Designing Incentives for Boolean Games

Ulle Endriss* Sarit Kraus+ Jérôme Lang† Michael Wooldridge‡

*University of Amsterdam, The Netherlands (ulle.endriss@uva.nl)
+Bar Ilan University, Israel (sarit@cs.biu.ac.il)
†Université Paris-Dauphine, France (lang@irit.fr)
‡University of Liverpool, United Kingdom (mjw@liv.ac.uk)

ABSTRACT

Boolean games are a natural, compact, and expressive class of logic-based games, in which each player exercises unique control over some set of Boolean variables, and has some logical goal formula that it desires to be achieved. A player's strategy set is the set of all possible valuations that may be made to its variables. A player's goal formula may contain variables controlled by other agents, and in this case, it must reason strategically about how best to assign values to its variables. In the present paper, we consider the possibility of overlaying Boolean games with *taxation schemes*. A taxation scheme imposes a cost on every possible assignment an agent can make. By designing a taxation scheme appropriately, it is possible to perturb the preferences of the agents within a society, so that agents are rationally incentivised to choose some socially desirable equilibrium that would not otherwise be chosen, or incentivised to rule out some socially undesirable equilibria. After formally presenting the model, we explore some issues surrounding it (e.g., the complexity of finding a taxation scheme that implements some socially desirable outcome), and then discuss possible desirable properties of taxation schemes.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems; I.2.4 [**Knowledge representation formalisms and methods**]

General Terms

Theory

Keywords

boolean games, incentives, taxation

1. INTRODUCTION

The computational aspects of game-theoretic mechanism design have received a great deal of attention over the past decade [11]. Particular attention has been paid to the Vickrey-Clarke-Groves (VCG) mechanism, which can be used to incentivise rational agents to truthfully report their private preferences in settings such as combinatorial auctions [5, 10]. The key point of interest of the VCG mechanism is that, because it incentivises agents to report their preferences truthfully, it allows us to compute outcomes that maximise social welfare, which would not in general be possible if agents could benefit from misrepresenting their preferences.

Ultimately, the VCG mechanism is a *taxation scheme*. Taxation schemes are used in human societies for several purposes. First, they are used to incentivise certain socially desirable behaviours, in much the same way that the VCG mechanism is used – for example, a government may tax car driving to encourage the use of environmentally friendly public transport. Second, they are used to raise revenue, typically with the intention that this revenue is then used to fund socially desirable projects (education, healthcare, etc). And finally, of course, they may be used for a combination of these purposes. Our aim in the present paper is to study the design of taxation schemes for incentivising behaviours in multi-agent systems. It is important to note that our focus in the present paper is *not* on the design of incentive compatible (truth-telling) mechanisms, and in this key respect, our work differs from the large body of work on computational and algorithmic mechanism design [11, 5, 10]. Of course, this is not to say that incentive compatibility is not important, or in any way to diminish the significance and value of the VCG mechanism; we are simply focussing on scenarios in which the preferences and actions of agents are known.

The setting for our study is the domain of *Boolean games* [6, 2, 4]. Boolean games are a natural, expressive, and compact class of games, based on propositional logic. Boolean games were introduced in [6], and their computational and logical properties have subsequently been studied by several researchers [2, 4]. In such a game, each agent i is assumed to have a goal, represented as a propositional formula γ_i over some set of variables Φ. In addition, each agent i is allocated some subset Φ_i of the variables Φ, with the idea being that the variables Φ_i are under the unique control of agent i. The choices, or strategies, available to i correspond to all the possible allocations of truth or falsity to the variables Φ_i. An agent will try to choose an allocation so as to satisfy its goal γ_i. Strategic concerns arise because whether i's goal is in fact satisfied will depend on the choices made by others.

In the present paper, we introduce the idea of imposing *taxation schemes* on Boolean games, so that various possible choices are taxed in different ways. Taxation schemes are designed by an agent external to the system known as the *principal*. The ability to impose taxation schemes enables the principal to *perturb the preferences of the players in certain ways*: all other things being equal, an agent will prefer to make a choice that minimises taxes. As discussed above, the principal is assumed to be introducing a taxation scheme so as to incentivise agents to achieve a certain socially desirable outcome; or to incentivise agents to rule out certain socially undesirable outcomes. We represent the outcome that the principal desires to achieve via a propositional formula Υ: thus, the idea is

Cite as: Designing Incentives for Boolean Games, U. Endriss, S. Kraus, J. Lang, and M. Wooldridge, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. XXX-XXX.

that the principal will impose a taxation scheme so that agents are rationally incentivised to make individual choices so as to collectively satisfy Υ. However, a fundamentally important assumption in what follows is that taxes do not give us absolute control over an agent's preferences. To assume that we were able to completely control an agent's preferences by imposing taxes would be unrealistic: to pick a perhaps rather morbid and slightly tongue in cheek example, no matter how much you propose to tax me, I would still choose to achieve my goal of being alive rather than otherwise. If we *did* have complete control over agents' preferences through taxation, then the problems we consider in this paper would indeed be rather trivial. In our setting specifically, it is assumed that no matter what the level of taxes, *an agent would still prefer to have its goal achieved than not*. This imposes a fundamental limit on the extent to which an agent's preferences can be perturbed by taxation.

We begin in the following section by introducing the model of Boolean games that we use throughout the remainder of the paper. We then introduce taxation schemes, and the incentive design problem. After investigating some properties of the incentive design problem, we go on to consider socially equitable properties of taxation schemes (such as minimising the total tax burden, etc). We conclude with a discussion and future work.

2. BOOLEAN GAMES

In this section, we introduce the model of Boolean games that we work with throughout the remainder of this paper. This model slightly generalises previous models of Boolean games [6, 2, 4], in that it explicitly represents the costs of each action. In what follows, we let $\mathbb{R}_{\geq}$ denote the set of real numbers greater than or equal to 0.

Propositional Logic: Throughout the paper, we make use of classical propositional logic, and for completeness, we thus begin by recalling the technical framework of this logic. Let $\mathbb{B} = \{\top, \bot\}$ be the set of Boolean truth values, with "$\top$" being truth and "$\bot$" being falsity. We will abuse notation a little by using $\top$ and $\bot$ to denote both the syntactic constants for truth and falsity respectively, as well as their semantic counterparts (i.e., the respective truth values). Let $\Phi = \{p, q, \ldots\}$ be a (finite, fixed, non-empty) vocabulary of Boolean variables, and let $\mathcal{L}$ denote the set of (well-formed) formulae of propositional logic over Φ, constructed using the conventional Boolean operators ("$\wedge$", "$\vee$", "$\rightarrow$", "$\leftrightarrow$", and "$\neg$"), as well as the truth constants "$\top$" and "$\bot$". We assume a conventional semantic consequence relation "$\models$" for propositional logic. A *valuation* is a total function $v : \Phi \rightarrow \mathbb{B}$, assigning truth or falsity to every Boolean variable. We write $v \models \varphi$ to mean that φ is true under, or satisfied by, valuation v, where the satisfaction relation "$\models$" is defined in the standard way. Let $\mathcal{V}$ denote the set of all valuations over Φ.

We write $\models \varphi$ to mean that φ is a tautology, i.e., is satisfied by every valuation. We denote the fact that formulae $\varphi, \psi \in \mathcal{L}$ are logically equivalent by $\varphi \Leftrightarrow \psi$; thus $\varphi \Leftrightarrow \psi$ means that $\models \varphi \leftrightarrow \psi$. Note that "$\Leftrightarrow$" is a meta-language relation symbol, which should not be confused with the object-language bi-conditional operator "$\leftrightarrow$".

Agents, Goals, and Controlled Variables: The games we consider are populated by a set $Ag = \{1, \ldots, n\}$ of *agents* – the players of the game. Each agent is assumed to have a *goal*, characterised by an $\mathcal{L}$-formula: we write γ_i to denote the goal of agent $i \in Ag$. Each agent $i \in Ag$ *controls* a (possibly empty) subset Φ_i of the overall set of Boolean variables (cf. [14]). By "control", we mean that i has the unique ability within the game to set the value (either $\top$ or $\bot$) of each variable $p \in \Phi_i$. We will require that $\Phi_1, \ldots, \Phi_n$ forms a partition of Φ, i.e., every variable is controlled by some agent and no variable is controlled by more than one agent ($\Phi_i \cap \Phi_j = \emptyset$ for $i \neq j$). Where $i \in Ag$, a *choice* for agent i is defined by a function $v_i : \Phi_i \rightarrow \mathbb{B}$, i.e., an allocation of truth or falsity to all the variables under i's control. Let $\mathcal{V}_i$ denote the set of choices for agent i. The intuitive interpretation we give to $\mathcal{V}_i$ is that it defines the *actions* or *strategies* available to agent i; the *choices* available to the agent.

An *outcome*, $(v_1, \ldots, v_n) \in \mathcal{V}_1 \times \cdots \times \mathcal{V}_n$, is a collection of choices, one for each agent. Clearly, every outcome uniquely defines a valuation, and we will often think of outcomes as valuations, for example writing $(v_1, \ldots, v_n) \models \varphi$ to mean that the valuation defined by the outcome $(v_1, \ldots, v_n)$ satisfies formula $\varphi \in \mathcal{L}$. Let $\varphi_{(v_1,\ldots,v_n)}$ denote the formula that uniquely characterises the outcome $(v_1, \ldots, v_n)$:

$$\varphi_{(v_1,\ldots,v_n)} = \left(\bigwedge_{\substack{p \in \Phi \\ (v_1,\ldots,v_n) \models p}} p \right) \wedge \left(\bigwedge_{\substack{q \in \Phi \\ (v_1,\ldots,v_n) \not\models q}} \neg q \right)$$

Let $succ(v_1, \ldots, v_n)$ denote the set of agents who have their goal achieved by outcome $(v_1, \ldots, v_n)$, i.e.,:

$$succ(v_1, \ldots, v_n) = \{i \in Ag \mid (v_1, \ldots, v_n) \models \gamma_i\}.$$

Costs: Intuitively, the actions available to agents correspond to setting variables true or false. We assume that these actions have *costs*, defined by a *cost function* $c : \Phi \times \mathbb{B} \rightarrow \mathbb{R}_{\geq}$, so that $c(p, b)$ is the marginal cost of assigning variable $p \in \Phi$ the value $b \in \mathbb{B}$. We let $c_{\mathbf{0}}$ denote the cost function that assigns zero cost to all assignments.

This notion of a cost function represents an obvious generalisation of previous presentations of Boolean games: costs were not considered in the original presentation of Boolean games [6, 2], and while costs were introduced in [4], it was assumed that only the action of setting a variable to $\top$ would incur a cost. (In fact, as we shall see later, costs are, in a technical sense, not required in our framework; we can capture the key strategic issues at stake without them. However, it is natural from the point of view of modelling to have costs for actions, and to think about costs as being imposed from within the game, and taxes, (defined below), as being imposed from without.)

Boolean Games: Collecting these components together, a *Boolean game*, G, is a $(2n + 3)$-tuple:

$$G = \langle Ag, \Phi, c, \gamma_1, \ldots, \gamma_n, \Phi_1, \ldots, \Phi_n \rangle,$$

where $Ag = \{1, \ldots, n\}$ is a set of agents, $\Phi = \{p, q, \ldots\}$ is a finite set of Boolean variables, $c : \Phi \times \mathbb{B} \rightarrow \mathbb{R}_{\geq}$ is a cost function, $\gamma_i \in \mathcal{L}$ is the goal of agent $i \in Ag$, and $\Phi_1, \ldots, \Phi_n$ is a partition of Φ over Ag, with the intended interpretation that Φ_i is the set of Boolean variables under the unique control of $i \in Ag$. We will say a game is *cost free* if it has cost function $c_{\mathbf{0}}$.

When playing a Boolean game, the primary aim of an agent i will be to choose an assignment of values for the variables Φ_i under its control so as to satisfy its goal γ_i. The difficulty is that γ_i may contain variables controlled by other agents $j \neq i$, who will also be trying to choose values for their variables Φ_j so as to get their goals satisfied; and their goals in turn may be dependent on the variables Φ_i. Note that if an agent has multiple ways of gettings its goal achieved, then it will prefer to choose one that minimises costs; and if an agent cannot get its goal achieved, then it simply chooses to minimise costs. These considerations are what give Boolean games their strategic character. For the moment, we will postpone the formal definition of the utility functions and preferences associated with our games.

3. DESIGNING INCENTIVES

We can now describe in more detail the overall problem that we consider in the remainder of the paper. Imagine a society populated by agents Ag, with each agent $i \in Ag$ having a goal $\gamma_i \in \mathcal{L}$ and actions corresponding to valuations to Φ_i. We assume an external *principal* has some goal $\Upsilon \in \mathcal{L}$ that it wants the society to achieve, and to this end, wants to incentivise the agents Ag to act collectively so as to bring about Υ. Incentives in our model are provided by *taxation schemes*.

Taxation Schemes: A taxation scheme defines additional (imposed) costs on actions, over and above those given by the marginal cost function c. While the cost function c is fixed and immutable for any given Boolean game, the principal is assumed to be at liberty to define a taxation scheme as they see fit. Agents will seek to minimise their overall costs, and so by assigning different levels of taxation to different actions, the principal can incentivise agents away from performing some actions and towards performing others; if the principal designs the taxation scheme correctly, then agents are incentivised to choose valuations $(v_1, \ldots, v_n)$ so as to satisfy Υ (i.e., so that $(v_1, \ldots, v_n) \models \Upsilon$).

How exactly should we model taxation schemes? One very general approach would be to levy taxes on the basis of *outcomes*. We could model such taxes by a function $\tau : Ag \times \mathcal{V}_1 \times \cdots \times \mathcal{V}_n \to \mathbb{R}_{\geq}$, with the intended interpretation that $\tau(i, v_1, \ldots, v_n)$ is the amount of tax that would be imposed on agent i if the outcome $(v_1, \ldots, v_n)$ was selected. However, for the purposes of the present paper, we choose a simpler, *additive* model of taxes, the idea being that taxes are levied on individual actions, and the total tax imposed on an agent i is the sum of the taxes on individual choices (assignments of truth or falsity to a variable) made in the outcome v_i chosen by i.

Formally, we therefore model a taxation scheme as a function $\tau : \Phi \times \mathbb{B} \to \mathbb{R}_{\geq}$, where the intended interpretation is that $\tau(p, b)$ is the tax that would be imposed on the agent controlling p if the value b was assigned to the Boolean variable p. The total tax paid by an agent i in choosing a valuation $v_i \in \mathcal{V}_i$ will be $\sum_{p \in \Phi_i} \tau(p, v_i(p))$.

We let τ_0 denote the taxation scheme that applies no taxes to any choice, i.e., $\forall x \in \Phi$ and $b \in \mathbb{B}$, $\tau_0(x, b) = 0$. Let $\mathcal{T}(G)$ denote the set of taxation schemes over G. We make one technical assumption in what follows, relating to the space requirements for taxation schemes in $\mathcal{T}(G)$. Unless otherwise stated explicitly, we will assume that we are restricting our attention to taxation schemes whose values can be represented with a space requirement that is bounded by a polynomial in the size of the game. This seems a reasonable requirement: realistically, taxation schemes requiring space exponential in the size of the game at hand could not be manipulated. It is important to note that this requirement relates to the *space requirements for taxes*, and not to the *size of taxes themselves*: for a polynomial function $f : \mathbb{N} \to \mathbb{N}$, the value $2^{f(n)}$ can be represented using only a polynomial number of bits (i.e., $f(n)$ bits).

Utilities and Preferences: One important assumption we make is that while taxation schemes can influence the decision making of rational agents, they cannot, ultimately, change the goals of an agent. That is, if an agent has a chance to achieve its goal, it will take it, no matter what the taxation incentives are to do otherwise. To understand this point, and to see formally how incentives work, we need to formally define the utility functions for agents, and for this we require some further auxiliary definitions. First, with a slight abuse of notation, we extend cost and taxation functions to partial valuations as follows:

$$c_i(v_i) = \sum_{p \in \Phi_i} c(p, v_i(p))$$

$$\tau_i(v_i) = \sum_{p \in \Phi_i} \tau(p, v_i(p))$$

Next, let v_i^e denote the most expensive possible course of action for agent i:

$$v_i^e \in \arg\max_{v_i \in \mathcal{V}_i} (c_i(v_i) + \tau_i(v_i)).$$

Let μ_i denote the cost to i of its most expensive course of action:

$$\mu_i = c_i(v_i^e) + \tau_i(v_i^e).$$

Given these definitions, we define the *utility* to agent i of an outcome $(v_1, \ldots, v_n)$, as follows:

$$u_i(v_1, \ldots, v_n) = \begin{cases} 1 + \mu_i - (c_i(v_i) + \tau_i(v_i)) & \text{if } (v_1, \ldots, v_n) \models \gamma_i \\ -(c_i(v_i) + \tau_i(v_i)) & \text{otherwise.} \end{cases}$$

Thus utility for agent i will range from $1 + \mu_i$ (the best outcome for i, where it gets its goal achieved by performing actions that have no tax or other cost) down to $-\mu_i$ (where i does not get its goal achieved but makes its most expensive choice). This definition has the following properties:

- an agent prefers all outcomes that satisfy its goal over all those that do not satisfy it;
- between two outcomes that satisfy its goal, an agent prefers the one that minimises total expense (= marginal costs + taxes); and
- between two valuations that *do not* satisfy its goal, an agent prefers to minimise total expense.

It is important to note that while utility functions provide a convenient numeric representation of preference relations, utility is not transferable *in our settings.*

Solution Concepts: Given this formal definition of utility, we can define solution concepts in the standard game-theoretic way [12]. In this paper, we focus on (pure) Nash equilibrium. (Of course, other solution concepts, such as dominant strategy equilibria, might also be considered, but for simplicity, in this paper we focus on Nash equilibria.) We say an outcome $(v_1, \ldots, v_i, \ldots, v_n)$ is a Nash equilibrium if for all agents $i \in Ag$, there is no $v_i' \in \mathcal{V}_i$ such that $u_i(v_1, \ldots, v_i', \ldots, v_n) > u_i(v_1, \ldots, v_i, \ldots, v_n)$. Let $NE(G, \tau)$ denote the set of all Nash equilibria of the game G with taxation scheme τ.

Before proceeding, let us consider some properties of Nash equilibrium outcomes. First, observe that an unsuccessful agent will choose a least cost course of action in any Nash equilibrium.

PROPOSITION 1. *Suppose* $(v_1^*, \ldots, v_i^*, \ldots, v_n^*) \in NE(G, \tau)$ *is such that* $i \notin succ(v_1^*, \ldots, v_i^*, \ldots, v_n^*)$. *Then*

$$v_i^* \in \arg\min_{v_i \in \mathcal{V}_i} c_i(v_i) + \tau_i(v_i)$$

PROOF. Agent i cannot make a choice v_i' that $(v_1^*, \ldots, v_i', \ldots, v_n^*) \models \gamma_i$, otherwise $u_i(v_1^*, \ldots, v_i', \ldots, v_n^*) > u_i(v_1^*, \ldots, v_i^*, \ldots, v_n^*)$, in which case $(v_1^*, \ldots, v_i^*, \ldots, v_n^*) \notin NE(G, \tau)$. So, the only way i could profitably deviate would be by making an alternative choice v_i' that reduced costs compared to v_i^*. But by definition, v_i^* minimises i's costs. □

The following is an obvious decision problem:

NASH OUTCOME VERIFICATION:
Instance: Boolean game G, taxation scheme τ, and outcome $(v_1, \ldots, v_n)$.
Question: Is $(v_1, \ldots, v_n) \in NE(G, \tau)$?

PROPOSITION 2. *NASH OUTCOME VERIFICATION is co-NP-complete, even for two player games with $\tau = \tau_0$ and where c assigns no costs.*

PROOF. Membership is immediate. For hardness, we reduce SAT to the complement problem. Given an instance φ of SAT over variables $x_1, \ldots, x_k$, define a game G with $Ag = \{1, 2\}$, $\Phi = \{x_1, \ldots, x_k, z\}$, (where z does not occur in φ), $\Phi_1 = \{x_1, \ldots, x_k\}$, $\Phi_2 = \{z\}$, $\gamma_1 = \varphi$, $\gamma_2 = z$, let $v_1(y) = \bot$ for all $y \in \Phi_1$, and let $v_2(z) = \top$. We claim $(v_1, v_2) \notin NE(G, \tau_0)$ iff φ is satisfiable. ($\rightarrow$) Suppose $(v_1, v_2) \notin NE(G, \tau_0)$. Then either agent 1 or agent 2 can benefit by deviating. Clearly agent 2 cannot benefit, since it gets its goal achieved through v_2 at no cost, which is optimal for 2. So 1 must be able to benefit by deviating. Since it incurs no cost through v_1, the only way agent 1 could benefit would be by achieving its goal, which would imply φ was satisfiable. ($\leftarrow$) Suppose φ is satisfiable. Then player 1 could benefit by choosing a valuation $v_1' \neq v_1$ satisfying φ. Hence $(v_1, v_2) \notin NE(G, \tau_0)$. □

Next, note that while being able to model costs in games explicitly is attractive from a modelling perspective, it is, in a sense, unnecessary from a purely technical point of view: we can always design a taxation scheme that simulates the costs and thus gives rise to the same set of Nash equilibria.

PROPOSITION 3. *Let G be a game with cost function c and let τ be a taxation scheme for G. Then there exists a taxation scheme τ' such that $NE(G, \tau) = NE(G', \tau')$ for the game G' we obtain by replacing c with c_0 in G.*

PROOF. Let $\tau'(p, b) = \tau(p, b) + c(p, b)$ for all $p \in \Phi$ and all $b \in \mathbb{B}$. Then the utility functions for (G', τ') are identical to those for (G, τ), and thus the Nash equilibria must coincide as well. □

Moreover, we can show that, for the analysis of Nash equilibria, it suffices to consider taxation schemes that only impose taxes on making a variable *true* (rather than *false*). Call a taxation scheme τ *positive* if $\tau(p, \bot) = 0$ for all $p \in \Phi$. Now consider two zero cost games G and G'[1]. We call G' a *variant* of G if G' is the same as G, except that for some $p \in \Phi$ all occurrences of p in the agents' goals γ_i have been replaced by $\neg p$ (but Υ has not been changed).

PROPOSITION 4. *Let G be a zero cost game and let τ be a taxation scheme for that game. Then there exists a variant G' of G and a positive taxation scheme τ' such that $NE(G, \tau) = NE(G', \tau')$.*

PROOF. We have to define G' and τ' with respect to each $p \in \Phi$. For all variables p we will have $\tau'(p, \bot) = 0$ (as τ' should be positive). So we have to define the values $\tau'(p, \top)$ and we have to specify whether p should occur in the goal formulas in G' as in G, or whether p should get flipped (i.e., whether it should get rewritten as $\neg p$).

1. If $\tau(p, \top) = \tau(p, \bot)$, then we set $\tau'(p, \top) = 0$ and we leave p untouched in the game.
2. If $\tau(p, \top) > \tau(p, \bot)$, then we set $\tau'(p, \top) = \tau(p, \top) - \tau(p, \bot)$ and we again leave p untouched in the game.
3. If $\tau(p, \top) < \tau(p, \bot)$, then we set $\tau'(p, \top) = \tau(p, \bot) - \tau(p, \top)$ and we flip p in the game.

The crucial feature of this construction is that the difference in tax between making p *true* or *false* remains $|\tau(p, \top) - \tau(p, \bot)|$ in the new game, and the new taxation scheme still "pushes in the same direction" as before. Therefore, the utility functions for (G', τ') are identical to those for (G, τ), and thus the Nash equilibria must coincide as well. □

[1]This restriction to games with zero cost is not required, but it does simplify exposition; and we have just seen that for the analysis of Nash equilibria it suffices to consider zero cost games.

Incentive Design: We now come to the main problems that we consider in the remainder of the paper. Suppose we have an agent, which we will call the principal, who is external to a game G. The principal is at liberty to impose taxation schemes on the game G. It will not do this for no reason, however: it does it because it wants to provide incentives for the agents in G to choose certain collective outcomes. Specifically, the principal wants to incentivise the players in G to choose rationally a collective outcome that satisfies an *objective*, which is represented as a propositional formula Υ over the variables Φ of G. We refer to this general problem – trying to find a taxation scheme that will incentivise players to choose rationally a collective outcome that satisfies a propositional formula Υ – as the *implementation problem*. It inherits concepts from the theory of Nash implementation in mechanism design [7], although our use of Boolean games, taxation schemes, and propositional formulae to represent objectives is quite different.

3.1 Weak Implementation

Let $\mathcal{WI}(G, \Upsilon)$ denote the set of taxation schemes over G that satisfy a propositional objective Υ in at least one Nash equilibrium outcome:

$$\mathcal{WI}(G, \Upsilon) = \{\tau \in \mathcal{T}(G) \mid \exists (v_1, \ldots, v_n) \in NE(G, \tau) \text{ s.t. } (v_1, \ldots, v_n) \models \Upsilon\}.$$

Given this definition, we can state the first basic decision problem that we consider in the remainder of the paper:

WEAK IMPLEMENTATION:
Instance: Boolean game G and objective $\Upsilon \in \mathcal{L}$.
Question: Is it the case that $\mathcal{WI}(G, \Upsilon) \neq \emptyset$?

If the answer to the WEAK IMPLEMENTATION problem (G, Υ) is "yes", then we say that Υ *can be weakly implemented in Nash equilibrium* (or simply: Υ can be weakly implemented in G). Let us see an example.

EXAMPLE 1. *Define a game G as follows: $Ag = \{1, 2\}$, $\Phi = \{p_1, p_2\}$, $\Phi_i = \{p_i\}$, $\gamma_1 = p_1$, $\gamma_2 = \neg p_1 \wedge \neg p_2$, $c(p_1, b) = 0$ for all $b \in \mathbb{B}$, while $c(p_2, \top) = 1$ and $c(p_2, \bot) = 0$. Define an objective $\Upsilon = p_1 \wedge p_2$. Now, without any taxes (i.e., with taxation scheme τ_0), there is a single Nash equilibrium, (v_1^*, v_2^*), which satisfies $p_1 \wedge \neg p_2$. Agent 1 gets its goal achieved, while agent 2 does not; and moreover $(v_1^*, v_2^*) \not\models \Upsilon$. However, if we adjust τ so that $\tau(p_2, \bot) = 10$, then we find a Nash equilibrium outcome (v_1', v_2') such that $(v_1', v_2') \models p_1 \wedge p_2$, i.e., $(v_1', v_2') \models \Upsilon$. Here, agent 2 is not able to get its goal achieved, but it can, nevertheless, be incentivised by taxation to make a choice that ensures the achievement of the objective Υ.*

So, what objectives Υ can be weakly implemented? At first sight, it might appear that the satisfiability of Υ is a sufficient condition for

implementability. Consider the following naive approach for constructing taxation schemes with the aim of implementing satisfiable objectives Υ:

> (*) Find a valuation v such that $v \models \Upsilon$ (such a valuation will exist since Υ is satisfiable). Then define a taxation scheme τ such that $\tau(p,b) = 0$ if $b = v(p)$ and $\tau(p,b) = k$ otherwise, where k is a suitably large number.

Thus, the idea is simply to make all choices other than selecting an outcome that satisfies Υ too expensive to be rational. In fact, this approach does not work, because of an important subtlety of the definition of utility. In designing a taxation scheme, the principal can perturb an agent's choices between different valuations, but it *cannot* perturb them in such a way that an agent would prefer an outcome that does not satisfy it's goal over an outcome that does. We have:

PROPOSITION 5. *There exist instances of the* WEAK IMPLEMENTATION *problem with satisfiable objectives* Υ *that cannot be weakly implemented.*

PROOF. Consider the following example. Define a game G as follows: $Ag = \{1\}$, $\Phi = \Phi_1 = \{p\}$, $\gamma_1 = p$, $c(p,b) = 0$ for all $b \in \mathbb{B}$. Let $\Upsilon = \neg p$. Suppose there is a taxation scheme τ such that $\exists v_1 \in NE(G,\tau)$ and $v_1 \models \Upsilon$. Clearly, $v_1(p) = \bot$, so $v_1 \not\models \gamma_1$ and thus $u_1(v) = -(c_1(v_1) + \tau(v_1)) = 0$. But consider the valuation $v_1'(p) = \top$, which since $v_1' \models \gamma_1$ would yield $u_1(v_1') = 1 + \mu_1 - (c_1(v_1') + \tau_1(v_1')) = 1$. Thus $u_1(v_1') > u_1(v_1)$; contradiction. □

What about tautologous objectives, i.e., objectives Υ such that $\Upsilon \Leftrightarrow \top$? Again, we might be tempted to assume that tautologies are trivially implementable. This is not in fact the case, however, as it may be that $NE(G,\tau) = \emptyset$ for all taxation schemes τ:

PROPOSITION 6. *There exist instances of the* WEAK IMPLEMENTATION *problem with tautologous objectives* Υ *that cannot be implemented.*

PROOF. Define a game G with $Ag = \{1,2\}$, $\Phi = \{p,q\}$, $\Phi_1 = \{p\}$, $\Phi_2 = \{q\}$, $\gamma_1 = (p \leftrightarrow q)$, $\gamma_2 = \neg(p \leftrightarrow q)$, and c assigns zero cost to all actions. Clearly $NE(G,\tau) = \emptyset$. For example, in the outcome (v_1, v_2) in which $v_1(p) = \top$ and $v_2(q) = \top$, agent 1 would prefer to change its valuation to $v_2'(q) = \top$. There is, in fact, no taxation scheme τ such that $NE(G,\tau) \neq \emptyset$. □

Tautologous objectives might appear to be of little interest, but we argue that this is not the case. Suppose we have a game G such that $NE(G,\tau_0) = \emptyset$. Then, in its unmodified condition, this game is *unstable*: it has no equilibria. Thus, we will refer to the problem of implementing $\top$ (= checking for the existence of a taxation scheme that would ensure at least one Nash equilibrium outcome), as the STABILISATION problem. The following example illustrates STABILISATION.

EXAMPLE 2. *Let* $Ag = \{1,2,3\}$*, with* $\varphi = \{p,q,r\}$*,* $\Phi_1 = \{p\}$*,* $\Phi_2 = \{q\}$*,* $\Phi_3 = \{r\}$*,* $\gamma_1 = \top$*,* $\gamma_2 = (q \wedge \neg p) \vee (q \leftrightarrow r)$*,* $\gamma_3 = (r \wedge \neg p) \vee \neg(q \leftrightarrow r)$*,* $c(p,\top) = 0$*,* $c(p,\bot) = 1$*, and all other costs are* 0*. For any outcome in which* $p = \bot$*, agent* 1 *would prefer to set* $p = \top$*, so no such outcome can be stable. So, consider outcomes* (v_1,v_2,v_3) *in which* $p = \top$*. Here if* $(v_1,v_2,v_3) \models q \leftrightarrow r$ *then agent* 3 *would prefer to deviate, while if* $(v_1,v_2,v_3) \not\models q \leftrightarrow r$ *then agent* 2 *would prefer to deviate. Now, consider a taxation scheme with* $\tau(p,\top) = 10$ *and* $\tau(p,\bot) = 0$ *and all other taxes are* 0*. With this scheme, the outcome in which all variables are set to* $\bot$ *is a Nash equilibrium. Hence this taxation scheme stabilises the system.*

Returning to the weak implementation problem, we can derive a *sufficient* condition for weak implementation, as follows.

PROPOSITION 7. *For all games G and objectives* Υ*, if the formula* Υ' *is satisfiable:*

$$\Upsilon' = \Upsilon \wedge \bigwedge_{i \in Ag} \gamma_i$$

then $\mathcal{WI}(G,\Upsilon) \neq \emptyset$*.*

PROOF. Assume $\Upsilon' = \Upsilon \wedge \bigwedge_{i \in Ag} \gamma_i$ is satisfiable. Let v be a valuation such that $v \models \Upsilon'$. The basic idea is to use the approach (*), described above, to build a taxation scheme ensuring that the valuation v is a rational choice. For all $i \in Ag$, $x \in \Phi_i$ and $b \in \mathbb{B}$, define:

$$\tau(x,b) = \begin{cases} 0 & \text{if } b = v(x) \\ 1 + c_i(v_i^e) & \text{otherwise.} \end{cases}$$

(Recall that v_i^e is the choice for i that has the highest marginal cost.) Let $(v_1^*,\ldots,v_n^*)$ be the outcome corresponding to the valuation v. Obviously, $(v_1^*,\ldots,v_n^*) \models \Upsilon$. We claim that $(v_1^*,\ldots,v_n^*) \in NE(G,\tau)$. For suppose that $(v_1^*,\ldots,v_n^*)$ is not a Nash equilibrium. Then some agent i can benefit by deviating. Since by construction $(v_1^*,\ldots,v_n^*) \models \gamma_i$, then i can only benefit from a choice that would decrease its overall costs. But the construction of τ ensures that any other choice would *increase* taxes more than any benefit gained by decreasing marginal costs. So, i cannot benefit by changing its choice, and so $(v_1^*,\ldots,v_n^*)$ is a Nash equilibrium. □

We know from [2] that the problem of checking for the existence of pure strategy Nash equilibria in cost-free Boolean games is Σ_2^p-complete. It turns out that the IMPLEMENTATION problem is no harder:

PROPOSITION 8. *The* STABILISATION *problem is* Σ_2^p*-complete, even if taxes are* 0*-bounded. As a consequence, the* WEAK IMPLEMENTATION *problem is also* Σ_2^p*-complete.*

PROOF. Membership requires evaluating the following condition:

$$\exists \tau \in \mathcal{T}(G), \exists (v_1,\ldots,v_n) \in \mathcal{V}_1 \times \cdots \times \mathcal{V}_n, \underbrace{(v_1,\ldots,v_n) \in NE(G,\tau)}_{(**)}.$$

Notice that the condition $(**)$ is a co-NP predicate, and that the existential quantifiers can be computed in NP (recall that we assume taxation schemes in $\mathcal{T}(G)$ require space at most polynomial in the size of G, and hence guessed in non-deterministic polynomial time). Thus the problem is in Σ_2^p. For hardness, we can trivially reduce the problem of checking for the existence of pure strategy Nash equilibria in cost-free Boolean games, which was proved Σ_2^p-complete in [2, Proposition 5]. Given a cost free game as in [2], we construct an instance of one of our games directly, setting all costs to 0; we then ask whether the system can be stabilised with a tax bound of 0. Clearly, the answer is "yes" iff the given Boolean game instance has a pure strategy Nash equilibrium. □

3.2 (Strong) Implementation

The fact that $\mathcal{WI}(G,\Upsilon) \neq \emptyset$ is good news of a kind – it tells us that we can impose a taxation scheme such that *at least one* rational (NE) outcome of the game satisfies Υ. However, it could be that there are many taxation schemes, and only one of them satisfies Υ. This motivates us to consider the *strong implementation* (or simply *implementation*) problem. Strong implementation corresponds closely to the notion of Nash implementation in the mechanism design literature [7]. Let $\mathcal{SI}(G,\Upsilon)$ denote the set of taxation schemes τ over G such that:

1. G, τ has at least one Nash equilibrium outcome;
2. all Nash equilibrium outcomes of G, τ satisfy Υ.

Formally:

$$\mathcal{SI}(G,\Upsilon) = \{\tau \in \mathcal{T}(G) \mid NE(G,\tau) \neq \emptyset \quad \& \quad \forall(v_1,\ldots,v_n) \in NE(G,\tau) : (v_1,\ldots,v_n) \models \Upsilon\}.$$

This gives us the following decision problem:

IMPLEMENTATION:
Instance: Boolean game G and objective $\Upsilon \in \mathcal{L}$.
Question: Is it the case that $\mathcal{SI}(G,\Upsilon) \neq \emptyset$?

It turns out that strong implementation is no harder than weak implementation:

PROPOSITION 9. IMPLEMENTATION *is* Σ_2^p*-complete.*

PROOF. Observe that the problem involves evaluating the following condition: is it the case that $\exists\tau \in \mathcal{T}(G) : NE(G,\tau) \neq \emptyset$ and $\forall(v_1,\ldots,v_n) \in NE(G,\tau)$ we have $(v_1,\ldots,v_n) \models \Upsilon$? Expanding out and re-arranging, it can be seen that this is equivalent to asking whether $\exists\tau \in \mathcal{T}(G), \exists(v_1,\ldots,v_n) \in \mathcal{V}_1 \times \cdots \times \mathcal{V}_n, \forall(v'_1,\ldots,v'_n) \in \mathcal{V}_1 \times \cdots \times \mathcal{V}_n$, we have $(v_1,\ldots,v_n) \in NE(G,\tau)$ and if $(v'_1,\ldots,v'_n) \in NE(G,\tau)$ we have $(v'_1,\ldots,v'_n) \models \Upsilon$. Clearly this is a Σ_2^p predicate. For hardness, we can reduce the STABILISATION problem as in Proposition 8. □

How are $\mathcal{WI}(G,\Upsilon)$ and $\mathcal{SI}(G,\Upsilon)$ related? It turns out that weak and strong implementation are indeed different:

PROPOSITION 10.

1. *For all games G and objectives* Υ *we have:*

$$\mathcal{SI}(G,\Upsilon) \subseteq \mathcal{WI}(G,\Upsilon).$$

2. *There exist games G and objectives* Υ *s.t.:*

$$\mathcal{WI}(G,\Upsilon) \not\subseteq \mathcal{SI}(G,\Upsilon).$$

PROOF. Item (1) is immediate: if a taxation scheme strongly implements Υ in G then it weakly implements it. For item (2), we give an example of an game and objective such that the objective can be weakly, but not strongly implemented. Let $Ag = \{1,2\}$, with $\Phi = \{p,q\}$, $\Phi_1 = \{p\}, \Phi_2 = \{q\}$, $\gamma_1 = \gamma_2 = (p \leftrightarrow q)$, with cost function c_0. Finally, let $\Upsilon = p \wedge q$. Now, the taxation function τ_0 with zero taxes will weakly implement Υ: there will be two Nash equilibria, one satisfying $p \wedge q$ and the other satisfying $\neg(p \vee q)$. However, $\Upsilon = p \wedge q$ cannot be strongly implemented, because the outcome satisfying $\neg(p \vee q)$ will be a Nash equilibrium for all taxation schemes τ. To see this, observe that for it not to be a Nash equilibrium, one agent would benefit by deviating; but by definition of the utility functions, such a deviation would involve an agent moving from positive to negative utility. □

Thus, to show that an objective Υ cannot be strongly implemented, it suffices to show that it cannot be weakly implemented.

One interesting question relates to the size of taxes required to for implementation. In some cases, it turns out that we only require very small amounts of tax:

PROPOSITION 11. *Let* $\varepsilon << 1$ *be any arbitrarily small positive number. If G is cost-free (i.e., with cost function* c_0*) then* Υ *is implementable (respectively, weakly implementable) in G iff* Υ *is implementable by a taxation scheme bounded by* ε*.*

PROOF. We first claim that for any arbitrarily small ε and any cost-free Boolean game G there is a cost-free Boolean game such that the taxation scheme is bounded by ε and the sets of Nash equilibria of the two games coincide. The idea is to define a new taxation scheme τ' by using ε to systematically scale down taxation values from τ. The transformation from (G,τ) to (G,τ') preserves the relative order on utilities, that is, $u_i^{G,\tau}(v_i) \geq u_i^{G,\tau}(v'_i)$ if and only if $u_i^{G,\tau'}(v_i) \geq u_i^{G,\tau'}(v'_i)$. Therefore, $v = (v_1,\ldots,v_n)$ is a Nash equilibrium of (G,τ) if and only if it is a Nash equilibrium of (G,τ'). As a consequence, Υ is implementable (respectively, weakly implementable) in G by taxation scheme τ if and only if it is implementable (resp. weakly implementable) in G by τ'. □

4. TAXATION AND SOCIAL WELFARE

In attempting to design a taxation scheme τ for a Boolean game G, the primary aim of a principal is to design the scheme so that agents are rationally motivated to choose an outcome satisfying the objective Υ. However, if it is possible to incentivise agents to satisfy Υ, then there will, in general, be multiple possible taxation schemes that incentivise the agents in this way, and not all of these taxation schemes will be equally desirable from the point of view of society. In this section, therefore, we consider different societal criteria that might be considered by a principal when choosing a taxation scheme; our discussion here is inspired by the literature on axioms for cooperative decision-making [9].

Utilitarian Social Welfare: The first idea we consider is the very well-known concept of *maximising utilitarian social welfare*. Formally, the social welfare of an outcome $(v_1,\ldots,v_n)$ is denoted $sw(v_1,\ldots,v_n)$:

$$usw(v_1,\ldots,v_n) = \sum_{i \in Ag} u_i(v_1,\ldots,v_n).$$

An outcome $(v_1^{usw},\ldots,v_n^{usw})$ that maximises utilitarian social welfare is thus one satisfying:

$$(v_1^{usw},\ldots,v_n^{usw}) \in \arg\max_{(v_1,\ldots,v_n)} usw(v_1,\ldots,v_n).$$

Of course, simply finding an outcome that maximises social welfare in itself is not much use if agents are rationally motivated to choose another outcome, which does not maximise social welfare. We therefore say a taxation scheme τ *weakly implements* utilitarian social welfare maximisation in a game G if

$$\exists(v'_1,\ldots,v'_n) \in NE(G,\tau) \text{ s.t. } (v'_1,\ldots,v'_n) \in \arg\max_{(v_1,\ldots,v_n)} usw(v_1,\ldots,v_n).$$

We can define strong implementation in the expected way. With a slight abuse of notation, let $\mathcal{WI}(G,usw)$ denote the set of taxation schemes that weakly implement utilitarian social welfare maximisation, and let $\mathcal{SI}(G,usw)$ denote the set of taxation schemes that strongly implement utilitarian social welfare maximisation. Can we always implement utilitarian social welfare maximisation? No:

PROPOSITION 12. *There are games G in which utilitarian social welfare maximisation cannot be weakly implemented (and hence cannot be strongly implemented).*

PROOF. Define a game G with $Ag = \{1,2,3\}$, $\Phi = \{p_1,p_2,p_3\}$, $\Phi_i = \{p_i\}$, and $\gamma_1 = p_1 \vee (p_2 \wedge p_3)$, $\gamma_2 = \neg p_2$, $\gamma_3 = \neg p_3$, $c(p_1,\top) = 20$, $c(p_1,\bot) = 1$, $c(p_2,\top) = 2$, $c(p_2,\bot) = 1$, $c(p_3,\top) = 2$, $c(p_3,\bot) = 1$. Now, with taxation scheme τ_0, there is a unique Nash equilibrium (v_1^*,v_2^*,v_3^*) in which agent 1 sets $p_1 = \top$, agent 2 sets $p_2 = \bot$, agent 3 sets $p_3 = \bot$. We have $u_1(v_1^*,v_2^*,v_3^*) = 1+20-20 = 1$, $u_2(v_1^*,v_2^*,v_3^*) = 1+2-1 = 2$,

and $u_3(v_1^*, v_2^*, v_3^*) = 1 + 2 - 1 = 2$, and so $usw(v_1^*, v_2^*, v_3^*) = 1 + 2 + 2 = 5$. Now consider the outcome (v_1, v_2, v_3) that satisfies $(\neg p_1) \wedge p_2 \wedge p_3$. Observe that $(v_1, v_2, v_3) \models \gamma_1$, while $(v_1, v_2, v_3) \not\models (\gamma_2 \vee \gamma_3)$ We have $u_1(v) = 1+20-1 = 20$, $u_2(v) = -2$, and $u_3(v) = -2$. Thus $usw(v_1, v_2, v_3) = 20+(-2)+(-2) = 16$. Clearly, outcome (v_1, v_2, v_3) maximises social welfare, but no taxation scheme can weakly implement this outcome: agents 2 and 3 will always prefer to get their goal achieved. □

This example also illustrates that maximising utilitarian social welfare is not the same as maximising the number of agents that get their goal achieved.

Notice that because we represent objectives Υ as logical formula, and these logical formula can completely characterise outcomes, we can directly model the problem of implementing utilitarian social welfare maximisation as a WEAK IMPLEMENTATION problem. The following is immediate:

PROPOSITION 13. *It is possible to weakly (respectively, strongly) implement utilitarian social welfare maximisation in game G iff $\mathcal{WI}(G, \Upsilon_{usw}) \neq \emptyset$ (respectively, $\mathcal{SI}(G, \Upsilon_{usw}) \neq \emptyset$), where:*

$$\Upsilon_{usw} = \bigvee_{(v_1^*,\ldots,v_n^*) \in \arg\max_{(v_1,\ldots,v_n) \in \mathcal{V}_1 \times \cdots \times \mathcal{V}_n} usw(v_1,\ldots,v_n)} \varphi_{(v_1^*,\ldots,v_n^*)}.$$

From the point of view of a principal, of course, the main concern is to implement an objective Υ; maximising utilitarian social welfare is a secondary concern. A very natural aim of the principal will therefore be to design a taxation scheme that implements an objective while at the same time maximises the worst case utilitarian social welfare of all possible Nash equilibrium outcomes. This yields the following decision problem,

> USW IMPLEMENTATION:
> *Instance*: Boolean game G, objective Υ, social welfare measure $w \in \mathbb{R}$.
> *Question*: Does there exist a taxation scheme $\tau \in \mathcal{SI}(G, \Upsilon)$ such that $\min\{usw(v, \ldots, v_n) \mid (v_1, \ldots, v_n) \in NE(G, \tau)\} \geq w$?

PROPOSITION 14. *The* USW IMPLEMENTATION *problem is Σ_2^p-complete.*

PROOF. We reduce WEAK IMPLEMENTATION. Where G, Υ be an instance of WEAK IMPLEMENTATION, we create an instance G, Υ, w of USW IMPLEMENTATION with a value for w that is guaranteed to be below the worst case utilitarian social welfare of any outcome. □

The corresponding function problem is to compute a taxation scheme maximising the worst case utilitarian social welfare of a Nash equilibrium outcome satisfying Υ. We will denote such a taxation scheme by $\tau_{usw}(G, \Upsilon)$:

$$\tau_{usw}(G, \Upsilon) \in \arg\max_{\tau \in \mathcal{SI}(G,\Upsilon)} \min\{usw(v, \ldots, v_n) \mid (v_1, \ldots, v_n) \in NE(G, \tau)\}.$$

Notice that in the case $\Upsilon \equiv \top$, finding $\tau_{usw}(G, \Upsilon)$ reduces to simply implementing utilitarian social welfare maximisation.

Egalitarian Social Welfare: A standard criticism of utilitarian social welfare is that it does not consider how utility is distributed amongst members of a society; it may allocate all utility to one agent, leaving all others with no utility. Egalitarian social welfare provides an alternative metric: it looks at how well off the least well off member of society is. The function $esw(\cdots)$ gives the egalitarian social welfare of an outcome:

$$esw(v_1, \ldots, v_n) = \min\{u_i(v_1, \ldots, v_n) \mid i \in Ag\}.$$

The taxation scheme implementing Υ while maximising egalitarian social welfare in G is denoted $\tau_{esw}(G, \Upsilon)$:

$$\tau_{esw}(G, \Upsilon) \in \arg\max_{\tau \in \mathcal{SI}(G,\Upsilon)} \min\{esw(v, \ldots, v_n) \mid (v_1, \ldots, v_n) \in NE(G, \tau)\}.$$

Minimising the Total Tax Burden: An alternative to measuring social welfare is to consider developing a taxation scheme that implements objective Υ while imposing the lowest possible tax burden on society. Broadly, we can think of this approach as minimising the degree of intervention of the principal in the operation of society. The function $tb(\cdots)$ gives the total tax burden of an outcome:

$$tb(v_1, \ldots, v_n) = \sum_{i \in Ag} \tau(v_i).$$

We define τ_{tb} in the obvious way:

$$\tau_{tb} \in \arg\min_{\tau \in \mathcal{SI}(G,\Upsilon)} \max\{tb(v_1, \ldots, v_n) \mid (v_1, \ldots, v_n) \in NE(G, \tau)\}$$

It is easy to construct examples showing that minimising the total tax burden may result in socially undesirable outcomes (but nevertheless, it seems such "least intervention" approaches are relatively popular in human societies).

5. TAXATION AND EQUITY

It is, of course, well-known that an outcome which maximises (for example) utilitarian social welfare may in fact be extremely undesirable from the point of view of the majority of agents in a system. For example, the social welfare maximising outcome might allocate all the utility in the system to one agent, leaving all others with none. This motivates us to consider a range of possible other notions of equity with respect to taxation schemes, inspired to some extent by the economics literature on taxation [3].

Minimising the Difference in Taxes: One very obvious (although arguably naive) notion of taxation equity is to simply try to ensure that agents are taxed at broadly the same level, i.e., to minimise the maximum difference in taxes levied on different agents. Let $md(v_1, \ldots, v_n)$ give the maximum difference in taxes between any two agents in outcome $(v_1, \ldots, v_n)$:

$$md(v_1, \ldots, v_n) = \max\{abs(\tau_i(v_i) - \tau_j(v_j)) \mid \{i, j\} \subseteq Ag\}$$

where $abs(x)$ denotes the absolute value of x. Let τ_{md} denote a taxation scheme that minimises this value over all possible Nash equilibria of taxation schemes that implement Υ in G:

$$\tau_{md} \in \arg\min_{\tau \in \mathcal{SI}(G,\Upsilon)} \max\{md(v_1, \ldots, v_n) \mid (v_1, \ldots, v_n) \in NE(G, \tau)\}$$

Horizontal Equity: Simply aiming to apply the same level of taxes across an entire society may appear to be equitable, but on closer examination, it has some definite drawbacks. In particular, it does not distinguish between agents that have their goals achieved and those that do not. In the literature on taxation, the term *horizontal equity* is used to describe the idea that those in the same circumstances should be taxed at the same level [3]. One could formalise this notion in several different ways for our model, but we will fo-

cus on the following idea: in any outcome, we have two "classes" of agents: those that get their goal achieved and those that do not. Thus, when looking at the differences in taxes paid, we only compare the taxes of agents that get their goal achieved against other agents that get their goal achieved, and we compare only compare agents that do not get their goal achieved against agents that do not get their goal achieved. The function $he(\cdots)$ denote the maximum difference in tax paid between agents in the same equivalence class:

$$he(v_1,\ldots,v_n) = \max(\{abs(\tau_i(v_i)-\tau_j(v_j)) \mid \{i,j\} \subseteq Ag \;\&\; (v_1,\ldots,v_n) \models \gamma_i \wedge \gamma_j\} \cup \{abs(\tau_i(v_i)-\tau_j(v_j)) \mid \{i,j\} \subseteq Ag \;\&\; (v_1,\ldots,v_n) \models \neg(\gamma_i \vee \gamma_j)\})$$

Then τ_{he} will denote an outcome that maximises horizontal equity (i.e., minimises the difference in taxes paid by agents in the same circumstances).

$$\tau_{md} \in \arg\min_{\tau \in \mathcal{SI}(G,\Upsilon)} \max\{he(v_1,\ldots,v_n) \mid (v_1,\ldots,v_n) \in NE(G,\tau)\}$$

6. CONCLUSIONS & FUTURE WORK

Taxation schemes, in the form of the VCG mechanism and variations thereof, have received an enormous amount of attention in the computer science literature over the past two decades [11]. Much of the current interest stems from the possibility, provided by VCG, of having inventive compatible mechanisms, i.e., mechanisms that incentivise agents to truthfully report their preferences, thereby allowing the computation of outcomes that maximise social welfare. There are, however, fundamental limits to what can be achieved with incentive compatible mechanisms, and it therefore seems worth considering the design of taxation schemes to incentivise behaviours in non incentive compatible settings. After all, taxation schemes in the real world are rarely incentive compatible. In the present paper, we have studied the use of taxation schemes to incentivise behaviours in Boolean games: a natural, expressive, and compact class of logic-based games. We showed how a principal could perturb the preferences of agents in a Boolean game by imposing a taxation scheme, and in so doing, how it could, in certain circumstances, incentivise agents to choose outcomes to satisfy some social objective Υ, represented as a Boolean formula. However, we saw that while an agent's preferences can be perturbed, they are not completely malleable: no matter what the taxation scheme, an agent would always prefer to get its goal achieved than otherwise. This means there are limits on the extent to which preferences can be perturbed by taxation, and hence limits on what objectives Υ can be achieved. We studied a number of questions around the question of implementing objectives Υ via taxation schemes, and also discussed some issues surrounding equitable taxation.

Our work relates to a number of other topics in the multi-agent systems community and beyond. Some consideration has been given to how a principal can change the equilibrium strategies of *specific* games by introducing penalties (a form of taxation) on some actions of the players. Interesting applications include information security [13] and analyzing the TCP protocol. In the multi-agent systems community, Monderer and Tennenholtz proposed the notion of k implementation [8], whereby a principal can make payments to players (negative taxes) to incentivise players to choose certain outcomes. The setting for k-implementation is one of payments, in contrast to the present paper, and our use of Boolean games and logical objectives Υ is rather different. A related idea is discussed in [1], which considers how much compensation would have to be paid to players in a cooperative game in order for certain outcomes to become core stable.

We believe the results of the present paper strongly indicate that there are important and interesting theoretical and practical questions relating to non-incentive compatible taxation schemes. Future work might consider, for example: a complete characterisation of the conditions under which an objective Υ can be implemented in a game G; consideration of the computation of taxation schemes τ for objectives Υ; and the use of taxation schemes to incentivise behaviour in other settings, beyond the Boolean games considered in the present paper.

7. REFERENCES

[1] Y. Bachrach, E. Elkind, R. Meir, D. Pasechnik, M. Zuckerman, J. Rothe, and J. S. Rosenschein. The cost of stability in coalitional games. In *Proceedings SAGT 2009*, 2009.

[2] E. Bonzon, M.-C. Lagasquie, J. Lang, and B. Zanuttini. Boolean games revisited. In *Proceedings of the Seventeenth European Conference on Artificial Intelligence (ECAI-2006)*, Riva del Garda, Italy, 2006.

[3] J. J. Cordes. Horizontal equity. In R. D. Ebel J. J. Cordes and J. G. Gravelle, editors, *The Encyclopedia of Taxation and Tax Policy*. Urban Institute Press, 1999.

[4] P. E. Dunne, S. Kraus, W. van der Hoek, and M. Wooldridge. Cooperative boolean games. In *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2008)*, Estoril, Portugal, 2008.

[5] E. Ephrati and J. S. Rosenschein. The Clarke tax as a consensus mechanism among automated agents. In *Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91)*, Anaheim, CA, 1991.

[6] P. Harrenstein, W. van der Hoek, J.-J.Ch. Meyer, and C. Witteveen. Boolean games. In J. van Benthem, editor, *Proceeding of the Eighth Conference on Theoretical Aspects of Rationality and Knowledge (TARK VIII)*, pages 287–298, Siena, Italy, 2001.

[7] E. Maskin. The theory of implementation in Nash equilibrium: A survey. MIT Department of Economics Working Paper, 1983.

[8] D. Monderer and M. Tennenholtz. K-implementation. *Journal of AI Research*, 21:37–62, 2004.

[9] H. Moulin. *Axioms of Cooperative Decision Making*. Cambridge University Press: Cambridge, England, 1988.

[10] N. Nisan and A. Ronen. Algorithmic mechanism design. In *Proceedings of the Thirty-first Annual ACM Symposium on the Theory of Computing (STOC-99)*, pages 129–140, May 1999.

[11] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, editors. *Algorithmic Game Theory*. Cambridge University Press: Cambridge, England, 2007.

[12] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. The MIT Press: Cambridge, MA, 1994.

[13] W. Sun, X. Kong, D. He, and X. You. Information Security Game Analysis with Penalty Parameter. In *Electronic Commerce and Security, 2008 International Symposium on*, pages 453–456, 2008.

[14] W. van der Hoek and M. Wooldridge. On the logic of cooperation and propositional control. *Artificial Intelligence*, 164(1-2):81–119, May 2005.