A Taxation Principle for Offenses

– Preliminary and Incomplete –

Francisco Poggi and Bruno Strulovici

Northwestern University*

November 12, 2020

1 Introduction

When an agent with private information and quasilinear utility chooses an action that results in a transfer, the taxation principle says that the regulator or mechanism designer does not need to directly inquire about the agent's private information, but can instead "tax" each action without incurring any loss of generality for the set of implementable actions.

Many applications, especially law-related ones, do not permit to tax each action. For example, suppose that the action of interest concerns whether to commit a crime. One would ideally like to "tax" (i.e., penalize) criminal behavior, but such taxation is possible only if the behavior is detected.

In some cases, one would like to tax the agent only if he possesses particular information. For example, a regulator may wish to encourage the launch of safe products and deter the launch of dangerous ones, but a product's potential for damage may be a firm's private information. Similarly, an action that is beneficial to the agent and harmful to a third party may still be socially desirable if the harm is small relative to the agent's benefit from the action, which is the agent's private information.

In legal settings, moreover, it is typically impossible to contract with the agent ex ante.

*Emails: franciscopoggi2015@u.northwestern.edu and b-strulovici@northwestern.edu. Strulovici gratefully acknowledges financial support from the National Science Foundation (grant 1151410)

For example, a criminal does not bargain with a prosecutor before committing a crime: bargaining occurs only after the crime was committed and only if the agent is apprehended.

This paper considers the following question: when is this inability to contract ex ante with the agent consequential? We compare two situations: one in which the agent must report his type ex ante and receives a report-contingent transfer if and when some outcome occurs, and one in which the agent does not report his type ex ante and transfer depends only on the outcome. We ask under which conditions "tariffs" that depend only on the outcome can replicate complicated schemes in which the agent report his type.

We introduce an *identifiability* property for social choice functions (i.e., maps from agent types to actions), such that *an identifiable social choice function is implementable if and only it is implementable by a tariff.* The property requires that (i) the distributions of contractible outcomes associated with actions in the range of the social choice function have disjoint support, and (ii) contractible outcomes associated with any given action outside of that range be compatible with at most one implemented action.

We apply the property 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.

2 Setting

An agent with type $\theta \in \Theta$ must choose an action a from some set A. This action generates an outcome $z \in Z$ that is either contractible ($z \in Z^c$) or not ($z \in Z \setminus Z^c$). Conditional on $z \in Z^c$, the distribution of z is assumed to be independent of θ. This assumption holds, for instance, if θ is a preference parameter of the agent that does not affect outcomes, or if θ affects the probability that the agent is caught (i.e., generates a "contractible" outcome), but not the evidence conditional on being caught. Formally, we assume that the distribution of z satisfies:

$$F(z|a, z \in Z^c, \theta) = F(z|a, z \in Z^c, \theta')$$

for all $\theta, \theta' \in \Theta$.

A designer wishes to induce specific type-dependent actions. Suppose first that the designer has perfect commitment power and can contract with the agent after the agent has observed his type and before any action is taken. Without loss of generality, the designer can restrict attention to direct revelation mechanisms, in which:

1. The agent reports his type.
2. The mechanism recommends an action to the agent.
3. The agent privately chooses an action.
4. An outcome is realized.
5. The agent receives a transfer that depends on his report and on the outcome. If the outcome is not contractible, the transfer is equal to zero.

Formally, a *contractible transfer* is a map $t : \Theta \times Z \to \mathbb{R}$ such that $t(\hat{\theta}, z) = 0$ for all $z \notin Z^c$ and report $\hat{\theta}$.

Given a contractible transfer t, an agent with type θ who chooses action a and reports $\hat{\theta}$ gets expected utility

$$u(\theta, a) + E[t(\hat{\theta}, z)|\theta, a].$$

Since z is independent of θ conditional on (Z^c, a), this expected utility is also equal to:

$$u(\theta, a) + Pr(Z^c|\theta, a)E[t(\hat{\theta}, z)|Z^c, a].$$

Given a transfer t, let

$$\tau(\hat{\theta}, a; t) = E[t(\hat{\theta}, z)|Z^c, a]. \tag{1}$$

If an agent chooses action a, this agent chooses a report $\hat{\theta}(a; t)$ that maximizes (1), regardless of the agent's actual type. This leads to the reduced-form utility

$$u(\theta, a) + P(Z^c|\theta, a)T(a; t)$$

where $T(a; t) = \max_{\hat{\theta}} E[t(\hat{\theta}, z)|Z^c, a]$.

A social choice function f is a map $f : \Theta \to A$. A social choice function f is:

- *implementable* if there exists a contractible transfer t such that for all $\theta \in \Theta$, $f(\theta)$ maximizes

  $$u(\theta, a) + P(Z^c|\theta, a)T(a; t)$$

  over $a \in A$;

- *truthfully implementable* if t can be chosen so that reporting $\hat{\theta} = \theta$ is optimal for all $\theta \in \Theta$;

- *tariff implementable* if t can be chosen so as to be independent of θ.

When a contractible transfer is independent of the agent's report, we will call it a *tariff*.

Because the designer can rarely contract with the agent ex ante, we wish to determine when implementable social choice functions are tariff implementable.

For any $A' \subset A$, let $Z(A')$ denote the *contractible consequences* of A': $Z(A')$ is the set contractible outcomes that can be generated by actions $a \in A'$. By assumption, this set is independent of the agent's type.

We start with a straightforward observation:

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

Proof. This result, which is a variation on the Revelation Principle, follows by choosing any transfer t that implements f and replacing it by the tariff $\hat{t}(\theta, z) = t(\hat{\theta}(f(\theta); t), z)$ for all $z \in Z(f(\theta))$ and $\hat{t}(\theta, z) = -M$ otherwise, where $-M$ is a lower bound on transfers is such a lower bound is imposed, and arbitrarily negative otherwise. With this new transfer $\hat{t}$, truthtelling is optimal for all types. ■

DEFINITION 1 *f is identifiable if there exists a partition $\mathcal{A} = \{A_k\}_{k=1}^K$ of A such that:*

(i) $Z(A_k) \cap Z(A_{k'}) = \emptyset$ for all $k \neq k'$.

(ii) $f(\Theta) \cap A_k$ has at most one element.

In words, f is identifiable if actions can be grouped so that (i) the principal can perfectly detect which group the action taken by the agent belongs to, (ii) each group contains at most one action in the range of f.

The first property imposes some structure on the environment, which is independent of f. For example, it rules out situations in which all actions lead to a full support over outcomes. One particular case is when $\mathcal{A}$ is the information partition of the principal, in which case the outcome can be identified with A_k. In general the principal could observe finer information than A_k. The set of partitions that satisfy property (i) is a primitive of the environment.

The second property says that one can choose such a partition, which is specific to f, so that f implements at most one action in each cell of this partition. The second property imposes a hierarchy over actions: within each action cell that is identifiable by the principal, there is one action that is singled out by the social choice function f.

THEOREM 1 *If f is implementable and identifiable, then it is tariff implementable.*

Proof. Suppose that t implements f. From Lemma 1, we can assume without loss that t is truthful. Let A_f denote the set of actions in the range of f and $Z_f = Z(A_f)$ denote the set of contractible outcomes that may arise when f is implemented. Since f is identifiable, an outcome $z \in A_f$ is only generated only actions that all lie in the same cell, A_k say, of the partition $\mathcal{A}$. Moreover, the types set Θ_k of who choose an action in A_k given transfer t all choose the same action a_k, and all these types $\theta \in \Theta_k$ must be indifferent between reporting their own type and any other type in Θ_k since, from Equation (1), these types have the same reporting incentives conditional on taking action a_k. Finally, by truthfulness, these types all these reports to any other report $\theta' \notin \Theta_k$.

Let θ_k denote an arbitrary element of Θ_k and let θ_0 denote an arbitrary element of Θ. We define a tariff $\tilde{t}$ as follows:

$$\tilde{t}(z) = \begin{cases} t(\theta_k, z) & \text{for all } z \in Z_f \\ t(\theta_0, z) & \text{for all } z \in Z \setminus Z_f \end{cases}$$

By construction, $\tilde{t}$ is a contractible transfer, and it is a tariff (i.e., independent of any report). Also by construction, truthtelling is optimal and implements the same function f. ■

3 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 the product. A firm's "action" thus consists in choosing (i) how to conduct the learning phase and (ii) whether to launch or 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 identifiability emerge: (i) Conditional on a damage, 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, 2020), 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.