# The (No) Value of Commitment

Nathan Hancart*

November, 2022

PRELIMINARY

**Abstract**

I provide a sufficient condition under which a principal does not benefit from commitment in economic situations. I focus on situations described by a constrained maximisation problem. I show that commitment has no value when the *marginal* contribution of the constraints is null in the problem with commitment. This condition also has bite when constraints are binding. I then apply this condition in a mechanism design setting. I show that a designer does not benefit from being able to contract over actions when his preferences are partially aligned with the agent's. Verifying the condition does not necessitate verifying explicitly that the strategy under commitment is a best-response to the information revealed in the economic problem.

*Nathan Hancart: Department of Economics, University College London, 30 Gordon St, London WC1H 0AN, UK, nathan.hancart.16@ucl.ac.uk.

# 1 Introduction

Commitment plays an important role in many economic models. The general insight of economic theory is that the value of commitment is positive: if a principal has commitment, he can replicate any action he would play without commitment. Moreover, commitment plays a key role in many standard tools used in economic theory such as the revelation principle (Myerson, 1982; Bester and Strausz, 2000; Doval and Skreta, 2022). However, commitment is usually a strong assumption and is sometimes hard to justify. Even when it is possible to justify the commitment assumption, it might be an undesirable feature of the model. For example, even if a regulator could commit to a rule, there might be reasons outside the model that require the government to maintain agency over this rule at any point in time.

In this paper, I provide a condition under which commitment has no value for a principal that faces a maximisation problem under constraints. That is I provide a condition under which, even when the principal can commit, he is better off best-replying to the information revealed in the economic problem. The usefulness of this result is twofold. First, as argued above, commitment can be an undesirable feature of economic models. Knowing that the condition provided is satisfied facilitates solving the model. Indeed, models assuming commitment are usually easier to solve as the number of constraints in the problem is smaller. When assuming commitment, the modeller does not need to make sure that the principal best-replies at the optimum. But if the condition holds, we are guaranteed that the omitted best-replies constraints of the principal will hold. Second, in the case commitment is actually assumed, it restricts the set of strategies the modeller has to look at. Even though assuming commitment can simplify the problem, the set of solutions the modeller needs to consider remain quite large. Knowing that the the value of commitment is zero restricts the set of potential solutions.

Consider the following maximisation problem. Let $\alpha$ describe the strategy of the principal and $\sigma$ the strategy of other agents in the economic problem.

$$V = \max_{\alpha,\sigma} v(\alpha,\sigma)$$
$$\text{s.t. } \text{Constraint}(\alpha,\sigma)$$
$$\alpha \text{ is a best-response to } \sigma$$

where $v$ denotes the payoffs of the principal. The problem above can represent many economic models. For example, the constraints can be incentive compatibility constraints of some agents in a mechanism design problem. Without commitment, there is an additional constraint guaranteeing that the principal's strategy is a best-reply to the information revealed in the economic interaction. With commitment, the principal can commit to information in an arbitrary way. My approach is to fix the principal's strategy and treat it as a parameter in the maximisation problem.

$$V(\alpha) = \max_{\sigma} v(\alpha,\sigma)$$
$$\text{s.t. Constraint}(\alpha,\sigma)$$

The problem $V(\alpha)$ is the principal's problem when he commits to the strategy $\alpha$. One can solve the problem above by finding a saddle-point of a Lagrangian:

$$\mathcal{L}(\sigma,\lambda;\alpha) = v(\alpha,\sigma) + \lambda \cdot \text{Constraint}(\alpha,\sigma),$$

where $\lambda$ is the Lagrangian multiplier associated with the constraints. For a solution $(\sigma^*,\lambda^*)$, we can apply an envelope theorem (Milgrom and Segal, 2002) on the Lagrangian to get that, omitting technical details,

$$\frac{dV(\alpha)}{d\alpha} = \frac{\partial\mathcal{L}(\sigma^*,\lambda^*;\alpha)}{\partial\alpha} = \frac{\partial v(\alpha,\sigma^*)}{\partial\alpha} + \lambda^* \cdot \frac{\partial\,\text{Constraint}(\alpha,\sigma^*)}{\partial\alpha}$$

Now note that if the last term $\lambda^* \cdot \frac{\partial \text{Constraint}(\alpha,\sigma^*)}{\partial \alpha} = 0$, then the total derivative of the value function is equal to its partial derivative:

$$\frac{dV(\alpha)}{d\alpha} = \frac{\partial v(\alpha, \sigma^*)}{\partial \alpha}.$$

This is exactly the condition needed to show that commitment does not have any value. Indeed, when the first-order condition is satisfied in the commitment problem, i.e., when taking into account the change it is going to induce in the constraint, it is also satisfied when the principal does not have commitment, i.e., does not take into account the change it induces in the constraints.

Note that if the constraints are slack at the optimum, this condition holds. But it can also hold when they are not. The key condition is not whether the constraints matter, i.e., they are slack, but whether the *marginal* contribution of the constraints to the Lagrangian is null. It is worth noting that to check whether the condition is satisfied, it is not needed to check explicitly whether the principal best-replies to the information revealed in the economic problem.

I apply this result to a mechanism design setting à la Myerson (1982) where some of the actions are non-contractible. In Proposition 1, I show that if the principal's preferences are partially aligned with the agent, he does not benefit from being able to contract over these actions. The proof of this result uses both a characterisation of the optimal mechanism and the condition for the value of commitment. The characterisation is not complete enough to conclude that commitment has no value by checking best-response conditions, but enough to check that the condition on the Lagrangian is satisfied. This latter result is similar to a result from Ben-Porath et al. (2021) that shows a non-commitment result in the context of mechanism design with evidence with partially aligned preferences.

More generally, the method presented here can be used to prove several results in the mechanism design with evidence literature that shows that the principal does not benefit from commitment. For example, I can use the Lagrangian condition to extend Glazer and Rubinstein (2004)'s result on the value of commitment. Vohra et al. (2021) also use the envelope theorem to show conditions under which the principal does not benefit from commitment. Their key assumption is to only allow environments where the constraints do not play a role when changing the uncommitted actions marginally. Instead, the condition presented here explicitly addresses the marginal effect of the uncommitted action on the constraints. Moreover, if their condition holds, so does mine. In Section 3, I show an example where commitment has no value and their condition does not hold.

# 2 General setup

A principal must solve the economic problem described as follows. For $m = 1, ..., M$, let $Y_m$ be a finite set and $A = \times_m \Delta(Y_m)$ with typical element $\alpha$. Let $S$ a subset of a convex compact subset of $\mathbb{R}^n$ with typical element $\sigma$. Let $v : A \times S \to \mathbb{R}$, $g : A \times S \to \mathbb{R}^k$ and $BR(\sigma) = \{\alpha : v(\alpha, \sigma) \geq v(\alpha', \sigma),\ \forall \alpha'\}$. The assumption that $A$ is the product set of simplexes allows for product set of intervals with the right normalisation by taking $Y_m$ to be binary.

Consider the following maximisation problem:

$$(\mathcal{V}) \qquad V = \max_{\alpha,\sigma} v(\alpha, \sigma)$$
$$\text{s.t. } g(\alpha, \sigma) \geq 0$$
$$\alpha \in BR(\sigma)$$

The function $v$ denotes the principal's payoff, the function $g$ describes a set of constraint he is facing and $BR(\sigma)$ describes the set of element of $A$ that are a best reply to $\sigma$.

If the principal could commit to $\alpha$, he would solve the following problem:

$$(\overline{\mathcal{V}}) \qquad \overline{V} = \max_{\alpha,\sigma} v(\alpha,\sigma) \quad \text{s.t. } g(\alpha,\sigma) \geq 0$$

The aim of this paper is to find condition under which $V = \overline{V}$.

To do so, I first introduce the following maximisation problem where the principal commits over some $\alpha$:

$$(\overline{\mathcal{V}}(\alpha)) \qquad \overline{V}(\alpha) = \max_{\sigma} v(\alpha,\sigma) \quad \text{s.t. } g(\alpha,\sigma) \geq 0$$

and the associated Lagrangian,

$$\mathcal{L}(\sigma,\lambda;\alpha) = v(\alpha,\sigma) + \lambda \cdot g(\alpha,\sigma)$$

We can now state our main theorem. Say that the *first-order conditions are sufficient for* $v$ if for each $\sigma \in S$, $\alpha^* \in \arg\max_\alpha v(\alpha,\sigma)$ implies that for all $m$, $y \in \operatorname{supp}\alpha^*(\cdot|m)$ only if $\frac{\partial v(\alpha^*,\sigma)}{\partial\alpha(y|m)} \geq \frac{\partial v(\alpha^*,\sigma)}{\partial\alpha(y'|m)}$ for all $y'$.

**Theorem 1.** *Suppose that each element of* $\nabla_\alpha v(\alpha,\sigma)$ *and* $\nabla_\alpha g(\alpha,\sigma)$ *is continuous in* $(\alpha,\sigma)$*, that first-order conditions are sufficient for* $v$ *and that the solution of* $\overline{\mathcal{V}}(\alpha)$ *can be obtained by finding a saddle-point of* $\mathcal{L}(\cdot,\cdot;\alpha)$ *for all* $\alpha$*.*

*Then* $V = \overline{V}$ *if there is* $\alpha^* \in \arg\max_\alpha \overline{V}(\alpha)$ *and saddle-point of* $\mathcal{L}(\cdot, \cdot; \alpha^*)$, $(\sigma, \lambda)$, *such that*

$$\lambda \cdot \nabla_\alpha g(\alpha^*, \sigma) = 0 \tag{1}$$

All proofs are relagated to the appendix.

This result tells us that commitment has no value if the *marginal* contribution of the constraints is zero. Note that if all the constraints were slack at the optimum, then $\lambda = 0$ and the condition is satisfied. This is what we would expect: if the principal is not effectively facing any constraint, he is better off best-replying to the information revealed. Theorem 1 tells us that what really matters is not that constraints do not matter but that their marginal contribution is null.

The condition that first-order conditions are sufficient is satisfied whenever $v(\alpha, \sigma)$ is linear in $\alpha$. This would be the case if $v(\alpha, \sigma)$ is the expected utility over some finite action set $\times_m Y_m$.

# 3 Application to mechanism design – Myerson (1982)

Following Myerson (1982), I consider a set-up where some actions are contractible and some are not. A mechanism can commit to mapping from input messages to distribution over output messages and contractibel actions. Formally, there is a principal and an agent. The principal has access to a set of action $X \times Y$ where both $X, Y$ are finite. The actions in $X$ are contractible whereas the action in $Y$ are not. There is also a finite set of messages $M$. The agent has private information $\theta \in \Theta$, $\Theta$ finite. The prior distribution over types is $\mu \in \Delta\Theta$. The principal and the agent have preferences $v : X \times Y \times \Theta \to \mathbb{R}$ and $u : X \times Y \times \Theta \to \mathbb{R}$.

A direct mechanism is a function $\sigma : \Theta \to \Delta(X \times M)$. A strategy for the principal is $\alpha : M \to \Delta Y$. These objects correspond to $\sigma$ and $\alpha$ defined in the previous section. The DM's payoff as a function of the mechanism and his actions is, abusing notation, $v(\alpha, \sigma) = \sum_\theta \sum_{x,y,m} \mu(\theta)\sigma(x, m|\theta)\alpha(y|m)v(x, y, \theta)$. Without loss of generality, we can take $M = Y$. Let $BR(\sigma)$ be the set of best-responses of the principal after observing the output messages given the mechanism $\sigma$.

**Example** (Regulation with externalities)**.** A government is contracting with a firm of unknown costs and externality $\Theta = C \times E \subset \mathbb{R}_+ \times \{-1, 1\}$ where $C$ is the cost parameter and $E$ is an externality parameter. The government can both decide on the scope of the project $x \subset \mathbb{R}$ and on whether to authorise it, $y \in \{0, 1\}$. Consider the following utility functions. For the firm $u(x, y, \theta) = y\big(x - c\frac{x^2}{2}\big)$ and the government has payoffs $v(x, y, \theta) = e \cdot y\big(x - c\frac{x^2}{2}\big)$. Intuitively the government cares positively about firms having a positive externality and vice-versa. After having decided the scope of the project, the government can always decide to shut down the project. A mechanism can be interpreted as an independent regulator that can decide on both the scope of the contract and authorisation if both are contractible or can only issue recommendations to the government regarding the authorisation. Would the government benefit from fully delegating the decision to the regulator? ■

In the following, I take a "partial revelation principle approach" in the sense that I assume that it is without loss to have the agent directly report his type (thus require classic incentive compatibility constraints) but I don't use the simplification of taking output messages as action recommendation (i.e., no obedience constraints).

If the principal cannot contract on $Y$, his problem is

$$\begin{aligned} V = &\max_{\alpha,\sigma} \sum_{\theta} \sum_{x,y,m} \mu(\theta)\sigma(x,m|\theta)\alpha(y|m)v(x,y,\theta) \\ \text{s.t.} \quad &\text{for all } \theta,\theta', \sum_{x,y,m} \big(\sigma(x,m|\theta) - \sigma(x,m|\theta')\big)\alpha(y|m)u(x,y,\theta) \geq 0 \\ &\alpha \in BR(\sigma) \end{aligned}$$

On the other hand, if the principal could contract on $Y$, his problem would be to solve

$$\begin{aligned} \overline{V} = &\max_{\sigma} \sum_{\theta} \sum_{x,y} \mu(\theta)\sigma(x,y|\theta)v(x,y,\theta) \\ \text{s.t.} \quad &\text{for all } \theta,\theta', \sum_{x,y} \big(\sigma(x,y|\theta) - \sigma(x,y|\theta')\big)u(x,y,\theta) \geq 0 \end{aligned}$$

When the principal commits to $\alpha$,

$$\begin{aligned} \overline{V}(\alpha) = &\max_{\sigma} \sum_{\theta} \sum_{x,y,m} \mu(\theta)\sigma(x,m|\theta)\alpha(y|m)v(x,y,\theta) \\ \text{s.t.} \quad &\text{for all } \theta,\theta', \sum_{x,y,m} \big(\sigma(x,m|\theta) - \sigma(x,m|\theta')\big)\alpha(y|m)u(x,y,\theta) \geq 0 \end{aligned}$$

I use $\lambda(\theta,\theta')$ to denote the Lagrange multiplier associated with the IC constraint $(\theta,\theta')$.

Satisfying condition (1) requires that there is $\alpha^* \in \arg\max_{\alpha} \overline{V}(\alpha)$ and a solution $(\sigma,\lambda)$ such that for all $m,y$,

$$\sum_{\theta,\theta'} \lambda(\theta,\theta') \sum_{x} \big(\sigma(x,m|\theta) - \sigma(x,m|\theta')\big)u(x,y,\theta) = 0.$$

Note that a solution to $\max_{\alpha} \overline{V}(\alpha)$ is always $\tilde{\alpha}(y|m) = 1$ iff $y = m$ as that would give the classic contracting problem. Thus solving this problem might give the sufficient condition for the theorem to hold. But it could be that it does not hold for $\tilde{\alpha}$ but for other $\alpha^* \in \arg\max \overline{V}(\alpha)$.

I apply Theorem 1 to show that when preferences are partially aligned, the designer does not benefit from being able to contract over actions in $Y$.

**Proposition 1.** *If $v(x, y, \theta) = \nu(\theta)u(x, y, \theta)$ for some $\nu : \Theta \to \mathbb{R}$, then the conditions of Theorem 1 holds.*

To prove Proposition 1, I partially solve for the optimal mechanism for each $\alpha$. I show how the optimal Lagrange multipliers and mechanism are related using a duality argument. I show the existence of an *auxiliary game* whose equilibrium determine the strategies and multipliers. In this game, the types aligned with the principal, $\nu(\theta) \geq 0$ choose a contractible action in $X$ and an ouput message in $M$. The types misaligned with principal choose an aligned type to mimic.

I define the auxiliary game in the following way. Let $\overline{\Theta} = \{\theta : \nu(\theta) \geq 0\}$ and $\underline{\Theta} = \{\theta : \nu(\theta) < 0\}$. The players are types in $\Theta$. The action space of $\theta \in \overline{\Theta}$ is $X \times M$, with strategy $\overline{s} : \overline{\Theta} \to \Delta(X \times M)$. The action space of $\theta \in \underline{\Theta}$ is $\overline{\Theta}$, with strategy $\underline{s} : \underline{\Theta} \to \Delta\overline{\Theta}$. The payoffs are

$$\text{for } \theta \in \overline{\Theta},\ \tilde{u}(x, m, \underline{s}|\theta) = \mu(\theta)\nu(\theta)u(x, m, \theta; \alpha) + \sum_{\underline{\theta} \in \underline{\Theta}} \mu(\underline{\theta})\nu(\underline{\theta})\underline{s}(\theta|\underline{\theta})u(x, m, \underline{\theta}; \alpha)$$

$$\text{for } \theta \in \underline{\Theta},\ \tilde{u}(\overline{\theta}, \overline{s}|\theta) = \sum_{x,m} \overline{s}(x, m|\overline{\theta})u(x, m, \theta; \alpha)$$

Note that the characterisation I provide is sufficient to show that condition (1) holds but not sufficient to show that the optimal mechanism induces a best-reply. Proposition 1 can be used to answer the question in the regulator example.

**Example** (Regulation with externalities - Continued)**.** Consider the following parametrisation of the regulator example: $\Theta = \{\theta_1 = (1/2, +1), \theta_2 = (1/3, +1), \theta_b = (1/2, -1)\}$ and suppose that $\mu(\theta_2) > \mu(\theta_b) > 2\mu(\theta_1)$ and $X = \{2, 3\}$. Note that $X$ is the set of efficient production level absent externality concerns.

One can show that an optimal mechanism with contract over $X \times Y$ is

$$\sigma(x,y|\theta_1) = \begin{cases} 3/4 & \text{if } x = 2, y = 1 \\ 1/4 & \text{if } x = 2, y = 0 \end{cases} \qquad \sigma(x,y|\theta_2) = 1 \text{ if } x = 3, y = 1$$

$$\sigma(x,y|\theta_b) = \begin{cases} \frac{\mu(\theta_1)}{\mu(\theta_b)} 3/4 & \text{if } x = 2, y = 1 \\ \frac{\mu(\theta_1)}{\mu(\theta_b)} 1/4 & \text{if } x = 2, y = 0 \\ 1 - \frac{\mu(\theta_1)}{\mu(\theta_b)} & \text{if } x = 3, y = 1 \end{cases}$$

It is then easy to check when the government observes $y$, it is better off playing $y$.

However, Proposition 1 gives us a direct way of knowing that there is no need to contract over $Y$ by noticing that $v(x, y, \theta) = \nu(\theta)u(x, y, \theta)$ with $\nu\big(\theta = (c, e)\big) = e$ without having to solve for the optimal mechanism.

Note that the condition for the zero value of commitment in Vohra et al. (2021) is that the optimal mechanism is partitional, in the sense that the type space can be partitioned such that all types withing the same element of the partition get the same allocation and strictly prefer their allocation to one of another element of the partition. Here type $\theta_b$ is indifferent between types $\theta_1$ and $\theta_2$'s allocation, thus their condition does not hold. ■

In Proposition 1, I showed that the principal does not need to be able to contract over actions when a mechanism can output messages like in Myerson (1982). In this section, I show that the optimal mechanism when preferences are partially aligned can be implemented by directly observing the input messages, and thus a randomising mechanism is not necessary. I then explain how these results can be used to generalise a result on commitment from Glazer and Rubinstein (2004).

A mechanism is now $\sigma : M \to \Delta X$, a strategy for the agent is $\xi : \Theta \to \Delta M$ and a strategy for the principal is $\alpha : M \to \Delta Y$. The principal can commit to $\sigma$ but not to $\alpha$. Because the the principal does not have access to a randomising mechanism, truth-telling is not necessarily an optimal strategy for the agent. Note that Bester and Strausz (2001) develop a method to solve this kind of problem with limited commitment, but I will not use their results directly in the following proposition.

**Proposition 2.** *If $v(x, y, \theta) = \nu(\theta)u(x, y, \theta)$ for some $\nu : \Theta \to \mathbb{R}$, the optimal mechanism can be implemented by directly observing the message of the agent.*

The proof proceeds by showing that there always is an equilibrium of the auxiliary game where types such that $\nu(\theta) \geq 0$ have a deterministic allocation. This is enough to show that all types distribution over messages is a best-reply as these types where the only ones not directly maximising their strategy in the auxiliary game.

I conclude this section by explaining how to use the Proposition 1 and Proposition 2 to show existing results in the literature on mechanism design with evidence. Glazer and Rubinstein (2004) study a setting where a speaker wants to persuade a listener to takes a certain action, accept. The listener on the other hand only wants to accept a subset of types, and wants to reject others. The speaker sends a message to a listener. Upon hearing the message, the listener chooses a test from an exogenously given set of verification technology. In Glazer and Rubinstein (2004), the verification technology is the perfect verification of one dimension of a multidimensional type. Here I allow for arbitrary, finite set of verification technology where a test is a mapping from types to distribution over signals. Formally, let $T \subset \{\pi : \Theta \to \Delta R\}$, where $|R| < \infty$ is a finite set of messages ($R$ stands for report). The action of the listener is accept or reject, $a \in \{0,1\}$. Abusing notation, the speaker's payoff is $u(a,\theta) = a$ and the listener's is $v(a,\theta) = \nu(\theta)a$.

A mechanism in Glazer and Rubinstein (2004) is a mapping $\alpha : M \to \Delta(T \times \{0,1\}^R)$, i.e., a mapping from input messages to distribution over verification technologies and decision for each realised report. A strategy for the speaker is a mapping $\sigma : \Theta \to \Delta M$.

Glazer and Rubinstein (2004) show that the optimal mechanism, $\alpha$, can be implemented without commitment. That is it is the outcome a Perfect Bayesian Equilibrium of a game where the speaker first makes a report, then the listener chooses a test based on the report and then based on the report and observed outcome of the test, takes an action. In Proposition 1, we have already shown that $\alpha$ is a best-response to $\sigma$ (here there are no contractible actions). To fully extend Glazer and Rubinstein (2004)'s result, we need to show that the optimal $\sigma$ is a best-repy to $\alpha$, which is guaranteed by Proposition 2.

# 4 Conclusion

I have presented a method leveraging the envelope theorem for saddle-point problems to show that commitment has no value in some economic problems. The advantage of this method is that it does not necessitate checking that the principal actually best-replies to the information revealed. Moreover it has a natural economic interpretation in terms of the marginal contribution of the constraints.

Many models are set up as maximisation problem under constraints with an explicit best-response constraints like in macroeconomic models of optimal policies (see e.g., Ljungqvist and Sargent, 2018) or mechanism design problems without commitments (Bester and Strausz, 2001; Doval and Skreta, 2022). The condition (1) can be used directly in these models to understand whether commitment has value or not.

Proposition 1 shows that commitment has no value when the principal's preferences are partially aligned with the agent's. A similar result has been shown in other contexts where the principal can learn about the agent's type (Glazer and Rubinstein, 2004; Ben-Porath et al., 2021; Hancart, 2022). The result indicates that the key assumption to show this result is on the preferences of the principal and not on the fact that hard information is revealed in the problem. I conjecture that other results from the literature on mechanism design with evidence can be proven using the method here like those of Ben-Porath et al. (2019) or Hart et al. (2017).

# References


Ben-Porath, E., Dekel, E. and Lipman, B. L. (2019), 'Mechanisms with evidence: Commitment and robustness', *Econometrica* **87**(2), 529–566.
**URL:** *https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA14991*

Ben-Porath, E., Dekel, E. and Lipman, B. L. (2021), 'Mechanism design for acquisition of/stochastic evidence'.

Bertsimas, D. and Tsitsiklis, J. N. (1997), *Introduction to linear optimization*, Vol. 6, Athena Scientific Belmont, MA.

Bester, H. and Strausz, R. (2000), 'Imperfect commitment and the revelation principle: the multi-agent case', *Economics Letters* **69**(2), 165–171.

Bester, H. and Strausz, R. (2001), 'Contracting with imperfect commitment and the revelation principle: the single agent case', *Econometrica* **69**(4), 1077–1098.

Doval, L. and Skreta, V. (2022), 'Mechanism design with limited commitment', *Econometrica* **90**(4), 1463–1500.

Glazer, J. and Rubinstein, A. (2004), 'On optimal rules of persuasion', *Econometrica* **72**(6), 1715–1736.

Hancart, N. (2022), 'Optimal menu of tests'.

Hart, S., Kremer, I. and Perry, M. (2017), 'Evidence games: Truth and commitment', *American Economic Review* **107**(3), 690–713.

Ljungqvist, L. and Sargent, T. J. (2018), *Recursive macroeconomic theory*, MIT Press.

Milgrom, P. and Segal, I. (2002), 'Envelope theorems for arbitrary choice sets', *Econometrica* **70**(2), 583–601.

Myerson, R. B. (1982), 'Optimal coordination mechanisms in generalized principal–agent problems', *Journal of mathematical economics* **10**(1), 67–81.

Vohra, R., Espinosa, F. and Ray, D. (2021), 'A principal-agent relationship with no advantage to commitment', *Pure and Applied Functional Analysis* **6**(5), 1043–1054.

# A Proof of Theorem 1

First note that $\overline{V} \geq V$. It is also true that $\max_\alpha \overline{V}(\alpha) = \overline{V} \geq V$. The goal of the proof is to show that $\max_\alpha \overline{V}(\alpha) = V$. To do so, I will show that if the condition of the theorem is satisfied for $(\sigma^*, \alpha^*)$ maximising $\max_\alpha \overline{V}(\alpha)$, then $\alpha^* \in BR(\sigma^*)$.

Let $\mathcal{L}(\sigma, \lambda; \alpha)$ be the Lagrangian associated with $\overline{\mathcal{V}}(\alpha)$.

From the assumptions of Theorem 1,

$$\overline{V}(\alpha) = \max_{\sigma} \min_{\lambda} \mathcal{L}(\sigma, \lambda; \alpha)$$

By Milgrom and Segal (2002), for any selection $\sigma \in \arg\max \min_{\lambda'} \mathcal{L}(\sigma', \lambda'; \alpha)$ and $\lambda \in \arg\min \max_{\sigma'} \mathcal{L}(\sigma', \lambda'; \alpha)$, for each element of $\alpha$, $\alpha_i$

$$\frac{d\overline{V}(\alpha)}{d\alpha_i} = \frac{\partial \mathcal{L}}{\partial \alpha_i} \quad \text{a.e.}$$

Furthermore, Milgrom and Segal (2002) show that both the left- and right-derivative exist.

If $A = \times_{i=1}^{M} \Delta(Y_i)$ for finite $Y_i$, then I can also show that the derivative exists at the optimal $\alpha^*$. Denote a typical element of $\alpha$ by $\alpha(y|m)$.

If $\alpha^*(y|m) \in \{0, 1\}$, then $\left.\frac{d\overline{V}(\alpha)}{d\alpha(y|m)}\right|_{\alpha^*} = \left.\frac{\partial \mathcal{L}}{\partial \alpha(y|m)}\right|_{\alpha^*}$ as $\overline{V}$ is left- and right-differentiable.

If there is $\alpha^*(y|m), \alpha^*(y'|m) \in (0, 1)$ (there can never be only one interior solution), a necessary condition for optimality is that

$$\frac{d^+V(\alpha)}{d\alpha(y|m)} + \frac{d^-V(\alpha)}{d\alpha(y'|m)} = 0 \quad \text{and} \quad \frac{d^-V(\alpha)}{d\alpha(y|m)} + \frac{d^+V(\alpha)}{d\alpha(y'|m)} = 0 \tag{2}$$

when evaluated at $\alpha^*$. To see why it is true, suppose for example that $\frac{d^+V(\alpha)}{d\alpha(y|m)} + \frac{d^-V(\alpha)}{d\alpha(y'|m)} > 0$.

Then we could increase $\alpha^*(y|m)$ by some small $\epsilon$ and decrease $\alpha^*(y'|m)$ by $\epsilon$ and get a feasible and strictly higher $\overline{V}(\alpha)$. A similar argument can be made for all possible contradictions of the statement above.

We want to show that for $\tilde{y} = y, y'$, $\frac{d^+V(\alpha)}{d\alpha(\tilde{y}|m)} = \frac{d^-V(\alpha)}{d\alpha(\tilde{y}|m)}$ at $\alpha^*$. Suppose they are different.

Note that (2) implies that $\frac{d^+V(\alpha)}{d\alpha(y|m)} - \frac{d^-V(\alpha)}{d\alpha(y|m)} = \frac{d^+V(\alpha)}{d\alpha(y'|m)} - \frac{d^-V(\alpha)}{d\alpha(y'|m)}$ when evaluated at $\alpha^*$.

If $\frac{d^+V(\alpha)}{d\alpha(y|m)}, \frac{d^-V(\alpha)}{d\alpha(y|m)} \geq 0$ with at least one strict inequality, (2) implies $\frac{d^+V(\alpha)}{d\alpha(y'|m)}, \frac{d^-V(\alpha)}{d\alpha(y'|m)} \leq 0$ with at least one strict inequality. But then there is a strict profitable deviation by increasing $\alpha(y|m)$ by some small $\epsilon$ and decreasing $\alpha(y'|m)$ by the same $\epsilon$.

If $\frac{d^+V(\alpha)}{d\alpha(y|m)} \geq 0 \geq \frac{d^-V(\alpha)}{d\alpha(y|m)}$ with at least one strict inequality, then $\frac{d^-V(\alpha)}{d\alpha(y'|m)} \geq 0 \geq \frac{d^+V(\alpha)}{d\alpha(y'|m)}$ with at least one strict inequality. But then $\frac{d^+V(\alpha)}{d\alpha(y|m)} - \frac{d^-V(\alpha)}{d\alpha(y|m)} > 0$ and $\frac{d^-V(\alpha)}{d\alpha(y'|m)} - \frac{d^+V(\alpha)}{d\alpha(y'|m)} > 0$, a contradiction.

All other possibilities can be proven similarly and therefore with have established that $\overline{V}(\alpha)$ is differentiable at $\alpha^*$.

If the condition $\lambda \cdot \nabla_\alpha g(\alpha^*, \sigma) = 0$ is satisfied, then $\frac{d\overline{V}(\alpha)}{d\alpha(y|m)} = \frac{\partial v(\alpha,\sigma)}{\partial\alpha(y|m)}$ at $\alpha^*$ for all $y, m$.

Because $\alpha^* \in \arg\max_\alpha \overline{V}(\alpha)$ and $V$ is differentiable at $\alpha^*$, the first-order conditions are satisfied and well-defined. Because first-order conditions are sufficient for $v$, then $\alpha^* \in BR(\sigma)$.

# B Proof of Proposition 1

Fix a strategy $\alpha : M \to \Delta Y$. Abusing notation, write $u(x, m, \theta; \alpha) = \sum_y \alpha(y|m)u(x, y, \theta)$. The principal's problem when he commits to $\alpha$ is

$$\begin{aligned}
\max_{\sigma \geq 0} \sum_\theta \sum_{x,m} \mu(\theta)\sigma(x, m|\theta)\nu(\theta)u(x, m, \theta; \alpha) \\
\text{s.t. for all } \theta, \theta', \sum_{x,m} \big(\sigma(x, m|\theta) - \sigma(x, m|\theta')\big)u(x, m, \theta; \alpha) \geq 0 \\
\text{for all } \theta, \sum_{x,m} \sigma(x, m|\theta) = 1
\end{aligned}$$

The dual program is

$$\begin{aligned}
&\min_{\lambda,\eta} \sum_\theta \eta(\theta) \\
&\text{s.t. for all } x, m, \theta,\ -u(x, m, \theta; \alpha) \sum_{\theta'} \lambda(\theta, \theta') + \sum_{\theta'} \lambda(\theta', \theta)u(x, m, \theta'; \alpha) + \eta(\theta) \geq \mu(\theta)\nu(\theta)u(x, m, \theta; \alpha) \\
&\lambda(\theta, \theta') \geq 0,\ \eta(\theta) \in \mathbb{R}
\end{aligned}$$

where $\lambda(\theta, \theta')$ is the dual variable associated with the IC constraint of type $\theta$ deviating to $\theta'$ and $\eta(\theta)$ is the dual variable associated with the feasibility constraint of type $\theta$.

I am going to prove the optimality of a solution by verifying complementary slackness condition. Specifically, I will (1) guess values for $\sigma, \lambda, \eta$, (2) verify that they are feasible in their respective problem and (3) verify complementary slackness conditions. If the variables are feasible and satisfy the complementary slackness conditions, then they are optimal (see e.g., Bertsimas and Tsitsiklis, 1997, Theorem 4.5).

To this end, first define the following normal-form game. Let $\overline{\Theta} = \{\theta : \nu(\theta) \geq 0\}$ and $\underline{\Theta} = \{\theta : \nu(\theta) < 0\}$. The players are types in $\Theta$. The action space of $\theta \in \overline{\Theta}$ is $X \times M$, with strategy $\overline{s}:\overline{\Theta}\to\Delta(X\times M)$. The action space of $\theta\in\underline{\Theta}$ is $\overline{\Theta}$, with strategy $\underline{s}:\underline{\Theta}\to\Delta\overline{\Theta}$. The payoffs are

$$\text{for } \theta\in\overline{\Theta},\ \tilde{u}(x,m,\underline{s}|\theta)=\mu(\theta)\nu(\theta)u(x,m,\theta;\alpha)+\sum_{\underline{\theta}\in\underline{\Theta}}\mu(\underline{\theta})\nu(\underline{\theta})\underline{s}(\theta|\underline{\theta})u(x,m,\underline{\theta};\alpha)$$

$$\text{for } \theta\in\underline{\Theta},\ \tilde{u}(\overline{\theta},\overline{s}|\theta)=\sum_{x,m}\overline{s}(x,m|\overline{\theta})u(x,m,\theta;\alpha)$$

Take an equilibrium of this normal-form game, $(\overline{s},\underline{s})$. We guess the following values:

for $\theta\in\overline{\Theta}$, $\sigma(x,m|\theta)=\overline{s}(x,m|\theta)$

for $\theta\in\underline{\Theta}$, $\sigma(x,m|\theta)=\sum_{\overline{\theta}}\underline{s}(\overline{\theta}|\theta)\overline{s}(x,m|\overline{\theta})$

for $\theta\in\overline{\Theta},\theta'\in\Theta$, $\lambda(\theta,\theta')=0$

for $\theta,\theta'\in\underline{\Theta}$, $\lambda(\theta,\theta')=0$

for $\theta\in\underline{\Theta},\theta'\in\overline{\Theta}$, $\lambda(\theta,\theta')=\mu(\theta)|\nu(\theta)|\underline{s}(\theta'|\theta)$

for $\theta\in\underline{\Theta}$, $\eta(\theta)=0$

for $\theta\in\overline{\Theta}$ and $(x_\theta,m_\theta)\in\operatorname{supp}\overline{s}(\cdot|\theta)$, $\eta(\theta)=\mu(\theta)\nu(\theta)u(x_\theta,m_\theta,\theta;\alpha)+\sum_{\underline{\theta}}\mu(\underline{\theta})\nu(\underline{\theta})u(x_\theta,m_\theta,\underline{\theta};\alpha)$

We can now verify that these guesses are feasible.

First let us check that the primal problem is feasible. Note that the allocation of of types in $\underline{\Theta}$ are convex combinations of allocations of type in $\overline{\Theta}$ so it is enough to check deviations to types in $\overline{\Theta}$ by linearity of the expected utility.

Let's check first incentives of types $\theta\in\overline{\Theta}$ to deviate. By the equilibrium conditions of the game defined above, any $\theta\in\overline{\Theta}$ is better off playing his equilibrium strategy over another

$\theta' \in \overline{\Theta}$:

$$\sum_{x,m} \overline{s}(x,m|\theta)\big[\mu(\theta)\nu(\theta)u(x,m,\theta;\alpha) + \sum_{\underline{\theta}\in\underline{\Theta}} \mu(\underline{\theta})\nu(\underline{\theta})\underline{s}(\theta|\underline{\theta})u(x,m,\underline{\theta};\alpha)\big]$$
$$\geq \sum_{x,m} \overline{s}(x,m|\theta')\big[\mu(\theta)\nu(\theta)u(x,m,\theta;\alpha) + \sum_{\underline{\theta}\in\underline{\Theta}} \mu(\underline{\theta})\nu(\underline{\theta})\underline{s}(\theta|\underline{\theta})u(x,m,\underline{\theta};\alpha)\big]$$

Rearranging, we get

$$\mu(\theta)\nu(\theta)\sum_{x,m}\big(\overline{s}(x,m|\theta) - \overline{s}(x,m|\theta')\big)u(x,m,\theta;\alpha)$$
$$\geq \sum_{\underline{\theta}\in\underline{\Theta}} \mu(\underline{\theta})\nu(\underline{\theta})\underline{s}(\theta|\underline{\theta})\sum_{x,m}\big(\overline{s}(x,m|\theta') - \overline{s}(x,m|\theta)\big)u(x,m,\underline{\theta};\alpha)$$

Note that the LHS is the IC constraint of type $\theta$ when considering deviating to $\theta'$. Moreover, $\underline{s}(\theta|\underline{\theta}) > 0$ implies that $\sum_{x,m}\big(\overline{s}(x,m|\theta') - \overline{s}(x,m|\theta)\big)u(x,m,\underline{\theta};\alpha) \leq 0$ from the equilibrium behaviour of $\underline{\theta}$. Because $\nu(\underline{\theta}) < 0$, the RHS is positive and so is the LHS.

We can now turn to the IC constraints of types in $\underline{\Theta}$. Again from the equilibrium behaviour in the normal-form game, for $\theta \in \underline{\Theta}$,

$$\sum_{x,m}\sigma(x,m|\theta)u(x,m,\theta;\alpha) = \sum_{\theta'\in\overline{\Theta}} \underline{s}(\theta'|\theta)\sum_{x,m}\overline{s}(x,m|\theta')u(x,m,\theta;\alpha) \geq \sum_{x,m}\sigma(x,m|\tilde{\theta})u(x,m,\theta;\alpha) = \sum_{x,m}\overline{s}(x,$$

for all $\tilde{\theta} \in \overline{\Theta}$. Thus IC constraints of the types in $\underline{\Theta}$ are also satisfied.

Let's now turn to the feasibility of the dual problem. For constraints $(x, m\theta)$ with $\theta \in \underline{\Theta}$, plugging in the guessed values gives

$$-u(x,m,\theta;\alpha)\sum_{\overline{\theta}}\mu(\theta)|\nu(\theta)|\underline{s}(\overline{\theta}|\theta) \geq \mu(\theta)\nu(\theta)u(x,m,\theta;\alpha)$$

Because $-|\nu(\theta)| = \nu(\theta)$ for $\theta \in \underline{\Theta}$ and $\sum_{\overline{\theta}}\underline{s}(\overline{\theta}|\theta) = 1$, this inequality holds with equality.

For constraints $(x, m, \theta)$ with $\theta \in \overline{\Theta}$, plugging in the guessed values gives

$$\mu(\theta)\nu(\theta)u(x_\theta, m_\theta, \theta; \alpha) + \sum_{\underline{\theta}} \mu(\underline{\theta})\nu(\underline{\theta})\underline{s}(\theta|\underline{\theta})u(x_\theta, m_\theta, \underline{\theta}; \alpha)$$
$$\geq \mu(\theta)\nu(\theta)u(x, m, \theta; \alpha) + \sum_{\underline{\theta}} \mu(\underline{\theta})\nu(\underline{\theta})\underline{s}(\theta|\underline{\theta})u(x, m, \underline{\theta}; \alpha)$$

which holds because it is derived from equilibrium behaviour of the normal-form game.

The last step to prove optimality is to verify complementary slackness constraints. That is we must verify that if a constraint is slack, its associated variable in the other problem is equal to zero. For the IC constraints of the types in $\overline{\Theta}$, the dual variable is always zero. For the IC constraint of types in $\underline{\Theta}$, the only potentially non-zero dual variables are those associated with a deviation to a type in $\overline{\Theta}$. If for $\tilde{\theta}^+ \in \overline{\Theta}$,

$$\sum_{x,m} \sigma(x, m|\theta)u(x, m, \theta; \alpha) = \sum_{\overline{\theta}} \sum_{x,m} \underline{s}(\overline{\theta}|\theta)\overline{s}(x, m|\overline{\theta})u(x, m, \theta; \alpha) > \sum_{x,m} \overline{s}(x, m|\tilde{\theta}^+)u(x, m, \theta; \alpha)$$

then $\underline{s}(\tilde{\theta}^+|\theta) = 0$ and thus $\lambda(\theta, \tilde{\theta}^+) = 0$.

In the dual problem, all constraints $(x, m, \theta)$ for $\theta \in \underline{\Theta}$ are binding.

For constraints $(x, m, \theta)$ for $\theta \in \overline{\Theta}$, if

$$\mu(\theta)\nu(\theta)u(x_\theta, m_\theta, \theta; \alpha) + \sum_{\underline{\theta}} \mu(\underline{\theta})\nu(\underline{\theta})\underline{s}(\theta|\underline{\theta})u(x_\theta, m_\theta, \underline{\theta}; \alpha)$$
$$> \mu(\theta)\nu(\theta)u(x, m, \theta; \alpha) + \sum_{\underline{\theta}} \mu(\underline{\theta})\nu(\underline{\theta})\underline{s}(\theta|\underline{\theta})u(x, m, \underline{\theta}; \alpha)$$

then $\overline{s}(x, m|\theta = \sigma(x, m|\theta) = 0$ from the equilibrium behaviour in the normal-form game.

Thus the guessed values are optimal. Let's now verify that the condition of Theorem 1 is satisfied. For all $m, y$,

$$\sum_{\theta,\theta'} \lambda(\theta,\theta') \sum_{x} \big(\sigma(x,m|\theta) - \sigma(x,m|\theta')\big) u(x,y,\theta) = 0$$

Using our characterisation above,

$$\sum_{\underline{\theta},x} \mu(\underline{\theta})|\nu(\underline{\theta})|u(x,y,\underline{\theta}) \sum_{\overline{\theta}} \underline{s}(\overline{\theta}|\underline{\theta}) \big( \sum_{\theta' \in \overline{\Theta}} \overline{s}(x,m|\theta')\underline{s}(\theta'|\underline{\theta}) - \overline{s}(x,m|\overline{\theta}) \big)$$
$$= \sum_{\underline{\theta},x} \mu(\underline{\theta})|\nu(\underline{\theta})|u(x,y,\underline{\theta}) \big( \sum_{\theta' \in \overline{\Theta}} \overline{s}(x,m|\theta')\underline{s}(\theta'|\underline{\theta}) - \sum_{\overline{\theta}} \underline{s}(\overline{\theta}|\underline{\theta})\overline{s}(x,m|\overline{\theta}) \big) = 0$$

using that $\sum_{\overline{\theta}} \underline{s}(\overline{\theta}|\underline{\theta}) = 1$.

# C Proof of Proposition 2

I will show that types in $\overline{\Theta}$ play a pure strategy in the auxiliary game. The proof follows closely the one of Hancart (2022) and is given here for completeness.

Note first that the auxiliary game can be represented by a saddle-point problem:

$$\max_{\overline{s}} \min_{\underline{s}} \sum_{\overline{\theta}} \sum_{x,m} \overline{s}(x,m|\overline{\theta}) \Big[ \mu(\overline{\theta})\nu(\overline{\theta})u(x,m,\overline{\theta};\alpha) + \sum_{\underline{\theta}} \mu(\underline{\theta})\nu(\underline{\theta})\underline{s}(\overline{\theta}|\underline{\theta})u(x,m,\underline{\theta};\alpha) \Big]$$

Note that the saddle-point is well define as the objective function is linear in both arguments and $\overline{s}, \underline{s}$ are elements of compact, covnex subset of $\mathbb{R}^n$.

Suppose there is $s^* \in \arg\max_{\overline{s}} \min_{\underline{s}} \sum_{\overline{\theta}} \sum_{x,m} \overline{s}(x,m|\overline{\theta})\tilde{u}(x,m,\underline{s}|\overline{\theta})$ such that $s^*(x,m|\overline{\theta}), s^*(x',m'|\overline{\theta}) > 0$ for some $\overline{\theta}$.

Assume that for all $(x, m)$, $(x', m')$ and $Z \subseteq \underline{\Theta}$,

$$(3) \quad \mu(\overline{\theta})\nu(\overline{\theta})u(x, m, \overline{\theta}; \alpha) + \sum_{\underline{\theta} \in Z} \mu(\underline{\theta})\nu(\underline{\theta})u(x, m, \underline{\theta}; \alpha) \neq \mu(\overline{\theta})\nu(\overline{\theta})u(x', m', \overline{\theta}; \alpha) + \sum_{\underline{\theta} \in Z} \mu(\underline{\theta})\nu(\underline{\theta})u(x', m', \underline{\theta}; \alpha)$$

Note that $s^*$ must be optimal for any selection of $\arg\min \sum_{\overline{\theta}} \sum_{x,m} \overline{s}(x, m|\overline{\theta})\tilde{u}(x, m, \underline{s}|\overline{\theta})$ and in particular for the following:

$$\underline{s}(\overline{\theta}|\underline{\theta}) = 1 \Leftrightarrow \sum_{x,m} \overline{s}(x, m|\overline{\theta})u(x, m, \underline{\theta}; \alpha) \geq \sum_{x,m} \overline{s}(x, m|\tilde{\theta})u(x, m, \underline{\theta}; \alpha), \text{ for all } \tilde{\theta} \in \overline{\Theta}$$

Note that under this selection, if a type $\underline{\theta}$ does not mimic $\overline{\theta}$, it means that it strictly prefers another type in $\overline{\Theta}$. But now observe that $\overline{\theta}$ can modify slightly its strategy and because (3) it would striclty increase his payoff. If the modification is small enough, it would not attract new types in $\underline{\Theta}$ as they all strictly prefer another type. Thus $\overline{\theta}$ plays must play a pure strategy.

Now note that any payoffs satisfying condition ( ) define a dense subset of the payoff space, $(u(x, m, \theta; \alpha)_{(x,m,\theta}$, using the usual metric for $\mathbb{R}^n$. Indeed, condition ( ) is a finite system of inequalities and perturbation to $u$ upsets any equality. Take a sequence in the payoff space such that for any member of the sequence, condition () is satisfied such that the sequence converges to an element of the payoff space where conditoin ( ) is not satisfied. Take an associated sequence of $s^{+,n}$ where $n$ indexes the sequence. $(s^{*,n}$ is a bounded sequence in a closed subset of $\mathbb{R}^n$ so it admits a converging subsequence. This subsequence contains only pure strategies so it must converge to a pure strategy. By upper hemicontinuity of the Nash Equilibrium correspondence, the limit is a Nash Equilibrium and thus there is an equilibrium $\overline{s}$ in pure strategy for any payoff.