Original citation:

Permanent WRAP URL:
http://wrap.warwick.ac.uk/97681

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:
© IJCAI Organization. 2017

A note on versions:
The version presented here is a working paper or pre-print that may be later published elsewhere. If a published version is known of, the above WRAP URL will contain details on finding it.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk
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, 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
one player. A choice (or strategy or action) for player \(i\) is an assignment of truth or falsity to all variables that \(i\) controls, that is, a function of the form \(v_i : \Phi_i \rightarrow \{\top, \bot\}\). 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 \((v_1, \ldots, v_n)\). The set of profiles is denoted \(V\) and we let \((v, v')\) abbreviate the profile \((v_1, \ldots, v_n, v'_1, \ldots, v'_n)\). Each profile \((v_1, \ldots, v_n)\) naturally defines to a unique valuation \(v : \Phi \rightarrow \{\top, \bot\}\) for the total set of propositional variables. We write \(pq\) 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 \(\varphi\) over \(\Phi\), we will thus write \(v \models \varphi\) to signify that \(v\) satisfies \(\varphi\), where \(\models\) is the standard propositional satisfaction relation. We also say that the valuation \(v = v_1 \cup \cdots \cup v_n\) is the outcome that results if profile \((v_1, \ldots, v_n)\) 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 \(c_i : V_i \rightarrow \mathbb{Q}_{\geq 0}\), 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, \(c^0_i(v) = 0\) for all players \(i\) and all valuations \(v\). Furthermore, we write \(B^0\) for \(B^d\) 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 \(\gamma_i\), we say that \(i\) weakly prefers outcome \(v\) to outcome \(v'\), in symbols \(v \succeq^w v'\), if

\[(i) \quad v \models \gamma_i \text{ and } v' \nparallel \gamma_i, \text{ or}
\]

\[(ii) \quad \text{both } v \models \gamma_i \text{ and if only if } v' \models \gamma_i, \text{ and } c_i(v) \leq c_i(v').\]

We use \(\succsim^w_i\) and \(\sim^w_i\) for, respectively, the strict and indifferent parts of \(\succeq^w_i\) in the usual way, and write \(\succeq^0_i\), \(\succsim^0_i\), and \(\sim^0_i\) if \(c = c^0\). Observe that \(v \succeq^0_i v'\) if and only if \(v' \models \gamma_i\) implies \(v \models \gamma_i\). As a consequence, \(\succsim^0_i\) refines \(\succsim^1_i\), in the sense that \(\succsim^0_i\) is a subset of \(\succsim^1_i\) and thus \(v \succeq^0_i v'\) implies \(v \succeq^1_i v'\).

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

---

第 2 特性下的 Boolean Games with Costs

我们使用了 Boolean 游戏与成本的框架，如在 [Wooldridge et al., 2013]。一种特殊的 Boolean 游戏（此处仅仅为“Boolean game”或“game”）由结构 \((N, \{\Phi_i\}_{i \in N}, \{\gamma_i\}_{i \in N}, c)\) 定义，其中 \(N = \{1, 2, \ldots, n\}\) 是一组玩家（或代理）。每个玩家 \(i\) 独立地控制一组 propositions \(\Phi_i\) 并且与一组 proposition \(\gamma_i\) 关联，一组 proposition 逻辑公式由 \(\Phi_i\) 的集合的总和 \(\Phi = \Phi_1 \cup \cdots \cup \Phi_n\) 构成。集合 \(\Phi_i\) 是由部分 \(\Phi_i\) 构成的，确保每个变量被控制由精确地

---

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].
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:

\[
v \succeq_i (v_1, \ldots, v_{i-1}, v'_i, v_{i+1}, \ldots, v_n).
\]

We denote the set of pure Nash equilibria of a Boolean game \( B \) by \( \text{NE}(B) \). 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 \( \succeq_f \) refines \( \succeq_0 \) for every cost function \( c \) and every player \( i \), it follows that generally \( \text{NE}(B^0) \subseteq \text{NE}(B^c) \). 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 a game \( B \) by \( \text{IN}(B) \), \( \text{HARD}(B) \), and \( \text{SOFT}(B) \), 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 \( pqr, \overline{pqr}, \overline{pqr} \), and \( pq\overline{r} \); 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 in Section 3.1 Manipulating Boolean Games. Let \( B \) be a Boolean game with cost function \( c \) and \( \tau \) a taxation scheme. Recall that then \( \text{NE}(B^\tau) \subseteq \text{NE}(B^0) \). Moreover, for every taxation scheme \( \tau \), there is another taxation scheme \( \tau' \) such that \( c + \tau = c^0 + \tau' \). Similarly, for every \( \tau' \) there is a \( \tau \) and constant \( k \) 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 \( v'' \in V_i \). 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 \( v \Rightarrow_i v' \)), if both \( v \rightarrow_i v' \) and \( v' \rightarrow_i v \);

- \( v \) is a hard deviation for \( i \) from \( v \), (denoted \( v \equiv_i v' \)), if \( v \rightarrow_i v' \) but not \( v' \rightarrow_i v \).

Thus, \( \Rightarrow_i \) and \( \equiv_i \) partition the initial deviation relation \( \rightarrow_i \) into a strict and a non-strict part, respectively. Also observe that \( \rightarrow_i \), \( \Rightarrow_i \), and \( \equiv_i \) are independent of the cost function and only depend on \( i \)’s goal formula. In particular, we have that \( v \equiv_i v' \) 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 $v' \triangleright^r_i v$ for some cost function $c$, 
(ii) $v \triangleright_i v'$ if and only if $v' \triangleright_i v$ for some cost function $c$ and $v \triangleright^r_i v'$ for another cost function $c'$, 
(iii) $v \triangleright_i v'$ if and only if $v' \triangleright_i v$ 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 profiles $v' \in V_i$ such that $(v_{-i}, v'_i) \not\in X$, it holds that $(v_{-i}, v'_i) \triangleright_i v$. 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 profile $v$ in any profile $v'$ outside $X$. In the game depicted in Figure 2, we thus find that $\{pqr, pqr, pqr\}$ is a dominating set, because $pqr \triangleright_1 pqr$, $pqr \triangleright_2 pqr$, as well as $pqr \triangleright_3 pqr$. 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 $v^{i} \rightarrow_{i} v^{i+1}$ 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 $\{pqr, pqr, pqr\}$ can be seen to contain no cycle. In the game depicted in Figure 3, however, $pqrs, pqrs, pqrs, pqrs, pqrs$ is a cycle in $V$, because

$$pqrs \rightarrow_i pqrs \rightarrow_3 pqrs \rightarrow_1 pqrs \rightarrow_3 pqrs.$$

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, $X \not\subseteq V$. 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 $m \geq 1$,

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

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

Observe that $X^1 \subseteq X^2 \subseteq X^3 \subseteq \ldots$. Moreover, as $X$ is nonempty and not dominating, $X^1 \not\subseteq \emptyset$. Also, for every $m \geq 1$, we have that $X^m \subseteq X$ and $X \setminus X^m$ is not dominating. Accordingly, $X^m \subseteq X^{m+1}$ provided that $X^m \not\subseteq X$. As $V$ is finite, it follows that $\bigcup_{m \geq 1} X^m = X$. Define $\tau$ 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 $v \triangleright_i v'$. If $m > 1$, there is a player $i$ and a $v' \in X^{m-1}$ such that $v \rightarrow_i v'$ and hence $v \triangleright_i v'$. We may conclude that $X$ contains no Nash equilibria of $B^*$, 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 $v \not\in X$. 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, $\tau$ be an taxation scheme that is X-local for some $X \subseteq V$, and $v$ a profile with $v \not\in X$. Then, $v \in \text{NE}(B)$ implies $v \in \text{NE}(B^*)$.

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 $\tau$ such that for every outcome $v \in X$ there is a player $i$ and a profile $v'_i \in X$ such that $(v_{-i}, v'_i) \in X$ and $v \triangleright^r_i (v_{-i}, v'_i)$, that is, if $\tau$ 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 ⊆ X ⊆ V such that Y is endogenously eliminable. Then, X is eliminable if and only if X \ Y is eliminable.

Proof. The “only if”-direction is immediate. For the opposite direction, let τX,Y be a taxation scheme that eliminates X \ Y and τY one that eliminates Y endogenously. Observe that by virtue of Lemma 6 we may assume that τX,Y is local on X,Y. Then, define τX such that for all players i and all v ∈ V,
\[ τ_i^X(v) = \begin{cases} τ_i^Y(v) & \text{if } v ∈ Y, \\ τ_i^X(v) + \max\{τ_j^Y(u) : y ∈ Y\} & \text{otherwise.} \end{cases} \]

Now, τX eliminates X \ 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 v0, v1, ..., v̅ a deviation cycle if v0 ≻ τ ̅, v0 ≻ τ ̅−1, ..., v0 ≻ τ ̅. 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 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 = B0. First assume that taxation scheme τ eliminates X. Having assumed that X is dominating, τ eliminates X endogenously. Hence, for every v ∈ X, there is a v′ ∈ X and player i, such that v ≻ τ i v′. As X is finite, it follows that there are v0, v1, ..., v̅ with v0 ≻ τ ̅, v0 ≻ τ ̅−1, ..., v0 ≻ τ ̅ and v0 = v̅. Then, by Lemma 2, also v0 → v1 → v2 → ... → vn−1 → v0, that is, v0, v1, ..., v̅ is a cycle in X. By assuming that v0, v1, ..., v̅ involves only one player i, we would obtain v0 ≻ τ i v̅ whereas v0 = v̅, a contradiction. We may therefore conclude that for every profile v and player i,
\[ τ_i^X(v) = τ_i^Y(v) + \ldots + τ_i^X(v). \]

Observe that τX is local on Y = Y1 ∪ · · · ∪ Yk. By virtue of Lemma 4, moreover, the sets Y1, ..., Yk are pairwise disjoint and, therefore, τX(v) = τX(v') if and only if v ∈ Yk. Some reflection then reveals that τX eliminates Y endogenously. Finally observe that X \ 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 ⊆ V. Then, X is inducible if and only if every minimal dominating subset Y ⊆ 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(Bτ) = X. Then, immediately, X ∈ INIT(B). Moreover, τ eliminates V \ X and hence, by Theorem 10, every minimal dominating set in V \ X contains a cycle involving at least two players.

For the opposite direction, assume that X ∈ INIT(B) and every minimal dominating subset Y ⊆ X contains a cycle involving at least two players. We show that X can be induced in Bτ. Lemma 1 then gives the result.

By Theorem 10, we find that V \ X is eliminable in B0. By Lemma 6, moreover, V \ X is eliminable in Bτ by a V \ X-local taxation scheme τ. Now observe that X ⊆ NE(Bτ). Accordingly, Lemma 5 yields X ⊆ NE(Bτ). Hence, NE(Bτ+τ) = X, that is, X is inducible in Bτ, as desired.
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 \( \tau \) map pairs in \( \Phi \times \{\bot, \top\} \) to a non-negative rational and, respectively, define outcome-based cost-functions \( c \) and outcome-based taxation schemes \( \tau \) such that, for every player \( i \) and outcome \( v \),

\[
\hat{c}_i(v) = \sum_{p \in \Phi : i(p) = \top} c(p, \top) + \sum_{p \in \Phi : i(p) = \bot} \hat{c}(p, \top) + \tau(p, \bot).
\]

This structure an additive cost function imposes on the valuations is more regular than that of outcome-based cost functions but nevertheless highly non-trivial. 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 \( X = V \setminus \{pqrs, p\bar{q}s, psq, \bar{p}qr\} \) can be eliminated by the outcome-based taxation scheme \( \tau' \) that levies taxes of 2 on the row player at \( psq \) and nil taxes otherwise. By contrast, \( X \) is not eliminable by any additive taxation scheme. To see this, observe that for any such scheme \( \tau \) should eliminate the equilibrium \( pqrs \). Hence, \( \hat{c}_i(psq) = \tau(psq) \geq 1 \). This, however, would imply that \( \tau(q, \bot) - \tau(q, \top) \geq 1 \). Hence, \( pqrs \geq \tau \) psq, causing \( pqrs \) to be an equilibrium under \( \tau \).

Interestingly, if we consider the game in Figure 4, which results from the one in Figure 3 when the rows \( pq \) and \( \bar{p}q \) 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 \( \rightarrow \) 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 \( \rightarrow \), 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.

3This 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 relevant 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., [Barber`a 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 Turriti acknowledges the support of the Imperial College Research Fellowship “Designing Negotiation Spaces for Collective Decision-Making” (DoC-AI1048).
References


