

Applying Abstract Argumentation Theory to Cooperative Game Theory

Anthony P. Young^[0000-0003-0747-3866], David Kohan Marzagão^[0000-0001-8475-7913], and Josh Murphy

Department of Informatics, King’s College London,
Bush House, Strand Campus, 30 Aldwych, WC2B 4BG
{peter.young, david.kohan, josh.murphy}@kcl.ac.uk

Abstract. We apply ideas from abstract argumentation theory to study cooperative game theory. Building on Dung’s results in his seminal paper, we further the correspondence between Dung’s four argumentation semantics and solution concepts in cooperative game theory by showing that complete extensions (the grounded extension) correspond to Roth’s subsolutions (respectively, the supercore). We then investigate the relationship between well-founded argumentation frameworks and convex games, where in each case the semantics (respectively, solution concepts) coincide; we prove that three-player convex games do not in general have well-founded argumentation frameworks.

Keywords: Abstract argumentation theory · argumentation semantics · cooperative game theory · solution concepts · convex games

1 Introduction

Argumentation theory is the branch of artificial intelligence (AI) that is concerned with the rational and transparent resolution of disagreements between arguments (e.g. [18]). *Abstract* argumentation theory, as articulated in Dung’s seminal paper [7], abstracts away from the contents of the arguments and the nature of their disagreements. The resulting directed graph (digraph) representation of arguments (nodes) and their disagreements (directed edges), called an *abstract argumentation framework* (AF), is simple yet powerful enough to resolve these disagreements and determine the sets of winning arguments.

Dung demonstrated the “correctness” of abstract argumentation by showing how abstract argumentation “can be used to investigate the logical structure of the solutions to many practical problems” [7, Section 3]. Specifically, he investigated two examples of problems from microeconomics (e.g. [14]): cooperative game theory and matching theory. In each case, Dung showed how an appropriate AF can represent a given cooperative game or a given instance of the stable marriage problem, and that the sets of winning arguments in such AFs correspond to meaningful solutions in both of these domains.

In this paper, we further demonstrate the “correctness” of abstract argumentation theory by investigating its relationship with cooperative game theory. *Cooperative game theory* (e.g. [6]) is the branch of game theory (e.g. [27])

that studies how normatively rational agents may cooperate in order to possibly earn more payoff than they would do as individuals. Cooperative game theory abstracts away from individual agents' strategies in such games for a simpler and "high level" view of their interactions.¹ Each *cooperative game* consists of finitely many agents that can cooperate as *coalitions*, and each coalition earns some payoff as measured by its *value*. Given certain standard assumptions on the values of coalitions, all agents should cooperate as a single coalition, called the *grand coalition*. However, this still leaves open the question of how the payoff obtained by the grand coalition (which is already maximised) should be distributed among the individual agents, such that no agent should want to defect from the grand coalition. Historically, the first *solution concept* for cooperative games that captures this idea is the *Von Neumann-Morgenstern (vNM) stable set*, where each such set consists of such payoff distributions interpreted as an "acceptable standard of behaviour" [27]. Subsequently, Dung showed that each possible payoff distribution can be interpreted as an argument in an AF. Payoff distributions "disagree" when agents can defect because they can earn strictly more. This argumentative interpretation of cooperative games allowed Dung to demonstrate that the stable extensions of the AF of each cooperative game correspond exactly to the game's vNM stable sets [7, Theorem 37].

However, just like that stable extensions of an AF may not exist, vNM stable sets for cooperative games also may not exist [11,12]. As a result, alternative solution concepts have been proposed. For example, Dung proposed that sets of payoff distributions that form preferred extensions could serve as an alternative solution concept, because preferred extensions of AFs always exist, and therefore this is well-defined for all cooperative games. Other possible alternative solution concepts from cooperative game theory include the *core* [8], the *subsolution* and the *supercore* [21]. Dung showed that the core corresponds to the set of unattacked arguments of the game's AF [7, Theorem 38]. This paper's first contribution is to finish the correspondence between Dung's four argumentation semantics and the various solution concepts from cooperative game theory by proving that the complete extensions (respectively, the grounded extension) of the AF correspond(s) to the subsolutions (respectively, the supercore) of the cooperative game. These correspondences allow us to characterise when the supercore of a cooperative game is non-empty via the *Bondareva-Shapley theorem* [3,24]: exactly when the game is *balanced* (see Section 3.2).

Shapley investigated a special class of cooperative games called *convex games*, which capture the intuition that agents have more incentive to join larger coalitions; the key property of each such game is that its core is its unique stable set [25, Theorem 8]. In abstract argumentation theory, a similar result by Dung states that if an AF is *well-founded*, its grounded extension is its unique stable set [7, Theorem 30]. Given these similar results, we would like to know whether convex games always give rise to well-founded AFs. Our second contribution in

¹ The nature of this cooperation is exogenous to the theory, but can be interpreted as groups of agents forming binding contracts. See, e.g. [6, page 7].

this paper is a negative answer to this question through a three-player counter-example.

The rest of this paper is structured as follows. In Section 2, we review abstract argumentation theory and cooperative game theory. In Section 3, we complete the correspondences between Dung’s four argumentation semantics with solution concepts in cooperative games and study the properties using well-known results from argumentation. In Section 4, we recap convex games and well-founded AFs, and give a counter-example of a three-player convex game that gives rise to a non-well-founded AF. In Section 5, we compare our results with related work in argumentation theory and game theory, and conclude with future work.

2 Background

Notation: Let X, Y, Z be sets. $\mathcal{P}(X)$ is the *power set of X* and $|X|$ is the *cardinality of X* . \mathbb{N} is *the set of natural numbers* (including 0) and \mathbb{R} is *the set of real numbers*. Further, \mathbb{R}^+ (respectively, \mathbb{R}_0^+) is *the set of positive (respectively, non-negative) real numbers*. For $n \in \mathbb{N}$, the *n -fold Cartesian power of X* is X^n . If $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ are appropriate functions, then $g \circ f : X \rightarrow Z$ is the *composition* of g then f , and functional composition is denoted with \circ . If X is a set, then for a function $f : X \rightarrow \mathbb{R}$, $f \geq 0$ abbreviates $(\forall x \in X) f(x) \geq 0$.

2.1 Abstract Argumentation Theory

An **(abstract) argumentation framework**² (AF) is a directed graph (di-graph) $\langle A, R \rangle$ where A is the **set of arguments** and $R \subseteq A^2$ denotes the **attack relation**, where for $a, b \in A$, $(a, b) \in R$, abbreviated by $R(a, b)$, denotes that a disagrees with b . Let $S \subseteq A$ be a set of arguments for the remainder of this subsection. Define $S^+ := \{b \in A \mid (\exists a \in S) R(a, b)\}$ to be the set of all arguments attacked by S . The **neutrality function** is $n : \mathcal{P}(A) \rightarrow \mathcal{P}(A)$ where $n(S) := A - S^+$ denotes the set of arguments *not* attacked by S , i.e. the set of arguments that S is *neutral* towards. We say S is **conflict-free** (cf) iff $S \subseteq n(S)$. Similar to S^+ , let $S^- := \{b \in A \mid (\exists a \in S) R(b, a)\}$, and for $a \in A$, $a^\pm := \{a\}^\pm$. The **defence function** is $d : \mathcal{P}(A) \rightarrow \mathcal{P}(A)$ where $a \in d(S)$ iff $a^- \subseteq S^+$.³ It can be shown that $d = n^2 := n \circ n$ [7, Lemma 45]. Let $U \subseteq A$ denote the **set of unattacked arguments**. It can be shown that $U = d(\emptyset)$. We say S is **self-defending** (sd) iff $S \subseteq d(S)$. Further, S is **admissible** iff it is cf and sd.

To determine which sets of arguments are justifiable we say S is a **complete extension** iff S is admissible and $d(S) \subseteq S$.⁴ Further, S is a **preferred extension** iff it is a \subseteq -maximal complete extension [28, Theorem 7.11], S is a **stable extension** iff $n(S) = S$ and S is the **grounded extension** iff it is

² When defining our terms in this paper, any words in between brackets may be omitted when using the term, e.g. in this case, the terms “argumentation framework” and “abstract argumentation framework” are interchangeable.

³ In [7, Section 2.2] this is called the **characteristic function**.

⁴ i.e. if one can defend a proposition then one is obliged to accept it as justifiable.

the \subseteq -least complete extension. These four semantics are collectively called the **Dung semantics**, and each defines sets of winning arguments given $\langle A, R \rangle$.

2.2 Cooperative Game Theory

We review the basics of cooperative game theory (see, e.g. [6]). Let $m \in \mathbb{N}$ and $N = \{1, 2, 3, \dots, m\}$ be our **set of players** or **agents**.⁵ We assume $m \geq 1$. A **coalition** C is any subset of N . The **empty coalition** is \emptyset and the **grand coalition** is N .

Example 1. Consider the set of players $N = \{1, 2, 3\}$, where $m = 3$. In our examples, agents 1, 2 and 3 are respectively named Josh, David and Peter. If all three decide to work together, then they form the grand coalition N . If only Josh and David work together and Peter works alone, then the resulting coalitions are, respectively, $\{1, 2\}$ and $\{3\}$.

A **valuation function** is a function $v : \mathcal{P}(N) \rightarrow \mathbb{R}$ such that $v(\emptyset) = 0$. The number $v(C)$ can be thought of as the coalition's payoff as a result of the agents' coordination of strategies; this payoff is in arbitrary units.

Example 2. (Example 1 continued) If Josh, David and Peter work together and earn 10 units of payoff, then $v(N) = 10$. If Josh and Peter will lose 50 units of payoff if they work together, then $v(\{1, 3\}) = -50$.

Given N and v , with $m := |N|$, a **(cooperative) (m -player) game (in normal form)** is the pair $G := \langle N, v \rangle$. The following five properties are standard for v . We say v is **non-negative** iff $v \geq 0$; this excludes valuation functions such as the one in Example 2. We say v is **monotonic** iff for all $C, C' \subseteq N$, if $C \subseteq C'$, then $v(C) \leq v(C')$. We say v is **constant-sum** iff $(\forall C \subseteq N) v(C) + v(N - C) = v(N)$. We say v is **super-additive** iff for all $C, C' \subseteq N$, if C and C' are disjoint then $v(C \cup C') \geq v(C) + v(C')$. We say v is **inessential** iff $\sum_{k=1}^m v(\{k\}) = v(N)$; inessential means there is no incentive to cooperate.

It is easy to show that if v is non-negative and super-additive, then v is monotonic, while the converse is not true (e.g. [5, Example 1]). For the rest of the games in this paper, we will assume v is non-negative, super-additive and **essential** (i.e. not inessential).⁶

Example 3. (Example 1 continued) Suppose if Josh, David or Peter earn no payoff if they work as individuals, but if any two of them work together they earn 10 units of payoff, and if all three work together they earn 20 units of payoff. This v is non-negative, super-additive, not constant-sum, and essential.

⁵ We use “ m ” instead of the more traditional “ n ” for the number of players to avoid confusion with the neutrality function $n : \mathcal{P}(A) \rightarrow \mathcal{P}(A)$ in argumentation.

⁶ When combined with super-additivity, it follows that there are two disjoint coalitions C and C' such that $v(C \cup C') > v(C) + v(C')$.

An **outcome** of a game is the pair (CS, \mathbf{x}) , where CS is a partition of N called a **coalition structure**, and $\mathbf{x} \in \mathbb{R}^m$ is a **payoff vector** that distributes the value of each coalition to the players in that coalition. As we have assumed that v is non-negative, super-additive and essential, then $v(N)$ has the (strictly) largest payoff among all coalitions by monotonicity. Agents are rational and want to maximise their payoff, and so they should seek to form the grand coalition. Therefore, we restrict our attention to outcomes where $CS = \{N\}$.

How should the amount $v(N)$ be distributed among the m players? In this paper, we consider **transferable utility games**, which allow for the distribution of $v(N)$ arbitrarily to the m players, e.g. by interpreting $v(N)$ as money, which all players should desire. This leads to the following properties of payoff vectors \mathbf{x} . We say $\mathbf{x} := (x_1, x_2, \dots, x_m)$ is **feasible** iff $\sum_{k \in N} x_k \leq v(N)$, **efficient** iff $\sum_{k \in N} x_k = v(N)$ and **individually rational** iff $(\forall k \in N) v(\{k\}) \leq x_k$. We call a payoff vector \mathbf{x} an **imputation** iff \mathbf{x} is efficient and individually rational; intuitively, imputations distribute all the money to every agent without waste, such that every agent earns at least as much as when they work alone. Following [7], we denote the **set of imputations for a game G** with $IMP(G)$, or just IMP if it is clear which cooperative game G we are referring to.

Example 4. (Example 3 continued) As $v(N) = 20$ (which we now measure in dollars (\$), as an example of transferable utility), a feasible payoff vector is $(5, 5, 5)$. An efficient payoff vector is $(10, 5, 5)$. Notice that both vectors are individually rational because $(\forall k \in N) v(\{k\}) = 0$. Therefore, $(10, 5, 5)$ is a valid imputation, in which case Josh receives \$10, while David and Peter receive \$5 each.

The solution concepts of cooperative games that we will consider are concerned with whether coalitions of agents are incentivised to defect from the grand coalition.⁷ Given a game $G = \langle N, v \rangle$, let C be a coalition and $\mathbf{x}, \mathbf{y} \in IMP$. We say \mathbf{x} **dominates** \mathbf{y} **via** C , denoted $\mathbf{x} \rightarrow_C \mathbf{y}$, iff (1) $(\forall k \in C) x_k > y_k$ and (2) $\sum_{k \in C} x_k \leq v(C)$. Intuitively, given imputation \mathbf{y} , it is possible for a subset of players to defect from N to form their own coalition C , where (1) they will each do strictly better because (2) they will earn enough to do so; the resulting payoff is \mathbf{x} . Note that for all $C \subseteq N$, \rightarrow_C is a well-defined binary relation on IMP .

Example 5. (Example 4 continued) Consider the two imputations $(20, 0, 0)$ and $(12, 4, 4)$. Clearly, $(12, 4, 4) \rightarrow_{\{2,3\}} (20, 0, 0)$, because David (agent 2) and Peter (agent 3) can defect to $\{2, 3\}$ and earn $4 + 4 = 8 \leq v(\{2, 3\}) = 10$, which is strictly better than both of them getting nothing in the imputation $(20, 0, 0)$.

It is easy to see from the definition of \rightarrow_C that it is irreflexive, acyclic (and hence asymmetric), and transitive. Further, some important special cases include $\rightarrow_N = \emptyset$ and $(\forall k \in N) \rightarrow_{\{k\}} = \emptyset$, i.e. it does not make sense for the grand coalition to defect to the grand coalition, and individual players cannot defect from N (the latter due to individual rationality). Also, $\rightarrow_\emptyset = IMP^2$, the total binary relation on IMP . We say imputation \mathbf{x} **dominates** imputation \mathbf{y} , denoted

⁷ See Section 5 for a brief mention of other solution concepts.

$\mathbf{x} \rightarrow \mathbf{y}$, iff $(\exists C \subseteq N) [C \neq \emptyset, \mathbf{x} \rightarrow_C \mathbf{y}]$. This is a well-defined irreflexive binary relation on IMP . However, it can be shown that this is not generally transitive, complete or acyclic (e.g. [26, Chapter 4]). Therefore, each cooperative game gives rise to an associated directed graph $\langle IMP, \rightarrow \rangle$, called an **abstract game** [21].

Let $I \subseteq IMP$ be a set of imputations. Following [27], we say I is **internally stable** iff no imputation in I dominates another in I . Further, I is **externally stable** iff every imputation not belonging to I is dominated by an imputation from I . A **(von Neumann-Morgenstern) stable set** is a subset of IMP that is both internally and externally stable. Taken together, a stable set of imputations contains the distributions of the amount of money $v(N)$ to the set of m players that are socially acceptable [27].

We now recapitulate a simplification to coalitional games that does not lose generality (e.g. [26, Chapter 4]). Let $G := \langle N, v \rangle$ and $G' := \langle N, v' \rangle$ be two games on the same set of players. We say G and G' are **strategically equivalent** iff $(\exists K \in \mathbb{R}^+) (\exists \mathbf{c} \in \mathbb{R}^m) (\forall C \subseteq N) v'(C) = Kv(C) + \sum_{k \in C} c_k$, for $\mathbf{c} := (c_1, \dots, c_m)$, and we denote this with $G \cong G'$; this is an equivalence relation between games. Further, the function $f : IMP(G) \rightarrow IMP(G')$ with rule $f(\mathbf{x}) := K\mathbf{x} + \mathbf{c}$ is a digraph isomorphism from $\langle IMP(G), \rightarrow \rangle$ to $\langle IMP(G'), \rightarrow' \rangle$, where \rightarrow' is the corresponding domination relation on $IMP(G')$. It follows that if $S \subseteq IMP(G)$ is a stable set of G , then its image set $f(S) \subseteq IMP(G')$ is also a stable set of G' . By setting $\frac{1}{K} = v(N) - \sum_{k \in N} v(\{k\})$ and $(\forall k \in N) c_k = -Kv(\{k\})$ for this K , then we can transform $G = \langle N, v \rangle$ to its **(0, 1)-normalised form**, $G^{(0,1)} := \langle N, v^{(0,1)} \rangle$, such that $(\forall k \in N) v(\{k\}) = 0$ and $v(N) = 1$.

Example 6. (Example 5 continued) We have that $\mathbf{c} = \mathbf{0}$ and $K = \frac{1}{20}$, therefore $v^{(0,1)}(C) = 0$ if $|C| \leq 1$, $v^{(0,1)}(C) = \frac{1}{2}$ if $|C| = 2$, and $v^{(0,1)}(N) = 1$.

This means $IMP(G^{(0,1)}) \subseteq \mathbb{R}^m$ is the standard $(m - 1)$ -dimensional topological simplex, which has the set $\{(x_1, x_2, \dots, x_m) \in (\mathbb{R}_0^+)^m \mid \sum_{k=1}^m x_k = 1\}$.

Corollary 1. *As a set, $IMP(G^{(0,1)})$ is uncountably infinite.*

Proof. Let \cong denote bijection between sets and \hookrightarrow denote an injective embedding between sets, then we have $\mathbb{R} \cong [0, 1] \hookrightarrow IMP(G^{(0,1)}) \subseteq \mathbb{R}^m \cong \mathbb{R}$, where \hookrightarrow in this case is the injective function with rule $t \mapsto K(1 - t, t, x_3, \dots, x_m)$, where $\frac{1}{K} := 1 + \sum_{k=3}^m x_k$ normalises the output to be on the simplex. By the Cantor-Schröder-Bernstein theorem [10, Theorem 3.2], $IMP(G^{(0,1)}) \cong \mathbb{R}$.

Therefore, any abstract game $\langle IMP, \rightarrow \rangle$ has uncountably infinitely many nodes. From now, we assume all games G have v that are non-negative, super-additive, not necessarily constant-sum, have transferable utility, and are $(0, 1)$ -normalised.

2.3 From Cooperative Game Theory to Abstract Argumentation

In this section we recap [7, Section 3.1]. We now understand how an abstract game $\langle IMP, \rightarrow \rangle$, which is also an uncountably infinite digraph, arises from a

game G . Dung interprets $\langle \text{IMP}, \rightarrow \rangle$ as an abstract AF, where each argument is an argument for a given payoff distribution among the agents, and each attack denotes the possibility for a subset of agents to defect from the grand coalition. Corollary 1 states that such an AF has uncountably infinite arguments, but this is not a problem because Dung's argumentation semantics and their properties hold for AFs of arbitrary cardinalities [2,7,28]. Dung then proved that various methods of resolving conflicts in $\langle \text{IMP}, \rightarrow \rangle$ as an AF correspond to meaningful solution concepts of G . The following result is straightforward to show.

Theorem 1. [7, Theorem 37] *Let G be a game and $\langle \text{IMP}, \rightarrow \rangle$ be its abstract game. If we view $\langle \text{IMP}, \rightarrow \rangle$ as an AF, then each of its stable extensions is a stable set of G , and each stable set of G is a stable extension of $\langle \text{IMP}, \rightarrow \rangle$.*

As stable sets may not exist for AFs, and in particular there are games without stable sets [11,12], Dung proposed that preferred extensions, as they always exist [7, Theorem 11(2)], can serve as an alternative solution concept for a game G in cases where stable sets do not exist, because the properties and motivations of preferred extensions also capture the imputations that are rational wealth distributions among the m players [7, Section 3.1].

Further, another important solution concept in cooperative game theory is the core [8]. Formally, the **core** of a cooperative game G is the set of imputations \mathbf{x} satisfying the system of inequalities $(\forall C \subseteq N) \sum_{k \in C} x_k \geq v(C)$. Intuitively, the core is the set of imputations where each agent is receiving at least as much payoff even if a subset of such agents were to defect to a new coalition (regardless of how the payoff is shared within that coalition). Therefore, no agent has an incentive to defect. It can be shown that the core is the subset of imputations that are not dominated by any other imputation. It follows that:

Theorem 2. [7, Theorem 38] *Let G be a game and $\langle \text{IMP}, \rightarrow \rangle$ be its abstract game. If we view $\langle \text{IMP}, \rightarrow \rangle$ as an AF, then its set of unattacked arguments corresponds exactly to the core.*

From argumentation theory, we thus conclude the well-known result from cooperative game theory that stable sets, if they exist, always contain the core.

Example 7. (Example 6 continued) We have $N = \{1, 2, 3\}$ and $v(C) = 0$ if $|C| \leq 1$ else $v(C) = \frac{1}{2}$, and $v(N) = 1$. The eight possible coalitions $C \subseteq N$ give rise to the three inequalities $x_1 + x_2 \geq \frac{1}{2}$, $x_2 + x_3 \geq \frac{1}{2}$ and $x_3 + x_1 \geq \frac{1}{2}$. Therefore, the core consists of all imputations $\mathbf{x} = (x_1, x_2, x_3)$ whose components satisfy these three inequalities, for example, $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $(\frac{1}{2}, \frac{1}{2}, 0)$.

3 Complete Extensions and the Grounded Extension

Given that stable extensions correspond to stable sets, and the set of unattacked arguments corresponds to the core, do the complete extensions (including the preferred extensions) and the grounded extension also correspond to solution concepts in cooperative games? We now show that the answer is yes.

3.1 Complete Extensions Correspond to Subsolutions

As preferred extensions are a subset of complete extensions, it is natural to ask whether complete extensions more generally correspond to solution concepts in cooperative games. In [21], motivated by the general lack of existence of stable sets [11,12], Roth considered abstract games arising from cooperative games (in his notation) $\langle X, > \rangle$, where X is the set of the game's **outcomes** and $> \subseteq X^2$ is an abstract **domination** relation. Let $u : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ be the function $u(S) := X - S^+$ ⁸, where in this case $S^+ := \{y \in X \mid (\exists x \in S) x > y\}$. Roth then defined the *subsolution* of such an abstract game as follows.

Definition 1. [21, Section 2] Let $\langle X, > \rangle$ be an abstract game. A **subsolution** is a set $S \subseteq X$ such that $S \subseteq u(S)$ and $S = u^2(S) := u \circ u(S)$.

Immediately we can see that by interpreting $\langle X, > \rangle$ as an abstract argumentation framework, subsolutions are precisely the complete extensions.

Theorem 3. Let $G = \langle N, v \rangle$ be a game and $\langle IMP, \rightarrow \rangle$ be its abstract game. When seen as an argumentation framework, the complete extensions $\langle IMP, \rightarrow \rangle$ are precisely the subsolutions.

Proof. $S \subseteq IMP$ is a complete extension of $\langle IMP, \rightarrow \rangle$ iff $S \subseteq n(S)$ and $S = d(S)$, where n is the neutrality function $n(S) = IMP - S^+$, but as $n^2 = d$ [7, Lemma 45], this is equivalent to saying that S is a subsolution, by identifying $X = IMP$, $> = \rightarrow$ and $u = n$.

Roth has shown that every abstract game, and hence every cooperative game, has a subsolution [21, Theorem 1].⁹ The \subseteq -maximal subsolutions of $\langle IMP, \rightarrow \rangle$ are exactly the preferred extensions of $\langle IMP, \rightarrow \rangle$ when seen as an abstract argumentation framework. Roth also showed that stable sets are subsolutions, and hence subsolutions generalise stable sets. This is well known in argumentation theory as stable extensions are also complete extensions. We can also apply further results from abstract argumentation theory to infer more properties of subsolutions. For instance, the core is contained in all subsolutions.

Corollary 2. Every subsolution is a superset of the core.

Proof. Interpreting $\langle IMP, \rightarrow \rangle$ as an argumentation framework, the core corresponds to the set of unattacked arguments, which is $d(\emptyset)$, where d is the defence function (Section 2.1). As a subsolution S is a complete extension, we know that $d(S) = S$. As d is \subseteq -monotonic and $\emptyset \subseteq S$, we have $d(\emptyset) \subseteq S$.

Further, subsolutions have a specific lattice-theoretic structure:

Theorem 4. The family of subsolutions of an abstract game form a complete semilattice that is also directed-complete.

Proof. This follows from [7, Theorem 25(3)] and [28, Theorem 6.30], respectively.

We will give an example of subsolutions in Section 4 (Example 8).

⁸ In [21], Roth uses U for this function. Here, we use u to avoid confusion with the set of unattacked arguments in an AF (defined in Section 2.1).

⁹ Also, see the abstract lattice-theoretic proof in [20].

3.2 The Supercore is the Grounded Extension

In [21, Example 5.1], Roth showed that subsolutions are in general not unique for abstract games; the supercore is one “natural” way of selecting a unique subsolution.

Definition 2. [21, Section 3] *The supercore of an abstract game is the intersection of all its subsolutions.*

Immediately we can conclude the following.

Theorem 5. *The supercore of an abstract game is its grounded extension when viewed as an argumentation framework.*

Proof. This follows from e.g. [7, Theorem 25(5)], or [28, Corollary 6.8].

It follows that the supercore exists and is unique for all abstract games, and hence cooperative games. Further, the supercore is a special case of a subsolution because the grounded extension is complete. Also, the supercore contains the core, because the grounded extension contains all unattacked arguments, by Corollary 2. We will give an example of the supercore in Section 4 (Example 8).

Therefore, for arbitrary cooperative games, if stable sets exist, they can serve as the possible sets of recommended payoff distributions, i.e. the “acceptable standards of behaviour” [27]. If they do not exist, then we may use subsolutions instead. If we desire a unique subsolution, we can use the supercore.

However, Roth noted that the supercore is empty iff the core is empty. This corresponds to the well-known result in argumentation theory that the set of unattacked arguments is empty iff the grounded extension is empty (e.g. [28, Corollary 6.9]). We can therefore completely characterise cooperative games with non-empty supercores using the *Bondareva-Shapley theorem* [3,24]. Let $G = \langle N, v \rangle$ be a cooperative game. A function $f : \mathcal{P}(N) \rightarrow [0, 1]$ is **balanced** iff $(\forall k \in N) \sum_{S \subseteq N - \{k\}} f(S \cup \{k\}) = 1$, i.e. if for every player the value under f of all coalitions containing that player sum to one. We call G **balanced** iff for every balanced function f , $\sum_{S \subseteq N} f(S) v(S) \leq v(N)$. Intuitively, each player k allocates a fraction of his or her time $f(S \cup \{k\})$ to the coalition $v(S)$, and that coalition receives a value proportional to that agent’s time spent there.

Theorem 6. (*Bondareva-Shapley*) *A game has a non-empty core iff it is balanced.*

It follows that:

Corollary 3. *A game has a non-empty supercore iff it is balanced.*

Proof. The core of a game is non-empty iff its supercore is non-empty, iff it is balanced, by the Bondareva-Shapley theorem (Theorem 6).

From both argumentation theory and abstract games in cooperative game theory, we have the following well-known containment relations between the solution concepts in cooperative games.

Theorem 7. *Let G be a cooperative game. Its stable sets are \subseteq -maximal subsolutions, which are subsolutions, and the supercore is also a subsolution.*

Proof. Immediate, because in argumentation theory, stable extensions are preferred, which are complete. Further, the grounded extension is also complete.

3.3 Summary

We summarise the correspondences between Dung’s argumentation semantics and the solution concepts of cooperative games in the Table 1, including the results presented in this paper.

Table 1. A Table Showing Concepts in Abstract Argumentation and the Corresponding Concept in Cooperative Games

Abstract Argumentation	Cooperative Game	Reference
Set of arguments A	Set of imputations IMP	[7, Section 3.1]
Attack relation R	Domination relation \rightarrow	[7, Section 3.1]
Argumentation Framework $\langle A, R \rangle$	Abstract Game $\langle IMP, \rightarrow \rangle$	[7, Section 3.1]
Unattacked arguments	The Core	[7, Thm. 38]
The Grounded Extension	The Supercore	Thm. 5
Complete Extensions	Subsolutions	Thm. 3
Preferred Extensions	\subseteq -maximal Subsolutions	[7, Section 3], Thm. 3
Stable Extensions	Stable Sets	[7, Thm. 37]

Further, we have shown that the supercore exists, and is unique and non-empty, iff the game is balanced (Corollary 3). Finally, the family of subsolutions form a complete semilattice that is also directed-complete (Theorem 4). These correspondences allow us to apply ideas from argumentation to cooperative game theory, as we will in the next section.

4 Convex three-player Cooperative Games and Well-Founded Argumentation Frameworks

Having shown how abstract argumentation can be used in cooperative game theory, we now investigate the relationship between well-founded AFs and convex games, because in both cases the semantics (respectively, solution concepts) coincide.

4.1 Convex Games and Coincidence of Solution Concepts

An important type of cooperative game is that of a convex game [25]. Formally, $G = \langle N, v \rangle$ is **convex** iff $(\forall C, C' \subseteq N) v(C \cup C') + v(C \cap C') \geq v(C) + v(C')$. Clearly, convex games are super-additive. This property is equivalent to [6,

Proposition 2.8]: for all $S, T \subseteq N$, if $T \subseteq S$, then $(\forall k \in N - S) v(T \cup \{k\}) - v(T) \leq v(S \cup \{k\}) - v(S)$. Intuitively, this means as a coalition grows in size, there is more incentive for agents not already in the coalition to join. Shapley calls this a “band-wagon” effect [25]. Further, if $G \cong G'$ and G is convex, then G' is also convex.¹⁰ The key property of convex games that we focus on is:

Theorem 8. [25, Theorem 8] *The core of a convex game is stable.*

From our results in Section 3, we can immediately show the following.

Corollary 4. *If G is convex, then the set of unattacked arguments of its AF $\langle \text{IMP}, \rightarrow \rangle$ is the only stable extension.*

Proof. Immediate from [7, Theorems 37 and 38].

Convex games exhibit a coincidence of the solution concepts so far considered.

Corollary 5. *If G is convex, then its core is also its supercore, which is also its unique subsolution.*

Proof. If G is convex, then viewing its abstract game $\langle \text{IMP}, \rightarrow \rangle$ as an AF, its set U of unattacked arguments is the unique stable extension [7, Theorem 37]. But if U is stable, then U is also the grounded extension - else the grounded extension, as a superset of U , is not conflict-free. Thus U is also the supercore of G (Theorem 5). As U is unique, it is the unique subsolution of G .

4.2 Well-Founded Argumentation Frameworks

Now recall from abstract argumentation theory we have a sufficient condition for an AF to have all four of Dung’s argumentation semantics coincide.

Theorem 9. [7, Theorem 30] *If $\langle A, R \rangle$ is well-founded, i.e. there is no A -sequence¹¹ $\{a^i\}_{i \in \mathbb{N}}$ such that $(\forall i \in \mathbb{N}) R(a^{i+1}, a^i)$ [7, Definition 29], then its grounded extension is the unique stable, complete and preferred extension.*

Given that the consequences of Theorem 9 and Corollaries 4 and 5 are the same, one may ask whether the AFs arising from convex games is in some sense “stronger” than well-founded AFs. Indeed, convex games refer to unattacked arguments U , while well-founded AFs refer to the grounded extension, a superset of U . Could it be that convex games always give rise to well-founded AFs? We answer this question in the following section.

¹⁰ This can be shown by writing out the definition of convex for the coalitions in G' given the definition of strategic equivalence in Section 2.2, and then applying the inclusion-exclusion principle for sets.

¹¹ Let X be a set, an X -sequence is a function $f : \mathbb{N} \rightarrow X$ written as $\{x^i\}_{i \in \mathbb{N}}$, so $(\forall i \in \mathbb{N}) x^i \in X$.

4.3 Three-Player Convex Games

To make this problem more tractable, we specialise to three-player convex games. We will assume the following canonical form without loss of generality:

Theorem 10. [13, Slide 19] *Every essential three-player game that is not necessarily constant-sum is strategically equivalent to the $(0,1)$ -normalised three-player game $\langle N, v^{(0,1)} \rangle$ where $v^{(0,1)}(C) = 0$ if $|C| \leq 1$, $v^{(0,1)}(N) = 1$, and for $a, b, c \in [0, 1]$, $v^{(0,1)}(\{1, 2\}) = a$, $v^{(0,1)}(\{2, 3\}) = b$ and $v^{(0,1)}(\{3, 1\}) = c$.*

Convexity constrains the parameters a, b and c as follows.

Corollary 6. *Every essential three-player convex game G that is not necessarily constant-sum is strategically equivalent to the $(0,1)$ -normalised game $\langle N, v^{(0,1)} \rangle$ defined in Theorem 10 iff $a + b \leq 1$, $b + c \leq 1$ and $c + a \leq 1$.*

Proof. (Sketch) (\Rightarrow) If G is convex and $G \cong \langle N, v^{(0,1)} \rangle$, then $\langle N, v^{(0,1)} \rangle$ is convex (Footnote 10). There are $2^3 = 8$ distinct coalitions, which we can use to write out the inequality in the definition of convex to conclude the resulting three inequalities on a, b and c . (\Leftarrow) If $\langle N, v^{(0,1)} \rangle$ is a game such that a, b and c satisfies the three inequalities, then $\langle N, v^{(0,1)} \rangle$ is convex (Section 4.1) and any game strategically equivalent to it, in particular G , is also convex.

Example 8. (Example 7 continued) Our game is convex, as $a = b = c = \frac{1}{2}$, by Corollary 6. Therefore, the core of this game, calculated in Example 7 to be the set of imputations $\mathbf{x} = (x_1, x_2, x_3)$ such that $x_1 + x_2 \geq \frac{1}{2}$, $x_2 + x_3 \geq \frac{1}{2}$ and $x_3 + x_1 \geq \frac{1}{2}$, is also the supercore, the only subsolution, and the only stable set.

The next result shows that three-player convex games do not always give rise to well-founded AFs.

Theorem 11. *The game in Examples 1 to 8 is an essential, non-constant-sum three-player convex game whose AF is not well-founded.*

Proof. Example 1 states this game is three-player, Example 8 states that this game is convex, and Example 3 states that this game is essential and not constant-sum. Let $\langle IMP, \rightarrow \rangle$ denote the abstract game from our examples, now seen as an AF. Consider the following IMP -sequence $\{\mathbf{x}^i\}_{i \in \mathbb{N}}$ where

$$\mathbf{x}^i := (x_1^i, x_2^i, x_3^i) = \left(\frac{1}{4} - \frac{1}{2(i+2)}, \frac{1}{4} - \frac{1}{2(i+2)}, \frac{1}{2} + \frac{1}{i+2} \right). \quad (1)$$

Clearly, this is a well-defined imputation for all $i \in \mathbb{N}$, because the three components sum to 1 and each component is non-negative.

We now show that $(\forall i \in \mathbb{N}) \mathbf{x}^{i+1} \rightarrow \mathbf{x}^i$, and hence $\langle IMP, \rightarrow \rangle$ is not a well-founded AF. We only need domination with respect to the coalition $\{1, 2\}$. For any $i \in \mathbb{N}$, we have $\mathbf{x}^{i+1} \rightarrow_{\{1, 2\}} \mathbf{x}^i$ because (1) the agents 1 and 2 do strictly better, i.e. $x_1^{i+1} > x_1^i$ and $x_2^{i+1} > x_2^i$, which in our case is

$$\frac{1}{4} - \frac{1}{2(i+2)} > \frac{1}{4} - \frac{1}{2(i+1)} \Leftrightarrow \frac{1}{i+2} < \frac{1}{i+1} \Leftrightarrow i+2 > i+1, \quad (2)$$

which is true for all $i \in \mathbb{N}$. Furthermore, (2) agents 1 and 2 earn enough payoff after defecting such that they can strictly better, because

$$x_1^{i+1} + x_2^{i+1} = \frac{1}{4} - \frac{1}{2(i+2)} + \frac{1}{4} - \frac{1}{2(i+2)} \leq v^{(0,1)}(\{1, 2\}) = \frac{1}{2}. \quad (3)$$

Equation 3 is equivalent to

$$\frac{1}{2} - \frac{1}{i+2} \leq \frac{1}{2} \Leftrightarrow \frac{1}{i+2} \geq 0, \quad (4)$$

which is true for all $i \in \mathbb{N}$. Therefore, $(\forall i \in \mathbb{N}) \mathbf{x}^{i+1} \rightarrow \mathbf{x}^i$ for $\langle IMP, \rightarrow \rangle$ of the convex game where $a = b = c = \frac{1}{2}$. Therefore, the abstract game from our running example game seen as an AF is not well-founded.

Therefore, not all convex games give rise to well-founded AFs. This result clarifies that the coincidence of solution concepts due to convexity and the coincidence of argumentation semantics due to well-foundedness are of different natures.

5 Discussion and Related Work

In this paper, we have proved that for the uncountably infinite AFs that arise from cooperative games, the complete extensions (respectively the grounded extension) correspond to that game's subsolutions (respectively, supercore), by Theorem 3 (respectively, Theorem 5). This allows for more results from argumentation theory to be applied to cooperative game theory, for example, the lattice-theoretic structure of the complete extensions (Theorem 4) or when the supercore is empty (Corollary 3). Both convex games and well-founded AFs result in a coincidence of, respectively, solution concepts and argumentation semantics, but convex games do not necessarily give rise to well-founded AFs (Theorem 11). To the best of our knowledge, these contributions are original.¹² These efforts strengthen the “correctness” of abstract argumentation by demonstrating its ability to reason about problems of societal or strategic concern.

Our first contribution completes the correspondence between Dung’s original four argumentation semantics with solution concepts in cooperative games. We can use this correspondence in future work to investigate the relationship between further abstract argumentation semantics not mentioned in [7] (e.g. those mentioned in [2]) and solution concepts in cooperative game theory. We can also investigate *continuum* AFs more generally, where the set of arguments $A \cong \mathbb{R}$, as cooperative game theory provides a natural motivation for them.

Our second contribution can be developed further. Intuitively, convex games do not have to be well-founded due to the continuum nature of the simplex and the order-theoretic underpinnings of domination, and hence one may divide

¹² The authors have checked all papers citing [21], which is the first paper defining the supercore and subsolutions from the abstract game of a cooperative game, and found no papers on argumentation theory among them.

extra payoffs into smaller and smaller units. But then one might justifiably ask whether it still makes sense for the first two agents to be sensitive to infinitesimal improvements in their payoffs when i is very large, and thus still maintain their desire to defect. We can attempt to answer this in future work.

There has been much work investigating the relationship between argumentation and game theory more generally. For example, Rahwan and Larson have used argumentation theory to analyse non-cooperative games [17] and study mechanism design [16]. Matt and Toni, and Baroni et al. have applied von Neumann’s minimax theorem [27] to measure argument strength [1,15]. Riveret et al. investigate a dialogical setting of argumentation by representing the dialogue in game-theoretic terms, allowing them to determine optimal strategies for the participants [19]. Roth et al. articulate a prescriptive model of strategic dialogue by using concepts from game theory [22]. Our paper is distinct from these as it applies ideas from argumentation theory to investigate *cooperative* games.

This paper builds on results from Dung’s seminal paper [7, Section 3.1]. There have been works applying ideas from cooperative game theory to non-monotonic reasoning and argumentation, specifically the *Shapley value* [23]. For example, Hunter and Konieczny have used the Shapley value to measure inconsistency of a knowledge base [9]. Bonzon et al. have used the Shapley value to measure the relevance of arguments in the context of multiagent debate [4]. The Shapley value, as a solution concept, is concerned with measuring the payoff to each agent given their *marginal contribution* in each coalition, averaged over all coalitions; we do not consider it here in this paper as we are concerned with solution concepts to do with *defection* rather than marginal contributions. Future work can build on the correspondences in this paper by considering which further solution concepts from cooperative games may be relevant for argumentation.

References

1. Pietro Baroni, Giulia Comini, Antonio Rago, and Francesca Toni. Abstract Games of Argumentation Strategy and Game-Theoretical Argument Strength. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 403–419. Springer, 2017.
2. Ringo Baumann and Christof Spanring. Infinite Argumentation Frameworks. In *Advances in Knowledge Representation, Logic Programming, and Abstract Argumentation*, pages 281–295. Springer, 2015.
3. Olga N. Bondareva. Some Applications of Linear Programming Methods to the Theory of Cooperative Games. *Problemy Kibernetiki*, 10:119–139, 1963.
4. Elise Bonzon, Nicolas Maudet, and Stefano Moretti. Coalitional games for abstract argumentation. In *COMMA*, pages 161–172, 2014.
5. Jean-François Caulier. A note on the monotonicity and superadditivity of TU cooperative games. 2009.
6. Georgios Chalkiadakis, Edith Elkind, and Michael Wooldridge. Computational Aspects of Cooperative Game Theory. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 5(6):1–168, 2011.
7. Phan Minh Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n -Person Games. *Artificial Intelligence*, 77:321–357, 1995.

8. Donald B. Gillies. Solutions to General Non-Zero-Sum Games. *Contributions to the Theory of Games*, 4(40):47–85, 1959.
9. Anthony Hunter and Sébastien Konieczny. Shapley Inconsistency Values. *KR*, 6:249–259, 2006.
10. Thomas Jech. *Set Theory, the Third Millennium Edition, Revised and Expanded*. Springer Monographs in Mathematics. Springer, Berlin, 2003.
11. William F. Lucas. A Game with No Solution. Technical report, RAND Corporation, Santa Monica, California, 1967.
12. William F. Lucas. The Proof that a Game may not have a Solution. *Transactions of the American Mathematical Society*, 137:219–229, 1969.
13. John C. S. Lui. CSC6480: Advanced Topics in Network Analysis lecture 10 - Cooperative Games (2). www.cse.cuhk.edu.hk/~cslui/CSC6480/cooperative_game2.pdf, 2008.
14. Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory*, volume 1. Oxford university press New York, 1995.
15. Paul-Amaury Matt and Francesca Toni. A game-theoretic measure of argument strength for abstract argumentation. In *European Workshop on Logics in Artificial Intelligence*, pages 285–297. Springer, 2008.
16. Iyad Rahwan and Kate Larson. Mechanism Design for Abstract Argumentation. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 2*, pages 1031–1038. International Foundation for Autonomous Agents and Multiagent Systems, 2008.
17. Iyad Rahwan and Kate Larson. Argumentation and Game theory. In *Argumentation in Artificial Intelligence*, pages 321–339. Springer, 2009.
18. Iyad Rahwan and Guillermo R. Simari. *Argumentation in Artificial Intelligence*, volume 47. Springer, 2009.
19. Régis Riveret, Henry Prakken, Antonino Rotolo, and Giovanni Sartor. Heuristics in argumentation: a game-theoretical investigation. In *Proceedings of the 2nd international conference on computational models of argument*, pages 324–335. IOS Press, 2008.
20. Alvin E. Roth. A Lattice Fixed-Point Theorem with Constraints. *Bulletin of the American Mathematical Society*, 81(1):136–138, 1975.
21. Alvin E. Roth. Subsolutions and the Supercore of Cooperative Games. *Mathematics of Operations Research*, 1(1):43–49, 1976.
22. Bram Roth, Régis Riveret, Antonino Rotolo, and Guido Governatori. Strategic argumentation: a game theoretical investigation. In *Proceedings of the 11th international conference on Artificial intelligence and law*, pages 81–90. ACM, 2007.
23. Lloyd S Shapley. A Value for n -Person Games. *Contributions to the Theory of Games*, 2(28):307–317, 1953.
24. Lloyd S. Shapley. On Balanced Sets and Cores. *Naval Research Logistics Quarterly*, 14(4):453–460, 1967.
25. Lloyd S. Shapley. Cores of Convex Games. *International Journal of Game Theory*, 1(1):11–26, 1971.
26. Lyn Carey Thomas. *Games, Theory and Applications*. Courier Corporation, 2012.
27. John Von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton university press, 1944.
28. Anthony P. Young. Notes on Abstract Argumentation Theory. *ArXiv preprint arXiv:1806.07709*, 2018.