Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence

Endogenous Boolean Games

Paolo Turrini

Department of Computing, Imperial College London

p.turrini@imperial.ac.uk

Abstract

In boolean games players exercise control over propositional variables and strive to achieve a goal formula whose realization might require the opponents' cooperation. Recently, a theory of *incentive engineering* for such games has been devised, where an external authority steers the outcome of the game towards certain *desirable* properties consistent with players' goals, by imposing a taxation mechanism on the players that makes the outcomes that do not comply with those properties less appealing to them. The present contribution stems from a complementary perspective and studies, instead, how boolean games can be transformed from inside, rather than from outside, by endowing players with the possibility of sacrificing a part of their payoff received at a certain outcome in order to convince other players to play a certain strategy. What we call here *endogenous boolean games* (EBGs) boils down to enriching the framework of boolean games with the machinery of side payments coming from game theory. We analyze equilibria in EBGs, showing the preconditions needed for desirable outcomes to be achieved without external intervention. Finally, making use of taxation mechanism, we show how to transform an EBG in such a way that desirable outcomes can be realized independently of side payments.

1 Introduction

Boolean games [Harrenstein *et al.*, 2001] have been studied for the past decade as a compact and computationally desirable representation of strategic interaction by means of logical formulas. Recently, a theory of *incentive engineering* for such games has been devised [Wooldridge *et al.*, 2013], where an external authority, i.e. the principal, steers the outcome of the game towards certain *desirable* properties, by imposing a taxation mechanism on the players that makes the outcomes that do not comply with those properties less appealing to them. However, boolean games are characterized by an important feature, which makes that of a principal a non trivial task: players are endowed with a goal formula, whose realization does not respond to incentives, i.e. there is no monetary compensation that can convince players to give up their goal. In all the other cases, though, players behave as cost-minimizers and desirable systemic properties satisfying players' goals have been shown to be implementable by the appropriate system of incentives.

The present contribution stems from a complementary perspective and studies, instead, how boolean games can be transformed from inside, rather than from outside, by endowing players with the possibility of sacrificing a part of their payoff received at a certain outcome in order to convince other players to play a certain strategy. What we call here *endogenous boolean games* (EBGs) boils down to enriching the framework of boolean games with the machinery of side payments coming from game theory [Jackson and Wilkie, 2005]. We illustrate our idea and, informally, the setting of EBGs in the following example.

Motivating example Consider two players, a and b, which can decide whether two light switches s_a and s_b are *on* or *off*. Let us assume player a to be in full control of s_a, player b of s_b, and that each player is unaware of the other player's final decision. Let us also assume that players have goals and actions have costs, in particular that a wants s_b to be *on*, that b wants both s_a and s_b to be *on*, and that the cost of turning s_a *on* is 5, of turning it *off* is 4, while the cost of turning s_b either *on* or *off* is 2. Finally, we assume that players always prefer to minimize the cost of the actions they take and that they always prefer outcomes satisfying their goal to outcomes that do not. It goes without saying that in a scenario of this kind player a, being indifferent between having s_a *on* or *off*, will simply look at the resulting costs and decide to turn the switch *off*. Likewise, player b will turn the switch *on* and this fact will not impose on him any extra cost. In the end, player a will have its goal satisfied, paying a cost of 4, while b will not, paying 2.

Suppose though that, before the game starts, players can commit to bear a part of the cost the opponent is incurring in at a certain outcome, should this outcome be reached. While in the previous scenario player b, although *depending* on the other player for the realization of his goal (cfr. [Bonzon *et al.*, 2009a] for a qualitative account of dependencies in boolean games), could not have any say on a's decision making, now he can adopt a richer strategy and offer a to bear some cost of s_a being *on*, say 3, while a might still refrain to bear any cost for b at any outcome. In the resulting situation both a and b will have their goal satisfied, a bearing a cost of 2 while b of 5, which is a more satisfactory solution for both players. Notice, however, that the solution is not stable, as player b has an incentive to deviate to more parsimonious offers in the preplay phase without compromising the realization of his own goal.

The added value of the analysis presented here, intuitively introduced in the example above, is two-fold:

- it complements the framework of incentive engineering for boolean games [Wooldridge *et al.*, 2013], studying those situations in which players can reach desirable properties *without* external intervention;
- it provides a quantitative resolution to dependence relations, broadly studied for the case of boolean games [Bonzon *et al.*, 2009a][Bonzon *et al.*, 2009b], allowing players to influence each other's decision making by the offer of monetary incentive.

We carry out the analysis studying EBGs in relation with both boolean games and endogenous games, focusing on the properties of the resulting equilibria.

Paper Structure In Section 2 we introduce boolean games and endogenous games. In Section 3 we bring them together, defining EBGs and studying their formal connection with related contributions in the game theory and AI literature. In Section 4 we carry out an equilibrium analysis of EBGs, showing results of pure strategy equilibrium survival. In Section 5 we integrate EBGs with taxation mechanisms, devising a procedure that ensures desirable properties to be reached. Finally, in Section 6 we wrap up the work pointing to possible future research directions.

2 Preliminaries

Boolean games

Definition 1 (Boolean Games). *A* **boolean game** *is a tuple* $(N, \Phi, c, \{\gamma_i\}_{i\in N}, \{\Phi_i\}_{i\in N})$ *where: N is a finite set of players; Φ a finite set of propositional atoms; $c : N \times \Phi \times \{\mathtt{tt}, \mathtt{ff}\} \to \mathbb{R}_+$ is a* **cost function***, associating to each player the cost he incurs in when determining the truth value of a given atomic proposition; γ_i is a boolean formula, constructed on the set Φ, denoting the goal of player i; $\Phi_i \subseteq \Phi$ is the set of atoms controlled by player i. As standard [van der Hoek and Wooldridge, 2005] we assume that for $j \neq i$, $\Phi_i \cap \Phi_j = \emptyset$ and that $\bigcup\{\Phi_i \mid i \in N\} = \Phi$, i.e. controlled atoms partition the whole space.*

A **choice** of player i is a function $v_i : \Phi_i \to \{\mathtt{tt}, \mathtt{ff}\}$, representing player i's decision concerning the atoms he controls. We denote V_i as the set of possible choices of player i. An **outcome** $v \in \prod_{i\in N} V_i$ of the boolean game $\mathcal{B}$ is a collection of choices, one per player. An outcome can be thought of as a valuation function [Blackburn *et al.*, 2001], assigning a boolean value to each propositional atom [van der Hoek and Wooldridge, 2005]. Therefore, we denote $V = \prod_{i\in N} V_i$ the set of all possible valuation functions and, for $v \in V$ and φ being a boolean formula constructed on Φ, we write $v \models \varphi$ (resp. v $\not\models \varphi$) to say that φ holds (resp. does not hold) under the valuation v. When a formula φ is satisfied by a unique valuation, we use v_φ to denote that valuation. Finally, we denote $c_i(v) = \sum_{p\in\Phi_i, v\models p} c_i(p, \mathbf{tt}) + \sum_{p\in\Phi_i, v\not\models p} c_i(p, \mathbf{ff})$ the cost that player i incurs in at valuation v.

Endogenous games As well-known, a strategic game $\mathcal{S}$ is a tuple $(N, \{\Sigma_i\}_{i\in N}, \pi)$, where N is a set of players, Σ_i a set of strategies for player i and $\pi : \prod_{i\in N} \Sigma_i \times N \to \mathbb{R}$ a utility function, assigning to each player his payoff at each strategy profile. Henceforth we abbreviate $\pi(\sigma, i)$ as $\pi_i(\sigma)$, $\prod_{i\in N} \Sigma_i$ as Σ, and denote $NE^{\Delta}(\mathcal{S})$ the set of mixed strategy Nash equilibria of strategic game $\mathcal{S}$. Following related work [Jackson and Wilkie, 2005] [1] we enrich a strategic game with a family $\{T_i\}_{i\in N}$ where, for each player i, T_i is a set of functions of the form $\tau_i : \Sigma \times N \to \mathbb{R}_+$, such that: $\tau_i(\sigma, i) = 0$ for each $\sigma \in \prod_{i\in N} \Sigma_i$. Each such function specifies how much payoff player i secures the other players in case some particular valuation obtains. We call each $\tau \in \prod_{i\in N} \tau_i$ a **transfer function** and a tuple $(\mathcal{S}, \{T_i\}_{i\in N})$ an **endogenous game**. It is useful to think of it as a game consisting of two phases:

- A preplay phase, where players simultaneously decide on their transfers to the other players;
- An actual game play, where the utility of $\mathcal{S}$ is updated taking the selected transfers into account.

Definition 2 (Update by side payments). *Let* $\mathcal{S} = (N, \{\Sigma_i\}_{i\in N}, \pi)$ *be a strategic game and let τ be a transfer function. The* **play of** τ **in** $\mathcal{S}$ *is the strategic game* $\tau(\mathcal{S}) = (N, \{\Sigma_i\}_{i\in N}, \pi')$ *where, for $\sigma \in \Sigma, i \in N$,*

$$\pi'_i(\sigma) = \pi_i(\sigma) + \sum_{j\in N}\{x \in \mathbb{R}_+ \mid \tau_j(\sigma, i) = x\}$$
$$- \sum_{j\in N}\{x \in \mathbb{R}_+ \mid \tau_i(\sigma, j) = x\}$$

In a nutshell, when a game is updated by a transfer function each player adds to his payoff at each outcome all the transfers that he receives from the other players at that outcome; subtracts from his payoff at each outcome all the transfers he makes to the other players at that outcome.

Let $\mathcal{E} = (\mathcal{S}, \{T_i\}_{i\in N})$ be an endogenous game. A pair (τ, σ), for τ being a transfer function and σ a strategy profile available at $\mathcal{S}$, is a **solution of** $\mathcal{E}$ if there is a strategy in the two phase game that is a subgame perfect equilibrium and where (τ, σ) is played on the equilibrium path. For a solution (τ, σ) of $\mathcal{E} = (\mathcal{S}, \{T_i\}_{i\in N})$ we say that σ is a **surviving equilibrium** if it is a Nash equilibrium of $\mathcal{S}$.

[1] While the approach to game transformation by side payments adopted here [Jackson and Wilkie, 2005] provides an elegant technical framework that well suits boolean games, we remind the reader of the existence of earlier related work in the game theory literature [Guttman, 1978], [Kalai, 1981], [Farrell, 1998]. Also, a more involved model of preplay negotiations in non-cooperative games, overcoming a number of limitations of the present approach, has recently been developed [Goranko and Turrini, 2012].

3 Boolean games with side payments

In this section we bring together the two approaches we have previously presented into a unified framework. We start off enriching each boolean game $\mathcal{B}$ with a family $\{B_i\}_{i\in N}$, where, for each player i, B_i is a set of functions of the form $\beta_i : V \times N \to \mathbb{R}_+$. Each such function, which we call **boolean transfer function**, specifies how much payoff player i secures the other players in case some particular valuation obtains. As in the case of endogenous games, we impose that for each each $i \in N, v \in V$, we have $\beta_i(v,i) = 0$, i.e. no player can make transfers to himself. Furthermore, for $v \in V$, we abbreviate the expression $\sum_{j\in N}\{x \in \mathbb{R}_+ \mid \beta_i(v,j) = x\} - \sum_{j\in N}\{x \in \mathbb{R}_+ \mid \beta_j(v,i) = x\}$, representing the *net loss* that player i incurs in at v under the boolean transfer function β, as $\beta_i(v)$. We call a tuple $(\mathcal{B}, \{B_i\}_{i\in N})$ an **endogenous boolean game** (EBG).

To describe equilibria and related notions in EBG we need to introduce the notion of utility, reflecting players' goals. We do this in the next paragraph, adapting the intuitions used by Wooldridge et al. [Wooldridge *et al.*, 2013] to our case, and showing how the setting of EBGs forces substantial differences with both endogenous games and boolean games with taxation mechanisms.

Utility Boolean games are characterized by dichotomous preferences, i.e. players always prefer states that satisfy their goal to states that do not. Preference dichotomy has been modelled [Wooldridge *et al.*, 2013] using a *boost* factor μ_i for states satisfying player i's goal, which gets his utility increased adding to the payoff already associated to each such state the cost and the taxes he would get at his worst possible outcome. As costs and taxes are never negative, this leads players' utility to be maximal at goal states.

In our case, however, a complication arises, demonstrating that a straightforward extension of boolean games with boolean transfer functions would not do. As the reader will have noticed, the expression $-\beta_i(v)$, describing the net gain that player i obtains from the transfer function β at v, can exceed $c_i(v)$, i.e. the cost i is incurring in at v. This means that, using side payments, players might incur in *de facto* negative costs for performing certain actions and, consequently, the boost factor associated to goal states might also be negative, disrupting preference dichotomy. To avoid this drawback we impose that *after a transfer is made, each game undergoes a positive affine transformation* [Osborne and Rubinstein, 1994], where each valuation v is assigned an extra correction factor κ. In words, when some player receives at some outcome a net gain that more than compensates his cost at that outcome, the game is automatically updated to nullify its effect and preserve preference dychotomy, by inflicting a punishment to all players at all outcomes of the same amount. [2].

[2] Clearly we had several other modelling choices at our disposal. The first that comes to mind is that of limiting the possibility of players to receive transfers not exceeding their costs. This would however put a serious constraint on players' strategic possibilities, especially in the case of three players or more, substantially forcing a coordination of offers among one's opponents.

Summing up, the utility function of a boolean game $\mathcal{B} = (N, \Phi, c, \{\gamma_i\}_{i\in N}, \{\Phi_i\}_{i\in N})$ endowed with a transfer function β is calculated using four parameters for each $i \in N$:

- a **net loss** $\beta_i(v)$, specified by the boolean transfer function for i at v;
- a **cost** $c_i(v)$, specified by the cost function for i at v.
- a **correction factor** κ, with $\kappa = max\{0, -min_{v\in V, i\in N}(c_i(v) + \beta_i(v)\}$ selecting the outcome and the player where the net gain from the transfer function most overcompensates the cost of that outcome for that player. If no such player and such outcome exist, i.e. it is never the case that an player receives offers that exceed his cost, k is set to 0.
- a **boost factor** μ_i, with $\mu_i = max_{v\in V}(c_i(v) + \beta_i(v))$, selecting the payoff of the worst outcome that can happen to player i, i.e. the updated valuation that is most costly to him;

Now we can define a dichotomous utility function for boolean games with side payments.

Definition 3 (Utility). *Let $\mathcal{B}$ be a boolean game, β a boolean transfer function and $\epsilon \in \mathbb{R}_+$ a sufficiently small positive real. The* **utility function** $u^{\beta(\mathcal{B})} : N \times V \to \mathbb{R}$ *assigning to each player the payoff $u_i^{\beta(\mathcal{B})}(v)$ he receives at valuation v is defined as follows.*

$$u_i^{\beta(\mathcal{B})}(v) = \begin{cases} \epsilon + \mu_i - (c_i(v) + \beta_i(v) + \kappa) & \text{IF } v \models \gamma_i \\ -(c_i(v) + \beta_i(v) + \kappa) & \textit{otherwise} \end{cases}$$

The definition basically says that at outcomes — i.e. valuations — satisfying players' goals, players sum to their costs and the appropriately corrected transfers made also an extra factor $\epsilon + \mu_i$. For the other outcomes instead, players receive their costs plus the appropriately corrected transfers made. The definition shows the specificity of boolean games, where players are cost-minimizers but never favour a cheap choice that does not reach a goal to an expensive one that instead does. Therefore, the expression $u_i(v)$ is always strictly positive in case $v \models \gamma_i$ and never strictly positive otherwise.

In this paper we introduce a novel feature with respect to the standard treatment of boolean games: we allow players to randomize over possible choices. This will make it possible to draw a comparison with the equilibrium existence results known for endogenous games, which rely on the well-known result on the existence of Nash equilibria with mixed strategies [Osborne and Rubinstein, 1994]. To compute expected utility we first denote $\Delta(V_i)$ the set of probability distributions over the choices of player i and, for $\delta \in \Delta(V_i)$, we denote $\delta(v_i)$ the probability that **mixed choice profile** δ assigns to v_i. We call the set $s(\delta) = \{v_i \in V_i \mid \delta(v_i) > 0\}$ the **support** of δ.

Definition 4 (Expected Utility). *Let δ be a mixed choice profile available at boolean game $\mathcal{B}$, β a boolean transfer function, $v \in s(\delta)$ a pure choice profile in the support of δ, and $\delta(v)$ the probability of v to occur according to δ. The* **expected utility** *of δ for player i is defined as follows.*

$$E_i^{\beta(\mathcal{B})}(\delta) = \sum_{v \in s(\delta)} u_i^{\beta(\mathcal{B})}(v)\delta(v)$$

Given a boolean game $\mathcal{B}$ and a boolean transfer function β we denote $NE^{\Delta}(\beta(\mathcal{B}))$ the set of its Nash equilibria and $NE(\beta(\mathcal{B}))$ its pure strategy subset. Given an EBG $\mathcal{E} = (\mathcal{B}, \{B_i\}_{i\in N})$, $\beta \in \prod_{i\in N} B_i$ and $v \in V$, we say that (β, v) is a surviving equilibrium if it is a subgame perfect equilibrium of the EBG $\mathcal{E}$ and $v \in NE(\beta^0(\mathcal{B}))$, where β^0 is the boolean transfer function where all transfers amount to 0 at every outcome. The following proposition establishes a first result on boolean games with side payments, showing that costs can be used to incorporate the effect of transfers.

Proposition 1 (Normalization). *Let β be a boolean transfer function on a boolean game $\mathcal{B}$. Then there is a boolean game $\mathcal{B}'$ such that, for all $i \in N$ and $v \in V$, we have that $u_i^{\beta(\mathcal{B})}(v) = u_i^{\beta^0(\mathcal{B}')}(v)$.*

Proof. Let $\mathcal{B} = (N, \Phi, c, \{\gamma_i\}_{i\in N}, \{\Phi_i\}_{i\in N})$, β be a boolean transfer function on $\mathcal{B}$, and let κ be the correction factor associated to $\beta(\mathcal{B})$. Let $\mathcal{B}' = (N, \Phi, c', \{\gamma_i\}_{i\in N}, \{\Phi_i\}_{i\in N})$ be a boolean game different from $\mathcal{B}$ only in the cost function and let $c'_i(v) = c_i(v) + \beta_i(v) + \kappa$. We have that for all $i \in N$ and $v \in V$, $u_i^{\beta(\mathcal{B})}(v) = u_i^{\beta^0(\mathcal{B}')}(v)$. □

In a nutshell, we can think of boolean transfer functions as cost modifiers. As we will see, the setting of EBGs allows to explicitly strategise over costs by a rational use of side payments, rather than taking them as immutable factor in a game.

Boolean Games and Strategic Games The reader may wonder what the difference actually is between a boolean game endowed with a cost function and a strategic games. After all, there seem to be a straightforward similarity between strategy profiles typical of strategic games and assignments to propositional atoms typical of boolean games. Surprisingly, a correspondence result between the two structures is not available in the literature. To fill the gap, the following propositions show that a rather straightforward correspondence exists between the two structures as long as we do not consider the iterated effect of dynamic transformations, such as taxation as done in the literature [Wooldridge *et al.*, 2013] or side payments as done here, when instead the symmetry breaks.

Proposition 2 (Static correspondence).

1. *Let $\mathcal{B} = (N, \Phi, c, \{\gamma_i\}_{i\in N}, \{\Phi_i\}_{i\in N})$ be a boolean game and β a boolean transfer function. Then there exist a strategic game $\mathcal{S} = (N, \{\Sigma_i\}_{i\in N}, \pi)$ and a bijective function $f : V \to \prod_{i\in N} \Sigma_i$ such that for all $i, u_i^{\beta(\mathcal{B})}(v) = \pi_i(f(v))$.*
2. *Let $\mathcal{S} = (N, \{\Sigma_i\}_{i\in N}, \pi)$ be a strategic game such that, for each $i \in N$ there is $n \in \mathbb{N} \setminus \{0\}$ with $|\Sigma_i| = 2^n$. Then there exist a boolean game $\mathcal{B} = (N, \Phi, c, \{\gamma_i\}_{i\in N}, \{\Phi_i\}_{i\in N})$, a $k \in \mathbb{N}$, and a bijective function $f : V \to \prod_{i\in N} \Sigma_i$ such that for all $i, u_i^{\beta^0(\mathcal{B})}(v) = \pi_i(f(v)) - k$.*

Proof. For [1], simply construct a strategic game such that f is a bijection and then set each $\pi_i(f(v))$ to return the value given by $u_i^{\beta(\mathcal{B})}(v)$. For [2], pick again f to be bijection, and, for a sufficiently large $k \in \mathbb{N}$, set each γ_i to $\perp$ — i.e. $p \wedge \neg p$ for some $p \in \Phi_i$ — and each $u_i^{\beta^0(\mathcal{B})}(v)$ to $\pi_i(f(v)) - k$. □

The construction given in the proof allows us to talk about the strategic game *corresponding* to a boolean game and the boolean game *corresponding* to a strategic game.

Proposition 3 (Dynamic asymmetry). *Let $\mathcal{B} = (N, \Phi, c, \{\gamma_i\}_{i\in N}, \{\Phi_i\}_{i\in N})$ be a boolean game, β^0, β' be two boolean transfer functions on $\mathcal{B}$ and $\mathcal{S}, \mathcal{S}'$ the strategic games corresponding to $\beta^0(\mathcal{B})$ and $\beta'(\mathcal{B})$, respectively. Let moreover f be the bijective function given in Proposition 2 and τ' a transfer function on $\mathcal{S}$ such that for all $v \in V$ and for all $i, j \in N$ we have that $\beta'_i(v, j) = \tau'_i(f(v), j)$. It is not necessarily the case that $\tau'(\mathcal{S}) = \mathcal{S}'$.*

Proof. Consider the boolean game $\mathcal{B} = (\{1, 2\}, c, \gamma_1, \gamma_2, \Phi_1, \Phi_2)$ where $\Phi_1 = q$, $\Phi_2 = p$, $\gamma_1 = p \wedge q$, $\gamma_2 = \perp$, $c_i(x, \mathbf{tt}) = c_i(x, \mathbf{ff}) = 10$ for all $i \in \{1, 2\}$ and $x \in \Phi_i$. Consider the transfer functions β^0 and β', such that $\beta'_1(v_{p\wedge q}, 2) = 0$ and $\beta'_1(v', 2) = 10$ for all v' such that $v' \not\models \gamma_1$, while $\beta'_2 = \beta^0_2$. We have that $u_1^{\beta^0(\mathcal{B})}(v_{p\wedge q}) = \epsilon$ and $u_1^{\beta'(\mathcal{B})}(v_{p\wedge q}) = 10 + \epsilon$. However $f(v_{p\wedge q})$ yields to player 1 payoff ϵ in both $\mathcal{S}$ and $\tau'(\mathcal{S})$. □

From the last proposition we gather that updating boolean games with transfer functions — but a similar argument extends to boolean games with incentives [Wooldridge *et al.*, 2013] in a straightforward fashion — destroys the direct correspondence with strategic games. The reason lies on the fact that the utility function is calculated in such a way that a player gets a boosted payoff in outcomes satisfying his goal (boosted by the factor μ_i).

The next paragraph compares instead boolean games endowed with transfer functions and boolean games endowed with taxation mechanisms.

Side payments and taxes Wooldridge et al. [Wooldridge *et al.*, 2013] define taxation mechanisms on a boolean game $\mathcal{B} = (N, \Phi, c, \{\gamma_i\}_{i\in N}, \{\Phi_i\}_{i\in N})$ as functions $\alpha_i : V \to \mathbb{R}_+$ where every player receives a monetary *sanction* at each particular outcome. To ease the connection with our previous definitions we define the boolean game $\alpha(\mathcal{B})$, i.e. the boolean game $\mathcal{B}$ to which the taxation mechanism $\alpha = \prod_{i\in N} \alpha_i$ is applied, as the boolean game $(N, \Phi, c', \{\gamma_i\}_{i\in N}, \{\Phi_i\}_{i\in N})$ where for each $i \in N, v \in V$ we have that $c'_i(v) = c_i(v) + \alpha_i(v)$. The utility of $\alpha(\mathcal{B})$ is calculated as in Definition 3 under the boolean transfer function β^0.

Transfer functions and taxation mechanisms are of a rather different kind. While the former ones consist of payoff redistributions among players at certain outcomes, without adding or subtracting to the players' total payoff, the latter ones explicitly inject new sanctions into the system to modify players' decision making. However, in a technical sense, we can always find a taxation mechanism having the same effect of a boolean transfer function on an underlying boolean game.

Proposition 4 (From side payments to taxes). *Let $\mathcal{B} = (N, \Phi, c, \{\gamma_i\}_{i\in N}, \{\Phi_i\}_{i\in N})$ be a boolean game, β a boolean transfer function. There exists a taxation mechanism α such that $NE(\beta(\mathcal{B})) = NE(\alpha(\mathcal{B}))$.*

Proof. Consider the boost factor μ_i^0 at $\beta^0(\mathcal{B})$ for player i, let κ be the correction factor at $\beta(\mathcal{B})$ and construct α as follows. For each outcome v of $\mathcal{B}$, and each $i \in N$, set $\alpha_i(v) := {\mu^0}_i + \beta_i(v) + \kappa$. Let now $u^{\alpha(\mathcal{B})}$ be the utility function induced by $\alpha(\mathcal{B})$. We can observe that the expression assigned to each $\alpha_i(v)$ is positive and that for each $i \in N$ and $v, v' \in V$ we have that $u_i^{\beta(\mathcal{B})}(v) \geq u_i^{\beta(\mathcal{B})}(v')$ if and only if $u_i^{\alpha(\mathcal{B})}(v) \geq u_i^{\alpha(\mathcal{B})}(v')$, and consequently that $NE(\beta(\mathcal{B})) = NE(\alpha(\mathcal{B}))$. □

Furthermore, we can transfer the following results to our framework.

Proposition 5 (Complexity of finding equilibria). *Let $\mathcal{B} = (N, \Phi, c, \{\gamma_i\}_{i\in N}, \{\Phi_i\}_{i\in N})$ be a boolean game and φ a boolean formula constructed on Φ. The following hold:*

1. *The problem of verifying whether, for $\beta \in \prod_{i\in N} B_i$ and $v \in V$, $v \in NE(\beta(\mathcal{B}))$ is co-NP complete;*
2. *The problem of verifying whether some boolean transfer function β and outcome $v \in NE(\beta(\mathcal{B}))$ exist such that $v \models \varphi$ is Σ_p^2 complete;*
3. *The problem of verifying whether some boolean transfer function β exists such that $NE(\beta(\mathcal{B})) \neq \emptyset$ and for all $v' \in NE(\beta(\mathcal{B}))$ we have that $v' \models \varphi$ is Σ_p^2 complete.*

Proof. The proofs of [Wooldridge *et al.*, 2013, Proposition 1], [Wooldridge *et al.*, 2013, Proposition 6], [Wooldridge *et al.*, 2013, Proposition 14] immediately carry over to our case. □

Proposition 4 shows that, in some respect, taxation mechanisms can *simulate* transfer functions. It must however be said that simulations of this kind only make sense when both taxation mechanisms and transfers functions are fixed. But unlike taxation mechanisms, that are decided externally, transfer functions bear further strategic considerations. As made clear in the motivating example, it is not enough to establish that a transfer function induces equilibria in the resulting game, but we also need to establish whether the transfer function itself is part of a larger equilibrium, i.e. whether players are not better off by switching to different transfers. The following section investigates the properties of Nash equilibria in the two phase game.

4 Equilibrium analysis

The classical results on equilibrium survival for the case of endogenous games [Jackson and Wilkie, 2005] are centered on the notion of solo payoff, i.e. the payoff that a single player i can guarantee if the opponents $-i$ do not make any transfer. Here we adapt the definition to fit EBGs.

Definition 5 (Solo payoff). *Let $\mathcal{E} = (\mathcal{B}, \{B_i\}_{i\in N})$ be an EBG and let $\beta^0_{-i} = \prod_{j\in N\setminus i} \beta^0_j$. The* **solo payoff** *$\hat{s}_i$ for player i, where sup is the least upper bound operator, is*

$$\hat{s}_i^{\mathcal{E}} = sup_{\beta_i}(min_{\rho \in NE^{\Delta}((\beta_i, {\beta^0}_{-i})(\mathcal{B}))} E_i^{(\beta_i, {\beta^0}_{-i})(\mathcal{B})}(\rho))$$

In words, the solo payoff of player i is given by the best transfer i can make, under the expectation that his opponents will not make any transfer and will play the worst for i Nash equilibrium in each subgame. When the underlying EBG is fixed and no confusion can arise, we use the notation $\hat{s}_i$.

Two important facts are known for endogenous games [Jackson and Wilkie, 2005]:

1. when $N \leq 2$, a Nash equilibrium survives if and only if each player gets at least his solo payoff;
2. when $N > 2$, every pure strategy Nash equilibrium survives.

It is not obvious at all that the stability results carry over to our case, due to the elaborated construction of the utility function (Definition 3). In fact we can show that the second statement does not hold for EBGs.

Proposition 6 (No survival). *There exist an EBG $(\mathcal{B}, \{B_i\}_{i\in N})$ with $|N| = 3$ and $\{v\} = NE(\beta^0(\mathcal{B}))$ such that:*

- *$v \models \gamma_i$, for each $i \in N$*
- *v is not a surviving equilibrium.*

Proof. Let $\mathcal{B} = (\{1, 2, 3\}, c, \gamma_1, \gamma_2, \gamma_3, \Phi_1, \Phi_2, \Phi_3)$ be an EBG such that $\Phi_1 = \{p\}$, $\Phi_2 = \{q\}$, $\Phi_3 = \{r\}$, $\gamma_1 = p, \gamma_2 = q, \gamma_3 = r$, $c_1(v_{\neg p\wedge\neg q\wedge\neg r}) = 10$, $c_1(v') = 0$ for $v' \neq v_{\neg p\wedge\neg q\wedge\neg r}$, $c_1(x) = c_2(x) = c_3(x)$ for each $x \in V$. The outcome $v_{p\wedge q\wedge r}$ is clearly a unique pure strategy Nash equilibrium. Suppose now that it is also a surviving equilibrium and that the transfer function β is part of the solution. But then there exists some player i for which $u_i^{\beta(\mathcal{B})}(v_{p\wedge q\wedge r}) < 30 + \epsilon$. This means that i is better off deviating to a transfer function (β'_i, β_{-i}) such that $\beta'_i(v_{\neg p\wedge\neg q\wedge\neg r}) + c_i(v_{\neg p\wedge\neg q\wedge\neg r}) = 30$. No matter what equilibrium will be played in the resulting subgame, i will be boosting his payoff at $v_{p\wedge q\wedge r}$ to $30 + \epsilon$. Contradiction. □

The proposition is quite striking as not only is it not true that in EBGs with more than two players pure strategy Nash equilibria always survive, but this is also not the case with pure strategy Nash equilibria that are unique common goal for all players. The proof reveals a curious case of *oscillating offers* between the players. Even in games that have a unique common goal players might end up disagreeing on the final payoff vector, by offering to bear their opponent's full cost at his worst possible outcome. This is clearly a specific feature of EBGs, where the utility function is constructed in such a way to allow strategising over the allocation of costs at undesirable outcomes for the sole reason of increasing the boost factor at desirable ones.

In spite of this negative result we are still able to show a sufficient condition for pure strategy Nash equilibria to survive in certain EBGs, independently of the number of players involved. Let us call an outcome v **shareable** if we can assign to each i with $v \models \gamma_i$ a unique outcome $\mathbf{v}^i$ such that:

- for all $j \in N$, $(v_j, \mathbf{v}^i_{-j}) \neq v$. In other words, we focus on outcomes that cannot be reached from v by an individual deviation.
- for all $v' \in V$, $\sum_{j \in N} c_i(v') \leq \sum_{j \in N} c_j(\mathbf{v}^i)$. In other words, each player realizing a goal in v is identified with a unique outcome where the aggregated cost is maximal.

We call it moreover **potentially shareable** if there exists a cost function c^* such that the outcome v of $(N, \Phi, c^*, \{\gamma_i\}_{i\in N}, \{\Phi_i\}_{i\in N})$ is shareable. EBGs with shareable outcomes display the property we are after.

Proposition 7 (Survival). *Let $(\mathcal{B}, \{B_i\}_{i\in N})$ be an EBG. A pure strategy Nash equilibrium v of $\mathcal{B}$ survives whenever v is shareable and $u_i(v) \geq \hat{s}_i$ for each $i \in N$.*

Proof. We need to construct a subgame perfect equilibrium of the two phase game where v is played. On the equilibrium path let players play the profile (β', v) where: [1] for all $j \notin \{i \in N \mid v \models \gamma_i\}, v' \in V$, we have that, for all $k \neq j$, $\beta'_j(v', k) = \beta^0_j(v', k)$; [2] for all $j \in \{i \in N \mid v \models \gamma_i\}$ we have that $\beta'_j(\mathbf{v}^j, k) = c_k(\mathbf{v}^j)$ and $\beta'_j(v', k) = \beta^0_j(v', k)$, for all $v' \neq \mathbf{v}^j$, $k \neq j$. Off the equilibrium path, if a single player deviates from β' pick the worst Nash equilibrium for that player and have that played in the continuation. If more than a player deviates from β', play any Nash equilibrium. We can observe that no player i can get more than $\hat{s}_i$ by deviating from β', given the choices played in each subgame. Moreover v is a pure strategy Nash equilibrium of $\beta'(\mathcal{B})$. □

To see that there are EBGs where pure strategy Nash equilibria do survive, notice that a large enough game, where all outcomes have zero cost for all players and where there is a common goal, has a shareable outcome and players get at least their solo payoff in the corresponding choice profile. [3]

5 An integrated framework

If we take the point of view of an external authority that would like a certain outcome of an EBG to be rationally chosen by players, the question is not only whether that outcome can be turned into a Nash equilibrium, but also whether that Nash equilibrium is bound to survive. The purpose of this section is to generate taxation mechanisms that guarantee equilibria to survive when players play rational transfers to one another and play rationally in the game that is updated both with the transfers made and with the taxation mechanism.

The procedure we present takes a potentially shareable outcome that is consistent with players' goals and turns it into a surviving Nash equilibrium by appropriately employing taxation mechanisms.

Algorithm 1.

Input *A potentially shareable outcome v of a boolean game $\mathcal{B} = (N, \Phi, c, \{\gamma_i\}_{i\in N}, \{\Phi_i\}_{i\in N})$ such that for each i either $v \models \gamma_i$ or for no v'_i we have that $(v_{-i}, v'_i) \models \gamma_i$.*

Output *A taxation mechanism α on $\mathcal{B}$.*

Steps

1. **Let** $\alpha_i(v') = \mu^0_i - c_i(v')$, *for each* $i \in N, v' \in V$;
2. **While** *for some* $i \in N$ *we have that* $u_i^{\alpha(\mathcal{B})}(v) < \hat{s}_i^{(\alpha(\mathcal{B}), \{B_i\}_{i\in N})}$, **do** $\alpha_i(v') := \alpha_i(v') + 1$, *for each* $v' \neq v \in V$;
3. **Return** α.

Proposition 8 (Survival by taxation). *Let v be potentially shareable outcome of a boolean game $\mathcal{B} = (N, \Phi, c, \{\gamma_i\}_{i\in N}, \{\Phi_i\}_{i\in N})$ such that for each i either $v \models \gamma_i$ or for no v'_i we have that $(v_{-i}, v'_i) \models \gamma_i$. There exists a taxation mechanism α such that v is a surviving equilibrium of the EBG $(\alpha(\mathcal{B}), \{B_i\}_{i\in N})$.*

Proof. To see that Algorithm 1 guarantees this fact we only need to observe that the construction of α at step 1 ensures that the outcome v of game $\alpha(\mathcal{B})$ is shareable, and the update of α at step 2 ensures that the payoff $u_i^{\alpha(\mathcal{B})}(v)$ will eventually reach $\hat{s}_i^{(\alpha(\mathcal{B}), \{B_i\}_{i\in N})}$ while keeping v shareable. □

6 Conclusion

We have studied boolean games where players are endowed with the possibility of offering side payments to their fellow players in order to convince them to play designated strategies. The perspective we have taken integrates the framework of boolean games studied in artificial intelligence with that of endogenous games with side payments studied in game theory. We have seen that the resulting games display specific properties that make them worth studying in their own sake (Propositions 2 and 3) and the classical results available on Nash equilibria survival do not generalize (Proposition 6). We have however provided sufficient conditions that Nash equilibria need to have in order to survive (Proposition 7), independently of the number of players involved. We have also shown that, with an appropriate use of taxation mechanisms, in large enough games every outcome consistent with players' goals can be turned into a surviving Nash equilibrium (Algorithm 1). Future research efforts will be devoted to studying the interaction between side payments in boolean games and more realistic taxation mechanism that carry out imperfect redistribution of wealth, i.e.. extract payoff units to some players at certain outcomes redistributing a part of it to possibly different players at possibly different outcomes. Attention will also be paid to the relation with mechanism design and the algorithmic properties of the procedures under study.

Acknowledgments

Paolo Turrini acknowledges the support of the IEF Marie Curie fellowship "Norms in Action: Designing and Comparing Regulatory Mechanisms for Multi-Agent Systems" (FP7-PEOPLE-2012-IEF, 327424 "NINA") and of the COFUND Marie Curie fellowship "Trust Games" (FP7-PEOPLE-2011-COFUND, 1196394 "TrustGames"), cofunded by the National Research Fund of Luxembourg.
He is, besides, immensely grateful to Valentin Goranko for their long discussions on side payments in strategic games.

[3] The size requirement is not particularly demanding. Every boolean game with $|\Phi_i| \geq 2$ for each i has a potentially shareable outcome. If all the costs are uniform, that outcome is also shareable.

References

[Blackburn *et al.*, 2001] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge Tracts in Theoretical Computer Science, 2001.

[Bonzon *et al.*, 2009a] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Dependencies between players in boolean games. *International Journal of Approximate Reasoning*, 50(6):899–914, 2009.

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

[Farrell, 1998] Joseph Farrell. Communication, coordination and nash equilibrium. *Economics Letters*, 27:209–214, 1998.

[Goranko and Turrini, 2012] Valentin Goranko and Paolo Turrini. Non-cooperative games with preplay negotiations. *CoRR*, abs/1208.1718, 2012.

[Guttman, 1978] Joel M. Guttman. Understanding collective action: Matching behavior. *American Economic Review*, 68(2):251–255, 1978.

[Harrenstein *et al.*, 2001] Paul Harrenstein, Wiebe van der Hoek, John-Jules Ch. Meyer, and Cees Witteveen. Boolean games. In *Proceeding of the Eighth Conference on Theoretical Aspects of Rationality and Knowledge (TARK VIII)*, pages 287–298, 2001.

[Jackson and Wilkie, 2005] Matthew O. Jackson and Simon Wilkie. Endogenous games and mechanisms: Side payments among players. *Review of Economic Studies*, 72(2):543–566, 2005.

[Kalai, 1981] Ehud Kalai. Preplay negotiations and the prisoner's dilemma. *Mathematical Social Sciences*, 1:375–379, 1981.

[Osborne and Rubinstein, 1994] Martin Osborne and Ariel Rubinstein. *A course in game theory*. MIT Press, 1994.

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

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