Taxation Principle with Moral Hazard

Francisco Poggi and Bruno Strulovici

University of Mannheim and Northwestern University*

September 15, 2025

Abstract

In principal-agent settings, a *Taxation Principle* is sometimes invoked to reduce the analysis to mechanisms that do not directly elicit the agent's private information. Standard versions of the principle assume that the principal can condition transfers on the agent's choices. We consider environments in which the agent's actions are not fully observable or contractible. For instance, a reckless agent may face a fine only if this recklessness results in an observable accident. We identify (i) conditions under which the Taxation Principle extends to such settings, (ii) weaker conditions in which the principal holds for a subset of social choice functions, and (iii) extensions of the principle to dynamic contracting.

*Emails: poggi@uni-mannheim.de and b-strulovici@northwestern.edu. Support from the National Science Foundation (Grant No. 1151410) and the German Research Foundation (DFG) through CRC TR 224 (Project B02) is gratefully acknowledged.

1 Introduction

When an agent with private information and quasilinear utility chooses an action that results in a transfer, the Taxation Principle (Hammond (1979); Guesnerie (1981)) asserts that a regulator or mechanism designer does not need to directly inquire about the agent's private information, but can instead "tax" each action without any loss of generality on the set of social choice functions that can be implemented.

In addition to their simplicity, these tax mechanisms are of particular interest when ex ante contracting is impossible, communication is impractical, or there are privacy concerns. In legal settings, for instance, it is typically impossible to contract with agents ex ante: a criminal does not bargain or communicate with a prosecutor before committing his crime. Communication occurs *only after the crime was committed, and only if the criminal is apprehended.*

Two key assumptions of the standard Taxation Principle are typically violated in such applications: (i) it is often impossible to perfectly observe the action chosen by the agent and, consequently, (ii) it is often impossible to "tax" (i.e., penalize) each action. For example, suppose that the action of interest concerns whether to commit a crime. One would ideally like to penalize criminal acts, but such acts are not perfectly observable and can be penalized only if the criminal is apprehended and found guilty. A third limitation of the Taxation Principle is its focus on quasilinear utility. In reality, agents face risk aversion with respect to money, sentences, or, more generally, any instrument available to the designer, and agents' valuations can depend on their types and other variables.

This paper considers the following question: Is there a result similar to the Taxation Principle when actions are partly unobservable and can be "taxed" only after specific events? We consider a general model in which an agent has private information and privately chooses an action whose stochastic outcome may or may not be contractible. The principal's only instrument to influence the agent's behavior is to choose a penalty after a contractible outcome.

We compare two classes of mechanisms: one in which the agent must report his type ex ante and receives a report-contingent (and outcome-contingent) penalty if and when a contractible outcome occurs, and one in which the agent does not report his type ex ante and the penalty depends only on the outcome. We ask under which conditions the set of implementable social choice functions (i.e., map from agent types to actions) in these two classes of mechanisms is the same.

This paper has three main results. Beginning with a static version of the problem, we first provide a sufficient condition on the environment for the Taxation Principle to hold. Intuitively, this condition guarantees any contractible outcome is a sufficient statistic for identifying the ex-ante distribution of contractible outcomes. Moreover, we show that it is the weakest sufficient condition that guarantees the Taxation Principle: when the condition is violated, it is possible to construct agent preferences and find an implementable social choice function that cannot be implemented with a tax mechanism.

The Taxation Principle holds when *every* implementable social choice function can be implemented by a tax mechanism, which is independent of the principal's objective function. In practice, however, the principal may care about only a subset of social choice functions. For instance, a regulator may aim to encourage firms to launch safer products and thus have no interest in the implementation of social choice functions where firms exclusively launch unsafe products. Our second main result derives simple, jointly sufficient conditions on the environment and the social choice function that guarantee tax implementability. Specifically, we show that when each realized outcome is sufficient to identify the ex-ante distribution of outcomes *conditional* on the action being prescribed by the social choice, the scf can be implemented with a tax mechanism. In Section 5, we apply these results to study the design of sentences in plea bargaining and the design of optimal liability rules when firms acquire information about product risks before deciding between launching the product or abandoning its development.

Many important applications, however, involve an agent making decisions sequentially. For instance, an agent may observe an evolving state and take actions at different points in time. Our third main result adapts our analysis to such dynamic settings. We show that the logic of the Taxation Principle carries over to dynamic environments, even when the agent's action influences the evolution of the state.

Literature (to be completed)

This paper contributes mainly to the literature on contract theory with private information and moral hazard. Our main results complement the classical Revelation Principle in Myerson (1982), which states that is without loss of generality to focus on direct recommendation mechanisms that satisfy obedience and truthful constraints. We complement this result by providing conditions under which even simpler mechanisms can be used to implement any implementable social choice function.

The Taxation Principle is deeply connected with the literature on optimal taxation. A seminal work is Mirrlees (1971), who studies optimal wealth taxes in the context of unobservable differences in productivity. Kocherlakota (2005) extends the work of Mirrlees to dynamic settings, and show that a symmetric constrained Pareto optimal allocation can be implemented using a tax system that is linear.

Perez-Richet and Skreta (2024) study mechanism design when there is soft information—that can be freely manipulated—and semi-hard information—that entails a cost for falsification. To do so, they rely on a *scoring-based mechanism*, which states that under certain conditions it is without loss to focus on mechanism that are only based on the final outcome, and not on any reports, as in the spirit of our Taxation Principle. They are interested in randomized mechanism, while in this paper we focus on deterministic mechanisms.

Our paper also applies our static Taxation Principle to prove conditions on dynamic environments under which there is no need to rely on interim communication with the agent in order to implement all desired social choice functions. In a recent paper, Liu (2025) study the design of rewards in a dynamic environment with stochastically evolving private information and where the action that the agent takes is publicly observable. Our main result for dynamic environments can be applied to simplify the design problem, even in conditions in which the state evolution depends on the history of actions chosen by the agent.

The rest of the paper is organized as follows: Section 1.1 motivates the paper with some illustrative examples that fit our framework. Section 2 introduces the general abstract setting and the mechanisms that the principal can use to shape the incentives of the agent. For completeness, we include the proof of the Taxation Principle when actions are perfectly observable by the principal. In Section 3, we relax the assumption of perfect observability and present the conditions for the Taxation Principle to hold in the main result of the paper, namely Theorem 1. Section 4 presents the result for specific social choice functions, which is applied in Section 5 to plea bargaining and experiments with caution. Finally, in Section 6, we extend our analysis to dynamic settings where the state evolves stochastically conditional on the agent's chosen actions to show that, when actions are contractible, a version of the Taxation Principle holds.

1.1 Examples

In the next examples, we explore different situations that will be covered in our framework.

Example I: Binary state A project can be risky ($\omega = 1$) or safe ($\omega = 0$). There is a common prior p denoting the probability of the project being risky. The agent follows these steps: First, he chooses an experiment e from a set of feasible experiments $\mathcal{E}$. Each experiment is represented by a map $s : \{0, 1\} \to \Delta(S)$, specifying the probability distribution over signal realizations S for each state ω. After selecting the experiment, the agent observes the signal realization s and chooses a level of precaution $x \in X$. If the project is risky, an accident occurs with probability $p(x)$. No accident occurs when the project is safe. Following an accident, the principal conducts an investigation and observes the experiment chosen e, the realized signal s, and the level of precaution x. Based on this information the principal chooses a sanction $t \in T$. The payoff of the agent is given by $-k(e, x, \theta) - t$, where k is a function that represents the cost of experimentation and precaution, which depends on the private type of the agent θ.

Example II: Binary state, constrained principal Consider the same binary-state framework of Example I, with the additional constraint on the principal's sanction strategy: The principal must ensure that for any two agents who obtained signals s, s' that produce the same objective evidence about the probability of an accident, in terms of the likelihood ratio of the state $\omega = 1$, the penalty in case of an accident is the same. This constraint reflects an objectivity and fairness consideration, ensuring that penalties are based solely on the evidence produced by the signals and not on any additional subjective factors.

Example III: Non-binary state, impact experiment Suppose now that the underlying state is two-dimensional (ω, w) where ω represents again whether there is a risk or not, and w the severity of the accident. Now, the agent chooses an experiment that is informative about the severity of the accident $s : \{\text{Mild}, \text{Severe}\} \to \Delta(S)$ and has private information about the probability θ. The payoff of the agent is given by $-k(e, c) - t$.

In all these examples, the agent has private information and takes actions privately. In general, the principal can design a mechanism in which the agent reports his private information and takes actions and, at the end of the day, the principal penalizes the agent as a function of the reports and outcomes. In this paper, we formalize a general setting that incorporates these examples and show that, when certain conditions are satisfied, the principal does not gain anything by eliciting information from the agent, meaning that for any actions that can be implemented, it is possible to construct a mechanism that does not require any reporting and that generates the same actions, and the same payoff for the agent.

2 General Setting

An agent of type $\theta \in \Theta$ chooses an action a from a set A. Each action generates a stochastic outcome $z \in Z$ that can be *contractible* ($z \in C$) or not ($z \in Z \setminus C$). If the outcome is contractible, the agent faces a penalty t from a set $T \subseteq \mathbb{R}$.

A designer wishes to induce type-dependent actions. To do so, the designer chooses the penalty map $g : Z \to T$ such that $g(z) \in \Gamma(z)$ where $\Gamma(z) \subset T$ represents the set of available penalties given outcome z.[1] If the outcome is non-contractible, the principal cannot penalize the agent. Let $\Gamma(z) = 0$ represent the situation in which there is no punishment.

The agent is an expected utility maximizer, with Bernoulli utility given by a function $u : \Theta \times A \times Z \times T \to \mathbb{R}$.

Definition 1 *The agent has preferences that are* separable in the penalties *if there are functions $v : \Theta \times A \times Z \to \mathbb{R}$, $q : \Theta \times A \to \mathbb{R}$, and $d : T \times Z \to \mathbb{R}$ such that:*

$$u(\theta, a, z, t) = v(\theta, a, z) - q(\theta, a) \cdot d(t, z)$$

Separability in the penalties implies that, conditional on the action a, all types $\theta \in \Theta$ share the same preference ranking over distributions of penalties. We formalize this observation next.

Let $\nu, \nu' \in \Delta(Z \times T)$. We write $\nu \succeq_{\{\theta,a\}} \nu'$ iff $E_\nu(u(\theta, a, g, z)) \geq E_{\nu'}(u(\theta, a, g, z))$

Observation 1 *Let $\nu, \nu' \in \Delta(Z \times T)$ with the same marginal distribution of outcomes $\mu \in \Delta(Z)$. If the preferences are separable in the penalties, then $\nu \succeq_{\{\theta,a\}} \nu'$ implies that $\nu \succeq_{\{\theta',a'\}} \nu'$ for all $(\theta', a') \in \Theta \times A$.*

Proof. In Appendix A.

Separability in the penalties is satisfied when agents have quasilinear preferences in t, but it also allows to represent richer preferences, for example in situations in which the cost of a sentence is higher when the agent is guilty.

Example. Let $\Theta = [0, 1]$ be the propensity of the agent to commit a crime. Let $a \in \{0, 1\}$ represent whether the agent commits the crime or not, and the outcome $z = (z_1, z_2)$ consists of two components: $z_1 \in \{0, 1\}$ indicates whether the individual is apprehended and z_1 is

[1] Available penalties might depend on the outcome z due to a physical constraint, such as the impossibility of taxing the agent more than his wealth, or to some other type of constraint. For example, a judge might be restricted by maximal sentences that depend on the verdict of the jury.

a signal that is correlated with the action a. Let $T = [0, \bar{t}]$ be the set of possible sentences where $\bar{t}$ is the maximal sentence. Let the utility of the agent take the form:

$$u(\theta, a, z, g) = \theta \cdot a - (1 - \gamma \cdot a)\ d(t)$$

The preferences represented by u satisfy the separability assumption for all parameters $\gamma \in [0, 1]$ and for every function $d : T \to \mathbb{R}$. Strictly positive γ implies that the cost of a sentence is higher for guilty agents, i.e., those agents that committed the crime.

The principal designs a mechanism to induce the agent to take a potentially type-dependent action. Two important considerations arise: First, the principal may wish to implement stochastic actions rather than deterministic ones. Second, the principal may also care directly about the rewards or penalties the agent faces, beyond its effect on the action. For simplicity, we abstract away from these concerns and define the social choice function accordingly.[2]

Definition 2 *A (deterministic)* social choice function *consists of a function $f : \Theta \to A$, that assigns an action to each type.*

In the following section, we explore which social choice functions can be implemented using different types of mechanisms.

2.1 Mechanisms

In general, a mechanism consists of a (potentially dynamic) game that the designer sets up, where the agent's private information and actions determine the outcomes. A social choice function f is implementable if there exists a mechanism and a Bayesian Nash Equilibrium such that each type θ of the agent plays action $f(\theta)$. Given the Revelation Principle, it is sufficient to consider mechanism in which the agent reports their type, and the mechanism recommends the action prescribed by $f(\theta)$.

[2]Extensions addressing these two concerns are provided in the appendix.

A *tax schedule* is a function $\tau : Z \to T$ satisfying the constraint $\tau(z) \in \Gamma(z)$ for each $z \in Z$. Let $\mathcal{T}$ denote the set of feasible tax schedules. A particularly interesting class of simple mechanisms is one in which the agent produces a report $\hat{\theta} \in \Theta$, and obtains a reward/penalty based on the report and the outcome z. We can characterize these mechanisms by the family of tax schedules induced by each report.

Definition 3 *A* direct mechanism $\{\tau_\theta\}_{\theta\in\Theta}$ *is one in which the agent produces a report* $\hat{\theta} \in \Theta$, *then chooses action* $a \in A$, *and for every realized outcome* $z \in Z$, *faces a penalty* $t = \tau_{\hat{\theta}}(z)$. *A direct mechanism is* truthful *if any type* θ *of the agent finds it optimal to report its true type.*

Revelation Principle. Any implementable social choice function can be implemented with a truthful direct mechanism.

The Revelation Principle is a foundational result in mechanism design. It allows us to reduce the space of mechanism to consider when we are interested in characterizing the set of implementable social choice functions. Moreover, it allows us to focus directly on the incentive compatibility constraints, reducing dramatically the complexity of the design problem.

A principle that is equivalent to the Revelation Principle states that we can focus without loss of generality on mechanisms where the agent chooses its own tax schedule. We denote this type of mechanisms *menu mechanisms*, as defined next.

Definition 4 *In a* menu mechanism, *the planner offers the agent a menu* $M \subseteq \mathcal{T}$, *from which the agent selects a tax schedule* τ. *The agent then takes the action* $a \in A$. *When the outcome* z *is realized, the agent pays* $\tau(z)$.

Menu Taxation Principle Any implementable social choice function can be implemented with a menu mechanism.

The Menu Taxation Principle can be directly derived from the Revelation Principle. Intuitively, starting from an implementable scf f and a direct revelation mechanism that implements it, it is possible to construct a menu mechanism where the menu corresponds to the tax schedules t_θ corresponding to each type θ. Since truthful reporting is optimal in the direct mechanism, it must also be optimal for the agent to choose t_θ in the menu mechanism, which replicates the same desired outcomes.

The Menu Taxation Principle is conceptually very close to the Revelation Principle. The key difference is that when the agent is choosing a message in a direct mechanism, it is choosing a type, and thus from a set of cardinality equal to the number of types. In a menu mechanism, the set from which the agent chooses is the set of tax schedules, which is potentially smaller. Indeed, there are social choice functions that can be implemented with a menu of cardinality one. Formally, we define these mechanisms as follows.

Definition 5 *A* tax mechanism $\tau \in \mathcal{T}$ *is one in which the agent chooses an action a. When outcome z is realized, the agent pays according to $\tau(z)$.*

Substantive Taxation Principle We say that the Substantive Taxation Principle holds if any implementable social choice function can be implemented with a tax mechanism.

When the substantive Taxation Principle holds, there is no need to communicate with the agent or offer a menu of any kind. Formally, the agent doesn't take any action other than the one stipulated by the framework of the model. The key to understanding how can we construct conditions under which the substantive taxation principle holds lies in the following observation.

Observation 2 *If preferences are separable in the penalties, for any direct mechanism the optimal report conditional on the action a is independent of the agent's true type.*

Proof. In Appendix A.

2.2 Substantive Taxation Principle with observable actions

In this section, we present the version of the substantive taxation principle in frameworks with observable actions due to Guesnerie (1981) and Hammond (1979). We assume the agent has quasilinear utility $u(\theta, a) - t$ and that $z = a$, meaning that the principal can directly observe the action that the agent chooses.

The (Substantive) Taxation Principle states that every implementable social choice function can be implemented with a tax mechanism. According to this principle, it is without loss of generality to focus on tax mechanisms for design purposes. The result is even stronger in the sense that for all implementable f and implementing mechanisms, there is a tax mechanism that implements f and such that *every type receives the same transfer than in the original mechanism.* For completeness, we state and prove the principle.

Proposition 1 *When actions are observable and the agent has quasilinear preferences, for every implementable social choice function f there is a tax mechanism that implements f and such that every type faces the same expected transfer.*

Proof. In the Appendix A.

This proposition hinges on two important assumptions: First, actions are observable and transfers can thus be tailored to each action. Second, preferences are quasilinear in the transfers. In the following sections, we study conditions under which the Taxation Principle holds when these assumptions are relaxed. We've enriched the benchmark model in three ways: (i) We include moral hazard: the action of the agent is not contractible, either because it is not observable by the principal or because it cannot be used directly to condition penalties. Instead, the principal can contract on some outcome or evidence that correlates with the action chosen by the agent; (ii) We allow the set of available penalties to depend on the ex-post observed outcome; (iii) The agent's preferences need not be quasilinear in the penalty, which encompasses situations in which penalties are non-monetary or in which the agent is risk averse.

Once we introduce moral hazard, the principal cannot condition transfers directly on the action. In this case, the report might be informative about the agents' true type, even when the preferences are separable in the penalties. The next example illustrates this.

Example Let the environment be such that there are four action $A = \{a_0, a_1, a_2, a_3\}$, there are three outcomes, all of which are contractible $Z = C = \{z_0, z_1, z_2\}$. The distribution of outcomes given actions is given by $\mu_{a_0}(z_0) = 1$. $\mu_{a_1}(z_1) = \mu_{a_2}(z_2) = 0.9$, $\mu_{a_1}(z_0) = \mu_{a_2}(z_0) = 0.1$. $\mu_{a_3}(z_1) = \mu_{a_3}(z_2) = 0.5$.

Finally, there are two types, denoted $\Theta = \{\theta_1, \theta_2\}$, and the preferences of the agent are represented by by $u(a,t) = 1_{\{a=a_0\}} + t$. Note that the preference of the agent is separable in the penalties. Finally, we impose no constraint in the set of penalties, i.e., $\Gamma(z) = \mathbb{R}$ for all $z \in Z$. Consider the following social choice function:

$$f(\theta) = \begin{cases} a_1 & \text{if} \quad \theta = \theta_1 \\ a_2 & \text{if} \quad \theta = \theta_2 \end{cases}$$

Note that f is implementable: consider the direct mechanism given by $\tau_{\theta_i}(z_j) = 1_{\{i=j\}} \cdot 2$. With this mechanism, each type does not benefit strictly from deviating from reporting the true type and taking the action prescribed by f.

However, we can show that f is not tax implementable: lets consider a tax mechanism $\tau(z_i)$. For tax implementability we require:

$$\begin{aligned} 0.9\tau(z_1) + 0.1\tau(z_0) &\geq \tau(z_0) + 1 \\ 0.9\tau(z_2) + 0.1\tau(z_0) &\geq \tau(z_0) + 1 \\ 0.9\tau(z_1) + 0.1\tau(z_0) &\geq 0.5[\tau(z_1) + \tau(z_2)] \\ 0.9\tau(z_2) + 0.1\tau(z_0) &\geq 0.5[\tau(z_1) + \tau(z_2)] \end{aligned}$$

Summing the first two inequalities and simplifying one gets:

$$\tau(z_1) + \tau(z_2) \geq 2\tau(z_0) + 2/0.9$$

But by summing the last two inequalities and simplifying one gets:

$$\tau(z_1) + \tau(z_2) \leq 2\tau(z_0)$$

These two inequalities are incompatible: so there is no tax mechanism τ that implements f.

The previous example evidences that having preferences that are separable in the penalties is not sufficient to guarantee that tariffs can implement all implementable social choice functions.

3 Sufficient conditions

In this section, we explore sufficient conditions for the Substantive Taxation Principle to hold, i.e. which conditions on the primitives of the problem guarantee that every implementable social choice function can be implemented with a tax mechanism. We start by defining a basic property of the distribution of outcomes.

Independence. For any action and type, let $\mu_{a,\theta} \in \Delta(C)$ be the outcome distribution conditional on it being contractible. We say that the independence axiom is satisfied if the distribution of contractible outcomes $\mu_{a,\theta}$ is constant in the type, i.e. if $\mu_{a,\theta} = \mu_{a,\theta'}$ for every $a \in A$ and $\theta, \theta' \in \Theta$.

Independence implies that the outcome z is potentially informative about the action that the agent took, but is not directly informative about the true type of the agent conditional on the action taken. The outcome can be indirectly informative about the type of the agent since, in equilibrium, different types will choose different actions. When the independence axiom is not satisfied, it is straightforward to construct a framework and an implementable social choice function that cannot be implemented with a tax mechanism, as the next example illustrates.

Example Let $\Theta = A = \{h, l\}$. Assume that the outcome z takes value θ with probability ϵ, and is equal to the action a and with probability $1 - \epsilon$. Let the agent get a payoff of

1 if the action mismatches the state, plus the transfer that he receives from the principal, i.e., $u(\theta, a, z, t) = 1_{\{a \neq \theta\}} + t$. Consider the social choice function in which the agent matches the state, i.e., $f(\theta) = \theta$ for all $\theta \in \Theta$. We will show that this social choice function can be implemented with a direct mechanism, but not with a tax mechanism.

To see this, we start by constructing a reward function that implements f. Consider paying the agent an amount X if z matches the report θ, and zero otherwise. If X is large enough, the agent has an incentive to tell the truth about their type and play $a = \theta$. Any tax that implements f, however, must pay the agent for both outcomes. The agent has therefore no incentive at all to actually match the state with the action.

When independence holds, we can define the following two properties for partitions of the action space, based on the support and distribution of contractible outcomes.

Observability. A partition $\mathcal{A}$ of the action space A is *observable* if for a, a' in different cells of $\mathcal{A}$, μ_a and $\mu_{a'}$ have disjoint supports. We denote by $\mathbf{A}$ the finest observable partition.

For any two actions in different cells of an observable partition, the principal can perfectly distinguish them, and thus tailor the reward to each action without affecting the rewards associated with the other one.

Invariance. A partition $\mathcal{A}$ of the action space A is *invariant* if for any cell A_i of the partition, and any two actions $a, a' \in A_i$ the distribution of contractible outcomes is the same, i.e., $\mu_a = \mu_{a'}$. We denote by $\mathbf{B}$ the coarsest invariant partition.

Invariance means that the distribution of contractible outcomes must be the same for any two actions in the same cell. Note that for quasilinear preferences, the expected payoff from following two actions in any tax mechanism would be proportional to the probability of a contractible outcome, and therefore the rewards and penalties are a very limited tool to affect the action within a cell. With more general preferences, the interaction between z, a, and t leads to potentially subtle and interesting effects.

Lemma 1 *Suppose that Independence holds. The following are equivalent:*

- *There is a partition $\mathcal{A}$ of the action space that is invariant and observable.*

- $\mathbf{A}$ *is invariant.*

- $\mathbf{B}$ *is observable.*

Proof. TBA.

Intuitively, $\mathbf{A}$ represents what the principal knows for sure about the action that the agent took, after observing the contractible outcome z. The partition $\mathbf{B}$ groups the actions according to their distribution of contractible outcomes. Having defined the partitions $\mathbf{A}$ and $\mathbf{B}$ of the action space, we are ready to state the conditions for the taxation principle to hold.

Theorem 1 *(**Substantive Taxation Principle**): Suppose that Independence holds. If* $\mathbf{A}$ *is invariant, any scf that is implementable by a direct mechanism is also implementable by a tax mechanism.*

Proof. in Appendix B.

When $\mathbf{A}$ is invariant, the principal can identify, for each contractible outcome realization, the distribution of contractible outcomes that is associated with the action that the agent took. In this case, asking the agent to report his private information is redundant, as any social choice function can be implemented with a tax mechanism. Note that this sufficient condition is completely independent of the utility of the agent, as long as is satisfies separability in the penalties. Moreover, it turns out that when $\mathbf{A}$ is not invariant, it is possible to construct a utility function such that separability is satisfied but there are implementable social choice functions that cannot be implemented with a tax mechanism, as formalized in the following result.

Proposition 2 *If* $\mathbf{A}$ *is not invariant, there is a set of types* Θ*, a set of feasible penalties* $\Gamma : Z \to R$*, a utility function* u*, and a social choice function* f *such that* f *is implementable but not tax implementable.*

Proof. in Appendix B.

In words, the equivalent conditions from Lemma 1 are the weakest conditions on the distribution of outputs that guarantee that every social choice function that is implementable can be implemented with a tariff, independently of the other aspects of the environment, such as utility function and penalty correspondence Γ.

4 Taxation Principle for specific SCF

The sufficient condition for the Taxation Principle to hold studied in Theorem 1, namely that $\mathbf{A}$ is invariant is a strong condition. There are two things to point out. First, notice that this is a condition on the environment, independent of everything else in the model, such as the utility function u, the constraints on the penalties given by Γ and the set of types Θ, or the distribution of outcomes beyond the conditional distribution for contractible ones.

Second, Theorem 1 shows that under this condition *all* implementable social choice functions can be implemented with tax mechanisms. In some contexts, however, only a subset of the social choice functions is relevant for the principal. For example, if we consider Examples I and II, it might not make sense for the principal to ask the agent to run an experiment if this experiment is not gonna be used to later condition the level of precautions. The principal might want agents to choose experiments that are valuable in the precaution decision, meaning that different levels of precautions are taken for different signal realizations. This consideration rules out certain social choice functions. To study these contexts, we derive conditions on f that can be used to determine whether the function can be implemented with a tax mechanism, as follows.

Definition 6 *Let $f(\Theta)$ denote the range of the social choice function f. A partition of the action space $\mathcal{A}$ is f-invariant if the partition $\{A_i \cap f(\Theta) : A_i \in \mathcal{A}\}$ is invariant.*

Intuitively, f-invariance requires that the principal can identify, based on the outcome z, the distribution of contractible outcomes assuming the actions falls in $f(\Theta)$, whenever possible. When this condition is satisfied for some implementable f, then f can also be implemented with a tax mechanism, as stated in the next theorem.

Theorem 2 *Suppose that Independence holds and f is implementable by a direct mechanism. Then, f is implementable by a tax mechanism if:*

- *There is an observable partition $\mathcal{A}$ that is f-invariant.*
- $\mathbf{A}$ *is f-invariant.*

Proof. In Appendix C.

In words, the conditions guarantee that, given f, any realized outcome z is a sufficient statistic for the distribution of outcomes conditional on that the outcome being contractible. These conditions are clearly satisfied for any f when $\mathbf{A}$ is invariant, as we would expect from Theorem 1. This result, however, is more versatile in applications in which the general condition does not hold, but there is a subset of social choice functions that are of particular interest to the principal. We discuss an application to Plea Bargaining in Section 5, which illustrates the usefulness of this result.

5 Applications

5.1 Offenses and Plea Bargaining

An agent decides whether to commit a crime and whether to plead guilty if arrested. The action space is therefore given by $A = \{g, i\} \times \{\hat{g}, \hat{i}\}$, where g and i represent guilty and innocent, and the hat variables represent what the agent pleads in case of being arrested. There are parameters related to the benefit from committing the crime and the dis-utility from jail time that are captured by θ, the private information of the agent. The outcome is given by $Z = \{\text{arrest}, -\text{arrest}\} \times \{\hat{g}, \hat{i}\} \times S$, where S represents a signal related to the true guilt of the agent. The outcome is contractible only if the agent is arrested. T is the set of durations of the sentence. Let the utility of the agent be given by:

$$u(\theta, a, z, t) = b(\theta)1_{\{a_1=g\}} - q(\theta, a) \cdot d(t)$$

Proposition 3 *To implement a truthful plea social choice function, there is no use in eliciting private information before the arrest.*

Proof. In a truthful plea social choice function, $f(\Theta) \subseteq \{(g, \hat{g}), (i, \hat{i})\}$. If we consider the partition of the actions according to what the agent pleads, i.e. $A_{\hat{g}} = \{(g, \hat{g}), (i, \hat{g})\}$ and $A_{\hat{i}} = \{(g, \hat{i}), (i, \hat{i})\}$, we obtain that $f(\Theta) \cap A_{\hat{g}}$ and $f(\Theta) \cap A_{\hat{i}}$ are singletons, therefore satisfying the condition for Theorem 2. This implies that f is implementable by a tax mechanism, i.e. there is no use in eliciting private information before the arrest.

5.2 Experiments and Caution

This application is based on the examples from Section 1.1. Translating the setting from Example I to our general framework, we have that the agent action set A is given by $A = \mathcal{E} \times X^S$. Where $\mathcal{E}$ is a set of experiments and X^S is the set of maps from signal realizations to levels of precautions. The outcome Z is given by $Z = \{\text{no accident}\} \cup \{(\text{accident}, e, s) : e \in \mathcal{E}, s \in S\}$, where $z = \{\text{no accident}\}$ corresponds to the only non-contractible outcome.

The distributions of contractible outcomes depends on the action $a \in A$ of the agent. In particular, the probability of $z = (\text{accident}, e, s)$ given accident, experiment e, and a map $x : S \to X$ is just $Pr(s|\omega = 1, e \in \mathcal{E})$ when the experiment is e and zero otherwise.

This framework, is application is therefore nested in the general framework that we study in this paper. In Poggi and Strulovici (2022), a companion paper, we explore the design of liability rules when the designer wants to incentivize firms to launch new products only if they are sufficiently safe. If a product is initially more likely to be risky, a firm should acquire stronger information about the product's safety before launching the product in order to gain a given degree of confidence in the product's safety.

If a product launched causes damage, the regulator can observe the strength of the evidence acquired on the product's safety. In this environment, we show that our Taxation Principle applies, but only for a subset of social choice functions: the ones that induce the agent to acquire information only if the information is going to be used. Since the regulator in this paper cares about wasting costs of information acquisition, we show that the regulator would generally not gain from the ability to elicit the firm's prior belief about the product's safety before the product is launched.

6 Dynamic Taxation Principle

In many contractual interactions, the agent takes actions over multiple periods, and information arises as time progresses. In this section, we extend our analysis to account for the dynamic relationship between an agent that chooses actions and a principal that can, at the end of the interaction, penalize or reward the agent based on the vector of chosen actions at every point in time.

For simplicity, we start by considering a dynamic environment with 2 periods. At each period $n = 1, 2$ (i) The agent privately observes θ_n, (ii) The agent takes action a_n from some actions set A_n, and (iii) a public outcome z_n is realized from a set Z_n, which can be observed by both the agent and principal.

We assume that the distribution of θ_2 is independent of θ_1, conditional on a_1, that is the action in the first period can affect the state in period 2, but not the state in period 1. Let $Z = Z_1 \times Z_2$. There is a set $Z^C \subseteq Z$ of contractible outcomes. After period 2 ends, if the outcome $z \in Z$ is contractible, the principal taxes the agent, choosing a transfer $t \in \mathbb{R}$. The final payoff of the agent is given by

$$t - c_1(\theta_1, a_1) - c_2(\theta_2, a_2)$$

The principal uses the tax to influence the action of the agent. Formally, the principal commits to a *contract* $g : \Theta \times Z \to \mathbb{R}$. At each period after learning θ_n and before taking the action, the agent reports the state $\hat{\theta_n} \in \Theta_n$ and gets an action recommendation $\hat{a}_n \in A_n$.

The tax that the agent pays depends on the vector of reports $\hat{\theta}$ and the vector of outcomes z, according to the contract g.

Fixing any contract g, the agent faces a dynamic decision problem. A (deterministic) dynamic social choice function $f = (f_1, f_2)$ consists of a maps $f_1 : \Theta_1 \to A_1$ and $f_2 : \Theta_1 \times \Theta_2 \to A_2$. f is implemented by a contract g if following actions prescribed by f is sequentially rational given the penalty function g.

A tax schedule is a map $t : A_1 \times A_2 \to \mathbb{R}$ which alternatively, can be interpreted as a contract that is constant in the reports of the agent.

Definition 7 *We say that actions are contractible if $Z_n = A_n$ and the distribution of outcomes is degenerate, taking $z_n = a_n$ with probability one.*

Intuitively, when actions are contractible the principal can directly observe a_1 and a_2, and set a transfer accordingly. When actions are contractible the following version of the taxation principle holds.

Proposition 4 ***(Dynamic taxation principle)****: If actions are contractible, any implementable social choice function f can be implemented with a tax mechanism.*

Proof. In Appendix D.

The proof works by backward induction. Intuitively, for any first-stage action and report, the problem boils down to a single-stage framework, for which we can directly apply Theorem 1 to obtain, for any implementable social choice function, a contract that implements it and that is constant in the second report. The main challenge arises with the first report, as our static result cannot be applied directly. After a careful mapping of the first-stage problem of the agent to a static setting, where the payoff function of the agent and outcomes have to be redefined, the Taxation Principle can be applied to finally construct a tax schedule that implements f.

Extension to $N > 2$ stages. The logic from the construction in 4 can be extended beyond 2 stages, as long as the process that determines the state at each stage θ_n depends on the history of actions, but is independent of the history of states.

7 Conclusion

In certain environments, the principal cannot directly condition rewards and penalties on the action chosen by the agent, and instead has to rely on imperfect measures. Designing rewards in this context, the principal may benefit from communicating with the agent before actions are taken, for example by making the agent write a report. We study conditions under which report-independent penalty functions are without loss of generality and show that these conditions hold in settings such as plea bargaining, and the design of liability transfers, where the impossibility of communication before actions take place is a plausible assumption. For static settings, the weakest necessary condition that guarantees this is that the principal must always be able to identify ex-post, and based on the realized outcome, the distribution of contractible outcomes associated with the action taken.

Although these conditions are quite strong, it is possible to determine whether an implementable social choice function can be implemented with report-independent incentives by looking at a similar condition on the equilibrium path. This result is useful in situations in which, given the objective of the principal, there is only a subset of the social choice function that makes sense to consider. Finally, we use our static results to extend our analysis to dynamic settings and show that communication is redundant when actions are contractible, even when the distribution of private information depends on previous periods' actions.

A Omitted Proofs

Proof of Observation 1

Proof.

$$\begin{aligned}\nu \succeq_{\{\theta',a'\}} \nu' &\Leftrightarrow E_\nu[v(\theta,a,z)] + q(\theta,a)E_\nu[d(t,z)] \geq E_{\nu'}[v(\theta,a,z)] + q(\theta,a)E_{\nu'}[d(t,z)] \\ &\Leftrightarrow E_\nu[d(t,z)] - E_{\nu'}[d(t,z)] \geq \frac{1}{q(\theta,a)}\underbrace{[E_{\nu'}[v(\theta,a,z)] - E_\nu[v(\theta,a,z)]]}_{=0} \\ &\Leftrightarrow E_\nu[d(t,z)] - E_{\nu'}[d(t,z)] \geq 0\end{aligned}$$

Proof of Observation 2

Proof. Let $\tau_{\hat{\theta}}$ be a direct mechanism, an agent with type θ who chooses action a and report m gets expected utility

$$E[u(\theta,a,\emptyset,z)|\theta,a,z \in Z \setminus C] + \Pr(z \in C|\theta,a) \cdot E_{\mu_a}\left[u(\theta,a,\tau_{\hat{\theta}}(z),z)\right].$$

Notice that the report only shows up on the last term, so the agent chooses the report that maximizes $E_{\mu_a}\left[u(\theta,a,\tau_{\hat{\theta}}(z),z)\right]$. Moreover, fixing the action a, for two different reports $\theta,\theta' \in \Theta$, the joint distributions of outcomes and penalties have the same marginal in the outcomes. By the previous observation, the ranking over distributions, and thus the optimal report, is independent of the agent's true type.

Proof of Proposition 1

To prove Proposition 1, we first apply the Revelation Principle which states that any implementable social choice function can be implemented with a direct mechanism that is incentive-compatible.

Lemma 2 *Let the direct mechanism $\{\tau_\theta\}$ implement f. Then, there is a map $r : A \to \Theta$ such that the tax $\hat{\tau}(a) = \tau_{r(a)}(a)$ implements f.*

Proof. For each $a \in A$, let

$$r(a) := \arg\min_{\theta \in \Theta} \tau_\theta(a)$$

By construction, we have for all $a \in A$ and $\theta \in \Theta$,

$$v(\theta, f(\theta)) - \hat{\tau}(f(\theta)) = v(\theta, f(\theta)) - \tau_{f(\theta)}(f(\theta)) \geq v(\theta, f(\theta)) - \tau_{\hat{\theta}}(f(\theta))$$

The first inequality holds because, by definition, $\hat{\tau}(a) = \tau_{r(a)}(a)$ for all $a \in A$. This implies that

$$v(\theta, f(\theta)) - \hat{\tau}(f(\theta)) \geq sup_{\theta \in \Theta} v(\theta, a) - \tau_\theta(a) = v(\theta, a) - \hat{\tau}(a) \qquad \forall a \in A. \qquad \blacksquare$$

B Proof of Theorem 1

We separate the proof of Theorem 1 in two parts. First, we show the key direction, namely that when A is invariant, it is possible to implement any implementable social choice function with a tax mechanism.

Lemma 3 *Let f be implementable by an incentive-compatible direct mechanism represented by $t : \Theta \times Z \to \mathbb{R}$. If independence holds and $\mathbf{A}$ is invariant, there exists a map $r : Z \to \Theta$ such that:*

- *for any action $a \in A$ and z, z' in the support of μ_a, $r(z) = r(z')$.*
- *the tariff $\hat{t}(z) := t(r(z), z)$ implements f.*

Moreover, the expected utility of each type under t and $\hat{t}$ is the same.

Proof. To construct $r(z)$, consider a cell A of $\mathbf{A}$ and let $Z_A \subseteq Z$ be the support of μ_a for every action $a \in A$. If $f(\Theta) \cap A$ is non-empty, select a type θ such that $f(\theta) \in A$ and set $r(z) = \theta$ for every $z \in Z_A$. If instead $f(\Theta) \cap A$ is empty, select an arbitrary type θ and set $r(z) = \theta$ for every outcome $z \in Z_A$.

We show next that, given our construction of $r(z)$, the tax mechanism $\hat{t}(z) := t(r(z), z)$ implements f. Since t is an incentive-compatible direct revelation mechanism that implements f, it must be all IC constraints are satisfied, meaning that

$$E_{z\sim\mu_{f(\theta)}}[v(\theta, f(\theta), z) - q(\theta, f(\theta)) \cdot d(\theta, z)] \geq E_{z\sim\mu_a}[v(\theta, a, z) - q(\theta, a) \cdot d(\hat{\theta}, z)] \qquad \forall\, \theta, \hat{\theta} \in \Theta \qquad \forall\, a \in A$$

By setting $a = f(\theta)$, we obtain that

$$E_{z\sim\mu_{f(\theta)}}[d(\theta, z)] \leq E_{z\sim\mu_{f(\theta)}}[d(\hat{\theta}, z)] \qquad \forall \hat{\theta} \in \Theta \tag{1}$$

Fix a type $\theta \in \Theta$ and an action $a \in A$. Let $\bar{\theta}$ be the type associated with the cell of $\mathbf{A}$ that contains the action $f(\theta)$, and θ' the type associated with the cell of $\mathbf{A}$ that contains action a, i.e., for every z in the support of $\mu_{f(\theta)}$, $r(z) = \bar{\theta}$ and for every z in the support of μ_a, $r(z) = \theta'$. Moreover, since $f(\bar{\theta})$ is in the same cell of $\mathbf{A}$ as $f(\theta)$ and because $\mathbf{A} = \mathbf{B}$, it must be that $\mu_{f(\bar{\theta})} = \mu_{f(\theta)}$. Let denote this distribution by μ.

Thus,

$$\begin{aligned}
E_{z\sim\mu}[v(\theta, f(\theta), z) - q(\theta, f(\theta)) \cdot d(r(z), z)] &= E_{z\sim\mu}[v(\theta, f(\theta), z) - q(\theta, f(\theta)) \cdot d(\bar{\theta}, z)] \\
&\geq E_{z\sim\mu}[v(\theta, f(\theta), z) - q(\theta, f(\theta)) \cdot d(\theta, z)] \\
&\geq E_{z\sim\mu_a}[v(\theta, a, z) - q(\theta, a) \cdot d(\theta', z)] \\
&= E_{z\sim\mu_a}[v(\theta, a, z) - q(\theta, a) \cdot d(r(z), z)]
\end{aligned}$$

Where the first equality holds since $r(z) = \bar{\theta}$ for all z in the support of μ. The first inequality is just an application of Eq. 1. The second inequality holds because of the incentive compatibility of the direct revelation mechanism. The final equality holds because the $r(z) = \theta'$ for all elements in the support of μ_a.

Proof of Proposition 2

Proof. The proof proceeds by construction. Since $\mathbf{A} \neq \mathbf{B}$, there must be a pair of actions $a_1, a_2 \in A$ with $Z(a_1) \cap Z(a_2) \neq \emptyset$ and $\mu_{a_1} \neq \mu_{a_2}$. Let $z_0 \in Z(a_1) \cap Z(a_2)$. Moreover, we can find two outcomes z_1, z_2 such that $\mu_{a_1}(z_1) > \mu_{a_2}(z_1)$ and $\mu_{a_1}(z_2) < \mu_{a_2}(z_2)$. Assume without loss that $\mu_{a_2}(z_0) \geq \mu_{a_1}(z_0)$.

Consider $\Theta = \{\theta_0, \theta_1, \theta_2\}$. We can make a_1 and a_2 the only relevant actions by setting the payoff for any other action $-\infty$ for all types. Similarly, we can make all outcomes z other than z_1 and z_2 irrelevant by setting $\Gamma(z) = \{\emptyset\}$.

Let $\Gamma(z) = \{H, L\}$ for $z \in \{z_0, z_1, z_2\}$ and u be as follows.

$$
\begin{aligned}
u(\theta_1, a, z, g) &= 1_{\{a=a_2\}}[\mu_{a_1}(z_1) + \mu_{a_1}(z_0)] + 1_{\{g=H\}}1_{\{z \in z_0 \cup z_1\}} \\
u(\theta_2, a, z, g) &= 1_{\{a=a_1\}}\mu_{a_2}(z_2) + 1_{\{g=H\}}1_{\{z \in z_2\}} \\
u(\theta_0, a, z, g) &= 1_{\{a=a_1\}}K_0 + 1_{\{g=H\}}1_{\{z \in z_0 \cup z_2\}}
\end{aligned}
$$

Consider the social choice function given by $f(\theta_1) = a_1$, $f(\theta_2) = a_2$, and $f(\theta_0) = a_1$.

Note that for f to be implementable, it must be that the outcome is $g = H$ when type is θ_1 and outcomes are z_0 or z_1, and when type is θ_2 and the outcome is z_2. Consider g given by $g(\theta_1, z_0) = g(\theta_1, z_1) = g(\theta_2, z_2) = H$ and $g(\theta, z) = L$ otherwise. This g truthfully implements f iff

$$
K_0 + \max\{\mu_{a_1}(z_0), \mu_{a_1}(z_2)\} \geq \max\{\mu_{a_2}(z_0), \mu_{a_2}(z_2)\}
$$

For tariff implementability, however, H must be assigned with probability 1 for outcomes z_0, z_1 and z_2. Thus, for IC of type θ_0, it must be that

$$
K_0 + \mu_{a_1}(z_0) + \mu_{a_1}(z_2) \geq \mu_{a_2}(z_0) + \mu_{a_2}(z_2)
$$

So, for all $K_0 \in (\max\{\mu_{a_2}(z_0), \mu_{a_2}(z_2)\} - \max\{\mu_{a_1}(z_0), \mu_{a_1}(z_2)\}, \mu_{a_2}(z_0) + \mu_{a_2}(z_2) - \mu_{a_1}(z_0) - \mu_{a_1}(z_2))$, f is implementable but not tariff implementable. We now show that this interval is non-empty.

$$
\begin{aligned}
\max\{\mu_{a_2}(z_0), \mu_{a_2}(z_2)\} - \max\{\mu_{a_1}(z_0), \mu_{a_1}(z_2)\} - \mu_{a_2}(z_0) - \mu_{a_2}(z_2) + \mu_{a_1}(z_0) + \mu_{a_1}(z_2)) = \\
\max\{\mu_{a_2}(z_0), \mu_{a_2}(z_2)\} - \mu_{a_2}(z_0) - \mu_{a_2}(z_2) - [\max\{\mu_{a_1}(z_0), \mu_{a_1}(z_2)\} - \mu_{a_1}(z_0) - \mu_{a_1}(z_2))] = \\
-\min\{\mu_{a_2}(z_0), \mu_{a_2}(z_2)\} - [-\min\{\mu_{a_1}(z_0), \mu_{a_1}(z_2)\}] = \\
\min\{\mu_{a_1}(z_0), \mu_{a_1}(z_2)\} - \min\{\mu_{a_2}(z_0), \mu_{a_2}(z_2)\} < 0
\end{aligned}
$$

Where the last inequality holds because $\mu_{a_2}(z_0) \geq \mu_{a_1}(z_0)$ and $\mu_{a_2}(z_2) > \mu_{a_1}(z_2)$.

C Proof of Theorem 2

Proof. Let f be implementable with a direct mechanism represented by $t(\theta, z)$ let $\mathcal{A}$ be an f-invariant partition of A. We will construct a tariff $\hat{t}$ that implements f by constructing a map $r : Z \to \Theta$, as we did in the proof of Theorem 1.

Consider a contractible outcome $z \in C$. By definition, we know that all actions that could have generated z belong to the same element of the partition A'. Suppose that no type chooses an action in A' according to f and θ_0 be an arbitrary type and consider replacing $\{\tau_\theta\}$ with $\{\tau'_\theta\}$ such that

$$
\tau'_\theta(z) = \begin{cases} \tau_\theta(z) & \text{if } z \notin Z_{A'} \\ \tau_{\theta_0}(z) & \text{if } z \in Z_{A'} \end{cases}
$$

This change does not affect the incentives for the agents to take their prescribed actions. Let $U^\theta(a; \tau))$ denote the expected utility of type θ, who takes action a and faces a tax τ. First, notice that

$$
U^\theta(f(\theta); \tau'_\theta) = U^\theta(f(\theta); \tau_\theta) \geq U^\theta(a; \tau_{\tilde{\theta}}) \qquad \text{for all } \theta, \tilde{\theta} \in \Theta \text{ and } a \in A. \tag{2}
$$

The first equality holds because given action $f(\theta)$ the probability of generating an outcome in $Z_{A'}$, where there would be a difference between τ'_θ and τ_θ, is zero. The inequality is the incentive compatibility constraint since $\{\tau_\theta\}$ truthfully implements f.

For $a \notin A'$, $U^{\theta}(a; g_{\tilde{\theta}}) = U^{\theta}(a; g'_{\tilde{\theta}})$ since no outcomes in $Z_{A'}$ are generated. For $a \in A'$, we can replace $\tilde{\theta}$ with θ_0 in eq. (2) and we have that $U^{\theta}(a, \tau_{\theta_0}) = U^{\theta}(a, \tau'_{\tilde{\theta}})$ by construction of τ'.

Thus, we have that $U^{\theta}(f(\theta); \tau'_{\theta}) \geq U^{\theta}(a; \tau'_{\tilde{\theta}})$ for all $\theta, \tilde{\theta} \in \Theta$ and $a \in A$. Each type prefers to continue reporting truthfully and taking the prescribed action.

Suppose instead that there is a non-empty set of types $\Theta' \subseteq \Theta$ such that $f(\theta) \in A'$ for all $\theta \in \Theta'$. Consider an arbitrary element θ' from this set. Replacing $\{\tau_{\theta}\}$ with $\{\tau'_{\theta}\}$ such that

$\tau'_{\theta}(z) = \begin{cases} \tau_{\theta}(z) & \text{if } z \notin Z_{A'} \\ \tau_{\theta'}(z) & \text{if } z \in Z_{A'} \end{cases}$ does not affect incentives:

$$U^{\theta}(f(\theta); \tau'_{\theta}) = U^{\theta}(f(\theta); \tau_{\theta}) \tag{3}$$

This is true for types outside Θ' because, as before, given action $f(\theta)$ the possibility of generating an outcome in $Z_{A'}$ is zero. For types in Θ', however, the reason is that, otherwise, $\{\tau_{\theta}\}$ would not truthfully implement f: since f is observably injective (part (ii)) $f(\theta)$ and $f(\theta')$ have the same distributions of contractible outcomes. If $U^{\theta}(f(\theta); \tau_{\theta'}) > U^{\theta}(f(\theta); \tau_{\theta})$ then type θ would find it optimal to report θ' instead of the truth. Combining Equation (3) with incentive compatibility we get:

$$U^{\theta}(f(\theta); \tau'_{\theta}) \geq U^{\theta}(a; \tau_{\tilde{\theta}}) \tag{4}$$

As before, for $a \notin A'$, $U^{\theta}(a; \tau_{\tilde{\theta}}) = U^{\theta}(a; \tau'_{\tilde{\theta}})$ since no outcomes in $Z_{A'}$ are generated. For $a \in A'$, we can replace $\tilde{\theta}$ with θ' in eq. (4) and we have that $U^{\theta}(a, \tau_{\theta'}) = U^{\theta}(a, \tau'_{\tilde{\theta}})$ by construction of τ'.

If this process is performed iteratively for all elements of the partition, we are left with a family that consists of repeated copies of the same element. ■

D Proofs of Proposition 4

To prove Proposition 4, we start by first showing a result equivalent to Lemma 3 for dynamic settings. The Proposition is a direct corollary to this lemma.

Lemma 4 *Let $f = (f_1, f_2)$ be implementable by an incentive compatible direct mechanism represented by $t : \Theta_1 \times A_1 \times \Theta_2 \times A_2 \to \mathbb{R}$. If θ_2 is independent of θ_1 conditional on a_1, there are functions $r_1 : A_1 \to \Theta_1$ and $r_2 : A_2 \to \Theta_2$ such that the tax:*

$$\hat{t}(a_1, a_2) := t(r_1(a_1), a_1, r_2(a_2), a_2)$$

also implements f.

Proof. Consider a two-stage framework, and an implementable social choice function $f = (f_1, f_2)$. Let t be an incentive-compatible direct mechanism that implements f. The proof works by backward induction. First, we show that f can be implemented with a direct mechanism that is constant in the last report.

Consider the sub-game that starts in the second stage. By fixing the first stage report-action pair $(\hat{\theta}_1, a_1)$, this sub-game can be analyzed as a single-stage game, in which the cost of the first stage action is sunk, and therefore the payoff is given by

$$t - c_2(\theta_2, a_2)$$

Since $t(\hat{\theta}_1, a_1, \cdot, \cdot)$ implements $f_2(\hat{\theta}_1, \cdot)$, and a_2 is observable, we can apply Lemma 3 and obtain that there exists a map $r_2 : A_2 \to \Theta_2$ such that the second-stage tax $\bar{t}(a_2) = t(\hat{\theta}_1, a_1, r_2(a_2), a_2)$ implements $f_2(\hat{\theta}_1, \cdot)$. Moreover, since the continuation payoff is the same for every second-stage type θ_2, the incentives in the first stage are not altered, and this DRM implements f as well.

Consider now the first stage. We can now map the problem to a single-stage framework in which the type space is Θ_1, the action space is A_1, the outcome $Z_1 = A_1 \times \Theta_2$ is given by $\tilde{z}_1 = (a_1, \theta_2)$, and the payoff by

$$t - c_1(\theta_1, a_1)$$

We know that with a contract $t_1(\hat{\theta}_1, \tilde{z}_1) = \bar{t}(\hat{\theta}_1, a_1, f_2(\hat{\theta}_1, \theta_2)) - c_2(\theta_2, f_2(\hat{\theta}_1, \theta_2))$ it is optimal for the agent to report the state θ_1 truthfully and take the action prescribed by f_1.

Note that for this problem Independence holds, since the distribution of outcomes (a_1, θ_2) conditional on the action a_1 is independent of θ_1. Moreover, $\mathbf{A}$ is the partition of A_1 of all actions, as a_1 is part of z_1, and $\mathbf{B} = \mathbf{A}$, as $\mathbf{B}$ is finer than $\mathbf{A}$ and there is no finer partition than $\mathbf{A}$.

Applying again Lemma 3, there exists a map $\hat{r}_1 : Z_1 \to \Theta_1$ such that (i) the tax $\bar{t}_1(z_1) = t_1(\hat{r}_1(z_1), z_1)$ implements f_1, and (ii) $\hat{r}_1(a_1, \theta_2)$ is constant in θ_2, and therefore we can write it as $r_1(a_1)$.

$$\bar{t}_1(z_1) = \bar{t}(\hat{r}_1(a_1), a_1, f(\theta_2)) - c_2(\theta_2, f(\theta_2))$$

Putting all together, the tax schedule given by $\hat{t}(a_1, a_2) := t(r_1(a_1), a_1, r_2(a_2), a_2)$ implements f. $\square$

References

Guesnerie, R. (1981). On Taxation and Incentives: Further Reflections on the Limits of Redistribution. . *Mimeo, Ecole des Hautes Etudes en Sciences Sociales, Paris.*

Hammond, P. (1979). Sraightforward Individual Incentive Compatibility in Large Economics. *Review of Economic Studies*, 46(2):263–282.

Kocherlakota, N. R. (2005). Zero expected wealth taxes: A mirrlees approach to dynamic optimal taxation. *Econometrica*, 73(5):1587–1621.

Liu, Y. (2025). Dynamic reward design.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies*, 38(2):175–208.

Myerson, R. (1982). Optimal coordination mechanisms in generalized principal- agent problems. *Journal of Mathematical Economics.*

Perez-Richet, E. and Skreta, V. (2024). Score-based mechanisms.

Poggi, F. and Strulovici, B. (2022). Liability Design with Information Acquisition. *Working Paper.*