Representing Pure Nash Equilibria in Argumentation

Bruno Yun [a,*], Srdjan Vesic [b] and Nir Oren [a]
[a] *Department of Computing Science, University of Aberdeen, Scotland*
E-mails: bruno.yun@abdn.ac.uk, n.oren@abdn.ac.uk
[b] *CNRS, Univ. Artois, CRIL, France*
E-mail: vesic@cril.fr

Abstract. In this paper we describe an argumentation-based representation of normal form games, and demonstrate how argumentation can be used to compute pure strategy Nash equilibria. Our approach builds on Modgil's Extended Argumentation Frameworks. We demonstrate its correctness, showprove several theoretical properties it satisfies, and outline how it can be used to explain why certain strategies are Nash equilibria to a non-expert human user.

Keywords: Argumentation, Game Theory, Nash equilibrium, Pure strategy

1. Introduction

Game theory studies how multiple rational decision-makers should act given interactions between their strategies, and preferences over the resultant outcomes. Game theory has been applied to myriad fields [1]. Within game theory, decision-makers (referred to as players), their strategies, preferences and outcomes are represented within a game, and the solutions to a game identify some form of rational outcome. One such solution concept is that of a *dominant* strategy, where a player has a strategy or a set of strategies that will always result in the best outcome for them, regardless of what other players do. However, such dominant strategies often do not exist. In this work, we consider instead the notion of a *Nash equilibrium*, which identifies optimal strategies given that other players also pursue their own optimal strategies. Such Nash equilibria therefore represent a form of best response, and provide a well understood solution concept in game theory. However, finding Nash equilibria is computationally difficult, and it is sometimes difficult for a non-expert to understand why a given strategy is (or is not) a Nash equilibrium. We believe that by providing an argumentation-based representation of games, dialogues can be used to explain a Nash equilibrium to such non-experts. While work such as [2] has considered game theory in the context of ABA, to our knowledge, this work is the first to link abstract argumentation and Nash equilibria. We consider only so-called *pure strategies* for *normal form games* and intend to relax this restriction in future work.

The remainder of the paper is structured as follows. In Section 2, we provide a brief overview of argumentation and game-theory concepts necessary to understand our article. In Section 3, we describe how a normal form game can be encoded using argumentation. Section 4 examines some formal properties of

[*]Corresponding author. E-mail: bruno.yun@abdn.ac.uk.

our approach. Section 5 shows how we can build upon the proposed framework to provide explanations to a user about whether a strategy profile is a Nash equilibrium or not. Lastly, we discuss related and future work in Section 6 before concluding.

2. Background

We begin by providing the necessary background in game theory and argumentation required for the rest of the paper.

2.1. Game Theory

In this paper, we use the usual *normal form* for games [3].

Definition 1. ***(Normal Game)*** *A (normal) game is* $G = (Ag, Ac, Av, Ou, Ef, \leqslant)$ *where* $Ag = \{0, 1, \ldots, n\}$ *is a finite set of players;* Ac *is a finite set of strategies;* $Av = [Ac_0, \ldots, Ac_n]$ *with* $Ac_i \subseteq Ac$ *denoting the strategies available to i;* $Ou = \{o_0, \ldots, o_m\}$ *is a set of possible outcomes;* $Ef : Ac^n \to Ou^n$ *captures the consequences of the joint strategies for each player; and* $\leqslant = [\leqslant_0, \ldots, \leqslant_n]$ *with* $\leqslant_i \subseteq Ou \times Ou$ *denoting the preference relation for player i.*

The notation $o_k \leqslant_i o_l$ means that player i prefers outcome o_l to o_k. As commonly done, we write $o_i <_i o_j$ iff $o_i \leqslant_i o_j$ and $o_j \nleqslant_i o_i$[1]. Likewise, we will use the notation $o_i \geqslant_i o_j$ iff $o_i \nless_i o_j$ and $o_i >_i o_j$ iff $o_i \nleqslant_i o_j$. A *pure strategy profile* S is a tuple containing one strategy from each player in the game. The set of all such pure strategy profiles is $S_G = \Pi_{i \in Ag} Ac_i$, and represents one joint strategy of all players. A *partial strategy profile* is a tuple containing a single strategy for a subset of the players. Given any pure strategy profile $S = [s_0, \ldots, s_n]$, we write S_{-i} to denote the *partial strategy profile* $[s_0, \ldots, s_{i-1}, \emptyset, s_{i+1}, \ldots, s_n]$, where the strategy for player i is not specified. We then write $S_{-i} \oplus s_i$ to denote strategy profile S. With a slight abuse of notation, for any $S, S' \in S_G$ we write that $S \leqslant_i S'$ iff $Ef(S)_i \leqslant_i Ef(S')_i$[2].

Example 1. *Let us consider the stag hunt game* $G = (\{0, 1\}, Ac, Av, Ou, Ef, \leqslant)$*, where* $Ac = \{stag, hare\}$*,* $Av = [Ac, Ac]$*,* $Ou = \{4, 3, 2, 1\}$*,* $\leqslant$ *is the standard less than relation over numbers. Table 1a graphically illustrates this game in normal form, and specifies* Ef*. For example, the tuple* $(1, 3)$ *in the column "hare" and row "stag" means that* $Ef([stag, hare]) = (1, 3)$*. Given the pure strategy profile* $S = [stag, hare]$*,* $S_{-0} = [\emptyset, hare]$ *and* $S_{-0} \oplus hare = [hare, hare]$*. Here* $[stag, hare] \leqslant_0 [hare, hare]$ *because* $(1, 3)_0 \leqslant_0 (2, 2)_0$ *but* $[hare, hare] \leqslant_1 [stag, hare]$*.*

In asking why a player should pursue some strategy, we must take into account the strategies of others. If each player has chosen a strategy, and no player can increase their own outcome by changing their strategy while the other players keep theirs unchanged, then the current pure strategy profile constitutes a Nash equilibrium.

[1] We assume that for all players i, $\leqslant_i$ is transitive and complete (each two outcomes are comparable). Thus, $\leqslant_i$ is acyclic. I.e., if $a <_i b <_i c$ then $c \nleqslant_i a$.

[2] The notation $Ef(S')_i$ means the i-th element of $Ef(S')$.

Table 1

Two games in normal form.

		Player 1	
		stag	*hare*
Player 0	*stag*	4 4	3 1
	hare	1 3	2 2

(a) Stag Hunt

		Player 1	
		heads	*tails*
Player 0	*heads*	-1 1	1 -1
	tails	1 -1	-1 1

(b) Matching pennies.

Definition 2. *Let $G = (Ag, Ac, Av, Ou, Ef, \leqslant)$, we say that $S \in S_G$ is a Nash equilibrium if for every $i \in Ag$ and for any strategy $s \in Ac_i$, it holds that $S_{-i} \oplus s \leqslant_i S$.*

A simple algorithm to identify all Nash equilibrium in the presence of pure strategies involves iterating through every player and identifying the best strategy profile (in terms of *Ef* for that player) given all other players' possible joint strategies. Any strategy profile which all players consider best is then a Nash equilibrium.

Given a game in normal form, the above algorithm involves – for a two player game – scanning down each column and marking the best strategy for the row player, and then doing the same for each row marking the best strategy for the column player. Each cell marked for both players is a Nash equilibrium. In the remainder of this paper, we show an argumentation-based alternative.

Example 2 (Cont'd)**.** *There are two Nash equilibria in the stag hunt game: $[stag, stag]$ and $[hare, hare]$. The strategy profile $[stag, stag]$ is a Nash equilibrium because $[hare, stag] \leqslant_0 [stag, stag]$ and $[stag, hare] \leqslant_1 [stag, stag]$. Similarly, $[hare, hare]$ is also a Nash equilibrium as $[stag, hare] \leqslant_0 [hare, hare]$ and $[hare, stag] \leqslant_1 [hare, hare]$.*

2.2. Argumentation

We encode normal form games in terms of arguments and attacks by building on Modgil's Extended Argumentation Frameworks (EAF) [4].

Definition 3. *An Extended Argumentation Framework is a triple $\langle \mathbb{A}, \mathbb{C}, \mathbb{D} \rangle$ where $\mathbb{A}$ is a set of arguments, $\mathbb{C} \subseteq \mathbb{A} \times \mathbb{A}$, $\mathbb{D} \subseteq \mathbb{A} \times \mathbb{C}$ and if $(z, (x, y)), (z', (y, x)) \in \mathbb{D}$ then $(z, z'), (z', z) \in \mathbb{C}$.*

Definition 4 (Defeat)**.** *Let $\mathcal{AS} = (\mathbb{A}, \mathbb{C}, \mathbb{D})$ be an EAF, $x, y \in \mathbb{A}$ and $Y \subseteq \mathbb{A}$. We say that y defeats x w.r.t. Y, denoted $y \rightarrow_Y x$ iff $(y, x) \in \mathbb{C}$ and there is no $z \in Y$ s.t. $(z, (y, x)) \in \mathbb{D}$.*

Definition 5 (Argumentation semantics)**.** *Let $\mathcal{AS} = (\mathbb{A}, \mathbb{C}, \mathbb{D})$ be an EAF and $E \subseteq \mathbb{A}$. We say that:*

- *E is conflict-free iff for every $x, y \in E$, if $(y, x) \in \mathbb{C}$ then $(x, y) \notin \mathbb{C}$, and there exists $z \in E$ s.t. $(z, (y, x)) \in \mathbb{D}$.*
- *$x \in \mathbb{A}$ is acceptable w.r.t. E iff for every $y \in \mathbb{A}$ s.t. $y \rightarrow_E x$, there exists $z \in E$ s.t. $z \rightarrow_E y$ and there exists $R_E = \{x_1 \rightarrow_E y_1, \ldots, x_n \rightarrow_E y_n\}$ s.t. for every $i \in \{1, \ldots, n\}, x_i \in E$, $z \rightarrow_E y \in R_E$ and for every $x_j \rightarrow_E y_j \in R_E$, for every y' s.t. $(y', (x_j, y_j)) \in \mathbb{D}$, there exists $x' \rightarrow_E y' \in R_E$*
- *E is an admissible extension iff every argument in E is acceptable w.r.t. E*

- *E is a preferred extension iff E is a maximal (w.r.t. $\subseteq$) admissible extension*
- *E is a stable extension iff for every $y \notin E$, there exists $x \in E$ such that $x \rightarrow_E y$.*

We will use the notation $Ext_s(\mathcal{AS})$ (resp. $Ext_p(\mathcal{AS})$) to denote the set of all stable (resp. preferred) extensions.

We note in passing that it is possible to *flatten* an EAF, that is, transform it to a standard abstract argumentation framework such that all arguments within an extension (according to some semantics) within the EAF are equivalently found in the extension of the abstract framework [5–7]. Therefore, standard argumentation solvers [8] can be applied — once flattened — to identify justified arguments within an EAF.

3. Argumentation-based approach for games

We consider an argumentation framework with multi-level arguments. At the base level, we consider all possible strategy profiles as arguments. Since only a single strategy profile can ever occur (as players execute one set of strategies in the interaction), every argument at this level must attack every other argument. We refer to such arguments as *game-based arguments*, and note that they are equivalent to pure strategy profiles.

Definition 6 (Game-based argument)**.** *Let $G = (Ag, Ac, Av, Ou, Ef, \leqslant)$ be a game, a game-based argument (w.r.t. G) is a pure strategy profile $S \in S_G$.*

The set of all game-based arguments for a game G is denoted by $\mathcal{A}_g(G)$.

Next, we introduce *preference arguments*. Intuitively, these can be interpreted as statements of the form: "Given that the other players are performing a given set of strategies, the remaining player's preferred strategy should be playing x".

Definition 7 (Preference argument)**.** *Let $G = (Ag, Ac, Av, Ou, Ef, \leqslant)$ be a game, $S \in S_G$ be a pure strategy profile and $i \in Ag$. A preference argument (w.r.t. G) is a tuple (S_{-i}, s), where $s \in Ac_i$.*

The set of preference arguments for a game G is denoted by $\mathcal{A}_p(G)$. A *cluster* of preference arguments is a maximal set of preference arguments sharing the same partial strategy profile.

Finally, we introduce *valuation arguments*, which can be interpreted as statements of the form: "Given that the other players are performing a given set of strategies, it is the case that the outcome of strategy s is better than the outcome of strategy s' for the remaining player".

Definition 8 (Valuation argument)**.** *Let $G = (Ag, Ac, Av, Ou, Ef, \leqslant)$ be a game, $i \in Ag$, $(S_{-i}, s), (S_{-i}, s') \in \mathcal{A}_p(G)$ be two preference arguments and $S_{-i} \oplus s' <_i S_{-i} \oplus s$. A valuation argument (w.r.t. G) is the pair $(S_{-i}, s' < s)$.*

The set of valuation arguments for a game G is denoted by $\mathcal{A}_v(G)$.

Example 3 (Cont'd)**.** *The sets of game-based, preference and valuation arguments w.r.t. G are shown in Table 2. The argument a_1 represents the case where player 0 chooses to hunt a stag and player 1 chooses to hunt a hare. The argument a_9 represents the argument: "Given that player 0 chooses to hunt a hare,*

Table 2

Arguments for the stag hunt game

Game-based arguments	Preference arguments	Valuation arguments
$a_1 = [stag, hare]$	$a_5 = ([stag, \emptyset], stag)$	$a_{13} = ([stag, \emptyset], stag > hare)$
$a_2 = [stag, stag]$	$a_6 = ([stag, \emptyset], hare)$	$a_{14} = ([\emptyset, stag], stag > hare)$
$a_3 = [hare, stag]$	$a_7 = ([\emptyset, stag], stag)$	$a_{15} = ([hare, \emptyset], hare > stag)$
$a_4 = [hare, hare]$	$a_8 = ([\emptyset, stag], hare)$	$a_{16} = ([\emptyset, hare], hare > stag)$.
	$a_9 = ([hare, \emptyset], stag)$	
	$a_{10} = ([hare, \emptyset], hare)$	
	$a_{11} = ([\emptyset, hare], stag)$	
	$a_{12} = ([\emptyset, hare], hare)$	

player 2's preferred strategy should be to hunt a stag". The argument a_{16} represents the argument: "Given that player 1 chooses to hunt a hare, the outcome of hunting a hare is better than the outcome of hunting a stag for player 0".

We now turn our attention to attacks. We note that preference and valuation arguments provide reasons why one argument should not attack another, and therefore introduce not only attacks between arguments, but also attacks on attacks.

Definition 9 (Attack). *For a game $G = (Ag, Ac, Av, Ou, Ef, \leqslant)$, $\alpha_1, \alpha_2 \in \mathcal{A}_g(G)$, $a_3 = (S_1, s_2), \alpha_4 = (S_3, s_4) \in \mathcal{A}_p(G)$ and $\alpha_5 = (S_5, s_6 > s_7) \in \mathcal{A}_v(G)$. We say that:*

- *α_1 attacks α_2, denoted $(\alpha_1, \alpha_2) \in \mathcal{C}_r(G)$, iff $\alpha_1 \neq \alpha_2$.*
- *α_3 attacks α_4, denoted $(\alpha_3, \alpha_4) \in \mathcal{C}_p(G)$, iff $S_1 = S_3$ and $s_2 \neq s_4$.*
- *α_3 attacks $(\alpha_1, \alpha_2) \in \mathcal{C}_r(G)$, denoted by $(\alpha_3, (\alpha_1, \alpha_2)) \in \mathcal{C}_u(G)$, iff there exists $s \in Ac$ such that $S_1 \oplus s = \alpha_1$ and $S_1 \oplus s_2 = \alpha_2$.*
- *α_5 attacks $(\alpha_3, \alpha_4) \in \mathcal{C}_p(G)$, denoted by $(\alpha_5, (\alpha_3, \alpha_4)) \in \mathcal{C}_v(G)$, iff $S_5 = S_3$, $s_6 = s_4$ and $s_7 = s_2$.*

The first attack captured within Definition 9 is between every two distinct game-based arguments. As each player has to choose exactly one strategy, different strategy profiles are clearly incompatible. The second bullet point represents attacks between preference arguments. In the stag hunt example for instance, a_5 attacks a_6 (and vice-versa) because in the event of player 0 hunting a stag, player 1 can either hunt a stag or a hare. The third type of attack captures attacks from preference arguments to attacks between game-based arguments. Within the stag hunt, a_5 attacks (a_1, a_2) because a_5 states that it is preferable for player 1 to hunt a stag when player 0 is also hunting a stag. Note that in general, the preference argument (S_1, s_2) attacks *all* attacks against the game-based argument $S_1 \oplus s_2$ coming from any other game-based arguments of the form $S_1 \oplus s'$, for any $s' \in Ac$ such that $s' \neq s_2$. The last type of attack captures attacks from valuation arguments to attacks between preference arguments. Returning to the stag hunt, a_{13} attacks (a_6, a_5) as a_{13} states that the strategy "hunt a stag" is better than the strategy "hunt a hare" for player 1 when player 0 is hunting a stag.

The arguments and attacks induce a very specific type of extended argumentation framework, where object-level (game-based) arguments have their attacks attacked by meta-arguments (preference arguments) at level one, and where attacks between these meta-arguments are attacked by meta-arguments at level two (valuation arguments).

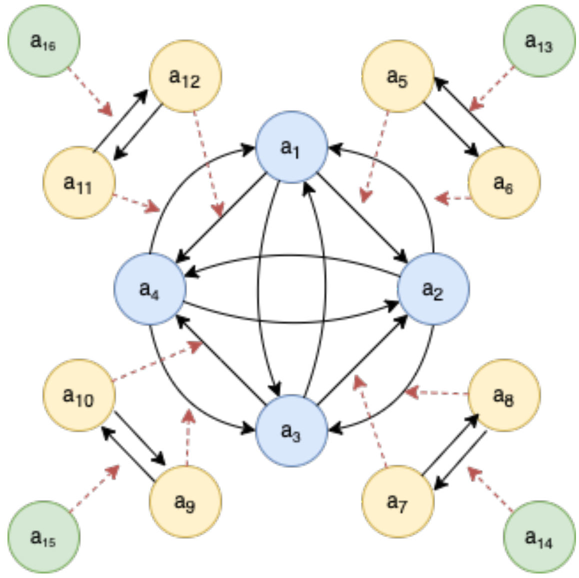

Fig. 1. Argumentation graph corresponding to stag hunt game

The first layer is needed to encode every possible outcomes, the second layer is useful for specifying outcomes that are comparable whereas the third layer returns an agent's preference between two outcomes.

Definition 10 (Argumentation framework)**.** *Let G be a game. The argumentation framework corresponding to G is the tuple* $\mathcal{AS}_G = (\mathbb{A}, \mathbb{C}, \mathbb{D})$ *where* $\mathbb{A} = \mathcal{A}_g(G) \cup \mathcal{A}_p(G) \cup \mathcal{A}_v(G)$*,* $\mathbb{C} = \mathcal{C}_r(G) \cup \mathcal{C}_p(G)$ *and* $\mathbb{D} = \mathcal{C}_u(G) \cup \mathcal{C}_v(G)$*.*

Example 4 (Example 3 Contd)**.** *Figure 1 represents the game-based, preference and valuation arguments of G using blue, yellow and green nodes respectively. The attacks between arguments (*$\mathbb{C}$*) and on attacks (*$\mathbb{D}$*) are represented using solid black arrows and dashed red arrows respectively.*

For our framework to be an EAF, it must satisfy some constraints, as described in [9], and we can easily show that this is the case.

Proposition 1. *Let G be a game and* $\mathcal{AS}_G = (\mathbb{A}, \mathbb{C}, \mathbb{D})$ *be the corresponding argumentation framework, it holds that if* $(z, (x, y)), (z', (y, x)) \in \mathbb{D}$ *then* $(z, z'), (z', z) \in \mathbb{C}$*.*

Proof. There are only two types of attacks on attacks: (1) attacks coming from valuation arguments to attacks between preference arguments and (2) attacks coming from preference arguments to attacks between game-based arguments. In the rest of this proof, we prove that Proposition 1 is satisfied for the two types of attacks on attacks.

- Considering (1), for a fixed partial strategy profile S_i, and fixed strategies $s_j, s_k \in Ac$, there is exactly one (or no) valuation argument of the form $(S_i, s_j > s_k)$ or $(S_i, s_k > s_j)$. As a result, the condition in Proposition 1 is trivially satisfied for attacks coming from valuation arguments.

- We now study the case (2) and show that Proposition 1 is also satisfied for attacks coming from preference arguments on attacks between game-based arguments. Assume that $(a_3, (x, y))$, $(a_4, (y, x)) \in \mathbb{D}$, where $a_3 = (S_1, s_2)$, $a_4 = (S_1, s_4)$, $x = S_1 \oplus s_4$ and $y = S_1 \oplus s_2$. By Definition 9, $s_2 \neq s_4$ thus $(a_3, a_4), (a_4, a_3) \in \mathcal{C}_p(G) \subseteq \mathbb{C}$.

  □

Since – given Proposition 1 – our argumentation system is an EAF, we can use EAF semantics to evaluate it.

Example 5 (Example 4 Contd)**.** *In our running example, a_5 defeats a_6 w.r.t. $\mathbb{A}$ as $(a_5, a_6) \in \mathbb{C}$ and there is no argument $z \in \mathbb{A}$ such that $(z, (a_5, a_6)) \in \mathbb{D}$. However, a_6 does not defeat a_5 w.r.t. $\mathbb{A}$ because $(a_{13}, (a_6, a_5)) \in \mathbb{D}$. All extensions contain arguments $\{a_{16}, a_{15}, a_{14}, a_{13}, a_{12}, a_{10}, a_7, a_5\}$, while one preferred extension contains $\{a_2\}$ and the other contains $\{a_4\}$.*

4. System Properties

Having described our system, we now consider its properties. The most important result we seek to show is the correspondence between argumentation semantics and Nash equilibria, and we begin by laying the groundwork for this. We then consider how many arguments will be generated for an arbitrary normal form game.

We begin by considering which preference arguments will appear in a preferred extension. This result is used in later proofs.

Lemma 1. *Let $G = (Ag, Ac, Av, Ou, Ef, \leqslant)$ be a game, and $\mathcal{AS}_G$ be the corresponding AS. For each preferred extension E of $\mathcal{AS}_G$, for each cluster C of preference arguments, there exists a unique argument $c \in C$ such that $c \in E$.*

Proof. Assume a partial strategy profile $S = [s_0, \dots, s_{i-1}, \emptyset, s_{i+1}, s_n]$ and the corresponding cluster of preference arguments C. Because our preferences are complete and acyclic, we know that there exists a strategy s^* such that for every $s \in Ac_i$, $S \oplus s \leqslant_i S \oplus s^*$. From the definition of the valuation argument, there are no valuation arguments attacking the attacks from the preference argument (S, s^*) to other preference arguments. As a result, we conclude that (S, s^*) is in a preferred extension E and that all the other arguments in C are not E. Moreover, you need to choose one such argument from the cluster C for each preferred extension to satisfy the maximality condition of the semantics. □

Next, we show that if there is a preferred extension with game-based arguments, then each such extension has exactly one game-based argument.

Lemma 2. *If any preferred extension of $\mathcal{AS}_G$ contains a game-based argument, then it contains exactly one game-based argument.*

Proof. Let E be a preferred extension containing game-based arguments. We prove by contradiction that it is not possible for E to have more than one game-based argument. Assume that E contains two game-based arguments a_1 and a_2. By definition of the attack relation, there is a symmetric attack between a_1 and a_2. Hence there must exist two preference arguments p_3 and p_4 with $(p_3, (a_1, a_2)), (p_4(a_2, a_1)) \in \mathbb{D}$

and $(p_3, p_4), (p_4, p_3) \in \mathbb{C}$. It is not possible for both (p_4, p_3) and (p_3, p_4) to be attacked by valuation arguments as this would require an inconsistency or cycle in $\leqslant$. By this observation, E contains only p_3 or p_4. Hence, $\{a_1, a_2\}$ is not conflict-free, contradiction.
□

We now show that a game-based argument which is not a Nash equilibrium will not appear in any preferred extension of the associated argumentation system.

Lemma 3. *Let $G = (Ag, Ac, Av, Ou, Ef, \leqslant)$ be a game, and $\mathcal{AS}_G$ be the corresponding AS. If $S \in S_G$ such that S is not a Nash equilibrium then for every preferred extension $E, S \notin E$.*

Proof. Assume there is a non-Nash equilibrium game-based argument $S' = [s'_0, \ldots, s'_n]$ in a preferred extension E. Then, from Lemma 2, E does not contain any other game-based arguments. Since S' is not a Nash equilibrium, there exists $i \in Ag$ and $s \in Ac_i$ such that $S'_{-i} \oplus s'_i <_i S'_{-i} \oplus s$. In the rest of this proof, we consider the strategy s^* such that for every $s \in Ac_i, S'_{-i} \oplus s \leqslant_i S'_{-i} \oplus s^*$. By definition, the attack from S' to $S'_{-i} \oplus s^*$ is attacked by the preference argument (S'_{-i}, s^*). Moreover, the preference argument (S'_{-i}, s^*) attacks all the other preference arguments (S'_{-i}, s'), where $s' \in Ac_i$ and $s' \neq s$. By definition of the valuation arguments, none of the attacks from (S'_{-i}, s^*) to those other preference arguments is defeated. As a result, we conclude that there is a preferred extension that contains (S'_{-i}, s^*). Let $s^+ = \{s \in Ac_i \mid S'_{-i} \oplus s \leqslant_i S'_{-i} \oplus s^* \text{ and } S'_{-i} \oplus s^* \leqslant_i S'_{-i} \oplus s\}$, we can conclude that there is at least one argument $(S'_{-i}, s_o), s_o \in s^+$ in E (Lemma 1) and (S'_{-i}, s_o) attacks the attack from S' to $S'_{-i} \oplus s_o$, contradiction. □

Corollary 1. *Let $G = (Ag, Ac, Av, Ou, Ef, \leqslant)$ be a game, and $\mathcal{AS}_G$ be the corresponding AS. If E is a preferred extension that contains a game-based argument S, then S is a Nash equilibrium.*

In the next proposition, we show that if a preferred extension contains a game-based argument, then it is a stable extension.

Proposition 2. *Let G be a game and $\mathcal{AS}_G = (\mathbb{A}, \mathbb{C}, \mathbb{D})$ be the corresponding argumentation framework. If $E \in Ext_p(\mathcal{AS}_G)$ and $E \cap \mathcal{A}_g(G) \neq \emptyset$ then $E \in Ext_s(\mathcal{AS}_G)$.*

Proof. We show that if a preferred extension possesses a game-based argument, then it is also a stable extension. Assume E contains a single game-based argument. By Lemma 2, E contains exactly one game-based argument. Therefore, all game-based arguments not in the extension are defeated by the game-based argument within the extension with respect to E, meaning that the game-based argument is a member (at the game-based level) of the stable extension. □

It may seem intuitive that the preferred and stable extension should coincide. However, this is not the case, as demonstrated by the following counter-example.

Example 6. *Consider the matching pennies game $G = (Ag, Ac, Av, Ou, Ef, \leqslant)$ where $Ag = \{0, 1\}, Ac = \{heads, tails\}, Av = [Ac, Ac], Ou = \{1, -1\}$, $\leqslant$ is defined as the "less-than relation" for each player, and Ef is defined in Table 1b.*

The set of arguments is $\mathbb{A} = \{b_1, b_2, b_3, \ldots, b_{16}\}$ and are listed in Table 3. There is only one preferred extension $\{b_{16}, b_{15}, b_{14}, b_{13}, b_{12}, b_{10}, b_8, b_6\}$ but no stable extensions.

Table 3

Arguments for the matching pennies game

Game-based arguments	Preference arguments	Valuation arguments
$b_1 = [heads, heads]$	$b_5 = ([heads, \emptyset], heads)$	$b_{13} = ([heads, \emptyset], tails > heads)$
$b_2 = [heads, tails]$	$b_6 = ([heads, \emptyset], tails)$	$b_{14} = ([\emptyset, tails], tails > heads)$
$b_3 = [tails, tails]$	$b_7 = ([\emptyset, tails], heads)$	$b_{15} = ([tails, \emptyset], heads > tails)$
$b_4 = [tails, heads]$	$b_8 = ([\emptyset, tails], tails)$	$b_{16} = ([\emptyset, heads], heads > tails)$
	$b_9 = ([tails, \emptyset], tails)$	
	$b_{10} = ([tails, \emptyset], heads)$	
	$b_{11} = ([\emptyset, heads], tails)$	
	$b_{12} = ([\emptyset, heads], heads)$	

Table 4

Three strategy variant of the matching pennies game.

		Player 1		
		heads	*tails*	*edge*
	heads	-1 1	1 -1	1 -1
Player 0	*tails*	1 -1	-1 1	-1 1
	edge	1 -1	-1 1	-1 1

Furthermore, even when multiple preferred extensions exist, these may not coincide with the stable extensions.

Example 7. *Let us consider the following variant of the matching pennies game with three strategies for each player. We have $G = (Ag, Ac, Av, Ou, Ef, \leqslant)$ where $Ag = \{0, 1\}, Ac = \{heads, tails, edge\}, Av = [Ac, Ac], Ou = \{1, -1\}$, $\leqslant$ is defined as the "less-than" relation for numbers for each player, and Ef is defined in Table 4. This variant of the game has eight distinct preferred extensions, but none contain any game-based arguments.*

We now turn to our main result, namely the equivalence of the Nash equilibrium with the game-based arguments found in the preferred extensions.

Proposition 3 (Equivalence)**.** *Let $G = (Ag, Ac, Av, Ou, Ef, \leqslant)$ be a game, and $\mathcal{AS}_G$ be the argument framework for the game. A strategy profile $S = [s_0, \dots, s_n] \in S_G$ is a Nash equilibrium iff there exists $E \in Ext_p(\mathcal{AS}_G)$ such that $S \in E$.*

Proof. We split this proof in two parts:

($\Rightarrow$) We need to show that if S is a Nash equilibrium, then it is within a preferred extension of $\mathcal{AS}_G$. Let us consider the set of arguments $E = \{S\} \cup \mathcal{A}_v(G) \cup \{(S_{-i}, s_i) \mid i \in Ag\}$. We now show that E is a preferred extension of $\mathcal{AS}_G$. It is clear that E is conflict-free as for every $x, y \in E, (x, y) \notin \mathbb{C}$. Every argument in $\mathcal{A}_v(G)$ is acceptable w.r.t. E as valuation arguments are not attacked. Every argument $a = (S_{-i}, s_i)$ is also acceptable w.r.t. E because for every $s' \in Ac_i$ and $s' \neq s_i$, the attacks from $a' = (S_{-i}, s')$ to a, is either not a defeat w.r.t. E (if there is a valuation argument that attacks (a', a)) or it is a defeat but a' is defeated by a w.r.t. E. The argument S is also acceptable w.r.t.

E because for every $S' \in S_G$ and $S' \neq S$, the attack from S' to S is not a defeat w.r.t. E as the arguments (S_{-i}, s_i) are attacking those attacks. We conclude that the set E is admissible. Following Lemma 2 and 1, we conclude that E is maximal for set inclusion as it contains all the valuation arguments, one preference argument per cluster and exactly one game-based argument.

($\Leftarrow$) We need to show that if S is within a preferred extension, then S is a Nash equilibrium. This follows directly from the result from Corollary 1.

□

Returning to the stable extensions, the following result shows that there is a one-to-one correspondence between the sets of Nash equilibria and the set of classes of stable extensions[3], where each Nash equilibrium S corresponds to the class of stable extensions containing argument S.

Corollary 2. *Let $G = (Ag, Ac, Av, Ou, Ef, \leqslant)$ be a game, and $\mathcal{AS}_G$ be the corresponding EAF. There is a bijection between $Y = \{S \in S_G \mid S \text{ is a Nash equilibrium}\}$ and $\{\{E \in Ext_s(\mathcal{AS}_G) \mid S' \in E\} \mid S' \in Y\}$*

Proof. Follows directly from Proposition 3 and Proposition 2. □

Finally, we consider how many arguments an argumentation system representing a normal form game will contain.

Proposition 4 (Number of arguments)**.** *Let $G = (Ag, Ac, Av, Ou, Ef, \leqslant)$ be a game s.t. $|Ag| = n$ and $m = \max_{i \in Ag} |Ac_i|$, the number of arguments in $\mathcal{AS}_G$ is in $\mathcal{O}(m^{n+1} \cdot n)$.*

Proof. The proof is split into three parts.

(1) Suppose n players and m strategies per player. Each game-based argument corresponds to a pure strategy profile, i.e., there are m^n game-based arguments.
(2) Consider the number of the preference arguments. There are $m^{n-1} \cdot n$ partial strategy profiles. Roughly speaking, a preference argument is obtained from a partial strategy profile by replacing the empty set with a strategy. Hence, there are up to $m^{n-1} \cdot n \cdot m = m^n \cdot n$ preference arguments.
(3) We estimate the number of valuation arguments. Each valuation argument is obtained from one partial strategy profile and one pair of different strategies. There are $m^{n-1} \cdot n$ partial strategy profiles and up to $m \cdot (m-1)$ pairs of different strategies. Furthermore, if a strategy x is preferred to strategy y, then y is not preferred to x. Thus, there are up to $\frac{m \cdot (m-1)}{2}$ possible combinations to consider. Hence, the total number of valuation arguments is limited by $\frac{m^{n-1} \cdot m \cdot (m-1) \cdot n}{2}$ which is in $\mathcal{O}(m^{n+1} \cdot n)$. Thus, the total number of arguments is in $\mathcal{O}(m^n) + \mathcal{O}(m^n \cdot n) + \mathcal{O}(m^{n+1} \cdot n)$ which is in $\mathcal{O}(m^{n+1} \cdot n)$.

□

We note that computing Nash equilibria is known to be computationally difficult, and the result regarding the number of arguments is therefore unsurprising.

[3]We say two stable extensions are equivalent iff they have the same game-based argument

5. Dialogue-based Explanations

In this section, we show how our framework can be used for determining whether a pure strategy profile is a Nash equilibrium or not. Let $G = (Ag, Ac, Av, Ou, Ef, \leqslant)$ be a game, and $\mathcal{AS}_G = (\mathbb{A}, \mathbb{C}, \mathbb{D})$ the corresponding AS. We consider a dialogue between two agents (the proponent P and the opponent O). The proponent's goal is to show that an argument A is a Nash Equilibrium and the opponent seeks to demonstrate that the proponent's game argument (A) is not a Nash equilibrium by proposing an alternative game-based argument (B) such that there is a player $i \in Ag$ for which $A_{-i} = B_{-i}$ and $A \neq B$ and for whom B yields a better outcome than A.

We now demonstrate the sequence of utterances dialogue participants should use to ensure that the proponent will win the dialogue if and only if A is a Nash equilibrium. However, argument B advanced by the opponent may not be a Nash Equilibrium. Therefore, multiple rounds of the dialogue may be required to identify such equilibria.

The dialogue consists of agents advancing locutions which refer to arguments, valuations and players. While a dialogue without locutions can be defined, we believe that such locutions aid the explanatory process without introducing additional complexity, and that the locutions' intuitive meaning is clear. We therefore do not provide a formal account of these locutions. There can be three possible scenarios for the dialogue:

(1) B is strictly better than A for an agent i, i.e. $A <_i B$. By construction, there will be two preference arguments A' and B' such that A' attacks $(B, A) \in \mathbb{D}$ and B' attacks $(A, B) \in \mathbb{D}$ respectively. Since B is strictly better than A for an agent i, there will be a valuation argument $V = (A_{-i}, s_i > s_i')$, where $A = A_{-i} \oplus s_i'$ and $B = B_{-i} \oplus s_i$, such that V attacks $(A', B') \in \mathbb{D}$. This line of reasoning is then captured by the dialogue shown in Table 5.

Table 5

The dialogue for Scenario 1

$P:$	$claim(A)$	Claim that A is a NE
$O:$	$alt(B, A, i)$	B is strictly better than A for player i
$P:$	$eq(B', A', i)$	The presence of A' and B' mean that A and B are of equal utility to player i
$O:$	$assert(V, A' \rightarrow B', i)$	The valuation argument V shows that B is strictly preferred to A as V attacks $A' \rightarrow B'$ for player i
$P:$	$concede(A)$	Concede that A is not a NE

(2) B is strictly worse than A for an agent i, i.e. $B <_i A$. By construction, there will be two preference arguments A' and B' such that A' attacks $(B, A) \in \mathbb{D}$ and B' attacks $(A, B) \in \mathbb{D}$ respectively. Since A is strictly better than B for an agent i, there will be a valuation argument $V = (A_{-i}, s_i > s_i')$, where $A = A_{-i} \oplus s_i$ and $B = B_{-i} \oplus s_i'$, such that V attacks $(B', A') \in \mathbb{D}$. This line of reasoning is then captured by the dialogue shown in Table 6.

(3) B is equivalent to A for an agent i, i.e. $B \leqslant_i A$ and $A \leqslant_i B$. By construction, there will be two preference arguments A' and B' such that A' attacks $(B, A) \in \mathbb{D}$ and B' attacks $(A, B) \in \mathbb{D}$ respectively. The attacks $(B', A'), (A', B') \in \mathbb{C}$ are not attacked. This line of reasoning is then captured by the dialogue shown in Table 7.

If the resultant dialogue evolves as per Scenario 1, then the proponent's game argument is not a Nash Equilibrium.

Table 6

The dialogue for Scenario 2

P : *claim*(A)	Claim that A is a NE
O : *alt*(B, A, i)	B is strictly better than A for player i
P : *assert*$(V, B' \to A', i)$	The valuation argument V shows that A is strictly preferred to B as V attacks $B' \to A'$ for player i
O : *concede*(B)	Concede that B is strictly worse than A for player i

Table 7

The dialogue for Scenario 3

P : *claim*(A)	Claim that A is a NE
O : *alt*(B, A, i)	B is strictly better than A for player i
P : *eq*(B', A', i)	The presence of A' and B' mean that A and B are of equal utility to player i
O : *concede*(B)	Concede that B is not strictly better than A for player i

6. Discussion, Related and Future Work

In this paper, we described how normal form games can be given an argumentation-based interpretation so as to allow – via argumentation semantics – for pure Nash equilibria to be computed. Intuitively, a Nash equilibrium identifies the best strategy a player can pursue given others' strategies. However, explaining – to a non-expert – why some set of strategies forms a Nash equilibrium is often difficult, and our argument-based interpretation is the first step towards an explanatory dialogue for such explanation. Other work has shown the utility of providing such dialogue-based explanations [10–12].

Our approach is based on extended argumentation frameworks, and Modgil [9] has proposed a proof dialogue for such frameworks. The dialogue presented in Section 5 is tailored for our framework and more specialised than Modgil's proof dialogue, but (we believe) provides a better explanation. In addition, while Modgil's dialogue specifies legal moves, it does not identify what arguments should be advanced by a dialogue participant, noting only that there exists a winning strategy to demonstrate that an argument is in the credulous preferred semantics. In contrast, our (simple) dialogue amalgamates both the legal moves that a player can make and the strategy that they must follow. This is best illustrated in Table 8, which shows two possible dialogues of the stag hunt game (shown in Table 1 and Figure 1) from Modgil's system. The left hand dialogue is analogous to Scenario 2 of our approach (cf. Section 5), but contains only the arguments themselves without explaining why they exist or attack other arguments (unlike our approach). The dialogue on the right demonstrates a non-winning but legal strategy in Modgil's system, which has no explanatory power.

Examining Tables 5-7, we note that the losing player will make a last `concede` move in all cases. This is similar to [4]'s proof dialogue where the winning player makes the last move. Furthermore, Tables 5-7 capture all possible evolutions of our explanatory dialogue.

If A is a Nash Equilibrium, then there is no dialogue whose first move by P is *claim*(A) and finishes with P conceding. Thus, P will win the dialogue and show that A is a Nash Equilibrium. Similarly, if A is not a Nash Equilibrium, then there is a dialogue whose first move by P is *claim*(A) and finishes by P conceding. Thus, P will lose the dialogue under perfect play. Therefore, our dialogue will identify whether a game argument is, or is not a NE. By running the dialogue over every game argument A, we are able to determine whether it is a NE. In other words, our dialogue is sound and complete. We note

Table 8

In the left dialogue, the proponent is demonstrating that argument a_2 is a Nash Equilibrium. In the right dialogue, both agents advance Nash equilibria.

P:	a_2
O:	a_1
P:	a_5
O:	a_6
P:	a_{13}

P:	a_2
O:	a_4

that the dialogue game of [4] is also sound and complete, making them — in some sense — equivalent in this context.

In the short term, we intend to empirically evaluate the explanatory capability of our dialogue with human subjects. Other extensions which we intend to investigate include providing an argumentation semantics for mixed Nash equilibria (perhaps through the use of some form of ranking semantics [13–15]), and investigating other solution concepts (e.g., Pareto optimality) for more complex types of games. Finally, there are clear links between game theory and group-based practical reasoning. Building on work such as [16, 17], we intend to investigate how an argument-based formulation to practical reasoning underpinned by game theory can be created.

In this work, we introduced three levels of argument to compute the Nash equilibria. An obvious alternative formulation would use a single level, where joint strategy profiles are arguments (equivalent to game-based arguments), and attacks are constructed based on the algorithm for computing equilibria. While this approach would yield similar results, it provides no explanation as to *why* the attacks appear (and therefore why something is a Nash equilibrium).In our formulation, we have arguments about the object level (i.e., game arguments), as well as arguments about preferences over these objects, which are themselves reasoned about. Modgil [4] demonstrates that the standard way of reasoning about such structures is through the use of meta-level argumentation, instantiated as an extended argumentation framework. By making use of this multi-level approach, we have shown how our dialogues can exploit this structure to provide explanation.

Several other authors have investigated some links between game theory and argumentation. For example, in his seminal paper, Dung [18] noted that the stable extension corresponds to the stable solution of an cooperative $n-$person game, but did not seem to deal with non-cooperative games as we do here. Game theory was also used to describe argument strength by Matt and Toni [15], and Rahwan and Larson [19] investigated the links between argumentation and game theory from a mechanism design point of view. Perhaps most closely related to the current work is Fan and Toni's work [2] exploring the links between dialogue and assumption-based argumentation (ABA). Here, the authors showed how admissible sets of arguments obtained from their ABA constructs are equivalent to Nash equilibria. In contrast to the current work, they only considered two player games and utilised structured argumentation, allowing them to describe a proof dialogue with associated strategies.

7. Conclusions

In this paper, we provided an argumentation-based interpretation of pure strategies in normal form games, demonstrating how argumentation semantics can be aligned with the Nash equilibrium as a solution concept, and examining some of the argumentation system's properties. We also formalised dialogues for our framework, highlighting how it can be used for real-word explanations of Nash Equilibria to non-experts.

We believe that this work has significant application potential in the context of argument-based explanation. At the same time, we recognise that there are significant open avenues for research in this area, but believe that the current work is an important step in investigating the linkages between the two domains.

References

[1] A. Matsumoto and F. Szidarovszky, *Game Theory and Its Applications*, Springer Japan, 2016. ISBN 978-4-431-54785-3. doi:10.1007/978-4-431-54786-0.

[2] X. Fan and F. Toni, On the Interplay between Games, Argumentation and Dialogues, in: *Proc. AAMAS-16*, 2016, pp. 260–268. ISBN 978-1-4503-4239-1.

[3] M. Osborne, *Introduction to Game Theory: International Edition*, OUP, 2009.

[4] S. Modgil, Reasoning about Preferences in Argumentation Frameworks, *Artificial Intelligence* **173**(9–10) (2009), 901–934. doi:10.1016/j.artint.2009.02.001.

[5] G. Boella, D.M. Gabbay, L.W.N. van der Torre and S. Villata, Meta-Argumentation Modelling I: Methodology and Techniques, *Stud Logica* **93**(2–3) (2009), 297–355.

[6] S. Modgil and T.J.M. Bench-Capon, Integrating Object and Meta-Level Value Based Argumentation, in: *Computational Models of Argument: Proceedings of COMMA 2008, Toulouse, France, May 28-30, 2008*, P. Besnard, S. Doutre and A. Hunter, eds, Frontiers in Artificial Intelligence and Applications, Vol. 172, IOS Press, 2008, pp. 240–251. http://www.booksonline.iospress.nl/Content/View.aspx?piid=9284.

[7] S. Modgil and T.J.M. Bench-Capon, Metalevel argumentation, *J. Log. Comput.* **21**(6) (2011), 959–1003. doi:10.1093/logcom/exq054.

[8] O. Rodrigues, E. Black, M. Luck and J. Murphy, On Structural Properties of Argumentation Frameworks: Lessons from ICCMA., in: *Proceedings of the Second International Workshop on Systems and Algorithms for Formal Argumentation (SAFA 2018) co-located with the 7th International Conference on Computational Models of Argument (COMMA 2018), Warsaw, Poland, September 11, 2018.*, 2018, pp. 22–35. http://ceur-ws.org/Vol-2171/paper_3.pdf.

[9] S. Modgil, Labellings and Games for Extended Argumentation Frameworks, in: *Proc. IJCAI-09*, 2009, pp. 873–878.

[10] M. Caminada, R. Kutlák, N. Oren and W.W. Vasconcelos, Scrutable Plan Enactment via Argumentation and Natural Language Generation, in: *AAMAS*, 2014.

[11] C. Kristijonas, S. Ken and T. Francesca, Explanation for Case-Based Reasoning via Abstract Argumentation, *Frontiers in Artificial Intelligence and Applications* (2016), 243–254. doi:10.3233/978-1-61499-686-6-243.

[12] N. Oren, K. van Deemter and W.W. Vasconcelos, Argument-Based Plan Explanation, in: *Knowledge Engineering Tools and Techniques for AI Planning*, M. Vallati and D. Kitchin, eds, Springer International Publishing, 2020, pp. 173–188. ISBN 978-3-030-38561-3.

[13] L. Amgoud, J. Ben-Naim, D. Doder and S. Vesic, Ranking Arguments With Compensation-Based Semantics, in: *KR*, 2016.

[14] E. Bonzon, J. Delobelle, S. Konieczny and N. Maudet, A Comparative Study of Ranking-Based Semantics for Abstract Argumentation, in: *Proc. AAAI-16*, 2016, pp. 914–920.

[15] P.-A. Matt and F. Toni, A Game-Theoretic Measure of Argument Strength for Abstract Argumentation, in: *Logics in Artificial Intelligence*, LNCS, 2008, pp. 285–297. ISBN 978-3-540-87803-2.

[16] K. Atkinson and T.J.M. Bench-Capon, Argument Schemes for Reasoning About the Actions of Others, in: *Proc. COMMA*, Vol. 287, 2016, pp. 71–82.

[17] Z. Shams, M.D. Vos, N. Oren and J. Padget, Argumentation-Based Reasoning about Plans, Maintenance Goals, and Norms, *ACM Trans. Auton. Adapt. Syst.* **14**(3) (2020). doi:10.1145/3364220.

[18] P.M. Dung, On the Acceptability of Arguments and Its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games, *Artificial Intelligence* **77**(2) (1995), 321–357. doi:10.1016/0004-3702(94)00041-X.

[19] I. Rahwan and K. Larson, Argumentation and Game Theory, in: *Argumentation in Artificial Intelligence*, G. Simari and I. Rahwan, eds, Springer US, Boston, MA, 2009, pp. 321–339. ISBN 978-0-387-98197-0.