal this decision, if the allocation net verification costs is profitable. If the agent appeals, the principal verifies the agent's claim and allocates if she finds that the agent told the truth.

Appealing is weakly dominant for the agent in this implementation game, even if the appeal has no chance of success. One can prevent this multiplicity of equilibria if the designer commits to allocating the good with small probability ϵ after a negative provisional decision that is not appealed by the agent. For any $\epsilon > 0$, the unique best response of the agent is to only appeal if the appeal will be successful. For ϵ converging to zero, the loss in efficiency for the principal (compared with the optimal transparent mechanism) goes to zero, as well.

Proof of Lemma 3.2. For a given $s \in \mathcal{S}$, let $x_{\cdot,s}$ denote the vector $(x_{t,s})_{\{t\in\mathcal{T}\}}$ and similarly for $z_{\cdot,s}$.

Step 0: For any $s \in \mathcal{S}$, the optimal $(x_{\cdot,s}, z_{\cdot,s})$ can be determined separately, as all constraints only involve allocation and verification probabilities for the same signal realization. The principal's optimal expected value is the weighted sum of the values of these subproblems:

$$(LP(s)) \qquad \max_{(x_{\cdot,s},z_{\cdot,s})\geq 0} \mathbb{E}_T \left[\, x_{T,s} \; v(T) \; + z_{T,s} \; (v(T) - c) \mid S = s \, \right]$$
$$\text{s.t. } \forall t, \hat{t} \in \mathcal{T} : \quad (EPIC(s)_{t,\hat{t}}) \qquad \text{and}$$
$$\forall t \in \mathcal{T} : \qquad x_{t,s} + z_{t,s} \leq 1.$$

Step 1: *For any $s \in \mathcal{S}$ and for all $t, \hat{t} \in \mathcal{T}$: $x_{t,s} = x_{\hat{t},s}$, i.e., the allocation probability $x_{\cdot,s}$ has to be constant in the report.*

Suppose to the contrary that there were reports t and $\hat{t}$ with $x_{\hat{t},s} > x_{t,s}$. Ex-post incentive compatibility implies that for all $\tilde{t} \in \mathcal{T}$, we have $x_{\tilde{t},s} + z_{\tilde{t},s} \geq x_{\hat{t},s} > x_{t,s}$. Hence, there could not be a type with a binding incentive constraint regarding the report t. This, in turn, implies that optimally, $z_{t,s} = 0$. If it were positive, $z_{t,s}$ could be lowered and $x_{t,s}$ could be increased, at least until the strict inequality above binds. This leaves the allocation probabilities unchanged but lowers verification costs.

The incentive constraints of type t now take the form $x_{t,s} + 0 \geq x_{\tilde{t},s}$ for all reports $\tilde{t}$ and, in particular, for report $\hat{t}$, contradicting the above hypothesis. Hence, we must have that for all t, $\hat{t}$: $x_{t,s} = x_{\hat{t},s} \equiv \chi_s$.

Step 2: With constant $x_{\cdot,s}$, all incentive constraints are automatically fulfilled, as the unverified allocation probability is the same for any possible report. The principal's problem reads as follows:

$$(\text{LP(s)}) \qquad \max_{(\chi_s, z_{\cdot,s})\geq 0} \sum_{t\in\mathcal{T}} f_{t,s}\left[\chi_s \; v(t) + z_{t,s} \; (v(t)-c)\right]$$
$$\text{s.t.} \quad \forall t \in \mathcal{T} : \chi_s + z_{t,s} \leq 1.$$

In this simplified program, $z_{t,s}$ will be set as high as possible, i.e., to $1-\chi_s$ if $(v(t,s)-c)$ is positive and to 0 otherwise, yielding the following:

$$(LP(s)) \qquad \max_{\chi_s\in[0,1]} \quad \chi_s \cdot \sum_{t\in\mathcal{T}} f_{t,s} \; v(t) \quad + \quad \sum_{t\in\mathcal{T}} f_{t,s} \; (1-\chi_s) \; (v(t)-c)^+.$$

Step 3: Expressed in terms of conditional expectations, the problem is linear in χ_s:

$$(LP(s)) \qquad \max_{\chi_s\in[0,1]} \quad \chi_s \cdot \mathbb{E}_T[v(T)| \; S = s \;] + (1-\chi_s)\cdot \mathbb{E}_T[(v(T)-c)^+| \; S = s \;].$$

Generically, the optimal value of χ is either 0 or 1, depending on which of the expectations is larger.

□

Because of the MLRP, the principal is more optimistic about the agent's type when she observes higher signals. This results in an intuitive cut-off form of the optimal transparent mechanism.

Corollary 3.3.
The optimal transparent mechanism is given by the following cut-off rule:

- *If the signal is above the cut-off $\bar{s}$ the good is allocated without verification (x=1).*
- *If the signal is below the cut-off $\bar{s}$, the good is allocated if and only if the allocation value net verification $v(t,s)-c$ is positive. In this case, the agent is always verified (z=1).*

Formally,

$$x_{t,s} = \mathbb{1}_{\{s \geq \bar{s}\}} \text{ and } z_{t,s} = \mathbb{1}_{\{s < \bar{s}\}} \cdot \mathbb{1}_{\{v(t) > c\}}.$$

The cut-off $\bar{s}$ is uniquely characterized by

$$\bar{s} = \min\left\{s \mid \mathbb{E}_T[\; (v(T) \mid S = s\;] > \mathbb{E}_T[\; (v(T) - c\;)^+ \mid S = s\;]\right\}.$$

Proof. As S and T are affiliated, the function $s \mapsto \mathbb{E}_T[v(T) - (v(T)-c)^+| S = s\;]$ is nondecreasing. Hence, there is a unique cut-off $\bar{s}$ such that

$$\begin{cases} E_T[v(T) - (v(T)-c)^+| S = s\;] \geq 0, & \forall s \geq \bar{s} \\ E_T[v(T) - (v(T)-c)^+| S = s\;] < 0, & \forall s < \bar{s}. \end{cases}$$

□

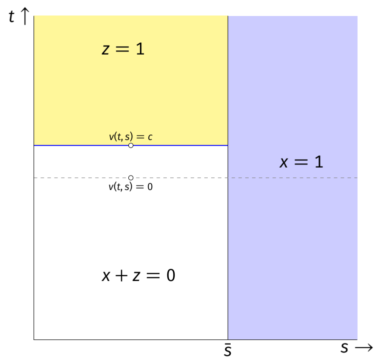


Figure 3.1. Transparent cut-off mechanism

Figure 3.1 sketches the cut-off mechanism. If the signal realization is above the cut-off, $\bar{s}$, the principal is optimistic about T and allocates the good to the agent without verification $(x = 1)$, irrespective of his reported type. If the signal is below the cut-off, the agent can receive the good only after his type report is verified $(z = 1)$ to be above a threshold so that $v(t,s) - c$ is positive. It is easy to see that this mechanism is transparently implementable. Either the signal is such that the agent gets the good independent of his type (the shaded blue area to the right of $\bar{s}$), or he can only get the good after being verified (the shaded yellow area above the horizontal line and to the left of $\bar{s}$) so that misreporting a higher or lower type cannot be beneficial even when the agent knows which signal s has realized.

3.5 Optimal BIC Mechanisms

This section shows that the optimal transparent mechanism is also optimal in the broader class of BIC mechanisms. This implies that the principal cannot exploit the privacy of her signal; transparency comes at no cost.

Proposition 3.4.
The optimal transparent mechanism is optimal in the class of Bayesian mechanisms.

To gain intuition on why the principal cannot exploit different types' beliefs, we refer back to the example in the introduction. Analogous to the method of Crémer and McLean (1988), the way in which the principal could potentially save verification costs with a nontransparent mechanism was to verify the agent who reports a high type only if the signal is low and not to reveal the signal realization to the agent.[11] Two features of the costly verification setting rule out that such an improvement exists. First, recall that in the example, the improvement was constructed by shifting verification probability to the lower signal, which is more likely to occur

11. As otherwise, a low type would misreport after observing a high signal.

from the low type's perspective, and that this was the only type with an incentive to deviate. The lack of monetary transfers implies that the relevant incentive constraints in verification mechanisms are those that impede reports from less-favorable toward more favorable types. Hence, it is favorable for all *relevant* incentive constraints to shift verification probability toward lower signals.[12] Second, the signal realization does not affect the principal's valuation conditional on the agent's type. This allows a shift in the allocation probability of lower types toward exactly those signal realizations after which a misreport may be fruitful (high signals that feature less verification).[13]

The optimal transparent mechanism in the last section already distributes allocation and verification probabilities in the way that this intuition suggests. The proof of Proposition 1 makes use of the first feature by relaxing the problem and considering only a specific class of upward incentive constraints. Within the relaxed problem, we make use of the second fact to apply feasible deviations to an arbitrary mechanism that make the principal better off, and finally yield the optimal transparent mechanism.

Proof. We show that the cut-off mechanism from Corollary 3.3 with

$$\bar{s} = \min\left\{s \mid \mathbb{E}_T[\ (v(T) \mid S = s\ ] > \mathbb{E}_T[\ (v(T) - c\ )^+ \mid S = s\ ]\right\}$$

solves the following relaxation of the problem, which proves that it is a solution to the original LP. Define the set of profitable types as those t with a positive allocation value,

$$\mathcal{T}^+ \equiv \{t \in \mathcal{T} | v(t) > 0\},$$

and the unprofitable types accordingly as $\mathcal{T}^- \equiv \mathcal{T} \backslash \mathcal{T}^+$. Both sets are non-empty by the assumption that v crosses 0. Otherwise, the optimal mechanism is trivial.

The relaxed problem includes only those incentive constraints that prevent types in $\mathcal{T}^-$ from misreporting types in $\mathcal{T}^+$. Hence, it reads as follows:

$$\text{(LP.r)} \qquad \max_{(x,z)\geq 0} \quad \sum_{t\in\mathcal{T}}\sum_{s\in\mathcal{S}} f_{t,s}\left[x_{t,s}\ \ v(t) + z_{t,s}\ \ (v(t)-c)\right]$$

$$\text{s.t.} \quad \forall t \in \mathcal{T}^-,\ \forall \hat{t} \in \mathcal{T}^+ : \quad (BIC_{t,\hat{t}}) \quad \text{and}$$

$$\forall (t,s) \in \mathcal{T} \times \mathcal{S} : \quad x_{t,s} + z_{t,s} \leq 1.$$

In the remainder of the proof, we derive feasible changes to a solution to the relaxed problem, which do not lower the principal's value and which finally lead to

12. If higher types had an incentive to deviate downward, it may save verification cost to verify after observing high signals for some type reports. This does not occur in our setting.

13. We want to highlight that in settings with monetary transfers, even payoff-irrelevant correlated signals allow for full surplus extraction (Riordan and Sappington, 1988). In the next section, we consider the case in which the signal has a direct effect on the principal's value.

the cut-off mechanism. We make repeated use of the following notation: we denote changes in the allocation probability by $dx_{t,s}$ so that the new probability after the change is given by $x_{t,s} + dx_{t,s}$. $dx_{t,s}$ may be positive or negative. Analogously for $dz_{t,s}$. Further, $d(BIC_{t,\hat{t}})$ denotes the change in surplus utility that type t receives from reporting the truth rather than misreporting $\hat{t}$, which is induced by a change of the above form. Recall that the constraint $(BIC_{t,\hat{t}})$ reads as $\sum_s f_{t,s}\left[x_{t,s} + z_{t,s} - x_{\hat{t},s}\right] \geq 0$ so that $d(BIC_{t,\hat{t}})$ denotes the change to the left-hand side of this inequality.

The value for the principal is given by

$$V = \sum_{t\in\mathcal{T}} \sum_{s\in\mathcal{S}} f_{t,s}\left[x_{t,s}\, v(t) + z_{t,s}\, (v(t) - c)\right],$$

and dV will denote the induced change to this value.

Step 1: The optimal mechanism in the relaxed problem features $\forall t \in \mathcal{T}^- \ \forall s \in \mathcal{S} : \ z_{t,s} = 0$:

Suppose $z_{t,s} > 0$ for some type $t \in \mathcal{T}^-$. Shifting probability mass from $z_{t,s}$ to $x_{t,s}$ such that the overall allocation probability stays constant,

$$0 \ < \ dx_{t,s} \ = \ -dz_{t,s},$$

saves the principal verification costs and does not distort the incentives, as type t's incentive to misreport remains the same, and all incentive constraints to misreport a type $t \in \mathcal{T}^-$ are ignored in the relaxed problem.

Step 2: There is an optimal mechanism in the relaxed problem featuring a cut-off form for $x_{\hat{t},\cdot}$:

$$\forall \hat{t} \in \mathcal{T}^+ \ \exists \tilde{s}(\hat{t}) \in \mathcal{S} : \ x_{\hat{t},s} \begin{cases} = 0 & \text{if } s < \tilde{s}(\hat{t}) \\ \in [0,1) & \text{if } s = \tilde{s}(\hat{t}) \\ = 1 & \text{if } s > \tilde{s}(\hat{t}) \end{cases}.$$

Take a feasible IC mechanism of the relaxed problem featuring that for some $\hat{t} \in \mathcal{T}^+$, $\exists s < s' \in \mathcal{S}$ such that $x_{\hat{t},s} > 0$, $x_{\hat{t},s'} < 1$.

Modify the mechanism only at two points, shifting allocation probability mass from $x_{\hat{t},s}$ to $x_{\hat{t},s'}$, i.e. $dx_{\hat{t},s} < 0$ and $dx_{\hat{t},s'} > 0$. Choose these shifts in a proportion, such that for the highest unprofitable type, $\tilde{t} \equiv \max \mathcal{T}^-$, the incentive to misreport $\hat{t}$ remains unchanged:

$$0 \stackrel{!}{=} d(IC_{\tilde{t},\hat{t}}) = -f_{\tilde{t},s}\, dx_{\hat{t},s} - f_{\tilde{t},s'}\, dx_{\hat{t},s'} = 0 \Leftrightarrow dx_{\hat{t},s} = -\frac{f_{\tilde{t},s'}}{f_{\tilde{t},s}}\, dx_{\hat{t},s'}.$$

For all types $t \in \mathcal{T}^-$, we have $t \leq \tilde{t}$, and, therefore,

$$d(BIC_{t,\hat{t}}) = -f_{t,s}\, dx_{\hat{t},s} - f_{t,s'}\, dx_{\hat{t},s'} = f_{t,s}\left[\frac{f_{\tilde{t},s'}}{f_{\tilde{t},s}} - \frac{f_{t,s'}}{f_{t,s}}\right] dx_{\hat{t},s'} \geq 0$$

by the monotone likelihood ratio property.

The principal's value changes in the following way:

$$\begin{aligned} dV =& f_{\hat{t},s}\ dx_{\hat{t},s}\ v(\hat{t}) + f_{\hat{t},s'}\ dx_{\hat{t},s'}\ v(\hat{t}) \ = \ f_{\hat{t},s}\ \left[-\frac{f_{\tilde{t},s'}}{f_{\tilde{t},s}}\ dx_{\hat{t},s'}\right] v(\hat{t}) + f_{\hat{t},s'}\ dx_{\hat{t},s'}\ v(\hat{t}) \\ =& f_{\hat{t},s}\left[\frac{f_{\hat{t},s'}}{f_{\hat{t},s}} - \frac{f_{\tilde{t},s'}}{f_{\tilde{t},s}}\right] dx_{\hat{t},s'}\ v(\hat{t}) \ \geq \ 0, \end{aligned}$$

since $dx_{\hat{t},s'} > 0$ and $\hat{t} \in \mathscr{T}^+$ which implies both $v(\hat{t}) \geq 0$ and $\hat{t} > \tilde{t}$.

The proposed shift is clearly feasible if in the original mechanism, $x_{\hat{t},s'} + z_{\hat{t},s'} < 1$.

In the case that $x_{\hat{t},s'} + z_{\hat{t},s'} = 1$, it can still be implemented by shifting in addition mass from $z_{\hat{t},s'}$ to $z_{\hat{t},s}$ to ensure that $x_{\hat{t},s'} + z_{\hat{t},s'}$ and $x_{\hat{t},s} + z_{\hat{t},s}$ remain constant:

$$dx_{\hat{t},s'} + dz_{\hat{t},s'} \ = \ 0 \quad \text{and} \quad dx_{\hat{t},s} + dz_{\hat{t},s} \ = \ 0.$$

This implies $dz_{\hat{t},s'} < 0$ and $dz_{\hat{t},s} > 0$. This is feasible, as $x_{\hat{t},s'} < 1$ and $x_{\hat{t},s'} + z_{\hat{t},s'} = 1$ imply that $z_{\hat{t},s'} > 0$. As $x_{\hat{t},s} > 0$, we must further have $z_{\hat{t},s} < 1$ by feasibility.

The above changes in x imply for z the following:

$$dx_{\hat{t},s} = -\frac{f_{\tilde{t},s'}}{f_{\tilde{t},s}} dx_{\hat{t},s'} \Leftrightarrow dz_{\hat{t},s} = \frac{f_{\tilde{t},s'}}{f_{\tilde{t},s}}(-dz_{\hat{t},s'}).$$

The incentives for any lower type to misreport his type as $\hat{t}$ are weakened in the same way as above because $z_{\hat{t},s}$ and $z_{\hat{t},s'}$ do not play a role in the constraints that prevent misreport $\hat{t}$.

Finally, the principal's value now changes by

$$\begin{aligned} dV =& f_{\hat{t},s}\left[dx_{\hat{t},s}\ \ v(\hat{t}) + dz_{\hat{t},s}\ \ (v(\hat{t}) - c)\right] + f_{\hat{t},s'}\left[dx_{\hat{t},s'}\ \ v(\hat{t}) + dz_{\hat{t},s'}\ \ (v(\hat{t}) - c)\right] \\ =& -c\left[f_{\hat{t},s}\ \ dz_{\hat{t},s} + f_{\hat{t},s'}\ \ dz_{\hat{t},s'}\right] \\ =& -c\ \ f_{\hat{t},s}\left[\frac{f_{\tilde{t},s'}}{f_{\tilde{t},s}} - \frac{f_{\hat{t},s'}}{f_{\hat{t},s}}\right](-dz_{\hat{t},s'}) \ \geq \ 0, \end{aligned}$$

as, by MLRP, the term in squared brackets is negative and, by assumption, $-dz_{\hat{t},s'} \geq 0$.

Step 3: There is an optimal mechanism in the relaxed problem featuring $x_{\hat{t},\cdot} = x_{\hat{\hat{t}},\cdot}$ for all $\hat{t}, \hat{\hat{t}} \in \mathscr{T}^+$:

By the cut-off structure established in Step 2, $x_{\hat{t},\cdot} = (0, \ldots, 0, x_{\hat{t},\tilde{s}(\hat{t})}, 1, \ldots, 1)$ for all $t \in \mathscr{T}^+$. Suppose to the contrary that $x_{\hat{t},\tilde{s}(\hat{t})} + \sum_{s>\tilde{s}(\hat{t})} 1 > x_{\hat{\hat{t}},\tilde{s}(\hat{\hat{t}})} + \sum_{s>\tilde{s}(\hat{\hat{t}})} 1$ for some $\hat{t}, \hat{\hat{t}} \in \mathscr{T}^+$.

Replacing $x_{\hat{\hat{t}},\cdot}$ by $x_{\hat{t},\cdot}$ does not generate new incentives to misreport, but it increases the principal's expected value, as it increases the allocation probability for profitable types. If feasibility is hurt, i.e., $x_{\hat{t},s} + z_{\hat{\hat{t}},s} > 1$ for some $s \in \mathscr{S}$, decrease $z_{\hat{t},s}$ until $x_{\hat{t},s} + z_{\hat{\hat{t}},s} = 1$. This is also a strict improvement for the principal, as she saves verification costs.

Step 4: There is an optimal mechanism in the relaxed problem featuring $x_{\hat{t},\cdot} = x_{\hat{\hat{t}},\cdot}$ for all $\hat{t}, \hat{\hat{t}} \in \mathscr{T} = \mathscr{T}^+ \cup \mathscr{T}^-$:

Fix some unprofitable type $t \in \mathscr{T}^-$. By Step 1, we have $z_{t,\cdot} = 0$. Optimally, the principal wants to choose the lowest possible allocation probability for the unprofitable types. However, she needs to grant him at least the same interim allocation probability that he could achieve by misreporting to be a profitable type $\hat{t} \in \mathscr{T}^+$ (By steps 2–3, we know that $x_{\hat{t},\cdot}$ is the same for all $\hat{t} \in \mathscr{T}^+$). As the signal realization has no effect on the allocation value, the principal is indifferent between any allocation vector $x_{t,\cdot}$ which induces the same interim allocation probability: $\mathbb{E}[x_{t,S}|T = t]$. Formally,

$$\mathbb{E}[v(t,S)x_{t,S}|T = t] = v(t)\mathbb{E}[x_{t,S}|T = t].$$

Therefore, she can grant the unprofitable types just the same allocation lottery, they would face if they would misreport to a profitable type: $x_{t,\cdot} = x_{\hat{t},\cdot}$.

In the section about optimal EPIC mechanisms, we have shown how, for given s, a constant allocation $x_{\cdot,s}$ in t implies that the principal's problem is reduced to the choice between allocating without verification at all reports and allocating after verification only if the report is such that $v(t) - c \geq 0$. This concludes the proof for the first case of the theorem, showing that the cut-off mechanism solves the relaxed problem and is therefore optimal in the original problem.

□

3.5.1 Application: pre-trial

Consider a defendant in court (the agent) who privately knows whether he is of guilty or innocent type $t \in \{G = 1, I = 0\}$. The judge (principal) privately observes the realization of a signal as a result of a pre-trial investigations and asks the agent to plead either guilty or innocent (report his type). The judge decides whether to conduct a costly trial (reveal the agent's type) and whether the defendant should be acquitted as a function of the type-signal report and the outcome of the trial, in case it is conducted. A trial requires verification cost $c > 0$. The defendant's utility is 1 from acquittal and 0 if he is convicted, irrespective of his type.

The system wishes justice to prevail. If the allocation decision is identified with acquitting the defendant the problem can be modeled by

$$v(t,s) = v(t) = t - r.$$

For some $r \in (0,1)$. If an innocent agent ($t = 0$) is convicted ($x = 0$) the net utility loss is given by $0 - v(1) = -(1 - r)$. If a guilty agent ($t = 0$) is acquitted ($x = 1$) it is given by $v(0) - 0 = -r$.

The code of procedure prescribes the mechanism in such a setting. The optimal mechanism we derive resembles the proceedings of a pretrial: The case is dismissed if the signal for the defendant's innocence is strong enough, i.e. the charge is weak. If the signal for innocence is below this cutoff, the agent can plead guilty and is convicted, or can request a trial by pleading not guilty, after which he is acquitted if he is indeed found to be not guilty and convicted otherwise.

An important implication from the transparency of the optimal mechanism is that the justice system cannot gain from keeping the charge secret during a pretrial. This is established practice in modern codes of procedures but was not always the case. Compare for example today's Austrian criminal code of procedure[14] with the code of 1803[15]. While the modern code grants the right to learn about all potential charges to the defendant, the version from 1803 gives the court of inquiry much more discretion in the extent of information release to the defendant, stating that he has to be informed only as far as necessary to notify him that he is accused.
Kittler (2003) argues that this observation is in line with the broader development of continental European criminal procedure from the medieval inquisitorial proceedings, which exhibited secret charges, to modern forms of criminal law proceedings.[16]

3.6 Direct Signal Effects

In the previous sections, we characterized optimal mechanisms under the assumption that the signal that the principal privately observes had no direct effect on her allocation value.

In this section, we deviate from this assumption: suppose that the principal derives valuation $v : \mathscr{T} \times \mathscr{S} \to \mathbb{R}$ when allocating the good to an agent of type t at signal s. We normalize the value she derives from not allocating again to 0 and assume that $t \mapsto v(t,s)$ is nondecreasing for any $s \in \mathscr{S}$. Note that the agent's problem does not change, so neither do the incentive constraints for the transparent mechanism ($EPIC(s)_{t,\hat{t}}$) nor for the Bayesian ($BIC_{t,\hat{t}}$). The only change is in the objective of the principal, which reads as follows:

$$\max_{(x,z)\geq 0} \mathbb{E}\left[\; x_{T,S}\;\; v(T,S) + z_{T,S}\;\; (v(T,S) - c)\;\; \right].$$

We first characterize the optimal transparent mechanism. This characterization and the proof (in the Appendix) are analogous to Lemma 3.2. The reason is that by transparency the optimal allocation and verification vector can again be determined

14. §6 (2) StPO: www.jusline.at/gesetz/stpo/paragraf/6
15. II.3 Von Untersuchung des Beschuldigten und dem Verhöre §331
16. Maybe the most famous defendant who is not informed about the charges he faces is Josef K., the protagonist in Franz Kafka's novel The Trial. In fact, Kittler (2003) suggests that Kafka, who took multiple courses in legal history before completing his law degree at the University of Prague, might have based this "process" not on his contemporary but the medieval proceeding standards.

for each signal realization $s \in \mathscr{S}$ separately so that the dependence of the value of the allocation on the signal does not change the optimal solution.

Lemma 3.5. *The optimal transparent mechanism is as follows: For all $s \in \mathscr{S}$,*

$$\begin{cases} x_{t,s} = 1, \ z_{t,s} = 0 & \text{if } \mathbb{E}_T[\ (v(T,s) - (v(T,s) - c\ )^+ \mid S = s\ ] > 0 \\ x_{t,s} = 0, \ z_{t,s} = \mathbb{1}_{\{v(t,s)>c\}} & \text{otherwise.} \end{cases}$$

The characterization of the optimal transparent mechanism in Lemma 3.5 holds true for any functional form of the principal's value $v(\cdot,\cdot)$, which is increasing in the agent's type. For the remainder of the paper, we distinguish between two cases in which the direct effect is either positive or negative. We refer to the setting where there is no direct effect as case 1.

Case 2: Positive Direct Effect.
First, we assume that for all t, $s \mapsto v(t,s)$ is increasing. For example the valuation could take the functional form $v(t,s) = t + s - r$, where the principal's value from allocating increases linearly in type and signal and r represents her reservation value from not allocating.

In the court example, this corresponds to a situation in which it was beneficial for the justice system that the final verdict confirm the initial charge.

Case 3: Negative Direct Effect.
By contrast, in the case in which for all t, $s \mapsto v(t,s)$ is decreasing, the direct effect is negative. Here, the direct and indirect effects go in opposite directions. This direction seems natural, for example if the principal observes the allocation value of an outside option s, which is affiliated with the allocation value of the agent t. In this example her net utility from allocating could read as follows: $v(t,s) = t - s$. We also assume that in this case, $v(t,s)$ is sufficiently decreasing such that,

$$\textbf{(Assumption 3)}\ \ s \mapsto \mathbb{E}[v(T,s)|S = s] \text{ is decreasing.}$$

3.6.1 Cut-off Mechanisms

First, we establish that in this case, the optimal transparent mechanisms again have a cut-off structure.

Corollary 3.6.

(i) *If the direct effect of the signal is positive (Case 2: $s \mapsto v(t,s)$ is increasing), then the optimal transparent mechanism is given by the following cut-off rule:*

$$x_{t,s} = \mathbb{1}_{\{s \geq \bar{s}\}} \text{ and } z_{t,s} = \mathbb{1}_{\{s < \bar{s}\}} \cdot \mathbb{1}_{\{v(t,s) > c\}}.$$

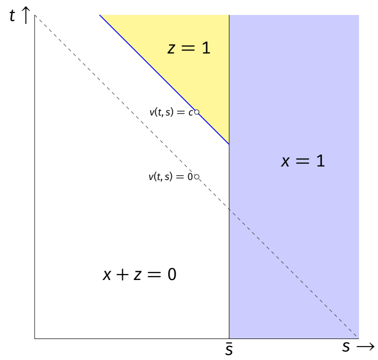


Figure 3.2. Transparent cut-off mechanisms with increasing $v(t, \cdot)$

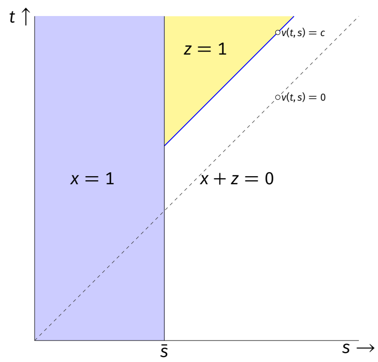


Figure 3.3. Transparent cut-off mechanisms with (sufficiently) decreasing $v(t, \cdot)$

(ii) If the direct effect of the signal is sufficiently negative (Case 3: $s \mapsto isv(t,s)$ is decreasing) and $s \mapsto \mathbb{E}[v(T,s)|S=s]$ is decreasing (Assumption 3), then the optimal transparent mechanism is given by the following cut-off rule:

$$x_{t,s} \;=\; \mathbb{1}_{\{s \;\leq\; \bar{s}\}} \text{ and } z_{t,s} \;=\; \mathbb{1}_{\{s \;>\; \bar{s}\}} \cdot \mathbb{1}_{\{v(t,s) \;>\; c\}}.$$

The cut-off $\bar{s}$ is, in both cases, uniquely characterized by

$$\bar{s} = \min\left\{s \mid \mathbb{E}_T[\ (v(T,s) \mid S=s\ ] > \mathbb{E}_T[\ (v(T,s)-c\ )^+ \mid S=s\ ]\right\}.$$

3.6.2 Optimal Bayesian Mechanisms

If the direct effect is positive, the optimal Bayesian mechanism is again given by the optimal transparent mechanism.

Proposition 3.7.
If the direct effect of the signal is positive (Case 2: $s \mapsto v(t,s)$), the optimal transparent mechanism is also optimal in the class of BIC mechanisms.

The proof is relegated to the Appendix. We again introduce a relaxation by defining a set of profitable types $\mathcal{T}^+$, but now, the relaxed problem has to include incentive constraints ensuring that no type t from the entire $\mathcal{T}$ has an incentive to misreport a $\hat{t} \in \mathcal{T}^+$, which is higher than t. That is, this relaxed problem ignores all downward incentive constraints and all constraints preventing misreports toward types within $\mathcal{T}^- \equiv \mathcal{T} \setminus \mathcal{T}^+$. This requires altering the construction of profitable perturbations, but the intuition is similar to the previous case. The transparent mechanism requires shifting allocation probability toward higher signals. If the principal's value is increasing in the signal, this shift is profitable and, therefore, already a feature of optimal BIC mechanisms.

The equivalence between transparent and Bayesian optimal mechanisms does not extend to the case where the direct effect is negative and $s \mapsto v(t,s)$ is decreasing (Case 3). In this case, the principal may incur loss when she chooses the EPIC optimal mechanism or, equivalently, when she publicizes the realization of the signal before the agent's report.

Proposition 3.8.
If the direct effect of the signal is negative (Case 3: $s \mapsto v(t,s)$ is decreasing), and the optimal transparent mechanism (x,z) features $x_{\hat{t},s} > 0$ and $z_{\hat{t},s'} > 0$ for some $\hat{t} \in \mathcal{T}$ and some $s < s' \in \mathcal{S}$, then there exists a mechanism $(\tilde{x},\tilde{z})$ with a strictly higher value, which is BIC. Hence, the principal profits strictly from S being private.

Note that under assumption 3 ($\mathbb{E}_T[\ (v(T,s) \mid S=s\ ]$ is decreasing in s), the optimal transparent mechanism has the properties stated in the proposition whenever it is nontrivial, i.e. whenever x and z are positive at some combinations (t, s). The proof is again relegated to the Appendix.

As noted previously, the principal saves verification costs by shifting verification probability at high reports toward low signals. Low types, who have an incentive to misreport, find such signals more likely. Transparency then requires a shift in the allocation probability for low types toward the other signal realizations to ensure that at those signals indicating no verification of high reports, the low types have no incentive to deviate. If the principal's value is decreasing in the signal, shifting allocation toward higher signals comes at a cost and, therefore, transparency comes at a cost.

3.7 Discussion and Related Literature

3.7.1 Literature

In settings where monetary transfers are feasible, the principal can design a lottery rewarding the agent for guessing the value of her privately observed signal correctly. Different agent types hold different beliefs over the signal distribution and, therefore, reveal their type by guessing the signal they deem most likely. If the agent's liability is not limited, the principal can increase reward and loss in the lottery to such an extent that the incentives to win the lottery exceed any incentives regarding the allocation decision. In doing so, she can learn the agent's type at arbitrarily small costs. Mechanisms with monetary transfers and correlated information have been discussed by Crémer and McLean (1988), Riordan and Sappington (1988), Johnson, Pratt, and Zeckhauser (1990), and McAfee and Reny (1992), who all establish conditions on the information structure that ensure full surplus extraction by the principal. As all surplus can be extracted, revenue maximization leads to ex-post efficient allocations. Neeman (2004) discusses the genericity of the above-mentioned conditions and shows that full surplus extraction is possible only if every preference type is "determined" by his belief over the correlated characteristics. Even though this condition is fulfilled, in our setting with costly verification instead of monetary transfers, full surplus extraction is not feasible and implementing the ex-post efficient allocation is not optimal for the principal.

The full surplus-extracting lotteries require potentially unbounded transfers. For the case of bounded transfers or limited liability, Demougin and Garvie (1991) show that the qualitative results, the application of rewards as bets on the signal, still apply. Different from our setting, the principal gains by maintaining her signal private.

In the absence of monetary transfers, Bhargava, Majumdar, and Sen (2015) show how positively correlated beliefs among voters allow overcoming the impossibility of nondictatorial voting rules established by Gibbard (1973) and Satterthwaite (1975).

Our result is in line with the findings in other settings where monetary transfers are not feasible but correlated information is absent. The literature (Glazer and

Rubinstein, 2004; Ben-Porath, Dekel, and Lipman, 2017; Hart, Kremer, and Perry, 2017; Erlanson and Kleiner, 2019; Halac and Yared, 2019) has found optimal mechanisms to take a simple cut-off structure and to be EPIC in the sense that the agents would also report truthfully if they were informed about the other agents' type realizations before their report.

The possibility for the mechanism designer to verify an agent's private information at a cost was first introduced by Townsend (1979) considering a principal-agent model for debt contracts, which was extended to a two-period model by Gale and Hellwig (1985). These early models of state verification feature both, monetary transfers and verification. Glazer and Rubinstein (2004) introduce a setting where the principal has to take a binary decision depending on the multidimensional private information of the agent. Here, the principal cannot use monetary transfers, but she can learn about one dimension before making her decision.

Our model is most closely related to that of Ben-Porath, Dekel, and Lipman (2014), who model costly verification and consider the case of allocating a good among finitely many agents whose types are independently distributed; see also the discussion section. Erlanson and Kleiner (2019) study a collective decision problem with costly verification and show that the optimal mechanism is EPIC and can be implemented by a simple weighted majority voting rule. Mylovanov and Zapechelnyuk (2017) consider an allocation problem without monetary transfers in which the principal learns the agents' types without cost but only posterior to the allocation decision and has the ability to punish untruthful reports up to a limit. Halac and Yared (2019) consider a delegation problem and specify conditions on the verification cost that ensure optimality of a threshold mechanism with an escape clause.

Erlanson and Kleiner (2019) show further that the equivalence between BIC and EPIC mechanisms holds more generally rather than only for optimal mechanisms. This relates to Gershkov, Goeree, Kushnir, Moldovanu, and Shi (2013) and Manelli and Vincent (2010), who show equivalence between BIC and DIC mechanisms in settings with monetary transfers. All these results have been derived under the assumption that the private information of players is independently distributed; we deviate from this assumption by introducing correlation between the agent's type and the principal's signal.

As the principal has private information, our model is also related to the informed principal problem cf. Myerson (1983) and Maskin and Tirole (1990). With monetary transfers, Severinov (2008) and Cella (2008) show that correlated information allows for an efficient solution to the informed principal problem. We assume that the principal designs mechanisms with full commitment over allocation procedures before observing the signal. A priori, there is no informed principal problem in our model, but in the discussion section, we show that the optimal mechanism we derive also solves the informed principal problem.

Roadmap. After an example highlighting the difficulties that correlation adds to the canonical verification setup, and showing how our findings advocate for transparency in pretrial investigations, section 2 sets up the model. The characterization of optimal mechanisms starts in section 3 for the class of transparent mechanisms, and section 4 shows that this mechanism is optimal in the broader class of BIC mechanisms. We then extend the analysis to a specification in which the principal's valuation may be affected by the signal realization, and discuss the relation to favored-agent mechanisms and to the informed principal problem.

3.7.2 Favored-agent mechanism

Ben-Porath, Dekel, and Lipman (2014) study the problem of allocating a single indivisible good among several agents. The principal's utility from allocating to an agent is this agent's private information which is assumed to be independently distributed across agents. The principal can perfectly verify any agent's type at an agent-specific cost. Ben-Porath, Dekel, and Lipman (2014) show that optimal mechanisms are so-called *favored-agent mechanisms*, which allocate the object to a predetermined (*favored*) agent if no other agent reports a type above his individual threshold. Whenever there is a type report above the threshold from an agent other than the favored agent, the highest type, net of verification costs, is verified and receives the good conditional on having reported the truth. In our setting, we can interpret the value s as coming from a second agent's type whose verification cost is zero.

We can interpret $v(t,s) = t + s$ as the (net) value of allocating to player one whose type is t. If we let $-s$ be the type of a second player, MLRP between type and signal means that the two players have negatively correlated values. Case 2 of our setting is fulfilled, and the EPIC mechanism from Figure 3.2 is optimal. This EPIC mechanism is essentially a favored-agent mechanism in which player one is the favored agent. If player two's type, $-s$, is low, i.e. s is above the cut-off, agent one always gets the good without being verified. If the signal is below a cut-off, agent one gets the good only if he is verified to have the highest type, net of verification costs, i.e., if $t - c \geq -s - 0$.

Similarly, we can interpret $v(t,s) = t - s$ as the value of allocating to player one if player two's type is s, so that MLRP implies positive correlation. As the utility function $v(t,s) = t - s$ is decreasing in s, case 3 is fulfilled and therefore (proposition 3.8) the EPIC favored agent mechanism is not generally, the optimal mechanism.

The difference between the present paper and Ben-Porath, Dekel, and Lipman (2014) is the correlation between players' types through MLRP. In contrast to their setting, here, different types hold different beliefs and therefore expect different interim allocations for a fixed type report. Hence, one cannot reduce the mechanism to interim allocation probabilities for each report.

Despite the technical differences in solving for the optimal mechanism, we find a close connection to favored-agent mechanisms (which are optimal under independence) and we infer that they remain optimal under negative correlation but not under positive correlation.

3.7.3 Informed principal problem

By our assumptions on the principal's commitment, the mechanism proposed by the designer does not convey information to the agent. However, the fact that the principal's signal can be made public without loss implies that the informed principal game has a separating equilibrium in which the agent perfectly learns the principal's type from the proposal. This implies that our mechanism constitutes a solution to the informed principal problem for Cases 1 and 2 when the EPIC mechanism is optimal.

3.8 Concluding Remarks

This paper studies the role of information in a mechanism design model, in which the principal may use costly verification instead of monetary transfers to incentivize the revelation of private information. We show that a transparent mechanism is optimal. It is without loss for the principal to make her information public before contracting with the agent. Our result gives a rationale for the use of transparent procedures in a variety of applications from hiring to procedural law. This is in contrast with results on correlation in mechanism design problems with money.

In an extension in which the principal's private information also affects her preferences, we characterize the mechanism and show that the above qualities remain if the information and direct effect work in the same direction. In the opposite case, we show how the principal can benefit by ensuring that her signal remains private. Interesting directions for future analysis include the question of whether, with correlation, the equivalence between BIC and transparent (EPIC) mechanisms holds more generally than for optimal mechanisms.

Appendix 3.A Revelation Principle

The revelation principle presented here is close to the revelation principle in Ben-Porath, Dekel, and Lipman (2014), but it takes into account possible issues arising from the correlation between the signal and the type realization.

Pick any (possibly dynamic) mechanism G and an agent strategy s_A that is a best response to this mechanism. Then, there is an equivalent incentive compatible, direct, two-stage mechanism characterized by the pair of functions (e, a),

$$\begin{aligned} e &: \mathscr{T} \times \mathscr{S} \rightarrow [0,1], \\ a &: \mathscr{T} \times \mathscr{T} \cup \{\emptyset\} \times \mathscr{S} \rightarrow [0,1], \end{aligned}$$

of the following form:

(1) The agent reports his type $\hat{t} \in \mathscr{T}$.

(2) Given her signal realization s, the principal verifies the agent's type with probability $e(\hat{t}, s)$.

(3) Depending on the result of this revision $t \in \mathscr{T} \cup \{\emptyset\}$, where $\emptyset$ encodes the event that there was no revision, the principal allocates the good to the agent with probability $a(\hat{t}, t, s)$.

Instead of G, the principal could commit to the following mechanism:

- The agent reports a type $\hat{t} \in \mathscr{T}$.
- Given this report and her signal's realization, s, the principal calculates the marginal probability of verification in the equilibrium in the original game under the condition that the agent's type was $\hat{t}$ and the principal's signal was s:[17]

$$e(\hat{t}, s) := \mathbb{P}(\text{there is verification} | s_A(\hat{t}), S = s).$$

- The principal verifies the agent's true type with this probability: $e(\hat{t}, s)$.
  - If she finds that the agent reported the truth, $\hat{t} = t$, or if she did not verify $t = \emptyset$, she allocates the good with probability that equals the marginal probability of allocation in the original mechanism, conditional on the type being equal to $\hat{t}$ and the signal being equal to s.

$$a(\hat{t}, \hat{t}, s) = a(\hat{t}, \emptyset, s) = \mathbb{P}(\text{allocation} | s_A(\hat{t}), S = s).$$

  - If she verifies and finds out that the agent misreported, i.e., $t \notin \{\hat{t}, \emptyset\}$, the allocation probability is determined in the following way:
    The principal constructs a lottery over all stages in the original mechanism, which have the principal verify the agent with positive probability in equilibrium, conditional on the event that the agent played according to $s_A(\hat{t})$ and the signal was s.
    The probabilities of the lottery are chosen such that they equal the probability of verifying at this stage for the first time, conditional on the event that there is verification at some point in the game.
    Now, she chooses one of these stages according to the above probabilities. She simulates the game from this point onward, assuming that the game had reached this stage and it was found at this point that the agent's true type was t, by letting the simulated agent behave according to what is described in $s_A(t)$ for behavior after this knot and the verification. The strategy s_A contains a plan for the behavior of the agent from this stage onward. The principal simulates his own behavior, as prescribed in the

17. This means the probability that there was verification at any point in the game, specified by G and played by the agent according to $s_A(\hat{t})$, under the condition that signal s realized.

original mechanism.

Note that given any signal realization, this reproduces exactly the same allocation profile as the following deviation strategy for type $t \neq \hat{t}$ (which he could play without knowing the true signal realization s):
The agent of type t imitates type $\hat{t}$'s behavior $s(\hat{t})$ until the first verification, and then sticks to the behavior that the equilibrium strategy prescribes for his type.

If the agent reports the truth, the marginal probabilities of verification and allocation and, therefore, the expected utilities of the agent and the principal are the same in both mechanisms.
However, truth-telling is optimal for the agent in the constructed mechanism, as misreporting yields the exact same outcome as the above-described deviation strategy in the original game and can therefore be not profitable.

There are two further observations that help simplify the class of possible optimal mechanisms. In short, in any optimal mechanism, the principal will chose he highest possible punishment for detected misreports and the highest possible reward for detected truth-telling.

(1) Maximal punishment: $t \notin \{\hat{t}, \emptyset\} \Rightarrow a(\hat{t}, t, s) = 0$
As the mechanism is direct, in equilibrium, the agent will not lie; therefore, decreasing $a(\hat{t}, t, s)$ for $t \notin \{\hat{t}, \emptyset\}$ will not affect the expected utility of the mechanism designer. This deviation only increases the incentives to report truthfully. Therefore in can be assumed WLOG that the optimal mechanism features maximal punishment.

(2) Maximal reward: $e(\hat{t}, s) > 0 \Rightarrow a(\hat{t}, \hat{t}, s) = 1$.
Suppose $a(\hat{t}, \hat{t}, s) < 1$. One could now lower the probability of verification $de(\hat{t}, s) < 0$ while increasing the probability of allocation after confirming the report as true $da(s, \hat{t}, \hat{t}) > 0$ such that $d(e(\hat{t}, s)a(\hat{t}, \hat{t}, s)) = 0$.
Lowering the verification probability would only increase the incentives to misreport and the overall allocation probability after report $\hat{t}$ and signal s, if there was allocation with positive probability conditional on no verification, i.e. $a(s, \hat{t}, \emptyset) > 0$. However, in this case, this allocation could be lowered $da(s, \hat{t}, \emptyset) < 0$ such that $d((1 - e(\hat{t}, s))a(s, \hat{t}, \emptyset)) = 0$, and the incentives to misreport and the overall allocation probability would remain constant. As these procedure would save verification costs while keeping all unconditional allocation probabilities constant, the fact that an optimal mechanism features non-maximal reward can be ruled out.

These observations fix the allocation after verification. Effectively the mechanism designer therefore has to choose only the verification probability $e(t, s)$ and the allocation probability, conditional on no verification $a(s, t, \emptyset)$.

For convenience, define $z_{t,s} = e(t,s)$, the joint probability of verification and allocation, and $x(t,s) = (1-e(t,s))a(t,\emptyset,s)$, the joint probability of no verification and allocation.
Note that the set of mechanisms described by

$$\{(x_{t,s}, z_{t,s})_{t\in\mathcal{T},s\in\mathcal{S}} | \forall t \in \mathcal{T} \ \forall s \in \mathcal{S} : \ 0 \leq x_{t,s} + z_{t,s} \leq 1\}$$

is equivalent to all maximal reward and punishment, two-stage, direct mechanisms.[18]

Appendix 3.B Proofs not included in the paper

3.B.1 Proof of proposition 3.7

Step 0: (Relaxation) Define the following cut-off in the signal space:

$$\bar{s} = \min\{s | \mathbb{E}[v(T,s)|S=s] > \mathbb{E}[(v(T,s)-c)^+]\},$$

where $(a)^+ = \max\{0,a\}$ and we use the convention that $\min\emptyset = \max\mathcal{S}$. Note that $v(t,s) - (v(t,s)-c)^+ = \min\{c, v(t,s)\}$ is increasing in both components. Due to the MLRP it follows that $E[v(T,s)|S=s] - \mathbb{E}[(v(T,s)-c)^+$ is increasing in s.

Next, define the set of profitable types $\mathcal{T}^+ = \{t \in \mathcal{T} | v(t,\bar{s}) \geq 0\}$. We denote all types that are not profitable by $\mathcal{T}^- = \mathcal{T} - \mathcal{T}^+$.

The relaxed problem ignores certain incentive constraints. It optimizes the same objective function but only subject to:

$$\forall \hat{t} \in \mathcal{T}^+ \ \forall t \in \mathcal{T} \text{ with } t < \hat{t}\text{: } (IC_{t\hat{t}})$$

Step 1: There is an optimal solution to the relaxed problem that takes a cut-off form in x for all $t \in \mathcal{T}$:

$$\forall t \in \mathcal{T} \ \exists \, \tilde{s}(t) \in \mathcal{S} : \ x_{t,s} = \begin{cases} 0 & \text{if } s < \tilde{s}(t) \\ x_{t,s} \in [0,1) & \text{if } s = \tilde{s}(t) \\ 1 & \text{if } s > \tilde{s}(t) \end{cases}.$$

Proof. Suppose there is a (relaxed) incentive compatible mechanism which has for some $t \in \mathcal{T}$ and $s' < s'' \in \mathcal{S}$, that $x_{t,s} > 0$ and $x_{t,s'} < 1$.

18. The inverse mapping is given by

$$e(t,s), a(t,\emptyset,s)) = \left(z_{t,s}, \frac{x_{t,s}}{1-z_{t,s}}\right).$$

Note that the value of $a(t,\emptyset,s)$ does not play any role in the mechanism if $e(t,s) = z_{t,s} = 1$ and can therefore be chosen arbitrarily.

In the following, we consider a shift in allocation probability from $x_{t,s'}$ to $x_{t,s''}$ that keeps the overall allocation probability for type t constant:

$$f_{t,s'}\, dx_{t,s'} + f_{t,s''}\, dx_{t,s''} = 0 \Leftrightarrow \underbrace{dx_{t,s'}}_{<0} = -\frac{f_{t,s''}}{f_{t,s'}} \underbrace{dx_{t,s''}}_{>0}.$$

First, for any type $t^- < t$ the probability of receiving the good without verification after a misreport t decreases in the new mechanism:

$$f_{\tilde{t},s'} dx_{t,s'} + f_{\tilde{t},s''} dx_{t,s''} = -f_{\tilde{t},s'} \left[\frac{f_{t,s''}}{f_{t,s'}} - \frac{f_{\tilde{t},s''}}{f_{\tilde{t},s'}}\right] dx_{t,s''} \leq 0.$$

The last inequality holds since the likelihood ratio is increasing. The shift yields type t the same allocation probability, so he cannot have a new incentive to misreport. Therefore, all relaxed incentive constraints survive.

Second, the modified mechanism yields the principal a higher expected value:

$$f_{t,s'}\, dx_{t,s'}\, v(t,s') + f_{t,s''}\, dx_{t,s''}\, v(t,s'') = f_{t,s''} dx_{t,s''} \left[-v(t,s') + v(t,s'')\right] > 0$$

The proposed shift is clearly feasible if in the original mechanism $x_{t,s''} + z_{t,s''} < 1$. In the case that $x_{t,s''} + z_{t,s''} = 1$, it can still be implemented by shifting in addition mass from $z_{t,s''}$ to $z_{t,s'}$ such that $x_{t,s''} + z_{t,s''}$ and $x_{t,s'} + z_{t,s'}$ stay constant:

$$dx_{t,s'} = -dz_{t,s'} \text{ and } dx_{t,s''} = -dz_{t,s''}.$$

As we assume $x_{t,s''} < 1$ and $x_{t,s''} + z_{t,s''} = 1$, we have that $z_{t,s'} > 0$. Since $x_{t,s'} > 0$ we also have $z_{t,s'} < 1$.

The incentives for any lower type to misreport his type as t are weakened in the same way as above since $z_{t,s}$ and $z_{t,s'}$ do not play a role in these constraints. The incentive for t to misreport is not affected since the total allocation probability $x + z$ is kept constant.

Further, the principal's expected value is not changed by this shifts:

$$\begin{aligned} f_{t,s'} \left[dx_{t,s'}\ \ v(t,s') + dz_{t,s'}\ \ (v(t,s') - c)\right] + f_{t,s''} \left[dx_{t,s''}\ \ v(t,s'') + dz_{t,s''}\ \ (v(t,s'') - c)\right] \\ = -c\ f_{t,s''} \left[\frac{f_{t,s''}}{f_{t,s'}} - \frac{f_{t,s''}}{f_{t,s'}}\right] (-dz_{t,s'}) \ = \ 0. \end{aligned}$$

The reason is, that the allocation $(x + z)$ remains the same with these shifts, so that only the verification cost changes. Yet, the change in verification is such that it does change the expected verification probability for the true type t and therefore neither the expected verification cost for the principal. □

Step 2: The optimal mechanism in the relaxed problem features,

$$\forall t \in \mathcal{T}^- \ \forall s \in \mathcal{S} : \ z_{t,s} = 0.$$

Proof. In the relaxed problem we disregard all incentive constraints that prevent the agent to misreport his type as $t \in \mathscr{T}^-$. If there were some $t \in \mathscr{T}^-$ and $s \in \mathscr{S}$ with $z_{t,s} > 0$, shifting probability mass from $z_{t,s}$ to $x_{t,s}$ by

$$\underbrace{dz_{t,s}}_{<0} = -\underbrace{dx_{t,s}}_{>0},$$

would save the principal verification costs while keeping the overall allocation probability constant. It would therefore not affect the incentive constraints in the relaxed problem. □

Step 3: We can assume that the optimal mechanism also takes a cut-off form in $x + z$:

$$\forall t \in \mathscr{T}\ \exists \underline{s}(t) \in \mathscr{S} : x_{t,s} + z_{t,s} = \begin{cases} 0 & \text{if } s < \underline{s}(t) \\ x_{t,s} + z_{t,s} \in [0,1) & \text{if } s = \underline{s}(t) \\ 1 & \text{if } s > \underline{s}(t) \end{cases}.$$

Proof. For $t \in \mathscr{T}^-$, this property follows immediately from the previous two steps with $\underline{s}(t) = \tilde{s}(t)$.

Suppose for $t \in \mathscr{T}^+$ that there exist $s' < s'' \in \mathscr{S}$ with $z_{t,s'} > 0$ and $x_{t,s''} + z_{t,s''} < 1$. To rule out this possibility, consider a shift in mass from $z_{t,s'}$ to $z_{t,s''}$ in a way that the allocation probability for a truth-telling agent of type t remains constant, i.e.,

$$\underbrace{dz_{t,s''}}_{>0} = \frac{f_{t,s'}}{f_{t,s''}}(-\underbrace{dz_{t,s'}}_{<0}).$$

Note that this shift is feasible by assumption and that it will keep all relaxed incentive constraints unchanged, since the true type t receives the same expected allocation probability, and $\mathbf{z}_{t,\cdot}$ does not play a role in the IC constraints preventing misreport t.

From the principal's point of view, it is favorable because it keeps the verification probability and thus the costs constant, while shifting allocation mass from (t,s') to the more favorable type–signal pair (t,s''), i.e.

$$dV = f_{t,s'}dz_{t,s'}[v(t,s') - c] + f_{t,s''}dz_{t,s''}[v(t,s'') - c] = 0 \cdot c + f_{t,s'}[v(t,s'') - v(t,s')](-dz_{t,s'}) > 0.$$

□

Step 4: The following restriction of the relaxed problem is without loss for the principal. Require, additionally, that the IC constraints hold point-wise. That is $\forall t \in \mathscr{T}$, $\forall \hat{t} \in \mathscr{T}^+$ with $t < \hat{t}$ and $\forall s \in \mathscr{S}$:

$$(EPIC(s)_{t,\hat{t}}) = x_{t,s} + z_{t,s} - x_{\hat{t},s} \geq 0.$$

An optimal solution to the restricted relaxed problem therefore also solves the relaxed problem.

Proof. By the above steps, the (Bayesian) IC constraints in the relaxed problem can be written as follows:[19]

$$\forall t \in \mathscr{T} \;\; \forall \hat{t} \in \mathscr{T}^+ \text{ with } \hat{t} > t :$$

$$(IC_{t,\hat{t}}) = \sum_{s \in \mathscr{S}} f_{t,s}(x_{t,s} + z_{t,s}) - \sum_{s \in \mathscr{S}} f_{t,s} x_{\hat{t},s}$$

$$= f_{t,\underline{s}(t)}(x_{t,\underline{s}(t)} + z_{t,\underline{s}(t)}) + \sum_{s > \underline{s}(t)} f_{t,s} \;\; 1 - (f_{t,\tilde{s}(\hat{t})} \;\; x_{\hat{t},\tilde{s}(\hat{t})} + \sum_{s > \tilde{s}(\hat{t})} f_{t,s} \;\; 1) \overset{!}{\geq} 0.$$

This condition clearly requires that $\underline{s}(t) \leq \tilde{s}(\hat{t})$ and, in the case of equality, $x_{t,\underline{s}(t)} + z_{t,\underline{s}(t)} \geq x_{\hat{t},\tilde{s}(t)}$. Because by the definition of $\underline{s}(t)$, $x + z$ is equal to 0 below and equal to 1 above this threshold, we can conclude that for all s, $x_{t,s} + z_{t,s} \geq x_{\hat{t},s}$, which implies $(EPIC(s)_{t,\hat{t}})$. □

Step 5: Consider an optimal mechanism of the above cut-off structure that satisfies $(EPIC(s)_{t,\hat{t}}) \geq 0$ for all $t \in \mathscr{T}$, for all $\hat{t} \in \mathscr{T}^+$ with $\hat{t} > t$, and for all $s \in \mathscr{S}$. Then, it also must hold for $t, \hat{t} \in \mathscr{T}^-$ with $t < \hat{t}$, that $(EPIC(s)_{t,\hat{t}}) \geq 0$. This means that no unprofitable type has an incentive to report a higher unprofitable type.

Proof. Assume that for some $s \in \mathscr{S}$ there are types $t < \hat{t} \in \mathscr{T}^-$ such that the constraint $(EPIC(s)_{t,\hat{t}})$ is violated.

Define $s' \equiv \min\{s \in \mathscr{S} | \exists t < \hat{t} \in \mathscr{T}^- : x_{t,s} + z_{t,s} < x_{\hat{t},s}\}$ to be the lowest signal for which some type t profits from a higher report $\hat{t} \in \mathscr{T}^-$. Let $t' \equiv \min\{t \in \mathscr{T}^- | \exists \hat{t} \in \mathscr{T}$ with $\hat{t} > t : x_{t',s'} < x_{\hat{t},s'}\}$ be the smallest type with $EPIC(s')$ incentives to misreport his type to some type $\hat{t} \in \mathscr{T}^-$.

Since $z_{t',s'} = 0$ for the unprofitable type t' (Step 3), this implies $x_{t,s'} < x_{\hat{t},s'}$. As $t', \hat{t} \in \mathscr{T}^-$, it follows that $\hat{t}$'s EPIC(s') constraints are slack for all reports in $\mathscr{T}^+$. Having $x_{\hat{t},s'} > x_{t',s'} \geq 0$ can therefore only be optimal in the relaxed problem if $v(\hat{t}, s') \geq 0$. This implies that $s' > \bar{s}$ since $\mathscr{T}^-$ is precisely defined as the set of types t with $v(t, \bar{s}) < 0$.

Since $x_{t',s'} < x_{\hat{t},s'} \leq 1$, taking the cut-off structure from step 1 into account we can infer that for all $s < s'$ it holds that $x_{t',s'} = 0$. In particular we have $x_{t',\bar{s}} = 0$.

By the minimality of s' we get that $0 = x_{t',\bar{s}} \geq x_{\tilde{t},\bar{s}}$ for all $\tilde{t} \in \mathscr{T}^-$ with $\tilde{t} > t'$. By the minimality of t' we get that $x_{\tilde{t},\bar{s}} \leq x_{t',\bar{s}} = 0$ for all $\tilde{t} \in \mathscr{T}^-$ with $\tilde{t} < t'$. Furthermore, by $EPIC(\bar{s})$ IC compatability, it follows that $0 = x_{t,\bar{s}} \geq x_{t',\bar{s}}$ for all $t' \in \mathscr{T}^+$. So we have that $x_{t,\bar{s}} = 0$ for all t which by definition of $\bar{s}$ cannot be optimal in the restricted problem (for which this mechanism is also optimal) as $\mathbb{E}_T[\, (v(T, \bar{s}) \mid S = \bar{s}\,] > \mathbb{E}_T[\, (v(T, \bar{s}) - c\,)^+ \mid S = \bar{s}\,] \geq 0$. □

Step 6: Consider an optimal mechanism of the above cut-off structure that satisfies $(EPIC(s)_{t,\hat{t}}) \geq 0$ for all $t \in \mathscr{T}$, for all $\hat{t} \in \mathscr{T}^+$ with $\hat{t} > t$, and for all $s \in \mathscr{S}$. Then,

19. Making use of the fact that $z_{t,s} = 0$ for all $t \in \mathscr{T}^-$.

for $t, \hat{t} \in \mathcal{T}$ with $t > \hat{t}$ it mus also satisfy $(EPIC(s)_{t,\hat{t}}) \geq 0$. This means that no type benefits from reporting any lower type.

Proof. Assume that there are types $t'' > t' \in \mathcal{T}$ such that $x_{t'',s} + z_{t'',s} < x_{t',s}$ for some s. WLOG let t'' be the lowest type for which such a downward deviation is profitable.

Optimality of the relaxed mechanism requires then that $\mathbb{E}_T[\, v(T,s)\mathbb{1}_{\{T \leq t'\}} \mid S = s\,] > 0$. Otherwise the principal would be better off by lowering x for all types below t'' (Note that $x_{t'',s} + z_{t'',s} < x_{t',s}$ implies that types $t < t''$ cannot have binding upwards constraints towards reports higher than t'' as this would violate the upward constraints for t''). Monotonicity of the value in the type in turn implies that $v(t,s) > 0$ for all $t > t'$. This contradicts optimality as the designer could increase $x_{t,s}$ for all higher types without violating any incentives. Either by just increasing $x_{t,s}$ if $x_{t,s} + z_{t,s} < 1$ or by lowering $z_{t,s}$ at the same time to save verification costs. □

Step 7: This concludes the proof. We have shown that optimal solution to the relaxed problem is EPIC (Step 4). Therefore the principal's expected value in the original problem cannot exceed the expected value from this optimal EPIC solution. In step 5 and 6 we ruled out the two possible violations of the original (ex-post) incentive constraints that can arise in a solution to the relaxed problem. Hence the candidate solution is also EPIC in the original problem. In particular, it is Bayesian incentive compatible and therefore a solution to the original problem.

3.B.2 Proof of Proposition 3.8

Proof. To prove the claim, we construct an improvement that will not violate the Bayesian incentive constraints. This suffices to show that the principal strictly profits from ensuring that the realization of S remains private because the improved mechanism will implement the same allocation at lower verification costs. Consider the shift of mass from $z_{\hat{t},s'}$ to $z_{\hat{t},s}$ and —in order to maintain the overall allocation $\boldsymbol{x} + \boldsymbol{z}$ unchanged— vice versa for $x_{\hat{t},s'}$ and $x_{\hat{t},s}$:

$$dx_{\hat{t},s'} + dz_{\hat{t},s'} \;=\; 0 \quad \text{and} \quad dx_{\hat{t},s} + dz_{\hat{t},s} \;=\; 0.$$

To ensure that the Bayesian incentive constraints of all types $t < \hat{t}$ are not violated by the shift, we require that

$$\forall t < \hat{t}:\; d(IC_{t,\hat{t}}) = -f_{t,s}\, dx_{\hat{t},s} - f_{t,s'}\, dx_{\hat{t},s'} \geq 0 \Leftrightarrow dx_{\hat{t},s'} \leq \frac{f_{t,s}}{f_{t,s'}} \left(-dx_{\hat{t},s}\right).$$

The proposed change has $(-dx_{\hat{t},s}) > 0$, and $\frac{f_{t,s}}{f_{t,s'}}$ is decreasing in t. Hence, the right-hand side of the above expression is minimized at $t' = \max\{t \in \mathcal{T} | t < \hat{t}\}$. Note that $\hat{t} \neq \min\{t \in \mathcal{T}\}$, as this would imply that at $v(t,s') > c > 0$ at all levels of t so that the optimal mechanism would allocate without verification after this signal.

Setting $dx_{\hat{t},s'} = \frac{f_{t',s}}{f_{t',s'}} \left(-dx_{\hat{t},s}\right)$ ensures that the incentives to misreport toward $\hat{t}$ are weakened for all lower types.

The above changes in x imply for z:

$$dx_{\hat{t},s} = \frac{f_{t',s'}}{f_{t',s}}(-dx_{\hat{t},s'}) \Leftrightarrow (-dz_{\hat{t},s}) = \frac{f_{t',s'}}{f_{t',s}} dz_{\hat{t},s'}.$$

The principal's value changes by:

$$\begin{aligned} dV &= f_{\hat{t},s}\left[dx_{\hat{t},s}\ \ v(\hat{t},s) + dz_{\hat{t},s}\ \ (v(\hat{t},s)-c)\right] + f_{\hat{t},s'}\left[dx_{\hat{t},s'}\ \ v(\hat{t},s') + dz_{\hat{t},s'}\ \ (v(\hat{t},s')-c)\right] \\ &= -c\left[f_{\hat{t},s}\ \ dz_{\hat{t},s} + f_{\hat{t},s'}\ \ dz_{\hat{t},s'}\right] \\ &= -c\ f_{\hat{t},s}\left[\frac{f_{t',s'}}{f_{t',s}} - \frac{f_{\hat{t},s'}}{f_{\hat{t},s}}\right](-dz_{\hat{t},s'})\ >\ 0. \end{aligned}$$

The second equality follows because the allocation remains the same with these shifts, so that only the verification cost changes.

Lastly, note that in the optimal EPIC mechanism, $z_{\hat{t},s} = 1$ implies $z_{t,s} = 1$ for all $t > \hat{t}$ and that $\boldsymbol{x}_{\cdot,s}$ is constant in the report at all s. Therefore, the fact that $z_{\hat{t},s} = 1$ in the original mechanism implies that the Bayesian IC constraints for higher types to lie downward to $\hat{t}$ are slack so that we can always find a shift in magnitude small enough to not violate these constraints.

The only case in which these constraints are not slack in the optimal EPIC mechanism is when several types receive exactly the same allocation. In this case, the above improvement can be applied to the highest report in this class. □

3.B.3 Proof of Lemma 3.5

Proof. Similar to Step 0 in Lemma 3.2, for any $s \in \mathcal{S}$, the optimal $(\boldsymbol{x}_{\cdot,s}, \boldsymbol{z}_{\cdot,s})$ can be determined separately, as all constraints only involve allocation and verification probabilities for the same signal realization.

This results in $|\mathcal{S}|$ separate problems, one for each possible signal realization $s \in \mathcal{S}$. In these separate problems $v(t,s)$ is a function of t only. Therefore, for all $s \in \mathcal{S}$, all steps in the proof of Lemma 3.2 can be replicated with $v(T)$ replaced by $v(T,s)$.

□

References

Akbarpour, Mohammad, and Shengwu Li. 2019. "Credible Auctions: A Trilemma." *Econometrica*, [109]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L Lipman. 2017. "Mechanisms with Evidence: Commitment and Robustness." *Working Paper*, [103, 124]

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman. 2014. "Optimal Allocation with Costly Verification." *American Economic Review* 104 (12): 3779–813. [103, 106, 109, 110, 124–126]

Bhargava, Mohit, Dipjyoti Majumdar, and Arunava Sen. 2015. "Incentive-compatible voting rules with positively correlated beliefs." *Theoretical Economics* 10 (3): 867–85. [103, 123]

Cella, Michela. 2008. "Informed principal with correlation." *Games and Economic Behavior* 64 (2): 433–56. [104, 124]

Crémer, Jaques, and Richard P. McLean. 1988. "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions." *Econometrica* 56 (6): 1247–57. [103, 106, 114, 123]

Demougin, Dominique M, and Devon A Garvie. 1991. "Contractual design with correlated information under limited liability." *RAND Journal of Economics*, 477–89. [103, 123]

Erlanson, Albin, and Andreas Kleiner. 2019. "Costly Verification in Collective Decisions." [103, 104, 108, 110, 124]

Gale, Douglas, and Martin Hellwig. 1985. "Incentive-Compatible Debt Contracts: The One-Period Problem." *Review of Economic Studies* 52 (4): 647–63. [103, 124]

Gershkov, Alex, Jacob K. Goeree, Alexey Kushnir, Benny Moldovanu, and Xianwen Shi. 2013. "On the Equivalence of Bayesian and Dominant Strategy Implementation." *Econometrica* 81 (1): 197–220. [104, 124]

Gibbard, Allan. 1973. "Manipulation of voting schemes: a general result." *Econometrica* 41 (4): 587–601. [103, 123]

Glazer, Jacob, and Ariel Rubinstein. 2004. "On Optimal Rules of Persuasion." *Econometrica* 72 (6): 1715–36. [103, 123, 124]

Halac, Marina, and Pierre Yared. 2019. "Commitment vs. Flexibility with Costly Verification." *Working Paper*, [103, 104, 124]

Hart, Sergiu, Ilan Kremer, and Motty Perry. 2017. "Evidence games: truth and commitment." *American Economic Review* 107 (3): 690–713. [103, 124]

Johnson, Scott, John W Pratt, and Richard J Zeckhauser. 1990. "Efficiency despite mutually payoff-relevant private information: The finite case." *Econometrica*, 873–900. [103, 123]

Kittler, Wolf. 2003. "Heimlichkeit und Schriftlichkeit: Das österreichische Strafprozessrecht in Franz Kafkas Roman Der Proceş." *Germanic Review: Literature, Culture, Theory* 78 (3): 194–222. [107, 119]

Manelli, Alejandro M., and Daniel R. Vincent. 2010. "Bayesian and Dominant-Strategy Implementation in the Independent Private-Values Model." *Econometrica* 78 (6): 1905–38. [104, 124]

Maskin, Eric, and Jean Tirole. 1990. "The principal-agent relationship with an informed principal: The case of private values." *Econometrica: Journal of the Econometric Society*, 379–409. [104, 124]

McAfee, R Preston, and Philip J Reny. 1992. "Correlated information and mecanism design." *Econometrica*, 395–421. [103, 123]

Myerson, Roger B. 1983. "Mechanism design by an informed principal." *Econometrica: Journal of the Econometric Society*, 1767–97. [104, 124]

Mylovanov, Tymofiy, and Andriy Zapechelnyuk. 2017. "Optimal Allocation with Ex Post Verification and Limited Penalties." *American Economic Review* 107 (9): 2666–94. [103, 124]

Neeman, Zvika. 2004. "The relevance of private information in mechanism design." *Journal of Economic Theory* 117 (1): 55–77. [103, 123]

Riordan, Michael H, and David EM Sappington. 1988. "Optimal contracts with public ex post information." *Journal of Economic Theory* 45 (1): 189–99. [103, 115, 123]

Satterthwaite, Mark Allen. 1975. "Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions." *Journal of economic theory* 10 (2): 187–217. [103, 123]

Severinov, Sergei. 2008. "An efficient solution to the informed principal problem." *Journal of Economic Theory* 141 (1): 114–33. [104, 124]

Townsend, Robert M. 1979. "Optimal contracts and competitive markets with costly state verification." *Journal of Economic Theory* 21 (2): 265–93. [103, 124]