

A Taxation Principle with Non-Contractible Events

Francisco Poggi and Bruno Strulovici

Mannheim University and Northwestern University*

November 25, 2021

Abstract

We study a principal-agent model with private information and moral hazard in which transfers, penalties, and other transactions between the principal and the agent can occur only after certain publicly observable events. For example, if an agent violates a law, penalties can be levied against the agent only if the violation is detected and the agent is apprehended. In these environments, we study when the principal can benefit from communicating with the agent *ex ante*, i.e., after the agent has learned his type but before taking any action, relative to a situation in which the principal can communicate with the agent after one of the publicly observable events has occurred. We characterize the set of environments for which all social choice functions that can be implemented *ex ante* can also be implemented when principal-agent interactions are restricted to contractible events. For environments outside of this set, we introduce a property of social choice functions, *observable injectivity*, such that any implementable social choice function satisfying this property can be implemented with the restriction.

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]) says that a

*Emails: poggi@uni-mannheim.de and b-strulovici@northwestern.edu. Strulovici gratefully acknowledges financial support from the National Science Foundation (Grant No. 1151410).

regulator or mechanism designer does not need to directly inquire about the agent’s private information, but can instead restrict attention to mechanisms that “tax” each action without affecting the set of social choice functions that can be implemented.

These taxation mechanisms are of special interest in settings for which 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*.

It is therefore valuable to understand whether a result similar to the Taxation Principle holds in such settings. In these settings, (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 private 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 outcomes.

We compare two situations: 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 it is possible to replicate any social choice function implemented by a mechanism in which the agent reports his type ex ante by a “tariff” in which the penalty depends only on the contractible outcome that occurred, if any.

We provide a necessary sufficient condition on the environment for this equivalence to hold. The condition requires that each possible contractible outcome reveals the distribution of

outcomes that was generated by the action chosen by the agent. The condition is necessary in the sense that, when it fails to hold, there are social choice functions that can be implemented with an ex ante mechanism but not with a tariff.

The Taxation Principle is sometimes unnecessarily demanding: it requires that *every* implementable social choice function (i.e., maps from agent types to actions) be implementable by a tariff. In practice, however, the designer may care only about a subset of social choice functions. For example, a regulator may wish to encourage firms to launch safe products and deter the launch of dangerous ones, but the Taxation Principle requires that mechanisms that induce the firm to launch only risky products be replicated with a tariff.

We introduce a class of social choice functions, which we call “observably injective,” such that *an observably injective social choice function is implementable if and only if it is implementable by a tariff*. The property requires that each outcome identifies the distribution of outcomes *given* the actions prescribed by the social choice function. In a companion paper, we apply this result to study optimal liability policy when firms can acquire information about the riskiness of their product before deciding between launching the product and abandoning its development.

Our first set of results does not require that agents have quasilinear preferences. Rather, it applies to a class of “separable” utility functions that allow for risk aversion, type-dependent preference for money (or sentences) and many other preferences.

Our second set of results focuses on environments in which the principal and the agent have quasilinear utility. In this case, requiring that the tariff exactly replicate the distribution of penalties as the original mechanism is unnecessary since agents and principal only care about the expected penalty. In ?? we establish a necessary condition for tariff implementation, namely that the distribution of outcomes satisfies a *no-convex-combination* property.

Outline In Section 2, we introduce a benchmark setting with private information and quasilinear preferences in which the classical Taxation Principle holds. In Section 3 we extend the setting to account for moral hazard, non-contractible events, and non-quasilinear preferences. We provide the main results characterizing the set of environments for which the Taxation Principle holds. In ?? we focus again on monetary penalties to study weaker conditions that are sufficient for tariff implementation of all implementable social choice functions.

2 The Standard Taxation Principle

We start the analysis by recalling the standard Taxation Principle. There is a set A of actions (e.g., how much of a good to consume) and a set Θ of private types (e.g., how much the agent values the good). An agent of type θ who chooses action a receives quasilinear utility $u(\theta, a) - t$ where t is a transfer made to the principal.

A *mechanism* consists of a message space M and a transfer policy $t : M \times A \rightarrow \mathbb{R}$, such that when the agent reports a message $m \in M$ and choose an action $a \in A$, it must pay a transfer $t_m(a)$ to the principal.

A mechanism (M, t) is a *tariff mechanism* if $t_m(a)$ is independent of m for all $a \in A$. Equivalently, a tariff mechanism is a mechanism for which M is a singleton.

A *social choice function* $f : \Theta \rightarrow A$ is *implementable* if there exists a mechanism under which each type θ finds it optimal to take action $f(\theta)$.

The Taxation Principle says that every implementable social choice function can be implemented with a tariff mechanism. According to this principle, it is without loss of generality to focus on tariff mechanisms for design purposes. The result is even stronger in the sense that for all implementable f and implementing mechanisms, there is a tariff mechanism that implements f and such that *every type receives the same transfer as in the original mechanism*.

For completeness, we state and prove the principle.

Proposition 1 (Taxation Principle) *For every implementable social choice function f and mechanism (M, t) that implements f with associated report function m , there is a tariff \hat{t} that implements f and such that $\hat{t}(f(\theta)) = t_{m(\theta)}(f(\theta))$.*

Proof. Suppose that f is implementable and consider any mechanism (M, t) that implements f . For each $a \in A$, let

$$\hat{t}(a) = \inf_{m \in M} t(m, a).$$

By construction, we have for all $a \in A$ and $m \in M$

$$u(\theta, f(\theta)) - \hat{t}(f(\theta)) \geq u(\theta, f(\theta)) - t(m(\theta), f(\theta)) \geq u(\theta, a) - t(m, a).$$

The first inequality holds because $\hat{t}(a) \leq t(m(\theta), a)$ for all $a \in A$ and the second inequality

holds by incentive compatibility of the mechanism (M, t) for type θ . This implies that

$$u(\theta, f(\theta)) - \hat{t}(f(\theta)) \geq \sup_{m \in M} u(\theta, a) - t(m, a) = u(\theta, a) - \hat{t}(a) \quad \forall a \in A. \quad \blacksquare$$

This result 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 remaining sections, we study conditions under which the Taxation Principle holds when these assumptions are relaxed. The present setting is enriched in three ways: (i) We add 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 introduce constraints describing how the set of penalties available to the social planner depends on the ex-post observed outcome. In particular, for certain outcomes no penalty can be imposed. For others, the set of available penalties might depend on the specific outcome; (iii) For the first part of our analysis, 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. In the second part of our analysis (??), we restrict attention to quasilinear preferences and derive general conditions for the Taxation Principle.

3 General Setting

An agent chooses an action from a set A . Each action generates a potentially random outcome $z \in Z$ that can be *contractible* ($z \in Y$) or not ($z \in Z \setminus Y$). Let $\mu_a \in \Delta(Y)$ be the distribution of contractible outcomes given the action a , i.e., the distribution of z conditional a and on the event $z \in Y$. We will refer to $(A, Z, Y, \{\mu_a\}_{a \in A})$ as the *environment*.

The agent is an expected-utility maximizer with private type $\theta \in \Theta$. Fixing an action a , and conditional on the outcome being contractible, the distribution of outcomes is given by μ_a that is independent of θ .¹

A designer wishes to induce type-dependent actions. To do so, the designer chooses, after

¹The fact that the distribution of contractible outcomes is independent of the type holds, for instance, if θ is a preference parameter of the agent that does not affect outcomes, or if θ affects the probability of being caught (i.e., generates a contractible outcome), but not the evidence conditional on being caught.

observing outcome z , a penalty g from a set of available penalties $\Gamma(z)$.² Let $\Gamma = \cup_{z \in Z} \Gamma(z)$ denote the set of all penalties. If the outcome z is non-contractible, we impose that $\Gamma(z)$ be a singleton, which means that the principal cannot choose a penalty for the agent. For simplicity we assume that in this case the penalty is set to a fixed value and let $g = 0$ represent the situation in which there is no punishment.

Suppose first that the designer has perfect commitment power and can contract with the agent after the agent has observed his type but before he takes his action. Without loss of generality, the designer can restrict attention to mechanisms in which:

1. The agent makes a report m from a space M .
2. The mechanism recommends an action to the agent \tilde{a} .
3. The agent privately chooses an action a .
4. An outcome z is realized.
5. The principal chooses a penalty $g \in \Gamma(z)$.³

A mechanism may usefully be described by a *penalty menu*, which consists of a message space M and a family of penalty maps $g_m : Z \rightarrow \Gamma$ that satisfy the feasibility condition $g_m(z) \in \Gamma(z)$ for all outcomes and messages. Notice that the menu allows duplicate maps: there can be multiple messages that entail the same penalty map: $g_m = g_{m'}$ for some messages $m \neq m'$. Let \mathcal{G} be the set of all maps from outcomes to penalties that satisfy the previous feasibility condition. A penalty menu can also be seen as choosing one of these maps for each report $\theta' \in \Theta$.

An agent of type θ and who takes action a , produces outcome z , and receives a penalty g , receives utility $u(\theta, a, z, g)$. For $\eta \in \Delta(Z \times \Gamma)$ We will write $v^\theta(a; g)$ for the expected utility of a type θ that takes action a and faces a penalty map g .

A social choice function f is a map $f : \Theta \rightarrow A$. We say that f is:

²Available 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 constrained by maximal sentences that depend on the verdict of the jury.

³The principal could in principle randomize among feasible penalties, in which case $\Gamma(z)$ would be equal to $\Delta(G(z))$ for some set of available penalties $G(z)$.

- *implementable* if there is a penalty menu $(M, \{g_m\}_{m \in M})$ such that all types find it incentive compatible to take the prescribed action. Formally, f is implementable if there exists a penalty menu such that

$$f(\theta) \in \arg \max_{a \in A} \max_{m \in M} v^\theta(a; g_m)$$

- *truthfully implementable* if $M = \Theta$ and each type finds optimal to report his own type truthfully;

$$(\theta, f(\theta)) \in \arg \max_{(\theta', a) \in \Theta \times A} v^\theta(a; \mu_{\theta'});$$

- *tariff implementable* if the penalty menu is such is such that the map from outcomes to distribution of penalties is independent of the agent's report. We call such a contractible menu a *tariff*. Formally, there exists a feasible penalty map $g : Z \rightarrow \Gamma$ such that

$$f(\theta) \in \arg \max_{a \in A} v^\theta(a; g).$$

We start with a straightforward observation:

Lemma 1 *If f is implementable, then it is truthfully implementable.*

Proof. This result follows from the Revelation Principle. Consider a penalty menu $(M, \{g_m\})$ and report strategy $m : \Theta \rightarrow M$ that implements f , and replacing this with the penalty menu (Θ, \hat{g}_θ) where $\hat{g}_\theta(z) = g_{m(\theta)}(z)$ for all $z \in Z$. With this new penalty menu, truth-telling is optimal for all types. ■

From here on, we focus without loss of generality on truthful implementation. A *direct* penalty menu is one in which $M = \Theta$.

Our objective is to determine conditions under which implementable social choice functions are tariff implementable. The planner might care, in addition to actions, about the joint distribution of outcomes and penalties. Thus, we also wish to determine under what conditions implementable social choice functions can be implemented with a tariff in a way that the on-path distribution of outcomes *and* penalties remains unchanged.

Assumption 1 (Separable Preferences) *There exist functions u_1 , u_2 and h such that*

$$u(\theta, a, z, g) = u_1(\theta, a, z) - h(\theta, a)u_2(g, z)$$

Separability is useful because it implies the following key property: conditional on the agent's action, all types have the same preference ranking over distributions of penalties. Formally, given two joint distributions $\nu, \nu' \in \Delta(Z \times \Gamma)$ over outcomes and penalties, the preferences of an agent with type θ and taking action a are given by

$$\nu \succeq_{\{\theta, a\}} \nu' \text{ iff } E_\nu(u(\theta, a, g, z)) \geq E_{\nu'}(u(\theta, a, g, z)).$$

Separability implies that these preferences are in fact independent of θ , as we prove next.

Observation 1 *Let $\nu, \nu' \in \Delta(Z \times \Gamma)$ denote two joint distributions that have the same marginal distribution $\mu \in \Delta(Z)$ over outcomes. If the agent has separable preferences, then $\nu \succeq_{\{\theta, a\}} \nu' \Leftrightarrow \nu \succeq_{\{\theta', a\}} \nu'$ for all $(\theta, \theta') \in \Theta^2$.*

Proof.

$$\begin{aligned} \nu \succeq_{\{\theta, a\}} \nu' &\Leftrightarrow E_\nu[u_1(\theta, a, z)] + h(\theta, a)E_\nu[u_2(g, z)] \geq E_{\nu'}[u_1(\theta, a, z)] + h(\theta, a)E_{\nu'}[u_2(g, z)] \\ &\Leftrightarrow E_\nu[u_2(g, z)] - E_{\nu'}[u_2(g, z)] \geq \frac{1}{h(\theta, a)} \underbrace{[E_{\nu'}[u_1(\theta, a, z)] - E_\nu[u_1(\theta, a, z)]]}_{=0} \\ &\Leftrightarrow E_\nu[u_2(g, z)] - E_{\nu'}[u_2(g, z)] \geq 0. \end{aligned}$$

■

Separability is satisfied when Γ consists of transfers and agents have quasilinear preferences in money, but it also allows for more general utility functions, for example that the cost of a sentence is higher when the agent is guilty.

Example 1 $\Theta = [0, 1]$ denotes the agent's gross benefit from committing a given crime, $a \in \{0, 1\}$ denotes the decision of whether to commit the crime, and z denotes the outcome, consisting of whether the agent is apprehended and of a signal correlated with the action. $\Gamma = [0, T]$ is a set of possible sentences. If the agent's utility take the form

$$u(\theta, a, z, g) = \theta a - (1 - \gamma a)f(g)$$

this satisfies Assumption 1 for all parameters $\gamma \in [0, 1]$ and for every $f : \Gamma \rightarrow \mathbb{R}$. □

Observation 2 *If preferences are separable, the optimal report conditional on the action a is independent of the agent's type.*

Proof. Given a penalty menu $(M, \{g_m\})$, an agent with type θ who chooses action a and report m gets expected utility

$$E[u(\theta, a, z, 0)|\theta, a, z \in Z \setminus Y] + \Pr(Y|\theta, a) \cdot E_{\mu_a}[u(\theta, a, z, g_m(z))].$$

The agent's report appears only in the last term. Therefore, the agent chooses a report that maximizes $E_{\mu_a}[u(\theta, a, g_m(z), z)]$. Moreover, fixing action a , for two different reports $m, m' \in M$, 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. ■

Observation 2 indicates that the agent's report cannot be very informative of the agent's true type given the action chosen by the agent. In fact, if the action is observable the agent's report is redundant (up to potential indifferences), as captured by the next result.

Proposition 2 (Observable Contractible Actions) *If $Y \subseteq A$ and $\mu_a(a) = 1$ for all $a \in Y$, then any implementable social choice function f is tariff implementable. Moreover, for any penalty menu $(M, \{g_\theta\})$ that implements f there is a tariff that implements f and gives every type the same expected utility as the penalty menu.*

Proof. We recall that $v^\theta(a; g)$ denotes the expected utility of a type θ that takes action a and faces penalty map $g : Z \rightarrow \Gamma$. Consider a direct penalty menu $\{g_\theta\}$ that truthfully implements f (by Lemma 1 there is one). And define the set $T(a) = \arg \min_{\theta' \in \Theta} E_\mu[u_2(z, g_{\theta'})]$ of reports that minimize the expected cost of penalty conditional on action a (by Observation 2 this set is independent of the true type $T(a)$). Finally, let $M(a)$ be the set of associated distribution of penalties for each action $G(a) := g(T(a), a)$. Incentive compatibility implies that

$$v^\theta(f(\theta); g_\theta(f(\theta))) \geq v^\theta(a, g_{\theta'}(a)) \quad \forall \theta, \theta' \in \Theta, \quad \forall a \in A$$

Fixing a and taking the supremum over reports yields

$$v^\theta(f(\theta); g_\theta(f(\theta))) \geq v^\theta(a; G(a))$$

Moreover, $\theta \in T(f(\theta))$, so $v^\theta(f(\theta); G(f(\theta))) \geq v^\theta(a; G(a))$. ■

When actions are not perfectly observable, however, the agent's report might be informative about the agents' true type even when preferences are separable, as the next example illustrates.

Example 2 Suppose that $A = \{a_0, a_1, a_2, a_3\}$, $Z = Y = \{z_0, z_1, z_2\}$, $\mu_{a_0}(z_0) = 1$, $\mu_{a_1}(z_1) = \mu_{a_2}(z_2) = 0.9$, $\mu_{a_1}(z_0) = \mu_{a_2}(z_0) = 0.1$, and $\mu_{a_3}(z_1) = \mu_{a_3}(z_2) = 0.5$. Also assume that $\Theta = \{\theta_1, \theta_2\}$, and that $u = 1_{\{a=a_0\}} + g$ with $\Gamma(z) = \mathbb{R}$ for all $z \in Y$.

Consider the social choice function

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

Then, f is implementable by the penalty map $g(\theta_i, z_j) = 1_{\{i=j\}} \cdot 2$. Agent's types do not strictly benefit from deviating from reporting their true type and taking the action prescribed by f .

However, f is not tariff implementable. To see this, let $g_i = \hat{g}(z_i)$ denote an arbitrary tariff. The social choice function is implementable by \hat{g} only if the following inequalities hold:

$$\begin{aligned} 0.9g_1 + 0.1g_0 &\geq g_0 + 1 \\ 0.9g_2 + 0.1g_0 &\geq g_0 + 1 \\ 0.9g_1 + 0.1g_0 &\geq 0.5(g_1 + g_2) \\ 0.9g_2 + 0.1g_0 &\geq 0.5(g_1 + g_2). \end{aligned}$$

Summing the first two inequalities and simplifying yields

$$g_1 + g_2 \geq 2g_0 + 2/0.9$$

while summing the last two inequalities and simplifying yields

$$g_1 + g_2 \leq 2g_0$$

These inequalities are incompatible, which shows that no tariff \hat{g} can implement f . \square

This example shows preference separability does not guarantee that tariffs can implement all implementable social choice functions. The next section explains what additional conditions are required to generalize Proposition 2.

We emphasize two points: First, the condition of observable contractible actions used in Proposition 2 is a condition on the environment, independent of other model primitives such as u, Γ, Θ , and that imposes no additional restriction on the distribution of outcomes beyond the conditional distribution for contractible outcomes. In this spirit, we will seek

general conditions on the environment guaranteeing the implementability of social choice functions with tariffs for all primitives.

Second, Proposition 2 provides conditions under which *all* implementable social choice functions can be implemented with tariffs. In some contexts, however, only a subset of the social choice functions is relevant for the principal. Therefore, it may also be useful to find conditions on specific social choice functions and the rest of the environment, under which the social functions are tariff implementable.

3.1 Tariff Implementability of Specific Social Choice Functions

Toward this end, we introduce a novel property of social choice functions, as follows. For any $A' \subset A$, let $Z(A')$ denote the set of contractible outcomes that can be generated by actions in A' . Formally, $Z(A')$ is union of the supports of the distributions μ_a for all $a \in A'$. Recall that by an assumption made early in our analysis (beginning of Section 3, this set is independent of the agent's type.

Definition 1 *f is observably injective if there exists a partition $\mathcal{A} = \{A_k\}_{k \in K}$ of A such that:*

- (i) $Z(A_k) \cap Z(A_{k'}) = \emptyset$ for all $k \neq k'$,
- (ii) $a, a' \in f(\Theta) \cap A_k$ implies that $\mu_a = \mu_{a'}$.

In words, f is observably injective if actions can be partitioned into cells so that conditional on observing a contractible outcome: (i) the principal can perfectly detect to which cell of the partition the action taken by the agent belongs, (ii) for each cell, there is at most one distribution of contractible outcomes associated with actions that are implemented by f and that belong to this cell. This means that, given f , any realized outcome z is a sufficient statistic for the distribution of outcomes conditional on that the outcome is observable.

The first requirement is trivially satisfied if the partition consists of single cell equal to A . In general however, finer partitions help satisfy the second requirement, which is more easily satisfied when cells are smaller. A particular instance of the first requirement is when \mathcal{A} describes the information partition of the principal, based on observable outcomes, in which case the outcome can be identified with A_k . In general however, the principal could observe finer information than A_k .

Equipped with this concept, we can state the main result of the paper.

Theorem 1 Taxation Principle with Non-Contractible Events: *If f is implementable and observably injective, then f is tariff implementable.*

Proof.

Let f be implementable with a direct penalty menu $\{g_\theta\}$ and observably injective with partition $\mathcal{A} = \{A_k\}_{k \in K}$. We will construct a tariff \hat{g} that implements f by selecting one type $\theta(z)$ for each outcome z and setting $\hat{g}(z) = g_{\theta(z)}(z)$

Consider a contractible outcome $z \in Y$. By condition (i) of observably injectivity, we know that all actions that could have generated z belong to the same element of the partition A' .

Suppose that there is no type that chooses an action in A' according to f and θ_0 be an arbitrary type and consider replacing $\{g_\theta\}$ with $\{g'_\theta\}$ such that

$$g'_\theta(z) = \begin{cases} g_\theta(z) & \text{if } z \notin Z(A') \\ g_{\theta_0}(z) & \text{if } z \in Z(A') \end{cases}$$

This change does not affect the incentives for the agents to take their prescribed actions. First, notice that

$$v^\theta(f(\theta); g'_\theta) = v^\theta(f(\theta); g_\theta) \geq v^\theta(a; g_{\tilde{\theta}}) \quad \text{for all } \theta, \tilde{\theta} \in \Theta \text{ and } a \in A. \quad (1)$$

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 g'_θ and g_θ , is zero. The inequality is the incentive compatibility constraint since $\{g_\theta\}$ truthfully implements f .

For $a \notin A'$, $v^\theta(a; g_{\tilde{\theta}}) = v^\theta(a; g'_\theta)$ since no outcomes in $Z(A')$ are generated. For $a \in A'$, we can replace $\tilde{\theta}$ with θ_0 in eq. (1) and we have that $v^\theta(a; g_{\theta_0}) = v^\theta(a; g'_\theta)$ by construction of g' .

Thus, we have that $v^\theta(f(\theta); g'_\theta) \geq v^\theta(a; g'_\theta)$ 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 nonempty 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 $\{g_\theta\}$ with $\{g'_\theta\}$ such that

$$g'_\theta(z) = \begin{cases} g_\theta(z) & \text{if } z \notin Z(A') \\ g_{\theta'}(z) & \text{if } z \in Z(A') \end{cases} \quad \text{does not affect incentives:}$$

$$v^\theta(f(\theta); g'_\theta) = v^\theta(f(\theta); g_\theta) \quad (2)$$

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, $\{g_\theta\}$ 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 $v^\theta(f(\theta); g_{\theta'}) > v^\theta(f(\theta); g_\theta)$ then type θ would find it optimal to report θ' instead of the truth. Combining Equation (2) with incentive compatibility we get:

$$v^\theta(f(\theta); g'_\theta) \geq v^\theta(a; g_{\tilde{\theta}}) \quad (3)$$

As before, for $a \notin A'$, $v^\theta(a; g_{\tilde{\theta}}) = v^\theta(a; g'_\theta)$ since no outcomes in $Z(A')$ are generated. For $a \in A'$, we can replace $\tilde{\theta}$ with θ' in eq. (3) and we have that $v^\theta(a, g_{\theta'}) = v^\theta(a, g'_\theta)$ by construction of g' .

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

Intuitively, given a truthful contractible mechanism, the conditions allow the Principal to select a reference type for each outcome in a way that he can leave each type that takes the prescribed action with the same payoff as before. Any deviations are also equivalent to a joint deviation of report and action that was also available with the truthful contractible mechanism, so cannot be optimal.

3.2 Tariff Implementability of All Social Choice Functions

Definition 2 *An environment $(A, Z, Y, \{\mu_a\}_{a \in A})$ is fully identifiable if for all $a, a' \in A$, either (i) $Z(a) \cap Z(a') = \emptyset$ or (ii) $\mu_a = \mu_{a'}$.*

In words, an environment is fully identifiable if the principal can identify the (conditional) distribution of contractible outcomes by observing the realized outcome.

Corollary 1 *If the environment is fully implementable, then all implementable f are tariff implementable.*

Proof. For arbitrary f , we just need to show that f is observably injective. Let \mathcal{A} be the partition of A according to the distribution of outcomes i.e. $a, a' \in A_k$ for some k if and only if $\mu_a = \mu_{a'}$. Then, if $a \in A_k$ and $a' \in A_{k'}$ for $k \neq k'$, it must be that $\mu_a \neq \mu_{a'}$ and thus

$Z(a) \cap Z(a') = \emptyset$. Thus, $Z(A_k) \cap Z(A_j) = \emptyset$ for all $k \neq j$. Finally, if $a, a' \in f(\Theta) \cap A_k$, it must be that $\mu_a = \mu_{a'}$, which is exactly condition (ii) in Definition 1. \blacksquare

Proposition 3 *If the environment (A, Z, Y, μ) is not fully identifiable, there exists a set Θ of types, a set $\Gamma(z)$ of penalties for each contractible outcome $z \in Y$, a utility function u and a social choice function f such that f is implementable but it is not tariff implementable.*

Proof. Since environment is not fully identifiable, there are actions $a_1, a_2 \in A$ with $Z(a_1) \cap Z(a_2) \neq \emptyset$ and $\mu_{a_1} \neq \mu_{a_2}$. Consider $\Theta = \{\theta_0, \theta_1, \theta_2\}$. And Let's make a_0, a_1 and a_2 the only relevant actions, by setting u of any other action to $-\infty$ for all types. Finally, let z_1, z_2 such that $\mu_{a_1}(z_1) > \mu_{a_2}(z_2)$ and $\mu_{a_1}(z_2) < \mu_{a_2}(z_2)$. These exist since $\mu_{a_1} \neq \mu_{a_2}$. Again, we can make all other outcomes irrelevant by setting $\Gamma(z) = \{\emptyset\}$ for all $z \notin \{z_1, z_2\}$.

Suppose $\max\{\mu_{a_1}(z_1), \mu_{a_2}(z_2)\} < 1$ (I think the other case can be easily accommodated adding some noise, but it requires non-deterministic g .) Let $\Gamma(z) = \{H, L\}$ for $z \in \{z_1, z_2\}$ and u be as follows.

$$\begin{aligned} u(\theta_1, a, z, g) &= 1_{\{a=a_0\}}K_1 + 1_{\{g=H\}}1_{\{z=z_1\}} \\ u(\theta_2, a, z, g) &= 1_{\{a=a_0\}}K_2 + 1_{\{g=H\}}1_{\{z=z_2\}} \\ u(\theta_0, a, z, g) &= 1_{\{a=a_0\}}K_0 + 1_{\{g=H\}} \end{aligned}$$

We want to implement $f(\theta_1) = a_1$, $f(\theta_2) = a_2$, and $f(\theta_0) = a_0$. Consider $g(\theta_1, z_1) = g(\theta_2, z_2) = H$ and $g(\theta_1, z_2) = g(\theta_2, z_1) = L$. f is implementable if:

$$K_1 \leq \mu_{a_1}(z_1)$$

$$K_2 \leq \mu_{a_2}(z_2)$$

$$K_0 \geq \max\{\mu_{a_1}(z_1), \mu_{a_2}(z_2)\}$$

Assume that the first two inequalities hold with equality. Two necessary conditions for tariff implementability are:

- $\Pr(H|z_1) = 1$.
- $\Pr(H|z_2) = 1$.

Thus, for IC of θ_0 ,

$$K_0 \geq 1.$$

So, for all $K_0 \in (\max\{\mu_{a_1}(z_1), \mu_{a_2}(z_2)\}, 1)$ f is implementable but not tariff implementable.

■

The conditions of Corollary 1 are satisfied when actions are ex-post observable ($Z(a) \cap Z(a') = \emptyset$ for all $a, a' \in A$) and when outcomes are completely uninformative about the action (z independent of a). However, they are also satisfied in less extreme situations. For example, if the principal observes a partition of the set of actions or other cases where there is a single observable outcome for each action. Let the action be choosing the false positive rate of a binary test with fixed false negative rates and where the outcome is contractible if and only if the true state is positive. Another example is a two-dimensional action $A = A_1 \times A_2$, where the first dimension is observable ex-post and affects the distribution of contractible outcomes and the second dimension is not observable ex-post and only affects the probability of having a contractible outcome. Another example is the application to information acquisition and product liability in the following section.

4 Application: Liability with Uncertain Product Riskiness

To illustrate Theorem 1, consider the following scenario: the agent is a firm with private information θ about the riskiness of a product. Prior to deciding whether to launch the product, the firm can acquire additional information about its riskiness. A firm's "action" thus consists in choosing (i) how to conduct the learning phase and (ii) whether to launch or to abandon the product at the end of the learning phase.

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 launched product causes damage, the regulator can observe the strength of the evidence acquired concerning the product's safety.

Combining these observations, the two components of observable injectivity emerge: (i) Conditional on damage occurrence, we can partition actions according to the strength of

evidence observed by the regulator. (ii) Different firm types (i.e., prior beliefs about product safety) should be associated with different strengths of evidence accumulated before the product's launch.

These ideas are formalized in our companion paper ([Poggi and Strulovici, 2021]), which models the learning environment as a continuous-time Wald problem, in which the agent's actions consist of a stopping time and a decision, both adapted to the filtration of a Brownian learning process.

In this environment, we show that the regulator would generally not gain from the ability to elicit the firm's prior belief about product's safety before the product is launched.

References

- [Guesnerie, 1981] Guesnerie, R. (1981). On taxation and incentives: Further reflections on the limits of redistribution. *Mimeo, Ecole des Hautes Etudes en Sciences Sociales, Paris*.
- [Hammond, 1979] Hammond, P. (1979). Straightforward individual incentive compatibility in large economies. *Review of Economic Studies*, pages 263–282.
- [Poggi and Strulovici, 2021] Poggi, F. and Strulovici, B. (2021). Liability design with information acquisition. *Working Paper*.