Verifiable Equilibria in Boolean Games

Thomas Ågotnes
University of Bergen
Norway
thomas.agotnes@uib.no

Paul Harrenstein
Oxford University
United Kingdom
paulhar@cs.ox.ac.uk

Wiebe van der Hoek
University of Liverpool
United Kingdom
wiebe@liverpool.ac.uk

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

Abstract

This work is motivated by the following concern. Suppose we have a game exhibiting multiple Nash equilibria, with little to distinguish them except that one of them can be *verified* while the others cannot. That is, one of these equilibria carries sufficient information that, if this is the outcome, then *the players can tell that an equilibrium has been played*. This provides an argument for this equilibrium being played, instead of the alternatives. Verifiability can thus serve to make an equilibrium a focal point in the game. We formalise and investigate this concept using a model of Boolean games with incomplete information. We define and investigate three increasingly strong types of verifiable equilibria, characterise the complexity of checking these, and show how checking their existence can be captured in a variant of modal epistemic logic.

1 Introduction

Equilibrium selection problems in game theory arise in situations where a game has multiple equilibria, in which case players must make individual choices so as to coordinate on a particular equilibrium [Binmore, 1992, p. 295]. Such problems are particularly important when the failure to coordinate has negative consequences for players. One of the best-known solutions proposed for the equilibrium selection problem is Schelling's concept of *focal points* [Schelling, 1980]. Schelling's idea was that some equilibria have distinguishing properties that are *independent* of their utility structure. A common example is that of two tourists on a day-trip to Paris, who become separated: where should they meet up? The tourists need to independently choose so as to coordinate on a single location; and with respect to utility, any location in the city is as good as any other. But nevertheless, most people suggest the Eiffel tower as a natural meeting place. This is an example of a focal point: an equilibrium that stands out for players in a game, enabling them to coordinate their actions.

We argue that fact that an equilibrium is *verifiable* can serve to make that equilibrium a focal point. The intuition is as follows. Suppose we have a game with two equilibrium points, A and B, and so players need to coordinate on one of these, and that in terms of utilities, A and B are identical. However, A and B differ with respect to the following property. Equilibrium A *carries sufficient information that, if this equilibrium is played, every player will be able to tell that an equilibrium has been played*, while B is such that, if it is played, one or more players would be unable to tell for sure that an equilibrium had been played. This, we argue, would be a reason for selecting A rather than B. For, if we chose B, then *some players would be unsure whether the actual outcome chosen was indeed an equilibrium*. We say that A is a *verifiable* equilibrium. For a more concrete example, consider the following. Two nations have agreed to eliminate their nuclear weapons, and there are two ways to do this: one of which is verifiable, the other of which is not. Here, the selection of the verifiable course of action is, we believe, more natural for all concerned.

In this paper, we introduce and formalise verifiable equilibria in *Boolean games*, an increasingly popular game theoretic model, with a natural computational interpretation [Harrenstein *et al.*, 2001; Bonzon *et al.*, 2006; Dunne *et al.*, 2008; Endriss *et al.*, 2011]. In a Boolean game, each player i has under its unique control a set of Boolean variables Φ_i, from an overall set of Boolean variables Φ. Player i can assign values to variables Φ_i in any way it chooses: the strategies available to i correspond to the possible Boolean assignments that can be made to variables Φ_i. The outcome of a Boolean game is a valuation for the variables Φ, made up of the choices of individual players. Each player i has a goal that it desires to be achieved, represented as a propositional formula γ_i, which may contain variables under the control of other players. Player i is satisfied with an overall outcome if that outcome satisfies γ_i, and is unsatisfied otherwise. We also assume that each player i is associated with a *visibility set*, $\Theta_i \subseteq \Phi$: player i is able to correctly perceive the values of the variables in Θ_i, but cannot observe the values of any other variables [van der Hoek *et al.*, 2011]. We then say that an outcome $(v_1, \ldots, v_n)$ for a game is a *verifiable equilibrium* if it "looks like" a Nash equilibrium to every player—that is, if $(v_1, \ldots, v_n)$ could (or would have to) be a Nash equilibrium, given i's view of the outcome $(v_1, \ldots, v_n)$ through its visibility set Θ_i. In the remainder of this paper, we formalise several variations of verifiable equilibrium, characterise the complexity of checking for these equilibria, and discuss conditions for their existence. In Section 4, we show how these concepts can

be represented using a *modal epistemic logic* interpreted over our game structures [Fagin *et al.*, 1995].

2 Boolean Games with Incomplete Information

We adapt the basic model of Boolean games [Harrenstein *et al.*, 2001; Bonzon *et al.*, 2006; Dunne *et al.*, 2008; Endriss *et al.*, 2011] to model incomplete information using the idea of visibility sets introduced in [van der Hoek *et al.*, 2011].

Propositional Logic: Let $\{\top, \bot\}$ be the set of Boolean truth values, with "$\top$" being truth and "$\bot$" being falsity; we use $\top$ and $\bot$ to denote both the syntactic constants for truth and falsity respectively, as well as their semantic counterparts. Let $\Phi = \{p, q, \ldots\}$ be a finite, fixed, non-empty vocabulary of Boolean variables, and let $\mathcal{L}$ denote the set of (well-formed) formulae of propositional logic over Φ, constructed using the conventional Boolean operators ("$\wedge$", "$\vee$", "$\rightarrow$", "$\leftrightarrow$", and "$\neg$"), as well as the truth constants "$\top$" and "$\bot$". Where $\varphi \in \mathcal{L}$, we let $vars(\varphi)$ denote the (possibly empty) set of Boolean variables occurring in φ (e.g., $vars(p \wedge q) = \{p, q\}$). A *valuation* is a total function $v : \Phi \rightarrow \{\top, \bot\}$, assigning truth or falsity to every Boolean variable. We write $v \models \varphi$ to mean that the propositional formula 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. We write $\models \varphi$ to mean that φ is a tautology. We denote the fact that $\models \varphi \leftrightarrow \psi$ by $\varphi \equiv \psi$.

Boolean Games: Our games are populated by a set $N = \{1, \ldots, n\}$ of *agents*—the players of the game. Each agent i is assumed to have a *goal*, which is represented by an $\mathcal{L}$-formula γ_i that i desires to have satisfied. Players i each *control* 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 ($\top$ or $\bot$) of each variable $p \in \Phi_i$. We will require that each Boolean variable is controlled by exactly one agent, i.e., $\Phi = (\Phi_1 \cup \cdots \cup \Phi_n)$ and $\Phi_i \cap \Phi_j = \emptyset$ for all $i \neq j$. A *Boolean game* is then defined as a tuple

$$(N, \Phi, \Phi_1, \ldots, \Phi_n, \gamma_1, \ldots, \gamma_n).$$

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 its control so as to satisfy its goal γ_i. The difficulty is that the truth-value of γ_i may depend on variables controlled by other agents $j \neq i$, who will also be trying to choose values for their variables in Φ_j to satisfy their own goals; and their goals in turn may depend on the variables Φ_i.

The set of *choices* available to an agent i is given by the set $\mathcal{V}_i$ of valuations $v_i : \Phi_i \rightarrow \{\bot, \top\}$ for the variables Φ_i under his control. Thus, i plays the game by choosing some $v_i \in \mathcal{V}_i$. An *outcome* is a collection of choices, one for each player. Formally, a *strategy profile* or *outcome* for a game is a tuple $\vec{v} = (v_1, \ldots, v_n)$ in $\mathcal{V}_1 \times \cdots \times \mathcal{V}_n$. Each outcome uniquely defines an overall valuation for the variables Φ and, given the sets $\Phi_1, \ldots, \Phi_n$, each valuation uniquely defines an outcome. We will, therefore, treat outcomes for games as valuations. We will also often abuse notation and go back and forth between valuations and outcomes, for example writing $(v_1, \ldots, v_n) \models \varphi$ to mean that the valuation defined by the outcome $\vec{v} = (v_1, \ldots, v_n)$ satisfies formula $\varphi \in \mathcal{L}$. The preferences of a player i are captured by a utility function $u_i : \mathcal{V}_1 \times \cdots \times \mathcal{V}_n \rightarrow \{0, 1\}$ such that for all outcomes $\vec{v}$,

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

Thus, the concepts of game theory are readily available to Boolean games as well. This holds in particular for the well-known notion of (pure strategy) Nash equilibrium. We say an outcome $(v_1, \ldots, v_i, \ldots, v_n)$ is a *Nash equilibrium* if there is no player $i \in N$ and choice $w_i \in \mathcal{V}_i$ for i such that $u_i(v_1, \ldots, w_i, \ldots, v_n) > u_i(v_1, \ldots, v_i, \ldots, v_n)$. Thus, an outcome is a Nash equilibrium if no player can unilaterally deviate to obtain a better outcome for herself, under the assumption that every other player stays with its choice. We denote the Nash equilibrium outcomes of a game G by $\mathcal{N}(G)$; of course, it could be that $\mathcal{N}(G) = \emptyset$ for a given game G.

Boolean Games of Incomplete Information: A *Boolean game of incomplete information* (hereafter simply a "Boolean game" or just "game") is a Boolean game augmented with *visibility sets* $\Theta_i \subseteq \Phi$ for each agent i. Each agent's visibility set Θ_i indicates that player i is able to correctly observe the values of the variables in Θ_i and no other variables. It is natural to require that each player is able to observe the values of the variables under its control, i.e., $\Phi_i \subseteq \Theta_i$ for all $i \in N$. Note, however, that the results we present below go through without this assumption. Formally, a Boolean game of incomplete information is a tuple

$$(N, \Phi, \Phi_1, \ldots, \Phi_n, \gamma_1, \ldots, \gamma_n, \Theta_1, \ldots, \Theta_n).$$

We will denote games by $G, G', G_1, \ldots$ etc. If G is such that for every player $i \in N$ we have $\Theta_i = \Phi$ then we say that G is a game of *complete information*. Thus, in a game of complete information, every player can see every variable. We say that an outcome is a Nash equilibrium in a Boolean game of incomplete information if it is a Nash equilibrium in the underlying Boolean game. Thus, the notion of Nash equilibrium is not dependent on the visibility sets Θ_i for players i. The following example illustrates the definitions above.

Example 1 (Two Professors) Two professors and a student have a joint paper accepted at a conference. The question is who of them is going to attend: the first professor (p_1), the second (p_2), and/or the student (q). Relationships between the professors are a little complicated. Their goals are as follows. Professor 1 wants himself to attend and professor 2 to stay at home; he does not care about the presence of the student. Professor 2's goal is that neither professor goes on his own. The first professor can tell the student what to do. Each professor (only) observes whether himself attends and whether the student attends. We model this situation as a game

$$GG = (\{1, 2\}, \Phi, \Phi_1, \Phi_2, \gamma_1, \gamma_2, \Theta_1, \Theta_2),$$

with $\Phi = \{p_1, p_2, q\}$; $\gamma_1 = p_1 \wedge \neg p_2$ and $\gamma_2 = \neg(p_1 \wedge \neg q \wedge \neg p_2) \wedge \neg(\neg p_1 \wedge \neg q \wedge p_2)$; $\Phi_1 = \Theta_1 = \{p_1, q\}$ and $\Phi_2 = \{p_2\}$ and $\Theta_2 = \{p_2, q\}$. The game GG is illustrated in Figure 1.

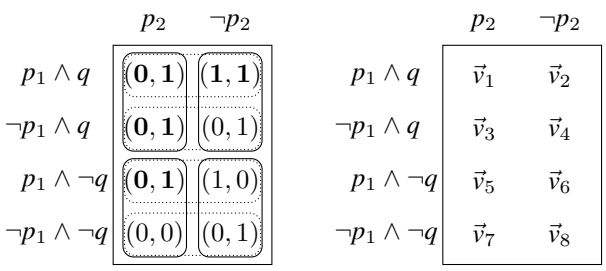

Figure 1: The normal-form game associated with the Boolean game *GG*. Actions/strategies are represented by formulae that characterise these choices. The first professor chooses rows, the second professor columns. The entries (x, y) represent the utility of professors 1 and 2, respectively. Nash equilibria are marked in bold. The matrix on the right indicates what the outcomes are called elsewhere in this paper. Epistemic indistinguishabilty relations for professors 1 and 2 are shown using dotted and solid lines, respectively.

3 Verifiable Equilibria

We now introduce the idea of *verifiable equilibria*. The motivation is as follows. A player *i* is part of a game *G*. The player can completely see the game—that is, it knows what the variables are, who controls what, and what the goals and visibility sets of each player are. However, when an outcome $\vec{v}$ is chosen, player *i* can only see the value of the variables in Θ_i. As a consequence, each player has some uncertainty about exactly what actions the other players have performed. We argue that, if the uncertainty is sufficiently large, then players that have *actually* played a Nash equilibrium collection of choices may in fact have no confidence that this is the case, because they cannot verify the actions of the others. As argued in the introduction, verifiable equilibria can be important in their own right, (such as in the example of nuclear disarmament given in the introduction), and they serve as natural focal points in general. So, we argue, for a Nash equilibrium to be both achievable and verifiable, not only must it satisfy the standard rationality requirements (that no player has any incentive to deviate) but the equilibrium choices of the players must convey sufficient information that each player has confidence that afterwards she can tell that the choices being made constitute a Nash equilibria. We formulate this idea as *verifiable feasibility*. We will shortly identify and investigate three types of verifiable feasibility; for this we will need some additional notation.

Where $\Delta \subseteq \Phi$ is a (sub)set of variables in a game *G*, we define an equivalence relation $\sim_\Delta$ over valuations, as follows:

$$v \sim_\Delta w \quad \text{iff} \quad \forall p \in \Delta : (v \models p \text{ iff } w \models p)$$

Thus $v \sim_\Delta w$ means that the valuations *v* and *w* agree on the values of all the variables in Δ. Clearly, this definition induces for every player $i \in N$ an "indistinguishability" relation $\sim_{\Theta_i}$ over valuations. Where there is no risk of confusion, we will write $\sim_i$ as a shorthand for $\sim_{\Theta_i}$. Alert readers will guess that the relations $\sim_i$ will later serve as epistemic accessibility relations [Fagin *et al.*, 1995].

Example 2 (Epistemic indistinguishability in *GG*) *The indistinguishability relations in the game GG are illustrated in Figure 1. Both agents know that in $\vec{v}_2$ both their goals are satisfied. Moreover, they can bring this about, and they know this. Still, they do not know that an equilibrium is played, e.g., since professor 2 considers it possible that the outcome in fact is $\vec{v}_4$, which is not an equilibrium. On the other hand, in $\vec{v}_1$ the outcome is known to be an equilibrium, although it does not satisfy professor 1's goal. Thus, although $\vec{v}_1$ is an equilibrium that can be* verified *by all players, it is not one that satisfies everybody's goals, whereas $\vec{v}_2$ is an equilibrium that makes everybody happy, although it cannot be verified.*

Weak Verifiable Equilibrium: We start with the simplest model of verifiable equilibrium. Suppose we are given a game *G* and an outcome $\vec{v} \in \mathcal{V}_1 \times \cdots \times \mathcal{V}_n$. We say that $\vec{v}$ is a *weak verifiable equilibrium* if every player considers it possible that $\vec{v}$ is a Nash equilibrium; that is, if for every player, there is some outcome $\vec{w}$ that is indistinguishable to *i* from $\vec{v}$ such that $\vec{w}$ is a Nash equilibrium. *It is important to note that weak verifiable equilibrium does not satisfy our criteria for a verifiable equilibrium, as an outcome $\vec{v}$ being a weak verifiable equilibrium does* not *mean that $\vec{v}$ is itself a Nash equilibrium.* Rather, *$\vec{v}$ could be*, as far as every player *i* is concerned. Formally, given game *G* and an outcome $\vec{v} \in \mathcal{V}_1 \times \cdots \times \mathcal{V}_n$, we say that $\vec{v}$ is a *weak verifiable equilibrium* if

$$\forall i \in N : \exists \vec{w} \in \mathcal{V} : \vec{v} \sim_i \vec{w} \text{ and } \vec{w} \in \mathcal{N}(G).$$

Let $\mathcal{W}(G)$ denote the weak verifiable equilibria of a game *G*.

Example 3 (Weak equilibria in *GG*) *In order to compute $\mathcal{W}(G)$, given that $\mathcal{N}(G) = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_5\} \subseteq \mathcal{W}(G)$, we only need to check membership of $\vec{v}_4, \vec{v}_6, \vec{v}_7$, and $\vec{v}_8$. In $\vec{v}_8$, the outcomes that professor 2 considers possible are $\vec{v}_8$ and $\vec{v}_6$ (in both, p_2 and q, the variables that 2 can see, are false). Neither of those outcomes is in $\mathcal{N}(G)$, and hence $\vec{v}_8 \notin \mathcal{W}(G)$. Symmetrically, given $\vec{v}_6$, professor 2 only considers $\vec{v}_8$ and $\vec{v}_6$ possible: neither are in $\mathcal{N}(G)$, so $\vec{v}_6 \notin \mathcal{W}(G)$. That $\vec{v}_7 \notin \mathcal{W}(G)$ follows similarly. Finally, in $\vec{v}_4$, professor 1 considers $\vec{v}_3$ possible and professor 2 considers $\vec{v}_2$ possible. Since both $\vec{v}_3$ and $\vec{v}_2$ are in $\mathcal{N}(G)$, we have $\vec{v}_4 \in \mathcal{W}(G)$.*

Let us now establish some basic properties of weak verifiable equilibria.

Proposition 1 *1. For all games G, we have $\mathcal{N}(G) \subseteq \mathcal{W}(G)$; for some games, the inclusion is strict.*

2. *For all games G, we have $\mathcal{W}(G) = \emptyset$ iff $\mathcal{N}(G) = \emptyset$.*
3. *For complete information games G, weak verifiable equilibria and Nash equilibria coincide: $\mathcal{W}(G) = \mathcal{N}(G)$.*
4. *If $\Theta_i = \emptyset$ for all players i, and $\mathcal{N}(G) \neq \emptyset$, then $\mathcal{W}(G) = \mathcal{V}$.*

We now analyse the problem of checking whether an outcome is a weak verifiable equilibrium and find it to be Σ_2^p-complete. The same applies to checking whether the set of weak verifiable equilibria is non-empty.

Theorem 1 *Given a game G and an outcome $\vec{v}$ for G, the problem of checking whether $\vec{v} \in \mathcal{W}(G)$ is Σ_2^p-complete;*

the problem of determining whether $\mathcal{W}(G) \neq \emptyset$ for a given game G is also Σ_2^p-complete.

Proof: For the first part, membership is by guess-and-check. For hardness, we reduce the problem of checking whether a Boolean game of complete information has a Nash equilibrium, which is known to be Σ_2^p-hard. Let G be a Boolean game of complete information with n players be given by $(N, \Phi, \Phi_1, \ldots, \Phi_n, \gamma_1, \ldots, \gamma_n)$, where $N = \{1, \ldots, n\}$. We construct a Boolean game of incomplete information G' with $n+1$ players given by $(N \cup \{n+1\}, \Phi', \Phi'_1, \ldots, \Phi'_{n+1}, \gamma'_1, \ldots, \gamma'_{n+1}, \Theta_1, \ldots, \Theta_{n+1})$ where $\Phi' = \Phi \cup \{a^*\}$, a^* is a variable not contained in Φ, $\Phi'_i = \Phi_i$ for all $1 \leq i \leq n$, $\Phi'_{n+1} = \{a^*\}$, $\Theta_i = \Phi'_i$ for all $1 \leq i \leq n+1$, and

$$\gamma'_i = \begin{cases} \gamma_i \vee a^* & \text{if } i \in N, \\ \top & \text{if } i = n+1. \end{cases}$$

Observe that for all $\vec{v} = (v_1, \ldots, v_{n+1})$ with $\vec{v}(a^*) = \bot$,

$$\vec{v} \in \mathcal{N}(G') \text{ if and only if } (v_1, \ldots, v_n) \in \mathcal{N}(G). \qquad (*)$$

To see this, let $\vec{v}(a^*) = \bot$ and assume that $\vec{v} = (v_1, \ldots, v_{n+1}) \in \mathcal{N}(G')$. Observe that $\vec{v} \not\models a^*$ and assume $(v_1, \ldots, v_n) \notin \mathcal{N}(G)$. Then, $(v_1, \ldots, v_n) \not\models \gamma_i$ and $(v_1, \ldots, w_i, \ldots, v_n) \models \gamma_i$ for some $i \in N$ and some w_i. Then, also $\vec{v} \not\models \gamma_i \vee a^*$. Let $\vec{w} = (v_1, \ldots, w_i, \ldots, v_n, v_{n+1})$. Then, $\vec{w} \models \gamma_i \vee a^*$. It follows that $\vec{v} \notin \mathcal{N}(G')$. For the opposite direction, assume $\vec{v} = (v_1, \ldots, v_{n+1}) \notin \mathcal{N}(G')$. Then, $\vec{v} \not\models \gamma_i \vee a^*$ and $(v_1, \ldots, w_i, \ldots, v_{n+1}) \models \gamma_i \vee a^*$ for some $i \in N \cup \{n+1\}$ and some w_i. Since $\gamma_i = \top$, we know that $i \neq n+1$. Also observe that $(v_1, \ldots, w_i, \ldots, v_{n+1}) \not\models a^*$ and, hence, $(v_1, \ldots, w_i, \ldots, v_{n+1}) \models \gamma_i$. Since, $vars(\gamma_i) \subseteq \Phi$, furthermore, $(v_1, \ldots, w_i, \ldots, v_n) \models \gamma_i$. It follows that $(v_1, \ldots, v_n) \notin \mathcal{N}(G)$.

To conclude the proof, let $\vec{v}_0$ denote the valuation that sets all variables in Φ' to $\bot$. We show that,

$$\vec{v}_0 \in \mathcal{W}(G') \text{ if and only if } \mathcal{N}(G) \neq \emptyset.$$

First, assume that $\vec{v}_0 \in \mathcal{W}(G')$ and consider player $n+1$. Then, there is some $\vec{v} = (v_1, \ldots, v_{n+1})$ such that both $\vec{v}_0 \sim_{n+1} \vec{v}$ and $\vec{v} \in \mathcal{N}(G')$. Hence, $\vec{v}(a^*) = \bot$ and with $(*)$ also $(v_1, \ldots, v_n) \in \mathcal{N}(G)$.

For the opposite direction, assume that $\vec{v}^* = (v_1^*, \ldots, v^*n) \in \mathcal{N}(G)$ and let $v_{n+1}^*(a^*) = \bot$. Thus, $\vec{v}_0 \sim_{n+1} (v_1^*, \ldots, v_n^*, v_{n+1}^*)$. In virtue of $(*)$, moreover, $(v_1^*, \ldots, v_n^*, v_{n+1}^*) \in \mathcal{N}(G')$. Finally, let $\vec{v}$ be the valuation that sets all variables in Φ to $\bot$ and a^* to $\top$. Obviously, $\vec{v}$ satisfies all players' goals. Hence, $\vec{v} \in \mathcal{N}(G')$. Moreover, for all $i \in N$, we have $\vec{v}_0 \sim_i \vec{v}$. Thus, for all $i \in N \cup \{i+1\}$, there is some $\vec{w} \in \mathcal{N}(G')$ with $\vec{v}_0 \sim_i \vec{w}$ and we may conclude that $\vec{v}_0 \in \mathcal{W}(G')$.[1] □

Note that the problem of verifying whether an outcome of complete information Boolean game is a Nash equilibrium is co-NP-complete, and thus, under standard complexity theoretic assumptions, *verifying weak verifiable equilibria is harder than the problem of verifying Nash equilibria in Boolean games of complete information*.

Strong Verifiable Equilibrium: Weak verifiable equilibria do not form a robust basis for action, as the fact that $\vec{v} \in \mathcal{W}(G)$ does not guarantee that $\vec{v} \in \mathcal{N}(G)$. We therefore introduce a stronger notion of verifiable equilibrium. We say $\vec{v}$ is a *strong verifiable equilibrium* if for every player i and for every outcome $\vec{w}$ that i cannot distinguish from $\vec{v}$, we have that $\vec{w}$ is a Nash equilibrium. Formally, we say that an outcome $\vec{v}$ of a game G is a *strong verifiable equilibrium* if

$$\forall i \in N : \forall \vec{w} \in \mathcal{V} : \vec{v} \sim_i \vec{w} \text{ implies } \vec{w} \in \mathcal{N}(G).$$

We let $\mathcal{S}(G)$ denote the strong verifiable equilibria of a game G. It may be readily appreciated that, for every game G, $\mathcal{S}(G) \subseteq \mathcal{N}(G)$, and for some G this inclusion is strict—e.g., for the game GG.

Example 4 (Strong equilibria in GG) *We know that $\mathcal{S}(G) \subseteq \mathcal{N}(G) = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_5\}$. In $\vec{v}_1$, professor 1 considers any outcome possible in which p_1 and q are true ($\vec{v}_1$ and $\vec{v}_2$), and we know that these outcomes are in $\mathcal{N}(G)$. Similarly, in $\vec{v}_1$ professor 2 considers $\vec{v}_1$ and $\vec{v}_3$ possible. These outcomes are also in $\mathcal{N}(G)$, hence $\vec{v}_1 \in \mathcal{S}(G)$. We have $\vec{v}_2 \sim_2 \vec{v}_4 \notin \mathcal{N}(G)$, hence $\vec{v}_2 \notin \mathcal{S}(G)$. Similarly, $\vec{v}_3 \sim_1 \vec{v}_4 \notin \mathcal{N}(G)$ implies that $\vec{v}_3 \notin \mathcal{S}(G)$. Finally, $\vec{v}_5 \sim_1 \vec{v}_6 \notin \mathcal{N}(G)$, hence $\vec{v}_5 \notin \mathcal{S}(G)$. Thus, $\mathcal{S}(G) = \{\vec{v}_1\}$.*

As with weak verifiable equilibria, strong verifiable equilibria and Nash equilibria coincide under complete information.

Observation 1 *If G is a game of complete information, then $\mathcal{S}(G) = \mathcal{W}(G) = \mathcal{N}(G)$.*

Theorem 2 *Given a game G and an outcome $\vec{v}$ for G, the problem of checking whether $\vec{v} \in \mathcal{S}(G)$ is co-NP complete.*

Proof: Membership is by showing the complementary problem is in NP, which can be achieved by guess-and-check. Given a certificate consisting of two outcomes $\vec{v}$ and $\vec{w}$ and two players i and j, it can be checked in polynomial time whether i wants to deviate from $\vec{v}$ to $\vec{w}$, that is, that $\vec{v}$ is not a Nash equilibrium. Moreover it is achievable in polynomial time to check whether player j can distinguish $\vec{v}$ and $\vec{w}$.

Hardness is shown by a reduction of the problem of checking whether a valuation $\vec{v}$ is a Nash equilibrium in a regular Boolean game or not, which is known to be co-NP hard. For a Boolean game $G = (N, \Phi, \Phi_1, \ldots, \Phi_n, \gamma_1, \ldots, \gamma_n)$ define Boolean game of incomplete information $G' = (N, \Phi, \Phi_1, \ldots, \Phi_n, \gamma_1, \ldots, \gamma_n, \Theta_1, \ldots, \Theta_n)$, such that $\Theta_i = \Phi$ for all players i, i.e., G' is the Boolean game of complete information corresponding to G. The result follows then immediately from Observation 1. □

Ideal Verifiable Equilibrium: Finally, we define *ideal verifiable equilibria*, which are outcomes that are commonly known to be equilibria, when they are played. One can argue that if there is a unique ideal verifiable equilibrium, this will be played: not only do I know that the outcome is an equilibrium, but I do know that you know it, and I know that you know that I know it, etc. Formally, let $\sim_C = (\bigcup_{i \in N} \sim_i)^+$ (where $+$ denotes transitive closure), and we say that an outcome $\vec{v}$ of a game G is an *ideal verifiable equilibrium* if

$$\forall i \in N : \forall \vec{w} \in \mathcal{V} : \vec{v} \sim_C \vec{w} \text{ implies } \vec{w} \in \mathcal{N}(G).$$

[1] We are indebted to an anonymous IJCAI reviewer for some simplifications of this proof.

Where G is a game, let $\mathcal{I}(G)$ denote the set of ideal verifiable equilibria of G. Clearly, $\mathcal{I}(G) \subseteq \mathcal{S}(G)$ for all G, and this inclusion is strict for some G (e.g., in the following example).

Example 5 (ideal verifiable equilibria in GG) *Since $\mathcal{I}(G) \subseteq \mathcal{S}(G) = \{\vec{v}_1\}$, we only need to check whether outcome $\vec{v}_1 \in \mathcal{I}(G)$. But since $\vec{v}_1 \sim_1 \vec{v}_2 \sim_2 \vec{v}_4 \notin \mathcal{N}(G)$, we conclude that $\vec{v}_1 \notin \mathcal{I}(G)$.*

Characterising Ideal Verifiable Equilibrium: Ideal verifiable equilibrium is a very strong concept: we will now show that many games do not have one. For this, we first give a characterization of the common knowledge relation $\sim_C$. Clearly, $\bigcap_{i \in n} \Theta_i$ is the set of variables *visible by all agents*. It now turns out that $\sim_C$ *is the universal relation on the set of valuations over variables not visible by all agents*.

Theorem 3 *For all outcomes $\vec{w}$ and $\vec{v}$ of a Boolean game: $\vec{w} \sim_C \vec{v}$ iff $\forall p \in \bigcap_{i \in N} \Theta_i : \vec{v}(p) = \vec{w}(p)$.*

In other words, two valuations are distinguishable by the common knowledge relation if and only if they disagree on at least one variable that is visible to all agents. This immediately gives us the following characterization of ideal verifiable equilibrium.

Corollary 1 *Let G be a game. $\vec{v} \in \mathcal{I}(G)$ iff $\vec{w} \in \mathcal{N}(G)$ for all $\vec{w}$ agreeing with $\vec{v}$ on the variables in $\bigcap_{i \in N} \Theta_i$.*

As a technical aside, we note that *reachability properties* (such as common knowledge) are often computationally hard to compute on succinctly represented graphs. Boolean games can be understood as providing a succinct representation for the relation $\sim_C$, and this might lead one to conclude that computational problems involving these relations would be computationally hard. However, by virtue of Corollary 1, we find that, (perhaps somewhat surprisingly), the ideal equilibrium case is in fact not computationally harder to check than weaker verifiable equilibria.

Theorem 4 *Given a game G and an outcome $\vec{v}$ for G, the problem of checking whether $\vec{v} \in \mathcal{I}(G)$ is co-NP-complete.*

Considering the set of variables visible to all agents, there are two borderline cases. First, if *all* variables are visible to all agents, then ideal verifiable equilibrium coincides with standard Nash equilibrium. Second, if *no* variable is visible to all agents, i.e., if every variable is invisible to at least on agent, $\sim_C$ is the universal relation:

Corollary 2 *If $\bigcap_{i \in N} \Theta_i = \emptyset$, i.e., if every variable is invisible to at least one agent, then $\vec{v} \sim_C \vec{w}$ for all $\vec{w}$ and $\vec{v}$.*

In this case an outcome is an ideal verifiable equilibrium if and only if *every* outcome is an ideal verifiable equilibrium, which in turn holds if and only if every outcome in the game is a Nash equilibrium.

Corollary 3 *For any game G, if every variable is invisible to at least one agent then $\mathcal{I}(G) = \emptyset$ if $\mathcal{N}(G) \neq \mathcal{V}_1 \times \cdots \times \mathcal{V}_n$, and $\mathcal{I}(G) = \mathcal{N}(G)$ otherwise.*

Thus, in the case that every variable is invisible to at least one agent, the games that have ideal verifiable equilibria are exactly those where every outcome is a Nash equilibrium. However, it is, we believe, plausible to assume that there are some commonly visible variables in a game. Indeed, one would intuitively expect that such variables provide the basis around which players will coordinate their actions.

In summary, we have seen that common knowledge of Nash equilibrium is a strong requirement, often difficult to obtain. That common knowledge is difficult to achieve is not surprising, as the situation is similar in related settings; for example in distributed systems where common knowledge is constant in every run [Meyer and van der Hoek, 1995, Corollary 2.2.6], and in public announcement games [Ågotnes and van Ditmarsch, 2011] where common knowledge of non-trivial Nash equilibria is impossible. Our framework demonstrates this once again.

4 Logical Characterisation

We now show how the *Epistemic Coalition Logic of Propositional Control* ("$\mathcal{ECL}$" for short) can be used to characterise and reason about the concepts we have introduced above. Introduced in [van der Hoek *et al.*, 2011], $\mathcal{ECL}$ combines the well-known modal epistemic language $\mathrm{S5}_n^C$ [Fagin *et al.*, 1995] with operators allowing us to represent the choices available to agents [van der Hoek and Wooldridge, 2005]. The language of $\mathcal{ECL}$ extends classical propositional logic with modal operators K_i and C for referring to the knowledge possessed by agents. A formula $K_i\varphi$, where $i \in N$ is an agent, is to be read "agent i knows that φ", while a formula $C\varphi$ is to be read "it is common knowledge that φ". In addition, the language contains operators $\Diamond_i\varphi$, where $i \in N$, with the intended interpretation that "assuming nothing else changes, then i has a choice such that φ holds". More precisely, $\Diamond_i\varphi$ means that i can assign values to the variables under its control in such a way as to make φ true.

Formally, the syntax of formula φ of $\mathcal{ECL}$ is defined with respect to a set $N = \{1, \ldots, n\}$ of agents and a set $\Phi = \{p, q, \ldots\}$ of Boolean variables by the following grammar:

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid \varphi \vee \varphi \mid K_i\varphi \mid C\varphi \mid \Diamond_i\varphi$$

where $p \in \Phi$ and $i \in N$.

We now give an interpretation of formulae of $\mathcal{ECL}$ with respect to pairs of the form $(G, \vec{v})$, called *pointed games*, where G is a game (with player set N and variable set Φ) and $\vec{v}$ is an outcome for G. We write $(G, \vec{v}) \models_{\mathcal{ECL}} \varphi$ to mean that the formula φ is true with respect to $(G, \vec{v})$, defined as follows:

$(G, \vec{v}) \models_{\mathcal{ECL}} p \quad \Leftrightarrow \vec{v} \models p$
$(G, \vec{v}) \models_{\mathcal{ECL}} \neg\varphi \quad \Leftrightarrow (G, \vec{v}) \not\models_{\mathcal{ECL}} \varphi$
$(G, \vec{v}) \models_{\mathcal{ECL}} \varphi \vee \psi \Leftrightarrow (G, \vec{v}) \models_{\mathcal{ECL}} \varphi$ or $(G, \vec{v}) \models_{\mathcal{ECL}} \psi$
$(G, \vec{v}) \models_{\mathcal{ECL}} K_i\varphi \quad \Leftrightarrow \forall \vec{w}$ with $\vec{v} \sim_i \vec{w}$: $(G, \vec{w}) \models_{\mathcal{ECL}} \varphi$
$(G, \vec{v}) \models_{\mathcal{ECL}} C\varphi \quad \Leftrightarrow \forall \vec{w}$ with $\vec{v} \sim_C \vec{w}$: $(G, \vec{w}) \models_{\mathcal{ECL}} \varphi$
$(G, \vec{v}) \models_{\mathcal{ECL}} \Diamond_i\varphi \quad \Leftrightarrow \exists \vec{w}$ with $\vec{v} \sim_{\Phi \setminus \Phi_i} \vec{w}$: $(G, \vec{w}) \models_{\mathcal{ECL}} \varphi$

Other propositional connectives $(\wedge, \rightarrow, \ldots)$ are defined in terms of $\neg$ and $\vee$ in the standard way. We also define $E\varphi = \bigwedge_{i \in N} K_i\varphi$ ("everyone knows that φ"). The *model checking problem* for $\mathcal{ECL}$ is as follows: we are given a pointed game $(G, \vec{v})$ and a formula φ, and asked whether $(G, \vec{v}) \models_{\mathcal{ECL}} \varphi$.

We now show how weak, strong and ideal verifiable equilibria can be quite naturally captured as formulae of our epistemic language. First, where G is a Boolean game, we define an $\mathcal{ECL}$ formula $\nu(G)$ (or just ν) as follows:

$$\nu(G) = \bigwedge_{i \in N} (\Diamond_i \gamma_i \rightarrow \gamma_i)$$

The key point about ν is that it characterises Nash equilibria:

Proposition 2 *For all pointed games $(G, \vec{v})$, we have: $(G, \vec{v}) \models_{\mathcal{ECL}} \nu$ iff $\vec{v} \in \mathcal{N}(G)$.*

Next, for each game G we define the following formulae:

$$\upsilon(G) = \bigwedge_{i \in N} \neg K_i \neg \nu \qquad \kappa(G) = C\nu \qquad \sigma(G) = E\nu$$

(We omit G when clear from context). With these definitions in place, we can state our logical characterisation results.

Proposition 3 *For all Boolean games G and outcomes $\vec{v}$: $\vec{v} \in \mathcal{W}(G)$ iff $(G, \vec{v}) \models_{\mathcal{ECL}} \upsilon$; $\vec{v} \in \mathcal{S}(G)$ iff $(G, \vec{v}) \models_{\mathcal{ECL}} \sigma$; $\vec{v} \in \mathcal{I}(G)$ iff $(G, \vec{v}) \models_{\mathcal{ECL}} \kappa$.*

Thus, the problem of checking verifiable equilibria can be reduced to $\mathcal{ECL}$ model checking problems.

Example 6 (Logical properties of *GG*) *In the game GG, we have $(GG, \vec{v}_4) \models_{\mathcal{ECL}} \neg\gamma_1 \wedge \neg\nu \wedge \neg K_1 \neg\nu \wedge \neg K_1 \nu \wedge Cq$: professor 1's goal γ_1 is not satisfied in $\vec{v}_4$; $\vec{v}_4$ is not a Nash equilibrium; professor 1 considers this possible, but he does not know it; and it is common knowledge that q holds. Also, $(GG, \vec{v}_1) \models_{\mathcal{ECL}} \gamma_2 \wedge \neg\gamma_1 \wedge \nu \wedge E\nu \wedge \neg C\nu$: $\vec{v}_1$ satisfies professor 2's goal but not professor 1's; v_1 is a Nash equilibrium and everybody knows this, still it is not common knowledge.*

Normal Forms: Above we gave a new interpretation of $\mathcal{ECL}$. We now turn to studying the resulting logic in more detail. Although the object language seems rather rich, we will now argue that in fact every formula φ of $\mathcal{ECL}$ is equivalent (on games) to a formula in propositional logic. More precisely, for every φ of $\mathcal{ECL}$ there is a formula π in propositional logic such that $(G, \vec{v}) \models_{\mathcal{ECL}} \varphi$ *iff* $\vec{v} \models \pi$.

We have to leave out the formal proof of this due to lack of space, and instead sketch the idea here based on examples. Let $\varphi(p_1, \ldots, p_k)$ be a formula where $p_1, \ldots, p_k$ are all the variables from Φ_i that occur in φ and let $\varphi(b_1, \ldots, b_k)$ denote φ with $b_i \in \{\bot, \top\}$ uniformly substituted for p_i. Then

$$\Diamond_i \varphi(p_1, \ldots, p_k) \equiv_{\mathcal{ECL}} \bigvee_{(b_1, \ldots, b_k) \in \{\bot, \top\}^k} \varphi(b_1, \ldots, b_k).$$

For examples, let us use atoms $p_i, q_i, \ldots$ to denote they are in Φ_i. Then, $\Diamond_1(p_1 \wedge (p_2 \vee q_2)) \equiv_{\mathcal{ECL}} (\bot \wedge (p_2 \vee q_2)) \vee (\top \wedge (p_2 \vee q_2)) \equiv_{\mathcal{ECL}} p_2 \vee q_2$. Indeed, if $\Phi_1 = \{p_1\}$, then for agent 1 to enforce $p_1 \wedge (p_2 \vee q_2)$, he needs to rely on the truth of those variables not in Φ_1, i.e., on $(p_2 \vee q_2)$. Similarly, $\Diamond_1 \Diamond_2 (p_1 \wedge \neg p_2 \wedge p_3)) \equiv_{\mathcal{ECL}} p_3 \equiv_{\mathcal{ECL}} \Diamond_1 (p_1 \wedge \Diamond (p_2 \wedge p_3))$.

Let $\varphi(q_1, \ldots, q_m)$ be a formula where $q_1, \ldots, q_m$ are all the variables in $\Phi \setminus \Theta_i$ occurring in φ. Then,

$$K_i \varphi(q_1, \ldots, q_m) \equiv_{\mathcal{ECL}} \bigwedge_{(b_1, \ldots, b_m) \in \{\bot, \top\}^m} \varphi(b_1, \ldots, b_m).$$

For example, suppose $\Theta_1 = \{p, q\}$ and $\Theta_2 = \{q, r\}$, while $\Phi = \{p, q, r, s\}$. Then $K_1(p \wedge q) \equiv_{\mathcal{ECL}} (p \wedge \top) \wedge (p \wedge \bot) \equiv \bot$: Indeed, 1 cannot know that q since $q \notin \Theta_1$. For another example, $K_1(K_2 q \wedge (K_2 r \vee K_2 \neg r)) \equiv_{\mathcal{ECL}} q$. If $\Theta_1 = \{p, s\}$ and $\Theta_2 = \{q, r\}$, then $K_2 \Diamond_1 (p \wedge (q \vee r)) \equiv_{\mathcal{ECL}} K_2 r \equiv_{\mathcal{ECL}} r$ while $K_1 \Diamond_1 (p \wedge (q \vee r)) \equiv_{\mathcal{ECL}} K_1 r \equiv_{\mathcal{ECL}} \bot$. For instance, note that we are not saying that $K_2 r \leftrightarrow r$ is a validity over all games even if it is true in games where $r \in \Theta_2$.

Since we have only a finite number (say k) of atoms, and the alternatives for the agents are valuations, we can finally replace every $C\varphi$ by $E\varphi \wedge EE\varphi \wedge \ldots \wedge EE \ldots E\varphi$, where the number of E operators in the last conjunct is 2^k.

This all implies that we do not need our rich object language to reason about even epistemic properties of Nash equilibria if we are only interested in expressivity. However, it is clear that the formulae of our object language are much more succinct than the equivalents we obtain in propositional logic.

5 Conclusions

In this paper, we proposed that verifiability can provide a means to distinguish between equilibria that are otherwise similar. We formalised verifiable equilibria by extending the standard model of Boolean games with sets of visible variables. As a consequence, players don't necessarily know which outcome they have ended up in. We formalised and studied three variants of verifiable equilibria. We characterised the complexity of computing them, gave conditions for their existence, and demonstrated how they can be characterised using modal logic. We emphasise that while we argue that verifiable equilibria are natural focal points, that does not mean that verifiable equilibria always are more reasonable than non-verifiable equilibria – there might be other reasons for selecting non-verifiable equilibria. For example, a game might have a non-verifiable equilibrium that consists of dominant strategies in addition to a verifiable equilibrium. The point is that verifiability can provide a solution to the problem of choosing between equilibria that otherwise are similar.

We looked at both strong verifiable equilibria, where all players know *a posteriori* that they are in an equilibrium, as well as ideal verifiable equilibria, where (in addition) it is common knowledge that an equilibrium has been played. In particular, we saw that the latter occur only in very special cases. It is important to note that the difference here is in the definition of verifiability: in the former case each agent knows that the outcome is an equilibrium but not necessarily that the other agents know that, unlike in the latter case. But in both cases it is common knowledge *before* the outcome is chosen, *which* outcomes are (strong or ideal) verifiable equilibria. The assumptions underpinning game theory with respect to what players in a game know about the game and each other, and the impact of relaxing or modifying these assumptions with respect to the outcomes of games that are predicted by game theoretic solution concepts, are studied in the field of *epistemic game theory* [Pacuit and Roy, 2012]. However, in this paper we are concerned with incomplete information about the *outcome*, i.e., about the actions chosen by all players *in the situation after the players have acted*.

For future research, it would be of interest to find more computationally tractable classes of games. We believe that the general framework in Section 4 can be used towards that end. Furthermore, other possible solution concepts can be of interest, taking the players' knowledge of whether or not they will be better off into account when defining best responses.

References

[Ågotnes and van Ditmarsch, 2011] Thomas Ågotnes and Hans van Ditmarsch. What will they say? – public announcement games. *Synthese (Special Section on Knowledge, Rationality and Action)*, 179(1):57–85, 2011.

[Binmore, 1992] K. Binmore. *Fun and Games: A Text on Game Theory*. D. C. Heath and Company: Lexington, MA, 1992.

[Bonzon *et al.*, 2006] E. Bonzon, 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.

[Dunne *et al.*, 2008] P. E. Dunne, 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.

[Endriss *et al.*, 2011] U. Endriss, S. Kraus, J. Lang, and M. Wooldridge. Designing incentives for boolean games. In *Proceedings of the Tenth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2011)*, Taipei, Taiwan, 2011.

[Fagin *et al.*, 1995] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning About Knowledge*. The MIT Press: Cambridge, MA, 1995.

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

[Meyer and van der Hoek, 1995] J.-J. Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge University Press: Cambridge, England, 1995.

[Pacuit and Roy, 2012] E. Pacuit and O. Roy. Epistemic foundations of game theory. Stanford Encyclopaedia of Philosophy, 2012.

[Schelling, 1980] Thomas Schelling. *The strategy of conflict*. Cambridge: Harvard University Press, 1980.

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

[van der Hoek *et al.*, 2011] W. van der Hoek, N. Troquard, and M. Wooldridge. Knowledge and control. In *Proceedings of the Tenth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2011)*, Taipei, Taiwan, 2011.