Characterising the Manipulability of Boolean Games

Paul Harrenstein
University of Oxford
United Kingdom
paul.harrenstein@cs.ox.ac.uk

Paolo Turrini
Imperial College London
United Kingdom
paolo.turrini@imperial.ac.uk

Michael Wooldridge
University of Oxford
United Kingdom
mjw@cs.ox.ac.uk

Abstract

The existence of (Nash) equilibria with undesirable properties is a well-known problem in game theory, which has motivated much research directed at the possibility of mechanisms for modifying games in order to eliminate undesirable equilibria, or introduce desirable ones. Taxation schemes are one mechanism for modifying games in this way. In the multi-agent systems community, taxation mechanisms for incentive engineering have been studied in the context of Boolean games with costs. These are games in which each player assigns truth-values to a set of propositional variables she uniquely controls with the aim of satisfying an individual propositional goal formula; different choices for the player are also associated with different costs. In such a game, each player prefers primarily to see the satisfaction of their goal, and secondarily, to minimise the cost of their choice. However, within this setting – in which taxes operate on costs only – it may well happen that the elimination or introduction of equilibria can only be achieved at the cost of simultaneously introducing less desirable equilibria or eliminating more attractive ones. Although this framework has been studied extensively, the problem of precisely characterising the equilibria that may be induced or eliminated has remained open. In this paper we close this problem, giving a complete characterisation of those mechanisms that can induce a set of outcomes of the game to be exactly the set of Nash equilibrium outcomes.

1 Introduction

Game theory is widely used in multi-agent systems and artificial intelligence to model and understand the behaviour of systems in which components are assumed to act in pursuit of individual preferences [Shoham and Leyton-Brown, 2008]. Probably the most important analytical concept in game theory is the notion of Nash equilibrium, representing a state of affairs such that no participant has any rational incentive to deviate. However, a standard problem in game theory is that strategic scenarios may have Nash equilibria with undesirable properties. In the Prisoner's Dilemma, for example,

r:

	q	$\bar{q}$
p	$0,\underline{3},\underline{1}$	$0,0,0$
$\bar{p}$	$1,1,\underline{2}$	$\underline{5},0,\underline{1}$

$\bar{r}$:

	q	$\bar{q}$
p	$0,3,1$	$\underline{0},\underline{0},\underline{0}$
$\bar{p}$	$1,1,1$	$5,0,1$

Figure 1: A three-player Boolean game with costs. The row player controls proposition p, the column player q, and the matrix player r. The numerical entries refer to the costs incurred by, respectively, the row, column, and matrix player at the corresponding outcome. An entry being underlined indicates that the corresponding player has her goal satisfied and Nash equilibria are encircled. The shaded cells are equilibria under *no* taxation scheme.

the unique Nash equilibrium outcome is strictly worse for all players than an alternative outcome. This problem – the presence of undesirable equilibria – has motivated research on the development of mechanisms for modifying games, with the goal of either eliminating undesirable equilibria, or inducing desirable ones.

Taxation mechanisms represent one natural class of mechanisms for manipulating games. The idea is that by levying taxes on the actions of agents, it is possible to incentivise an agent to avoid or choose a particular action (cf., e.g., [Cordes, 1999; Tobin, 1978; Meade, 1952; Coase, 1960]). In the multi-agent systems literature, this idea has been investigated in the context of Boolean games with costs [Wooldridge *et al.*, 2013]. In such a game, each player exercises unique control over a set of propositional variables, in the sense that the player can choose to assign values (true or false) to these variables as they wish. Preferences in the game are defined by associating with each agent a propositional formula representing a goal that the player desires to see satisfied. Different assignments of values to variables induce different costs for the corresponding agent, and while players are primarily motivated to seek the satisfaction of their goal, they are secondarily motivated to minimise costs. Because taxation schemes can apply additional costs to actions (or subsidise actions), designing such a scheme can influence the rational behaviour of players, making it possible to eliminate some equilibria or introduce new ones. However, as players always prefer to get their goal achieved than otherwise, there is an inherent limit

	q	$\bar{q}$
p	$2, \underline{3}, \underline{1}$	$0, 0, 0$
$\bar{p}$	$1, 1, \underline{2}$	$\underline{5}, 2, \underline{1}$

r

	q	$\bar{q}$
p	$2, 3, 1$	$\underline{0}, \underline{0}, \underline{0}$
$\bar{p}$	$1, 1, 1$	$5, 2, 1$

$\bar{r}$

Figure 2: An attempt to raise taxes to eliminate the undesirable equilibria pqr and $\bar{p}\bar{q}r$ of the game in Figure 1. As a result, $\bar{p}qr$ – which is less desirable for every player than either pqr and $\bar{p}\bar{q}r$ in the original game – becomes a Nash equilibrium. Notice that no such attempt will be successful: the set $\{pqr, \bar{p}qr, \bar{p}\bar{q}r\}$ is dominating but contains no cycle involving at least two players.

to the manipulation that is possible through taxation: a player can never be incentivised away from achieving her goal.

The basic framework of taxation mechanism for Boolean games with costs was introduced by [Wooldridge *et al.*, 2013], and has since been investigated and extended by other authors [Harrenstein *et al.*, 2014; Turrini, 2013; 2016]. However, fundamental questions remain unanswered about the extent to which Boolean games can be manipulated through taxation. One issue in particular is that, while it may be possible to eliminate some undesirable equilibrium from a game, this process may inadvertently introduce new undesirable equilibria. Consider for example the Boolean game in Figure 1 with three equilibria. The top-right outcome is an equilibrium in which all players have their goals achieved at minimum cost. The other two equilibria are less desirable for all players. Yet, as illustrated in Figure 2, any attempt at eliminating these with the same taxation scheme will result in the bottom-left outcome becoming an equilibrium. This outcome, however, is worse for all players than the original two that were eliminated.

To date, the problem of precisely characterising the sets of equilibria that can be eliminated or introduced through taxation has remained open. It is this problem that we address, and settle, in the present paper: we give for the first time a complete characterisation of the sets of outcomes in Boolean games with costs that may be induced or eliminated through taxation mechanisms.

2 Boolean Games with Costs

We use the framework of Boolean games with costs as presented in [Wooldridge *et al.*, 2013].[1] Formally, a **Boolean game with costs** (hereafter just "Boolean game" or "game") is given by a structure, $(N, \{\Phi_i\}_{i\in N}, \{\gamma_i\}_{i\in N}, c)$. Here, $N = \{1, 2, \dots, n\}$ is a set of **players** (or **agents**). Each player i uniquely controls a set of propositional variables Φ_i and is associated with a **goal** γ_i, a propositional logic formula constructed from the total set of propositional variables $\Phi = \Phi_1 \cup \dots \cup \Phi_n$. The sets $\Phi_1, \dots, \Phi_n$ are assumed to partition Φ, ensuring that each variable is controlled by precisely one player. A **choice** (or **strategy** or **action**) for player i is a an assignment of truth or falsity to all variables that i controls, that is, a function of the form $v_i : \Phi_i \to \{\top, \bot\}$. The set of all such choices for player i is denoted V_i. Players, independently and simultaneously, make individual choices, giving rise to a **profile (of choices)** of the form $(v_1, \dots, v_n)$. The set of profiles is denoted V and we let (v_{-i}, v_i') abbreviate the profile $(v_1, \dots, v_{i-1}, v_i', v_{i+1}, \dots, v_n)$. Each profile $(v_1, \dots, v_n)$ naturally defines to a unique **valuation** $v\colon \Phi \to \{\top, \bot\}$ for the total set of propositional variables. We write $p\bar{q}\bar{r}$ to denote the profile in which variable p is set to true and variables q and r are set to false, and similarly for other valuations. We will generally not notationally distinguish between profiles and valuations. For a profile v and formula φ over Φ, we will thus write $v \models \varphi$ to signify that v satisfies φ, where $\models$ is the standard propositional satisfaction relation. We also say that the valuation $v = v_1 \cup \dots \cup v_n$ is the **outcome** that results if profile $(v_1, \dots, v_n)$ is played.

At each profile every player incurs a cost. These costs are modelled by an **(outcome-based) cost function** c, which for each player i specifies a function $c_i : V \to \mathbb{Q}_{\geq 0}$, associating a non-negative rational cost with each profile. (Here we deviate from the more restrictive additive notion of cost in [Wooldridge *et al.*, 2013], where costs are associated with setting a propositional variables to a specific Boolean value: the present model is more expressive.) If B is a Boolean game with cost function c and d is another cost function, then B^d denotes the Boolean game that results from B by replacing c by d. We denote by c^0 the **zero-cost function**, which assigns cost 0 to every player i and every valuation, that is, $c_i^0(v) = 0$ for all players i and all valuations v. Furthermore, we write B^0 for B^{c^0} to avoid cluttered notation.

As discussed in the introduction, players prioritise goal realisation over cost minimisation, that is, each will prefer outcomes that satisfy her goal to outcomes that do not, no matter the respective costs, and prefer cheaper outcomes to more expensive ones, otherwise. Accordingly, costs refine the dichotomous preferences each player has over the outcomes on basis of her goal alone. Formally, we model the preferences of player i as a complete and transitive relation over outcomes. Thus, given a Boolean game with cost function c and a player i with goal γ_i, we say that i **weakly prefers** outcome v to outcome v', in symbols $v \succeq^c v'$, if

(*i*) $v \models \gamma_i$ and $v' \not\models \gamma_i$, or

(*ii*) both $v \models \gamma_i$ if and only if $v' \models \gamma_i$, and $c_i(v) \leq c_i(v')$.

We use $\succ_i^c$ and $\sim_i^c$ for, respectively, the **strict** and **indifferent** parts of $\succeq_i^c$ in the usual way, and write $\succeq_i^0$, $\succ_i^0$, and $\sim_i^0$ if c is c^0. Observe that $v \succsim_i^0 v'$ if and only if $v' \models \gamma_i$ implies $v \models \gamma_i$. As a consequence $\succeq_i^c$ refines $\succeq_i^0$, in the sense that $\succeq_i^c$ is a subset of $\succeq_i^0$ and thus $v \succeq_i^c v'$ implies $v \succeq_i^0 v'$.[2]

Defined thus, Boolean games represent strategic games and as such the standard solution concepts from game theory are

[1] The framework of Boolean games was initiated by [Harrenstein *et al.*, 2001]. Important follow-up work includes [Bonzon *et al.*, 2006; 2009], [Dunne *et al.*, 2008], [Grant *et al.*, 2011], [Mavronicolas *et al.*, 2015], and [Clercq *et al.*, 2015].

[2] The opposite direction also holds: for every preference (i.e., reflexive, transitive, and complete) relation $\succeq_i$ over the outcomes that refines $\succeq_i^0$, there is an outcome-based cost function c such that $\succeq_i$ equals $\succeq_i^c$. This, however, does not generally hold for the additive cost functions of [Wooldridge *et al.*, 2013].

available for their analysis. The solution concept we work with is pure strategy Nash equilibrium. Formally, a profile $v = (v_1, \ldots, v_n)$ of a Boolean game B with cost function c is a **(pure strategy) Nash equilibrium** if for all players i and all choices $v_i' \in V_i$, we have:

$$v \succeq_i^c (v_1, \ldots, v_{i-1}, v_i', v_{i+1}, \ldots, v_n).$$

We denote the set of pure Nash equilibria of a Boolean game B by $NE(B)$. Pure Nash equilibria are not guaranteed to exist, and, when they exist, they need not be unique.

The Nash equilibria of a Boolean game depend both on the players' goals and on the cost function. Thus, [Harrenstein *et al.*, 2014] distinguished hard, soft, and initial equilibria. A profile v is an **initial equilibrium** of a Boolean game B if it is an equilibrium of B^0. Having observed that $\succeq_i^c$ refines $\succeq_i^0$ for every cost function c and every player i, it follows that generally $NE(B^c) \subseteq NE(B^0)$. Accordingly, profile v is an initial equilibrium of a Boolean game B if and only if v is an equilibrium of B^c for some cost function c. By contrast, a **hard equilibrium** of B is a profile v that is a Nash equilibrium in B^c for *every* cost function c. Finally, v is a **soft equilibrium** of B if it is an initial equilibrium of B that is not hard. We denote the initial, hard, and soft equilibria of game B by $\textsc{init}(B)$, $\textsc{hard}(B)$, and $\textsc{soft}(B)$, respectively. It is easy to see that the stability of hard equilibria only depends on the players' goals, whereas it is the cost function that determines whether a soft equilibrium is also actually a Nash equilibrium of a given game. For example, the initial equilibria of the game in Figure 1 are pqr, $\bar{p}qr$, $\bar{p}\bar{q}r$, and $p\bar{q}\bar{r}$; while the former three are soft, the latter is hard.

3 Manipulating Boolean Games

Over the past few years, a number of contributions have focussed on understanding how a system designer can incentivise players in a Boolean game by manipulating the cost function by levying taxes. This line of research started with [Wooldridge *et al.*, 2013] and was pursued further by, e.g., [Harrenstein *et al.*, 2014; 2016]. For the purposes of this paper, we have an **(outcome-based) taxation scheme** τ associate with every player i a function τ_i that maps each profile to a non-negative rational number, that is,

$$\tau_i \colon V \to \mathbb{Q}_{\geq 0}.$$

A taxation scheme τ modifies the cost function of a Boolean game, meaning that it increases the cost for every player i at every profile v according to τ. Given Boolean game B with cost function c, a taxation scheme τ gives rise to a cost function $c^\tau = c + \tau$ defined such that, for each player i and each profile v,

$$c_i^\tau(v) = c_i(v) + \tau_i(v).$$

Thus, the application of a taxation scheme τ to a Boolean game B with cost function c results in the Boolean game B^{c^τ} with cost function c^τ. Every taxation scheme induces a cost transformation and, moreover, every cost transformation can be achieved by a taxation scheme modulo positive affine transformations. We write B^τ for B^{c^τ} and $v \succeq_i^\tau v'$ for $v \succeq_i^{c^\tau} v'$, whenever c is known from the context. For τ and τ' taxation schemes, we write $\tau + \tau'$ for the taxation scheme τ'' such that $\tau_i''(v) = \tau_i(v) + \tau_i'(v)$.

Previous work on incentive engineering for Boolean games, in particular [Wooldridge *et al.*, 2013], has focussed on taxation schemes that eliminate individual soft equilibria or induce such to become Nash equilibria. However, if v can be eliminated under some taxation scheme and v' under another, it does not generally follow that v and v' can both be eliminated by the *same* taxation scheme. As we saw in Figures 1 and 2, the elimination of some (undesirable) equilibrium may inevitably result in another, perhaps worse, equilibrium coming into being.

For Boolean game B and set of profiles X, we say that a taxation scheme τ **induces** X if $NE(B^\tau) = X$ and **eliminates** X if no $v \in X$ is a Nash equilibrium in B^τ, that is, if $NE(B^\tau) \cap X = \emptyset$. Our main research question can then be phrased as *which sets of profiles can be eliminated by some (outcome-based) taxation scheme and which sets can be induced*. In Section 4, we answer this question by providing characterisations of inducible and eliminating sets in Boolean games that do not involve cost functions or taxation schemes and only pertains to the players' goals. For reasons of space, we will omit some of the easier proofs.

We find that attention can be restricted to games with the zero cost function c^0. Let B be a game with cost function c and τ a taxation scheme. Recall that then $NE(B^\tau) \subseteq NE(B^0)$. Moreover, for every taxation scheme τ, there is another taxation scheme τ' such that $c + \tau = c^0 + \tau'$. Similarly, for every τ' there is a τ and constant k and such that $c^0 + \tau' + k = c + \tau$. As raising costs by a constant everywhere does not affect the preference structure, we have the following lemma.

Lemma 1. *Let B be a Boolean game with cost function c and $X \subseteq V$ a subset of profiles. Then,*

(i) *X is inducible in B if and only if X is inducible in B^0,*

(ii) *X is eliminable in B if and only if X is eliminable in B^0.*

4 Characterising Inducible and Eliminable Sets

[Harrenstein *et al.*, 2016] identified necessary and sufficient conditions for hard, soft, and initial equilibria that only depend on the players' goals and not on the cost functions. In order to achieve the same for inducible and eliminable sets of equilibria, we first distinguish *initial*, *soft*, and *hard* deviations. Let v, v' be two *distinct* outcomes and i a player and $v' = (v_{-i}, v_i'')$ for some $v_i'' \in V_i$. We then say that:

- v' is an **initial deviation** for i from v, (denoted $v \rightarrow_i v'$), if $v \models \gamma_i$ implies $v' \models \gamma_i$,
- v' is a **soft deviation** for i from v, (denoted by $v \leftrightarrows_i v'$), if both $v \rightarrow_i v'$ and $v' \rightarrow_i v$.
- v' is a **hard deviation** for i from v, (denoted by $v \rightrightarrows_i v'$), if $v \rightarrow_i v'$ but not $v' \rightarrow_i v$.

Thus, $\rightrightarrows_i$ and $\leftrightarrows_i$ partition the initial deviation relation $\rightarrow_i$ into a strict and a non-strict part, respectively. Also observe that $\rightarrow_i$, $\leftrightarrows_i$, and $\rightrightarrows_i$ are independent of the cost function and only depend on i's goal formula. In particular, we have that $v \rightrightarrows_i v'$ if and only if $v \not\models \gamma_i$ and $v' \models \gamma_i$. The following lemma, moreover, is an immediate consequence of the lexicographic nature of preferences in Boolean games with costs.

Lemma 2. *Let B be a Boolean game. Then, for all* distinct *outcomes $v, v' \in V$, each of the following hold.*

(*i*) *$v \rightarrow_i v'$ if and only if $v' \succ_i^c v$ for some cost function c,*

(*ii*) *$v \leftrightarrows_i v'$ if and only if $v' \succ_i^c v$ for some cost function c and $v \succ_i^{c'} v'$ for another cost function c',*

(*iii*) *$v \rightrightarrows_i v'$ if and only if $v' \succ_i^c v$ for all cost functions c.*

Accordingly, a profile v is an initial equilibrium if and only if there is no hard deviation from v, and that v is a hard equilibrium if and only if there is no initial deviation from v. A soft equilibrium, moreover, is an initial equilibrium from which there is at least one soft deviation. As our main result we show that in a similar way we can characterise inducible and eliminable sets in terms of the initial deviation relations $\rightarrow_i$.

We first generalise the concept of a hard equilibrium to sets of profiles and say that a *nonempty* set X is **dominating** if for all $v \in X$, all players i, and all $v_i'' \in V_i$ such that $(v_{-i}, v_i'') \notin X$, it holds that $(v_{-i}, v_i'') \rightrightarrows_i v$. Equivalently, a nonempty set X is dominating if and only if there is *no* player i for whom there is an initial deviation from some v in X to some profile v' outside X. In the game of Figure 2, we thus find that $\{pqr, \bar{p}qr, \bar{p}\bar{q}r\}$ is a dominating set, because $p\bar{q}r \rightrightarrows_1 \bar{p}\bar{q}r$, $p\bar{q}r \rightrightarrows_2 pqr$, as well as $pq\bar{r} \rightrightarrows_3 pqr$, $\bar{p}q\bar{r} \rightrightarrows_3 \bar{p}qr$, and $\bar{p}\bar{q}\bar{r} \rightrightarrows_3 \bar{p}\bar{q}r$. It can easily be seen that V is (trivially) a dominating set and that dominating sets are closed under union. By a **(set-inclusion) minimal dominating set** we understand a subset of outcomes that is dominating and contains no dominating sets as strict subsets.

A **cycle in** X, moreover, we define as a sequence $v^0, v^1, \ldots, v^k$ of $k \geq 3$ distinct profiles in X such that $v^0 = v^k$ and $v^0 \rightarrow_{i_1} v^1 \rightarrow_{i_2} \ldots \rightarrow_{i_k} v^k$ for some players $i_1, \ldots, i_k \in N$. We say that a cycle $v^0, \ldots, v^k$ in X **involves player** i if $i = i_m$ for some $1 \leq m \leq k$. Thus, in Figure 2, the set $\{pqr, \bar{p}qr, \bar{p}\bar{q}r\}$ can be seen to contain no cycle. In the game depicted in Figure 3, however, $\bar{p}q\bar{r}s, \bar{p}\bar{q}\bar{r}s, \bar{p}\bar{q}\bar{r}\bar{s}, \bar{p}q\bar{r}\bar{s}, \bar{p}q\bar{r}s$ is a cycle in V, because

$$\bar{p}q\bar{r}s \rightarrow_1 \bar{p}\bar{q}\bar{r}s \rightarrow_3 \bar{p}\bar{q}\bar{r}\bar{s} \rightarrow_1 \bar{p}q\bar{r}\bar{s} \rightarrow_3 \bar{p}q\bar{r}s.$$

After a couple of technical lemmas, we will be in a position to prove that a set X of profiles is eliminable if and only if all dominating sets in X contain a cycle involving at least two players (Theorem 10) and that X is inducible if and only if X is a subset of initial equilibria and the complement $V \setminus X$ is eliminable (Theorem 11).

4.1 Characterisation

The characterisation of inducible sets relies on the characterisation of eliminable sets and we concentrate on the latter first. We thus find that sets that do not contain any dominating sets are always eliminable. Observe that this excludes the set V of all profiles, which is dominating itself.

Lemma 3. *Let B be a Boolean game and assume $X \subseteq V$ contains no dominating sets. Then, X is eliminable.*

Proof. By Lemma 1, we may assume that $B = B^0$. If X is empty we are done immediately. Now assume that X is nonempty. As V is dominating, $X \neq V$. Hence, $V \setminus X$ is non-empty. We now construct inductively a sequence $X^1, X^2, X^3 \ldots$ of subsets of V such that, for every $m \geq 1$,

$$X^1 = \{v \in X : v \rightarrow_i v' \text{ for some } v' \in V \setminus X \text{ and } i \in N\},$$

$$X^{m+1} = \{v \in X : v \rightarrow_i v' \text{ for some } v' \in X^m \text{ and } i \in N\} \cup X^m.$$

Observe that $X^1 \subseteq X^2 \subseteq X^3 \subseteq \cdots$. Moreover, as X is nonempty and not dominating, $X^1 \neq \emptyset$. Also, for every $m \geq 1$, we have that $X^m \subseteq X$ and $X \setminus X^m$ is not dominating. Accordingly, $X^m \subsetneq X^{m+1}$ provided that $X^m \neq X$. As V is finite, it follows that $\bigcup_{m \geq 1} X^m = X$. Define τ such that, for each $v \in V$ and each player i,

$$\tau_i(v) = \begin{cases} \min\{m \geq 1 : v \in X^m\} & v \in X, \\ 0 & \text{otherwise.} \end{cases}$$

Now consider an arbitrary $v \in X$. Then, there is a *minimal* $m \geq 1$ with $v \in X^m$. If $m = 1$, there is a player i and a $v' \in V \setminus X$ such that $v \rightarrow_i v'$ and hence $v \prec_i^\tau v'$. If $m > 1$, there is a player i and a $v' \in X^{m-1}$ such that $v \rightarrow_i v'$ and hence $v \prec_i^\tau v'$. We may conclude that X contains no Nash equilibria of B^τ, signifying that X is eliminable. □

We now consider the case in which the set X to be eliminated does contain dominating sets. Having assumed a finite number of profiles, every dominating set contains at least one minimal dominating set. Although dominating sets may overlap, distinct *minimal* dominating sets will be disjoint.

Lemma 4. *Let B be a Boolean game and $X, Y \subseteq V$ be overlapping dominating sets. Then, $X \cap Y$ is also a dominating set. Therefore, distinct minimal dominating sets are disjoint.*

We now introduce the following auxiliary concept. For X a set of profiles, we say a taxation scheme is **local on** X (or X**-local**) if $\tau_i(v) = 0$, for all players i and all profiles $v \notin X$. Thus, a taxation scheme is local if it only raises taxes on valuations in X and no taxes on valuations outside X. The following two lemmas show that taxes that are local on X cannot eliminate equilibria that lie outside X, and that a set X can be eliminated if and only if it can be eliminated by an X-local taxation scheme.

Lemma 5. *Let B be a Boolean game, τ be an taxation scheme that is X-local for some $X \subseteq V$, and v a profile with $v \notin X$. Then, $v \in NE(B)$ implies $v \in NE(B^\tau)$.*

Lemma 6. *Let B be a Boolean game and $X \subseteq V$. Then, X is eliminable if and only if X is eliminable by an X-local taxation scheme.*

Introducing a second auxiliary concept, we say that a set X is **endogenously eliminable** if there is a taxation scheme τ such that for every outcome $v \in X$ there is a player i and $v_i' \in X$ such that $(v_{-i}, v_i') \in X$ and $v \succ_i^\tau (v_{-i}, v_i')$, that is, if τ induces profitable deviations from every outcome in X to another outcome in X. Observe that, as a consequence of Lemmas 2 and 6, dominating sets are eliminable only if they are endogenously eliminable by an X-local taxation scheme.

Lemma 7. *Let B be a Boolean game and X a dominating set. Then, X is eliminable if and only if X is endogenously eliminable by an X-local taxation scheme.*

We moreover have the following simple but useful lemma.

Lemma 8. *Let B be a Boolean game with cost function c and $Y \subseteq X \subseteq V$ such that Y is endogenously eliminable. Then, X is eliminable if and only if $X \setminus Y$ is eliminable.*

Proof. The "only if"-direction is immediate. For the opposite direction, let $\tau^{X\setminus Y}$ be a taxation scheme that eliminates $X \setminus Y$ and τ^Y one that eliminates Y endogenously. Observe that by virtue of Lemma 6 we may assume that $\tau^{X\setminus Y}$ is local on $X \setminus Y$. Then, define τ^X such that for all players i and all $v \in V$,

$$\tau_i^X(v) = \begin{cases} \tau_i^Y(v) & \text{if } v \in Y, \\ \tau_i^{X\setminus Y}(v) + \max\{\tau_i^Y(u) : \ y \in Y\} & \text{otherwise.} \end{cases}$$

Now, τ^X eliminates $X \setminus Y$ and Y, the latter endogenously. □

We can now establish necessary and sufficient conditions for the eliminability of minimal dominating sets. To appreciate our results, call a cycle $v^0, v^1, \ldots, v^k$ a **deviation cycle** if $v^k \succ_{i_k}^c v^{k-1} \succ_{i_{k-1}}^c \cdots \succ_{i_1}^c v^0$. Obviously, no profile contained in a deviation cycle can be an equilibrium, irrespective of the costs on the other profiles. The intuition underlying Lemma 9 is that, for every cycle in a minimal dominating set involving at least two players, we can find a taxation scheme that turns it into a deviation cycle and, consequently, eliminates it endogenously. Lemma 8 then guarantees that the minimal dominating set itself can be eliminated. We find, moreover, that this sufficient condition for eliminability is also necessary.

Lemma 9. *Let B be a Boolean game and X a minimal dominating set. Then, X is eliminable if and only if X contains a cycle involving at least two distinct players.*

Proof. By Lemma 1, we may assume that $B = B^0$. First assume that taxation scheme τ eliminates X. Having assumed that X is dominating, τ eliminates X endogenously. Hence, for every $v \in X$, there is a $v' \in X$ and player i, such that $v \prec_i^\tau v'$. As X is finite, it follows that there are $v^0, v^1, \ldots, v^m$ with $v^0 \prec_{i_1}^\tau v^1 \prec_{i_2}^\tau \cdots \prec_{i_m}^\tau v^m$ and $v^0 = v^m$. Then, by Lemma 2, also $v^0 \to_{i_1} v^1 \to_{i_2} \ldots \to_{i_m} v^m$, that is, $v^0, v^1, \ldots, v^m$ is a cycle in X. By assuming that $v^0, v^1, \ldots, v^m$ involves only one player i, we would obtain $v^0 \prec_i^\tau v^m$ whereas $v^0 = v^m$, a contradiction. We may therefore conclude that $v^0, v^1, v^2, \ldots, v^m$ is a cycle in X involving at least two players.

For the opposite direction, assume that X contains a cycle $v^0, v^1, \ldots, v^m$ that involves at least two players i and j. Then, $v^0 \to_{i_1} v^1 \to_{i_2} \ldots \to_{i_m} v^m$. Now define a taxation scheme τ_j for each player j as follows. As $v^0, v^1, \ldots, v^m$ involves at least two players, there is a least $1 \le k \le m$ such that $j \ne i_k$. Let

$$w^1, \ldots, w^m = v^k, \ldots, v^m, v^1, \ldots, v^{k-1}$$

and then define $\tau_j(w^{m'}) = m - m'$ for every $1 \le m' \le m$. Accordingly, $w^1 \prec_j^\tau \cdots \prec_j^\tau w^m$. Intuitively, the cycle is linearised by breaking it at the kth edge and decreasing taxes on j are raised along $v^k, \ldots, v^m, v^1, \ldots, v^{k-1}$. This induces deviations for j. Thus, j incurs low taxes at v^{k-1} and relatively high ones at v^k and j does not want to deviate from v^{k-1} to v^k. That does not matter, however, as exactly that deviation is to be performed by player i_k, which was assumed to be distinct from j. Having defined τ in this way, it follows that $v^0 \prec_{i_1}^\tau v^1 \prec_{i_1}^\tau \cdots \prec_{i_m}^\tau v^m$. Accordingly, τ eliminates $\{v^0, v^1, \ldots, v^m\}$ endogenously. To finish the proof, recall that we had assumed X to be minimally dominating. Hence, the set $X \setminus \{v^0, v^1, \ldots, v^m\}$ contains no dominating set and is, by virtue of Lemma 3, eliminable. As, moreover, τ endogenously eliminates $\{v^0, v^1, \ldots, v^m\}$, it follows by Lemma 8 that X is eliminable itself. □

The results obtained so far put us in a position to prove our first main result and provide a complete characterisation of eliminable sets of equilibria. Here, the crucial part is to observe that, if two minimal dominating sets are eliminable separately, there is also a taxation scheme that eliminates them both.

Theorem 10. *Let B be a Boolean game and $X \subseteq V$. Then, X is eliminable if and only if every minimal dominating subset $Y \subseteq X$ contains a cycle involving at least two players.*

Proof. For the "only if"-direction, let Y be a minimal dominating subset of X and assume for contraposition that Y contains no cycle involving at least two players. By Lemma 9, it follows that Y is not eliminable and, hence, neither is X.

For the opposite direction, let $Y^1, \ldots, Y^m$ be all the minimal dominating sets contained in X and assume for each $1 \le k \le m$ that Y^k contains a cycle involving at least two players. By Lemmas 7 and 9, each Y^k is endogenously eliminable by some Y_k-local taxation scheme τ^k. Define τ^Y such that, for every profile v and player i,

$$\tau_i^Y(v) = \tau^{Y_1}(v) + \cdots + \tau_i^{Y_k}(v).$$

Observe that τ^Y is local on $Y = Y_1 \cup \cdots \cup Y_k$. By virtue of Lemma 4, moreover, the sets $Y^1, \ldots, Y^m$ are pairwise disjoint and, therefore, $\tau_i^Y(v) = \tau^{Y_k}(v)$ if and only if $v \in Y_k$. Some reflection then reveals that τ^Y eliminates Y endogenously. Finally observe that $X \setminus Y$ contains no dominating sets and thus, by Lemma 3, is eliminable. By Lemma 8, we may conclude that X is eliminable as well, as desired. □

Our second main result, which provides a characterisation of inducible sets, now follows as a corollary of Theorem 10.

Theorem 11. *Let B be a Boolean game and $X \subseteq V$. Then, X is inducible if and only if $X \subseteq \textsc{init}(B)$ and every minimal dominating subset $Y \subseteq V \setminus X$ contains a cycle involving at least two players.*

Proof. For the "only if"-direction, assume that there is a taxation scheme τ such that $NE(B^\tau) = X$. Then, immediately, $X \subseteq \textsc{init}(B)$. Moreover, τ eliminates $V \setminus X$ and hence, by Theorem 10, every minimal dominating set in $V \setminus X$ contains a cycle involving at least two players.

For the opposite direction, assume that $X \subseteq \textsc{init}(B)$ and every minimal dominating subset $Y \subseteq V \setminus X$ contains a cycle involving at least two players. We show that X can be induced in B^0. Lemma 1 then gives the result.

By Theorem 10, we find that $V \setminus X$ is eliminable in B^0. By Lemma 6, moreover, $V \setminus X$ is eliminable in B^0 by a $V \setminus X$-local taxation scheme τ. Now observe that $X \subseteq NE(B^0)$. Accordingly, Lemma 5 yields $X \subseteq NE(B^{0+\tau})$. Hence, $NE(B^{0+\tau}) = X$, that is, X is inducible in B^0, as desired. □

	r	$\bar{r}$
pq	$\underline{3}, \underline{1}, \underline{1}$	$\underline{3}, 0, \underline{1}$
$p\bar{q}$	$\underline{2}, \underline{1}, 1$	$2, 0, \underline{1}$
$\bar{p}q$	$1, 1, 1$	$1, \underline{0}, \underline{1}$
$\bar{p}\bar{q}$	$0, 1, 1$	$0, \underline{0}, 1$

s

	r	$\bar{r}$
pq	$3, \underline{1}, 0$	$3, 0, 0$
$p\bar{q}$	$\underline{2}, \underline{1}, \underline{0}$	$2, 0, 0$
$\bar{p}q$	$1, 1, \underline{0}$	$\underline{1}, \underline{0}, 0$
$\bar{p}\bar{q}$	$\underline{0}, 1, \underline{0}$	$\underline{0}, \underline{0}, \underline{0}$

$\bar{s}$

Figure 3: A Boolean game with additive costs in which the set $V \setminus \{pq\bar{r}\bar{s}, p\bar{q}\bar{r}\bar{s}, \bar{p}qrs, \bar{p}\bar{q}rs\}$ can be eliminated by an outcome-based taxation scheme, but not by an additive one.

	r	$\bar{r}$
pq	$\underline{3}, \underline{1}, \underline{1}$	$\underline{3}, 0, \underline{1}$
$p\bar{q}$	$\underline{2}, \underline{1}, 1$	$2, 0, \underline{1}$
$\bar{p}q$	$1, 1, 1$	$1, \underline{0}, 1$
$\bar{p}\bar{q}$	$0, 1, 1$	$0, \underline{0}, \underline{1}$

s

	r	$\bar{r}$
pq	$3, \underline{1}, 0$	$3, 0, 0$
$p\bar{q}$	$\underline{2}, \underline{1}, \underline{0}$	$2, 0, 0$
$\bar{p}q$	$\underline{1}, 1, \underline{0}$	$\underline{1}, \underline{0}, \underline{0}$
$\bar{p}\bar{q}$	$0, 1, \underline{0}$	$\underline{0}, \underline{0}, 0$

$\bar{s}$

Figure 4: The game of Figure 3 with players' goal satisfaction in rows $\bar{p}q$ and $\bar{p}\bar{q}$ interchanged.

4.2 On Additive Costs and Taxes

Our definitions of outcome-based cost functions and taxation schemes diverge slightly from the additive variants used by [Wooldridge *et al.*, 2013]. Both **additive cost-functions** c and **additive taxation schemes** τ map pairs in $\Phi \times \{\bot, \top\}$ to a non-negative rational and, respectively, define outcome-based cost-functions $\hat{c}$ and outcome-based taxation schemes $\hat{\tau}$ such that, for every player i and outcome v,

$$\hat{c}_i(v) = \sum_{p \in \Phi_i :\, v(p)=\top} c(p,\top) + \sum_{p \in \Phi_i :\, v(p)=\bot} c(p,\bot) \quad \hat{\tau}_i(v) = \sum_{p \in \Phi_i :\, v(p)=\top} \tau(p,\top) + \sum_{p \in \Phi_i :\, v(p)=\bot} \tau(p,\bot).$$

The structure an additive cost function imposes on the valuations is more regular than that of outcome-based cost functions but nevertheless highly non-trivial.[3] The problem of characterising eliminable and inducible sets of equilibria is correspondingly more complicated in this setting.

Consider the Boolean game in Figure 3, where the additive cost function c is such that $c(p,\top) = 2$, $c(q,\top) = c(r,\top) = c(s,\top) = 1$, and $c(x,\bot) = 0$ for all variables x. The set $X = V \setminus \{pq\bar{r}\bar{s}, p\bar{q}\bar{r}\bar{s}, \bar{p}qrs, \bar{p}\bar{q}rs\}$ can be eliminated by the outcome-based taxation scheme τ' that levies taxes of 2 on the row player at $\bar{p}\bar{q}\bar{r}\bar{s}$ and nil taxes otherwise. By contrast, X is *not* eliminable by any *additive* taxation scheme. To see this, observe that for any such scheme τ should eliminate the equilibrium $\bar{p}\bar{q}\bar{r}\bar{s}$. Hence, $\hat{\tau}_{row}(\bar{p}\bar{q}\bar{r}\bar{s}) - \hat{\tau}_{row}(\bar{p}q\bar{r}\bar{s}) > 1$. This, however, would imply that $\tau(q,\bot) - \tau(q,\top) > 1$. Hence, $pqrs \succ^{\hat{\tau}}_{row} p\bar{q}rs$, causing $pqrs$ to be an equilibrium under $\hat{\tau}$.

Interestingly, if we consider the game in Figure 4, which results from the one in Figure 3 when the outcomes the rows $\bar{p}q$ and $\bar{p}\bar{q}$ are interchanged with respect to the players' goal satisfaction, we find that the set X is eliminable by the additive taxation scheme that levies zero taxes all around. Yet, the graphs on the valuations induced by the initial deviation relations $\rightarrow_i$ in both games are identical up to permuting the valuations. Consequently, eliminability, and therewith inducibility, of sets of outcomes by additive taxation schemes does not only depend on the (graph theoretic) structure of the initial deviation relations $\rightarrow_i$, but also on the very propositions that are set to true or false in the valuations. This reveals a fundamental mathematical distinction between the outcome-based and additive settings.

[3]This issue is closely related to *additive conjoint measurement* as studied in measurement theory (see, e.g., [Suppes and Zinnes, 1963; Krantz *et al.*, 1971; Roberts, 1979; Slinko, 2009]).

5 Conclusion

We have studied equilibrium elimination and introduction via incentive engineering in the context of Boolean games. The problem of fully characterising the conditions under which this is possible was an open problem we inherited from the related literature, notably [Wooldridge *et al.*, 2013], which had only focussed on induction and elimination strategies for single Nash equilibria. We have settled the problem for the general case outcome-based taxation mechanisms (that is, unconstrained transformations of the players' cost function). Our characterisations reduce to the presence of cycles of possible (that is, initial) deviations in some fully separated subsets of outcomes. Still, a number of research questions remain.

First, there is need to characterise eliminability and inducibility under the more restrictive additional taxation mechanisms. We observed how the characterisation under outcome-based taxation mechanisms does not carry over to the additive setting. This does of course not show that no such characterisation can be obtained, but we have to leave the issue as an open problem. A similar point concerns the side-payment schemes as studied by [Harrenstein *et al.*, 2014].

A second point that deserves attention is how to compare the desirability of the sets of outcomes that are induced under different taxation or side-payment schemes. If viewed from the perspective of the players' welfare, this requires to raise preferences over outcomes to preferences over sets of outcomes. In the social choice literature there have been several proposals for such metrics (cf. e.g., [Barberà *et al.*, 2004]).

One of the main advantages of Boolean games lies in their computational aspects and connections with logic. A third issue is therefore to establish the computational complexity of deciding whether a given set of outcomes is inducible or eliminable by a taxation scheme. We trust our results provide deeper insight into the structure of these problems.

Acknowledgments

Michael Wooldridge and Paul Harrenstein are supported by the ERC under Advanced Grant 291528 ("RACE"). Paolo Turrini acknowledges the support of the Imperial College Research Fellowship "Designing Negotiation Spaces for Collective Decision-Making" (DoC-AI1048).

References

[Barberà *et al.*, 2004] Salvador Barberà, Walter Bossert, and Prasanta K. Pattanaik. Ranking sets of objects. In S. Barberà, P. J. Hammond, and C. Seidl, editors, *Handbook of Utility Theory*, volume II, chapter 17, pages 893–977. Kluwer Academic Publishers, 2004.

[Bonzon *et al.*, 2006] Elise Bonzon, Marie-Christine Lagasquie, Jérôme Lang, and Bruno Zanuttini. Boolean games revisited. In *Proceedings of the Seventeenth European Conference on Artificial Intelligence (ECAI-2006)*, pages 265–269, 2006.

[Bonzon *et al.*, 2009] Elise Bonzon, Marie-Christine Lagasquie, Jérôme Lang, and Bruno Zanuttini. Compact preference representation and Boolean games. *Autonomous Agents and Multi-Agent Systems*, 18(1):1–35, 2009.

[Clercq *et al.*, 2015] Sophie De Clercq, Steven Schockaert, Ann Nowé, and Martine De Cock. Multilateral negotiation in boolean games with incomplete information using generalized possibilistic logic. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-2015)*, pages 2890–2896, 2015.

[Coase, 1960] Ronald H. Coase. The problem of social cost. *Journal of Law and Economics*, pages 1–44, 1960.

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

[Dunne *et al.*, 2008] Paul E. Dunne, Sarit Kraus, Wiebe van der Hoek, and Michael Wooldridge. Cooperative Boolean games. In *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2008)*, pages 1015–1022, 2008.

[Grant *et al.*, 2011] John Grant, Sarit Kraus, Michael Wooldridge, and Inon Zuckerman. Manipulating Boolean games through communication. In *Proceedings of the 20th International Conference on Artificial Intelligence (IJCAI-2011)*, pages 210–215, 2011.

[Harrenstein *et al.*, 2001] Paul Harrenstein, Wiebe van der Hoek, John-Jules Ch. Meyer, and Cees 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, 2001.

[Harrenstein *et al.*, 2014] Paul Harrenstein, Paolo Turrini, and Michael Wooldridge. Hard and soft equilibria in Boolean games. In A. Lomuscio, P. Scerri, A. Bazzan, and M. Huhns, editors, *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2014)*, pages 845–852, 2014.

[Harrenstein *et al.*, 2016] Paul Harrenstein, Paolo Turrini, and Michael Wooldridge. Hard and soft preparation sets in Boolean games. *Studia Logica*, 104(4):813–847, 2016.

[Krantz *et al.*, 1971] David H. Krantz, R. Duncan Luce, Patrick Suppes, and Amos Tversky. *Foundations of Measurement. Volume I: Additive and Polynomial Representations*. Academic Press, 1971.

[Mavronicolas *et al.*, 2015] Marios Mavronicolas, Burkhard Monien, and Klaus W. Wagner. Weighted Boolean formula games. In *Algorithms, Probability, Networks, and Games — Scientific Papers and Essays Dedicated to Paul G. Spirakis on the Occasion of His 60th Birthday*, pages 49–86, 2015.

[Meade, 1952] James Meade. External economies and diseconomies in a competitive situation. *Economic Journal*, 62(245):54–67, 1952.

[Roberts, 1979] Fred S. Roberts. *Measurement Theory*, volume 7 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley, 1979.

[Shoham and Leyton-Brown, 2008] Yohav Shoham and Kevin Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2008.

[Slinko, 2009] Arkadii Slinko. Additive representability of finite measurement structures. In S. J. Brams, W. V. Gehrlein, and F. S. Roberts, editors, *The Mathematics of Preference, Choice and Order—Essays in Honor of Peter C. Fishburn*, Studies in Choice and Welfare, pages 113–133. Springer, 2009.

[Suppes and Zinnes, 1963] Patrick Suppes and Joseph L. Zinnes. Basic measurement theory. In R. D. Luce, R. R. Bush, and E. Galanterl, editors, *Handbook of Mathematical Psychology*, volume I, chapter 1. Wiley and Sons, 1963.

[Tobin, 1978] James Tobin. A proposal for international monetary reform. *Eastern Economic Journal*, 4(3–4):153–159, 1978.

[Turrini, 2013] Paolo Turrini. Endogenous Boolean games. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence (IJCAI-2013)*, pages 390–396, 2013.

[Turrini, 2016] Paolo Turrini. Endogenous games with goals: side-payments among goal-directed agents. *Autonomous Agents and Multi-Agent Systems*, 30(5):765–792, 2016.

[Wooldridge *et al.*, 2013] Michael Wooldridge, Ulle Endriss, Sarit Kraus, and Jérôme Lang. Incentive engineering for Boolean games. *Artificial Intelligence*, 195:418–439, 2013.