Hard and Soft Equilibria in Boolean Games

Paul Harrenstein
Dept. of Computer Science
University of Oxford
paulhar@cs.ox.ac.uk

Paolo Turrini
Department of Computing
Imperial College London
p.turrini@imperial.ac.uk

Michael Wooldridge
Dept. of Computer Science
University of Oxford
mjw@cs.ox.ac.uk

ABSTRACT

A fundamental problem in game theory is the possibility of reaching equilibrium outcomes with undesirable properties, e.g., inefficiency. The economics literature abounds with models that attempt to modify games in order to eliminate such undesirable equilibria, for example through the use of subsidies and taxation, or by allowing players to undergo a preplay negotiation phase. In this paper, we consider the effect of such transformations in Boolean games with costs, where players are primarily motivated to seek the satisfaction of some goal, and are secondarily motivated to minimise the costs of their actions. The preference structure of these games allows us to distinguish between *hard* and *soft* equilibria, where hard equilibria arise from goal-seeking behaviour, and cannot be eliminated from games by, e.g., taxes or subsidies, while soft equilibria are those that arise from the desire of agents to minimise costs. We investigate several mechanisms which allow groups of players to form coalitions and eliminate undesirable equilibria from the game, even when taxes or subsidies are not a possibility.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems; J.4 [**Computer Applications**]: Social and Behavioral Sciences - Economics

General Terms

Economics, Theory

Keywords

Game theory (cooperative and non-cooperative), Boolean games, Nash equilibria, Externalities

1. INTRODUCTION

A fundamental problem in the theory of games is that a game may contain Nash equilibria with undesirable properties. To take a famous example, in the Prisoner's Dilemma the unique pure strategy Nash equilibrium, i.e., mutual defection, is the only outcome of the game that is not Pareto optimal, and it is besides strictly worse for both players than the alternative solution of mutual cooperation. From an outside perspective, mechanisms can be devised to incentivise the players to play certain actions, e.g., by means of subsidies, or to disincentivise them to do so, e.g., by imposing taxes, de facto modifying the game. Likewise, the situation may allow for various ways in which the players themselves can modify the game they are playing, e.g., by making agreements, transferring money to one other, or joining forces in coalitions to their mutual benefit. Both these solutions have been studied in the economic literature since its early stages [4, 3, 9, 13].

Boolean games (of the form studied in [15, 12]) represent an important domain for investigating these issues, because preferences in such games have a particular structure: players are primarily motivated to achieve a goal and they are only secondarily motivated to minimise the cost of actions required to achieve that goal. In particular, it is assumed that a player will *always* prefer to achieve her goal than otherwise. Such so-called *quasi-dichotomous* preferences, besides inducing nonstandard properties in game play [12], are quite natural in many application domains for multi-agent systems. For example, consider a robot programmed to perform a particular task in an automated warehouse. Operating the robot involves energy consumption, which we might want to minimise, but at the same time we do not want to compromise the successful execution of the task. In other words, we *primarily* want the robot to successfully carry out the task, and only *secondarily* to minimise its energy consumption. Given such preference structures, it turns out there are limits on way such a game can be manipulated. A player *cannot* be incentivised to choose a course of action that would not lead to his goal to be satisfied over a course of action that would.

Following an introduction to the formal framework of Boolean games with costs, we distinguish between *hard* and *soft* equilibria in our games. A hard equilibrium is one that will remain an equilibrium no matter what costs are being imposed. In contrast, a soft equilibrium is one that can be eliminated from or introduced to a game by the introduction of some cost incentive. Clearly, the presence of hard equilibria with undesirable properties would be bad news, as it would be impossible to incentivise players not to choose this outcome if they want to. However, we show that hard equilibria are in fact rather rare in the sense that the conditions required for their presence are rather strong. In addition, we show that hard equilibria, when present, do in fact have desirable social properties.

We give logical classifications of both hard and soft equilibria and discuss their properties. We then turn to the issue of managing equilibria, using well-known ideas from the economics literature [13, 4, 3, 9, 8]. We first consider the possibility of groups of players to engineer side-payments so as to motivate another player to act in a way that is beneficial to the group. This is the idea of (a group of) players encouraging another player to increase the positive externalities or reduce the negative externalities it induces [4, 9, 8]. Second, we study the possibility of a player taking into ac-

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

count the undesirable consequences his choices have for others by *merging* that player with some of the other players. This is one way of what is known in the economics literature as *internalising externalities* [13, 3]. We investigate how such a mechanism can affect the equilibria of a game. In particular, we show that by allowing players to merge in coalitions hard equilibria can be eliminated.

2. BOOLEAN GAMES WITH COSTS

Boolean games are based on propositional logic, and have a natural computational interpretation, which is highly relevant to the multi-agent systems domain (see, e.g., [6, 2, 5, 15, 12]). In this paper, we use the Boolean games model with cost functions, in which the players have quasi-dichotomous preferences, as in [15]. Thus, each player is primarily interested in satisfying a goal, which is expressed by a Boolean formula, classifying each outcome as desirable or undesirable. A player's secondary concern is with costs: if two outcomes both satisfy or both do not satisfy a player's goal, the one is preferred that has the lower cost. This quasi-dichotomous character of the preferences allows us to accentuate certain aspects of the effects of side-payments and coalition merging in the context of externalities.

Let $\mathbb{B} = \{\top, \bot\}$ be the set of Boolean truth values, with "$\top$" being truth and "$\bot$" being falsity. Let, furthermore, $\Phi = \{p, q, \ldots\}$ be a fixed, finite, and non-empty vocabulary of Boolean variables and $\mathcal{L}$ the set of well-formed formulae of propositional logic over Φ with the conventional Boolean operators ("$\wedge$", "$\vee$", "$\rightarrow$", "$\leftrightarrow$", "$\neg$") as well as the truth constants "$\top$" and "$\bot$". A *valuation* is a function $v : \Phi \rightarrow \mathbb{B}$, assigning truth or falsity to every Boolean variable. We write $v \models \varphi$ to mean that φ is true under, or satisfied by, valuation v, where the satisfaction relation "$\models$" is defined in the standard way. Let $\mathcal{V}$ denote the set of all valuations over Φ.

The games we consider are populated by a set $N = \{1, \ldots, n\}$ of *agents*, the players of the game. Each agent $i \in N$ is assumed to have a *goal*, which is characterised by an $\mathcal{L}$-formula γ_i. Each agent $i \in N$ *controls* a (possibly empty) subset Φ_i of the overall set of Boolean variables. By "control", we mean that i has the unique ability within the game to set the value (either $\top$ or $\bot$) of each variable $p \in \Phi_i$. We will require that $\Phi_1, \ldots, \Phi_n$ forms a partition of Φ, i.e., $\Phi_i \cap \Phi_j = \emptyset$ for $i \neq j$ and $\Phi_1 \cup \cdots \cup \Phi_n = \Phi$. Every variable is controlled by precisely one agent. A *choice* for agent $i \in N$ is defined by a function $v_i : \Phi_i \rightarrow \mathbb{B}$, i.e., an allocation of truth or falsity to all the variables under i's control. If $\Phi_i = \emptyset$, player i has only one action and is called a *dummy player*. Let $\mathcal{V}_i$ denote the set of choices for agent i. The intuitive interpretation we give to $\mathcal{V}_i$ is that it defines the *actions* or *strategies* available to agent i, i.e., the *choices* available to the agent.

An *outcome* $\vec{v} = (v_1, \ldots, v_n)$ in $\mathcal{V}_1 \times \cdots \times \mathcal{V}_n$ is a collection of choices, one for each agent. Clearly, every outcome uniquely defines a valuation, and we will abuse notation by treating outcomes as valuations and valuations as outcomes. So, for example, we will write $\vec{v} \models \varphi$ to mean that the valuation defined by the outcome $\vec{v}$ satisfies formula φ. We let $\vec{\mathcal{V}}$ denote the set of outcomes, and we write $(\vec{v}_{-i}, v_i')$ for the outcome $(v_1, \ldots, v_{i-1}, v_i', v_{i+1}, \ldots, v_n')$.

When playing a Boolean game, the primary aim of an agent i will be to choose an assignment of values for the variables Φ_i under her control so as to satisfy her goal γ_i. The difficulty is that γ_i may contain variables controlled by other agents $j \neq i$, who will also be trying to choose values for their variables Φ_j so as to get their goals satisfied. As their goals in turn may be dependent on the variables in Φ_i, they may have to take into account how player i will act when making their choice. And so on. In our setting, moreover, outcomes are associated with costs to the players. Minimising these costs is another important, but secondary, concern to the players. Thus, if an agent has multiple ways of getting his goal achieved, then *he will prefer to choose one that minimises his costs*, whereas, if an agent cannot get his goal achieved, then *he simply chooses to minimise his costs*.

To capture these preferences, we introduce two types of cost function: *global cost functions* and *local cost functions*. The former associate with each outcome a cost for each of the players, whereas the latter associate costs with setting propositional variables to one of the two truth-values. Formally, we define a *global cost function* as a function

$$c : N \times \vec{\mathcal{V}} \rightarrow \mathbb{Q}_{\geq},$$

which associates each player i and each outcome $\vec{v}$ with a non-negative rational number, intuitively representing the amount by which player i is taxed when $\vec{v}$ is the outcome of the game. We also write $c_i(\vec{v})$ for $c(i, \vec{v})$. Wooldridge *et al.* [15] assumed a natural *additive* (and more concise) model for costs given by *local cost functions* of the form

$$\hat{c} : \Phi \times \mathbb{B} \rightarrow \mathbb{Q}_{\geq}.$$

Intuitively, $\hat{c}(p, b)$ is the marginal cost of assigning the value $b \in \mathbb{B}$ to variable $p \in \Phi$. Given this definition, we can extend the local cost function $\hat{c}$ to outcomes $\vec{v} = (v_1, \ldots, v_n)$, as follows:

$$\hat{c}(i, \vec{v}) = \begin{cases} \sum_{p \in \Phi_i} \hat{c}(p, v_i(p)) & \text{if } \Phi_i \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

Notice that this model implies that the cost a player incurs *only depends on the choice that this player makes*. With a slight abuse of notation, we therefore also write $\hat{c}_i(v_i)$ for $\hat{c}(i, \vec{v})$ where $\vec{v} = (v_1, \ldots, v_n)$ and $\hat{c}$ is induced by a local cost function. Observe that every local cost function defines a global cost function, but not necessarily the other way round. In particular, the local costs for a player i will be the same for any two outcomes $\vec{v} = (v_1, \ldots, v_n)$ and $\vec{v}' = (v_1', \ldots, v_n')$ whenever $v_i = v_i'$. This need not be the case for global cost functions.

We can now introduce the utility functions that model the players' preferences. Let $\vec{v}^e$ be an outcome with maximal cost for player i, and let μ_i denote the cost to i of this most expensive outcome, i.e.,

$$\vec{v}^e \in \arg\max_{\vec{v} \in \vec{\mathcal{V}}} c_i(\vec{v}) \quad \text{and} \quad \mu_i = c_i(\vec{v}^e).$$

The *utility* to agent i of an outcome $\vec{v} = (v_1, \ldots, v_i, \ldots, v_n)$ is then defined as follows:

$$u_i(\vec{v}) = \begin{cases} 1 + \mu_i - c_i(\vec{v}) & \text{if } \vec{v} \models \gamma_i, \\ -c_i(\vec{v}) & \text{otherwise.} \end{cases}$$

Thus, utility to agent i will range from $1 + \mu_i$ (for an outcome in which i gets his goal achieved at the lowest cost) down to $-\mu_i$ (for outcomes with the highest cost to i in which i's goal is not satisfied).

Formally, a *Boolean game (with costs)* is then given by a structure

$$G = (N, \Phi, c, (\gamma_i)_{i \in N}, (\Phi_i)_{i \in N}),$$

where $N = \{1, \ldots, n\}$ is a set of agents, $\Phi = \{p, q, \ldots\}$ is a finite set of Boolean variables, c is a (global or local) cost function, $\gamma_i \in \mathcal{L}$ is the goal of agent $i \in N$, and $\Phi_1, \ldots, \Phi_n$ is a partition of Φ over N, with the intended interpretation that Φ_i is the set of Boolean variables under the unique control of $i \in N$.

Boolean games represent games in strategic form, with choices of players as their actions and the utility function as defined above representing their preferences. Accordingly, the standard game-theoretic solution concepts are available for the analysis of Boolean

	$q \wedge r$	$q \wedge \neg r$	$\neg q \wedge r$	$\neg q \wedge \neg r$
p	1,2,3 $1, 11, 0$	1,2,3 $1, 11, 0$	2 $1, 2, 0$	$1, 2, 0$
$\neg p$	1,2,3 $1, 11, 0$	1,2,3 $1, 11, 0$	2 $1, 2, 0$	$1, 2, 0$

Figure 1: The game of Example 1, in which player 1 chooses rows and player 2 columns. Player 3 is a dummy player. The figures in the top right corners of the cells indicate the players that have their goals satisfied in the respective outcome. The three figures x, y, z in the centre of each cell denote the *costs* to player 1, player 2, and player 3, respectively.

games [11]. We focus on Nash equilibrium. An outcome $\vec{v} = (v_1, \ldots, v_n)$ is a *(pure) Nash equilibrium* if for all agents $i \in N$, there is no $v_i' \in \mathcal{V}_i$ such that

$$u_i(\vec{v}_{-i}, v_i') > u_i(\vec{v}).$$

Let $\text{NE}(G)$ denote the set of all Nash equilibria of the game G. Let us consider an example.

EXAMPLE 1. *Suppose we have a game G_1 with $N = \{1, 2, 3\}$, $\Phi = \{p, q, r\}$, $\Phi_1 = \{p\}$, $\Phi_2 = \{q, r\}$, $\Phi_3 = \emptyset$, and a local cost function. The cost for player 2 of setting $q = \top$ is 10 and all other costs in the game are 1. Finally, we have $\gamma_1 = \gamma_3 = q$, and $\gamma_2 = q \vee r$. Also see Figure 1. Now, although player 2 could choose an action that satisfies not only his goal but also those of players 1 and 3—e.g., setting q and r both to true—he will not rationally do so: by setting q to true and r to false he would achieve his goal at a lower cost. Then, however, players 1 and 3 are left without their goals being satisfied. It is easy to see that for all $\vec{v} \in \text{NE}(G_1)$ we have $\vec{v} \models \neg\gamma_1 \wedge \gamma_2 \wedge \neg\gamma_3$, i.e., in all Nash equilibria, players 1 and 3 fail to get their goals achieved. This is a simple example of a Nash equilibrium with undesirable social properties. Moreover, this equilibrium is troubling, because there is an alternative outcome of the game in which* all *players's goals are satisfied.*

3. HARD AND SOFT EQUILIBRIA

Recall that an agent's preferences are driven by two components: the primary one is her goal γ_i, the secondary one is cost minimisation. It is important to emphasise that cost minimisation is strictly secondary to goal achievement: an agent will *always* prefer an outcome that satisfies her goal over one that does not, irrespective of what the cost implications are.

In this section, we will show that the fact that there are two distinct drivers behind an agent's preferences gives a two-tier structure to the Nash equilibria of Boolean games. Specifically, we distinguish between *hard* and *soft* equilibria. Informally, a hard equilibrium is one that is present in game irrespective of the cost function of the game. In contrast, a soft equilibrium is one whose presence in a game is contingent upon the cost function of the game. As a consequence, if an equilibrium is soft, then it can potentially be eliminated from the game if it is viewed as undesirable—e.g., through taxes [13, p.656]—or introduced to the game by providing appropriate incentives, if it is seen as desirable. To formalise this intuition, we need some more notation.

First, given a game G with cost function c, we denote by $G^{c'}$ the game obtained from G by replacing the cost function c with cost function c'. Thus, in $G^{c'}$, the primary drivers behind each player's preferences (i.e., goal achievement) remain the same as in G, but the secondary drivers (i.e., cost reduction) may be different.

$$\begin{array}{ccccc} \text{HARD}(G) & \subseteq & \text{NE}(G) & \subseteq & \text{INIT}(G) \\ & & \cup\!| & & \cup\!| \\ & & \text{PRESENT}(G) & \subseteq & \text{SOFT}(G) \\ & & & & \cup\!| \\ & & & & \text{ABSENT}(G) \end{array}$$

Figure 2: Containment relations between types of equilibria.

Important in the discussion that follows is the *zero cost function* $c^{\mathbf{0}}$, which assigns cost 0 to all players in all outcomes. Thus, in a game G with cost function $c^{\mathbf{0}}$, which we will also denote by $G^{\mathbf{0}}$, players are indifferent between outcomes on the basis of costs: the *only* driver for an agent is to achieve his goal.

Let us now define the set $\text{INIT}(G)$ of *initial equilibria* of a game G to consist of those equilibria that are present in the game $G^{\mathbf{0}}$, i.e.,

$$\text{INIT}(G) = \text{NE}(G^{\mathbf{0}}).$$

The reason for singling out this set and giving it its name is illustrated by the following observation.

OBSERVATION 2 ([15]). *For all games G, $\text{NE}(G) \subseteq \text{INIT}(G)$.*

Thus, the game $G^{\mathbf{0}}$ contains a *maximal* set of Nash equilibria with respect to G. In particular, if for some outcome $\vec{v}$ we have $\vec{v} \notin \text{NE}(G^{\mathbf{0}})$, then there is no possibility of introducing it to G via the imposition of some cost function. Marginal costs defined within cost functions c serve to *eliminate* Nash equilibria from a game $G^{\mathbf{0}}$.

We can now define the hard equilibria of a game G. Formally, the set of hard equilibria of G are those equilibria that are present in G no matter what cost function we assign to the game, i.e.,

$$\text{HARD}(G) = \bigcap_{c:N\times\vec{\mathcal{V}}\to\mathbb{Q}_{\geq}} \text{NE}(G^c).$$

Thus, if $\vec{v} \in \text{HARD}(G)$, then $\vec{v}$ is "immune" to any cost considerations, because no matter what we do to the cost function of G, the outcome $\vec{v}$ will remain an equilibrium in the game.

In contrast, a soft equilibrium is one that is present in a game for some assignment of costs in the game, but is absent for some other assignment of costs. We can thus think of soft equilibria as being the "malleable" part of a game: it is these equilibria that we can eliminate from or introduce to games. Formally, $\text{SOFT}(G)$ denotes the set of soft equilibria of G, i.e.,

$$\text{SOFT}(G) = \text{INIT}(G) \setminus \text{HARD}(G).$$

To understand this definition, recall that $\text{INIT}(G)$ is the maximal set of Nash equilibria that could be present in a game. Accordingly, an outcome can be a soft equilibrium in a game without being a Nash equilibrium. It will, however, be a Nash equilibrium for the game with another cost function. For this reason, we will distinguish between soft equilibria that are present and those that are absent in a game. Thus, we let $\text{PRESENT}(G)$ denote the set of soft equilibria of G that are present in G, and let $\text{ABSENT}(G)$ denote the set of soft equilibria that are not present in G, i.e.,

$$\text{PRESENT}(G) = \text{NE}(G) \setminus \text{HARD}(G),$$
$$\text{ABSENT}(G) = \text{SOFT}(G) \setminus \text{PRESENT}(G).$$

Figure 2 illustrates the containment relations between these sets; these all follow directly from the definitions presented above together with Observation 2. Let us see an example.

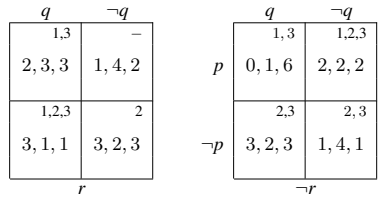

Figure 3: A three-player game, in which player 1 controls p, player 2 controls q, and player 3 controls r. The notational conventions are as in Figure 1.

EXAMPLE 3. *Consider the Boolean game in Figure 3 with* $N = \{1, 2, 3\}$, $\Phi_1 = \{p\}$, $\Phi_2 = \{q\}$, *and* $\Phi_3 = \{r\}$. *The goals of the players are given by* $\gamma_1 = (r \rightarrow q) \wedge (\neg r \rightarrow p)$, $\gamma_2 = p \rightarrow (\neg q \wedge \neg r)$, *and* $\gamma_3 = r \rightarrow q$. *Here, as in the further examples, we will refer to the outcome satisfying* $p \wedge \neg q \wedge r$ *as* $\vec{v}^{p\neg qr}$ *and for the other other outcomes similarly.*

The game has two Nash equilibria, viz., $\vec{v}^{pqr}$ *and* $\vec{v}^{p\neg q\neg r}$. *The former is soft and present, whereas the latter is hard. Outcome* $\vec{v}^{\neg pqr}$ *is the third initial equilibrium and is achieved by the empty cost function* $c^{\mathbf{0}}$. *It is, however, soft and absent.*

The concept of a hard equilibrium is defined with respect to *all* (global) cost functions. The following lemma, however, shows that one only need to consider the *local* cost functions to decide whether an outcome is a hard equilibrium. As the zero cost function $c^{\mathbf{0}}$ can obviously be seen as being induced by the local cost function that assigns cost zero to every variable, this also holds for the initial, soft, absent, and present equilibria.[1]

LEMMA 4. *Let G be a game and* $\vec{v}$ *an outcome. Then,* $\vec{v} \in \text{HARD}(G)$ *if and only if* $\vec{v} \in \text{NE}(G^c)$ *for all* local *cost functions c.*

Notice that in the game G_1 of Example 1 there are no hard equilibria. The following proposition establishes that this is in fact no coincidence: hard equilibria are rather scarce in Boolean games. For an outcome to be a hard equilibrium, every player must get their goal achieved in that outcome, and any deviation from that outcome by a player must result in that player's goal being unsatisfied. The following proposition states this fact more formally.

PROPOSITION 5. *Let* $\vec{v} = (v_1, \ldots, v_i, \ldots, v_n)$ *be an outcome of a game G. Then,* $\vec{v} \in \text{HARD}(G)$ *if and only if both of the following conditions are satisfied:*

(*i*) $\vec{v} \models \gamma_i$ *for all non-dummy players i in G, and*

(*ii*) *for all players i in G and all choices* $v_i' \in \mathcal{V}_i$ *with* $v_i' \neq v_i$, *we have* $(\vec{v}_{-i}, v_i') \not\models \gamma_i$.

PROOF. For the "only if"-direction, we show that, if either of the right-hand side conditions is not satisfied, then $\vec{v} \notin \text{HARD}(G)$. Suppose condition (*i*) is not satisfied and let *i* be a non-dummy player for which $\vec{v} \not\models \gamma_i$. Let $v_i' \in \mathcal{V}_i$ with $v_i' \neq v_i$ and fix a local cost function $\hat{c}$ such that that v_i' is strictly the cheapest choice for *i*. Then v_i' represents a beneficial deviation for *i* from $\vec{v}$ in $G^{\hat{c}}$, and so $\vec{v} \notin \text{HARD}(G)$. Now suppose condition (*ii*) is not satisfied. Then, some player *i* has a choice $v_i' \neq v_i$, such that $(\vec{v}_{-i}, v_i') \models \gamma_i$. Fix a (local) cost function $\hat{c}$ that makes v_i' the strictly cheapest choice for *i*. Then, v_i' is a beneficial deviation for *i*: if $\vec{v} \not\models \gamma_i$, then the deviation to v_i' benefits *i* because she gets her goal achieved, while, if $\vec{v} \models \gamma_i$, then deviating to v_i' benefits *i* because it reduces costs compared to v_i.

For the opposite direction, suppose (*i*) and (*ii*) are satisfied but that $\vec{v} \notin \text{HARD}(G)$. From the latter assumption, there is a cost function $c : \Phi \times \mathbb{B} \rightarrow \mathbb{R}$ such that $\vec{v} \notin \text{NE}(G^c)$. Then, some player *i* has a beneficial deviation v_i' from $\vec{v}$ in G^c. Now, by (*i*), all non-dummy players have their goals achieved with $\vec{v}$, and, by (*ii*), any deviation, in particular to v_i', would result in their goals being unsatisfied. Hence, v_i' cannot be a beneficial deviation and we obtain a contradiction. □

The significance of Proposition 5 may not be immediately apparent. We argue, however, that it is a positive result rather than a negative one. Hard equilibria cannot be eliminated from games through cost functions, and so the presence of hard equilibria with undesirable properties would be bad news indeed. But Proposition 5 establishes that, first, hard equilibria are rare in games, in the sense that the condition required for their presence is very strong, and second, where they are present, hard equilibria in fact have properties that can be viewed as very desirable: all players have their goals achieved, and hence obtain positive utility. Thus, hard equilibria can be understood as maximising *qualitative social welfare* [15]. Also notice that Proposition 5 is *purely logical*. The condition on the right-hand side is expressed solely in terms of valuations and goal formulae; no reference is made to cost functions.

The following proposition characterises the set of soft equilibria.

PROPOSITION 6. *Let* $\vec{v} = (v_1, \ldots, v_i, \ldots, v_n)$ *be an outcome of a game G. Then,* $\vec{v} \in \text{SOFT}(G)$ *if and only if both* $\vec{v} \in \text{INIT}(G)$ *and there is a player i and an outcome* $\vec{v}' = (v_1', \ldots, v_n')$ *with* $\vec{v}' \neq \vec{v}$ *such that the following conditions are satisfied:*

(*i*) $\vec{v} \models \gamma_i$ *if and only if* $\vec{v}' \models \gamma_i$, *and*

(*ii*) $\vec{v}_{-i} = \vec{v}'_{-i}$.

PROOF. For the "if"-direction, assume $\vec{v}'$ satisfies the right-hand side of the condition. Define a local cost function $\hat{c}$ such that, for each propositional variable *p* and boolean $b \in \{\bot, \top\}$,

$$\hat{c}(p, b) = \begin{cases} 0 & \text{if } \vec{v}'(p) = b, \\ 1 & \text{otherwise.} \end{cases}.$$

With this cost function, the least-cost choice for an agent *i* is to choose his part v_i' of $\vec{v}'$; any other choice will incur higher cost. In this case, since $\vec{v} \models \gamma_i$ if and only if $\vec{v}' \models \gamma_i$, then *i* will have a beneficial deviation from $\vec{v}$ to $\vec{v}'$, and so $\vec{v} \notin \text{NE}(G^{\hat{c}})$. Since $\vec{v} \in \text{NE}(G^{\mathbf{0}})$ and $\vec{v} \notin \text{NE}(G^{\hat{c}})$, then $\vec{v} \in \text{SOFT}(G)$.

For the opposite direction, assume $\vec{v} \in \text{SOFT}(G)$. Then, there is a cost function $c : \Phi \times \mathbb{B} \rightarrow \mathbb{Q}_{\geq}$ such that $\vec{v} \in \text{INIT}(G)$ but $\vec{v} \notin \text{NE}(G^c)$. Since $\vec{v} \notin \text{NE}(G^c)$, then some player *i* has a beneficial deviation v_i' from $\vec{v}$ in G^c. We claim that $\vec{v} \models \gamma_i$ if and only if $\vec{v}' \models \gamma_i$. For suppose not, then either (*a*) $\vec{v} \models \gamma_i$ and $\vec{v}' \not\models \gamma_i$, or else (*b*) $\vec{v} \not\models \gamma_i$ and $\vec{v}' \models \gamma_i$. In case (*a*), we would have $u_i(\vec{v}_{-i}, v_i') < u_i(\vec{v})$ in G^c, and so v_i' cannot be a beneficial deviation. In case (*b*), since *i* would not get her goal achieved in $\vec{v}$ but would with $(\vec{v}_{-i}, v_i')$, choice v_i' would be a beneficial deviation from $\vec{v}$ for player *i* in the game $G^{\mathbf{0}}$. Hence, $\vec{v} \notin \text{NE}(G^{\mathbf{0}})$, i.e., $\vec{v} \notin \text{INIT}(G)$. In both cases we have a contradiction. Hence, $\vec{v} \models \gamma_i$ if and only if $\vec{v}' \models \gamma_i$. Thus, the conditions of the right-hand side are satisfied. □

Again, this characterisation is purely logical and makes no reference to cost functions.

[1] Some of the proofs in this paper we omit due to space restrictions.

4. EXTERNALITIES IN BOOLEAN GAMES

In this section, we will see how the concepts we introduced above can be used to understand and manage *externalities* in Boolean games. The term "externality" in economics is used to refer to a situation where the actions of one agent can affect the well-being of one or more other agents [13]. An example of (negative) externality is a factory discharging industrial effluent into a river upstream of a fish farm, thereby reducing the quality and quantity of the fish that the farm can produce. An example of (positive) externality is a honey producer keeping bee hives in a field that happens to be close to an orchard [9, 8]: the orchard owner benefits from the presence of the bees, who pollinate the apple trees.

There are two standard approaches in economics to deal with externalities. The first is to allow players to provide monetary compensation, or *side-payments*, to encourage or discourage certain actions to be taken. In the example of the beekeeper and the apple grower, if side-payments are allowed, the apple grower will compensate the beekeeper for his positive externality, provided the beekeeper is effectively able to prevent his bees from pollinating the apple trees [8]. Economic theorists like Coase, Meade, and Maskin have studied under what conditions this possibility allows efficient outcomes to be reached [4, 9, 8, 7]. The second approach to dealing with externalities is to have players *internalise* externalities, that is, to somehow incentivise them to take externalities into consideration when they make their choices. In the factory-fish farm example, above, if we *merge* the fish farm and factory into a single company, then it is in this company's own interest to take into account the negative effects of the pollution it causes. As such, merging players can be seen as one way to internalise externalities.

Neither of these approaches is always realisable in practice, e.g., due to the absence of communication channels among the parties involved or the lack of appropriate legislation. It is, however, interesting to study the many cases in which they are.

In Boolean games, externalities arise from the fact that the satisfaction of one player's goal can depend on the choices made by the other players. By choosing a particular valuation, a player can either help or hinder other players achieving their goals. In the next section we adapt the two approaches described above to the framework of Boolean games.

4.1 Side-payments

In this section, we investigate what groups of players can achieve if, before the game starts, they are allowed to make binding offers to their fellow coalition members to persuade them to play designated strategies. Turrini [12] studied a preplay phase preceding a Boolean game as a second game taking place before the actual game starts. Our approach here is different: given a Boolean game, we focus on the ability of *coalitions* to engineer side-payments in order to escape unsatisfactory outcomes. The question we are especially interested in is which equilibria—hard or soft—can be eliminated from the game in this manner. Consider, for example, the game in Figure 3. There, player 2 does not have her goal achieved in the equilibrium satisfying $p \wedge q \wedge r$, but she could try to incentivise player 1 to set p to $\bot$ by offering him compensation for the additional costs he incurs if he were to do so.

Following [7, 12], we formalise side-payments by means of so-called *transfer functions*, i.e., functions of the form

$$\tau : N \times N \times \vec{\mathcal{V}} \to \mathbb{Q}_{\geq}.$$

Intuitively, $\tau(i,j,\vec{v})$ is the compensation player j receives from player i for the costs j incurs at outcome $\vec{v}$. Thus, after the transfer, player i's cost at $\vec{v}$ is increased by $\tau(i,j,\vec{v})$, whereas player j's cost at the same outcome is decreased by the same amount. Importantly, it may very well be that $\tau(i,j,\vec{v}) \neq \tau(j,i,\vec{v})$.

We say that a transfer function τ *only involves coalition* C if all transfers to and from players not in C are zero at all outcomes, i.e., if, for all players $i \in N$ and $j \in N \setminus C$ and all outcomes $\vec{v}$,

$$\tau(i,j,\vec{v}) = \tau(j,i,\vec{v}) = 0.$$

Furthermore, we let $\tau_i(\vec{v})$ abbreviate the term

$$\sum_{j \in N} \tau(j,i,\vec{v}) - \sum_{j \in N} \tau(i,j,\vec{v}),$$

i.e., the net transfer received by player i under τ. It is important to observe that every transfer to a player means an equally large transfer from the other player. Therefore, each transfer $\tau(i,j,\vec{v})$ occurs once (negatively) in $\tau_i(\vec{v})$ and once (positively) in $\tau_j(\vec{v})$. In particular, if τ only involves C, we have

$$\sum_{i \in C} \tau_i(\vec{v}) = 0.$$

We restrict our attention to *admissible* transfer functions, i.e., transfer functions such that, for all players i and all outcomes $\vec{v}$,

$$\tau_i(\vec{v}) \leq c_i(\vec{v}).$$

In words, at no outcome the amount of what a player receives from others minus what he gives to them can exceed his cost. Thus, the cost a player incurs at an outcome cannot be overcompensated, i.e., it cannot end up being negative as result of preplay negotiation. For instance, if player i's cost $c_i(\vec{v}) = 3$ and player j is the only other player in the game, then it cannot be that $\tau(j,i,\vec{v}) = 5$ and $\tau(i,j,\vec{v}) = 1$. This restriction is mainly of a technical nature and preserves the quasi-dichotomous character of the preferences in games transformed by transfer functions.[2]

Thus, a transfer function transforms the cost function of a Boolean game. Let τ be a transfer function. For G a Boolean game with cost function c, we then define c^τ as the cost function such that, for all players i and all outcomes $\vec{v}$,[3]

$$c_i^\tau(\vec{v}) = c_i(\vec{v}) - \tau_i(\vec{v}).$$

The utility function of player i in game G with cost function c^τ we will henceforth denote by u_i^τ.

We are particularly interested in the equilibria of a game that can be eliminated by groups of players making side-payments to one another. Intuitively, an equilibrium $\vec{v}$ can be eliminated if a coalition can engineer a transfer function that makes it attractive for one of its members i to deviate from $\vec{v}$ to another outcome $(\vec{v}_{-i}, v_i')$. Moreover, such a side-payment scheme has to benefit all players of the coalition. That is, all coalition members should prefer $(\vec{v}_{-i}, v_i')$ after the side-payments have been made to $\vec{v}$ before. Formally, we say that a coalition C *blocks outcome* $\vec{v}$ if there is some transfer function τ only involving C, some player $i \in C$, and some $v_i' \in \mathcal{V}_i$ such that the following two conditions hold:

(i) $u_i^\tau(\vec{v}_{-i}, v_i') > u_i^\tau(\vec{v})$, and

(ii) $u_j^\tau(\vec{v}_{-i}, v_i') > u_j(\vec{v})$ for all players $j \in C$.

Condition (i) ensures that player i is incentivised to deviate from $\vec{v}$ to $(\vec{v}_{-i}, v_i')$ *after* transfers have taken place, whereas condition (ii)

[2] One could also assume a base level of unavoidable costs and model *rewards* (as distinguished from *reparations*) as compensation beyond this level.

[3] It is worth observing that transfer functions τ can be applied to both local and global cost functions c. However, if c is a local cost function, it is not necessarily the case that c^τ is as well.

guarantees that all players in C are better of in $(\vec{v}_{-i}, v_i')$ *after* transfers have taken place than they were in $\vec{v}$ *before*. Furthermore, we say an outcome $\vec{v}$ is *blocked (via side-payments)* if there is a coalition blocking $\vec{v}$. In case a coalition blocks an initial equilibrium $\vec{v}$, we also say that $\vec{v}$ is *eliminable via side-payments*.

EXAMPLE 7. *Consider again the game in Figure 3. At outcome $\vec{v}^{pqr}$ player 2's goal is not satisfied, whereas she is at $\vec{v}^{\neg pqr}$. Let τ be such that*

$$\tau(i,j,\vec{v}) = \begin{cases} x & \text{if } i = 2, j = 1, \text{ and } \vec{v} = \vec{v}^{\neg pqr}, \\ 0 & \text{otherwise.} \end{cases}$$

Then, τ would incentivise player 1 to deviate to $\vec{v}^{\neg pqr}$ provided that $x > 1$. Moreover, player 2 would gladly make any such transfer in order to satisfy her goal. Accordingly, $\{1,2\}$ is a coalition blocking the outcome $\vec{v}^{pqr}$. In a similar way, player 1 might want to induce player 3 to deviate to outcome $\vec{v}^{pq\neg r}$. That, however, would require compensating player 3 for the additional costs of 3 that player 3 incurs at $\vec{v}^{pq\neg r}$. Player 1, having his goals achieved at both $\vec{v}^{pqr}$ and $\vec{v}^{pq\neg r}$, however, is not prepared to do so, as his marginal gain in costs (before transfer) would only be 2. Still, player 2 would also like to see player 3 deviate to $\vec{v}^{pq\neg r}$. Moreover, together players 1 and 2 can compensate player 3 sufficiently for him to do so. For instance, this could be achieved by the transfer function τ', defined as

$$\tau'(i,j,\vec{v}) = \begin{cases} 1\frac{3}{4} & \text{if } i \in \{1,2\}, j = 3, \text{ and } \vec{v} = \vec{v}^{pq\neg r}, \\ 0 & \text{otherwise.} \end{cases}$$

Accordingly, $\{1,2,3\}$ is also a coalition blocking outcome $\vec{v}^{pqr}$. We may therefore conclude that $\vec{v}^{pqr}$ is eliminable by side-payments.

As transfer functions operate on cost functions only, it is immediate that hard equilibria cannot be eliminated via side-payments.

OBSERVATION 8. *Let $\vec{v}$ be a hard equilibrium of a game G, i.e., $\vec{v} \in \text{HARD}(G)$. Then, $\vec{v}$ cannot be eliminated via side-payments.*

Moreover, if an outcome $\vec{v}$ fails to be a Nash equilibrium, there is some player i and some $v_i' \in \mathcal{V}_i$ with $u_i(\vec{v}_{-i}, v_i') > u_i(\vec{v})$. Then $\vec{v}$ is eliminable by the singleton coalition $\{i\}$ via the transfer function that assigns cost zero to all players at all outcomes.

OBSERVATION 9. *Let $\vec{v}$ be an outcome of game G such that $\vec{v}$ is not a Nash equilibrium. Then $\vec{v}$ is eliminable via side-payments.*

The previous two observations show that in every game there is a class of outcomes that can never be eliminated via side-payments (the hard equilibria) as well as a class of outcomes that can always be eliminated via side-payments (the outcomes that are not Nash equilibria). There may, however, very well be outcomes in a game that do not belong to either of these classes. Together with Observations 8 and 9, the following result establishes a full characterisation of all outcomes that are eliminable via side-payments in a game.

PROPOSITION 10. *Let $\vec{v}$ be a present equilibrium of a Boolean game G with a* global *cost function c. Then, $\vec{v}$ is eliminable via side-payments if and only if there is a coalition C, a player $i \in C$ with $c_i(\vec{v}) > 0$, and a $v_i' \in \mathcal{V}_i$ such that $\vec{v} \models \gamma_j$ implies $(\vec{v}_{-i}, v_i') \models \gamma_j$ for all $j \in C$, and one of the following conditions holds:*

(i) $\vec{v} \not\models \gamma_j$ and $(\vec{v}_{-i}, v_i') \models \gamma_j$ for some player $j \in C$, or

(ii) $\sum_{j \in C} c_j(\vec{v}_{-i}, v_i') < \sum_{j \in C} c_j(\vec{v})$.

PROOF. First assume that $\vec{v}$ is eliminable via side-payments. Then, there is some coalition C, some $i \in C$, some $v_i' \in \mathcal{V}_i$, and some admissible transfer function τ only involving C such that,

$$u_i^\tau(\vec{v}_{-i}, v_i') > u_i^\tau(\vec{v}), \text{ and}$$
$$u_j^\tau(\vec{v}_{-i}, v_i') > u_j(\vec{v}) \quad \text{for all players } j \in C.$$

From the latter follows that $\vec{v} \models \gamma_j$ implies $(\vec{v}_{-i}, v_i') \models \gamma_j$ for all $j \in C$. Assume that $\vec{v} \not\models \gamma_j$ and $(\vec{v}_{-i}, v_i') \models \gamma_j$ for no $j \in C$. It follows that $\vec{v} \not\models \gamma_j$ if and only if $(\vec{v}_{-i}, v_i') \models \gamma_j$ for all $j \in C$. Since $u_j^\tau(\vec{v}_{-i}, v_i') > u_j(\vec{v})$ for all players $j \in C$, we also have that,

$$\sum_{j \in C} c_j^\tau(\vec{v}_{-i}, v_i') < \sum_{j \in C} c_j(\vec{v}).$$

By admissibility of τ, in particular, $c_i^\tau(\vec{v}_{-i}, v_i') \geq 0$. Hence, also $c_i(\vec{v}) > 0$. Recall that $\sum_{j \in C} \tau_j(\vec{v}_{-i}, v_i') = 0$. Therefore,

$$\begin{aligned} \sum_{j \in C} c_j^\tau(\vec{v}_{-i}, v_i') &= \sum_{j \in C} (c_j(\vec{v}_{-i}, v_i') - \tau_j(\vec{v}_{-i}, v_i')) \\ &= \sum_{j \in C} c_j(\vec{v}_{-i}, v_i') - \sum_{j \in C} \tau_j(\vec{v}_{-i}, v_i') \\ &= \sum_{j \in C} c_j(\vec{v}_{-i}, v_i'). \end{aligned}$$

Thus, finally we obtain that

$$\sum_{j \in C} c_j(\vec{v}_{-i}, v_i') < \sum_{j \in C} c_j(\vec{v}).$$

For the opposite direction, assume $c_i(\vec{v}) > 0$ and let C be a coalition, $i \in C$, and $v_i' \in \mathcal{V}_i$. Also assume that $\vec{v} \models \gamma_j$ implies $(\vec{v}_{-i}, v_i') \models \gamma_j$ for all $j \in C$. First, let $\vec{v} \not\models \gamma_j$ and $(\vec{v}_{-i}, v_i') \models \gamma_j$ for some player $j \in C$. Observe that $j \neq i$, otherwise $\vec{v}$ would not be a present equilibrium of G. As $\vec{v} \models \gamma_i$ implies $(\vec{v}_{-i}, v_i') \models \gamma_i$, we may also assume that $c_i(\vec{v}_{-i}, v_i') - c_i(\vec{v}) \geq 0$. Let ϵ be such that $0 < \epsilon \leq c_i(\vec{v})$ and define the transfer function τ such that

$$\tau(j, i, (\vec{v}_{-i}, v_i')) = c_i(\vec{v}_{-i}, v_i') - c_i(\vec{v}) + \epsilon$$

and $\tau(k, k', \vec{w}) = 0$ for all $k \neq j$, $k' \neq i$, and $\vec{w} \neq (\vec{v}_{-i}, v_i')$. Observe that τ is admissible. Then, coalition $\{i, j\}$ blocks $\vec{v}$ via τ. Finally, we may assume that $\vec{v} \models \gamma_j$ if and only if $(\vec{v}_{-i}, v_i') \models \gamma_j$ for all $j \in C$ and that $\sum_{j \in C} c_j(\vec{v}_{-i}, v_i') < \sum_{j \in C} c_j(\vec{v})$. Without loss of generality we may assume that $c_j(\vec{v}) - c_j(\vec{v}_{-i}, v_i') > 0$ for all $j \in C \setminus \{i\}$. As $\vec{v}$ is a present equilibrium, moreover, $c_i(\vec{v}_{-i}, v_i') - c_i(\vec{v})) > 0$. Now, for some positive ϵ,

$$c_i(\vec{v}_{-i}, v_i') - c_i(\vec{v}) + \epsilon = \sum_{j \in C \setminus \{i\}} c_j(\vec{v}) - c_j(\vec{v}_{-i}, v_i').$$

Let $0 < \delta \leq \min(c_i(\vec{v}), \epsilon)$ and define the transfer function τ such that for every $j \in C \setminus \{i\}$,

$$\tau(j, i, w) = \frac{(c_i(\vec{v}_{-i}, v_i') - c_i(\vec{v}) + \delta)(c_j(\vec{v}) - c_j(\vec{v}_{-i}, v_i')}{\sum_{k \in C \setminus \{i\}} (c_k(v) - c_k(\vec{v}_{-i}, v_i'))}.$$

and $\tau(k, k', \vec{w}) = 0$ for all $k \neq j$, $k' \neq i$, and $\vec{w} \neq (\vec{v}_{-i}, v_i')$. Observe that τ is admissible. Moreover, it can now easily be established that for every $j \in C$, we have that $u_j^\tau(\vec{v}_{-i}, v_i') > u_j(\vec{v})$. As, moreover, $c_i^\tau(\vec{v}) = c_i(\vec{v})$, we may conclude that $\vec{v}$ can be eliminated by coalition C via τ. □

Notice that this result characterises present equilibria that are eliminable via side-payments *without making reference to transfer functions.* Also, a closer inspection of the proof shows that, if a present

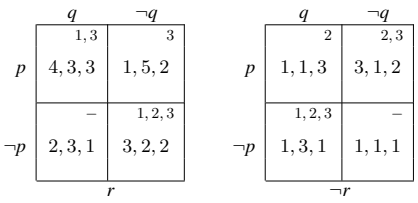

Figure 4: A three-player game illustrating some features of coalition merging

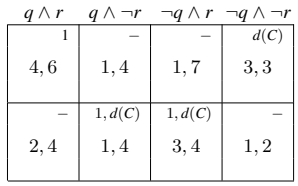

Figure 5: The reduced game resulting from the game in Figure 4 by merging players 2 and 3 into one player $d(C)$

equilibrium can be eliminated, then a coalition need compensate only one of its members. This is exactly the player that has to be incentivised to deviate from the equilibrium in question.

COROLLARY 11. *If $\vec{v}$ is a present equilibrium of a game G that can be eliminated via side-payments, then this can be achieved by means of a transfer function τ for which side-payments are made to only one player at only one outcome $(\vec{v}_{-i}, v_i')$, where $v_i \in \mathcal{V}_i$.*

The corollary tells us something about the number of coalition members receiving the transfer, but it says very little about the number of coalition members making the payments. In fact, the compensation costs required to bring about the intended deviation may be so high that we need many players to bear them. In some specific cases, however, also the number of coalition members compensating the deviating player can be extremely small.

PROPOSITION 12. *Let $\vec{v}$ be a present equilibrium of a Boolean game G with a* local cost function $\hat{c}$. *Then, $\vec{v}$ is eliminable via side-payments if and only if there are two distinct players i and j, and some $v_i' \in \mathcal{V}_i$ such that both*

(i) $\vec{v} \models \gamma_i$ if and only if $(\vec{v}_{-i}, v_i') \models \gamma_i$, and

(ii) $\vec{v} \not\models \gamma_j$ and $(\vec{v}_{-i}, v_i') \models \gamma_j$.

The main idea underlying this proposition is that, if a player i can be incentivised to deviate from a present equilibrium $\vec{v}$ to $(\vec{v}_{-i}, v_i')$ by some coalition C, then i must receive some compensation to do so. Some other player j in C must then be willing to carry at least part of this cost. As the cost function $\hat{c}$ is local, however, we have $\hat{c}_j(\vec{v}) = \hat{c}_j(\vec{v}_{-i}, v_i')$. Thus, the costs for j in $(\vec{v}_{-i}, v_i')$ are *higher* after the side-payments took place than before. Therefore, for j to be willing to participate in the coalition C, he should have his goal γ_j satisfied at $(\vec{v}_{-i}, v_i')$, but not at $\vec{v}$. Then, however, player j would be prepared to pay any price to make i deviate from $\vec{v}$ to $(\vec{v}_{-i}, v_i')$. It follows that the two-player coalition $\{i, j\}$ would also block outcome $\vec{v}$.

It is worth observing that the right-hand side of Proposition 12 only involves conditions relating to the realisation of players' goals. In the setting with local cost functions, whether an outcome is eliminable via side-payments is quite independent of the actual cost function.

Summarising, Corollary 11 tells us that blocking coalitions can be *greedy*, i.e., they can design such transfers only paying one player and without transferring money among themselves. Proposition 12 tells us that, when goal realisation is at stake, blocking coalitions can be *small*: a player i whose goal is not satisfied is prepared to compensate any additional costs that another player j incurs if j (unilaterally) were to deviate to an outcome that does satisfy i's goal.

4.2 Coalition merging

In this section, we explore the idea of *internalising* externalities by forming large enough coalitions to eliminate the potential interference among the players in a Boolean game. In particular, we consider the extent to which we can facilitate positive externalities and eliminate negative externalities by merging players.

Just as we do when imposing taxation schemes on games or by allowing coalitions to make side-payments, by merging players we can transform the structure of the game. In particular, we can modify its original equilibria. To explore the properties of coalition merging, we need to establish some notational conventions first.

Let $G = (N, \Phi, c, (\gamma_i)_{i\in N}, (\Phi_i)_{i\in N})$ be a Boolean game and C a subset of players in G. We denote by G_C *the game obtained from G by merging the players in C into a single player*. In this context merging the players in C means that the coalition C operates as a single player $d(C)$ aiming to satisfy all its members' goals at the same time, while controlling all their variables simultaneously. The costs incurred by C are then the joint costs of C. Formally,

$$G_C = (N', \Phi, c', (\gamma_i')_{i\in N'}, (\Phi_i')_{i\in N'}),$$

where $N' = (N \setminus C) \cup \{d(C)\}$ for some $d(C) \notin N$, and

$$\Phi'_{d(C)} = \bigcup_{j\in C} \Phi_j \qquad\qquad \gamma'_{d(C)} = \bigwedge_{i\in C} \gamma_i.$$

For all $i \notin C$, we have $\Phi_j' = \Phi_j$ and $\gamma_j' = \gamma_j$. The cost function c' is such that, for all outcomes $\vec{v}$,

$$c_i'(\vec{v}) = \begin{cases} \sum_{i\in C} c_i(\vec{v}) & \text{if } i = d(C), \\ c_i(\vec{v}) & \text{otherwise.} \end{cases}$$

In words, the cost function of player i in the updated game G_C yields, at each outcome, the sum of costs of all members of coalition C at that outcome, if player i is the player $d(C)$, i.e., if player i is the result of the merging of coalition C. Otherwise, it leaves the costs for player i unchanged.

We refer to G_C as a *reduced* game, and the game G from which G_C is derived as the *original* game. Let us first consider an example to illustrate the concept.

EXAMPLE 13. *Consider the game depicted in Figure 4, where player 1 chooses values for p, player 2 for q, and player 3 for r. The original game has three Nash equilibria, viz., the outcomes $\vec{v}^{pqr}$, $\vec{v}^{\neg p\neg qr}$, and $\vec{v}^{\neg pq\neg r}$, the latter two of which are hard. By merging all players into one, only $\vec{v}^{\neg pq\neg r}$ remains as an equilibrium. It satisfies all players' goals and, hence also $d(N)$'s. Although outcome $\vec{v}^{\neg p\neg qr}$ does this as well, it does so at a considerably higher cost to $d(N)$, viz., 7 versus 5. This shows that coalition merging can eliminate hard equilibria as well as the hardness of equilibria: both $\vec{v}^{\neg p\neg qr}$ and $\vec{v}^{pq\neg r}$ are soft equilibria in G_N.*

Now, consider the (soft) equilibrium $\vec{v}^{pqr}$. Observe that this equilibrium cannot be eliminated via side-payments. Player 1 will have

part nor parcel in any blocking coalition. Player 3 *cannot be incentivised to deviate to* $\vec{v}^{pq\neg r}$ *nor would he be willing to compensate player* 2 *sufficiently if she were to deviate to* $\vec{v}^{p\neg q r}$*. Observe, however, that, if player* 2 *and* 3 *were to merge, as depicted in Figure 5, they could deviate to* $\vec{v}^{p\neg q\neg r}$*, benefitting both players.*

Example 13 illustrates several properties of coalition merging. First and foremost, it shows that it can lead to hard equilibria of the original game being removed as equilibria from the reduced game.

OBSERVATION 14. *Coalition merging does not preserve hard equilibria, i.e., there are games G, outcomes* $\vec{v}$*, and coalitions C such that* $\vec{v} \in \text{HARD}(G)$ *and* $\vec{v} \notin \text{NE}(G_C)$.

In the extreme case in which the grand coalition of all players are merged, we find that the conditions for a hard equilibrium in game G to be preserved as a hard equilibrium in G_N are very restrictive indeed. This is shown by the following proposition, which is an almost immediate consequence of Proposition 5.

PROPOSITION 15. *Let G be a game and* $\vec{v}$ *an outcome. Then,* $\vec{v}$ *is a hard equilibrium in* G_N *if and only if* $\vec{v}$ *is the* only *outcome in G such that* $\vec{v} \models \gamma_i$ *for all* $i \in N$.

The operation G_C applied to an original game G is well defined for every coalition C. One might, however, consider only the classes of games where a certain group of players have compatible objectives. We will say that a group of players C are *compatible* if they have mutually consistent goals, i.e., the formula $\bigwedge_{i\in C} \gamma_i$ is satisfiable. We believe that focusing on coalitions that have compatible goals—and even impose compatibility as a requirement for coalition merging—is a desirable and intuitive feature. While we leave a thorough exploration of goal-compatible coalition merging to future work, we provide a result for the G_N case next.

PROPOSITION 16. *For every game G in which the set of all players N is compatible, we have* $\text{NE}(G_N) \neq \emptyset$*, and for every* $\vec{v} \in \text{NE}(G_N)$ *we have* $\vec{v} \models \bigwedge_{i\in N} \gamma_i$.

Finally, we focus on the relation of equilibrium eliminability via coalition merging and via side-payments. Observation 14, together with Observation 8, immediately shows that coalition merging can eliminate equilibria that cannot be eliminated via side-payments.

OBSERVATION 17. *There are games with equilibria that can be eliminated by merging coalitions but not by side-payments.*

Thus, one might suspect that every equilibrium that can be eliminated via side-payments can also be eliminated by merging coalitions. Example 18 shows that this is not the case.

EXAMPLE 18. *Consider again the game in Figure 4, but now assume that in* $\vec{v}^{\neg p\neg q r}$ *only player* 1 *achieves his goal and in and in* $\vec{v}^{\neg pq\neg r}$ *only players* 1 *and* 3 *theirs. Then, outcome* $\vec{v}^{\neg p\neg q\neg r}$ *is an equilibrium, be it one in which none of the players' goals is satisfied. Clearly, this outcome can eliminated via side-payments. In fact, every non-singleton coalition is blocking outcome* $\vec{v}^{\neg p\neg q\neg r}$ *and contains players who are prepared to compensate fully the costs incurred by some player deviating from* $\vec{v}^{\neg p\neg q\neg r}$*. However, no matter how you merge coalitions, outcome* $\vec{v}^{\neg p\neg q\neg r}$ *will remain a Nash equilibrium in the reduced game due to its low costs.*

Thus, we can make the following final observation, showing that side-payments and coalition merging are complementary tools to eliminate undesirable equilibria.

OBSERVATION 19. *There are games with equilibria that can be eliminated via side-payments but not by merging coalitions.*

5. SUMMARY

The problems of eliminating undesirable equilibria and facilitating desirable equilibria are fundamental in economics and multi-agent systems. By focussing on Boolean games with costs, in which the players have quasi-dichotomous preferences, we were able to distinguish between hard and soft equilibria in games, i.e., equilibria that are equilibria irrespective of the cost function and those that may or may not be equilibria, depending on the cost function. We studied techniques by which undesirable equilibria can be eliminated, i.e., ways in which the game can be modified so that these outcomes are no longer stable: coalitions making side-payments and merging coalitions. We found that these two ways behave quite differently. In particular, even if by coalition merging hard equilibria may get eliminated, coalition merging is not stronger than side-payments: there may be equilibria that can be eliminated via side-payments but not via coalition merging.

Acknowledgments

Michael Wooldridge and Paul Harrenstein are supported by the ERC under Advanced Grant 291528 ("RACE"). 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").

6. REFERENCES

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

[2] E. Bonzon, M. Lagasquie, J. Lang, and B. Zanuttini. Boolean games revisited. In *Proc. ECAI-2006*.

[3] R. H. Coase. The nature of the firm. *Economica*, 4(16):386–405, 1937.

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

[5] P. E. Dunne, S. Kraus, W. van der Hoek, and M. Wooldridge. Cooperative Boolean games. In *Proc. AAMAS-2008*.

[6] P. Harrenstein, W. van der Hoek, J.-J. Meyer, and C. Witteveen. Boolean games. In *Proc. TARK VIII*, 2001.

[7] M. O. Jackson and S. Wilkie. Endogenous games and mechanisms: Side payments among players. *Rev. of Econ. Stud.*, 72(2):543–566, 2005.

[8] E. S. Maskin. The invisible hand and externalities. *Amer. Econ. Rev.*, 84(2):333–337, 1994.

[9] J. Meade. External economies and diseconomies in a competitive situation. *Econ. Jnl.*, 62(245):54–67, 1952.

[10] N. Nisan. Introduction to mechanism design (for computer scientists). In *Algorithmic Game Theory*. Cambridge UP, 2007.

[11] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.

[12] P. Turrini. Endogenous Boolean games. In *Proc. IJCAI-2013*.

[13] H. Varian. *Microeconomic Analysis (3rd edition)*. W. W. Norton, 1992.

[14] M. Wooldridge. Bad equilibria (and what to do about them). In *Proc. ECAI-2012*, 2012.

[15] M. Wooldridge, U. Endriss, S. Kraus, and J. Lang. Incentive engineering for Boolean games. *Artif. Intell.*, 195:418–439, 2013.