Paul Harrenstein
Paolo Turrini
Michael Wooldridge

Hard and Soft Preparation Sets in Boolean Games

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 avoid such undesirable properties, for example through the use of subsidies and taxation, or by allowing players to undergo a bargaining phase before their decision. In this paper, we consider the effect of such transformations in Boolean games with costs, where players control propositional variables that they can set to true or false, and are primarily motivated to seek the satisfaction of some goal formula, while secondarily motivated to minimise the costs of their actions. We adopt *(pure) preparation sets* (prep sets) as our basic solution concept. A preparation set is a set of outcomes that contains for every player at least one best response to every outcome in the set. Prep sets are well-suited to the analysis of Boolean games, because we can naturally represent prep sets as propositional formulas, which in turn allows us to refer to *prep formulas*. The preference structure of Boolean games with costs makes it possible to distinguish between *hard* and *soft* prep sets. The hard prep sets of a game are sets of valuations that would be prep sets in that game no matter what the cost function of the game was. The properties defined by hard prep sets typically relate to goal-seeking behaviour, and as such these properties cannot be eliminated from games by, for example, taxation or subsidies. In contrast, soft prep sets can be eliminated by an appropriate system of incentives. Besides considering what can happen in a game by unrestricted manipulation of players' cost function, we also investigate several mechanisms that allow groups of players to form coalitions and eliminate undesirable outcomes from the game, even when taxes or subsidies are not a possibility.

Keywords: Boolean games, Set-valued solutions concepts, Preparation sets, Nash equilibrium, Externalities, Coalition formation.

1. Introduction

A fundamental problem in the theory of games is that a game may contain equilibria with undesirable properties. To take a famous example, in the Prisoner's Dilemma the unique pure strategy Nash equilibrium—mutual

Special Issue: Logical Aspects of Multi-Agent Systems
Edited by **Nils Bulling and Wiebe van der Hoek**

Studia Logica (2016) 104: 813–847
DOI: 10.1007/s11225-015-9629-9

defection—is the only outcome of the game that is not Pareto optimal, and it is, moreover, strictly worse for both players than the alternative outcome of mutual cooperation. From an external 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. Such mechanisms thus modify 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 another, or joining forces in coalitions to their mutual benefit. Both solutions have been studied in the economic literature since its early stages [7,8,17,23].

Boolean games (of the form studied in [22,25]) represent an important domain for investigating these issues, because preferences in such games have a special 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 [22], 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 the 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.

The possibility of manipulating Boolean games with costs in order to eliminate undesirable Nash equilibria was studied by Wooldridge et al. [25], who considered the possibility of introducing taxation schemes to influence the behaviour of players towards or against certain outcomes, and by Turrini [22], who considered the possibility of a pre-play bargaining phase for such games. While the present paper is also concerned with manipulating Boolean games with costs, it extends previous work in three key respects.

- First, we adopt as our basic analytical solution concept the notion of a *preparation set* ("prep set"). Introduced by Voorneveld [24], a prep set is a set of game outcomes that contains for every player at least one best response to every outcome in the set. As preparation is a set-valued solution concept, we find prep sets to be well-suited to the analysis of Boolean games.[1] In particular, because propositional formulas characterise sets of outcomes, we can naturally use propositional formulas to characterise prep sets themselves. We thus refer to "prep formulas" to mean propositional formulas that denote prep sets.

- Second, we investigate prep sets that are immune to manipulation by cost-based incentives. We refer to a *hard* prep set as a prep set that will be present in the game no matter what the cost function of the game is. Hard prep sets are important because they cannot be eliminated from the game through taxation schemes or by other ways of manipulating the cost structure of the game. In contrast, a *soft* prep set is a prep set that is present in a game for some cost functions, but not all. Given the correspondence between propositional formulas and sets of outcomes, we find it natural to refer to *hard formulas* and *soft formulas*, with the obvious interpretations. Clearly, the presence of hard prep sets with undesirable properties would be bad news: no cost incentives would be able to tempt players away from such prep sets. But, dually, hard prep sets with desirable properties are good news, as their achievement does not depend on the availability of material resources. We will see that when hard formulas are satisfied at only one outcome, what we call *hard equilibria*, these outcomes satisfy desirable properties from the point of view of society. The terminology of hard and soft equilibria was introduced by Harrenstein et al. [11]; here we extend and generalise that work.
- Third, we turn to the issue of managing these sets of outcomes, using ideas from the economics literature [7,8,16,17,23]. We first consider the possibility of groups of players engineering 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 [8,16,17]. Second, we study the possibility of a player taking into account 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* [7,23]. We investigate how such a mechanism can affect the set of rational states of a game. In particular, we show that by allowing players to merge in coalitions, hard formulas can sometimes be eliminated.

We begin, in the following section, with an introduction to the basic game model we work with in the remainder of the paper.

---

[1] We thank an anonymous reviewer for drawing our attention to Voorneveld's paper.

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., [4,6,10,12,22,25]). Although their primary interest is as an abstract theoretical model, Boolean games have an proved to have an increasing number of applications. For example, Levit et al. use Boolean games for modelling recharging schemes for electric vehicles [14], while in another paper Levit et al. Boolean games to model traffic signalling systems [15].

In this paper, we use the Boolean games model with cost functions, in which the players have quasi-dichotomous preferences, as in [25]. Thus, each player is primarily interested in satisfying a goal, which is expressed by a Boolean formula; this goal formula classifies each outcome as either desirable or undesirable. Costs form a secondary concern for players: between two outcomes that either both satisfy or both fail to satisfy the player's goal, the player will prefer the outcome that minimises costs.

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 L the set of well-formed formulas 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. Where v is a valuation and φ is a propositional formula, 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. If a formula φ is satisfied by at least one valuation, we say that φ is **satisfiable**. If φ is satisfied at every valuation, φ is said to be a **tautology**.

We will frequently exploit the fact that a propositional formula corresponds to a set of valuations, its **truth set**. Where φ is a propositional formula, we denote by φ^G the set of valuations that satisfy it. We frequently find it convenient to refer to formulas as if they were sets of valuations, and to sets of valuations as if they were formulas, understanding by this terminology that we are referring to the relationship above.

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 L-formula γ_i. Every agent $i \in N$, moreover, **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 $\Phi_1 \cup \cdots \cup \Phi_n = \Phi$. A **choice** for agent $i \in N$ is defined by a function $v_i : \Phi_i \to \mathbb{B}$, i.e., an allocation of truth or falsity to all the variables under i's control. Let $\mathcal{V}_i$ denote the set of choices for agent i.

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 outcome $\vec{v}$ satisfies formula φ. We let $\vec{\mathcal{V}}$ denote the set of outcomes. Where $\vec{v} = (v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_n)$ and $v_i' \in V_i$, 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 her goal achieved, then *she will prefer to choose one that minimises her costs*, whereas, if an agent cannot get her goal achieved, then *she simply chooses to minimise her costs*.

To capture these preferences, we introduce two types of cost functions: *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

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

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. [25] assumed a natural *additive* (and more concise) model for costs given by **local cost functions** of the form

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

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:

$$\hat{c}(i, \vec{v}) = \sum_{p \in \Phi_i} \hat{c}(p, v_i(p)).$$

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 now introduce the utility functions that model the players' preferences. Let μ_i denote the cost of the most expensive outcome for agent i, i.e.,

$$\mu_i = \max_{\vec{v} \in \vec{V}} u_i(\vec{v}).$$

The **utility** to agent i of an outcome $\vec{v}$ 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). It is worth observing that the expression $1 + \mu_i$ is merely used to ensure that the valuations satisfying i's goal are guaranteed to get a higher utility than the ones which do not. In other words the construction of our utility function is only meant to encode an ordinal preference relation among outcomes. The following observation illustrates this.

OBSERVATION 1. *Let i be an agent, c_i a cost function, and γ_i a goal formula. For each pair of outcomes $\vec{v}$ and $\vec{v}'$, we have that $u_i(\vec{v}) > u_i(\vec{v}')$ if and only if either (i) both $\vec{v}' \models \neg\gamma_i$ and $\vec{v} \models \gamma_i$, or (ii) both $\vec{v}' \models \gamma_i$ if and only if $\vec{v} \models \gamma_i$ and $c_i(\vec{v}') > c_i(\vec{v})$.*

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, $\gamma_i \in L$ 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, standard game-theoretic solution concepts are available for the analysis of Boolean games [19]. We first recall the notion of 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 $\textsc{ne}(G)$ denote the set of all Nash equilibria of the game G. Let us now consider an example.

Example 2. Consider the Boolean game with $N = \{1, 2, 3\}$, $\Phi_1 = \{p\}$, $\Phi_2 = \{q\}$, and $\Phi_3 = \{r\}$. The goals of the players are given by $\gamma_1 = (r \wedge q) \vee (p \wedge \neg r)$, $\gamma_2 = (r \rightarrow \neg p) \wedge (q \rightarrow r)$, 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 $p\bar{q}r$ and for the other outcomes likewise. We adopt a similar notational convention for choices of players, writing, for instance, $\bar{p}$ for the choice of the player controlling p to set p to false. Figure 1 gives graphical representation of this game and also specifies the cost function.

This game has two Nash equilibria, namely pqr and $p\bar{q}\bar{r}$. In the latter, all players achieve their goals, whereas in pqr player 2 fails to get her goal satisfied. If player 2 and player 3 choose to set q and r to true respectively,

r:

	q	$\bar{q}$
p	^{1,3} 2, 3, 3	$^{-}$ 1, 4, 2
$\bar{p}$	^{1,2,3} 3, 1, 1	^{2} 3, 2, 3

$\bar{r}$:

	q	$\bar{q}$
p	^{1,3} 0, 1, 6	^{1,2,3} 2, 2, 2
$\bar{p}$	^{3} 3, 2, 3	^{2,3} 1, 4, 1

Figure 1. A three-player game, in which player 1 controls p and chooses rows, player 2 controls q and chooses columns, and player 3 controls r and chooses matrices. 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 players 1, player 2, and player 3, respectively

player 1 *could* choose an action that satisfies not only his goal but also those of players 1 and 3—e.g., setting p to false—he will not rationally do so: by setting p to true he would achieve his goal at a lower cost. It could be argued that the equilibrium pqr is socially undesirable, because there is an alternative outcome of the game in which *all* players's goals are satisfied.

This paper is concerned with formulas representing more or less desirable properties in a game situation. In our logical setting, properties correspond to sets of outcomes. It makes therefore good sense to generalise the point-valued notion of Nash equilibrium, which applies to single outcomes, to a set-valued solution concept. In the context of game theory, point-valued concepts like Nash equilibrium may appear to be predominant. While making their case for set-valued solution concepts, however, Duwfenberg et al. [9] note in passing that in the early stages of game theory set-valued concept were actually prevailing [9, p. 120, footnote 6]. One main advantage of set-valued concepts is that their existence is generally guaranteed, even in absence of randomised strategies.[2]

This chimes well with the literature on Boolean games, where usually only pure strategies are considered. In this context, Basu and Weibull's *CURB sets* are worthy of special mention. CURB sets are defined as sets of outcomes that contain for every player all best responses to every outcome in the set [2]. CURB sets combine the guaranteed existence of (mixed) Nash equilibrium with the stability of strict Nash equilibrium. In this paper, however, we adopt Voorneveld's related notion of *prep sets*, where prep stands for "preparation" [24]. A prep set is defined as CURB sets, except that a prep set only needs to contain one best response of every player to every outcome it contains. As such "the minimal prep notion can be seen as a set-valued extension of the pure Nash equilibrium concept" [24, p. 407]. As we will see, this feature makes them especially suitable for the purposes of the present paper. Although prep sets were originally defined for settings in which players can randomise over their choices, we restrict our attention here to pure (non-randomised) prep sets.

First, we say for an agent i that a choice $v_i' \in V_i$ is a **best response** to an outcome $\vec{v} \in V$ if for all $v_i'' \in V_i$,

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

[2] Norde et al. [18] show that Nash equilibrium is the only point-valued solution concept for strategic games whose existence is guaranteed, is consistent with utility maximising behaviour, and satisfies consistency, an intriguing property (also see [1, pp. 478–479]). This result has sparked new interest in set-valued concepts among game theorists.

The set of best responses of a player i to an outcome $\vec{v}$, we denote by $BR_i(\vec{v})$.[3] Notice that an outcome $\vec{v} = (v_1, \ldots, v_n)$ is a Nash equilibrium if and only if v_i is a best response to $\vec{v}$ for all players i. Then, a **(pure) prep set** is defined as a non-empty X of outcomes such that,

(i) X is a product set, i.e., $X = X_1 \times \cdots \times X_n$ for (non-empty) subsets $X_1 \subseteq V_1, \ldots, X_n \subseteq V_n$, and

(ii) for every player i and every outcome $\vec{v} \in X$, there is a best response $v'_i \in BR_i(\vec{v})$ of i to $\vec{v}$ such that $(\vec{v}_{-i}, v'_i) \in X$, i.e., $\{(\vec{v}_{-i}, v'_i) : v'_i \in BR_i(\vec{v})\} \cap X \neq \emptyset$.

We let $\textsc{prep}(G)$ denote the set of all prep sets of the game G. If φ is a propositional formula such that $\varphi^G \in \textsc{prep}(G)$, then we say that φ is a *prep formula*, since its satisfying valuations characterise a prep set. Given this, we will frequently abuse notation by writing $\varphi \in \textsc{prep}(G)$ for $\varphi^G \in \textsc{prep}(G)$. As prep sets are defined as product sets, prep formulas likwise have a specific structure. Given a game G, it can easily be appreciated that φ^G is a product set if and only if φ is equivalent to a conjunction $\varphi_1 \wedge \cdots \wedge \varphi_n$, where each φ_i is a propositional formula over the set Φ_i of propositional variables controlled by player i. Such formulas we will refer to as **product fomulas (for G)**.[4] Observe that whether a formula is a product formula may depend on the game that is being considered. Having assumed the set Φ of propositional variables to be finite, it can also easily be seen that every product set $X = X_1 \times \cdots \times X_n$ of pure strategies in a game G can be expressed by a product formula φ, i.e., $\varphi^G = X$.

There is an obvious connection between pure prep sets and the (pure) Nash equilibria of a Boolean game: the set of pure Nash equilibria coincide with the singleton prep sets. This is in line with Voorneveld's original notion and is reflected in the following lemma.

---

[3] Voorneveld [24] defines prep sets in terms of (pure) best responses to profiles of *randomised* strategies. A profile $\vec{\sigma}_{-i} = (\sigma_1, \ldots, \sigma_{i-1}, \sigma_{i+1}, \ldots, \sigma_n)$ of randomised strategies, i.e., σ_j is a probability distribution over V_j, can then be interpreted as a *belief* of player i as to which actions his opponents may play. Thus, prep sets are also related to sets of rationalisable strategies in the sense of Bernheim [3] and Pearce [20]. For the purposes of this paper we restrict ourselves to profiles $(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n)$ of pure strategies, which may be interpreted as beliefs of player i that his opponents will play a certain profile with certainty.

[4] If in the definition of a prep set the restriction to product sets were dropped, all results in this paper would apply to general formulas.

LEMMA 3. *Let* $\vec{v} = (v_1, \ldots, v_n)$ *be an outcome of game* G. *Then,*

$$\vec{v} \in \text{NE}(G) \text{ if and only if } \{\vec{v}\} \in \text{PREP}(G).$$

PROOF. First assume that $\vec{v} \in \text{NE}(G)$ and consider $\{\vec{v}\}$ along with an arbitrary player i. By definition of Nash equilibrium, v_i is a best response to $\vec{v}$ and, obviously, $(\vec{v}_{-i}, v_i) \in \{\vec{v}\}$.

For the opposite direction, assume that $\{\vec{v}\}$ is a PREP set and consider an arbitrary player i. As v_i is the only choice v_i' for i such that $(\vec{v}_{-i}, v_i') \in \{\vec{v}\}$, it follows that v_i is a best response to $\vec{v}$. We may conclude that $\vec{v}$ is a Nash equilibrium in G. ■

By a similar argument, it can easily be appreciated that every subset of Nash equilibria of a game also constitutes a prep set, although the implication in the opposite direction does not hold. As we see in the next example, a prep set need not even contain a single Nash equilibrium.

EXAMPLE 4. Consider again the game of Example 2 and let us now concentrate on prep sets. We saw above that pqr and $p\bar{q}\bar{r}$ are Nash equilibria, and by Lemma 3, both $\{pqr\}$ and $\{p\bar{q}\bar{r}\}$ are prep sets. Trivially, the set of all outcomes is a prep set, and thus $\top$ is a prep formula. The game contains many other prep sets. For instance, both $\{pqr, \bar{p}qr\}$ and $\{p\bar{q}\bar{r}, \bar{p}\bar{q}\bar{r}\}$ are, whereas $\{p\bar{q}\bar{r}, pq\bar{r}\}$ is not. To see the latter, observe that player 3's best response at $pq\bar{r}$ would be to set r to true: that would result in pqr, which lies outside $\{p\bar{q}\bar{r}, pq\bar{r}\}$ but satisfies his goal at a lower cost. In this game, all prep sets contain a Nash equilibrium. This, however, is not necessarily the case. Consider for instance the three-player game depicted in Figure 2, which has one Nash equilibrium, namely, $\bar{p}\bar{q}\bar{r}\bar{s}$. Nevertheless, $p \wedge s$—representing $\{pqrs, p\bar{q}rs, pq\bar{r}s, p\bar{q}\bar{r}s\}$—is also a prep set but does not contain the equilibrium $\bar{p}\bar{q}\bar{r}\bar{s}$.

3. The Robustness of Prep Sets Under Costs

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 once more 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 show that the fact that there are two distinct drivers behind an agent's preferences gives a two-tier structure to the rational outcomes of Boolean games. Specifically, we distinguish between *hard* and

	r	$\bar{r}$
pq	2, 3 2, 1, 3	1, 3 3, 3, 4
$p\bar{q}$	1, 2, 3 1, 3, 3	2, 3 1, 2, 3
$\bar{p}q$	1, 2, 3 5, 1, 2	2 1, 3, 4
$\bar{p}\bar{q}$	2 6, 5, 2	2 4, 6, 5

s

	r	$\bar{r}$
pq	2 3, 2, 4	– 2, 2, 5
$p\bar{q}$	2 1, 4, 4	2 3, 2, 2
$\bar{p}q$	2, 3 2, 2, 1	2 2, 4, 3
$\bar{p}\bar{q}$	1, 2, 3 1, 5, 1	2 2, 3, 4

$\bar{s}$

Figure 2. A three-player game (notational conventions as in Figure 1). Here $p \wedge s$ is a prep formula but none of its satisfying outcomes, $pqrs$, $pq\bar{r}s$, $p\bar{q}rs$, and $p\bar{q}\bar{r}s$, is a Nash equilibrium. Despite $p \wedge s$ being soft, it is not eliminable via side-payments

soft prep sets. Informally, a hard prep set is one from which rational agents cannot escape, in the sense that, no matter what the cost function is, every best reply to an outcome in the set will lead to an outcome in the set. If we think of prep sets as formulas, then a hard formula is one in which every rational deviation from an outcome satisfying the formula will simply lead to another outcome satisfying the formula.

In contrast, a soft prep set is one whose presence in a game is contingent upon the cost function of the game. As a consequence, if a formula is soft, then it can potentially be eliminated from the game if it is viewed as undesirable—e.g., through taxes [23, p. 656]—or introduced to the game by providing appropriate incentives, if it is seen as desirable. To formalise this intuition, we first need some more notation.

3.1. Hard and Soft Prep Sets

Given a game G with cost function c, we denote by $G^{c'}$ the game obtained from G by replacing 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.

In the discussion that follows, the **zero cost function** $c^{\mathbf{0}}$ will be important. This is the cost function that 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.

We now define the set $\textsc{init}(G)$ of **initial prep sets** (briefly *initial sets*) of a game G to consist of the prep sets of the game $G^{\mathbf{0}}$, i.e.,

$$\textsc{init}(G) = \textsc{prep}(G^{\mathbf{0}}).$$

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

LEMMA 5. *For every game G,* $\textsc{prep}(G) \subseteq \textsc{init}(G)$.

PROOF. Consider an arbitrary game G with cost function c along with an arbitrary product set $X \subseteq \vec{V}$. Assume that $X \notin \textsc{prep}(G^{\mathbf{0}})$. Then, there is some outcome $\vec{v} \in X$ and some player i such that $(\vec{v}_{-i}, v'_i) \notin X$ for all best responses v'_i of i to $\vec{v}$ in $G^{\mathbf{0}}$. This implies $u^{\mathbf{0}}_i(\vec{v}) < u^{\mathbf{0}}_i(\vec{v}_{-i}, v'_i)$. As in $G^{\mathbf{0}}$ all costs are zero, it follows that $(\vec{v}_{-i}, v''_i) \not\models \gamma_i$ for all choices v''_i such that $(\vec{v}_{-i}, v''_i) \in X$. Moreover, as the set of best responses is always non-empty, we also have $(\vec{v}_{-i}, v'_i) \models \gamma_i$ for some choice $v'_i \in V_i$ with $(\vec{v}_{-i}, v'_i) \notin X$. For G, this means that $(\vec{v}_{-i}, v''_i) \notin X$ for all best responses v''_i of i to $\vec{v}$ and, hence $X \notin \textsc{prep}(G)$. ■

Thus, the game $G^{\mathbf{0}}$ contains the *maximal* set of prep sets with respect to G and all possible cost functions for G. In particular, if for some formula φ we have $\varphi^G \notin \textsc{prep}(G^{\mathbf{0}})$, then there is no possibility of introducing φ to G by imposing a cost function. Marginal costs defined within cost functions c serve to *eliminate* prep sets from a game $G^{\mathbf{0}}$.

We can now define the hard prep sets of a game G. Formally, the set of **hard prep sets** of game G are those sets of outcomes that are prep no matter what cost function we assign to the game, i.e.,

$$\textsc{hard}(G) = \bigcap_{c: N \times \vec{V} \to \mathbb{Q}_{\geq}} \textsc{prep}(G^c).$$

Thus, if $X \in \textsc{hard}(G)$, then X is a set of valuations that is "immune" to any cost considerations, because no matter what we do to the cost function of G, X will remain a prep set in the game.

In contrast, a soft prep set 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 prep sets as being the "malleable" part of a game: it is these sets that we can potentially eliminate from or introduce to games. Formally, $\textsc{soft}(G)$ denotes the set of **soft prep sets** of G, i.e.,

$$\textsc{soft}(G) = \textsc{init}(G) \setminus \textsc{hard}(G).$$

To understand this definition, recall that $\textsc{init}(G)$ is the maximal set of prep sets that could be present in a game. Note that a set of outcomes can be

$$\begin{array}{ccccc} \textsc{hard}(G) & \subseteq & \textsc{prep}(G) & \subseteq & \textsc{init}(G) \\ & & \cup\!\!\mid & & \cup\!\!\mid \\ & & \textsc{present}(G) & \subseteq & \textsc{soft}(G) \\ & & & & \cup\!\!\mid \\ & & & & \textsc{absent}(G) \end{array}$$

Figure 3. Containment relations between types of formulas

a soft prep set in a game without actually being a prep set in that game—but it will, however, be a prep set for the game with another cost function. For this reason, we will distinguish between soft prep sets that are present and those that are absent in a game.

Thus, we let $\textsc{present}(G)$ denote the set of soft prep sets of G that are present in G, and let $\textsc{absent}(G)$ denote the set of soft prep sets that are not present in G, i.e.,

$$\textsc{present}(G) = \textsc{prep}(G) \setminus \textsc{hard}(G),$$
$$\textsc{absent}(G) = \textsc{soft}(G) \setminus \textsc{present}(G).$$

Figure 3 illustrates the containment relations between these sets of sets; these all follow directly from the definitions presented above together with Lemma 5. We extend this terminology to single outcomes. Thus an outcome $\vec{v}$ is said to be *initial*, *hard*, *soft*, and *present* if $\{\vec{v}\}$ is a hard, soft, absent, and present prep set, respectively. On basis of Lemma 3, we also refer to initial, hard, soft, and present outcomes as **hard**, **soft**, **present**, and **absent equilibria**. With a slight abuse of notation we will sometimes omit braces when dealing with outcomes and write, e.g., $\vec{v} \in \textsc{prep}(G)$ instead of $\{\vec{v}\} \in \textsc{prep}(G)$.

As we did earlier, we also find it useful to refer to **initial**, **hard**, **soft**, and **present formulas**. Thus, when we say a formula φ is in hard we mean that the set of outcomes satisfying φ is a hard prep set and write $\varphi \in \textsc{hard}(G)$ for $\varphi^G \in \textsc{hard}(G)$. We adopt similar terminological conventions for initial, soft, and present formulas. Let us see an example.

Example 6. Consider again the game in Figure 1. There are several formulas of interest in this game. Consider for instance $\neg q \wedge \neg r$, which, notice, is insensitive to the choice of the row player. This formula is hard, and the reason is that all best replies that players have to outcomes in this set never leave the set itself. For player 1 we only need to inspect outcomes satisfying $\neg q \wedge \neg r$. His best response to both $\bar{p}\bar{q}\bar{r}$ and $p\bar{q}\bar{r}$ is setting p to true, either leading to $p\bar{q}\bar{r}$, which is contained in the truth set of $\neg q \wedge \neg r$. For players 2 and 3, it suffices to observe that both of them have their goal achieved in all outcomes satisfying $\neg q \wedge \neg r$, and that, if either of them were to deviate from any of these outcomes, a state would be reached that does not satisfy her goal. This means that, no matter what the cost function is, the best response of either 2 or 3 to any outcome satisfying $\neg q \wedge \neg r$ is the choice defined by the outcome itself. Notice that this would not be true if 2 and 3 were acting as a coalition, as they could reach for instance outcome $\bar{p}rq$ from outcome $\bar{p}\bar{r}\bar{q}$. In this example, all hard formulas are equivalent to either $p \wedge \neg q \wedge \neg r$ or $\neg q \wedge \neg r$.

Another interesting formula is p. This formula is satisfied at four outcomes, among which the two Nash equilibria in the game, viz., pqr and $p\bar{q}\bar{r}$. A closer inspection reveals that the former is soft and present, whereas the latter is hard. The formula as a whole is soft and present, as we could act on the first equilibrium and eliminate it.

There are many initial formulas. Clearly $\neg p \wedge q \wedge r$ is an initial formula, as $\bar{p}qr$ is an initial equilibrium achieved by the empty cost function $c^{\mathbf{0}}$. Still, $\bar{p}qr$ is not an equilibrium in the game itself. Thus, $\neg p \wedge q \wedge r$ is an absent formula. Not all product formulas, however, are initial. For instance, every product formula consistent with $\neg p \wedge q \wedge \neg r$ but not with $\neg p \wedge \neg q \wedge \neg r$ cannot be initial, as player 2 will always deviate from the outcome satisfying the first to the outcome satisfying the second.

The game in Figure 2 also has a number prep formulas—for instance $p \wedge s$ and $\neg(p \vee s)$—none of which, apart from $\top$, is hard. That is, all prep formulas are soft. This does not mean that all formulas are initial. For instance $\neg(r \vee s)$ is not: no matter which cost function, player 2 will want to deviate from $pq\bar{r}\bar{s}$ to $pqr\bar{s}$ by setting r to $\top$.

The concept of a hard formula is defined with respect to *all* (global) cost functions. The following lemma, however, shows that one only needs to consider the *local* cost functions to decide whether an outcome is a hard formula. 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 formulas.

LEMMA 7. *Let G be a game and φ be a satisfiable product formula. Then, $\varphi \in \textsc{hard}(G)$ if and only if $\varphi \in \textsc{prep}(G^c)$ for all* local *cost functions c.*

PROOF. The "only if"-direction is trivial. For the opposite direction, assume that $\varphi \notin \textsc{hard}(G)$. Then there is an outcome $\vec{v} \in \varphi^G$, a player i, and a cost function c such that $(\vec{v}_{-i}, v_i') \notin \varphi^G$ for all best responses v' of i to $\vec{v}$ in G^c. We may assume that such a best response v_i' to $\vec{v}$ exists. Observe that $\vec{v} \models \gamma_i$ then implies that $(\vec{v}_{-i}, v_i') \models \gamma_i$. Now define a local cost function $\hat{c}$ with c_i

such that for all $p \in \Phi_i$ and $b \in \{\top, \bot\}$,

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

Then, v_i' is i's single least expensive choice. It is now not hard to see that v_i' is the unique best response of i to $\vec{v}$ in $G^{\hat{c}}$. Since $(\vec{v}_{-i}, v_i') \notin \varphi^G$, it follows that $\vec{v} \notin \textsc{hard}(G^{\hat{c}})$. ■

The initial formulas of a game can be determined by inspecting the prep formulas in the corresponding game with the zero cost function $c^{\mathbf{0}}$. We find that, when the formula we want to know the hardness of is *given*, we can likewise restrict our attention to one particular cost function. To see this, let c^φ be the cost function that assigns to every player i and every outcome $\vec{v}$ cost 1 if v satisfies φ and assigns 0 otherwise, that is, for every player i,

$$c_i^\varphi(\vec{v}) = \begin{cases} 1 & \text{if } \vec{v} \models \varphi, \\ 0 & \text{otherwise.} \end{cases}$$

Then, we have the following proposition.

Proposition 8. *Let G be a game and φ a satisfiable product formula. Then, $\varphi \in \textsc{hard}(G)$ if and only if $\varphi \in \textsc{prep}(G^{c^\varphi})$.*

Proof. The "only if"-direction follows directly from the definition of hard formulas. The proof for the opposite direction is by contraposition. Assume that $\varphi \notin \textsc{hard}(G)$. This means that there exists a cost function c such that $\varphi \notin \textsc{prep}(G^c)$. Hence, there is some outcome $\vec{v}$ with $\vec{v} \models \varphi$ and some player i such that $(\vec{v}_{-i}, v_i') \not\models \varphi$ for all best responses v_i' of i to $\vec{v}$ in G^c. Observe that we may assume the existence of such a best response, which we will denote by v_i^*. We distinguish two cases: (i) $(\vec{v}_{-i}, v_i'') \not\models \gamma_i$ for all $v_i'' \in V_i$, or (ii) $(\vec{v}_{-i}, v_i'') \models \gamma_i$ for some $v_i'' \in V_i$.

First assume that case (i) obtains. Then, the best responses of i to $\vec{v}$ in G^{c^φ} are given by $\{v_i'' \in V_i \colon (\vec{v}, v_i'') \not\models \varphi\}$. This is immediate by the definition of c^φ and the fact that $(\vec{v}_{-i}, v_i^*) \not\models \varphi$. If, on the other hand, (i) holds, then $(\vec{v}_{-i}, v_i^*) \models \gamma_i$. Now again by the definition of c^φ and the fact that v_i^* exists, the best responses of i to $\vec{v}$ in G^{c^φ} are given by $\{v_i'' \in V_i \colon (\vec{v}_{-i}, v_i'') \models \neg\varphi \wedge \gamma_i\}$. In either case, we may conclude that $\varphi \notin \textsc{prep}(G^{c^\varphi})$, as desired. ■

3.2. Characterising Hard and Soft Formulas

Whether a set of outcomes is prep in a given game may well depend on the associated cost function. We find, however, that the sets of initial and hard formulas of a game can be characterised solely in terms of valuations and goal formulas without making reference to cost functions. In this sense, our characterisations could be said to be of a purely logical nature. On basis of their definitions, it also follows that similar characterisations can also be obtained for soft, present, and absent formulas.

First, we characterise what it means for a formula to be an initial prep set of a game.

PROPOSITION 9. *Let G be a game and φ a satisfiable product formula. Then, $\varphi \in \textsc{init}(G)$ if and only if for all players $i \in N$ and all outcomes $\vec{v} \in \varphi^G$,*

$$(\vec{v}_{-i}, v_i') \models \gamma_i \wedge \neg\varphi \text{ for some } v_i' \in V_i \text{ implies } (\vec{v}_{-i}, v_i'') \models \gamma_i \wedge \varphi \text{ for some } v_i'' \in V_i.$$

PROOF. For the "only if"-direction assume for contraposition that there is an outcome $\vec{v} \in \varphi^G$, a player $i \in N$, and a choice $v_i' \in V_i$ such that $(\vec{v}_{-i}, v_i') \models \gamma_i \wedge \neg\varphi$ whereas $(\vec{v}_{-i}, v_i'') \models \gamma_i \wedge \varphi$ for no $v_i'' \in V_i$. It follows that $(\vec{v}_{-i}, v_i^*) \not\models \varphi$ for all best responses v_i^* of i to $\vec{v}$ in $G^{\mathbf{0}}$. Hence, $\varphi \notin \textsc{prep}(G^{\mathbf{0}})$, that is, $\varphi \notin \textsc{init}(G)$ either.

For the opposite direction, assume, again for contraposition, that $\varphi \notin \textsc{init}(G)$, that is, $\varphi \notin \textsc{prep}(G^{\mathbf{0}})$. Accordingly, there is an outcome $\vec{v}$ and a player i such that $\vec{v} \models \varphi$ and $(\vec{v}_{-i}, w_i) \not\models \varphi$ for all best responses w_i of i to $\vec{v}$ in $G^{\mathbf{0}}$. As in $G^{\mathbf{0}}$ costs are everywhere equal and players are only interested in satisfying their goal, there is a best response v_i' of i to $\vec{v}$ in $G^{\mathbf{0}}$ with $(\vec{v}_{-i}, v_i') \models \gamma_i \wedge \neg\varphi$, whereas for all choices $v_i'' \in V_i$ with $(\vec{v}_{-i}, v_i'') \models \varphi$ we have $(\vec{v}_{-i}, v_i'') \not\models \gamma_i$. Hence, $(\vec{v}_{-i}, v_i'') \models \gamma_i \wedge \varphi$ for no choice $v_i'' \in V_i$, as desired. ■

For some formulas it is easy to see whether they are hard or not in a given game. For instance, $\top$ is a hard formula in every game. For other cases the task of identifying the hard formulas requires considerably more effort. The following proposition helps in this task, by establishing a purely logical characterisation of hard formulas in Boolean games.

PROPOSITION 10. *Let G be a game and φ a satisfiable product formula. Then, $\varphi \in \textsc{hard}(G)$ if and only if for all $\vec{v} \in \varphi^G$, all players $i \in N$, and all $v_i' \in V_i$, the following two conditions are satisfied:*

(i) $(\vec{v}_{-i}, v_i'') \not\models \gamma_i$ *for all* $v_i'' \in V_i$ *implies* $(\vec{v}_{-i}, v_i') \models \varphi$,

(ii) $(\vec{v}_{-i}, v_i') \models \gamma_i$ *implies* $(\vec{v}_{-i}, v_i') \models \varphi$.

PROOF. The "only if"-direction is by contraposition. Assume that for some $\vec{v} \in \varphi^G$, some player i and some $v_i' \in V_i$ either (i') $(\vec{v}_{-i}, v_i'') \not\models \gamma_i$ for all $v_i'' \in \mathcal{V}_i$ and $(\vec{v}_{-i}, v_i') \not\models \varphi$, or (ii') both $(\vec{v}_{-i}, v_i') \models \gamma_i$ and $(\vec{v}_{-i}, v_i') \not\models \varphi$. Define $\hat{c}$ as the local cost function with $\hat{c}_i$ such that, for all $p \in \Phi_i$ and

$b \in \{\bot, \top\}$,

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

Thus, v_i' is the strictly cheapest choice for i under $\hat{c}$, that is, $\hat{c}_i(\vec{v}_{-i}, v_i') > \hat{c}_i(\vec{v}_{-i}, v_i'')$ for all $v_i'' \in \mathcal{V}_i \setminus \{v_i'\}$. Both in case (i') and in case (ii'), choice v_i' is the unique best response of i to $\vec{v}$ under $\hat{c}$. As $(\vec{v}_{-i}, v_i') \not\models \varphi$, we obtain in either case that $\varphi \notin \text{PREP}(G^{\hat{c}})$ and subsequently that $\varphi \notin \text{HARD}(G)$.

For the "if"-direction, assume that for all $\vec{v} \in \varphi^G$, all players i, and all $v_i \in V_i$ both (i) and (ii) are satisfied. Let c be an arbitrary cost function; we show that $\varphi \in \text{PREP}(G^c)$. To this end, consider an arbitrary outcome $\vec{v}$ with $\vec{v} \models \varphi$, an arbitrary player i, and an arbitrary choice $v_i' \in V_i$ such that v_i' is a best response of i to $\vec{v}$. If $(\vec{v}_{-i}, v_i') \not\models \gamma_i$, it follows that $(\vec{v}_{-i}, v_i'') \not\models \gamma_i$ for all $v_i'' \in \mathcal{V}_i$ and, with condition (i), that $(\vec{v}_{-i}, v_i') \models \varphi$. On the other hand, if $(\vec{v}_{-i}, v_i') \models \gamma_i$, condition (ii) yields $(\vec{v}_{-i}, v_i') \models \varphi$. In either case, we find that $(\vec{v}_{-i}, v_i') \in \varphi^{G^c}$. Hence, $\varphi \in \text{PREP}(G^c)$, as desired. ■

Condition (i) says that if a player cannot satisfy his goal from some outcome, then a hard formula should be consistent with all his actions. Condition (ii) says that the formula must comprise *all* players' deviations to states satisfying his goal. Let us explain the result further. Consider a formula that is hard and an outcome satisfying it. If this outcome does not satisfy a goal of a player, and there was an action of his that would allow a deviation to a valuation not satisfying the formula while leaving the actions of the other players unchanged, then it would be easy to find a cost function making such deviation profitable for the player. This is the reason why we need condition (i). If instead the outcome does satisfy the goal of a player, all outcomes which do should also satisfy the formula. If some outcome did satisfy the goal but not the formula, then we would again be able to make this outcome the most profitable deviation. This is why we need condition (ii).

Propositions 9 and 10 provide *purely logical* characterisations: the characterising conditions are expressed in terms of valuations and goal formulas but cost functions are not referred to. Recall that the soft, absent, and present formulas are defined on the basis of the sets of initial and hard formulas using the customary set-theoretic constructs. Accordingly, similarly logical characterisations of soft, absent, and present formulas can straightforwardly be obtained by combining the characterisations provided by Propositions 9 and 10.

Characterising Hard and Soft Equilibria. We have already seen that Nash equilibria are prep sets. Accordingly, Propositions 9 and 10 can also be used to characterise initial and hard equilibria. The results in this section follow as corollaries, but we find that the characterisations are simpler and also provide more conceptual insight into the nature of hard equilibria. First, we give the result for initial equilibria.

COROLLARY 11. *Let $\vec{v}$ be an outcome of a game G. Then, $\vec{v} \in \textsc{init}(G)$ if and only if for all players $i \in N$ and all $v_i' \in V_i$,*

$$v_i' \neq v_i \text{ and } (\vec{v}_{-i}, v_i') \models \gamma_i \text{ together imply } \vec{v} \models \gamma_i.$$

PROOF. By virtue of Lemma 3 and Proposition 9, it suffices to show that, if φ^G is a singleton, for all $\vec{v} \in \varphi^G$, all players i, and all $v_i \in V_i$, the following two conditions are equivalent:

$(A1)$ $v_i' \neq v_i$ and $(\vec{v}_{-i}, v_i') \models \gamma_i$ together imply $\vec{v} \models \gamma_i$,

$(A2)$ $(\vec{v}_{-i}, v_i') \models \gamma_i \wedge \neg\varphi$ implies $(\vec{v}_{-i}, v_i'') \models \gamma_i \wedge \varphi$ for some $v_i'' \in V_i$.

Observe that $\varphi = \{\vec{v}\}$. First assume $(A1)$ as well as that $(\vec{v}_{-i}, v_i') \models \gamma_i \wedge \neg\varphi$. Then, $(\vec{v}_{-i}, v_i') \notin \varphi^G$. Therefore, $v_i' \neq v_i$. Moreover, $(\vec{v}_{-i}, v_i') \models \gamma_i$. By $(A1)$, then $\vec{v} \models \gamma_i$. Hence, there is some $v_i'' \in V_i$ with $(\vec{v}_{-i}, v_i'') \models \gamma_i \wedge \varphi$.

For the opposite direction, assume $(A2)$ as well as that $v_i' \neq v_i$ and $(\vec{v}_{-i}, v_i') \models \gamma_i$. Then, $(\vec{v}_{-i}, v_i') \models \gamma_i \wedge \neg\varphi$. By $(A2)$, also $(\vec{v}_{-i}, v_i'') \models \gamma_i \wedge \varphi$ for some $v_i'' \in V_i$. As $(\vec{v}_{-i}, v_i'') \models \varphi$, we have that $v_i'' = v_i$. Therefore, also $\vec{v} \models \gamma_i$, as desired. ■

Also for hard equilibria we find a neat logical characterisation.

COROLLARY 12. *Let $\vec{v} = (v_1, \ldots, v_i, \ldots, v_n)$ be an outcome of a game G. Then, $\vec{v} \in \textsc{hard}(G)$ if and only if for all players i and all choices $v_i' \in V_i$ with $v_i' \neq v_i$, both $\vec{v} \models \gamma_i$ and $(\vec{v}_{-i}, v_i') \not\models \gamma_i$.*

PROOF. Consider an arbitrary outcome $\vec{v}$ and assume that $\varphi^G = \{\vec{v}\}$. By virtue of Lemma 3 and Proposition 10, it suffices to show that conditions

(i) $(\vec{v}_{-i}, v_i'') \not\models \gamma_i$ for all $v_i'' \in \mathcal{V}_i$ implies $(\vec{v}_{-i}, v_i') \models \varphi$, and

(ii) $(\vec{v}_{-i}, v_i') \models \gamma_i$ implies $(\vec{v}_{-i}, v_i') \models \varphi$

are satisfied for all players i and all $v_i' \in V_i$ if and only if for all players i, all choices v_i' with $v_i' \neq v_i$ both $\vec{v} \models \gamma_i$ and $(\vec{v}_{-i}, v_i') \not\models \gamma_i$ hold.

To this end, first assume that for all players i, all choices v_i' with $v_i' \neq v_i$ both $\vec{v} \models \gamma_i$ and $(\vec{v}_{-i}, v_i') \not\models \gamma_i$. Consider an arbitrary player i along with an arbitrary choice $v_i' \in V_i$ and assume, for contraposition, that $(\vec{v}_{-i}, v_i') \not\models \varphi$. Then, $v_i' \neq v_i$ and, therefore, $(\vec{v}_{-i}, v_i') \not\models \gamma_i$. This yields (ii). Moreover,

$\vec{v} \models \gamma_i$. Hence, $(\vec{v}_{-i}, v_i'') \models \gamma_i$ for some $v_i'' \in V_i$. It follows that (i) is satisfied as well.

For the opposite direction, assume that for all players i and all $v_i' \in V_i$ both (i) and (ii) hold. Let i be an arbitrary player and v_i' an arbitrary choice for i such that $v_i' \neq v_i$. As $\varphi^G = \{\vec{v}\}$, it immediately follows that $(\vec{v}_{-i}, v_i') \not\models \varphi$. By virtue of the contrapositive of (ii), we may conclude that $(\vec{v}_{-i}, v_i') \not\models \gamma_i$. Moreover, the contrapositive of (i) yields $(\vec{v}_{-i}, v_i'') \models \gamma_i$ for some $v_i'' \in V_i$. By the above $(\vec{v}_{-i}, v_i') \not\models \gamma_i$ for all v_i' with $v_i' \neq v_i$. Hence $v_i'' = v_i$ and we may conclude that $\vec{v} \models \gamma_i$. ■

The significance of Corollary 12 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 Corollary 12 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. Thus, hard equilibria can be understood as maximising *qualitative social welfare* [25].

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 [23]. 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 [16, 17]: 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 [16]. Economic theorists like Coase, Meade, and Maskin have studied under what conditions this possibility allows efficient outcomes to be reached [8,13,16,17]. 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.

The main question we focus on is the extent to which prep sets, characterised by propositional formulas, can be eliminated from games. That is, suppose we have a game G and formula φ such that $\varphi \in \textsc{prep}(G)$. Then, is there a mechanism (either side payments or merging, for example, as discussed above) which results in a game G' such that $\varphi \notin \textsc{prep}(G')$? It is important to note that the problem of eliminating a certain formula φ from a game in this sense reduces to the possibility of eliminating one of the outcomes satisfying it, by means of an outcome not satisfying it. That is, for φ to fail to be a prep formula, we only need to introduce a deviation to an outcome outside the set of outcomes denoted by φ.

Before proceeding, we note that our notion of an externality is related to the notion of *dependence* in Boolean games as studied by Bonzon et al. [5] and Sauro and Villata [21]. Intuitively, a player i is dependent on a player j if there is some situation in which j can act in such a way as to obtain a better or worse outcome for i. Formally, a player i is said to be dependent on a player j if there are two valuations v_1 and v_2 differing only in the values assigned to the variables controlled by j, such that either i strictly prefers v_1 over v_2, or i strictly prefers v_2 over v_1.

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 [22] 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 (sets of) outcomes. The question we are especially interested in is which formulas—hard or soft—can be eliminated from the game in this manner. Consider, for example, the game in Figure 1. There, player 2 does not have her goal achieved in the equilibrium outcome pqr, but she could incentivise player 1 to set p to false by offering him compensation for the additional costs he incurs if he were to do so.

Following [13, 22], 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. Accordingly, $\tau(i, j, \vec{v}) > \tau(j, i, \vec{v})$ whenever the compensation i receives from j is smaller than the compensation j receives from i.

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, that is, 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.

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 $\vec{v}$,[5]

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

[5] 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.

The utility function of player i in game G with cost function c^τ we will henceforth denote by u_i^τ. To avoid cluttered notation, we also denote the game G^{c^τ} by G^τ and the set of best responses of i to $\vec{v}$ in G^τ by $BR_i^\tau(\vec{v})$.

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 for every outcome $\vec{v}$ that $\sum_{i\in C} \tau_i(\vec{v}) = 0$ and, hence,

$$\sum_{i\in C} c(\vec{v}) = \sum_{i\in C} c^\tau(\vec{v}).$$

We restrict our attention to **admissible** transfer functions, i.e., transfer functions such that $\tau_i(\vec{v}) \leq c_i(\vec{v})$ for all players i and all outcomes $\vec{v}$. In words, in no outcome can the amount that a player receives from others minus what he gives to them 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 of a purely technical nature and preserves the quasi-dichotomous character of the preferences in games transformed by transfer functions.[6] For similar purely technical reasons, we will restrict our attention to games with **positive cost functions**, i.e., cost functions that assigns at all outcomes a strictly positive cost to every player.

We are particularly interested in formulas that can be eliminated by groups of players making side-payments to one another and also have a mutual interest in doing so. First, however, we make the following observation.

PROPOSITION 13. *Let G be a game with at least two players and with a global and positive cost function c and φ a satisfiable product formula. Then, $\varphi \in \text{PREP}(G^\tau)$ for all transfer functions τ if and only if $\varphi \in \text{HARD}(G)$.*

PROOF. First observe that c^τ is a properly defined cost function for every transfer function τ. The "if"-direction then follows immediately from the definition of hard equilibria. For the opposite direction assume that $\varphi \notin \text{HARD}(G)$. Then, by Proposition 10, there is some outcome $\vec{v}$, some player i and some choice v'_i such that either $(i')\ (\vec{v}_{-i}, v''_i) \not\models \gamma_i$ for all $v''_i \in \mathcal{V}_i$ and

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

$(\vec{v}_{-i}, v_i') \not\models \varphi$, or (ii') both $(\vec{v}_{-i}, v_i') \models \gamma_i$ and $(\vec{v}_{-i}, v_i') \not\models \varphi$. Let j be a player distinct from i and define the transfer function τ such that for all players k, k' and all outcomes $\vec{w}$,

$$\tau(k, k', \vec{w}) = \begin{cases} c_i(\vec{v}_{-i}, v_i') & \text{if } k = j,\ k' = i, \text{ and } \vec{w} = (\vec{v}_{-i}, v_i'), \\ 0 & \text{otherwise.} \end{cases}$$

Having assumed positive cost functions, it follows that v_i' is the unique best response of i to $\vec{v}$ in G^τ if (i') as well as if (ii'). Hence, $\varphi \notin \text{PREP}(G^\tau)$, as desired. ∎

To illustrate this result, consider once more the game depicted in Figure 2. We saw $p \wedge s$ is a prep formula but not a hard one. By virtue of Proposition 13, there are transfer functions that render $p \wedge s$ non-prep. Still, none of these appear to be particularly attractive for the coalition involved. The transfer function τ with $\tau(2, 1, 4\frac{1}{2})$, for instance, would make $\bar{p}q$ player 1's unique best response to $p\bar{q}r$, but if 1 were to deviate in this way, player 2 would be worse off after transfers than she was before. Observe that the situation would not improve if player 3 were to join the coalition. It can readily be appreciated that there are similar concerns with every other transfer function and no coalition can come to a mutually profitable understanding on how to divide costs so as to eliminate $p \wedge s$. In the remainder of this section we will therefore restrict our attention to the elimination of prep formulas via transfer functions that are beneficial to all members of the coalition involved.

Thus, intuitively, a formula φ can be eliminated via side-payments if there exists some outcome $\vec{v}$ satisfying φ along with some coalition that can engineer a mutually beneficial transfer scheme that induces one of its members i to only play actions leading to outcomes not satisfying φ by compensating i for any additional costs if i does so. There are different ways of formally defining this. We adopt the following concept, which compares the set X of outcomes $(\vec{v}_{-i}, v_i'')$ satisfying φ that can be reached by i playing a best response to $\vec{v}$ in the original game with the set Y outcomes that are reached by i playing a best response after having been compensated: all coalition members should prefer all outcomes in Y after transfer to any outcome in X before.

Formally, a formula φ can be **eliminated via side-payments**, if there is a coalition $C \subseteq N$, a transfer function τ only involving C, an outcome $\vec{v}$, a player $i \in C$, such that for all $v' \in BR_i^\tau(\vec{v})$ and for all $v_i'' \in BR_i(\vec{v})$ with $(\vec{v}_{-i}, v_i'') \in \varphi^G$:

(C1) $(\vec{v}_{-i}, v_i') \notin \varphi^G$, and

(C2) $u_j^\tau(\vec{v}_{-i}, v_i') > u_j(\vec{v}_{-i}, v_i'')$ for all $j \in C$.

We also say that under the circumstances specified thus, coalition C **blocks** formula φ and, when φ^G consists of a single outcome $\vec{v}$, also that C blocks $\vec{v}$. Condition (C1) ensures that all of i's best responses *after transfers* have taken place lead to outcomes outside φ^G, while condition (ii) guarantees that all players in C are better off in any of the outcomes that may arise if i best-responses to $\vec{v}$ *after* transfers have taken place than they were in any of the outcomes that ensues if i best-responses to $\vec{v}$ *before*. Intuitively, this concept incorporates the idea of a blocking coalition that engineers a transfer scheme such that the resulting set of best responses of one of its members becomes more attractive to all than the original one.

We illustrate these definitions by means of two examples. The first one concentrates on elimination of equilibria by side-payments, whereas the second also deals with the more general case of eliminating formulas.

Example 14. Consider once more the game in Figure 1. At outcome pqr, which, recall, is an equilibrium, player 2's goal is not satisfied, whereas it is at $\bar{p}qr$. Let τ be such that

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

Then, τ would incentivise player 1 to deviate to $\bar{p}qr$ provided that $x > 1$. Moreover, player 2 would prefer to make any such transfer in order to satisfy her goal. Accordingly, $\{1, 2\}$ is a coalition blocking the outcome pqr in set $\{pqr\}$. In a similar way, player 1 might want to induce player 3 to deviate to outcome $pq\bar{r}$. That, however, would require compensating player 3 for the additional costs of 3 that player 3 incurs at $pq\bar{r}$. Player 1, having his goals achieved at both pqr and $pq\bar{r}$, 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 $pq\bar{r}$. 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

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

Accordingly, $\{1, 2, 3\}$ is also a coalition blocking outcome pqr. We may therefore conclude that pqr is not eliminable by side-payments.

Example 15. Consider the game in Figure 2. We have already argued that the soft prep formula $p \wedge s$ is not eliminable by side-payments. Now consider the formula $\neg(p \vee s)$, which is also prep. There are various ways in which $\neg(p \vee s)$ can be eliminated by side-payments. For instance, at $\bar{p}qr\bar{s}$ player 3 plays a best response and, in order to have his goal achieved, 1 would be prepared to compensate 3 with any amount required. Similarly, player 3 plays a best response at $\bar{p}q\bar{r}\bar{s}$ and 1 and 2 can induce him to set s to true by compensating him each with $\frac{3}{4}$ at $\bar{p}q\bar{r}s$.

Finally, consider the outcomes $\bar{p}q\bar{r}\bar{s}$ and $\bar{p}\bar{q}\bar{r}\bar{s}$. At both of these outcomes player 1 plays a best response and a blocking coalition engineering a new best response for i will have to make sure that all its members will be better off if 1 plays this new best response than if they were to stay at either $\bar{p}q\bar{r}\bar{s}$ or $\bar{p}\bar{q}\bar{r}\bar{s}$. Thus, although 2 could induce 1 to move from $\bar{p}q\bar{r}\bar{s}$ to $p\bar{q}\bar{r}\bar{s}$ by compensating him with, say $1\frac{1}{2}$ at the latter outcome, this would not make her better off with respect to $\bar{p}\bar{q}\bar{r}\bar{s}$. All three players together, however, form a blocking coalition against $\neg(p \vee s)$ as witnessed by the transfer function τ'' defined as

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

We now move towards a characterisation of formulas that can be eliminated by side-payments. Before doing so, however, we first make the following observation. If a formula φ fails to be prep, there is some outcome v, player i and $v'_i \in V_i$ such that $v \models \varphi$, $(\vec{v}_{-i}, v'_i) \models \neg\varphi$, $u_i(\vec{v}_{-i}, v'_i) > u_i(\vec{v})$ and $u_i^\tau(\vec{v}_{-i}, v'_i) \geq u_i^\tau(\vec{v}_{-i}, v''_i)$ for all $v''_i \in \mathcal{V}_i$. Then φ is eliminable by the singleton coalition $\{i\}$ via the trivial transfer function for which all transfers between all players at all outcomes are zero.

Observation 16. *Let φ be a product formula that is not prep in a game G. Then, φ is eliminable via side-payments.*

Thus, there is a class of formulas that can never be eliminated via side-payments—the hard formulas—as well as a class of outcomes that can always be eliminated via side-payments—the formulas that are not prep. There may, however, very well be formulas in a game that do not belong to either of these classes: the present formulas. Together with Proposition 13 and Observation 16, the following result establishes a full characterisation of all product formulas that are eliminable via side-payments in a game. Notice that this characterisation *makes no reference to transfer functions.*

Proposition 17. *Let φ a present prep formula in Boolean game G with a positive cost function c. Then φ is eliminable via side-payments if and only if there is a coalition $C \subseteq N$, an outcome $\vec{v} \in \vec{V}$, a player $i \in N$, and a choice $v'_i \in V_i$ such that,*

(i) $(\vec{v}_{-i}, v'_i) \not\models \varphi$,

(ii) *for all* $j \in C$ *and* $v''_i \in BR_i(\vec{v})$ *with* $(\vec{v}_{-i}, v''_i) \models \varphi$, *if* $(\vec{v}_{-i}, v''_i) \models \gamma_j$ *then* $(\vec{v}_{-i}, v'_i) \models \gamma_j$,

(iii) $\displaystyle\sum_{j\in C} c_j(\vec{v}_{-i}, v'_i) < \sum_{j\in C} \inf_{\vec{w}\in X_j} c_j(\vec{w})$,

where, for every $j \in C$,

$$X_j = \{(\vec{v}_{-i}, w_i) \in \varphi^G : w_i \in BR_i(\vec{v}) \text{ and if } (\vec{v}_{-i}, v'_i) \models \gamma_j \text{ then } (\vec{v}_{-i}, w_i) \models \gamma_j\}.$$

Proof. First assume that φ^G is eliminable via side-payments. Then, there is a coalition $C \subseteq N$, a transfer function τ only involving C, an outcome $\vec{v}$, a player $i \in C$, such that for all $v' \in BR_i^\tau(\vec{v}_{-i})$ and for all $v''_i \in BR_i(\vec{v}_{-i})$ with $(\vec{v}_{-i}, v'_i) \in \varphi^G$ both (C1) and (C2) hold, that is, $(\vec{v}_{-i}, v'_i) \notin \varphi^G$, and $u_j^\tau(\vec{v}_{-i}, v'_i) > u_j(\vec{v}_{-i}, v''_i)$ for all $j \in C$. Let $v'_i \in BR_i^\tau(\vec{v})$. We may assume that v'_i exists. Through (C1) we immediately obtain (i), that is, $(\vec{v}_{-i}, v'_i) \notin \varphi^G$. Consider an arbitrary $j \in C$ and equally arbitrary $v''_i \in BR_i(\vec{v}_{-i})$ with $(\vec{v}_{-i}, v''_i) \models \varphi$. By virtue of (C2), it moreover follows that $u_j^\tau(\vec{v}_{-i}, v'_i) > u_j(\vec{v}_{-i}, v''_i)$. Therefore, $(\vec{v}_{-i}, v''_i) \models \gamma_j$ implies $(\vec{v}_{-i}, v'_i) \models \gamma_j$, and (ii) follows. For (iii), consider an arbitrary $(\vec{v}_{-i}, w_i) \in X_j$. Then, $(\vec{v}_{-i}, w_i) \models \gamma_j$ if and only if $(\vec{v}_{-i}, v'_i) \models \gamma_j$. Again by (C2), we have $c_j^\tau(\vec{v}_{-i}, v'_i) < c_j(\vec{v}_{-i}, w_i)$. Having chosen $(\vec{v}_{-i}, w_i)$ arbitrarily, furthermore, $c_j^\tau(\vec{v}_{-i}, v'_i) < \inf_{\vec{w}\in X_j} c_j(\vec{w})$. Hence,

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

As τ only involves C, also $\sum_{j\in C} c_j^\tau(\vec{v}_{-i}, v'_i) = \sum_{j\in C} c_j(\vec{v}_{-i}, v'_i)$. Therefore,

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

For the opposite direction, assume that there is a coalition $C \subseteq N$, an outcome $\vec{v}$, a player i, and a choice v'_i for which (i), (ii), and (iii) hold. For ease of notation, let $c_k^* = \inf_{\vec{w}\in X_k} c_k(\vec{w})$ for every $k \in C$. Notice that without loss of generality we may assume that C maximises the difference $\sum_{j\in C} \inf_{\vec{w}\in X_j} c_j(\vec{w}) - \sum_{j\in C} c_j(\vec{v}_{-i}, v'_i)$, that is, $c_j(\vec{v}, v'_i) < c_j^*$ for all $j \in C \setminus \{i\}$.[7] Consider an arbitrary $\vec{w} \in X_i$. By (ii), then $(\vec{v}_{-i}, v'_i) \models \gamma_i$ if and only if $\vec{w} \models \gamma_i$. Therefore, $0 < c_i(\vec{w}) \leq c_i(\vec{v}_{-i}, v'_i)$. Moreover, with $\vec{w}$

[7] Also recall that $\inf(\emptyset) = +\infty$. This covers the case in which there is some $j \in C$ with $(\vec{v}_{-i}, v'_i) \models \gamma_j$ but $(\vec{v}_{-i}, v''_i) \not\models \gamma_j$ for all $v''_i \in BR_i(\vec{v})$ with $(\vec{v}_{-i}, v''_i) \models \varphi$.

having been chosen arbitrarily, also $0 < c_i^* \leq c_i(\vec{v}_{-i}, v_i')$. Thus, i needs to be compensated with at least $c_i(\vec{v}_{-i}, v_i') - c_i^*$ at $(\vec{v}_{-i}, v_i')$ for v_i' to become a best response of i to $\vec{v}$, and slightly more to become the only one. Notice that (iii) can be rewritten as

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

Therefore, there is some real $r \in \mathbb{R}$ with $0 < r \leq c_i^*$ such that

$$c_i(\vec{v}_{-i}, v_i') - c_i^* + r < \sum_{j \in C \setminus \{i\}} c_j^* - \sum_{j \in C \setminus \{i\}} c_j(\vec{v}_{-i}, v_i').$$

Now, define the transfer function τ such that, for every $j \in C \setminus \{i\}$,

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

and $\tau(k, k', \vec{w}) = 0$ if either $k \notin C \setminus \{i\}$, $k \neq i$, or $\vec{w} \neq (\vec{v}_{-i}, v_i')$. As $r \leq c_i^*$, we find that τ is admissible. Moreover, $c_i^\tau(\vec{v}_{-i}, v_i') = c_i^* - r$. As $u_i^\tau(\vec{w}) = u_i(\vec{w})$ for all $\vec{w} \neq (\vec{v}_{-i}, v_i')$, it follows that $BR_i^\tau(\vec{v}) = \{v_i'\}$. Moreover, for all $j \in C$ we have $c_j^\tau(\vec{v}_{-i}, v_i') < c_j^*$. By (ii), it follows that $u_j^\tau(\vec{v}_{-i}, v_i') > u_j(\vec{v}_{-i}, v_i'')$ for all $v'' \in BR_i(\vec{v})$ with $(\vec{v}_{-i}, v_i'') \in \varphi^G$ and $j \in C$. As by (ii), finally, $(\vec{v}_{-i}, v_i') \notin \varphi^G$, we may conclude that C blocks φ^G and, therefore, that φ is eliminable by side-payments, as desired. ■

Sufficient and necessary conditions for a present *equilibrium* to be eliminable via side-payments follow of course immediately from Proposition 17, but we find that they can be formulated much more concisely.

Corollary 18. *Let $\vec{v}$ be a present equilibrium of a Boolean game G with a positive* global *cost function c. Then, $\vec{v}$ is eliminable via side-payments if and only if there is a coalition C, a player $i \in C$, and a choice $v_i' \in V_i$ such that the following two conditions hold:*

(I) $\vec{v} \models \gamma_j$ *implies* $(\vec{v}_{-i}, v_i') \models \gamma_j$ *for all* $j \in C$,

(II) *if* $(\vec{v}_{-i}, v_i') \models \gamma_j$ *implies* $\vec{v} \models \gamma_j$ *for all* $j \in C$, *then* $\sum_{j \in C} c_j(\vec{v}_{-i}, v_i') < \sum_{j \in C} c_j(\vec{v})$.

Proof. Let $\varphi^G = \{\vec{v}\}$. Let furthermore C be a coalition, i be a player in C, and $v_i' \in \mathcal{V}_i$. We show that the conditions (I) and (II) are equivalent to the conditions (i), (ii), and (iii) in Proposition 17.

First assume that (I) and (II) hold. For contradiction, also assume that $(\vec{v}_{-i}, v_i') \models \varphi$. Then, $v_i' = v_i$. It follows that the antecedent of (II) is satisfied and its consequent then gives the contradiction. Thus (i) is satisfied. Observe furthermore that we may assume that φ is a present formula and hence that the conditions on v_i'' in the antecedent of (ii) imply that $v_i'' = v_i$. Thus, (ii) follows immediately from (I). Finally, observe that under our assumptions, either $X_j = \emptyset$ or $X_j = \{\vec{v}\}$ for all $j \in C$. If the former holds for some $j \in C$, it follows that $\inf_{\vec{w} \in X_j} c_j(\vec{w}) = +\infty$, which immediately yields (iii). Otherwise, for all $j \in C$ we have that $(\vec{v}_{-i}, v_i') \models \gamma_j$ implies $\vec{v} \models \gamma_j$ as well as that $\inf_{\vec{w} \in X_j} c_j(\vec{w}) = c_j(\vec{v})$. Now (iii) follows from (II).

For the opposite direction, assume (i), (ii), and (iii) hold. As we may assume that φ is a present formula, both $v_i \in BR_i(\vec{v})$ and $\vec{v} \models \varphi$. Thus, (I) follows from (ii). Finally, assume that $(\vec{v}_{-i}, v_i') \models \gamma_j$ implies $\vec{v} \models \gamma_j$ for all $j \in C$. Then, $X_j = \{\vec{v}\}$ for all $j \in C$, and (II) follows from (iii). ■

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

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

r:

	q	$\bar{q}$
p	$1,3$ $4,3,3$	3 $1,5,2$
$\bar{p}$	– $2,3,1$	$1,2,3$ $3,2,2$

$\bar{r}$:

	q	$\bar{q}$
p	2 $1,1,3$	$2,3$ $3,1,2$
$\bar{p}$	$1,2,3$ $1,3,1$	– $1,1,1$

Figure 4. A three-player game illustrating coalition merging (notational conventions as in Figure 1)

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

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

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. It is worth observing that if φ is a product formula in a game G, then it remains a product formula in a merged game G_C. The implication in the opposite direction, however, does not generally hold.

Some features of coalition merging are presented most clearly by focussing on the elimination of equilibria, as in the following example.

Example 19. Consider the game depicted in Figure 4, where player 1 is able to choose values for p, player 2 for q, and player 3 for r. The original game has three Nash equilibria, viz., the outcomes pqr, $\bar{p}\bar{q}r$, and $\bar{p}q\bar{r}$, the latter two of which are hard. By merging all players into one, only $\bar{p}q\bar{r}$ remains as an equilibrium. It satisfies all players' goals and, hence also $d(N)$'s. Although outcome $\bar{p}\bar{q}r$ 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 $\bar{p}\bar{q}r$ and $pq\bar{r}$ are soft equilibria in G_N.

Now, consider the (soft) equilibrium 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 $pq\bar{r}$ nor would he be willing to compensate player 2 sufficiently if she were to

	qr	$q\bar{r}$	$\bar{q}r$	$\bar{q}\bar{r}$
	1	$-$	$-$	$d(C)$
p	$4,6$	$1,4$	$1,7$	$3,3$
	$-$	$1,d(C)$	$1,d(C)$	$-$
$\bar{p}$	$2,4$	$1,4$	$3,4$	$1,2$

Figure 5. The reduced game resulting from the game in Figure 4 by merging players 2 and 3 into one player $d(C)$ (notational conventions as in Figure 1)

deviate to $p\bar{q}r$. Observe, however, that, if player 2 and 3 were to merge, as depicted in Figure 5, they could deviate to $p\bar{q}\bar{r}$, benefitting both players.

Example 19 illustrates several interesting properties of coalition merging. First and foremost, it shows that it can lead to hard formulas of the original game being removed as prep sets from the reduced game.

Observation 20. *Coalition merging does not preserve hard formulas, i.e., there are games G, product formulas φ, and coalitions C such that $\varphi \in$* HARD(G) *and $\varphi \notin$* PREP(G_C).

In the extreme case in which *all* players are merged into one coalition, 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 result, which is an immediate consequence of Corollaries 12 and 11 and the observation that in G_N there is only one player, $d(N) = N$, that $V = \vec{V} = V_{d(N)}$, and that $(\vec{v}_{-d(N)}, v'_{d(N)}) = v'_{d(N)}$.

Corollary 21. *Let G be a game and $\vec{v}$ an outcome. Then, $\vec{v}$ is an initial equilibrium in G_N if and only if $\bigwedge_{i\in N}\gamma_i$ being satisfiable implies $\vec{v} \models \bigwedge_{i\in N}\gamma_i$. Moreover, $\vec{v}$ is a hard equilibrium in G_N if and only if $\vec{v}$ is the* unique *outcome in G such that $\vec{v} \models \gamma_i$ for all $i \in N$.*

Proof. The result follows immediately by translating the characterising conditions for initial and hard equilibria—see Corollary 12 and Corollary 11, respectively—to the special case of G_N. ■

For general formulas, the situation is only slightly more complicated and the logical concepts of a tautology and a formula being satisfiable can be employed to obtain suitable characterisations.

Proposition 22. *Let G be a game and φ a satisfiable product formula. Then,*

(i) $\varphi \in \textsc{init}(G_N)$ *if and only if* $\bigwedge_{i \in N} \gamma_i \wedge \neg\varphi$ *being satisfiable implies that* $\bigwedge_{i \in N} \gamma_i \wedge \varphi$ *is satisfiable as well.*

(ii) $\varphi \in \textsc{hard}(G_N)$ *if and only if* $\bigwedge_{i \in N} \gamma_i \rightarrow \varphi$ *is a tautology and either* $\bigwedge_{i \in N} \gamma_i$ *is satisfiable or* φ *is a tautology.*

Proof. Both (i) and (ii) follow almost immediately from Proposition 9 and 10. For G_N the characterising condition for initial equilibria in Proposition 9 reduces to:

$$v' \models \gamma_{d(N)} \wedge \neg\varphi \text{ for some } v \in V \text{ implies } v'' \models \gamma_{d(N)} \wedge \varphi \text{ for some } v'' \in V.$$

Recall that $\gamma_{d(N)} = \bigwedge_{i \in N} \gamma_i$. This yields (i).

For (ii), observe that in G_N the characterisation of hard formula in Proposition 10 reduces to that for all $v' \in V$ both

(i') if $v''_i \not\models \gamma_{d(N)}$ for all $v'' \in V$, then $v' \models \varphi$, and (ii') if $v'_i \models \gamma_i$, then $v'_i \models \varphi$. This is equivalent to the conjunction of (i'') if $v''_i \models \gamma_{d(N)}$ for some $v'' \in V$ or $v' \models \varphi$ for all $v' \in V$, and (ii'') for all $v' \in V$, if $v'_i \models \gamma_i$, then $v'_i \models \varphi$. Recall that $\gamma_{d(N)} = \bigwedge_{i \in N} \gamma_i$ and the result straightforwardly follows. ■

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 state the following result.

Corollary 23. *Let G be a game in which the players' goals are compatible and φ a satisfiable product formula. Then,*

(i) $\varphi \in \textsc{init}(G_N)$ *if and only if* $\bigwedge_{i \in N} \gamma_i \wedge \varphi$ *is satisfiable,*

(ii) $\varphi \in \textsc{hard}(G_N)$ *if and only if* $\bigwedge_{i \in N} \gamma_i \rightarrow \varphi$ *is a tautology.*

Proof. The result is obtained immediately from Proposition 22 through some elementary truth-functional reasoning about satisfiability in propositional logic. ■

We furthermore find that the equilibria and prep sets of G_N have some desirable properties if the players goals are compatible.

Proposition 24. *Let G be a game in which the players' goals are compatible and $\varphi \in \textsc{prep}(G_N)$. Then, there is some outcome $\vec{v} \in \varphi^G$ with $\vec{v} \models \bigwedge_{i \in N} \gamma_i$. Consequently, $\vec{v} \models \bigwedge_{i \in N} \gamma_i$ for every $\vec{v} \in \textsc{ne}(G_N)$.*

Proof. We may assume that there is some valuation v with $v \models \bigwedge_{i \in N} \gamma_i$. Moreover, $v = v_{d(N)} \in V_{d(N)}$. As prep sets are non-empty, we may also assume that there is some outcome $\vec{v}'$ with $\vec{v}' \models \varphi$. As $\varphi \in \textsc{prep}(G_N)$, there is some $v''_{d(N)} \in BR_{d(N)}(\vec{v}')$ such that $(\vec{v}'_{-d(N)}, v''_{d(N)}) \models \varphi$. Then, $u_{d(C)}(\vec{v}'_{-d(N)}, v''_{d(N)}) \geq u_{d(N)}(\vec{v}'_{-d(N)}, v_{d(N)})$. Observe that $(\vec{v}'_{-d(N)}, v_{d(N)}) = v_{d(N)}$. Hence, $(\vec{v}'_{-d(N)}, v''_{d(N)}) \models \bigwedge_{i \in N} \gamma_i$ as well (otherwise $v''_{d(N)}$ would not be a best response). Recall that $(\vec{v}'_{-d(N)}, v''_{d(N)}) \models \varphi$ and the result follows.

Observe by the above that if $\varphi^G = \{\vec{v}\}$, also $\vec{v} \models \bigwedge_{i \in N} \gamma_i$. The second part then follows immediately from Lemma 3. ■

Finally, we focus on the relation of equilibrium eliminability via coalition merging and via side-payments. Observation 14, together with Proposition 13 immediately shows that coalition merging can eliminate equilibria that cannot be eliminated via side-payments.

Observation 25. *There are games with product formulas that can be eliminated by merging coalitions but not by side-payments.*

Thus, one might suspect that formulas that can be eliminated via side-payments can also be eliminated by merging coalitions. Example 26 shows that this is not the case.

Example 26. Consider again the game in Figure 4, but now assume that in $\bar{p}\bar{q}r$ only player 1 achieves his goal and in $\bar{p}q\bar{r}$ only players 1 and 3 theirs. Then, outcome $\bar{p}\bar{q}\bar{r}$ 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 $\bar{p}\bar{q}\bar{r}$ and contains players who are prepared to compensate fully the costs incurred by some player deviating from $\bar{p}\bar{q}\bar{r}$. However, no matter how you merge coalitions, outcome $\bar{p}\bar{q}\bar{r}$ 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 properties in Boolean Games.

Observation 27. *There are games with product formulas that can be eliminated via side-payments but not by merging coalitions.*

5. Summary

The problems of eliminating undesirable properties of an interaction and facilitating desirable ones are fundamental in economics and multi-agent systems. By focussing on Boolean games with costs, in which players have quasi-dichotomous preferences, Harrenstein et al. [11] were able to distinguish between hard and soft equilibria in games, i.e., outcomes that are Nash equilibria irrespective of the cost function and those that may or may not be Nash equilibria, depending on the cost function. We extended this work to pure prep sets, a set-valued solution concept based on the notion of prep set as originally proposed by Voorneveld [24]. A product set of outcomes is prep if it contains for each player at least one best response to every outcome in the set. We thus came to consider hard and soft prep sets, which are defined for sets of outcomes in an analogous way as hard and soft equilibria are for single outcomes. We studied techniques by which undesirable properties 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 though by coalition merging hard formulas may get eliminated, coalition merging is not stronger than side-payments: there may be formulas that can be eliminated via side-payments but not via coalition merging.

Acknowledgements. The authors would like to thank an anonymous reviewer for drawing their attention to Voorneveld's preparation sets and numerous suggestions to improve the paper. Michael Wooldridge and Paul Harrenstein were 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").

References

[1] Aumann, R. J., Game theory, in J. Eatwell, M. Milgate, and P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*, vol. 2, MacMillan, London, 1987, pp. 460–482.

[2] Basu, K., and J. Weibull, Strategy subsets closed under rational behavior, *Economic Letters* 36:141–146, 1991.

[3] Bernheim, B. D., Rationalizable strategic behavior, *Econometrica* 52(4):1007–1028, 1984.

[4] Bonzon, E., M.-C. Lagasquie, J. Lang, and B. Zanuttini, Boolean games revisited, in *Proceedings of the Seventeenth European Conference on Artificial Intelligence (ECAI-2006)*, Riva del Garda, Italy, 2006, pp. 265–269.

[5] Bonzon, E., M.-C. Lagasquie-Schiex, and J. Lang, Dependencies between players in Boolean games, *International Journal of Approximate Reasoning* 50:899–914, 2009.

[6] Bonzon, E., M.-C. Lagasquie-Schiex, J. Lang, and B. Zanuttini, Compact preference representation and Boolean games, *Autonomous Agents and Multi-Agent Systems* 18(1):1–35, 2009.

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

[8] Coase, R. H., The problem of social cost, *Journal of Law and Economics* 3:1–44, 1960.

[9] Dufwenberg, M., H. Norde, H. Reijnierse, and S. Tijs, The consistency principle for set-valued solutions and a new direction for normative game theory, *Mathematical Methods of Operations Research* 54(1):119–131, 2001.

[10] Dunne, P. E., S. Kraus, W. van der Hoek, and M. Wooldridge, Cooperative Boolean games, in *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2008)*, Estoril, Portugal, 2008, pp. 1015–1022.

[11] Harrenstein, P., P. Turrini, and M. Wooldridge, Hard and soft equilibria in Boolean games, in A. Lomuscio, P. Scerri, A. Bazzan, and M. Huhns (eds.), *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2014)*, Paris, France, 2014, pp. 845–852.

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

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

[14] Levit, V., T. Grinshpoun, and A. Meisels, Boolean games for charging electric vehicles, in *Proceedings of 2013 IEEE/WIC/ACM International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT)*, 2013.

[15] Levit, V., T. Grinshpoun, A. Meisels, and A. Bazzan, Taxation search in Boolean games, in *Proceedings of 12th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2013)*, St Paul, MN, 2013.

[16] Maskin, E. S., The invisible hand and externalities, *American Economic Review* 84(2):333–337, 1994.

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

[18] Norde, H., J. Potters, H. Reijnierse, and D. Vermeulen, Equilibrium selection and consistency, *Games and Economic Behavior* 12(2):219–225, 1996.

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

[20] Pearce, D. G., Rationalizable strategic behavior and the problem of perfection, *Econometrica* 52(4):1029–1050, 1984.

[21] Sauro, L., and S. Villata, Dependency in cooperative Boolean games, *Journal of Logic and Computation* 23(2):425–444, 2013.

[22] Turrini, P., Endogenous Boolean games, in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI-13)*, Beijing, China, 2013, pp. 390–396.

[23] Varian, H., *Microeconomic Analysis*, 3rd ed., W. W. Norton, New York, 1992.

[24] Voorneveld, M., Preparation, *Games and Economic Behaviour* 48:403–414, 2004.

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

P. Harrenstein, M. Wooldridge
Department of Computer Science
University of Oxford
Oxford, UK
peh@cs.ox.ac.uk

M. Wooldridge
mjw@cs.ox.ac.uk

P. Turrini
Department of Computing
Imperial College London
London, UK
p.turrini@imperial.ac.uk