A User's Theory
How to Model Agents of Online Debates

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakulät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von
Hilmar Schadrack
aus Düsseldorf

Düsseldorf, Februar 2019

aus dem Institut für Informatik
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1.

2.

Tag der mündlichen Prüfung:

In caverns deep and dark and cold
Where shadows do with shadows mate
Where ancient books dream dreams untold
Of being trees before their fate
Where coal begets bright diamond-stone
And mercy is an unknown thing
There in the deep sits on his throne
The one is called the Shadow-King.

Right in his grasp, under his guard
There is a chest, so big and dark.
You try to look, but you are blind
Right as a voice cuts through your mind:

"This chest you see here on my side",
The voice is dark and cold and deep
Just as the shadows here do creep,
"Is filled with secrets urged to hide,
With all the world could ever know
And wonders you could not comprise."
The voice is rumbling in your head,
This presence fills you up with dread,
Your body tingles, you wanna leave,
But cannot move, not even breathe.
"But I will give you for your eyes
One tiny piece to make you grow".

Just for a blink you saw a light,
Your mind is cleared and shines so bright
And were the chest was short before,
there stands now just a simple door.

You wished to know just so much more,
So couldn't help yourself and then
You started rushing through the door
Before it closes once again.

Acknowledgments

I would like to thank Dorothea Baumeister and Jörg Rothe, as well as my colleagues and family for their ongoing support. This research was part of the Ph.D.-programme "Online Participation", supported by the North Rhine-Westphalian funding scheme "Fortschrittskolleg".

Erklärung

Ich versichere an Eides Statt, dass die vorliegende Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Des Weiteren erkläre ich, dass ich eine Dissertation in der vorliegenden oder in ähnlicher Form noch bei keiner anderen Institution eingereicht habe.

Teile dieser Arbeit wurden bereits in den folgenden Schriften veröffentlicht beziehungsweise zur Publikation angenommen: [54], [55], [56], [40], [62], [8], [9], [11], [12] und [13].

Düsseldorf, 05.02.2019

Hilmar Schadrack

Abstract

This thesis deals with hedonic games and abstract argumentation, both lively research fields that belong to the areas of multiagent systems and artificial intelligence. Hedonic games is a specialized subfield of coalition formation, which is concerned with questions regarding the grouping of agents. More specifically, coalition formation is about when, how, and why agents work together and what it needs that these so-called coalitions do not break up again. This includes the modeling of utilities for the agents, as well as thresholds and other means that specify when agents might have an incentive to leave their current coalition. This field is also influenced by other areas of computer science and artificial intelligence, as, for example, computational social choice with notions of manipulation and control. In hedonic games, we assume that agents are only concerned about their own coalition and its members. Hence, we do not need a specific utility notion, as we assume the agents to provide a preference ranking over all possible coalitions including themselves.

Our work includes research on hedonic games in two ways. First, we study the computational complexity of a special case for hedonic games in which agents only express sets of friends and enemies, but not total rankings; we narrow existing gaps in complexity regarding the existence of the strict core in these games with the help of closely related graph problems. Second, we introduce a new model for hedonic games that allows for a compact, yet quite expressive representation. Sadly, our model suffers from a problem of incomparable coalitions. We address this issue on the one hand with the help of possibility and necessity notions, and on the other through means of comparability functions that work similar to the Borda-scoring vectors from social choice theory.

In detail, we will show that in the first case with enemy-oriented preferences, it is at least DP-hard to decide whether the strict core of a given hedonic game is nonempty, and that its complexity cannot rise above the second level of the boolean hierarchy. The same holds for the encountered graph problem of deciding whether an undirected graph contains a wonder-

fully stable partition. For the second case, we will provide a comprehensive analysis of the model's properties, as well as one possibility of how to extend the resulting incomplete preference over coalitions for each player in form of the polarized responsive extension principle. Our analysis also includes a classification of the computational complexity of the verification and existence problems for several stability concepts including the (strict) core, Nash stability, perfectness, Pareto optimality and more. This includes hardness and completeness results for NP and Σ_2^p, but also feasibility results in both cases, i.e., for the notions of possibility and necessity, as well as for Borda-induced hedonic games.

Abstract argumentation takes a different view on agents' behavior. It belongs to the intersection of social sciences and computer science, and uses an abstraction approach on topics from argumentation theory. In argumentation theory, scientists try to precisely analyze the behavior of human agents in debates and discussions. Abstract argumentation takes a more distant view and assumes the agents to not only be human, but arbitrary, and also abstracts from the internal structure of given arguments. Instead, arguments are seen as nodes in a network, and the connection between nodes represents their correlation, which is derived in an earlier step that is no longer of interest. In past research it turned out that focusing on conflicting behavior instead of allowing for multiple or complex interactions between arguments is enough to express a wide range of situations. The goal in this model is to find so-called extensions, i.e., subsets of the arguments, that satisfy certain criteria, such as having no internal conflicts or being able to defend against incoming attacks.

Here, we focus on an expansion of the basic model such that we can deal with situations in which complete information over all given arguments and their interactions is not given beforehand. That is, we introduce another set of arguments and attacks for which it is not clear from the start whether they will be part of the debate or not. We then use notions of possibility and necessity to deal with this high degree of uncertainty and analyze this model in terms of computational complexity. More specifically, we will show that the verification problem for argumentation frameworks can be extended to fit to the new model, and that its computational complexity can rise up to Σ_2^p-completeness. However, this increase only happens in situations in which we have to deal with alternating quantifiers. In all other cases we can find shortcuts that allow for a faster solution to the problem.

Both parts of this work illuminate different views on the behavior of agents. When focusing on an abstract level of indirect interaction as in hedonic games, we can concentrate more on the bigger picture instead of the overwhelming density of argumentative approaches. However, on the side of abstract argumentation, we directly use this comprehensive information regarding a specific argumentation process to directly elicit a solution, while abstracting from external influences. Both views are needed to correctly and completely model all possible actions of any kind of agent. This thesis takes another step on this agenda.

Contents

CHAPTER 1

Introduction

The term 'agent' is used in computer science as a placeholder for a possibly heterogeneous group of entities. When talking about agents, one could talk about voters or candidates in elections, computers in distributed systems, judges in a court, players of a game, or something completely different. Given the fact that many formal models of computer science are derived from such real-world applications, but are also valuable in many other real-world situations, it is understandable why 'agents' are a central point of interest in research. However, an arguably big flaw is the often taken assumption that agents are acting rational and that they are describable in some self-contained mathematical model. Agents in such models are predictable and the worlds these models describe are assumed to be known completely. There are a lot of good reasons for these assumptions and the easiest might be that it is currently not possible to model every aspect of the behavior of agents in open world scenarios. Anticipating the content of this thesis, we will probably not break out of this shell anytime in the near future. However, trying to shatter the bonds of these restrictions is one of the most important and prominent tasks of modern-day research.

In this work, we join the idea of giving agents more freedom in their actions. Our main field of interest is the question of how agents form, join and leave groups, communicate, exchange information and 'argue' about every day topics. More importantly, we investigate how to get a formal look on such situations, not only, but also to get a better understanding of the behavior of human agents. Our approach is to take close-world models and extend them to meaningful models with more freedom for agents, based on interesting situations that arise in the real world. One major goal is to achieve results that somehow can be translated back to the real world such that they can be used to improve in turn the future outcome of these interesting situations that we discovered earlier. We do this with the help of the interlapping research fields of computational social choice, game theory, and abstract argumentation.

Computational social choice, a field being part of the union of the social sciences and computer science, consists of the analysis of problems that arise if we want to computationally aggregate opinions into a collective decision. It contains topics such as voting, fair division, judgment aggregation, and others. The common ground is the idea of collecting a priori formalized opinions from agents, and digest these opinions into a commonly acceptable collective opinion. An example for this would be an election, in which an authority collects the preferences of the voters over a given set of candidates, and identifies with the help of these votes a winner. For more information on computational social choice, we refer the reader to [18], a book that serveys computational social choice from past to present, and to [31], a relatively new book that serves as an excelent introductory book to the matter of computation social choice.

Games theory is a subfield of both, computer science and mathematics, fueled by problems that arose from economics and related studies. Scientists tried to explain the behavior of economic agents such as salespersons or markets, with the help of mathematical models and game-like simulations. In game theory, agents are often called players, and one tries to estimate the strategic behavior of two such players in conflicting situations, if, for example, both players try to maximize their profit in a simple negotiation process. Here, profit is often referred to as the 'utility' a player can get, which itself is not specified any further (except for the assumption that the utility results in a natural number and higher numbers means higher utility). The field of game theory became increasingly popular in 1944 due to von Neumann and Morgenstern [68] as result of the increasing interest in mathematics during the second world war. In 1950, Nash finished his seminal work [45] regarding non-cooperative games, i.e., mathematical models to analyze the strategic behavior of individual agents who are trying to win over an opponent in game-like situations. In contrast to this, in cooperative games agents profit from forming coalitions, i.e., acting cooperatively. In such cooperative games we can now distinguish between games with or without transferable utility. The former describes cooperative games in which agents can directly transfer utility to other members of their coalition without restrictions. In the latter, one focuses mainly on the coalitions itself, and the utility of an agent is directly derived from the grouping of all agents. A common assumption is that the utility of an agent only depends on the coalition that she is a member of but not on how the rest of the players group together. Such games are called hedonic games and are commonly investigated for different notions of stability, such as core or Nash stability. A common ground of the most stability notions is the idea of agents not having an incentive to leave her current coalition. An early application of such hedonic games is the stable marriage problem due to Gale and Shapley [32]. The idea has been developed further to be of use for matching students to residency programs as of Roth and Peranson [59], to study the composition of teams as by Alcalde and Revilla [1] or distributed task allocation due to Saad et al. [63], among many other applications.

Abstract argumentation, on the other hand, is a relatively young branch of the scientific environment, where researchers use mathematical methods to shed light on the highly complicated field of argumentation. Considering the age of the field, going back to the idea of democracy of the ancient Greece and the need to analyze speech regarding its influencing effects, it is surprising that a formal, mathematical analysis of this field did not become increasingly popular before 1995 through Dung [29]. Before that time, mainly scientists from philosophy or social, political or legal sciences saw an incentive to do research in the field of argumentation. It is hard to determine the beginning of modern scientific analysis of the field of argumentation, but one might see the deeper philosophical analysis of rhetoric in the early 19th century by Whately [73] as an important turning point. At this time, argumentation was defined as the question-answer dialog of two agents with controversial positions. Another strong candidate for the origin of argumentation being part of modern science originates in legal sciences. An important task here is to determine which party has the burden of proof (Walton [72]), i.e., the responsibility to prove (or disprove) a presumption. In a simplified legal proceeding, this burden of proof would alternate between prosecutor and defendant until one party is not able to repel it successfully, and therefore, lose the court proceedings.

Argumentation processes have been recorded in many different ways, ranging from simply listing the arguments (possibly in two lists, one for pro, one for contra arguments), to using complex tools such as argumentation maps as by Rinner [58]. To provide a simple model for mathematical analysis, Dung assumed in his seminal work [29] that the internal information of arguments can be used to derive conflicting links, so-called *attacks*, between them. Then, he could abstract from the internal structure of the given discussion, and therefore simplify it. Now, one could use this simplified discussion to compute winning subsets of the arguments, i.e., subsets that do not conflict and fulfill a priori defined properties, such as the ability to defend its members from incoming attacks. He would then call those subsets extensions and pronounce them to be solutions of the discussion. His work has staggering influence on the field of argumentation until today, and led to a high number of papers building on his idea, such as argumentation frameworks with recursive attacks [7], abstract dialectical frameworks [19, 20], bipolar argumentation frameworks [3], value-based argumentation [16, 28], preference-based argumentation [2], extended argumentation framworks [43, 44], and probabilistic argumentation frameworks [41].

As all three fields, computational social choice, game theory, and abstract argumentation were born from the interest to analyze problems from sociology, economics, politics, law or philosophy with the help of formal methods, it is understandable that scientists from all fields exchange methods to improve the quality of their research, or even establish new branches such as argument games [69, 37].

Outline

Chapter 2 provides the general background of the investigated topics, i.e., preliminary information on online participation, computational complexity, hedonic games, and abstract argumentation. To give a deeper understanding, we anticipate some of the definitions of the attached papers and present more detailed explanations beforehand to compensate for the limited space in the papers.

In Chapters 3, 4, 5, and 6, we summarize and include the papers that have been part of this Ph.D. project, as well as give a detailed overview over the contribution of the author of this dissertation to each included paper. Our work described in Chapter 3 is on the topic of classical hedonic games, for which we investigated open gaps in computational complexity regarding strict core stability in enemy-oriented hedonic games, as well as related problems. In Chapter 4 we introduce a new idea on how to represent hedonic games compactly, that merges two existing approaches. We also investigate this model in terms of their computational complexity for several stability concepts. This model is then extended in Chapter 5 with the help of comparability functions that suit as a tool to close the gap between the partial rankings that our new model contains, and total rankings that are needed for hedonic games. Again, we did a comprehensive complexity analysis of this model. Finally, in Chapter 6, we use abstract argumentation to analyze the behavior of agents. Our contribution is a new model that is capable of displaying incomplete information on both, the arguments and the attacks, of a classical argumentation framework. In Chapter 7 we conclude our results and make a note on future work.

CHAPTER 2

Background

2.1 Online Participation

Online participation is a widely used term for all processes in which people take part of an collaborative decision or discussion with the help of the internet. Decisions derived from this participation process often have advising character for the authority that started the process, but in some situations such decisions might be obligatory. Such processes are often started by governments that ask for opinions of their citizens on manifold situations, e.g., how to distribute a specific budget, where potholes can be found, additional zebra crossings are needed, or when more bus lines are necessary. Citizens then usually contribute by writing comments in which they, e.g., point to specific spots in the city that they think should be dealt with. In the past, those participation processes were only used by a minority of people, as they required citizens to personally come to the town hall, write a letter, or contribute by other means. With the growing possibilities of the internet, such processes became more and more flexible and authorities started online participation processes to receive more and better feedback. It turned out that some of these processes have been really successful, while others seem to be attracting almost no one. To shed light on the reasons for this unpredictable behavior is a major topic in current research.

Simultaneously, a second problem arose, namely the problem of how to find a 'common sense', or an 'aggregated opinion' of the participating users to get to a consensus. Having in mind that such processes might be used by governments with possibly thousands of people attending, one quickly loses track of the discussions and suggestions that users make. Therefore, we need an automated tool to analyze such discussion and find those suggestions that withstand the critical debate and arose as the 'strongest'. On the other hand, we might not want to know what the single strongest suggestion is, or the strongest suggestions are, but instead want to get a general idea on how the discussion went, what parts were the most debatable ones and what seemed to be easily agreed on. And last but not least, all those decisions have to be transparent for the user, such that she can always follow up on how her opinion had been taken into consideration.

To answer these questions, the graduate school 'Online Participation' was founded by the North Rhine-Westphalian Ministry for Innovation, Science, and Research. Involved in this graduate school are the faculties of Law, Mathematics and Natural Sciences, Arts and Humanities, and Business Administration and Economics. Currently, fifteen Ph.D. students are members of this graduate school, including four associated members. One major goal is to do intra- and interdisciplinary research on the topic of civic participation, to lighten up the possibilities and limits of enabling citizens to have direct impact on political and administrative decisions. From the perspective of theoretical computer science, we have the goal to provide a strong and scalable mathematical foundation for all emerging problems. It turned out, that creating one expressive model to cover all aspects one could encounter would be infeasible. Hence, we had to stick to the idea of using several different models to cover the most important aspects and analyze them separately. We refer the reader to `https://www.fortschrittskolleg.de/en/` for more information on the graduate school Online Participation.

Dialog-Based Argumentation System

The Dialog-Based Argumentation System is a platform for distributed users to discuss a common topic. It was created by Krauthoff et al. [39] at the Heinrich-Heine-Universität in Düsseldorf, Germany, and designed to improve certain flaws of existing platforms for deliberation or argumentation. It's intriguing approach is to simulate a time-shifted dialog of the users, guided by the system to lead the user to the most interesting parts of the discussion. While this guiding algorithm needs to be designed carefully in terms of transparency, it opens up for a broad spectrum of possibilities, such as guiding the user to parts of the discussion that she seems to be most interested in, that needs more discussion from an authorities point of view, or simply to resolve conflicts. Some of the results of this thesis can be incorporated into the design of the guiding algorithm, for example, to locate arguments that prevent the existence of particular solutions. First steps in this direction have already been done by Neugebauer [46].

2.2 Computational Complexity

To understand computational complexity, we first have to understand complexity theory. In complexity theory we basically try to partition all mathematically describable problems into certain groups. Each group only contains those problems that are 'equally complex' to solve. At first sight, there is no direct restriction on what kind of problems we could investigate. However, we focus on problems that can be fully expressed with the help of formal methods or mathematical models. This usually leads away from problems expressible by natural language to problems derived by mathematical abstraction. These problems are then put into groups, i.e., collections or simply sets of equally hard problems. Where exactly the border of these groups is drawn, and what 'equally hard' means, depends on the focus of the analysis. An interesting example could be, that we simply put all riddles of the known world into two sets. The first set contains all riddles that can be solved by every single member of the group of people that contains only the upper 5% of the most intelligent people on earth. The other set contains all riddles that can be solved by at least one of all remaining people. Admittedly, this example is highly unrealistic, as it is probably not only impossible to create a trustworthy ranking of the intelligence of all human beings but also to ask them all to try to solve a riddle. However, this setting raises interesting questions: First, what set will be larger, the one that contains riddles that have to be solved by the generally more intelligent group of people, in which, however, no one is allowed to fail in finding a solution; or the one that contains those riddles that can be solved by at least one of the less intelligent people? Second, is one set contained in the other, or can we at least determine if there are riddles, that belong to both sets, and if the answer is yes, which are those? Can we maybe also find a rule or a common ground that easily identifies the riddles of each set? Third, what are the hardest riddles of each set, and what makes them so hard to solve?

The computational part of computational complexity results in the idea of using computing devices, such as Turing machines or computers as measurements for the complexity of given problems. As measure, we usually use the running time or consumed space of the device in dependence of the size of the input. To structure the infinitely large number of problems, we sort these a priori into different types. To explain this matter further, we take a look at the classical traveling salesman problem, that describes the problem of a salesperson who wants to visit all cities of the country by using only already existing roads and without visiting one city twice. Given this problem, we can easily extract four types of problems, that all get the same input (a map of the country including the location of the cities and roads between them), but different outputs: First, one can ask whether there is at least one solution to the given input. Second, we might want to know one route that solves the given instance of the problem. Third, it could be of interest to count the number of possible solutions. And fourth, the salesperson probably wants to know the shortest path. The first problem type is called *decision problem*, as we only try to decide the given problem instance, without the need to specifically output a solution. The second is called *function problem*, in which we want to compute one solving solution, if it exists. The third type is named *counting problem*, as we want to count the number of the solutions, but, similar to decision problems, not output a specific one. And the fourth type we call *optimization problem*, as we want to output an optimal solution. Please note, that the fourth type uniquely uses implicit information of the input, namely the distance between the cities

It is easy to see that some of these types of problems are connected in some way. For example, if we know the answer of a counting problem, we can immediately answer the respective decision problem as well. However, knowing the answer to the decision problem has only immediate consequences to the answer of the counting problem if the answer is 'no', as a 'yes'-answer only implies the existence of at least one solution, but not the exact number of solutions. Another connection can be found between the function version and the decision version of a problem. Assume, that we already know a solution to the problem, we trivially also know that there exists at least one solution. However, knowing that there has to be a solution does no easily help us with finding it. A similar connection exists between the optimization version and the decision version, as knowing an optimal solution not only tells us that there is at least one solution but also that there is no 'better' one.

Despite these relatively easy connections, different types of problems are often connected even if their input is not exactly the same. To understand this, we need to explain the second dimension of the distinction of complexity classes, the measure. We already mentioned, that we want to measure the difficulty of a problem by the use of different measurements. The two most prominent measures are the running time and the consumed space, and for the sake of simplicity, we assume that we use Turing machines[1] as computing device and the running time as a measure. Now we can define different complexity classes by assigning to each of these classes a mathematical function, such as a root, polynomial, or exponential function, a faculty, a logarithmic function, or else. Here, we are not interested in the exact growth rate of the

[1]We assume the reader to be familiar with the basic computing concept of Turing machines. For more information, see the book by Rothe [60].

given function, but only in its asymptotic growth, i.e., for this matter the functions $f(x) = 2x + 5$ and $g(x) = x$ are equivalent. As we now have identified each complexity class with an asymptotic growth rate, we can have a look at a fixed problem to decide to which class it belongs. Therefore, we assume that we have a Turing machine that solves this fixed problem for all inputs, then we can compute a function that describes its asymptotic running time on worst case inputs. This Turing machine now serves as a witness for the given problem and proves the membership of this problem in exactly that complexity class that is represented by the asymptotic running time of the function that describes the running time of the Turing machine.

As we now have all the information we need, we can turn back to the different problem types and the idea that they are connected even if their inputs are different. Assume, for example, a special kind of decision problem, namely *verification problems*. Here, the input consists of two parts, the original input and another object. The question is whether we can verify something for that object on the original input. Let us again have a look at the traveling salesman problem. The first part of the input is the map of the country including the location of the cities and the roads between them. The second part could be a route that the salesperson thinks to be a possible solution for the problem. Now, we can ask whether this route solves the problem, therefore verifying it as a solution. This seems to be an easy task, and indeed, this verification version of the traveling salesman problem is solvable by a Turing machine that runs in polynomial time with respect to the input. As polynomial time is usually considered to be an acceptable running time, we can say that this verification problem is easy to solve. Now, one could try to use this result to find an optimal solution to the problem, thus solving the optimization version of the problem, by iteratively applying the Turing machine for the verification version of the problem to all possible routes. Even though this definitively provides a solution to the optimization version, its running time would be too high, as the number of possible routes grows exponentially in the size of the input map, therefore resulting in an asymptotic worst case running time that is exponential in the input size. As even small exponential functions grow extremely fast if the input is sufficiently large enough, an exponential running time is considered to be infeasible. However, there are problems for which this naive approach works in the sense that the resulting Turing machine has a running time that is at most polynomial.

Another crucial part of computational complexity is the notion of hardness. Until now, we only discussed membership in a certain complexity class. However, as our introductory example already suggested, we are also interested in the question, what the toughest problems of a given class are. For this, we make use of the notion of hardness. We call a problem *hard* for a complexity class, if it is provably at least as difficult as every problem in that class.[2] To make this proof easier, we usually do not compare all possible problems in that class with our designated one, but instead show that even one of the hardest problems of that class is not harder than our problem. To compare the difficulty of two problems, we use different notions of reducibility, and for exemplary reasons, we stick to the following idea: Assume, we could find for a problem of a fixed complexity class an easy way, e.g., an algorithm[3] running in polynomial time, to transform every instance[4] of that problem into an instance of a second problem that we want to prove being hard for the given complexity class. Then we know, that the second problem is at least as hard to solve as the first one, because if we could solve an instance of the first problem, we could apply our transformation resulting in a solution, respectively an answer to the second problem. Such a transformation is called a *reduction* from the first to the second problem. Now assume further, that we can find such a reduction from a problem that we know to be one of the hardest problems of the investigated complexity class. In that case we would have been able to show that our designated problem is at least as hard as every other problem in that class. The only issue that remains is to find a problem to start with, i.e., a problem that is provably one of the hardest problems of a complexity class. Luckily that was already done in the past for most of the known classes, resulting in a flood of papers that prove huge numbers of problems to be among the hardest of their classes.

In this work, we only investigate decision problems building on the classes P, NP, and coNP, and use the basic notions of hardness and completeness (based on polynomial-time many-one reducibility, $\leq_{\mathrm{m}}^{\mathrm{p}}$). As the formal definition of these classes and notions is not the main focus of this work, we refer the reader to the books by Papadimitriou [48], Rothe [60], and Arora and Barak [4] for more information, and assume from now on that the reader is familiar with those terms.

A class that is defined with the help of the class NP is DP, a class introduced by Papadimitriou and Yannakakis [49]. DP is the second level of the boolean hierarchy over NP (see the articles by Cai et al. [21, 22] for a concise

[2]Please note, that we do not require that every input for a problem is hard to solve (which is, nevertheless, unrealistic, as there always is at least one trivial case). Instead, we need one worst case input that is difficult to solve.

[3]An *algorithm* is a specification of how to solve a class of problems, e.g. how to perform calculations, process data or answer reasoning tasks. A Turing machine is, for example, a generally accepted attempt to formalize the intuitive idea of an algorithm.

[4]An *instance* of a fixed problem is a tuple that contains all parts of the input that are specified in the problem's definition.

analysis of the boolean hierarchy). For natural complete problems of DP but also for other levels of the boolean hierarchy, see the survey by Riege and Rothe [57] and, more recently, the work of Nguyen et al. [47] and of Reisch et al. [53], both in the field of computational social choice. DP is defined as the class of problems that can be described as the intersection of an NP problem with a coNP problem. Equivalently, any problem in DP can also be defined as the difference of two NP problems. It is known (and easy to see by its definition), that DP is a superset of NP $\cup$ coNP.

Another class building on the basic complexity classes is $\mathrm{P}^{\mathrm{NP}[\log]}$, introduced by Papadimitriou and Zachos [50]. It was defined as the class of problems that can be solved in polynomial time by asking $\mathcal{O}(\log n)$ sequential Turing queries to an NP oracle[5]. Problems of this class can also be solved by asking a polynomial number of parallel Turing queries to an NP oracle, and vice versa. This equivalence was shown independently by Hemachandra [33] and Köbler, Schöning, and Wagner [38], and the class of problems with the latter structure is known as $\mathrm{P}^{\mathrm{NP}}_{\|}$. $\mathrm{P}^{\mathrm{NP}[\log]} = \mathrm{P}^{\mathrm{NP}}_{\|}$, belongs to the Θ_2^p level of the polynomial hierarchy. Structural research of this class goes back to Köbler, Schöning, and Wagner [38], Hemachandra [33], Wagner [71], Beigel, Hemachandra, and Wechsung [15], and Beigel [14]. Several authors also focus on proving completeness of natural problems in it, for example, versions of classical graph problems as Clique, Colorability, Independent Set, Vertex Cover or Traveling Salesman as by Wagner [70], versions of the famous Satisfiability problem (also Wagner [70]), and the winner problems for the voting systems by Dodgson, Young, and Kemeny, due to Hemaspaandra, Hemaspaandra, and Rothe [34], Rothe, Spakowski, and Vogel [61], and Hemaspaandra, Spakowski, and Vogel [36]. We also refer the reader to the survey by Hemaspaandra, Hemaspaandra, and Rothe [35] for more interesting research on that topic.

The second level of the polynomial hierarchy (see the work by Stockmeyer [66], and Meyer and Stockmeyer [42]) is named $\Sigma_2^p = \mathrm{NP}^{\mathrm{NP}}$. It is defined for decision problems for which yes-instances, i.e., problem instances for which the answer to the respective question can be answered with 'yes', can be verified in nondeterministic polynomial time with access to an NP oracle. Natural complete problems of the polynomial hierarchy, especially of Σ_2^p, have been surveyed by Schaefer and Umans [64, 65]. Recent results on the complexity of core stability in hedonic games are due to Woeginger [75] (see also his survey [74]). It holds that $\mathrm{P} \subseteq \mathrm{NP} \cup \mathrm{coNP} \subseteq \mathrm{DP} \subseteq \Theta_2^p \subseteq \Sigma_2^p$, and none of these inclusions is known to be strict.

[5]See the book by Rothe [60] for more information on the Bachmann-Landau notation and oracle access.

2.3 Hedonic Games

Taking the real world as an example, people tend to form groups to perform certain tasks or answer specific questions. The size, structure, cohesion, and goal of these groups depend on each situation and can vary from small groups of two people with equal rights that stay together for a lifetime to form a bond from which both benefit (as, for example, in relationships) to large groups of thousands of people that belong to a clear hierarchy in which individual members leave and new members join frequently with the aim to earn money (as, for example, in large companies). The mathematical term for these groups, *coalitions*, originates from political parties and the forming of clusters of similar political attitude. Today, we use this term as a general notion for all kinds of grouping agents, even if the agents are not of human nature.

The first studies on coalitions and their behavior cannot be dated exactly. However, we can say that the greek democratic society constituted a huge demand of research on this topic. It was this era of great philosophers that started and accelerated many different kinds of scientific fields, and this also happened to the fields of game theory and coalition formation. Until today the demand of research on this topic is rising, which led to the founding of subfields and interconnection fields and the interchangeable use of methods their methods an idea. The field of hedonic games is one of these children that was born in the need of more comprehensive and accurate research for a very specific problem.

In coalition formation in general, we are mainly interested in the coalition formation process, i.e., in the understanding of when and why agents join or leave coalitions, and whether we can reach some kind of equilibrium, i.e., a situation in which no one wants to deviate from their current coalition. Special for hedonic games is the assumption, that agents are only interested in the coalition that they (could) belong to, but not how other agents group together. Formally, an agent expresses this interest in the context of hedonic games with a preference ranking over all possible coalitions that she could belong to; no other input is given. However, for some analysis this input is too large, as each of these rankings consists of a list of a major subset of the power set over the agents, which is exponential in the number of the agents. Therefore, researchers invented several so-called *encodings* to represent hedonic games more compactly without significantly decreasing their expressivity. The idea is to let agents only express a smaller part of their preference relation, for example, only the other agents instead of coalitions (which is called *singleton encoding* in the literature), and then extend this information to total ranking over coalitions, recreating a hedonic game. Naturally, not every hedonic game is representable by every encoding as compact representations always lead to information loss. However, depending on the domain, the idea of using encodings is extremely valuable.

One major goal in hedonic games is to make presumptions about the stability of coalition structures, i.e., collections of several coalitions that together contain all agents of the investigated game. For the stability notion, a wide range of meaningful ideas has been studied in the literature, ranging from very basic notions as *individual rationality*, which secures that every player[6] ends in a coalition that she prefers to being alone, to more complex notions that incorporate the idea that no individual should have an incentive to leave her current coalition (e.g., *Nash stability*), or that no group of players wants to deviate (e.g., *(strict) core stability*).

2.3.1 Preliminary Definitions

A *hedonic game* consists of a finite set $N = \{1, \dots, n\}$ of *players* and a profile $\succeq = (\succeq_1, \dots, \succeq_n)$ of *preference relations*, where $\succeq_i$ denotes player i's preference relation. Each such preference relation $\succeq_i$ defines a weak preference order over all *coalitions* (i.e., subsets of *players*) that contain player i. By $\mathcal{N}_i$ we denote all coalitions of N that contain player i. For two coalitions $A, B \in \mathcal{N}_i$, we say that *i weakly prefers A to B* if $A \succeq_i B$. Additionally, we say *i prefers A to B* (denoted by $A \succ_i B$) if $A \succeq_i B$, but not $B \succeq_i A$, and *i is indifferent between A and B* (denoted by $A \sim_i B$) if $A \succeq_i B$ and $B \succeq_i A$[7]. A *coalition structure* Γ for a given game is a partition of N into disjoint coalitions, and for each player $i \in N$, Γ_i denotes the unique coalition containing i, i.e., $\Gamma_i = \mathcal{N}_i \cap \Gamma$.

Example 2.1 *Assume a situation, in which three players have to decide over how they want to cooperate to deal with a given task. For the sake of this example of hedonic games, the task itself is not of interest for us, as we are only interested in the coalition formation process. We further assume, that the three players have a specific opinion over the coalitions they could be part of. This could, for example, lead to the hedonic game $H = (N, \succeq)$, where $N = \{1, 2, 3\}$ is the set of the three players 1, 2 and 3, and $\succeq = (\succeq_1, \succeq_2, \succeq_3)$ is a profile of one preference relation for each player. A possible preference*

[6] In game theory, agents are commonly called *players*. In this work, we use both terms interchangeably.

[7] In some literature, the term *i strictly prefers A to B* is used for the notion $A \succ_i B$, while *i prefers A to B*, the version without adjective, is used for $A \succeq_i B$. Also, $A \sim_i B$ is often called *equally preferred*.

profile is, for example,

$$\begin{aligned}
\succeq_1\colon\quad & \{1,2\} \succ_1 \{1\} \succ_1 \{1,3\} \sim_1 \{1,2,3\},\\
\succeq_2\colon\quad & \{1,2\} \succ_2 \{2\} \sim_2 \{1,2,3\} \succ_2 \{2,3\},\\
\succeq_3\colon\quad & \{1,2,3\} \succ_3 \{2,3\} \succ_3 \{1,3\} \succ_3 \{3\}.
\end{aligned}$$

This profile implies, that player 1 *prefers being together with player* 2 *than being alone, while she seems to dislike any constellation in which she has to be together with player* 3*. Player* 2 *also prefers to work together with* 1*, but strictly refuses to be paired with* 3*. However, a group of all three, meaning that she is not alone with* 3*, is somewhat ok for her. Player* 3 *prefers any coalition in which he does not have to be alone, and, more specifically, prefers* 2 *to* 1 *if he has to chose. His most preferred option is, however, the grand coalition.*

Since the number of coalitions in a player's preference relation is exponential in the number of players, it is reasonable to consider compactly represented hedonic games (as already mentioned in the preface); see the next section for an overview of various possible encodings, as well as the survey of Woeginger [74].

In this thesis, we tackle two questions regarding hedonic games. In Chapter 3, we have a look at so-called enemy-oriented preferences as introduced by Dimitrov et al. [26]. In their setting, every player $i \in N$ reports a set of friends and a set of enemies, and this information is extended to a total ranking over all possible coalitions containing player i in two versions, one with focus on the friends, and one with focus on the enemies. We approach the question on how hard it is to decide whether a given hedonic game with enemy-oriented preferences has a strictly core stable coalition structure. Chapters 4 and 5 are about a new compact encoding for hedonic games that is more expressive than many known encodings, including friend- or enemy-oriented hedonic games. We then analyze this new model in regards to axiomatic properties and computational complexity.

2.3.2 Representations of Hedonic Games

For a concise representation of a hedonic game, the players should express their preferences in a compact manner. On the other hand, they should be able to express their opinion as precise as possible. Therefore, a high number of suggestions has been made in the literature. In chapters 4 and 5 we will give a new representation that unites some of the advantages of existing ideas. Below, we will list some of the known representations of hedonic games.

We start with a very powerful class of hedonic games that was introduced by Banerjee et al. [6]. An *additively separable hedonic game (ASHG)* is given by a pair (N, w), where $N = \{1, \dots, n\}$ is a set of *players* and $w = (w_1, \dots, w_n)$ is a collection of *value functions*, one for each player. Each of these value functions $w_i : N \to \mathbb{R}$ assign real values to each player (with $w_i(i) = 0$ for every $i \in N$). Then, each player i's preferences over all $A, B \in \mathcal{N}_i$ is computed by

$$A \succeq_i B \iff \sum_{j \in A} w_i(j) \geq \sum_{j \in B} w_i(j),$$

yielding the corresponding hedonic game $(N, \succeq)$.

Example 2.2 *We continue with Example 2.1 and will show, that some, but not all preference relations can be expressed by the compact representation using the value functions of additively separable hedonic games. Therefore, we define an additively separable hedonic game $H_{\text{AS}} = (N, w)$ using the following values that define the value functions:*

	1	2	3
w_1 :	0	3	-6
w_2 :	5	0	-5
w_3 :	4	7	0

It is important to notice that for large numbers of players, this representation is much more compact than listing all possible subsets of $\mathcal{N}_i$ for each player i as in Example 2.1. However, this compact representation is not as expressive as the original representation is. This can easily be seen in the above example by deriving the real preferences from the information that the above values give us. Even though this results in the same preferences for player 2 *and* 3 *as in Example 2.1, the preference relation of player* 1 *is different, namely* $\{1, 2\} \succ_1 \{1\} \succ_1 \{1, 2, 3\} \succ_1 \{1, 3\}$. *We can even prove that it is not possible the find a value function for player* 1 *that represents the preference relation* $\{1, 2\} \succ_1 \{1\} \succ_1 \{1, 3\} \sim_1 \{1, 2, 3\}$ *from Example 2.1:* $\{1, 2\} \succ_1 \{1\}$ *indicates, that player* 2 *has to get a value strictly larger than zero in player* 1*'s value function, while this no longer allows the indifference* $\{1, 3\} \sim_1 \{1, 2, 3\}$ *to be derived for any value for* 3. *Therefore, we have shown that the original representation of hedonic games is strictly more expressive than the additively separable encoding.*

Another impactful representation is due to Dimitrov et al. [26]. It is based on so-called friend- and enemy-oriented preference extensions and provides a subclass of additively separable hedonic games. We distinguish between *friend-oriented hedonic games (FHG)* and *enemy-oriented hedonic games (EHG)*, and in both each player has to partition the other players into a set of friends and a set of enemies. Then, their preferences over two coalitions are then determined by the number of friends and enemies in these coalitions. Formally, every player i reports in both versions a set $F_i \subseteq N$, including herself, as her set of friends. $E_i = N \setminus F_i$ is then her automatically derived set of enemies. Let $A, B \in \mathcal{N}_i$, then, under *friend-oriented preferences*, $A \succeq_i B$ if $|A \cap F_i| > |B \cap F_i|$ (stating that A contains more friends than B) or if $|A \cap F_i| = |B \cap F_i|$ and $|A \cap E_i| \leq |B \cap E_i|$ (stating, that if the number of friends is equal, A contains at most as many enemies as B). For *enemy-oriented preferences*, we have $A \succeq_i B$ if $|A \cap E_i| < |B \cap E_i|$ (stating that A contains less enemies than B) or if $|A \cap E_i| = |B \cap E_i|$ and $|A \cap F_i| \geq |B \cap F_i|$ (stating that if the number of enemies is equal, A contains at least as many friends as B).

Example 2.3 *We will again take Example 2.1 as a reference. A friend- or enemy-oriented hedonic game is given by $H_{\text{FE}} = (N, F)$, while $F = (F_1, F_2, F_3)$ is a profile of the sets of friends of every player. Let, for example, $F_1 = \{1, 2\}$, $F_2 = \{1, 2\}$, and $F_3 = \{1, 2, 3\}$, then, the sets of enemies are automatically derived via $E_i = N \setminus F_i$, therefore resulting in $E_1 = \{3\}$, $E_2 = \{3\}$, and $E_3 = \emptyset$. As in Example 2.2, this representation is extremely compact, but less expressive than standard hedonic games. The latter follows immediately from the fact, that the friend- or enemy-oriented enncodings are a special case of the additively separable encoding. Simply set $w_i(j) = |N|$ for every $j \in F_i \setminus \{i\}$ and $w_i(j) = -1$ for every $j \in E_i$ in the friend-oriented case, and $w_i(j) = -|N|$ for every $j \in E_i$ and $w_i(j) = 1$ for every $j \in F_i \setminus \{i\}$ in the enemy-oriented case.*

In friend-oriented hedonic games, we focus on the number of friends in every coalition. The number of enemies is only taken into account if the number of friends is the same in the two compared coalitions. This results in the following preference relations:

$$\begin{aligned}
\succeq_1&: \ \{1,2\} \succ_1 \{1,2,3\} \succ_1 \{1\} \succ_1 \{1,3\} \\
\succeq_2&: \ \{1,2\} \succ_2 \{1,2,3\} \succ_2 \{2\} \succ_2 \{2,3\} \\
\succeq_3&: \ \{1,2,3\} \succ_3 \{1,3\} \sim_3 \{2,3\} \succ_3 \{3\}
\end{aligned}$$

In the enemy-oriented case, we only focus on the number of enemies, and the number of friends only matters in ties. This results in the following preference *relations:*

$$\begin{aligned}
\succeq_1&: \{1,2\} \succ_1 \{1\} \succ_1 \{1,2,3\} \succ_1 \{1,3\}\\
\succeq_2&: \{1,2\} \succ_2 \{2\} \succ_2 \{1,2,3\} \succ_2 \{2,3\}\\
\succeq_3&: \{1,2,3\} \succ_3 \{1,3\} \sim_3 \{2,3\} \succ_3 \{3\}
\end{aligned}$$

A different approach is taken by Cechlárová and Romero-Medina [25] (see also Cechlárová and Hajduková [23, 24]), who suggest the *singleton encoding*, i.e., each player $i \in N$ only has to provide a small part of her usual preference relation $\succeq_i$ that is equivalent to a ranking over all players. Formally, we assume in such *singleton encoded hedonic games (SHG)*, that every player $i \in N$ reports a preference relation $\succeq_i^{\text{SG}}$ over N[8]. This relation over N is formally equivalent to a total preference relation over $\{\{i,j\} \mid j \in N\}$, which itself corresponds to a partial, i.e., not total, preference relation over $\mathcal{N}_i$. Then, this relation $\succeq_i^{\text{SG}}$ over N is extended to a total relation over $\mathcal{N}_i$ the following way: For any coalition $A \in \mathcal{N}_i$, let $\mathcal{B}_i(A)$ be any *best player* $j \in A$ from i's view, i.e., $j \succeq_i^{\text{SG}} k$ for each $k \in A$; and let $\mathcal{W}_i(A)$ be any *worst player* $j \in A \setminus \{i\}$ from i's view, i.e., $k \succeq_i^{\text{SG}} j$ for each $k \in A$. (For the special case of $A = \{i\}$, let $\mathcal{W}_i(A) = i$.) Now, for any $A, B \in \mathcal{N}_i$, we say A *is* $\mathcal{B}$*-preferred by* i *over* B (stating $A \succeq_i B$ in the *best player* case) if $\mathcal{B}_i(A) \succ_i^{\text{SG}} \mathcal{B}_i(B)$ or if $\mathcal{B}_i(A) \sim_i^{\text{SG}} \mathcal{B}_i(B)$ and $|A| \leq |B|$, and we say A *is* $\mathcal{W}$*-preferred by* i *over* B (stating $A \succeq_i B$ in the *worst player* case) if $\mathcal{W}_i(A) \succeq_i^{\text{SG}} \mathcal{W}_i(B)$[9].

Example 2.4 *Let* $H_{\text{SG}} = (N, \succeq_i^{\text{SG}})$ *be a hedonic game with singleton encoding and* $N = \{1,2,3\}$*. Then, we receive a similar hedonic game to the one from Example 2.1 by letting*

$$\begin{aligned}
\succeq_1^{\text{SG}}&: 2 \succ_1^{\text{SG}} 1 \succ_1^{\text{SG}} 3,\\
\succeq_2^{\text{SG}}&: 1 \succ_2^{\text{SG}} 2 \succ_2^{\text{SG}} 3,\\
\succeq_3^{\text{SG}}&: 2 \succ_3^{\text{SG}} 1 \succ_3^{\text{SG}} 3.
\end{aligned}$$

With focus on the best player, this extends to the preference relations

$$\begin{aligned}
\succeq_1&: \{1,2\} \succ_1 \{1,2,3\} \succ_1 \{1\} \succ_1 \{1,3\},\\
\succeq_2&: \{1,2\} \succ_2 \{1,2,3\} \succ_2 \{2\} \succ_2 \{2,3\},\\
\succeq_3&: \{2,3\} \succ_3 \{1,2,3\} \succ_3 \{1,3\} \succ_3 \{3\},
\end{aligned}$$

[8] We use the superscript SG to refer to a preference relation over N, in contrast to the notion without superscript that refers to a preference relation over $\mathcal{N}_i$.

[9] Please note, that in [25] the definitions are slightly different. However, our definition is equivalent

and the worst player case results in

$$
\begin{aligned}
\succeq_1\colon &\quad \{1,2\} \succ_1 \{1\} \succ_1 \{1,3\} \sim_1 \{1,2,3\},\\
\succeq_2\colon &\quad \{1,2\} \succ_2 \{2\} \succ_2 \{2,3\} \sim_2 \{1,2,3\},\\
\succeq_3\colon &\quad \{2,3\} \succ_3 \{1,3\} \sim_3 \{1,2,3\} \succ_3 \{3\}.
\end{aligned}
$$

Again, we can prove that this singleton encoding is strictly less expressive than standard hedonic games. In the best player case this can easily be proven equivalently to our argumentation in Example 2.2: The original preference relation of player 1 *from Example 2.1 , namely* $\{1,2\} \succ_1 \{1\} \succ_1 \{1,3\} \sim_1 \{1,2,3\}$*, cannot be derived from any singleton encoded preference relation* $\succeq_1^{\text{SG}}$*, as* $\{1,2\} \succ_1 \{1\} \succ_1 \{1,3\}$ *is only achievable via the preferences* $2 \succ_1^{\text{SG}} 1 \succ_1^{\text{SG}} 3$ *or* $2 \succ_1^{\text{SG}} 1 \sim_1^{\text{SG}} 3$*, which stands in conflict with* $\{1,3\} \sim_1 \{1,2,3\}$*. For the worst player case we have to focus on player* 2 *with her preference relation* $\{1,2\} \succ_2 \{2\} \sim_2 \{1,2,3\} \succ_2 \{2,3\}$ *in Example 2.1: The partial preference* $\{1,2\} \succ_2 \{2\}$ *indicates* $1 \succ_2^{\text{SG}} 2$ *and the partial preference* $\{2\} \succ_2 \{2,3\}$ *indicates* $2 \succ_2^{\text{SG}} 3$*. In total we must have* $1 \succ_2^{\text{SG}} 2 \succ_2^{\text{SG}} 3$*, which does not lead to the partial preference* $\{2\} \sim_2 \{1,2,3\}$ *of player* 2*'s preference relation from Example 2.1.*

2.3.3 Stability Concepts

An important solution concept for the study of hedonic games is the notion of stability of a coalition structure. There are several known so-called stability concepts, and we can divide them into three major groups: One group that deals with avoiding a player to deviate to another (possibly empty) existing coalition (in the following all concepts from *perfectness* to *contractual individual stability*), another group that has the goal that there is no blocking coalition, i.e., no group of players that want to deviate together (e.g., *(strict) core stability* or *Pareto optimality*), and a third group that takes a global cardinal approach of securing some kind of stability (e.g., *(strict) popularity*). For more explanations on the general concept of stability in hedonic games, we refer the reader to the article by Bogomolnaia and Jackson [17] and to the book chapter of Aziz and Savani [5]. For the interested reader, we refer to the work of Banerjee et al. [6] for interesting properties and natural restrictions of hedonic games. In the following definition, we give a brief overview over well-known stability concepts. Please note, that the last concept was introduced in our work [40].

Definition 2.5 *Let $(N, \succeq)$ be a hedonic game. A coalition structure Γ is called*

- perfect *if each player i weakly prefers Γ_i to every other coalition containing i;*
- individually rational *if each player $i \in N$ weakly prefers Γ_i to being alone in $\{i\}$;*
- Nash stable *if for each player $i \in N$ and for each coalition $C \in \Gamma \cup \{\emptyset\}$, $\Gamma_i \succeq_i C \cup \{i\}$ (that is, no player wants to join another coalition);*
- individually stable *if for each player $i \in N$ and for each coalition $C \in \Gamma \cup \{\emptyset\}$, it holds that $\Gamma_i \succeq_i C \cup \{i\}$, or there exists a player $j \in C$ such that $C \succ_j C \cup \{i\}$ (that is, no player can join another coalition without making some player in the new coalition objecting to this switch);*
- contractually individually stable *if for each player $i \in N$ and for each coalition $C \in \Gamma \cup \{\emptyset\}$, it holds that $\Gamma_i \succeq_i C \cup \{i\}$, or there exists a player $j \in C$ such that $C \succ_j C \cup \{i\}$, or there exists a player $k \in \Gamma_i \setminus \{i\}$ such that $\Gamma_i \succ_k \Gamma_i \setminus \{i\}$ (that is, no player can join another coalition without making some player in the new coalition or in the old coalition objecting to this switch);*
- core stable *if for each coalition $C \subseteq N$, there exists a player $i \in C$ such that $\Gamma_i \succeq_i C$ (that is, no coalition $C \subseteq N$* blocks *Γ);*
- strictly core stable *if for each coalition $C \subseteq N$, there exists a player $i \in C$ such that $\Gamma_i \succ_i C$, or for each player $i \in C$, we have $\Gamma_i \sim_i C$ (that is, no coalition $C \subseteq N$* weakly blocks *Γ);*
- Pareto optimal *if for each coalition structure Δ, there exists a player $i \in N$ such that $\Gamma_i \succ_i \Delta_i$, or for each player $j \in N$, we have $\Gamma_j \sim_j \Delta_j$ (that is, no other coalition structure Δ* Pareto-dominates *Γ);*
- popular *if for each coalition structure Δ, the number of players i with $\Gamma_i \succ_i \Delta_i$ is at least as large as the number of players j with $\Delta_j \succ_j \Gamma_j$;*
- strictly popular *if for each coalition structure Δ, the number of players i with $\Gamma_i \succ_i \Delta_i$ is strictly larger than the number of players j with $\Delta_j \succ_j \Gamma_j$;*

Example 2.6 *To better understand the different stability concepts for hedonic games, we take a look at the game H from Example 2.1, where we have* *the three players* $1, 2,$ *and* 3, *and the preference profile* $\succeq = (\succeq_1, \succeq_2, \succeq_3)$ *with the individual preferences*

$$\begin{aligned}
\succeq_1 &: \{1,2\} \succ_1 \{1\} \succ_1 \{1,3\} \sim_1 \{1,2,3\}, \\
\succeq_2 &: \{1,2\} \succ_2 \{2\} \sim_2 \{1,2,3\} \succ_2 \{2,3\}, \quad \textit{and} \\
\succeq_3 &: \{1,2,3\} \succ_3 \{2,3\} \succ_3 \{1,3\} \succ_3 \{3\}.
\end{aligned}$$

It seems reasonable to consider the coalition structure $\Gamma = \{\{1,2\},\{3\}\}$, *as both, player* 1 *and player* 2, *prefer* $\Gamma_1 = \Gamma_2 = \{1,2\}$ *to all other possible coalitions. We will now investigate* Γ *in regards to the stability concepts defined above:*

- Perfectness: Γ *is not perfect, as player* 3 *does not prefer* $\Gamma_3 = \{3\}$ *to every other coalition from* $\mathcal{N}_3$. *In fact, there cannot exist a perfect coalition structure in this example, as never both,* $\{1,2\}$ *and* $\{1,2,3\}$, *can be part of a coalition structure.*

- Individual rationality: Γ *is individually rational, as player* 1 *and* 2 *prefer* $\Gamma_1 = \Gamma_2$ *to being alone, and player* 3 *obviously is indifferent between* Γ_3 *and* $\{3\}$, *as both coalitions are the same.*

- Nash stability: Γ *is not Nash stable, as player* 3 *would prefer to join* $\{1,2\}$. *In fact, there cannot exist a perfect coalition structure in this example, as player* 1 *and* 2 *both prefer being in a coalition without* 3, *while player* 3 *always wants to join* 1 *or* 2 *or both.*

- Individual stability: Γ *is individually stable, as* 1 *and* 2 *are already in their most preferred coalition, and* 3 *cannot join* $\{1,2\}$, *as both other players would be worse off with the grand coalition.*

- Contractual individual stability: Γ *is, with the same argumentation as used for individual stability, contractually individually stable.*

- Core stability: Γ *is core stable, as there cannot exist a blocking coalition containing player* 1 *or* 2, *as they both already are in their most preferred coalition. The only remaining possibility for a blocking coalition is* $\{3\}$, *which is already part of* Γ.

- Strict core stability: Γ *is, with the same argumentation as used for core stability and the fact that all involved rankings are strict, strictly core stable.*

- Pareto optimality: Γ *is Pareto optimal, as for all coalition structures* Δ *that do not contain* $\{1, 2\}$ *both, player* 1 *and* 2*, prefer their coalition* $\{1, 2\}$ *in* Γ *to any possible coalition* Δ_1*, respectively* Δ_2*, and for all coalition structures* Δ *that contain* $\{1, 2\}$*, player* 3 *must be alone, which leads to* $\Gamma = \Delta$.

- Popularity: Γ *is popular, as for all coalition structures* Δ *that do not contain* $\{1, 2\}$*, two out of three players (namely player* 1 *and player* 2*) prefer their coalition in* Γ *to their coalition in* Δ*, and for all coalition structures* Δ *that contain* $\{1, 2\}$*, player* 3 *must be alone, which leads to* $\Gamma = \Delta$.

- Strict popularity: Γ *is, with the same argumentation as used for popularity and the fact that all involved rankings are strict, strictly popular.*

Example 2.6 indicates a connection between these stability concepts (except for (strict) popularity, which is as a cardinal approach not connected to the other, ordinal ideas). A perfect coalition structure is, for example, also Nash stability and a member of the strict core. All connections are illustrated in Figure 2.1, in which an arrow symbolizes an implication. If there is no arrow between two concepts, this does not mean that it is impossible for a single coalition structure to fulfill both criteria at the same time. In fact, there are hedonic games in which a single coalition structure fulfills all concepts simultaneously.

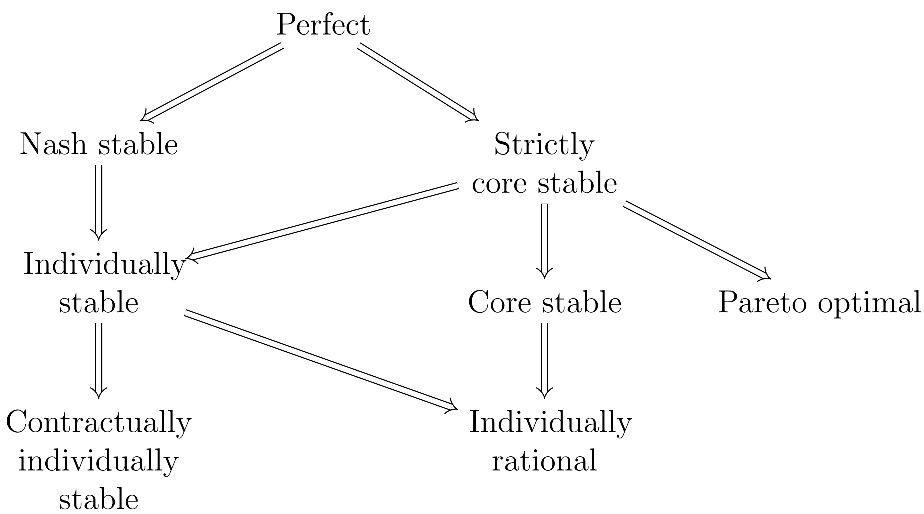


Figure 2.1: Relations among various stability concepts for hedonic games

2.3.4 Computational Results

Natural decision problems from the field of hedonic games are usually tied to a specific stability concept and a specific encoding. However, many papers focus on a specific encoding and investigate it under several stability concepts. Therefore, the name of the most prominent decision problems only specify the investigated stability concept, but not the used encoding, as it usually is clear from the context. One of the most basic problems is the verification problem, which asks for a given hedonic game and a coalition structure whether the coalition structure satisfies the stability concept γ:

γ-Verification	
Given:	A hedonic game H and a coalition structure Γ.
Question:	Does Γ satisfy γ?

Obviously, this decision problem can be investigated for each mentioned encoding separately simply by restricting the hedonic game of the input to be of a fixed encoding. The same holds true for the following decision problems, which asks whether a given hedonic game has at least one coalition structure that satisfies the fixed stability concept γ:

γ-Existence	
Given:	A hedonic game H.
Question:	Does there exist a coalition structure that satisfies γ?

We know that if for a stability concept γ the problem γ-Verification is in P, then γ-Existence belongs to NP by simply guessing a coalition structure and verifying it in polynomial time with the algorithm for the verification problem. The difficulty of the respective problems range from trivial (for instance, for friend- or enemy-oriented hedonic games, in which there always exists a core stable coalition structure) to Σ_2^p-completeness (for instance, for core stability in additively-separable hedonic games). For an overview of known results we refer to the comprehensive survey by Woeginger [74].

2.4 Abstract Argumentation

Two people communicating in a such a way that one tries to convince the other of some opinion, is already an exquisite example for the basic idea of the field of argumentation, which is naturally part of every society. The founding of argumentation theory, a field that tries to formally capture the ideas of persuasion and discussion, probably goes hand in hand with the peak stage of philosophy in the Greek society, and is until today a highly interesting field for researchers from the whole scientific spectrum. Many scientists from mathematics and computer science are mainly interested in the abstract and computational part of argumentation theory, and therefore, call the respective subfield *abstract argumentation*. In abstract argumentation we try to explain chosen parts of a discussion process with the help of mathematical models and a formal analysis.

In 1995 Phan Minh Dung [29] revolutionized the idea of analyzing discussions just by abstraction from the content of the given arguments. If, for example, in some family the father says "I don't want dogs, because they are dirty" this would just translate to a placeholder argument with its variable name a. However, Dung suggest to use the inside and its internal structure to derive an attack relation between arguments. Let us assume, the daughter answers with "But I want a dog, because they are so cute and fluffy", then, in Dung's setting, we could just call this argument b and derive a mutual attack between those two arguments a and b, as they exclude each other, i.e., we cannot expect both to be part of a suitable solution of the discussion.

In the next step, Dung proposed to define semantics, i.e., collections of properties that can be fulfilled by subsets of the argument set, with the goal to identify arguments that are, in some sense, stronger than others, such that they can be accepted, while others have to be rejected. The most basic property—already mentioned in the above example—is called *conflict-freeness*, and it is fulfilled by any subset of the arguments that only contains arguments that do not attack another element of that subset. All other properties extend conflict-freeness and describe more elaborated concepts.

Abstract argumentation is a field that is far from being completed, and is therefore highly interesting. However, how deep and complicated research in this field may get, the basic notions are extremely simple. We will explain those basic notions together with some fundamental connections in the following paragraph.

2.4.1 Preliminary Definitions

In this section, we give formalizations of the basic notions of abstract argumentation. While we adopt some notation from the book chapter by Dunne and Wooldridge [30], the underlying concepts are due to Dung [29]. For several more explanations and ideas regarding abstract argumentation, we refer the reader to the book of Rahwan and Simari [52].

Definition 2.7 *An* argumentation framework AF *consists of a set of* arguments $\mathcal{A}$ *and binary relation* $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$*, thus, forming a pair* $\langle \mathcal{A}, \mathcal{R} \rangle$*. We say that* a attacks b *if* $(a, b) \in \mathcal{R}$*.*

A graph is a pair of a set of vertices V and a set of directed edges E on these vertices V. Therefore, an argumentation framework can be visualized easily and directly via a graph $G_{AF} = (V, E)$, by identifying arguments with vertices and attacks with directed edges, i.e., $V = \mathcal{A}$ and $E = \mathcal{R}$, as in the following example:

Example 2.8 *Let us assume a simple argumentation with three arguments: The two abstract arguments from our introductory example regarding a family discussing the necessity of getting a dog, and a third argument c from the mother telling her husband, that "I will take care of the additional dirt." This results in the argumentation framework* $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ *with* $\mathcal{A} = \{a, b, c\}$ *and* $\mathcal{R} = \{(a, b), (b, a), (c, a)\}$*. Then, Figure 2.2 displays the graph representation of the argumentation framework.*

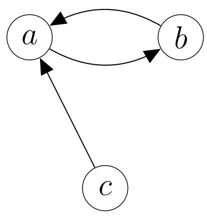


Figure 2.2: A simple argumentation framework

Dung has, as already mentioned, introduced in his seminal paper [29] semantics, which have been defined to evaluate the acceptability status of sets of arguments. The following definition contains his ideas, that became fundamental for the analysis of abstract argumentation.

Definition 2.9 *Let* $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ *be an argumentation framework. A set* $S \subseteq \mathcal{A}$ *is called*

- conflict-free *if* $\forall a, b \in S$ *it holds that* $(a, b) \notin \mathcal{R}$*, i.e., no argument in* S *attacks an argument in* S*,*
- admissible *if* S *is conflict-free and* $\forall b \in \mathcal{A}$*,* $a \in S$ *and* $(b, a) \in \mathcal{R}$ *it holds that* $\exists c \in S$ *with* $(c, b) \in \mathcal{R}$*, i.e., every argument in* S *is* defended *by an argument in* S *against incoming attacks,*

- preferred *if S is admissible and $\nexists S' \subseteq \mathcal{A}$ with S' is admissible and $S \subset S'$, i.e., S is a maximal (with respect to set inclusion) admissible set,*

- stable *if S is conflict-free and $\forall b \in \mathcal{A} \setminus S$ it holds that $\exists a \in S$ with $(a, b) \in \mathcal{R}$, i.e., every argument outside of S is attacked by an argument inside of S,*

- complete *if S is admissible and, $\forall a \in \mathcal{A}$, if a is defended by an argument in S it holds that $a \in S$, i.e., all arguments that are successfully defended by arguments in S also belong to S, and*

- grounded *if $S = F^*_{AF}(\emptyset)$, where $F_{AF} : 2^{\mathcal{A}} \rightarrow 2^{\mathcal{A}}$ is the* characteristic function *of AF, defined by*

  $$F_{AF}(S) = \{a \in \mathcal{A} \mid a \text{ is defended by an argument of } S\}, \text{ and}$$

  $F^*_{AF}(\emptyset)$ is the least fixed point of F^*_{AF}.

Since the characteristic function is monotonic with respect to set inclusion if applied to admissible sets, i.e., $S \subseteq F_{AF}(S)$, there always is a least fixed point, therefore securing the existence of a (unique) grounded set. Admissibility and completeness can also be defined via the characteristic function: If a subset of the arguments S is conflict-free and $S \subseteq F_{AF}(S)$ holds, then it is admissible, and if $S = F_{AF}(S)$ holds, then it is complete. The latter also states, that the complete sets of an argumentation framework are exactly the fixed points of F_{AF}—in particular this implies, that the grounded set is complete. Dung [29] also proved several other correlations between semantics, that can be easily verified with the help of the characteristic function. Among others, he showed that every admissible set is a subset of a preferred set, that there always is at least one (maybe empty) preferred set, that every stable set is preferred, and every preferred set is complete. It is not hard to prove that a preferred or grounded set does not have to be stable, and it is easy to show that each of the above defined semantics secures conflict-freeness and admissibility. Figure 2.3 displays all relations among these semantics. The arrow from ST to PR indicates, for example, that all sets of arguments that are stable must also be preferred, and so on. If there is no arrow between two semantics this does not mean, that it is impossible for a subset to fulfill both semantics. It is, for example, possible that one single argument set fulfills all semantics simultaneously.

Dung [29] also uses the notion of *extensions* of an argumentation framework as a term for those sets that fulfill the criteria of a semantics. This

$$\begin{array}{ccccccc} \text{Stable} & & \text{Grounded} & & & & \\ \Downarrow & & \Downarrow & & & & \\ \text{Preferred} & \Longrightarrow & \text{Complete} & \Longrightarrow & \text{Admissible} & \Longrightarrow & \text{Conflict-free} \end{array}$$

Figure 2.3: Relations among various semantics for sets of arguments

means, that an argument set is called an **s**-extension, if it fulfills the criteria of the semantics **s**. However, Dung does not consider conflict-freeness and admissibility to be semantics, as those are basic requirements in his eyes. As a result, he also does not call conflict-free or admissible sets "extensions". However, for convenience, we might do this sometimes.

Example 2.10 *Example 2.8 has exactly four conflict-free extensions, namely all three singletons* $\{a\}, \{b\}$, *and* $\{c\}$ *and the pair* $\{b, c\}$. *No argument set that contains argument a and another arbitrary argument can be conflict-free, as a attacks b and is attacked by b and c. Among those conflict-free extensions, all but* $\{a\}$ *are also admissible, as they directly defend any incoming attack, if one exists. Please note, that it especially is okay that in* $\{b\}$ *the incoming attack* $(a, b) \in \mathcal{R}$ *is defended by b itself through* $(b, a) \in \mathcal{R}$. *In this example exists only one preferred extension, one complete extension, and one stable extension, which all coincide with the unique grounded extension* $\{b, c\}$.

2.4.2 Computational Results

In the field of abstract argumentation we can naturally find decision problems from many common complexity classes. Here, complexity does not only depend on the problem's structure, but also mainly on the investigated semantics. Let us have a look at the definition of standard decision problems for abstract argumentation. The probably most basic one gets an argumentation framework and a subset of the arguments as input, and the question is whether the given subset is an extension for the a priori fixed semantics **s**:

s-VERIFICATION	
Given:	An argumentation framework $\langle \mathcal{A}, \mathcal{R} \rangle$ and a subset $S \subseteq \mathcal{A}$.
Question:	Is S an **s** extension of AF?

In this work, more specifically in Chapter 6, we only focus on the six semantics from the previous paragraph, and for better readability we write

CF for *conflict-freeness*, AD for *admissibility*, PR for *preferredness*, ST for *stability*, CP for *completeness*, and GR for *groundedness*.

Dunne and Wooldridge [30] surveyed several decision problems, including the verification problem from above, and they also mention known complexity results. This includes the membership of **s**-VERIFICATION in P for all mentioned semantics except for PR, for which the respective decision problem is coNP-complete, shown by Dimopoulos and Torres [27]. The complexity of other problems mentioned in [30] go up to Π_2^p-completeness, suggesting that abstract argumentation bears very hard problems. In Chapter 6, we naturally extend the basic model of argumentation frameworks and investigate corresponding versions of the verification problem, and also obtain, among others, hardness results for the classes of the second level of the boolean hierarchy.

CHAPTER 3

Toward the Complexity of the Existence of Wonderfully Stable Partitions and Strictly Core Stable Coalition Structures in Enemy-Oriented Hedonic Games

Summary

In this paper we discuss the computational complexity of several decision problems based on hedonic games in the restricting case that each agent i does not provide complete preferences over all possible coalitions, but instead a single subset F_i of the agents N that she would call her *friends*, including herself. This subset contains all players of the game that she would like to cooperate with. All other agents belong to $E_i = N \setminus F_i$, the set of the enemies of agent i. Then, we use an extension principle to extend this information to a preference ranking of all possible coalitions containing player i, therefore creating a classical hedonic game. This representation of a hedonic game is very compact, yet not fully expressive, as there are hedonic games that are not representable this way. In this paper, we concentrate on the extension principle that focuses on the number of enemies in the coalitions; resulting games are called *enemy-oriented hedonic games*. It is a special case of the additive separable representation. We have chosen this representation for this paper, as the literature does not provide sufficient results for the investigated decision problems, i.e., the question of the existence of a strict core stable

coalition structure in the given hedonic game.

Woeginger [74] already suggested upper and lower bounds for some of the investigated decision problems, as well a connection of strict core stability in enemy-oriented hedonic games and the purely graph theoretic concept of wonderfully stable partitions in undirected graphs. The connection is made by identifying the vertices with the players and the arcs with all the symmetric friendship relations, which origins in the fact that, when investigating the strict core in enemy-oriented hedonic games, only mutual friendships matter. In our paper we continue his research on this connection and also on upper and lower bounds, and are able to tighten the bounds for both major decision problems up to the fact that it remains to show hardness for DP to establish hardness for Θ_2^p.

Contribution and Preceding Versions

The idea, model, and writing was done jointly with my coauthors, as well as Lemma 1, Property 1 and the quantifier representations of the investigated decision problems, Theorems 1, 2, 5, 6 and the proof of Theorem 3. Theorems 4, 7 and 9, and Proposition 1 is part of my contribution. This paper merges and extends the preliminary papers [54] and [55].

Publication - Rey, Rothe, Schadrack, and Schend [56]

A. Rey, J. Rothe, H. Schadrack, and L. Schend. Toward the complexity of the existence of wonderfully stable partitions and strictly core stable coalition structures in enemy-oriented hedonic games. *Annals of Mathematics and Artificial Intelligence*, 77(3-4):317–333, 2016

Ann Math Artif Intell (2016) 77:317–333
DOI 10.1007/s10472-015-9461-y

Toward the complexity of the existence of wonderfully stable partitions and strictly core stable coalition structures in enemy-oriented hedonic games

Anja Rey[1] · **Jörg Rothe**[1] · **Hilmar Schadrack**[1] ·
Lena Schend[1]

Published online: 2 June 2015

Abstract We study the computational complexity of the existence and the verification problem for wonderfully stable partitions (WSPE and WSPV) and of the existence problem for strictly core stable coalition structures (SCSCS) in enemy-oriented hedonic games. In this note, we show that WSPV is NP-complete and both WSPE and SCSCS are DP-hard, where DP is the second level of the boolean hierarchy, and we discuss an approach for classifying the latter two problems in terms of their complexity.

Keywords Game Theory · Hedonic games · Strict core stability · Wonderful stability

Mathematics Subject Classification (2010) 68Q15 · 68Q17 · 91A12

1 Introduction

Hedonic games are an interesting model combining the central ideas of, on the one hand, cooperative game theory (see, e.g., the textbooks by Peleg and Sudhölter [25] and Chalkiadakis et al. [9]) where players form coalitions in order to manage certain tasks as a team, and, on the other hand, voting scenarios (see, e.g., the book chapters by Brams and Fishburn [5] and Brandt et al. [6]) where players give their preferences over several alterna-

✉ Lena Schend
schend@cs.uni-duesseldorf.de

Anja Rey
rey@cs.uni-duesseldorf.de

Jörg Rothe
rothe@cs.uni-duesseldorf.de

Hilmar Schadrack
schadrack@cs.uni-duesseldorf.de

1 Institut für Informatik, Heinrich-Heine-Universität Düsseldorf 40225 Düsseldorf, Germany

tives in order to elect mutually desirable alternatives by aggregating their preferences. In a hedonic game, the alternatives are groups (coalitions) of players and players "vote" on coalitions they want to join by expressing their preferences. Hedonic games have been studied from a computational perspective, for example by Dimitrov et al. [11], Sung and Dimitrov [32], Aziz et al. [1] and Woeginger [37]. In his survey, Woeginger [36] gives an overview of several core stability concepts in hedonic games and their analysis.

We in particular focus on the concepts of wonderfully stable partitions and strictly core stable coalition structures that have been considered in this survey. A partition of the vertices of an undirected graph is called *wonderfully stable* if each vertex is assigned to a clique of largest size that contains the vertex. In the context of hedonic games, this notion can be interpreted to express the following scenario. If the players are represented by the vertices in a graph and there is an undirected edge between two vertices if and only if the two related players like each other, then—under so-called enemy-oriented preferences [11]—a largest clique corresponds to the coalition that is most preferred by each player in the coalition, among those coalitions not containing any enemies. A wonderfully stable partition for this graph hence corresponds to a coalition structure where each player ends up in her most preferred coalition among those without enemies. In the same domain, intuitively, a coalition structure is *(strictly) core stable* if no group of players has an incentive to form a different coalition, thus breaking away from the given coalition structure.

1.1 Related work and our contribution

Besides enemy-oriented preferences, there are several other ways to represent a hedonic game compactly. Additively separable hedonic games, for example, are represented by numerical values for each player evaluating each other player; preferences of a player over coalitions are derived from the particular sum of values of this player for the players in a coalition. It is known that, for additively separable hedonic games, the problem as to whether a given coalition structure is core stable is NP-complete [32], and the corresponding problem of whether such a coalition structure exists in a given game was first shown to be NP-hard by Sung and Dimitrov [33], even for the case of symmetric additive preferences (see the work of Aziz et al. [1]), and was finally shown to be Σ_2^p-complete by Woeginger [37]. For friend-oriented preferences—defined similarly to enemy-oriented preferences—it is known that there always exists a core-stable partition [11]. Under enemy-oriented preferences, there always exists a core stable coalition structure in a given game [11], and deciding whether a given coalition structure is core stable or strictly core stable is strongly NP-complete [32, 36].

Let WSPE be the problem of deciding whether there exists a wonderfully stable partition in a given graph, and let SCSCS be the problem of deciding whether there exists a strictly core stable coalition structure in a given enemy-oriented hedonic game. The exact complexity of these problems is unknown so far. Woeginger [36] points out that these interesting open issues might be difficult to solve. The best known upper bounds are Θ_2^p for WSPE and Σ_2^p for SCSCS (where Θ_2^p and Σ_2^p are levels of the polynomial hierarchy), and Woeginger [36] conjectures that they are complete for these classes.

Raising the known lower bounds, we establish DP-hardness for both problems, where DP is the second level of the boolean hierarchy over NP. This is a first step toward classifying these two problems in terms of their complexity. We also provide arguments for why they cannot be complete for any level of the boolean hierarchy higher than the second level (unless this hierarchy collapses, which is considered unlikely). Moreover, we show that proving coDP-hardness for them would already suffice to establish their Θ_2^p-hardness.

2 Preliminaries

In this section, we introduce the concept of hedonic game, describe links between such games and appropriate graph-theoretic concepts, define the corresponding stability concepts, and give the needed background from complexity theory.

2.1 Hedonic games

A *hedonic game* consists of a finite set $N = \{1, \dots, n\}$ of *players* and a profile $\succeq = (\succeq_1, \dots, \succeq_n)$ of *preference relations*, where $\succeq_i$ denotes player i's preference relation. Each such preference relation $\succeq_i$ defines a weak preference order over all *coalitions* (i.e., subsets of N) that contain player i. Let A and B be coalitions containing i. We say that i weakly prefers A to B if $A \succeq_i B$, and we say i prefers A to B (denoted by $A \succ_i B$) if $A \succeq_i B$, but not $B \succeq_i A$.

Since the number of coalitions in a player's preference order is exponential in the number of players, it is reasonable to consider compactly represented hedonic games; see the survey of Woeginger [36] for an overview of various possible encodings. We consider so-called enemy-oriented preferences as introduced by Dimitrov et al. [11]. In their setting, every player $i \in N$ has a set of friends and a set of enemies, and that is all that is needed to represent i's preferences over all coalitions.

Definition 1 For a set $N = \{1, \dots, n\}$ of players, define the enemy-oriented preference profile $\succeq = (\succeq_1, \dots, \succeq_n)$ of a hedonic game $\mathcal{G} = (N, \succeq)$ as follows. Let $i \in N$ be a player with friends $F_i \subseteq N$ (including i herself) and enemies $E_i = N \setminus F_i$, and let $A, B \subseteq N$ be two coalitions that both contain i.

1. We say *i weakly prefers A to B under enemy-oriented preferences* (denoted by $A \succeq_i B$) if $|A \cap E_i| < |B \cap E_i|$ (i.e., i has fewer enemies in A than in B), or $|A \cap E_i| = |B \cap E_i|$ and $|A \cap F_i| \geq |B \cap F_i|$ (i.e., i has the same number of enemies in A and in B, but at least as many friends in A as in B).
2. We say *i prefers A to B under enemy-oriented preferences* if $A \succ_i B$.

In the following, we often omit the phrase "under enemy-oriented preferences" and simply say that a player *prefers* or *weakly prefers* one coalition to another.

An enemy-oriented hedonic game can be represented by an undirected[1] graph G, where the set $N = \{1, \dots, n\}$ of players corresponds to the vertex set $V = \{v_1, \dots, v_n\}$ of G, and for each $i, j \in N$, $i \neq j$, there is an edge $\{v_i, v_j\}$ in G if and only if i and j are friends. A *clique in G* is a subset $C \subseteq V$ such that each two distinct vertices in C are connected by an edge. For each vertex v of G, let $\omega_G(v)$ denote the *clique number of v in G*, which is the size of a largest clique in G that contains v.

A *coalition structure for a hedonic game* $\mathcal{G} = (N, \succeq)$ is a partition $\Gamma = \{C_1, \dots, C_k\}$ of the players into $k \geq 1$ coalitions $C_1, \dots, C_k \subseteq N$ (i.e., $\bigcup_{i=1}^{k} C_i = N$ and $C_i \cap C_j = \emptyset$ for $i \neq j$). For a coalition structure Γ, we denote the coalition that contains player i by $\Gamma(i)$. In the associated graph G, a coalition structure corresponds to a *partition* Π of the vertices of G, and we denote the set in Π that contains a vertex v_i by $\Pi(v_i)$.

[1] As Woeginger [36] points out, in the context of stability only symmetric friendship relations matter in the enemy-oriented scenario, so we assume that a player $j \in N$ is player i's friend if and only if i is j's friend.

2.2 Stability concepts

We consider the following stability concepts for hedonic games (see the survey by Woeginger [36] for more details).

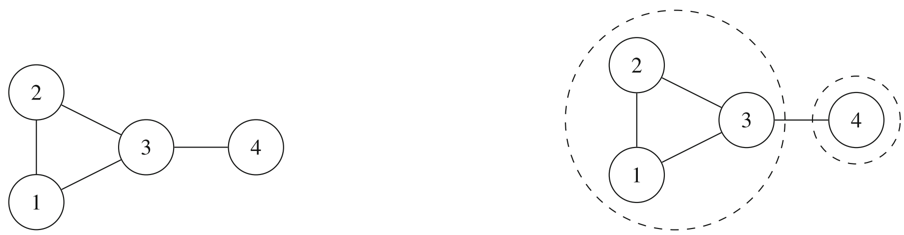
Definition 2 1. A coalition $C \subseteq N$ *blocks a coalition structure* Γ if each player $i \in C$ prefers C to $\Gamma(i)$ (i.e., $C \succ_i \Gamma(i)$).
2. A coalition structure Γ is *core stable* if there is no nonempty coalition $C \subseteq N$ that blocks Γ.
3. A coalition $C \subseteq N$ *weakly blocks a coalition structure* Γ if each player $i \in C$ weakly prefers C to $\Gamma(i)$ (i.e., $C \succeq_i \Gamma(i)$), and at least one player $j \in C$ prefers C to $\Gamma(j)$ (i.e., $C \succ_j \Gamma(j)$).
4. A coalition structure Γ is *strictly core stable* if there is no coalition $C \subseteq N$ that weakly blocks Γ.

Example 1 Consider the hedonic game $(\{1, 2, 3, 4\}, \succeq)$ with four players that have enemy-oriented preferences, given by their sets of friends $F_1 = \{1, 2, 3\}$, $F_2 = \{1, 2, 3\}$, $F_3 = \{1, 2, 3, 4\}$, and $F_4 = \{3, 4\}$. Figure 1a shows the graph G corresponding to this game. Now consider the coalition structure $\Gamma = \{\{1, 2, 3\}, \{4\}\}$ that is illustrated by the dashed lines in Fig. 1b. Γ is a strictly core stable coalition structure: Players 1 and 2 are in their unique most preferred coalition; thus they cannot be part of any weakly blocking coalition for Γ. Coalition $\{3, 4\}$ does not block Γ because of player 3 who prefers her coalition $\Gamma(3) = \{1, 2, 3\}$ to $\{3, 4\}$, since both coalitions have the same number of 3's enemies (namely, none—kind-hearted 3 is enemies with no one) but the former contains more of 3's friends. Finally, both player 3 and player 4 do not prefer to be alone under enemy-oriented preferences. That is, coalition $\{3\}$ does not weakly block Γ because of $\Gamma(3) = \{1, 2, 3\} \succ_3 \{3\}$, and coalition $\{4\}$ does not weakly block Γ because for its only member, player 4, it is not true that $\{4\} \succ_4 \{4\} = \Gamma(4)$ (even though, of course, $\{4\} \succeq_4 \{4\} = \Gamma(4)$ does hold).

Note that in a hedonic game with enemy-oriented preferences, a core stable coalition structure always corresponds to a partition into cliques in the associated graph. Recall from Section 1 that the concept of wonderfully stable partition in hedonic games has a purely graph-theoretic interpretation:

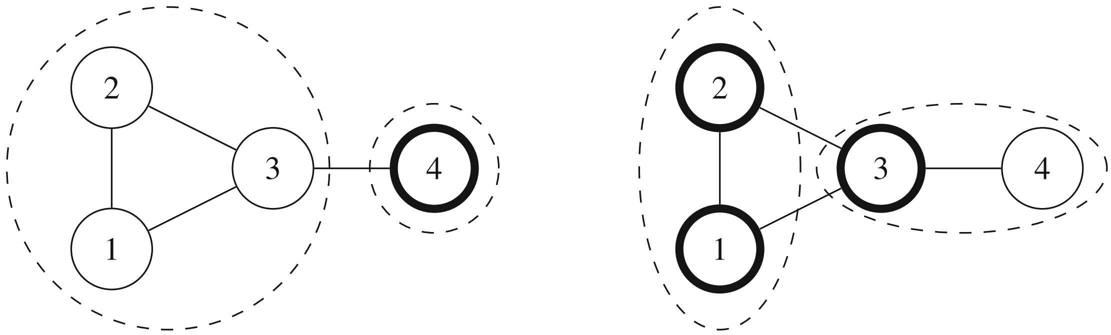
Definition 3 Given a graph $G = (V, E)$, a partition Π of the vertex set of G is called *wonderfully stable* if each $P \in \Pi$ is a clique and $|\Pi(v)| = \omega_G(v)$ for each vertex $v \in V$.

Adopting the notation from core stability in hedonic games, we say that a clique $P \subseteq V$ *blocks* a partition Π into cliques if there exists a vertex $v \in P$ with $\omega_G(v) > |\Pi(v)|$.



(a) Graph G corresponding to the game $(\{1,2,3,4\}, \succeq)$ (b) Strictly core stable coalition structure for the game $(\{1,2,3,4\}, \succeq)$

Fig. 1 Graph G corresponding to a game with a strictly core stable coalition structure



(a) Π_1 is blocked by clique $\{4\}$ (b) Π_2 is blocked by cliques $\{1,2\}$ and $\{3,4\}$

Fig. 2 Graph G that does not have a wonderfully stable partition

By definition of clique number, $\omega_G(v) \geq |\Pi(v)|$ for each vertex $v \in V$, since $\Pi(v)$ is a clique that contains v. Furthermore, note that the problem of whether there exists a partition into a limited number of cliques in a graph is NP-hard (see, e.g., the book by Garey and Johnson [12]). If, however, the number of cliques is not limited, a partition into cliques can easily be found.

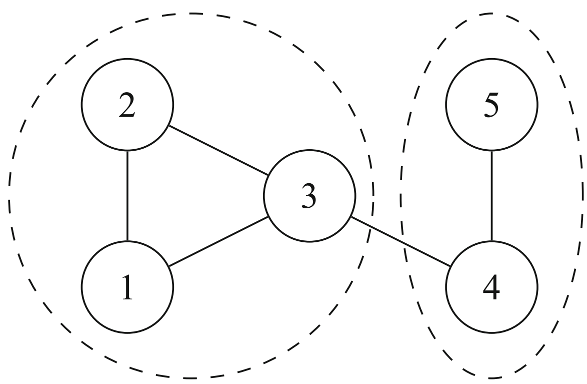
Example 2 Recall graph G from Fig. 1, which corresponds to the hedonic game defined in Example 1. We can see that the vertices 1,2, and 3 each have a clique number of 3, and vertex 4 has a clique number of 2. Figure 2 shows two possible partitions into cliques, $\Pi_1 = \{\{1, 2, 3\}, \{4\}\}$ and $\Pi_2 = \{\{1, 2\}, \{3, 4\}\}$. Neither of them is wonderfully stable. In Π_1, which is shown in Fig. 2a, vertex 4 forms a 1-clique in partition Π_1 and is thus blocking it. In Π_2, on the other hand, we have that the vertices 1, 2, and 3 each are in a 2-clique (see Fig. 2b), and the 2-cliques $\{1, 2\}$ and $\{3, 4\}$ both block the partition Π_2. The boldfaced vertices in Fig. 2 indicate that these vertices are not in a maximum-size clique containing them.

Now consider graph G' and the partition Π into cliques indicated by the dashed lines, both shown in Fig. 3. This partition is wonderfully stable since every vertex is in a clique of maximum size.

The following lemma provides a relation between strictly core stable coalition structures and wonderfully stable partitions.

Lemma 1 *Let $G = (V, E)$ be the graph representing an enemy-oriented hedonic game $\mathcal{G}$. Let Π be a partition of V and let Γ be the corresponding coalition structure in $\mathcal{G}$.*

Fig. 3 Graph G' that has a wonderfully stable partition Π



1. *If* Π *is a wonderfully stable partition for* G*, then* Γ *is a strictly core stable coalition structure for* $\mathcal{G}$*.*
2. *If there is an integer* $c \in \mathbb{N}$ *such that* $\omega_G(v) = c$ *for all vertices* $v \in V$ *and* Γ *is a strictly core stable coalition structure for* $\mathcal{G}$*, then* Π *is a wonderfully stable partition for* G*.*

Proof The first implication holds by definition: If a coalition C weakly blocks a coalition structure that corresponds to a partition into cliques, C has to be a clique with a larger cardinality and hence blocks the partition.

Second, assume that there is a blocking clique C for Π, i.e., there exists some vertex $v_i \in C$ with $\omega_G(v_i) > |\Pi(v_i)|$. Since $\omega_G(v_i) = c$, there is a clique D with $C \subseteq D$ and $|D| = c$. Now, the corresponding coalition $\tilde{D} = \{i \mid v_i \in D\}$ is a weakly blocking coalition for Γ, because $\tilde{D} \succ_i \Gamma(i)$ and $\tilde{D} \succeq_j \Gamma(j)$ for each $j \in \tilde{D}$, which follows from the fact that the number of friends in $\Gamma(i)$ is at most $c - 1$ and the number of friends in $\Gamma(j)$ is at most c, respectively. □

Note the following useful property that holds by definition for graphs consisting of several independent components.

Property 1 Let G be the graph representing an enemy-oriented hedonic game $\mathcal{G}$ and let G consist of k independent components G_i, $1 \leq i \leq k$, corresponding to games $\mathcal{G}_i$. There exists a wonderfully stable partition Π for G (respectively, a strictly core stable coalition structure Γ for $\mathcal{G}$) if and only if there exist wonderfully stable partitions Π_i for all components G_i of G (respectively, strictly core stable coalition structures Γ_i for all games $\mathcal{G}_i$), $1 \leq i \leq k$.

We will analyze the following decision problems.

Strictly Core Stable Coalition Structure (SCSCS)	
Given:	A hedonic game $\mathcal{G} = (N, \succeq)$ with enemy-oriented preferences.
Question:	Does there exist a strictly core stable coalition structure in G?
Wonderfully Stable Partition Existence (WSPE)	
Given:	A graph $G = (V, E)$.
Question:	Does there exist a wonderfully stable partition of V for G?
Wonderfully Stable Partition Verification (WSPV)	
Given:	A graph $G = (V, E)$ and a partition Π of V into cliques.
Question:	Does there exist a clique $P \subseteq V$ that blocks Π?

Just as the (existence and verification) core stability problems considered by Woeginger [36], the latter two problems are, by definition, related to each other. The verification problem can be characterized by an existential quantifier, and the existence problem can be characterized by an existential quantifier followed by a universal quantifier:

$$(G, \Pi) \in \text{WSPV} \iff (\exists P)\,[P \text{ blocks } \Pi], \tag{1}$$

$$G \in \text{WSPE} \iff (\exists \Pi)\,(\forall P)\,[\neg(P \text{ blocks } \Pi)]. \tag{2}$$

2.3 Complexity theory

We assume the reader is familiar with the basic notions of complexity theory, such as the complexity classes P, NP, and coNP and the notions of hardness and completeness (based on the polynomial-time many-one reducibility, $\leq_{\mathrm{m}}^{\mathrm{p}}$).

DP was introduced by Papadimitriou and Yannakakis [23] as the class of differences of any two NP problems; DP is also known as the second level of the boolean hierarchy over NP [7, 8]. For natural complete problems in the levels of the boolean hierarchy, and especially in DP, see the survey by Riege and Rothe [27] and, more recently, the work of Nguyen et al. [22] on social welfare optimization in multiagent resource allocation and of Reisch et al. [26] on the margin of victory in Schulze, cup, and Copeland elections.

$\mathrm{P}^{\mathrm{NP[log]}}$ was introduced by Papadimitriou and Zachos [24] as the class of problems that can be solved in polynomial time by asking $\mathcal{O}(\log n)$ sequential Turing queries to an NP oracle. This class is also known as capturing "parallel access to NP" (denoted by $\mathrm{P}_{\|}^{\mathrm{NP}}$) where polynomially many oracle queries may be asked in parallel; the equality of $\mathrm{P}^{\mathrm{NP[log]}}$ and $\mathrm{P}_{\|}^{\mathrm{NP}}$ has been shown independently by Hemachandra [13] and Köbler et al. [20]. $\mathrm{P}^{\mathrm{NP[log]}}$ constitutes the Θ_2^p level of the polynomial hierarchy and has been studied by many authors. While some of the earlier papers explore the properties of this class and its relation to other complexity classes [3, 4, 13, 20, 34, 35], both the early and more recent work focuses on proving completeness of natural problems in it, including various graph and satisfiability problems [34], the problems of whether certain heuristics can find constant-factor approximations for certain NP-complete graph problems [14, 18], the winner problems for Dodgson, Young, and Kemeny elections [15, 17, 28] (see also the survey by Hemaspaandra et al. [16]), and minimal upward or downward covering sets [2].

$\Sigma_2^p = \mathrm{NP}^{\mathrm{NP}}$ is the second level of the polynomial hierarchy [21, 31]. Natural complete problems in the levels of the polynomial hierarchy, and especially in Σ_2^p, have been surveyed by Schaefer and Umans [29, 30]. Recent Σ_2^p-completeness results on the complexity of core stability in hedonic games are due to Woeginger [37] (see also his survey [36]). It holds that $\mathrm{P} \subseteq \mathrm{NP} \subseteq \mathrm{DP} \subseteq \Theta_2^p \subseteq \Sigma_2^p$, and none of these inclusions is known to be strict.

The following two lemmas are due to Wagner [34] and provide sufficient conditions for proving lower bounds for DP and Θ_2^p. They will be applied in the proofs of Theorem 4 and Proposition 1, respectively.

Lemma 2 (Wagner [34]) *Let A be some* NP*-hard problem, and let B be any set. If there exists a polynomial-time computable function f such that, for any two instances x_1 and x_2 of A for which $x_2 \in A$ implies that $x_1 \in A$, we have*

$$|\{i|x_i \in A\}| \text{ is odd} \iff f(x_1, x_2) \in B, \tag{3}$$

then B is DP*-hard.*

Lemma 3 (Wagner [34]) *Let A be some* NP*-hard problem, and let B be any set. If there exists a polynomial-time computable function f such that, for all $k \geq 1$ and any $2k$ instances $x_1, \ldots, x_{2k}$ of A for which $x_j \in A$ implies that $x_i \in A$ for $i < j$, we have*

$$|\{i|x_i \in A\}| \text{ is odd} \iff f(x_1, x_2, \ldots, x_{2k}) \in B, \tag{4}$$

then B is Θ_2^p-hard.

3 Hardness of WSPV, WSPE, and SCSCS

We now turn to the main results of this paper, proving hardness results for the problems WSPV, WSPE, and SCSCS, first for the general problems and then for WSPE and SCSCS restricted to special graph classes. We start with the existence and verification problems for wonderfully stable partitions.

3.1 General hardness results for WSPV and WSPE

Just as for the core stability problems, the verification problem for wonderfully stable partitions, WSPV, belongs to NP due to the characterization stated in (1), since it can be tested in polynomial time whether a given subset of vertices is a clique and, if so, whether it blocks a given partition. Consequently, due to (2), the existence problem, WSPE, belongs to Σ_2^p. As a (potentially) better upper bound, Woeginger [36] shows membership of WSPE in Θ_2^p and conjectures that WSPE is Θ_2^p-hard.

Let us first consider WSPV. To pinpoint its complexity, we make use of the same proof technique that Sung and Dimitrov [32] used for the core stability problem in hedonic games with enemy-oriented preferences.

Theorem 1 WSPV *is* NP-*complete.*

Proof NP membership is obvious, as stated above. NP-hardness is shown via a reduction from CLIQUE as in the work of Sung and Dimitrov [32]. Given an instance of CLIQUE (which, for an undirected graph $G = (V, E)$ and a positive integer k, asks whether G has a clique of size at least k), we construct the following graph $G' = (V', E')$. The vertex set V' is obtained from V by adding, for each $v \in V$, $k - 2$ vertices. We connect each of the $k - 2$ new vertices and v to form a clique of size $k - 1$, for each $v \in V$. The edge set E' consists of these new edges and all edges in E. Let Π be the partition into $|V|$ cliques such that each $(k - 1)$-clique as constructed above forms one part. This can obviously be achieved in polynomial time. We claim that there is a clique of size k in G if and only if there exists a clique $P \subseteq V'$ that blocks Π in G'.

Only if: If there is a clique P of size k in G, the same clique can be found in G'. The vertices $v \in P$ thus have a clique number $\omega_{G'}(v)$ of at least k. Since the size of all cliques in Π is $k - 1$, there exists a vertex v in the clique P with $\omega_{G'}(v) > |\Pi(v)|$; therefore, P blocks Π in G'.

If: If there is no clique of size k in G, there is no clique of size k in G', either, and $\omega_{G'}(v) = k-1$ holds for each $v \in V'$. Furthermore, $|\Pi(v)| = k-1$, for each $v \in V'$. Thus, there is no blocking clique for Π in G'. □

We now turn to the problem WSPE, seeking to raise its lower bound step by step. We start by showing coNP-hardness; the construction presented in this proof will be used later on in the proof of Theorem 4.

Theorem 2 WSPE *is* coNP-*hard.*

Proof Again, we reduce from CLIQUE, but this time to the complement of WSPE. Given an instance (G, k) of CLIQUE, with $G = (V, E)$, we construct the same graph G' as in the proof of Theorem 1 as an instance for the complement of WSPE. We may assume that $k \geq 3$; otherwise, we could test in polynomial time whether E is empty or not and reduce to an appropriate trivial instance. We now show that there is a clique of size k in G if and only if there is no wonderfully stable partition for G'.

Only if: If there is a clique P of size k in G, the same clique can be found in G'. As in the proof of Theorem 1, P blocks the partition that consists of the $|V|$ cliques of size $k - 1$ constructed in the reduction. On the other hand, if a partition contains P, then each of the $(k - 1)$-cliques mentioned above blocks this partition, since the new vertices are now in a clique of size at most $k - 2$, but their clique number is $k - 1$.

If: If there is no clique of size k in G, the partition as in the proof of Theorem 1 is wonderfully stable, since there is no blocking clique. □

Next, we show that WSPE is also NP-hard, which was already mentioned without proof by Woeginger [36]. Thus, it is unlikely that the problem is in either NP or coNP (otherwise, the polynomial hierarchy would collapse). For completeness and since it will also be used in the upcoming proof of Theorem 4, we provide a proof of this result.

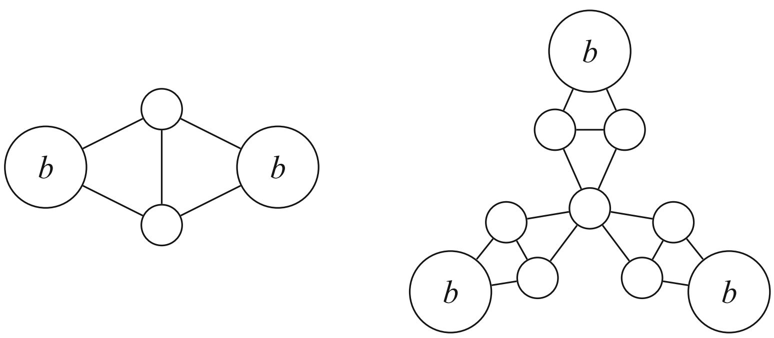
Theorem 3 (Woeginger [36]) WSPE *is* NP-*hard.*

Proof We show NP-hardness via a reduction from the well-known NP-hard problem EXACT COVER BY THREE-SETS (see, e.g., [12]), which we refer to as X3C. The input of this problem is a base set $B = \{b_1, \dots, b_{3k}\}$, $k > 0$, and a collection $\mathscr{S} = \{S_1, \dots, S_m\}$ of 3-element subsets of B, and the question is whether B can be exactly covered by k sets from $\mathscr{S}$. Given an X3C instance $(B, \mathscr{S})$, we may assume that each element of B occurs at most three times in any of the sets in $\mathscr{S}$ (see the book by Garey and Johnson [12]). Furthermore, we can assume that each element occurs at least once; otherwise, we could reduce to a trivial no-instance of WSPE.

Construct the following graph $G = (V, E)$ from $(B, \mathscr{S})$. For each $S_i \in \mathscr{S}$, add three vertices to V that are connected to each other as a 3-clique. Label the three vertices with the three elements of S_i. For each element $b \in B$, consider the following three cases. First, if b occurs only once in a set of $\mathscr{S}$, no changes are made. Second, if b occurs twice, the subgraph in Fig. 4a is inserted between the two vertices labeled with b. Third, if b occurs three times, the subgraph in Fig. 4b is inserted between the three vertices labeled with b. Since it is easy to determine how often an element of B occurs in a set of $\mathscr{S}$ and the number of new vertices is limited by $7|B|$, G can be constructed in polynomial time.

We now show that there is an exact cover of B by sets in $\mathscr{S}$ if and only if there is a wonderfully stable partition for G.

Only if: If there exists an exact cover of B by $k = |B|/3$ sets in $\mathscr{S}$, include the 3-cliques corresponding to these sets into the partition Π that shall be wonderfully stable. The remaining vertices (those from the inserted connecting subgraphs, and those corresponding to the S_i that are not part of the exact cover) are distributed as follows. Again, consider the three cases of occurrence: If an element b occurs only once, the only vertex labeled with b is already in a clique in Π. If an element b occurs twice, one vertex labeled b remains.



(a) Two vertices (b) Three vertices

Fig. 4 Construction between vertices labeled $b \in B$

This vertex forms a 3-clique with the two connecting vertices as in Fig. 4a. Put this 3-clique into Π. If an element b occurs three times, two vertices with the same label remain. From the structure of the connecting subgraph as in Fig. 4b, the two vertices connected to the vertex that is already in a part of the partition, form a 3-clique with the vertex in the middle. The other two pairs of vertices again form 3-cliques with the remaining vertices labeled b. If these three cliques are added to Π, the partition is complete. It remains to show that Π is wonderfully stable. Since each part of Π is a clique of size 3 and each vertex in G has clique number 3, the conditions for a wonderfully stable partition are satisfied.

If: If there exists a wonderfully stable partition Π in G, all cliques in Π have size 3, since by construction each vertex $v \in V$ has a clique number $\omega_G(v) = 3$. Since the connecting subgraphs from Figs. 4a and 4b are constructed such that exactly one labeled vertex is not part of a 3-clique, we have that, for each element $b \in B$, the one corresponding vertex has to be part of another 3-clique that does not contain an unlabeled vertex. Thus, there exist exactly $|B|/3$ cliques that consist of three labeled vertices, corresponding to sets in $\mathscr{S}$ in which each element of B occurs exactly once. That is, there exists an exact cover of B in $\mathscr{S}$. □

In order to prove DP-hardness of WSPE, we make use of Property 1 and Wagner's sufficient condition stated in Lemma 2, applying the constructions presented in the proofs of Theorems 2 and 3.

Theorem 4 WSPE *is* DP-*hard.*

Proof Again, consider the NP-hard problem X3C. Given two instances of X3C, $(B_1, \mathscr{S}_1)$ and $(B_2, \mathscr{S}_2)$, where $(B_2, \mathscr{S}_2) \in$ X3C implies $(B_1, \mathscr{S}_1) \in$ X3C, we construct the following graph $G = (V, E)$. G consists of two disconnected subgraphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, that is, $G = (V_1 \cup V_2, E_1 \cup E_2)$. G_1 is obtained from $(B_1, \mathscr{S}_1)$ by the construction given in the proof of Theorem 3. G_2 is built in two steps. First, the X3C instance $(B_2, \mathscr{S}_2)$ is transformed into an instance of CLIQUE: For each set $S_i \in \mathscr{S}$, create a vertex v_i. If two sets S_i and S_j are disjoint, connect the corresponding vertices by an edge $\{v_i, v_j\}$. Let $k = |B|/3$. In the second step, add $k-2$ vertices for each vertex corresponding to a set of $\mathcal{S}$, and edges as in the proof of Theorem 2. This construction can obviously be done in polynomial time. Note that, again, the proof only works for $k \geq 3$. If $k \leq 2$, reduce to an approriate trivial WSPE instance. We claim that $(B_1, \mathscr{S}_1) \in$ X3C and $(B_2, \mathscr{S}_2) \notin$ X3C if and only if there exists a wonderfully stable partition for G. Note that, since $(B_2, \mathscr{S}_2) \in$ X3C implies $(B_1, \mathscr{S}_1) \in$ X3C, this is enough to establish equivalence (3) in Lemma 2.

Only if: Suppose $(B_1, \mathscr{S}_1) \in$ X3C and $(B_2, \mathscr{S}_2) \notin$ X3C. Since $(B_1, \mathscr{S}_1)$ is in X3C, G_1 has a wonderfully stable partition by the proof of Theorem 3. Since additionally $(B_2, \mathscr{S}_2) \notin$ X3C, there are no $k = |B|/3$ pairwise disjoint sets in $\mathscr{S}$, thus there is no clique of size k in G. By the proof of Theorem 2, G_2 then also has a wonderfully stable partition. Since G_1 and G_2 are not connected, that is, the clique number of each vertex remains unchanged ($\omega_G(v) = \omega_{G_1}(v)$ if $v \in V_1$, and $\omega_G(v) = \omega_{G_2}(v)$ if $v \in V_2$), and since there are no additional vertices in G, G has a wonderfully stable partition as well.

If: We prove the contrapositive, i.e., if $(B_1, \mathscr{S}_1) \notin$ X3C or $(B_2, \mathscr{S}_2) \in$ X3C, then there is no wonderfully stable partition for G. Indeed, if $(B_1, \mathscr{S}_1) \notin$ X3C, then by the proof of Theorem 3, there is no wonderfully stable partition for G_1. On the other hand, if $(B_2, \mathscr{S}_2) \in$ X3C, there exists an exact cover of B in $\mathscr{S}$, that is, there are $k = |B|/3$ pairwise disjoint sets in $\mathscr{S}$. By construction, these sets are represented by k vertices in G_2, each connected to one another, thus forming a k-clique. By the proof of Theorem 2, it follows that there is no wonderfully stable partition for G_2. By construction, since there is no wonderfully stable partition for G_1 or G_2, there is no wonderfully stable partition for G either.

By Lemma 2, WSPE is DP-hard. □

3.2 General hardness results for SCSCS

We now turn to SCSCS, first showing its coNP-hardness by a reduction from CLIQUE to the complement of SCSCS.

Theorem 5 SCSCS *is* coNP-*hard.*

Proof Let (G, k) be a CLIQUE instance with a graph $G = (V, E)$ and an integer $k \geq 4$. Construct an SCSCS instance represented by the graph $G' = (V', E')$. Let $V' = V \cup V_1 \cup V_2$, where V_1 contains $k - 2$ new vertices for each of the vertices $v \in V$ and V_2 contains $k - 3$ new vertices for each $v \in V$, so $|V'| = |V| + |V|(2k - 5)$. Every vertex $v \in V$ is connected to its $k - 2$ associated vertices from V_1, any two of which are also connected by an edge, thus forming a $(k - 1)$-clique with "their" vertex v. Moreover, the $k - 3$ vertices from V_2 associated with v are connected to one of the vertices from V_1 in the $(k - 1)$-clique containing v, and they are also connected among each other, thus forming a $(k - 2)$-clique with the single vertex v' from V_1 they are connected to. E' contains all edges from E and the additional edges described above. See Fig. 5 for an illustration.

We claim that G has a clique of size at least k if and only if there is no strictly core stable coalition structure in the game $\mathcal{G}'$ represented by G'.

Only if: Assuming that there is a clique P of size k in G, this clique also exists in G'. Every possible strictly core stable coalition structure Γ has to contain a coalition corresponding to a clique P' at least as large as P, since otherwise the coalition corresponding to P would block Γ. Consider an arbitrary vertex $v \in P'$ and the vertices from $V_1 \cup V_2$ connected to v. The player corresponding to the single vertex v' from V_1 that is connected to v and vertices in V_2 can form a coalition of size $k - 2$ with the players corresponding to v''s neighbors either in V_1 or in V_2. In both cases, the one coalition with the player corresponding to v' that is not contained in Γ weakly blocks Γ: While the player corresponding to v' is indifferent, the other players strictly prefer to be in a coalition with her. Thus, there can be no strictly core stable coalition structure for the game represented by G'. Note that this argument does not work for the nonstrict core.

If: Assuming that there is no clique of size k in G, there is no such clique in G' either. Construct a strictly core stable coalition structure Γ for $\mathcal{G}'$ by letting each player corresponding to $v \in V$ form a coalition with the players corresponding to v's neighbors in V_1, and letting the players corresponding to the vertices from V_2 form a coalition with the players corresponding to their $k - 4$ neighbors from V_2. □

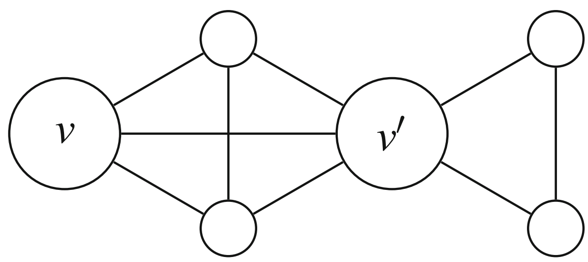


Fig. 5 Construction of $G' = (V', E')$ from $G = (V, E)$: Connecting vertices from V_1 and V_2 to $v \in V$ for $k = 5$

Recalling Lemma 1, we know that in graphs where all vertices have the same fixed clique number, every wonderfully stable partition Π of G corresponds to a strictly core stable coalition structure in the game represented by G, and vice versa. Hence, NP-hardness for SCSCS straightforwardly follows from the proof of Theorem 3, which states NP-hardness for WSPE.

Theorem 6 SCSCS *is* NP-*hard.*

Proof Use the reduction from the proof of Theorem 3 to construct a graph from a given X3C instance $(B, \mathcal{S})$. In this graph, all vertices have the same clique number, so with Lemma 1 we have that $(B, \mathcal{S}) \in$ X3C if and only if the game represented by G has a strictly core stable coalition structure. □

By using Wagner's sufficient condition from Lemma 2, DP-hardness of SCSCS can now be shown. We state this result without proof and refer to the proof of Theorem 4. The construction can be transferred directly to SCSCS by using the reduction showing CLIQUE $\leq_{\mathrm{m}}^{\mathrm{p}}$ $\overline{\text{SCSCS}}$ (see the proof of Theorem 5) to construct G_2 from a given X3C instance.

Theorem 7 SCSCS *is* DP-*hard.*

3.3 A result for a special graph class

Consider the class of graphs where all vertices have the same fixed clique number k. We can show NP membership of WSPE restricted to instances of this graph class (denoted by k-WSPE; note that k is not given as part of the input to k-WSPE; rather, it is a fixed constant, i.e., we study the problems 1-WSPE, 2-WSPE, etc. separately). Together with a lower bound that follows from the construction for proving Theorem 3, this gives NP-completeness.

Theorem 8 *For each* $k \geq 3$, k-WSPE *is* NP-*complete.*

Proof By assumption, all vertices in the given graph G have the same clique number k. The graph has to have $\ell \cdot k$ vertices for $\ell \in \mathbb{N}$; otherwise, a wonderfully stable partition could never be found. Thus, the problem of deciding whether G has a wonderfully stable partition is equivalent to the problem of deciding whether there is a clique cover of size ℓ for G, which is an NP-complete problem [19]. Therefore, NP membership of k-WSPE is shown by nondeterministically guessing a partition of the vertices into ℓ sets and, for each partition guessed, testing whether these sets are cliques.

For the lower bound, it follows from Theorem 3 that WSPE on graphs with a fixed clique number of $k = 3$ is NP-hard. We can extend this NP-hardness to any fixed clique number $k \geq 3$ by reducing k-WSPE to $(k + 1)$-WSPE. We may assume that an instance for k-WSPE has $\ell \cdot k$ vertices (otherwise, we reduce to a trivial no-instance). Given such a graph, we construct an instance of $(k+1)$-WSPE by adding ℓ vertices to the original graph. We connect each new vertex to each original vertex and leave the new vertices unconnected among each other. It is easy to see that there is a wonderfully stable partition into ℓ k-cliques in the original graph if and only if there is a wonderfully stable partition into ℓ cliques of size $k + 1$ each in the constructed graph. □

Since by Lemma 1 the problems WSPE and SCSCS are equivalent for graphs in this class, NP-completeness can also be shown for the SCSCS problem restricted to instances with a fixed clique number, k-SCSCS, $k \geq 3$.

Corollary 1 *For each $k \geq 3$, k-*SCSCS *is* NP*-complete.*

4 Conclusions and future work

We have shown that it is NP-complete to verify whether a given partition into cliques in a given graph can be blocked by some clique (thus preventing this partition from being wonderfully stable), and that it is DP-hard to decide whether or not a given graph has a wonderfully stable partition into cliques. Wonderful stability can be translated to a stability concept for enemy-oriented hedonic games. For a weaker stability concept in such games, strict core stability, we have also shown DP-hardness for the existence problem. In the case of friend-oriented preferences, the verification problem for core stability is an open question, suspected to be decidable in polynomial time by Woeginger [36]. Friend-oriented preferences, however, do not possess the property that a partition into cliques cannot be blocked by incomplete subgraphs in the corresponding graphs.[2] Therefore, wonderfully stable partitions carry over to hedonic games only for enemy-oriented preferences which we have focused on.

The main results of this paper (raising the lower bounds for WSPE and SCSCS to DP-hardness) should only be seen as a first step toward classifying them in terms of their complexity. We will now discuss possible ways toward showing Θ_2^p-hardness for them (as conjectured by Woeginger [36]) and will then conclude this paper by presenting a challenge: For showing Θ_2^p-hardness of these two problems, it would be enough to prove them coDP-hard.

4.1 Toward Θ_2^p-Hardness of WSPE and SCSCS

Chang and Kadin [10] define the following property: A problem A *has* AND$_\omega$ *functions*[3] if there exists a polynomial-time computable function f such that for all $n \in \mathbb{N}$ and for all instances $x_1, x_2, \ldots, x_n$ for A, it holds that $x_i \in A$ for each i, $1 \leq i \leq n$, if and only if $f(x_1, x_2, \ldots, x_n) \in A$.

Lemma 4 (Chang and Kadin [10]) 1. *If a problem is* NP*-complete, it has* AND$_\omega$ *functions.*
2. *If a problem is* DP*-complete, it has* AND$_\omega$ *functions.*
3. *If a problem is complete for any class of the boolean hierarchy higher than the second level, it cannot have* AND$_\omega$ *functions, unless the boolean hierarchy collapses to the second level.*
2. *If a problem is Θ_2^p-complete, it has* AND$_\omega$ *functions.*

Note that WSPE has AND$_\omega$ functions by Property 1. By Lemma 4(3), we thus can conclude that WSPE cannot be complete for any level of the boolean hierarchy higher

[2] Note that even games with a nonsymmetric friendship relation might allow stable partitions.

[3] Note that this is a different ω than the clique number, used here for consistency with the literature. Which ω is meant will always be clear from the context.

than the second level: WSPE is either complete for DP or Θ_2^p (or is something completely different). A similar statement applies to SCSCS.

In this section, we discuss a way for how to approach the as yet open issues of showing that WSPE and SCSCS are Θ_2^p-hard. To apply Lemma 3, the idea would be to generalize the construction for showing their DP-hardness (see Theorems 4 and 7), which we will elaborate on exemplarily for WSPE. From $2k$ given instances $x_1, \ldots, x_{2k}$ of an NP-hard problem A such as X3C, we construct a WSPE instance as a graph G with $k+1$ independent components G_i, $1 \le i \le k+1$. Then again, we can use Property 1 to conclude that G has a wonderfully stable partition if and only if each G_i has one. The single components G_i are constructed in the following way: The first one, G_1, is constructed from the first A instance x_1, the last one, G_{k+1}, is constructed from the last A instance x_{2k}, and the remaining $k-1$ components G_i, $2 \le i \le k$, are constructed from pairs (x_{2i-2}, x_{2i-1}) of A instances (see Fig. 6 for an illustration). For the thus constructed subgraphs, we need the following properties to hold.

Property 2 Let $x_1, \ldots, x_{2k}$ be given instances of an NP-hard problem A. Construct graphs $G_1, \ldots, G_{k+1}$ as follows:

1. Construct G_1 from x_1 such that $x_1 \in A \iff G_1 \in$ WSPE.
2. Construct G_i, $2 \le i \le k$, from x_{2i-2} and x_{2i-1} such that $(x_{2i-2}, x_{2i-1} \in A)$ or $(x_{2i-2}, x_{2i-1} \notin A) \iff G_i \in$ WSPE.
3. Construct G_{k+1} from x_{2k} such that $x_{2k} \in A \iff G_{k+1} \notin$ WSPE.

Proposition 1 *Let A be an* NP*-hard problem and let $x_1, \ldots, x_{2k}$ be any $2k$ instances of A such that $x_j \in A$ implies $x_i \in A$ for $i < j$. If $G_1, \ldots, G_{k+1}$ are graphs that can be constructed from $x_1, \ldots, x_{2k}$ in polynomial time such that Property 2 is satisfied, then* WSPE *is Θ_2^p-hard.*

Proof Let f be a polynomial-time computable function such that $f(x_1, \ldots, x_{2k}) = G$, where G is the graph consisting of $k+1$ independent components $G_1, \ldots, G_{k+1}$ that satisfy Property 2. To apply Lemma 3, we have to show equivalence (4) stated in that lemma:

$$|\{x_i | x_i \in A, 1 \le i \le 2k\}| \text{ is odd} \iff G \in \text{WSPE}.$$

G_1	G_2	G_3	$\cdots$	G_{k-1}	G_k	G_{k+1}
$\uparrow$	$\uparrow$	$\uparrow$		$\uparrow$	$\uparrow$	$\uparrow$
x_1	(x_2, x_3)	(x_4, x_5)	$\cdots$	(x_{2k-4}, x_{2k-3})	(x_{2k-2}, x_{2k-1})	x_{2k}
$+$	$(+,+)$	$(+,+)$	$\cdots$	$(+,+)$	$(+,+)$	$+$
$+$	$(+,+)$	$(-,-)$	$\cdots$	$(-,-)$	$(-,-)$	$-$
$+$	$(+,+)$	$(+,+)$	$\cdots$	$(+,-)$	$(-,-)$	$-$
$-$	$(-,-)$	$(-,-)$	$\cdots$	$(-,-)$	$(-,-)$	$-$

Fig. 6 Illustration of the reduction using Lemma 3. The last rows show possible cases of yes/no-instances due to the relation between the x_i, "+" denotes a yes-instance, and "−" denotes a no-instance

Only if: Assume that $|\{x_i | x_i \in A, 1 \le i \le 2k\}|$ is odd. Since $x_j \in A$ implies that $x_i \in A$ for $i < j$, neither $x_1 \notin A$ nor $x_{2k} \in A$ can hold.[4] By Property 2, we have that both G_1 and G_{k+1} have a wonderfully stable partition. Since $x_1 \in A$ and $x_{2k} \notin A$, there exists an index s (which we call the *separation index*) such that $x_i \in A$ for $i \le s$, and $x_i \notin A$ for $i > s$. Again, since $x_j \in A$ implies that $x_i \in A$ for $i < j$, only the following three cases can occur for each pair (x_{2i-2}, x_{2i-1}) of the remaining instances:

Case 1: both x_{2i-2} and x_{2i-1} are in A,
Case 2: neither x_{2i-2} nor x_{2i-1} are in A, or
Case 3: x_{2i-2} is in A, yet x_{2i-1} is not.

Case 3 implies that the separation index is of the form $s = 2i - 2$ for some i (see the third row of Fig. 6), which leads to a contradiction, since that would mean that there is an even number of yes-instances. So all pairs have to be of the form stated in Case 1 or Case 2 (see the second row of Fig. 6). By Property 2, each component G_i, $2 \le i \le k$, has a wonderfully stable partition and so has G.

If: Assume that G has a wonderfully stable partition. This implies that every component G_i, $1 \le i \le k+1$, does as well. By Property 2, we have that $x_1 \in A$, $x_{2k} \notin A$, and for all pairs (x_{2i-2}, x_{2i-1}), $2 \le i \le k$, either both x_{2i-2} and x_{2i-1} are in A, or neither x_{2i-2} nor x_{2i-1} are in A. In total, we have an odd number of yes-instances among $x_1, \dots, x_{2k}$.

By Lemma 3, WSPE is Θ_2^p-hard. □

4.2 Challenge

With the reduction presented in the DP-hardness proof for WSPE (see Theorem 4), the subgraphs G_1 and G_{k+1} can be constructed from given X3C instances such that the first and the third statement of Property 2 hold. To complete the Θ_2^p-hardness proof, we would have to construct the remaining subgraphs $G_2, \dots, G_k$ so as to satisfy the second statement of Property 2.

Looking closely at this statement and letting the NP-complete set A from Lemma 3 be 3-SAT, we are searching for a polynomial-time reduction f such that for two given 3-SAT instances, φ_1 and φ_2, we have:

$$(\varphi_1, \varphi_2 \in \text{3-SAT}) \text{ or } (\varphi_1, \varphi_2 \notin \text{3-SAT}) \iff f(\varphi_1, \varphi_2) \in \text{WSPE}. \tag{5}$$

Papadimitriou and Yannakakis [23] introduced the well-known DP-complete problem SAT-UNSAT: Given two boolean formulas in 3-CNF, φ_1 and φ_2, is it true that φ_1 is satisfiable (i.e., $\varphi_1 \in$ 3-SAT) and φ_2 is not satisfiable (i.e., $\varphi_2 \notin$ 3-SAT)? We may assume that $\varphi_2 \in$ 3-SAT implies $\varphi_1 \in$ 3-SAT. By Lemma 2, this restriction of SAT-UNSAT is also DP-complete. Then (5) simplifies to:

$$(\varphi_1, \varphi_2) \notin \text{SAT-UNSAT} \iff f(\varphi_1, \varphi_2) \in \text{WSPE}$$

It follows that in order to prove Θ_2^p-hardness—and thus Θ_2^p-completeness—of WSPE, it suffices to show coDP-hardness of WSPE. To summarize, we have shown the following result.

Theorem 9 WSPE *is* Θ_2^p*-complete if and only if it is* coDP*-hard.*

[4] Indeed, looking at the top and the bottom row of Fig. 6, we see that if either $x_{2k} \in A$ or $x_1 \notin A$, then either all $x_1, \dots, x_{2k}$ would be in A or none of them, contradicting the assumption that $|\{x_i | x_i \in A, 1 \le i \le 2k\}|$ is odd.

Essentially the same argument works for SCSCS as well: For proving a Θ_2^p-hardness lower bound, it would suffice to establish a coDP-hardness lower bound. Whether one can show coDP-hardness for WSPE and SCSCS is left as an open problem.

Acknowledgments We thank Gerhard Woeginger for drawing our attention to his survey [36] and the interesting open questions raised therein. We thank the anonymous AMAI and LOFT-2014 reviewers for helpful comments. We thank Francesca Rossi and Brent Venable for inviting us to present our work at the Special Session on Computational Social Choice at the 13th International Symposium on Artificial Intelligence and Mathematics (ISAIM 2014). This work was supported in part by DFG grants RO-1202/14-1 and RO1202/15-1 and a grant for gender-sensitive universities funded by the NRW Ministry for Innovation, Science, and Research.

References

1. Aziz, H., Brandt, F., Seedig, H.: Computing desirable partitions in additively separable hedonic games. Artif. Intell. **195**, 316–334 (2013)
2. Baumeister, D., Brandt, F., Fischer, F., Hoffmann, J., Rothe, J.: The complexity of computing minimal unidirectional covering sets. Theory of Computing Systems **53**(3), 467–502 (2013)
3. Beigel, R.: Bounded queries to SAT and the boolean hierarchy. Theor. Comput. Sci. **84**(2), 199–223 (1991)
4. Beigel, R., Hemachandra, L., Wechsung, G.: On the power of probabilistic polynomial time: $\mathrm{P}^{\mathrm{NP[log]}} \subseteq PP$. In: Proceedings of the 4th Structure in Complexity Theory Conference, pp. 225–227. IEEE Computer Society Press, June (1989)
5. Brams, S., Fishburn, P. In: Arrow, K., Sen, A., Suzumura, K. (eds.): Handbook of Social Choice and Welfare, volume 1, pages 173–236. North-Holland (2002)
6. Brandt, F., Conitzer, V., Endriss, U.: Computational social choice. In: Weiß, G. (ed.): Multiagent Systems, pp. 213–283. MIT Press, second edition (2013)
7. Cai, J., Gundermann, T., Hartmanis, J., Hemachandra, L., Sewelson, V., Wagner, K., Wechsung, G.: The boolean hierarchy I: Structural properties. SIAM J. Comput. **17**(6), 1232–1252 (1988)
8. Cai, J., Gundermann, T., Hartmanis, J., Hemachandra, L., Sewelson, V., Wagner, K., Wechsung, G.: The boolean hierarchy II: Applications. SIAM J. Comput. **18**(1), 95–111 (1989)
9. Chalkiadakis, G., Elkind, E., Wooldridge, M.: Computational Aspects of Cooperative Game Theory. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan and Claypool Publishers (2011)
10. Chang, R., Kadin, J.: On computing boolean connectives of characteristic functions. Mathematical Systems Theory **28**(3), 173–198 (1995)
11. Dimitrov, D., Borm, P., Hendrickx, R., Sung, S.: Simple priorities and core stability in hedonic games. Soc. Choice Welf. **26**(2), 421–433 (2006)
12. Garey, M., Johnson, D.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H, Freeman and Company (1979)
13. Hemachandra, L.: The strong exponential hierarchy collapses. J. Comput. Syst. Sci. **39**(3), 299–322 (1989)
14. Hemaspaandra, E., Rothe, J.: Recognizing when greed can approximate maximum independent sets is complete for parallel access to NP. Inf. Process. Lett. **65**(3), 151–156 (1998)
15. Hemaspaandra, E., Hemaspaandra, L., Rothe, J.: Exact analysis of Dodgson elections: Lewis Carroll's 1876 voting system is complete for parallel access to NP. J. ACM **44**(6), 806–825 (1997a)
16. Hemaspaandra, E., Hemaspaandra, L., Rothe, J.: Raising NP lower bounds to parallel NP lower bounds. SIGACT News **28**(2), 2–13 (1997b)
17. Hemaspaandra, E., Spakowski, H., Vogel, J.: The complexity of Kemeny elections. Theor. Comput. Sci. **349**(3), 382–391 (2005)
18. Hemaspaandra, E., Rothe, J., Spakowski, H.: Recognizing when heuristics can approximate minimum vertex covers is complete for parallel access to NP. R.A.I.R.O. Theoretical Informatics and Applications **40**(1), 75–91 (2006)
19. Karp, R.: Reducibility among combinatorial problems. In: Miller, R., Thatcher, J. (eds.): Complexity of Computer Computations, pp. 85–103. Plenum Press (1972)
20. Köbler, J., Schöning, U., Wagner, K.: The difference and truth-table hierarchies for NP. *R.A.I.R.O.* Informatique théorique et Applications **21**, 419–435 (1987)

21. Meyer, A., Stockmeyer, L.: The equivalence problem for regular expressions with squaring requires exponential space. In: Proceedings of the 13th IEEE Symposium on Switching and Automata Theory, pages 125–129. IEEE Computer Society Press (1972)
22. Nguyen, N., Nguyen, T., Roos, M., Rothe, J.: Computational complexity and approximability of social welfare optimization in multiagent resource allocation. Journal of Autonomous Agents and Multi-Agent Systems **28**(2), 256–289 (2014)
23. Papadimitriou, C., Yannakakis, M.: The complexity of facets (and some facets of complexity). J. Comput. Syst. Sci. **28**(2), 244–259 (1984)
24. Papadimitriou, C., Zachos, S.: Two remarks on the power of counting. In: Proceedings of the 6th GI Conference on Theoretical Computer Science, pages 269–276. Springer-Verlag Lecture Notes in Computer Science #145 (1983)
25. Peleg, B., Sudhölter, P.: Introduction to the theory of cooperative games. Kluwer Academic Publishers (2003)
26. Reisch, Y., Rothe, J., Schend, L.: The margin of victory in Schulze, cup, and Copeland elections: Complexity of the regular and exact variants. In: Proceedings of the 7th European Starting AI Researcher Symposium, pp. 250–259. IOS Press, August (2014)
27. Riege, T., Rothe, J.: Completeness in the boolean hierarchy: Exact-Four-Colorability, minimal graph uncolorability, and exact domatic number problems – a survey. Journal of Universal Computer Science **12**(5), 551–578 (2006)
28. Rothe, J., Spakowski, H., Vogel, J.: Exact complexity of the winner problem for Young elections. Theory of Computing Systems **36**(4), 375–386 (2003)
29. Schaefer, M., Umans, C.: Completeness in the polynomial-time hierarchy: Part I: A compendium. *SIGACT News* **33**(3), 32–49 (2002a)
30. Schaefer, M., Umans, C.: Completeness in the polynomial-time hierarchy: Part II. *SIGACT News* **33**(4), 22–36 (2002b)
31. Stockmeyer, L.: The polynomial-time hierarchy. Theor. Comput. Sci. **3**(1), 1–22 (1976)
32. Sung, S., Dimitrov, D.: On core membership testing for hedonic coalition formation games. Oper. Res. Lett. **35**(2), 155–158 (2006)
33. Sung, S., Dimitrov, D.: Computational complexity in additive hedonic games. Eur. J. Oper. Res. **203**(3), 635–639 (2010)
34. Wagner, K.: More complicated questions about maxima and minima, and some closures of NP. Theor. Comput. Sci. **51**, 53–80 (1987)
35. Wagner, K.: Bounded query classes. SIAM J. Comput. **19**(5), 833–846 (1990)
36. Woeginger, G.: Core stability in hedonic coalition formation. In: Proceedings of the 39th Conference on Current Trends in Theory and Practice of Computer Science, pages 33–50. Springer-Verlag Lecture Notes in Computer Science #7741 (January 2013a)
37. Woeginger, G.: A hardness result for core stability in additive hedonic games. Math. Soc. Sci. **65**(2), 101–104 (2013b)

CHAPTER 4

Representing and Solving Hedonic Games with Ordinal Preferences and Thresholds

Summary

Different representations of hedonic games have different advantages and disadvantages, and it often boils down to weighting the compactness of the representation against its expressiveness. Common representations, such as the additive encoding or singleton encoding, are very compact, but lack the ability to represent hedonic games as soon as qualitative user inputs should be taken into account. On the other side of the spectrum are the most expressive representations, which are, however, not compact, such that a single preference order often already needs exponential space in the number of agents. In this paper, we merge two compact approaches, namely the singleton encoding and the friends-and-enemies encoding, and create an encoding that still is compact, i.e., it only needs polynomial space in the number of agents to specify a complete hedonic game, but also much more expressive than any of the two original encoding itself. The former, the singleton encoding, allows agents to rank only other agents, but not the entire coalitions containing her. A preference order over all such coalitions is then derived by only looking at the best (resp. worst) ranked agent in a given coalition. The friends-and-enemies encoding has already been discussed in Chapter 3, and basically works by counting the number of friends and/or enemies in the coalitions.

In the resulting encoding, called *weak ranking with double threshold* in the paper, every agent i first puts the other agents into three groups, her friends

A_i^+, her enemies A_i^-, and neutral agents A_i^0, i.e., agents that she does not care about. Then, she ranks her friends and enemies, leading to a compact representation of her opinion over the other agents. The resulting hedonic game, i.e., a hedonic game in which every agent's opinion is represented by a weak ranking with double threshold, is called FEN-hedonic game. This representation, however, has one disadvantage: It cannot be easily extended to a ranking over the coalitions containing i. To cover this disadvantage, we take two steps. First, we use a polarized version of the so called *responsive extension principle*, which results in an incomplete ranking over all coalitions containing agent i. Second, we use the principles of two modularities that describe whether there is a way to fill the open gaps in the rankings in such a way that the desired outcome is achieved, versus whether each such completion leads to the desired outcome. Such terms are often referred to as 'possibility' and 'necessity', leading to two versions of each investigated decision problems. One Example would be POSSIBLE-γ-VERIFICATION, which asks, for a given coalition structure in a hedonic game and a fixed stability concept γ, whether there is at least one possible way to extend the incomplete rankings, such that the coalition structure is stable in regards to the given stability concept γ and the resulting game. In contrast to this, NECESSARY-γ-VERIFICATION asks, whether the given coalition structure is stable regarding γ in all ways to extend the incomplete rankings.

Hence, the following paper contains an analysis of the verification and existence decision problems, both in the possible and necessary case, and for the stability concepts of perfectness, individual rationality, (contractual) individual stability, Nash stability, as well as some additional analysis regarding (strict) core stability, Pareto optimality, and (strict) popularity. It offers several hardness results up to the class of NP, but also leaves open gaps in complexity for future work. Please note, that Paragraph 5.2 of [40] displays an earlier version of our research regarding Borda-like comparability functions. This research is represented in more detail in Chapter 5.

Contribution

The idea, model, and writing was done jointly with my coauthors. Additionally, I contributed research regarding the analysis of FEN-hedonic games that did not make it into the paper but will be featured in an upcoming version. These results are based on properties introduced by Peters and Elkind in [51] that will also be used in Chapter 5. This includes axiomatic properties lead to NP-completeness results for individual stability and Nash stability, and to NP-hardness results for core stability. For a detailed analysis of my contribution regarding Borda-like comparability functions, please see Chapter 5.

Publication - Lang, Rey, Rothe, Schadrack, and Schend [40]

J. Lang, A. Rey, J. Rothe, H. Schadrack, and L. Schend. Representing and solving hedonic games with ordinal preferences and thresholds. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems*, pages 1229–1237. IFAAMAS, May 2015

Representing and Solving Hedonic Games with Ordinal Preferences and Thresholds

Jérôme Lang
LAMSADE, Université
Paris-Dauphine
75775 Paris Cedex 16, France
lang@lamsade.dauphine.fr

Anja Rey
Heinrich-Heine-Universität
Düsseldorf
40225 Düsseldorf, Germany
rey@cs.uni-duesseldorf.de

Jörg Rothe
Heinrich-Heine-Universität
Düsseldorf
40225 Düsseldorf, Germany
rothe@cs.uni-duesseldorf.de

Hilmar Schadrack
Heinrich-Heine-Universität
Düsseldorf
40225 Düsseldorf, Germany
schadrack@cs.uni-duesseldorf.de

Lena Schend
Heinrich-Heine-Universität
Düsseldorf
40225 Düsseldorf, Germany
schend@cs.uni-duesseldorf.de

ABSTRACT

We propose a new representation setting for hedonic games, where each agent partitions the set of other agents into friends, enemies, and neutral agents, with friends and enemies being ranked. Under the assumption that preferences are monotonic (respectively, antimonotonic) with respect to the addition of friends (respectively, enemies), we propose a bipolar extension of the Bossong–Schweigert extension principle, and use this principle to derive the (partial) preferences of agents over coalitions. Then, for a number of solution concepts, we characterize partitions that necessarily (respectively, possibly) satisfy them, and identify the computational complexity of the associated decision problems. Alternatively, we suggest cardinal comparability functions in order to extend to complete preference orders consistent with the generalized Bossong–Schweigert order.

Categories and Subject Descriptors

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

General Terms

Economics, Theory

Keywords

Computational Social Choice, Coalition Formation, Game Theory

1. INTRODUCTION

Hedonic games are strategic games where agents, from a set A, are free to form coalitions. Each agent has a preference relation over the set of all coalitions containing her; various solution concepts—such as individual rationality, Nash stability, individual contractual stability, core stability, and so on—have been proposed and studied. However, an important bottleneck is how the agents' preferences over all coalitions that contain them are expressed. As there are exponentially many coalitions containing agent i, it is unreasonable to expect that agent i should express explicitly a ranking (or a utility function) over all these coalitions. This issue is often addressed by assuming that only a small part of the preference relation is expressed by the agent, and that this small part is then extended into a complete preference relation over coalitions. Various assumptions about the nature of the input (what the agents express) and the preference extension have been made in the literature (for a survey, see Woeginger [23]):

1. The *individually rational encoding* [4]: Each agent ranks only the coalitions she prefers to herself being alone.

2. The *additive encoding* [21, 22, 3, 24]: Each agent gives a valuation (positive or negative) of each other agent; preferences are additively separable, and the extension principle is that the valuation of a set of agents, for agent i, is the sum of the valuations i gives to the agents in the set (and then the preference relation is derived from this valuation function).

3. The *"friends and enemies" encoding* [15, 21]: Each agent partitions the set of other agents into two sets (her friends and her enemies); under the *friend-oriented preference extension*, coalition X is preferred to coalition Y if X contains more friends than Y, or as many friends as Y and fewer enemies than Y; under the *enemy-oriented preference* extension, X is preferred to Y if X contains fewer enemies than Y, or as many enemies as Y and more friends than Y.

4. The *singleton encoding* [12, 10, 11]: Each agent ranks only single agents; under the optimistic (respectively, pessimistic) extension, X is preferred to Y if the best (respectively, worst) agent in X is preferred to the best (respectively, worst) agent in Y.

5. The *anonymous encoding* [4, 13]: Each agent specifies only a preference relation over the number of agents in her coalition (and does not care about the identities of these agents).

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.*

6. *Hedonic coalition nets* [16]: Each agent specifies her utility function over the set of all coalitions via (more or less) a set of weighted logical formulas.

7. *Fractional hedonic games* [2]: Each agent assigns a value to each other agent (and 0 to herself); an agent's utility of a coalition is the average value she assigns to the members of the coalition. A coalition X is preferred to Y if the utility of X is greater than that of Y.

Naturally, compact representation either does not avoid exponential-size representations in the worst case (Case 1 and, to a lesser extent, Case 6), or comes with a loss of expressivity, corresponding to a demanding domain restriction, such as separable preferences (Cases 2 and 4), anonymous preferences (Case 5), or other domain restrictions that do not bear a specific name (Case 3).

In Cases 2 and 6, preferences are expressed numerically: Agents do explicitly express numbers. In all other cases, they are expressed ordinally. Advantages of ordinal preferences in social choice have been discussed many times and we want to stick here to ordinality. We do not want to make the very demanding anonymity assumption, which does not allow to distinguish between agents. The individually rational encoding is not compact in general. So there remain only the "friends and enemies" and singleton encodings. The problem with "friends and enemies" is that an agent cannot express preferences inside the friend set nor inside the enemy set: Preferences over individual agents are dichotomous (but preferences between coalitions are not, because they depend on the number of friends and enemies). The problem with the singleton encoding is that having simply a rank $\rhd_i$ for each agent i does not tell us which agents i would like to see in her coalitions and which agents she would like not to: For instance, if $\rhd_1$ is $2 \rhd_1 3 \rhd_1 4$, we know that 1 prefers 2 to 3 and 3 to 4, but nothing tells us whether 1 prefers to be with 2 (respectively, 3 and 4) to being alone, that is, if the absolute desirability of $2, 3$, and 4 is positive or negative (of course, if it is negative for 3, it is also negative for 4, etc.). So, both ways are insufficiently informative: Specifying only a partition into positive and negative agents ("friends" and "enemies") does not tell which of her friends i prefers to which other agents, and which of her enemies she wants to avoid most. On the other hand, specifying a ranking over agents does not say which agents i prefers to be with rather than being alone. Here we propose a model that integrates the models of Cases 1, 3, and 4: Each agent i first subdivides the other agents into three groups, her friends, her enemies, and an intermediate type of agents on which she has neither a positive nor a negative opinion and then specifies a ranking of her friends and enemies. Based on this representation, we consider a natural extension to a player's preference, the generalized Bossong–Schweigert extension (see [8, 14]), which is a partial order over coalitions containing the player. A related model can be found in the context of matching theory: Responsive preferences are studied in bipartite many-to-one matching markets and consider the comparison of one participant to another,[1] although not in distinction of friends or enemies (see, e.g., [19, 20]). In the following, we consider different ways of how to deal with incomparabilities within these partial orders. A first approach is to leave incomparabilities open and define notions such as "possible" and "necessary" stability concepts. A second approach is to define comparability functions in order to determine the relation between incomparable coalitions that extend the generalized Bossong–Schweigert extension to a total preference order for each player. Questions of interest include appropriate characterizations of stability concepts and a computational study of the related problems.

[1] In the context of many-to-one matching markets, an agent on the one side has *responsive preferences* over assignments of the agents on the other side if, for any two assignments that differ in only one agent, the assignment containing the most preferred agent is preferred.

2. PRELIMINARIES

Generally, a *hedonic game* is a tuple (A, P) consisting of a set of *players* (or *agents*) $A = \{1, 2, \dots, n\}$ and a profile of preference relations $P = (\succeq_1, \succeq_2, \dots, \succeq_n)$ defining for each player a weak preference order over all possible *coalitions* $C \subseteq A$ containing the player herself. For two coalitions $C, D \subseteq A$, both containing player i, we say that *i weakly prefers C to D* if $C \succeq_i D$; *i prefers C to D*, denoted by $C \succ_i D$, if $C \succeq_i D$, but not $D \succeq_i C$; and *i is indifferent between C and D*, denoted by $C \sim_i D$, if both $C \succeq_i D$, and $D \succeq_i C$. A *coalition structure* Γ for a given game (A, P) is a partition of A into disjoint coalitions, and for each player $i \in A$, $\Gamma(i)$ denotes the unique coalition in Γ containing i.

An important solution concept for the study of hedonic games is the notion of stability of a coalition structure. There are several known such stability concepts [7, 3, 1]. In this paper we focus on concepts that deal with avoiding a player to deviate to another (possibly empty) existing coalition. Relatedly, other commonly studied concepts consider group deviations, such as core stability with the goal that there is no blocking coalition. A third group of stability concepts, such as Pareto optimality and popularity, is based on a relation comparing different coalition structures. Further restrictions of games as well as properties can be found amongst others in [5].

A coalition structure Γ is called

- *perfect* if each player i weakly prefers $\Gamma(i)$ to every other coalition containing i,
- *individually rational* if each player $i \in A$ weakly prefers $\Gamma(i)$ to being alone in $\{i\}$,
- *Nash stable* if for each player $i \in A$, $\Gamma(i) \succeq_i A' \cup \{i\}$ holds for each coalition $A' \in \Gamma \cup \emptyset$,
- *individually stable* if for each player $i \in A$ and for each coalition $A' \in \Gamma \cup \emptyset$, it holds that $\Gamma(i) \succeq_i A' \cup \{i\}$ or there exists a player $j \in A'$ such that $A' \succ_j A' \cup \{i\}$,
- *contractually individually stable* if for each player $i \in A$ and for each coalition $A' \in \Gamma \cup \emptyset$, it holds that $\Gamma(i) \succeq_i A' \cup \{i\}$, or there exists a player $j \in A'$ such that $A' \succ_j A' \cup \{i\}$, or there exists a player $j' \in \Gamma(i)$ such that $\Gamma(i) \succ_{j'} \Gamma(i) \setminus \{i\}$.

3. DERIVING PREFERENCES OVER COALITIONS FROM PREFERENCES OVER SINGLE FRIENDS AND ENEMIES

We define a new representation of preferences combining ordinal rankings with friend and enemy sets. We suggest deriving a player's preference over coalitions by generalizing the Bossong–Schweigert extension principle.

3.1 Ordinal Preferences with Thresholds

DEFINITION 1. *Let $A = \{1, 2, \dots, n\}$ be a set of agents. For each $i \in A$, a* weak ranking with double threshold *for agent i, denoted by $\trianglerighteq_i^{+0-}$, consists of a partition of $A \setminus \{i\}$ into three sets:*

- *A_i^+ (i's friends), together with a weak order $\trianglerighteq_i^+$ over A_i^+,*
- *A_i^- (i's enemies), together with a weak order $\trianglerighteq_i^-$ over A_i^-, and*

- A_i^0 *(the neutral agents, i.e., the agents i does not care about).*

We also write $\trianglerighteq_i^{+0-}$ *as* $(\trianglerighteq_i^+ \,|\, j_1 \cdots j_k \,|\, \trianglerighteq_i^-)$ *for* $A_i^0 = \{j_1, \ldots, j_k\}$.

Not having an order of the neutral agents can be interpreted as being indifferent about them all: $j_a \sim_i j_b$ for all $j_a, j_b \in A_i^0$. Agent i strictly prefers all her friends to her neutral agents, and those to her enemies. The *weak order induced by* $\trianglerighteq_i^{+0-}$ is therefore defined via $f \triangleright_i j$, for each $f \in A_i^+$ and $j \in A_i^0$, $j_1 \sim_i j_2 \sim_i \cdots \sim_i j_k$, and $j \triangleright_i e$, for each $j \in A_i^0$ and $e \in A_i^-$.

EXAMPLE 2. *Let* $A = \{1, 2, \ldots, 11\}$. *Then,*

$$\trianglerighteq_1^{+0-} = (2 \triangleright_1 3 \sim_1 4 \,|\, 567 \,|\, 8 \triangleright_1 9 \sim_1 10 \triangleright_1 11)$$

means that 1 *likes* 2*,* 3*, and* 4 *(and prefers* 2 *to both* 3 *and* 4*, and is indifferent between* 3 *and* 4*);* 1 *does not care about* 5*,* 6*, and* 7 *(and is indifferent between them); and* 1 *does not like* 8*,* 9*,* 10*, and* 11 *(but still prefers* 8 *to* 9 *and* 10*, is indifferent between* 9 *and* 10*, and prefers* 9 *and* 10 *to* 11*). The weak order* $\trianglerighteq_1$ *induced by* $\trianglerighteq_1^{+0-}$ *is* $2 \triangleright_1 3 \sim_1 4 \triangleright_1 5 \sim_1 6 \sim_1 7 \triangleright_1 8 \triangleright_1 9 \sim_1 10 \triangleright_1 11$. *Note that here the preference between a friend and a neutral player is strict, because we assume below that a coalition containing a friend instead of a neutral player is preferred. Analogously, the preference between a neutral player and an enemy is strict, because a player does not care about having a neutral player in a coalition but is less happy with having an enemy in the coalition instead.*

3.2 Generalizing Bossong–Schweigert Extensions

DEFINITION 3. *Let* $\trianglerighteq_i^{+0-}$ *be a weak ranking with double threshold for agent i. The extended order* $\succeq_i^{+0-}$ *is defined as follows: For every* $X, Y \subseteq A$, $X \succeq_i^{+0-} Y$ *if and only if the following two conditions hold:*

1. *There is an injective function* σ *from* $Y \cap A_i^+$ *to* $X \cap A_i^+$ *such that for every* $y \in Y \cap A_i^+$*, we have* $\sigma(y) \trianglerighteq_i y$.
2. *There is an injective function* θ *from* $X \cap A_i^-$ *to* $Y \cap A_i^-$ *such that for every* $x \in X \cap A_i^-$*, we have* $x \trianglerighteq_i \theta(x)$.

Finally, $X \succ_i^{+0-} Y$ *if and only if* $X \succeq_i^{+0-} Y$ *and not* $(Y \succeq_i^{+0-} X)$.

Intuitively speaking, for a fixed coalition C adding a further friend makes the coalition strictly more valuable while adding an enemy causes the opposite. When exchanging two friends, the valuation of the coalition changes depending on the relation between the exchanged players (the same holds when two enemies are exchanged). When both a friend and an enemy are added or are both removed, the original and the new coalition are incomparable with respect to the Bossong–Schweigert extension principle.

Thus, to construct the *generalized Bossong–Schweigert extension* (GBS-extension, for short) for a player i, we start with the coalition containing i and her set of friends (which is the most preferred coalition) and then construct all directly comparable coalitions by adding enemies, removing friends, or exchanging enemies or friends. For each newly obtained coalition we repeat this procedure until we reach the least preferred coalition containing all of i's enemies. Note that the elements of A_i^0 are disregarded as their adding to or removing from a coalition does not change the value of a coalition. The following examples illustrate the just presented extension principle.

EXAMPLE 4. *For* $A = \{1, 2, \ldots, 6\}$*, consider*

$$\trianglerighteq_1^{+0-} = (2 \triangleright_1 3 \sim_1 4 \,|\, \,|\, 5 \triangleright_1 6).$$

The graph in Figure 1 shows the generalized Bossong–Schweigert extension of this preference, where an arc from coalition X to coalition Y implies that $X \succ_1^{+0-} Y$*. Hence, any path leading from* X' *to* Y' *implies* $X' \succ_1^{+0-} Y'$*, whereas coalitions that are not connected by a path, such as* $\{1,2,3\}$ *and* $\{1,2,3,4,5\}$*, are incomparable.*

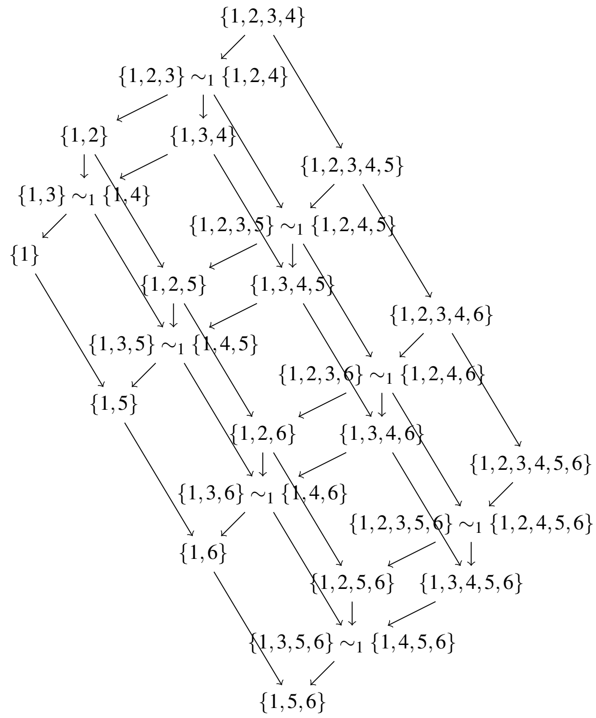


Figure 1: The generalized Bossong–Schweigert extension of $\trianglerighteq_1^{+0-} = (2 \triangleright_1 3 \sim_1 4 \,|\, \,|\, 5 \triangleright_1 6)$**.**

Note that if there were additional players $j > 6$ in A considered as neutral by player 1, the general picture would be the same with indifferences at each level, for any $C \subseteq \{2, \ldots, 6\}$, between each $\{1\} \cup C \cup N$ for $N \subseteq A \setminus \{1, \ldots, 6\}$.

EXAMPLE 5. *Consider* $A = \{1,2,3,4,5\}$ *and the first players' preference* $\trianglerighteq_1^{+0-} = (2 \triangleright_1 3 \,|\, \,|\, 4 \triangleright_1 5)$*. The graph in Figure 2 shows the generalized Bossong–Schweigert extension of this preference using the same notation as in Example 4.*

Using the generalized Bossong–Schweigert extension principle, we can extend the given preferences of the players to a preference over the possible coalitions. However, this preference over the coalitions might be incomplete; there are coalitions that remain incomparable. We consider two possibilities to deal with these incomparabilities: Leave them open and consider every possible extension that does not conflict with transitivity; alternatively, determine the relation between incomparable coalitions by adapting the Borda scoring rule, which is well-known from voting theory.

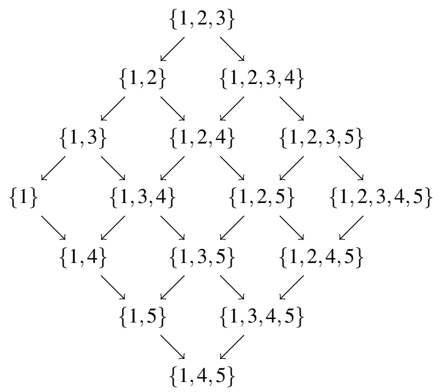


Figure 2: The generalized Bossong–Schweigert order of $\trianglerighteq_1^{+0-} = (2 \rhd_1 3 \mid \mid 4 \rhd_1 5)$.

Intuitively, the relation between two coalitions C and D ($C \succ_i D$, $D \succ_i C$, $C \sim_i D$, or undecided) from player i's point of view can be determined by the following characterizations. These characterizations are inspired by Bouveret et al. [9] who show characterizations for the original Bossong–Schweigert order in the context of fair division.

PROPOSITION 6. *1. Let $\trianglerighteq_i^{+0-}$ be a weak ranking with double threshold for agent i, and let C and C' be two coalitions containing i. Consider the orders $f_1 \trianglerighteq_i f_2 \trianglerighteq_i \cdots \trianglerighteq_i f_\mu$ with $\{f_1, f_2, \dots, f_\mu\} = C \cap A_i^+$ and $f'_1 \trianglerighteq_i f'_2 \trianglerighteq_i \cdots \trianglerighteq_i f'_{\mu'}$ with $\{f'_1, f'_2, \dots, f'_{\mu'}\} = C' \cap A_i^+$, as well as $e_1 \trianglerighteq_i e_2 \trianglerighteq_i \cdots \trianglerighteq_i e_\nu$ with $\{e_1, e_2, \dots, e_\nu\} = C \cap A_i^-$ and $e'_1 \trianglerighteq_i e'_2 \trianglerighteq_i \cdots \trianglerighteq_i e'_{\nu'}$ with $\{e'_1, e'_2, \dots, e'_{\nu'}\} = C' \cap A_i^-$. Then, $C \succ_i^{+0-} C'$ if and only if*

(a) $\mu \geq \mu'$ and $\nu \leq \nu'$,

(b) for each k, $1 \leq k \leq \mu'$, it holds that $f_k \trianglerighteq_i f'_k$, and

(c) for each ℓ, $1 \leq \ell \leq \nu$, it holds that $e_{\nu-\ell+1} \trianglerighteq_i e'_{\nu'-\ell+1}$.

2. Say that $w_i : A \to \mathbb{R}$ is compatible with $\trianglerighteq_i^{+0-}$ if and only if

- *for each $j \in A_i^+$, we have $w_i(j) > 0$;*
- *for each $j \in A_i^-$, we have $w_i(j) < 0$;*
- *for each $j \in A_i^0$, we have $w_i(j) = 0$; and*
- *for all $j,k \in A_i^+ \cup A_i^-$, we have $j \rhd_i k$ if and only if $w_i(j) > w_i(k)$.*

Then, $C \succ_i^{+0-} C'$ if and only if for any w_i compatible with $\trianglerighteq_i^{+0-}$, we have $\sum_{j \in C} w_i(j) > \sum_{j' \in C'} w_i(j')$.

PROOF. 1. Obviously, if (a) to (c) hold, the two injective functions $\sigma : C' \cap A_i^+ \to C \cap A_i^+$, and $\theta : C \cap A_i^- \to C' \cap A_i^-$ mapping $f'_k \mapsto f_k$ for each k, $1 \leq k \leq \mu'$, and $e_{\nu-\ell+1} \mapsto e'_{\nu'-\ell+1}$ for each ℓ, $1 \leq \ell \leq \nu$, satisfy $\sigma(f'_k) \trianglerighteq_i f'_k$ and $e_{\nu-\ell+1} \trianglerighteq_i \theta(e_{\nu-\ell+1})$, for the same range of k and ℓ. On the other hand, if there are two injective functions with the desired requirements, (a) holds. If there was a k with $f'_k \rhd_i f_k$ (or an ℓ with $e'_{\nu'-\ell+1} \rhd_i e_{\nu-\ell+1}$), this would imply $\sigma(f'_k) = f_j$ for a $j < k$ (or $\theta(e_{\nu-\ell+1}) = e'_{\nu-j+1}$ with $j > \ell$, respectively). This, however, implies that either a requirement is violated for f'_1 (or e_ν), or that σ (or θ) is not injective, a contradiction.

2. Assume that $C \succ_i^{+0-} C'$, that is, $C \succeq_i^{+0-} C'$ and not $C' \succeq_i^{+0-} C$. For the set of friends A_i^+, with $F = C \cap A_i^+$ and $F' = C' \cap A_i^+$, it follows that there is an injective function $\sigma : F' \to F$ such that for each $y \in F'$, we have $\sigma(y) \trianglerighteq_i y$. Hence, for each compatible w_i, $w_i(\sigma(y)) \geq w_i(y)$. Thus, since σ is injective,

$$\sum_{j \in F} w_i(j) \geq \sum_{j \in \sigma(F') \subseteq F} w_i(j) = \sum_{j' \in F'} w_i(\sigma(j')) \geq \sum_{j' \in F'} w_i(j'). \quad (1)$$

Similarly, for A_i^-, with $E = C \cap A_i^-$ and $E' = C' \cap A_i^-$, and θ injective, it holds that

$$0 \geq \sum_{j \in E} w_i(j) \geq \sum_{j \in E} w_i(\theta(j)) = \sum_{j' \in \theta(E) \subseteq E'} w_i(j') \geq \sum_{j' \in E'} w_i(j'). \quad (2)$$

Since $C' \succeq_i^{+0-} C$ does not hold, at least one of the inequalities (1) and (2) is strict, since one preference $(\sigma(j') \rhd_i j'$ or $j \rhd_i \theta(j))$ or one inclusion $(\sigma(F') \subset F$ or $\theta(E) \subset E')$ is strict. For each player $j \in A_i^0$, we have $w_i(j) = 0$; therefore, in total,

$$\sum_{j \in C} w_j > \sum_{j' \in C'} w_{j'}. \quad (3)$$

Now assume that for each compatible w_i, (3) holds. Thus,

$$\sum_{j \in F} w_i(j) - \sum_{j' \in E'} w_i(j') > \sum_{j' \in F'} w_i(j') - \sum_{j \in E} w_i(j).$$

Assume there were no injective function mapping from each summand from the right-hand side to one at least as large on the left hand side; then, there exists an assignment to the values of w_i compatible with $\trianglerighteq_i^{+0-}$ that does not satisfy the inequality, a contradiction. Hence, such a function must exist, and this function induces the mappings σ and θ, showing $C \succeq_i^{+0-} C'$. Additionally, because the inequality is strict in (3), $C' \succeq_i^{+0-} C$ does not hold, which completes the proof.

This completes the proof. ❑

4. POSSIBLE/NECESSARY STABILITY

As we have seen above, the generalized Bossong–Schweigert extension can leave uncertainties between two coalitions in a player's preference order.

DEFINITION 7. *A complete preference relation $\succeq_i$ over all coalitions containing i* extends *$\succeq_i^{+0-}$ if and only if it contains it; that is, if $C \succeq_i^{+0-} D$ implies $C \succeq_i D$ for all coalitions C, D. Let $\mathrm{Ext}(\succeq_i^{+0-})$ be the set of all complete preference relations extending $\succeq_i^{+0-}$.*

Now we can define games where each player has friends, enemies, and neutral co-players, and preferences over the former two sets such that we can derive each player's preference relation as introduced in the previous section.

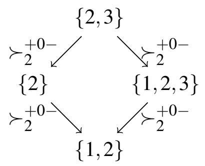
DEFINITION 8. *An* FEN-hedonic game *is a tuple $H = \langle A, \trianglerighteq_1^{+0-}, \dots, \trianglerighteq_n^{+0-} \rangle$, where $A = \{1, 2, \dots, n\}$ is a set of players, and $\trianglerighteq_i^{+0-}$ gives the ordinal preferences with thresholds of player $i \in A$ as defined in Definition 1.*

Definition 9. *Let α be a stability concept for hedonic games, $\langle A, \trianglerighteq_1^{+0-}, \dots, \trianglerighteq_n^{+0-} \rangle$ be an FEN-hedonic game and Γ be a coalition structure. Γ satisfies* possible α *if and only if there exists a profile $\langle \succeq_1, \dots, \succeq_n \rangle$ in $\times_{i=1}^{n} \mathrm{Ext}(\succeq_i^{+0-})$ such that $\langle A, \succeq_1, \dots, \succeq_n \rangle$ satisfies α. Γ satisfies* necessary α *if and only if for each $\langle \succeq_1, \dots, \succeq_n \rangle$ in $\times_{i=1}^{n} \mathrm{Ext}(\succeq_i^{+0-})$, $\langle A, \succeq_1, \dots, \succeq_n \rangle$ satisfies α.*

Example 10. *Let $A = \{1,2,3\}$, $\trianglerighteq_1^{+0-} = (2 \triangleright_1 3 \mid \mid)$, $\trianglerighteq_2^{+0-} = (3 \mid \mid 1)$, and $\trianglerighteq_3^{+0-} = (1 \mid 2 \mid)$.*
The generalized Bossong–Schweigert orders are

$$\{1,2,3\} \succ_1^{+0-} \{1,2\} \succ_1^{+0-} \{1,3\} \succ_1^{+0-} \{1\}$$

for player 1,



for player 2, and for player 3

$$\{1,3\} \sim_3^{+0-} \{1,2,3\} \succ_3^{+0-} \{3\} \sim_3^{+0-} \{2,3\}.$$

So, two preferences are already complete, and there are three complete preferences extending $\succeq_2^{+0-}$, one setting $\{2\} \succ_2 \{1,2,3\}$, another setting $\{2\} \sim_2 \{1,2,3\}$, and the third setting $\{1,2,3\} \succ_2 \{2\}$, leaving all other relations the same.

4.1 Properties and Characterizations

Observe first that there always exists a necessarily individually rational coalition structure (namely, the coalition structure where every agent is alone).

Proposition 11. *Consider an FEN-hedonic game $\langle A, \trianglerighteq_1^{+0-}, \dots, \trianglerighteq_n^{+0-} \rangle$.*

1. *A coalition structure Γ is (necessarily and possibly) perfect if and only if for each player i, $A_i^+ \subseteq \Gamma(i)$ and $A_i^- \cap \Gamma(i) = \emptyset$.*[2]

2. *A coalition structure Γ is possibly individually rational if and only if for each $i \in A$, $\Gamma(i)$ contains at least a friend of i's or only neutral agents.*

3. *A coalition structure Γ is necessarily individually rational if and only if for each $i \in A$, $\Gamma(i)$ does not contain any enemies of i's.*

4. *A coalition structure Γ is necessarily individually stable if and only if it is necessarily individually rational and no player i can join a coalition that she would possibly prefer and the members of which do not see her as an enemy.*

Proof. 1. A coalition structure is perfect if and only if each player is in one of her favorite coalitions, that is, each player is together with all her friends and no enemies.

2. For each $i \in A$, i necessarily prefers $\{i\}$ to $\Gamma(i)$ if and only if $\Gamma(i)$ contains no friend and at least one enemy of i's.

3. For each $i \in A$, i possibly prefers $\{i\}$ to $\Gamma(i)$ if and only if $\Gamma(i)$ contains an enemy of i's.

4. Note that a player j possibly prefers a coalition C to $C \cup \{i\}$ if and only if j necessarily prefers C to $C \cup \{i\}$ if and only if i is an enemy of j's. Assume that Γ is necessarily individually stable. Then, for each $i \in A$, if i prefers to move to another (possibly empty) coalition C in Γ, there is a player in C that prefers player i not being in the coalition. If C is empty, there is no such player, thus, Γ has to be individually rational. Hence, C is nonempty and there has to be a player in C that sees i as an enemy. Now assume that Γ is not individually stable, that is, there is a player i and a coalition $C \in \Gamma \cup \{\emptyset\}$ such that i prefers $C \cup \{i\}$ to $\Gamma(i)$ and, for each $j \in C$, $C \cup \{i\} \succeq_j C$. If $C = \emptyset$, then Γ is not individually rational. Otherwise, each j does not see i as an enemy.

This completes the proof. ❑

[2]As a consequence, a possibly perfect coalition structure in an FEN-hedonic game is always necessarily perfect.

Note that a similar characterization holds for contractually individual stability, where additionally to the conditions of individual stability, it is required that no j in $\Gamma(i)$ considers i a friend.

Example 12. *Consider the FEN-hedonic game from Example 10. Observe that there does not exist a (possibly) perfect coalition structure. While $\{\{1,2,3\}\}$ is possibly Nash stable, there does not exist a necessarily Nash stable coalition structure, as in each of five cases, player* 1 *or player* 2*, at least possibly, wants to move to another coalition. Coalition structure $\{\{1,2,3\}\}$ is possibly individually rational, but not necessarily due to player* 2*; $\{\{1,2\},\{3\}\}$ is not possibly individually rational; the other three coalition structures are necessarily individual rational.*

For $\{\{1,3\},\{2\}\}$ it holds that player 2 *possibly wants to move to $\{1,3\}$ and* 1 *and* 2 *do not see* 2 *as an enemy, thus necessary individual stability is not satisfied. Also, since in $\{2\}$ there is no other player who considers* 2 *a friend, contractually individual stability is not satisfied either. Observe that this coalition structure is, however, possibly individually stable.*

Coalition structure $\{\{1\},\{2,3\}\}$ is not necessarily individually stable, as player 3 *wants to move to $\{1,3\}$ where* 1 *welcomes him. Player* 2*, however, considers* 3 *a friend, thus, as* 2 *does not want to move, and* 1 *is considered an enemy by* 2 *when moving to $\{2,3\}$, this coalition structure is contractually individually stable.*

4.2 Complexity of Possible and Necessary Stability Problems

We are interested in axiomatic properties and characterizations of stability concepts in FEN-hedonic games. However, for some concepts no general statements can be made as to whether there exists a coalition structure satisfying a stability concept α (possibly or necessarily). In these cases we ask how hard it is to decide whether for a given FEN-hedonic game a given coalition structure possibly or necessarily satisfies α, and to decide whether there exists a coalition structure in a given FEN-hedonic game that possibly or necessarily satisfies α. Similar questions are often analyzed in the context of hedonic games [24, 3, 18]. Here, we redefine the verification and existence problems to the notions of possible and necessary existence.

Note that two interpretations of necessary existence can be distinguished, the first one asking whether there always exists a coalition structure that satisfies α, while the second one is asking whether a particular coalition structure necessarily satisfies α. Intuitively this distinction makes sense, since in the first case the setting might provide a central authority with partial knowledge of the agents' preferences and require the knowledge that whatever the possible preferences are, there is always some coalition structure satisfying α; in the second case, the choice of coalition structure is independent of the agents' possible preferences.

EXAMPLE 13. *For example, consider the following game with three players,* $A = \{1,2,3\}$*, with* $\trianglerighteq_1^{+0-} = (2 \mid 3 \mid)$*,* $\trianglerighteq_2^{+0-} = (1 \mid 3 \mid)$*, and* $\trianglerighteq_3^{+0-} = (1 \mid \mid 2)$*. We obtain the following generalized Bossong–Schweigert orders:* $\{1,2\} \sim_1 \{1,2,3\} \succ_1 \{1\} \sim_1 \{1,3\}$*,* $\{1,2\} \sim_2 \{1,2,3\} \succ_2 \{2\} \sim_2 \{2,3\}$*, and* $\{1,3\} \succ_3 \{3\} \succ_3 \{2,3\}$ *and* $\{1,3\} \succ_3 \{1,2,3\} \succ_3 \{2,3\}$*, while* 3 *is undecided between* $\{3\}$ *and* $\{1,2,3\}$*. Any coalition structure in which players* 1 *and* 2 *are not in the same coalition cannot possibly be Nash stable. On the one hand,* $\{\{1,2\},\{3\}\}$ *is Nash stable if and only if an extension provides* $\{3\} \succeq_3 \{1,2,3\}$*. On the other hand,* $\{\{1,2,3\}\}$ *is Nash stable if and only if* $\{1,2,3\} \succeq_3 \{3\}$ *in an extension. Thus, for every extension, there certainly exists a Nash stable coalition structure. However, there is no necessarily Nash stable coalition structure.*

Here, we focus on the second interpretation. Possible existence is unambiguous, asking whether there is some coalition structure satisfying α for some extension.

PROPOSITION 14. *All problems regarding perfection are in* P.

PROOF. Verfication of whether a coalition structure is possibly and necessarily perfect is easy by Proposition 11.

Existence can be decided by, e.g., the following algorithm: Start with player 1 and let $\Gamma(1) := \{1\} \cup A_1^+$. Sequentially, for each $i \in \Gamma(1)$, add A_i^+ to $\Gamma(1)$ until there are no further possible changes. Check whether, for each $i \in \Gamma(1)$, $A_i^- \cap \Gamma(1) = \emptyset$. If not, output "there is no perfect coalition structure"; if so, start over with $A \smallsetminus \Gamma(1)$. It might be the case that a friend cannot be added, because he is already assigned to another coalition. If he is on his own, add him anyway; otherwise, output "there is no perfect coalition structure." Continue until each player is allocated to a coalition. Then, output "there is a perfect coalition structure."

Note that this algorithm works in polynomial time. ❑

All problems regarding individual rationality are in P by the characterizations in Proposition 11 and the observation preceding it.

Proposition 11 does not provide a characterization of Nash stability. Nevertheless, it can be verified in polynomial time whether a given coalition structure in a given FEN-hedonic game is necessarily Nash stable.

LEMMA 15. *The verification problem for possible Nash stability is in* P.

PROOF. Given an FEN-hedonic game and a coalition structure Γ, verify the following steps for each $i \in A$: For each (of at most n coalitions) $C \in \Gamma \cup \{\emptyset\}$, $C \neq \Gamma(i)$, determine the relation between $\Gamma(i)$ and $C \cup \{i\}$. This can be done in polynomial time by Proposition 6.1. If $C \cup \{i\} \succ_i \Gamma(i)$, output "$\Gamma$ is not Nash stable." If the relation is undecided, output "Γ is possibly not Nash stable." Otherwise, if this is not true for any player or coalition in $\Gamma \cup \{\emptyset\}$, output "$\Gamma$ is necessarily Nash stable." ❑

By the characterizations in Proposition 11, similar algorithms work for individual and contractually individual stability. Note that this cannot easily be transferred to possible Nash stability, since resolving an undecided relation might influence another relation for the same player.

THEOREM 16. *The problem of whether there exists a possibly Nash stable coalition structure in a given FEN-hedonic game is* NP*-complete.*

PROOF. The problem belongs to NP, since it is enough to check whether there exists a coalition structure of A and an extension persuing the GBS-extension such that for each player $i \in A$ and each coalition $C \in \Gamma$, $\Gamma(i) \succeq_i C \cup \{i\}$. The latter can be tested in time polynomial in $n = \|A\|$, since there are at most n coalitions in Γ and the relation between two coalitions from a common player's perspective can be decided in polynomial time by Proposition 6.1.

NP-hardness can be shown via a polynomial-time many-one reduction from EXACT-COVER-BY-THREE-SETS (X3C, see [17]): Given a set R with $3m$ elements and a family $\mathscr{S}$ of subsets $s \subseteq R$ with $\|s\| = 3$, is there an exact cover of R in $\mathscr{S}$, that is, is there a subset $S \subseteq \mathscr{S}$ such that $\cup_{s \in S} s = R$ and $\|S\| = m$? Without loss of generality it can be assumed that $m \geq 2$ and each element in R occurs at most three times in a set in $\mathscr{S}$. Given such an X3C instance, we construct the following game. This construction is inspired by the construction of the proof that it is NP-hard to decide whether there exists a Nash stable coalition structure in an additively separable hedonic game [22, Theorem 3]. Here, however, several adjustments have to be made in order to guarantee necessary preferences over coalitions.[3] Let

$$A = \{\alpha_i \mid 1 \leq i \leq 3m-1\} \cup \{\beta_r \mid r \in R\} \cup \{\zeta_{s,k} \mid s \in \mathscr{S}, 1 \leq k \leq 3m-2\}$$

and let the players' preferences be defined as follows, where in player i's preference and for a set $X = \{a_1, a_2, \ldots, a_x\}$, $X_\sim$ denotes $a_1 \sim_i a_2 \sim_i \cdots \sim_i a_x$

- $\trianglerighteq_{\alpha_i}^{+0-} = \big(\alpha_{i+1} \mid \{\alpha_j : i \neq j \neq i+1\}_\sim \mid \{\text{other players}\}_\sim\big)$, for each i, $1 \leq i \leq 3m-2$,

  $\trianglerighteq_{\alpha_{3m-1}}^{+0-} = \big(\mid \{\alpha_j : j \neq 3m-1\}_\sim \mid \{\text{other players}\}_\sim\big)$,

- $\trianglerighteq_{\beta_r}^{+0-} = \big(\{\alpha_i : 1 \leq i \leq 3m-1\}_\sim \rhd_{\beta_r} \bigcup_{r \in s} Q_{s\sim} \rhd_{\beta_r} \{\beta_{r'} : r' \neq r\}_\sim \mid \mid \{\text{other players}\}_\sim\big)$, for each $r \in R$,

- $\trianglerighteq_{\zeta_{s,k}}^{+0-} = \big(\zeta_{s,k+1} \mid \{\zeta_{s,k'} : k \neq k' \neq k+1\} \cup \{\beta_r : r \in s\}_\sim \mid \{\text{other players}\}_\sim\big)$, for each $s \in \mathscr{S}$, and k, $1 \leq k \leq 3m-3$,

  $\trianglerighteq_{\zeta_{s,3m-2}}^{+0-} = \big(\mid \{\zeta_{s,k'} : k' \neq 3m-2\} \cup \{\beta_r : r \in s\}_\sim \mid \{\text{other players}\}_\sim\big)$, for each $s \in \mathscr{S}$

where $Q_s = \{\zeta_{s,k} \mid 1 \leq k \leq 3m-2\}$ for each $s \in \mathscr{S}$. Moreover, let $P_s = \{\beta_r \mid r \in s\} \cup Q_s$. This profile can be constructed in polynomial time, since there are $n \leq 3m + 3m + 3m \cdot (3m-2)$ players, and each player's preference can be written in linear time in n.

We now show that $(R, \mathscr{S})$ is a positive instance for X3C if and only if there exists a possibly Nash stable coalition structure in the GBS-extension of the constructed game.

Only if: Assume there exists a solution S for $(R, \mathscr{S})$. Consider the coalition structure

$$\Gamma = \{\{\alpha_i \mid 1 \leq i \leq 3m-1\}\} \cup \{P_s \mid s \in S\} \cup \{Q_s \mid s \notin S\}.$$

[3]Consider, e.g., a coalition $\{i, f, e\}$ where player i has a positive value for f, and a negative value for e. In comparison to $\{i\}$ this coalition is preferred by player i if f has a greater absolute value than e in the additively separable representation, is considered indifferent if f and e have the same absolute value, and is less preferred otherwise. If we do not provide values but ordinal preferences and thresholds and consider f as a friend and e as an enemy of i's, $\{i, f, e\}$ and $\{i\}$ are incomparable from i's perspective; thus, all three scenarios are possible in an extension persuing GBS.

By a close look at all possibly empty coalitions in Γ it can be seen that no α_i, $1 \leq i \leq 3m-1$, and no $\zeta_{s,k}$, $s \in \mathscr{S}$, $1 \leq k \leq 3m-2$, wants to move, and each β_r, $r \in R$, possibly does not want wo move, thus, Γ is possibly Nash stable.

If: Assume there is a possibly Nash stable coalition structure Γ. Ruling out, one by one, coalitions that cannot be contained in Γ, it can be shown that for each $r \in R$, there exists an $s \in \mathscr{S}$ such that $\Gamma(\beta_r) = P_s$, which means that there is an exact cover of R in $\mathscr{S}$. ❑

By similar, but not trivially the same methods we can show that the problem of necessary Nash stable existence is NP-complete.

5. CHALLENGES

In order to give a prospect to future work we provide initial thoughts on further stability concepts as well as comparability functions in order to deal with incomparabilities.

5.1 Further Stability Concepts

So far we have focused on single-player deviations. In this section, we give a prospect to other stability concepts such as group deviations, Pareto optimality, and popularity. A coalition structure Γ is called *core stable* if for each coalition $A' \subseteq A$, there exists a player $i \in A'$ such that $\Gamma(i) \succeq_i A'$. A coalition structure Γ is called *Pareto-optimal* if for each coalition structure Δ, there exists a player $i \in A$ such that $\Gamma(i) \succ_i \Delta(i)$, or for each player $j \in A$, $\Gamma(j) \sim_j \Delta(j)$. A coalition structure Γ is called *popular* if for each coalition structure Δ, the number of players i with $\Gamma(i) \succ_i \Delta(i)$ is at least as large as the number of players j with $\Delta(j) \succ_j \Gamma(j)$. We furthermore introduce the notion of *strict popularity*. A coalition structure Γ is called *strictly popular* if it *beats* each other coalition structure Δ *in pairwise comparison*,[4] that is,

$$\|\{i \in A \mid \Gamma(i) \succ_i \Delta(i)\}\| > \|\{j \in A \mid \Delta(j) \succ_j \Gamma(j)\}\|.$$

For each extension there exists a Pareto-optimal coalition structure (perhaps a different one for different extensions). Observe that if there exists a necessarily strictly popular coalition structure, it is unique, whereas there can be more than one possibly strictly popular coalition structure.

If there exists a necessarily strictly popular coalition structure, it is necessarily Pareto optimal. If there exist possibly strictly popular coalition structures, each of them is possibly Pareto-optimal. A necessarily strictly popular coalition structure does not need to be possibly individually rational. Even if the possible core is nonempty, a necessarily strictly popular coalition structure does not need to be possibly core stable. The same holds for the concepts of Nash stability, individual stability, contractual individual stability, and strict core stability. If there exists a unique perfect partition, it is necessarily the unique necessarily strictly popular coalition structure.

With techniques related to those in the proof of Theorem 16, we can show that the questions of whether a given coalition structure is possibly strictly popular or popular or Pareto-optimal are coNP-hard, necessarily strictly popular or popular or Pareto-optimal are coNP-complete, and it is coNP-hard to decide whether there exists a strictly popular coalition structure, for both, the possible and the necessary case.

Moreover, coNP-hardness of the problems of whether a given coalition structure is core stable or strictly core stable can be shown with help of the reduction from CLIQUE to the core stability verification problem in the enemy-based representation [21]. Note that this representation is a special case of the representation with ordinal preferences and thresholds, where there are no neutral agents and only indifferences between all friends and between all enemies in a player's preference. Furthermore, note that the enemy-based-extension [15] is a possible extension in $\times_{i=1}^{n} \mathrm{Ext}(\succeq_i^{+0-})$. While a "clique" of friends is necessarily preferred by all members to a coalition containing fewer friends or even more enemies, there is not necessarily a blocking coalition in the construction if there is no such clique (for example, there is no blocking coalition in the enemy-based extension).

[4] This notion is adapted from the voting-theoretic term of *Condorcet winner*: Such a candidate wins an election if and only if she beats each other candidate in pairwise comparison.

5.2 Breaking Incomparabilities with Borda-Like Scoring Vectors

In this section, we present a mechanism for determining the relation between coalitions that are not comparable via the ordering that the Bossong–Schweigert extension induces.

Every player has to evaluate a total preference order over all possible coalitions she might be part of, so we define a so-called *comparability function* (short CF) for a fixed player, say $i \in A$. One possibility to do so is to use scoring vectors that assign values to the players in $A \setminus \{i\}$ depending on the position they have in the weak ranking with double threshold of player i. In particular, for the notions presented in Definition 1, we define the following variants of Borda-like scoring vectors.

We define scoring vectors $w_i : A \to \mathbb{Z}$ assigning points to the players in the sets of friends, neutral agents, and enemies of agent i, according to their positions in ranking $\trianglerighteq_i^{+0-}$, compatible with $\trianglerighteq_i^{+0-}$ as in Proposition 6. In more detail, we propose the following possibilities, distinguishing between an "optimistic" and a "pessimistic" case (see also the optimistic and pessimistic scoring model for modified Borda voting, due to Baumeister et al. [6]), and for each we have a regular and a strong variant. Recall that we have n agents in total. Suppose that i's friends, A_i^+, are ordered as follows: $\trianglerighteq_i^+ = A_{i,1}^+ \triangleright_i^+ A_{i,2}^+ \triangleright_i^+ \cdots \triangleright_i^+ A_{i,\ell}^+$, where each $A_{i,j}^+$ contains some agents i is indifferent about. Similarly, suppose that i's enemies, A_i^-, are ordered as follows: $\trianglerighteq_i^- = A_{i,1}^- \triangleright_i^- A_{i,2}^- \triangleright_i^- \cdots \triangleright_i^- A_{i,m}^-$, where each $A_{i,j}^-$ contains agents i is indifferent about. Here, we do not explicitly define all 16 combinations of (strictly) friend/enemy-optimistic/pessimistic scoring vectors. For instance, consider the cases of a strongly friend-optimistic and a strongly enemy-pessimistic setting.

DEFINITION 17. *Let A be a set of players and $\trianglerighteq_i^{+0-}$ be player i's preference relation. Let $w_i : A \to \mathbb{Z}$, compatible with $\trianglerighteq_i^{+0-}$, assign n points to each agent in $A_{i,1}^+$, $n-1$ points to each agent in $A_{i,2}^+$, ..., and $n-\ell+1$ points to each agent in $A_{i,\ell}^+$. Moreover, let each agent in $A_{i,m}^-$ get $-n$ points, each agent in $A_{i,m-1}^-$ get $-n+1$ points, ..., and each agent in $A_{i,1}^-$ get $-(n-m+1)$ points. Then, we call w_i* strongly friend-optimistic *and* strongly enemy-pessimistic.

We now define a numerical comparability function that captures the notion of Borda-like scoring.

DEFINITION 18. *For each fixed agent $i \in A$ and for every fixed choice of scoring vectors w_i, the Borda-like CF*

$$f_{\mathrm{Borda}}^i : \{C \subseteq A \mid i \in C\} \to \mathbb{Z}$$

maps every coalition C containing i to the sum of the scores the agents in C obtain from w_i. The value of a coalition *$C \subseteq A$ is defined as $F_{\mathrm{Borda}}(C) = \sum_{i \in C} f_{\mathrm{Borda}}^i(C)$.*

	$\{1,2,3,4\}$	$\{1,2,3\} \sim \{1,2,4\}$	$\{1,2,3,4,5\}$
v_1	16	11	11

Table 1: Values of some coalitions in player 1's view for the scoring vector $\mathbf{v_1} = (*, \mathbf{6}, \mathbf{5}, \mathbf{5}, -\mathbf{5}, -\mathbf{6})$

EXAMPLE 19. *Let $A = \{1,2,3,4,5,6\}$ and the preference with thresholds from Example 2: $\trianglerighteq_1^{+0-} = (2 \rhd_1 3 \sim_1 4 | \mid 5 \rhd_1 6)$. Figure 1 shows the graph corresponding to the Bossong–Schweigert extension of this preference. For six agents and $\trianglerighteq_1^{+0-}$, the scoring vector in the strongly friend-optimistic and strongly enemy-pessimistic setting is $v_1 = (*, 6, 5, 5, -5, -6)$.*

Table 1 shows the scores of some of the coalitions from agent 1's view with scoring vector v_1.

To determine the overall value of all coalitions, the individual scores of the other five agents have to be determined as well.

The following observation follows directly from the definitions above.

OBSERVATION 20. *For each player $i \in A$, the comparability function f^i_{Borda} preserves those rankings that are induced by the Bossong–Schweigert extension.*

Furthermore, a game that is induced by comparability function F_{Borda} (as an extension) is additively separable.

This observation allows us to use known results for the complexity of the various stability problems in general additive separable hedonic games (ASHGs, for short), which have been studied intensely (see, e.g., the work by Aziz et al. [3] for a comprehensive overview). Upper bounds can be transferred directly from known results for general ASHGs. Whether the known lower bounds also hold for our special games, however, has to be checked separately. For certain settings of scoring vectors (often all 16 combinations at once), we were able to adapt known hardness proofs for some of the stability concepts to our setting. Although the cardinaliziation of the ordinal preferences might suggest that verification and existence of a stability concept become more tractable. However, for the strongly friend-pessimistic and strongly friend-optimistic case, we obtain the same complexity results as for Nash stability: verification is decidable in P, existence NP-complete. The problem of whether there exists a core stable coalition structure in a given FEN-hedonic game is even Σ_2^p-complete.

6. CONCLUSIONS AND FUTURE WORK

In this paper we introduce a new representation of preferences in hedonic games using the Bossong–Schweigert principle to extend the players' preferences over the other players to preferences over the coalitions. This generalized Bossong–Schweigert extension principle to positive and negative items (here called friends and enemies), and neutral items, is new and it is original in itself, independently of its use in hedonic games.

We have then looked at several stability concepts in hedonic games with such preferences. The problem of remaining incomparabilities is tackled in two ways: Firstly, by letting these incomparabilities unresolved and introducing known stability concepts with respect to notions of necessity and possibility, and secondly by introducing a comparability function based on Borda-like scoring vectors.

For both approaches we analyze for the induced games the complexity of the existence and verification of well-known stability concepts. So far, with the help of these solution concepts we can verify if a coalition structure is a "good" solution, compare two coalition structures, and decide, whether there even exists such a coalition structure—sometimes at great cost in terms of complexity.

Besides completing the analysis initiated here (such as considering other solution concepts and solving remaining open problems), we suggest for future work introducing the notion of *partition correspondences* with the purpose to actually identify "good" coalition structures as an output. In contrast to the original idea of hedonic games where coalitions form in a decentralized manner, here a central correspondence is used, in order to decide which coalitions will work together. This might, for example, be the case in a setting where the head of a department has to divide a group of employees into teams. The teams should be stable, in the sense that the team members are as happy as possible with their group to create a good working atmosphere.

Acknowledgments

We thank the anonymous reviewers for their helpful comments. This work was supported in part by DFG grants RO-1202/14-1 and RO-1202/15-1, by the ANR project 14-CE24-0007-01 - CoCoRICo-CoDec, by the DAAD-PPP / PHC PROCOPE programme entitled "Fair Division of Indivisible Goods: Incomplete Preferences, Communication Protocols and Computational Resistance to Strategic Behaviour," and by COST Action IC1205 on Computational Social Choice.

REFERENCES

[1] H. Aziz, F. Brandt, and P. Harrenstein. Pareto optimality in coalition formation. *Games and Economic Behavior*, 82:562–581, 2013.

[2] H. Aziz, F. Brandt, and P. Harrenstein. Fractional hedonic games. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems*, pages 5–12. IFAAMAS, May 2014.

[3] H. Aziz, F. Brandt, and H. Seedig. Computing desirable partitions in additively separable hedonic games. *Artificial Intelligence*, 195:316–334, 2013.

[4] C. Ballester. NP-completeness in hedonic games. *Games and Economic Behavior*, 49(1):1–30, 2004.

[5] S. Banerjee, H. Konishi, and T. Sönmez. Core in a simple coalition formation game. *Social Choice and Welfare*, 18:135–153, 2001.

[6] D. Baumeister, P. Faliszewski, J. Lang, and J. Rothe. Campaigns for lazy voters: Truncated ballots. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, pages 577–584. IFAAMAS, June 2012.

[7] A. Bogomolnaia and M. Jackson. The stability of hedonic coalition structures. *Games and Economic Behavior*, 38(2):201–230, 2002.

[8] U. Bossong and D. Schweigert. Minimal paths on ordered graphs. *Mathematica Slovaca*, 56(1):23–31, 2006.

[9] S. Bouveret, U. Endriss, and J. Lang. Fair division under ordinal preferences: Computing envy-free allocations of indivisible goods. In *Proceedings of the 19th European Conference on Artificial Intelligence*, pages 387–392. IOS Press, Aug. 2010.

[10] K. Cechlárová and J. Hajduková. Computational complexity of stable partitions with B-preferences. *International Journal of Game Theory*, 31(3):353–364, 2003.

[11] K. Cechlárová and J. Hajduková. Stable partitions with $\mathscr{W}$-preferences. *Discrete Applied Mathematics*, 138(3):333–347, 2004.

[12] K. Cechlárová and A. Romero-Medina. Stability in coalition formation games. *International Journal of Game Theory*, 29(4):487–494, 2001.

[13] A. Darmann, E. Elkind, S. Kurz, J. Lang, J. Schauer, and G. Woeginger. Group activity selection problem. In *Proceedings of the 8th International Workshop on Internet & Network Economics*, pages 156–169. Springer-Verlag *Lecture Notes in Computer Science #7695*, Dec. 2012.

[14] C. Delort, O. Spanjaard, and P. Weng. Committee selection with a weight constraint based on a pairwise dominance relation. In *Proceedings of the 2nd International Conference on Algorithmic Decision Theory*, pages 28–41. Springer-Verlag *Lecture Notes in Artificial Intelligence #6992*, Oct. 2011.

[15] D. Dimitrov, P. Borm, R. Hendrickx, and S. Sung. Simple priorities and core stability in hedonic games. *Social Choice and Welfare*, 26(2):421–433, 2006.

[16] E. Elkind and M. Wooldridge. Hedonic coalition nets. In *Proceedings of the 8th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 417–424. IFAAMAS, May 2009.

[17] M. Garey and D. Johnson. *Computers and Intractibility: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, 1979.

[18] A. Rey, J. Rothe, H. Schadrack, and L. Schend. Toward the complexity of the existence of wonderfully stable partitions and strictly core stable coalition structures in hedonic games. In *Proceedings of the 11th Conference on Logic and the Foundations of Game and Decision Theory*, July 2014.

[19] A. Roth. Common and conflicting interests in two-sided matching markets. *European Economic Review*, 27(1):75–96, 1985.

[20] A. Roth and M. Sotomayor. *Two-sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press, 1992.

[21] S. Sung and D. Dimitrov. On core membership testing for hedonic coalition formation games. *Operations Research Letters*, 35(2):155–158, 2007.

[22] S. Sung and D. Dimitrov. Computational complexity in additive hedonic games. *European Journal of Operational Research*, 203(3):635–639, 2010.

[23] G. Woeginger. Core stability in hedonic coalition formation. In *Proceedings of the 39th Conference on Current Trends in Theory and Practice of Computer Science*, pages 33–50. Springer-Verlag *Lecture Notes in Computer Science #7741*, Jan. 2013.

[24] G. Woeginger. A hardness result for core stability in additive hedonic games. *Mathematical Social Sciences*, 65(2):101–104, 2013.

CHAPTER 5

Borda-induced Hedonic Games with Friends, Enemies, and Neutral Players

Summary

In contrast to the approach of using the two modularities 'possibility' and 'necessity' from the previous chapter, we tackle the problem of how to determine meaningful outcomes with the help of ideas from social choice theory. In particular, we use the idea of scoring vectors that originate in voting theory. A scoring vector is an ordered n-dimensional vector with natural numbers as elements, used to derive points from votes over candidates in a voting scenario. More specifically, we use the idea of the Borda scoring vector, which is a strictly declining scoring vector. In the following paper we will use the principles of these Borda scoring vectors once we have to dissolve incomparabilities between coalitions. To this end, we define eight principles, four for the part of the weak rankings with double threshold that describe the friends, and one for the part that describes the enemies of the players. Those principles deviate in the way they promote friends and enemies, with the help of different scores assigned to different positions in the rankings. In the end, one can simply add up such scores to receive a total score of a coalition for a fixed player, which afterwards can be compared to the coalitions that were not comparable beforehand. The following paper continues with an analysis of the verification and existence decision problems for Borda-induced hedonic games in regards to the stability concepts of perfectness, (contractual) individual stability, Nash stability, and (strict) core stability.

Contribution and Preceding Versions

The idea, model, and writing was done jointly with my coauthors. Additionally, Examples 2, 4, 7 and 14, Lemmata 9 and 11, Propositions 20 and 21, and the result for the combination $(\mathbf{fo}, \mathbf{eo})$ of Theorems 28 and 29 has to be attributed to my contribution. This paper extends the preliminary paper [40]

Publication - Rothe, Schadrack, and Schend [62]

J. Rothe, H. Schadrack, and L. Schend. Borda-induced hedonic games with friends, enemies, and neutral players. *Mathematical Social Sciences*, 96:21–36, 2018

Mathematical Social Sciences 96 (2018) 21–36

Contents lists available at ScienceDirect

Mathematical Social Sciences

journal homepage: www.elsevier.com/locate/mss

Borda-induced hedonic games with friends, enemies, and neutral players

Jörg Rothe *, Hilmar Schadrack, Lena Schend

Institut für Informatik, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany

H I G H L I G H T S

- In FEN-hedonic games, players are divided into friends, enemies, and neutral players.
- We propose Borda-induced FEN-hedonic games to extend partial to complete orders.
- We study the complexity of existence and verification for common solution concepts.

A R T I C L E   I N F O

Article history:
Received 20 July 2017
Received in revised form 30 March 2018
Accepted 20 August 2018
Available online 29 August 2018

A B S T R A C T

In a FEN-hedonic game, each player partitions the set of other players into friends, enemies, and neutral players and ranks her friends and enemies. Assuming that preferences are monotonic with respect to adding friends and antimonotonic with respect to adding enemies, we use bipolar responsive extensions to lift the players' rankings of players to their (partial) preferences over coalitions. We propose cardinal comparability functions in order to extend partial to complete preference orders consistent with these polarized responsive orders, in particular focusing on Borda-induced FEN-hedonic games. For a number of common solution concepts, we study the computational complexity of the existence and the verification problem.

1. Introduction

In a hedonic game, each player has preferences over the coalitions she can join, and a central question is which coalition structure will form and remain stable. Among the well-known stability concepts we will study for hedonic games are Nash stability, individual stability, contractual individual stability, and core stability. However, a critical issue is how to represent the players' preferences over all coalitions containing them. For each of n players, there are 2^{n-1} coalitions containing this player, so listing them all explicitly to express one's preferences does not make sense. This issue has been addressed in previous work, for example, by assuming that just a small part of the preference relation is expressed by each player, which then is extended to a complete preference relation over coalitions via some appropriate extension principle.

The literature on hedonic games (see, e.g., the recent book chapters by Aziz and Savani, 2016 and Elkind and Rothe, 2015 or the survey by Woeginger, 2013a) contains various assumptions about what kind of input the players are required to specify (and, if appropriate, what kind of preference extension is to be used). For example, Ballester (2004) proposed the *individually rational encoding* where players give their preferences only over those coalitions they prefer to being alone, and also the *anonymous encoding* (see also Darmann et al., 2012) where players' preferences depend only on the number of players in their coalition (and not on who these players are). Both these encodings are ordinal.[1] Some representations make use of logical formulas, such as the *hedonic coalition nets* (a cardinal encoding proposed by Elkind and Wooldridge, 2009 where players specify their utilities for coalitions via a set of weighted logical formulas), or the *boolean hedonic games* (a dichotomous encoding proposed and studied by Aziz et al., 2016b and Peters, 2016 where players partition the coalitions containing them into two classes, preferring one to the other while being indifferent between the coalitions inside each class).

Other encodings of hedonic games are based on requiring each player to specify a ranking or a numerical evaluation of *single*

* Corresponding author.
E-mail address: rothe@cs.uni-duesseldorf.de (J. Rothe).
URLs: http://ccc.cs.uni-duesseldorf.de/~rothe (J. Rothe), https://ccc.cs.uni-duesseldorf.de/~schadrack (H. Schadrack), https://ccc.cs.uni-duesseldorf.de/~schend (L. Schend).

https://doi.org/10.1016/j.mathsocsci.2018.08.003
0165-4896/

[1] The advantages of representing preferences ordinally have been extensively discussed both in social choice theory (see, e.g., the work of Caragiannis and Procaccia, 2011) and, more recently, in fair division (see, e.g., the work of Baumeister et al., 2017 and Nguyen et al., 2018).

players only, which is then extended to rank or evaluate *coalitions* of players via some extension principle. For example, the *singleton encoding* due to Cechlárová and Romero-Medina (2001) (see also Cechlárová and Hajduková, 2003, 2004) is an ordinal approach that extends the ranking of players to preferences over coalitions in an optimistic or a pessimistic way (see Section 2.1 for the formal definition). The well-studied *additive encoding* (Sung and Dimitrov, 2007, 2010; Aziz et al., 2013b; Woeginger, 2013b) and the more recent notion of *fractional hedonic game* (Aziz et al., 2014; Bilò et al., 2014, 2015) are cardinal approaches that require each player to assign numerical values to players from which appropriate utilities for coalitions are derived. And the *friends-and-enemies encoding* due to Dimitrov et al. (2006) (see also Sung and Dimitrov, 2007; Rey et al., 2016; Nguyen et al., 2016) is a dichotomous approach where players partition the other players into a set of their friends and a set of their enemies: Under the *friend-oriented preference extension*, coalition C is preferred to coalition D if the player has more friends in C than in D, or has the same number of friends in C and D but fewer enemies in C than in D, whereas under the *enemy-oriented preference extension*, C is preferred to D if the player has fewer enemies in C than in D, or has the same number of enemies in C and D but more friends in C than in D. All these encodings of hedonic games have their advantages and their disadvantages; for example, the individual rational encoding may still be exponential-size in the worst case, while the singleton encoding as well as additive and fractional hedonic games require some domain restriction and so are not fully expressive.

A downside of the friends-and-enemies encoding, on the other hand, is that players cannot express ordinal preferences inside their sets of friends or enemies. For instance, if player 1 considers 3 to be a friend and 2 and 4 to be enemies, it is clear that 1 prefers being with 3 to being with either of 2 or 4, but we do not know which of 2 and 4 is despised more by 1. Such a ranking of players is provided by the singleton encoding; however, this ranking does not allow a player to distinguish between friends (whom she would like to join in a coalition) and enemies (whom she would like to avoid in a coalition). For instance, if player 1 ranks her fellow players 3, 2, and 4 in this order, it is clear that 1 would rather be together with 3 than with 2 and would also prefer being with 2 to being with 4, but we do not know whether 1 would like to join any of them or would rather stay alone. To avoid both shortcomings, our approach (originally proposed in the conference predecessor Lang et al., 2015 of this paper) is to combine the singleton encoding with the friends-and-enemies encoding: First, each player partitions the other players into three groups – her friends, her enemies, and her neutral players (whom she does not care about) – and then specifies a ranking of her friends and a ranking of her enemies. We refer to these as *FEN-hedonic games*. To obtain preferences over coalitions of players in such games, we then apply a natural generalization of the *responsive extension principle* (sometimes referred to as the *Bossong–Schweigert extension principle* Bossong and Schweigert, 2006, see also Delort et al., 2011), which gives a partial order over coalitions containing the player at hand. We call this generalization the *polarized responsive extension*.

Responsive preferences have been studied, for example, in the context of bipartite many-to-one matching markets (see, e.g., the work of Roth, 1985 and Roth and Sotomayor, 1992) where participants are compared with one another, even though not by distinguishing friends from enemies. In this context, each agent on the one side has responsive preferences over assignments of the agents on the other side if the assignment containing the most preferred agent is preferred for any two assignments that differ in only one player. Responsive preferences have also been studied for allocation problems, in particular in the context of strategy-proofness (see, e.g., Ehlers and Klaus, 2003; Hatfield, 2009; Nguyen et al., 2018; Aziz et al., 2016a). Informally, under responsive preferences, a set X of items is preferred to another set Y of items if X contains an additional item or if some item in Y is replaced in X by a better item.

One issue with the polarized responsive preferences in FEN-hedonic games is that coalitions in these partial orders can be incomparable (see also the conference version Lang et al., 2015 for details). Our approach to deal with this issue is to define comparability functions in order to determine the relation between incomparable coalitions, focusing on *Borda-induced* FEN-hedonic games.[2] We then study, for various common stability concepts, the existence and the verification problem for Borda-induced FEN-hedonic games in terms of their computational complexity. To this end, we will apply useful metatheorems due to Peters and Elkind (2015), which allows us to close some of the complexity gaps that have been left open in the conference version of this paper (Lang et al., 2015).

Interestingly, as described by Woeginger (2012) in detail, the extensively studied stable matching and stable roommates problems can be seen as special cases of hedonic games where all coalitions are restricted to be of size two. The players present their (additive) preferences simply by ranking the other players. More precisely, in an instance of the *stable matching problem*, we have the same number of male and female players, the male players rank the female players and vice versa, and the goal is to find a *stable* matching between the men and women, i.e., a partition into man–woman pairs that is not blocked by any pair of a man m and a woman w: (m, w) would be blocking a partition if m would prefer w to his current partner and w would prefer m to her current partner in the partition. On the other hand, in an instance of the *stable roommate problem*, we have an even number of (unisex) players, so everyone can be paired with everyone else, and stability again is defined via nonexistence of blocking pairs. Known (complexity) results about these two problems depend on the underlying preferences that can be strict (no ties in the players' rankings) or not and can be complete or not. For complete preferences, stable matchings always exist and can be found in polynomial time, no matter whether they are strict (Gale and Shapley, 1962) or not (Irving, 1994). For strict, incomplete preferences, by slightly modifying the famous Gale–Shapley algorithm (Gale and Shapley, 1962) one can show that stable matchings still always exist and can be found in polynomial time. However, a stable matching may not be *perfect*: There might be matched pairs and, in addition, some singletons with players who could not be assigned an appropriate partner.[3] For the most general case (incomplete preferences with ties), stable matchings still always exist and can be found in polynomial time, but deciding whether there exists a *perfect* stable matching is NP-complete (Manlove et al., 2002), even if every player ranks no more than three acceptable partners (Irving et al., 2009). Regarding the stable roommate problem, Irving's algorithm (Irving, 1985) can be used to decide in polynomial time whether there exists a stable matching whenever we have strict preferences, no matter whether they are complete or not. However, with ties allowed, Ronn (1990) showed that the stable roommate problem is NP-complete (see also the work of Irving and Manlove, 2002). Our study of Borda-induced FEN-hedonic games is remotely related to the classical stable matching and stable roommates problems, but our approach is more general as we allow coalitions of arbitrary size. As is common in the study of hedonic games, we allow ties in the preferences (more to the point, our model is based on the

[2] Borda scoring (Borda, 1781), originally proposed for elections, has also been used, for example, in fair division (Brams et al., 2003; Brams and King, 2005; Bouveret et al., 2010; Baumeister et al., 2017; Nguyen et al., 2017; Kuckuck and Rothe, 2018).

[3] Interestingly, Gale and Sotomayor (1985) show that in *each* stable matching, the same set of men and women are paired up, leaving the same (complementary) set of men and women single.

players' "weak rankings with double threshold" as explained in Definition 1, and these weak rankings are complete).

This paper is organized as follows. In Section 2, we will introduce the needed notions of hedonic games and will give some background on complexity theory. In Section 3, we will first describe FEN-hedonic games and the polarized responsive extension principle and then present the metatheorems due to Peters and Elkind (2015) that we will use later on. To deal with incomparabilities that can result from the polarized responsive extension principle, we will introduce and study Borda-induced FEN-hedonic games in Section 4, and we will study their properties in Section 5 and the computational complexity of the related problems in Section 6. We conclude in Section 7 with stating some open problems and directions of future research.

2. Preliminaries

We provide some background from the theory of hedonic games in Section 2.1 and from complexity theory in Section 2.2.

2.1. Hedonic games

A *hedonic game* $(A, \succeq)$ has a set of players, $A = \{1, 2, \dots, n\}$,[4] and a profile of the players' preferences, $\succeq = (\succeq_1, \succeq_2, \dots, \succeq_n)$, each $\succeq_i$ a weak preference order over all possible coalitions $C \subseteq A$ including this player. More formally, denoting the set of coalitions containing player $i \in A$ by $\mathscr{A}_i$ and letting $C, D \in \mathscr{A}_i$ be two coalitions, we say that *i weakly prefers C to D* if $C \succeq_i D$; we say that *i prefers C to D* (and write $C \succ_i D$) if $C \succeq_i D$ but not $D \succeq_i C$; and we say that *i is indifferent between C and D* (and write $C \sim_i D$) if both $C \succeq_i D$ and $D \succeq_i C$. Given a hedonic game $(A, \succeq)$, a *coalition structure* is a partition Γ of A into coalitions, and $\Gamma(i)$ is the unique coalition in Γ containing player $i \in A$.

Since each player expresses preferences over 2^{n-1} coalitions, the question arises how one can represent hedonic games compactly. Below we list some of the known representations from the literature that will be used to describe our new model.

In an *additively separable hedonic game*, due to Banerjee et al. (2001), each player assigns some real value to each player, i.e., there is a value function $w_i : A \to \mathbb{R}$ for each $i \in A$. The players' preferences in the profile $\succeq = (\succeq_1, \succeq_2, \dots, \succeq_n)$ can then be derived by setting $B \succeq_i C$ if and only if $\sum_{j \in B} w_i(j) \geq \sum_{j \in C} w_i(j)$ for each $i \in A$ and for any two coalitions $B, C \in \mathscr{A}_i$.

The friend- and enemy-oriented preference extensions are due to Dimitrov et al. (2006). Every player $i \in A$ partitions the other players into a set $F_i \subseteq A \setminus \{i\}$ of friends and a set $E_i = A \setminus (F_i \cup \{i\})$ of enemies. Let $B, C \in \mathscr{A}_i$. In the *friend-oriented preference extension, i weakly prefers B to C* ($B \succeq_i C$) if and only if B either contains more of i's friends than C or, if B and C have the same number of i's friends, B has at most as many enemies of i's as C, i.e., $\|B \cap F_i\| > \|C \cap F_i\| \vee (\|B \cap F_i\| = \|C \cap F_i\| \wedge \|B \cap E_i\| \leq \|C \cap E_i\|)$. Analogously, in the *enemy-oriented preference extension, i weakly prefers B to C* ($B \succeq_i C$) if and only if B either contains fewer of i's enemies than C or, if B and C have the same number of i's enemies, B has at least as many of i's friends as C, i.e., $\|B \cap E_i\| < \|C \cap E_i\| \vee (\|B \cap E_i\| = \|C \cap E_i\| \wedge \|B \cap F_i\| \geq \|C \cap F_i\|)$. Both friend- and enemy-oriented hedonic games are additively separable, by letting each player assign the value $\|A\|$ to her friends and the value -1 to her enemies in the friend-oriented case, and by letting each player assign the value 1 to her friends and the value $-\|A\|$ to her enemies in the enemy-oriented case.

In the *singleton encoding*, due to Cechlárová and Romero-Medina (2001) (see also Cechlárová and Hajduková, 2003, 2004), each player $i \in A$ reports a complete ranking $\trianglerighteq_i$ over all players. For each coalition $B \in \mathscr{A}_i$, $\mathscr{B}_i(B)$ denotes any *best* player in B according to i's ranking (i.e., a player $j \in B$ such that $j \trianglerighteq_i k$ for each $k \in B$), and $\mathscr{W}_i(B)$ denotes any *worst* player in B according to i's ranking (i.e., $\mathscr{W}_i(B) = i$ if $B = \{i\}$, and otherwise a player $j \in B \setminus \{i\}$ such that $k \trianglerighteq_i j$ for each $k \in B$). For any $B, C \in \mathscr{A}_i$, B is *$\mathscr{B}$-preferred by i over C* if $\mathscr{B}_i(B) \triangleright_i \mathscr{B}_i(C)$ or ($\mathscr{B}_i(B) \sim_i \mathscr{B}_i(C)$ and $\|B\| < \|C\|$), and B *is $\mathscr{W}$-preferred by i over C* if $\mathscr{W}_i(B) \triangleright_i \mathscr{W}_i(C)$.

We will focus on well-known notions of stability for coalition structures in hedonic games (Bogomolnaia and Jackson, 2002; Aziz et al., 2013b) (see Aziz and Savani, 2016; Elkind and Rothe, 2015 for a survey) that are based either on avoiding that a single player has an incentive to deviate to another (possibly empty) existing coalition (e.g., Nash stability), or on avoiding groups of players having an incentive to deviate from the current coalition structure (e.g., core stability). For other restrictions of games and other properties, we refer, e.g., to the work of Banerjee et al. (2001) and Aziz et al. (2013a). We say a coalition structure Γ is

1. *perfect* if every player i weakly prefers $\Gamma(i)$ to every other coalition i is contained in;
2. *individually rational* if every player $i \in A$ weakly prefers $\Gamma(i)$ to $\{i\}$;
3. *Nash stable* if no player wants to move to another (possibly empty) coalition in Γ (i.e., $\Gamma(i) \succeq_i C \cup \{i\}$ for every player $i \in A$ and for every coalition $C \in \Gamma \cup \{\emptyset\}$);
4. *individually stable* if no player prefers another (possibly empty) coalition in Γ or can move to another such coalition without some player in the new coalition objecting to it (i.e., for every player $i \in A$ and for every coalition $C \in \Gamma \cup \{\emptyset\}$, $\Gamma(i) \succeq_i C \cup \{i\}$ or there is some player $j \in C$ with $C \succ_j C \cup \{i\}$);
5. *contractually individually stable* if no player prefers another (possibly empty) coalition in Γ or can move to another such coalition without some player in the new or in the old coalition objecting to it (i.e., for every player $i \in A$ and for every coalition $C \in \Gamma \cup \{\emptyset\}$, we have $\Gamma(i) \succeq_i C \cup \{i\}$ or $C \succ_j C \cup \{i\}$ for some player $j \in C$ or $\Gamma(i) \succ_k \Gamma(i) \setminus \{i\}$ for some player $k \in \Gamma(i) \setminus \{i\}$);
6. *core stable* if no coalition blocks Γ (i.e., for every coalition $C \subseteq A$, $\Gamma(i) \succeq_i C$ for some player $i \in C$);
7. *strictly core stable* if no coalition weakly blocks Γ (i.e., for every coalition $C \subseteq A$, $\Gamma(i) \succ_i C$ for some player $i \in C$ or $\Gamma(i) \sim_i C$ for each player $i \in C$);

2.2. Complexity theory

For a stability concept γ as defined above, we study the question of how hard it is to decide whether a given solution for a given game is γ-stable (the *verification problem*) and how hard it is to decide whether there exists a γ-stable outcome in a given game (the *existence problem*). We denote the verification problem for γ by γ-Verification and define it formally as follows: Given a hedonic game H and a coalition structure Γ, is Γ stable in H in the sense of γ? The existence problem for γ, γ-Existence, is defined as: Given a hedonic game H, does there exist a coalition structure that is stable in H in the sense of γ?

We assume the reader to be familiar with the basics of complexity theory, such as the complexity classes P, NP, and coNP and the notions of (polynomial-time many-one) reducibility, hardness, and completeness. It is easy to see that membership of γ-Verification in P implies membership of γ-Existence in NP: Guess a coalition structure and verify whether it satisfies γ. However, other direct connections between these two problems are not known to hold (see the survey by Woeginger, 2013a for further discussion).

In Section 6, we will study these problems in terms of their complexity for several stability concepts in FEN-hedonic games, using

4 We sometimes may give the players names other than numbers.

reductions from the following well-known NP-complete problems (see, e.g., Garey and Johnson, 1979):

In Exact-Cover-by-Three-Sets (X3C), we are given a set $B = \{b_1, b_2, \ldots, b_{3m}\}$, $m > 1$, and a collection $\mathscr{S} = \{S_1, S_2, \ldots, S_n\}$ of subsets $S_i \subseteq B$ such that $\|S_i\| = 3$ for each i, $1 \le i \le n$, and we ask whether there is a subcollection $\mathscr{S}' \subseteq \mathscr{S}$ that exactly covers B (i.e., each element of B occurs in exactly one set in $\mathscr{S}'$). Note that X3C is NP-complete even if each element in a set from $\mathscr{S}'$ occurs in at most three sets in $\mathscr{S}$ (see Garey and Johnson, 1979).

In Clique, we are given an undirected graph $G = (V, E)$ and a positive integer k, and we ask whether G has a clique (i.e., a subset $V' \subseteq V$ such that there is an edge between any two vertices in V') of size at least k.

We will also study problems in the second level of the polynomial hierarchy, $\Sigma_2^p = \mathrm{NP}^{\mathrm{NP}}$. Meyer and Stockmeyer (1972) (see also Stockmeyer, 1976) characterized this class by two alternating, polynomially length-bounded quantifiers: $B \in \Sigma_2^p$ if and only if there are a set $C \in \mathrm{P}$ and a polynomial p such that for each input x,

$$x \in B \iff (\exists y : |y| \le p(|x|))(\forall z : |z| \le p(|x|))\,[(x, y, z) \in C],$$

where the length of a string s is denoted by $|s|$.

Stockmeyer (1976) showed Σ_2^p-completeness of the following problem, which we will also use in Section 6: In 2-Quantified-3-DNF-SAT, we are given two sets, $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, y_2, \ldots, y_n\}$, of boolean variables and a boolean formula $\varphi(X, Y)$ over $X \cup Y$ in disjunctive normal form, with exactly three literals per disjunct. We ask whether there exists a truth assignment τ_X for the variables in X such that for every truth assignment τ_Y for the variables in Y the formula φ evaluates to true under τ_X and τ_Y.

3. Groundwork: FEN-hedonic games and some sufficient conditions for hardness of stability

In Section 3.1, we will present the model of FEN-hedonic games that we introduced in Lang et al. (2015) and illustrate the relevant notions by examples, and in Section 3.2 we will state some useful results from the work of Peters and Elkind (2015) to be applied later on.

3.1. FEN-hedonic games and the polarized responsive extension principle

Let us first give a rough, high-level outline of how we proceed to define FEN-hedonic-games. First, similarly to the singleton encoding, to hedonic games with $\mathscr{W}$-preferences, and to the friend- and enemy-oriented encoding (formally defined in Section 2), we assume that each player $i \in A$ expresses her preferences over the other players, and we denote these preferences by $\trianglerighteq_i^{+0-}$. We then lift these preferences over players to preferences over coalitions, denoted by $\succeq_i^{+0-}$ for each $i \in A$, by generalizing the *responsive extension principle* as formally defined below. Note that $\succeq_i^{+0-}$ can be incomplete because there might be pairs of coalitions for which $\succeq_i^{+0-}$ does not tell which of them is preferred by player i. Finally, we will define the set $\mathrm{Ext}\left(\succeq_i^{+0-}\right)$ of possible complete extensions of $\succeq_i^{+0-}$, and the collection of $\succeq_i \in \mathrm{Ext}\left(\succeq_i^{+0-}\right)$ for each $i \in A$ will then define the class of FEN-hedonic games. More concretely, we assume that each player considers each other player as either a friend, or a neutral player, or an enemy, where the friends and enemies are to be ranked (with indifferences allowed) and the player is indifferent about the neutral players. This is formalized as follows.

Definition 1. Every player $i \in A = \{1, 2, \ldots, n\}$ provides a *weak ranking with double threshold*, denoted by $\trianglerighteq_i^{+0-}$, by partitioning $A \setminus \{i\}$ into three sets: the set A_i^+ of i's friends, along with a weak order $\trianglerighteq_i^+$ over A_i^+, the set A_i^0 of neutral players for i (i.e., i is indifferent about them: $j \sim_i k$ for all $j, k \in A_i^0$), and the set A_i^- of i's enemies, along with a weak order $\trianglerighteq_i^-$ over A_i^-. We write $\trianglerighteq_i^{+0-}$ as $(\trianglerighteq_i^+ \mid A_i^0 \mid \trianglerighteq_i^-)$.

In addition, we assume that every player i (strictly) prefers her friends to her neutral players and her neutral players to her enemies. Define the *weak order* $\trianglerighteq_i$ *induced by* $\trianglerighteq_i^{+0-}$ as follows: $\trianglerighteq_i$ coincides with $\trianglerighteq_i^+$ on A_i^+; $f \vartriangleright_i j$ for each $f \in A_i^+$ and $j \in A_i^0$; $j_1 \sim_i j_2 \sim_i \cdots \sim_i j_k$ for $A_i^0 = \{j_1, j_2, \ldots, j_k\}$; $j \vartriangleright_i e$ for each $j \in A_i^0$ and $e \in A_i^-$; and $\trianglerighteq_i$ coincides with $\trianglerighteq_i^-$ on A_i^-.

A *FEN-hedonic game* is a pair $H = (A, (\trianglerighteq_1^{+0-}, \ldots, \trianglerighteq_n^{+0-}))$, where $A = \{1, 2, \ldots, n\}$ is a set of players, and $\trianglerighteq_i^{+0-}$ gives the weak ranking with double threshold of player $i \in A$.

Notation: If a player i is indifferent about all players in a set $X = \{a_1, a_2, \ldots, a_x\} \subseteq A \setminus \{i\}$, we write $X_{\sim_i}$ as a shorthand for $a_1 \sim_i a_2 \sim_i \cdots \sim_i a_x$. When i is clear from context, we sometimes omit the subscript i and simply write $X_\sim$ instead of $X_{\sim_i}$. Whenever a player i has no friends or no enemies, we will slightly abuse notation and denote the empty preference $\trianglerighteq_i^+$ or $\trianglerighteq_i^-$ by $\emptyset$.

Example 2. Let $A = \{1, 2, \ldots, 11\}$. Then the weak ranking with double threshold

$$\trianglerighteq_1^{+0-} = (2 \sim_1 3 \vartriangleright_1 5 \mid \{10, 11\} \mid 6 \vartriangleright_1 7 \vartriangleright_1 8 \vartriangleright_1 4 \vartriangleright_1 9)$$

means that player 1 likes 2, 3, and 5 (and is indifferent between 2 and 3, but prefers both to 5); 1 does not care about 10 and 11 (and is indifferent between them); and 1 does not like 4, 6, 7, 8, and 9 (but still prefers 6 to 7, 7 to 8, 8 to 4, and 4 to 9). The weak order $\trianglerighteq_1$ induced by $\trianglerighteq_1^{+0-}$ is $2 \sim_1 3 \vartriangleright_1 5 \vartriangleright_1 10 \sim_1 11 \vartriangleright_1 6 \vartriangleright_1 7 \vartriangleright_1 8 \vartriangleright_1 4 \vartriangleright_1 9$.

Using a bipolar variant of the *responsive extension principle*, which is sometimes referred to as the *Bossong–Schweigert extension principle* (Bossong and Schweigert, 2006; Delort et al., 2011), we now define a player i's preferences over coalitions she is contained in. This *polarized responsive extension* induced by i's weak ranking with double threshold $\trianglerighteq_i^{+0-}$ is a partial order over coalitions containing i.

Definition 3. Let $\trianglerighteq_i^{+0-}$ be player i's weak ranking with double threshold. Define the *extended order* $\succeq_i^{+0-}$ as follows. For any two coalitions $X, Y \in \mathscr{A}_i$, i weakly prefers X to Y ($X \succeq_i^{+0-} Y$) if and only if

1. there is an injective function $\sigma : Y \cap A_i^+ \to X \cap A_i^+$ such that $\sigma(y) \trianglerighteq_i y$ for each $y \in Y \cap A_i^+$, and
2. there is an injective function $\theta : X \cap A_i^- \to Y \cap A_i^-$ such that $x \trianglerighteq_i \theta(x)$ for each $x \in X \cap A_i^-$.

Further, we write $X \succ_i^{+0-} Y$ if and only if $X \succeq_i^{+0-} Y$ and not $Y \succeq_i^{+0-} X$, and we write $X \sim^{+0-} Y$ if and only if $X \succeq^{+0-} Y$ and $Y \succeq^{+0-} X$.

Intuitively speaking, adding friends to a coalition makes it strictly more valuable, whereas adding enemies makes it strictly less valuable. Replacing a friend in a coalition by another friend that i prefers increases its value, and similarly so when replacing an enemy by another enemy that i prefers. However, when both a friend and an enemy are added to a coalition or when both are removed, the two coalitions are incomparable with respect to the responsive extension principle.

To construct the polarized responsive extension for a player i, we start with the coalition consisting of i and all her friends – this is i's most preferred coalition. We then construct all directly comparable coalitions by adding enemies, removing friends, or exchanging enemies or friends. For each newly obtained coalition, we repeat this procedure until we reach i's least preferred coalition consisting of i and all of her enemies. We may safely disregard the neutral players (elements of A_i^0) in this process because adding or removing them to or from a coalition does not change i's value of the coalition. To illustrate, we give some examples.

Example 4. Let $A = \{1, 2, \dots, 6\}$. Player 1's weak ranking with double threshold is given by $\trianglerighteq_1^{+0-} = (2 \sim_1 3 \triangleright_1 5 \mid \emptyset \mid 6 \triangleright_1 4)$. The polarized responsive extension of 1's preference is shown by the graph in Fig. 1, where an arc from coalition X to coalition Y means that $X \succ_1^{+0-} Y$. Therefore, each path from X' to Y' implies $X' \succ_1^{+0-} Y'$ (e.g., $\{1, 2, 3\} \succ_1^{+0-} \{1, 2, 4, 5\}$), yet coalitions that are not connected by a path (e.g., $\{1, 2, 3\}$ and $\{1, 2, 3, 5, 6\}$) are incomparable.

Inspired by the work of Aziz et al. (2015) and of Bouveret et al. (2010) that establishes characterizations for the original responsive order in the context of fair division, we now provide some characterization of $C \succeq_i^{+0-} D$ for any two coalitions C and D, from player i's perspective.

Proposition 5. *Let $\trianglerighteq_i^{+0-}$ be a weak ranking with double threshold for player i, and let C and D be any two coalitions containing i. Define $w_i : A \to \mathbb{R}$ to be compatible with $\trianglerighteq_i^{+0-}$ if (a) for each $j \in A_i^+$, we have $w_i(j) > 0$; (b) for each $j \in A_i^-$, we have $w_i(j) < 0$; (c) for each $j \in A_i^0$, we have $w_i(j) = 0$; and (d) for all $j, k \in A_i^+ \cup A_i^-$, we have $j \triangleright_i k$ if and only if $w_i(j) > w_i(k)$. Then $C \succeq_i^{+0-} D$ if and only if $\sum_{j \in C} w_i(j) \geq \sum_{j \in D} w_i(j)$ for each w_i compatible with $\trianglerighteq_i^{+0-}$.*

Proof. Assume that $C \succeq_i^{+0-} D$. For the set of friends A_i^+, we have $\sigma : D \cap A_i^+ \to C \cap A_i^+$ such that for each $y \in D \cap A_i^+$, we have $\sigma(y) \trianglerighteq_i y$. Hence, for each compatible w_i, $w_i(\sigma(y)) \geq w_i(y)$. Thus, since σ is injective, $\sum_{j \in C \cap A_i^+} w_i(j) \geq \sum_{j \in \sigma(D \cap A_i^+) \subseteq C \cap A_i^+} w_i(j) = \sum_{j \in D \cap A_i^+} w_i(\sigma(j)) \geq \sum_{j \in D \cap A_i^+} w_i(j)$. Similarly, for A_i^- and injective mapping $\theta : C \cap A_i^- \to D \cap A_i^-$, it holds that $0 \geq \sum_{j \in C \cap A_i^-} w_i(j) \geq \sum_{j \in C \cap A_i^-} w_i(\theta(j)) = \sum_{k \in \theta(C \cap A_i^-) \subseteq D \cap A_i^-} w_i(k) \geq \sum_{j \in D \cap A_i^-} w_i(j)$. For each player $j \in A_i^0$, we have $w_i(j) = 0$; therefore, in total, $\sum_{j \in C} w_i(j) \geq \sum_{j \in D} w_i(j)$.

Now assume that $\sum_{j \in C} w_i(j) \geq \sum_{j \in D} w_i(j)$ holds for each compatible w_i. Thus

$$\sum_{j \in C \cap A_i^+} w_i(j) - \sum_{j \in D \cap A_i^-} w_i(j) \geq \sum_{j \in D \cap A_i^+} w_i(j) - \sum_{j \in C \cap A_i^-} w_i(j). \quad (1)$$

Assume there were no injective function mapping from each summand from the right-hand side to one at least as large on the left hand side. Then there exists an assignment to the values of w_i compatible with $\trianglerighteq_i^{+0-}$ that does not satisfy the above inequality (1), a contradiction. □

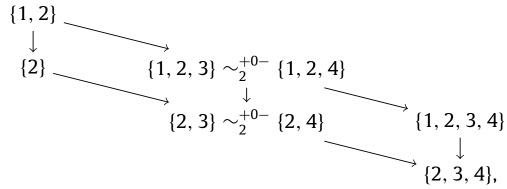
Since the preference relations $\succeq_i^{+0-}$ can be incomplete, we consider their extensions to complete relations, each preserving the already defined comparisons.

Definition 6. A complete preference relation $\succeq_i$ over $\mathscr{A}_i$, *extends* $\succeq_i^{+0-}$ if it contains it: For all $C, D \in \mathscr{A}_i$ $C \succ_i^{+0-} D$ implies $C \succ_i D$, and $C \sim_i^{+0-} D$ implies $C \sim_i D$. Let $\mathrm{Ext}\left(\succeq_i^{+0-}\right)$ be the set of all complete preference relations extending $\succeq_i^{+0-}$.

We will see that weak rankings with double threshold can have various complete extensions.

Example 7. Let $(A, (\trianglerighteq_1^{+0-}, \dots, \trianglerighteq_4^{+0-}))$ be a FEN-hedonic game with players $A = \{1, 2, 3, 4\}$ and the following weak rankings with double threshold: $\trianglerighteq_1^{+0-} = (\emptyset \mid \emptyset \mid 2 \triangleright_1 3 \sim_1 4)$, $\trianglerighteq_2^{+0-} = (1 \mid \emptyset \mid 3 \sim_2 4)$, $\trianglerighteq_3^{+0-} = (2 \mid \{1, 4\} \mid \emptyset)$, and $\trianglerighteq_4^{+0-} = (1 \triangleright_4 2 \mid \{3\} \mid \emptyset)$. This gives the following polarized responsive order for

- player 1: $\{1\} \succ_1^{+0-} \{1, 2\} \succ_1^{+0-} \{1, 3\} \sim_1^{+0-} \{1, 4\} \succ_1^{+0-} \{1, 2, 3\} \sim_1^{+0-} \{1, 2, 4\} \succ_1^{+0-} \{1, 3, 4\} \succ_1^{+0-} \{1, 2, 3, 4\}$,
- player 2 (using the same notation as in Example 4):



- player 3: $\{2, 3\} \sim_3^{+0-} \{1, 2, 3\} \sim_3^{+0-} \{2, 3, 4\} \sim_3^{+0-} \{1, 2, 3, 4\} \succ_3^{+0-} \{3\} \sim_3^{+0-} \{1, 3\} \sim_3^{+0-} \{3, 4\} \sim_3^{+0-} \{1, 3, 4\}$, and
- player 4: $\{1, 2, 4\} \sim_4^{+0-} \{1, 2, 3, 4\} \succ_4^{+0-} \{1, 4\} \sim_4^{+0-} \{1, 3, 4\} \succ_4^{+0-} \{2, 4\} \sim_4^{+0-} \{2, 3, 4\} \succ_4^{+0-} \{4\} \sim_4^{+0-} \{3, 4\}$.

Note that three preferences (namely, $\succeq_1^{+0-}$, $\succeq_3^{+0-}$, and $\succeq_4^{+0-}$) are already complete. There are eleven complete preferences extending $\succeq_2^{+0-}$, obtained by specifying the relation between $\{2\}$ and $\{1, 2, 3\} \sim_2^{+0-} \{1, 2, 4\}$, $\{2, 3\} \sim_2^{+0-} \{2, 4\}$ and $\{1, 2, 3, 4\}$, and $\{2\}$ and $\{1, 2, 3, 4\}$. Setting $\{2\} \succ_2 \{1, 2, 3\} \sim_2 \{1, 2, 4\}$ or $\{2\} \sim_2 \{1, 2, 3\} \sim_2 \{1, 2, 4\}$ also implies $\{2\} \succ_2 \{1, 2, 3, 4\}$; then, we can still freely choose between $\{2, 3\} \sim_2 \{2, 4\} \succ_2 \{1, 2, 3, 4\}$, $\{2, 3\} \sim_2 \{2, 4\} \sim_2 \{1, 2, 3, 4\}$, or $\{1, 2, 3, 4\} \succ_2 \{2, 3\} \sim_2 \{2, 4\}$, which gives six possible complete preferences extending $\succeq_2^{+0-}$. On the other hand, if $\{1, 2, 3\} \sim_2 \{1, 2, 4\} \succ_2 \{2\}$, we are not restricted regarding our decision on the relation between $\{2, 3\} \sim_2 \{2, 4\}$ and $\{1, 2, 3, 4\}$. However, if $\{1, 2, 3, 4\} \succ_2 \{2, 3\} \sim_2 \{2, 4\}$, the relation between $\{2\}$ and $\{1, 2, 3, 4\}$ is not yet determined and leaves us with three additional choices. Therefore, we have three instead of one possible complete preferences extending $\succeq_2^{+0-}$ in the latter case plus two for the first two other possibilities regarding $\{2, 3\} \sim_2 \{2, 4\}$ and $\{1, 2, 3, 4\}$, resulting in five additional complete preferences extending $\succeq_2^{+0-}$. Overall, by adding up all those possibilities, we have eleven valid complete preferences extending $\succeq_2^{+0-}$.

3.2. Some useful results on hardness of stability obtained from properties of preference extensions

Peters and Elkind (2015) established some useful links between the properties of players' preferences in hedonic games and NP-hardness of a number of problems related to whether there exist stable coalition structures in these games. They assume that each player $i \in A$ reports a ranking $\trianglerighteq_i$ over A that is used to partition $A \setminus \{i\}$ into a set of enemies, $A_i^- = \{j \neq i \mid i \triangleright_i j\}$, and a set of friends; note that their notion of "i's friends" also includes what we call "i's neutral players": $A_i^+ \cup A_i^0 = \{j \neq i \mid j \trianglerighteq_i i\}$. They also assume that each ranking $\trianglerighteq_i$ of players can be extended to a preference $\succeq_i$ over coalitions. Moreover, they assume that each player is allowed to have arbitrary orderings of size-2 coalitions; we refer to this property as *arbitrary ordering of players*.[5] Finally, they assume that the preference profile $\succeq = (\succeq_1, \dots, \succeq_n)$ of the hedonic game $(A, \succeq)$ can be obtained from $\trianglerighteq = (\trianglerighteq_1, \dots, \trianglerighteq_n)$ in deterministic polynomial time; we will say that this hedonic game is *induced by* $\trianglerighteq$.

Peters and Elkind (2015) define the following properties of preference extensions, which can be used to obtain hardness results for certain stability problems.

[5] It is easy to see that the class of all possible preference extensions for FEN-hedonic games allows to order the players arbitrarily.

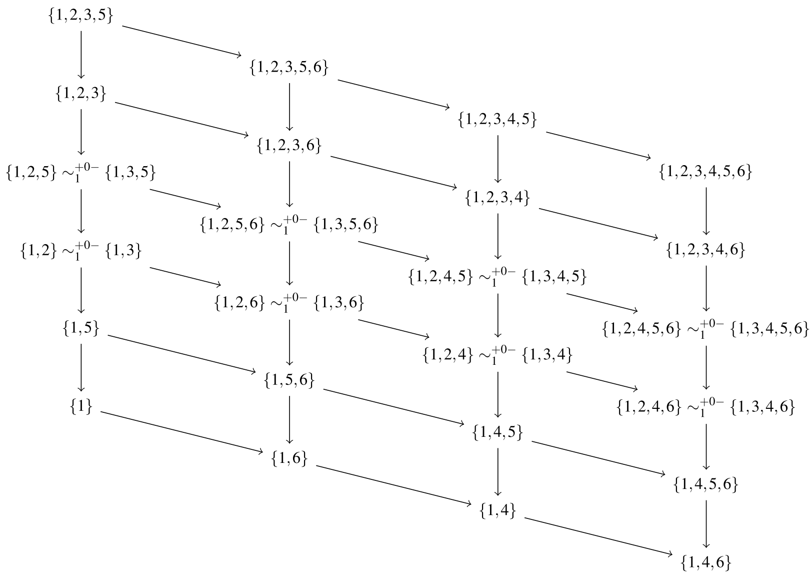


Fig. 1. The polarized responsive extension of $\trianglerighteq_1^{+0-} = (2 \sim_1 3 \rhd_1 5 \mid \emptyset \mid 6 \rhd_1 4)$.

Definition 8 (*Peters and Elkind, 2015*)**.** Let $a, b \in \mathbb{N}$. A hedonic game $(A, \succeq)$ (with n players and preferences induced by a profile $\trianglerighteq = (\trianglerighteq_1, \ldots, \trianglerighteq_n)$ of rankings over players) is said to be

(a) *consistent on pairs* if for all $i \in A$ and for all $j, k \in A_i^+ \cup A_i^0 \cup \{i\}$, it holds that $\{i, j\} \succeq_i \{i, k\}$ if and only if $j \trianglerighteq_i k$;
(b) *a-b-toxic* if for all $i \in A$ and for each $S \subseteq A$, it holds that $\{i\} \succeq_i S$ if $\|S \cap (A_i^+ \cup A_i^0)\| = a$ but $\|S \cap A_i^-\| \geq b$;
(c) *strictly a-b-toxic* if for all $i \in A$ and for each $S \subseteq A$, it holds that $\{i\} \succ_i S$ if $\|S \cap (A_i^+ \cup A_i^0)\| = a$ but $\|S \cap A_i^-\| \geq b$; and
(d) *weakly a-b-toxic* if for all $i \in A$ and for each $S \subseteq A$, it holds that $\{i, j\} \succ_i S$ for all $j \in A_i^+ \cup A_i^0$ if $\|S \cap (A_i^+ \cup A_i^0)\| = a$ but $\|S \cap A_i^-\| \geq b$.

A class of hedonic games fulfills any one of these properties if for each set of n players and every profile $\trianglerighteq = (\trianglerighteq_1, \ldots, \trianglerighteq_n)$ of rankings over players, there is a hedonic game $(A, \succeq)$ in this class that is induced by $\trianglerighteq$ and satisfies this property.

The following lemma, proven in the appendix, shows how these properties are related to each other.

Lemma 9.

(a) *Strict a-b-toxicity implies a-b-toxicity.*
(b) *Strict a-b-toxicity together with consistency on pairs implies weak a-b-toxicity.*
(c) *a-b-toxicity implies a-c-toxicity for all $c > b$. The same holds for strict and weak toxicity.*

We will also make use of the following results due to Peters and Elkind (2015). Some notation is needed first: A profile $\trianglerighteq = (\trianglerighteq_1, \ldots, \trianglerighteq_n)$ of preference orderings on A is said to be *strict* if each $\trianglerighteq_i$ is antisymmetric (i.e., if $j \trianglerighteq_i k$ and $k \trianglerighteq_i j$, then $j = k$), and it is said to be *mutual* if $j \in A_i^+ \cup A_i^0$ is equivalent to $i \in A_j^+ \cup A_j^0$.

Theorem 10 (*Peters and Elkind, 2015*)**.** *For each class of hedonic games that allows arbitrary ordering of players, it holds that*

1. Core-Stability-Existence *is* NP-*hard if for every n and every profile $\trianglerighteq = (\trianglerighteq_1, \ldots, \trianglerighteq_n)$ of mutual preferences, where each player has at most three friends, the class contains an induced hedonic game that is consistent on pairs,* 0-1-*toxic, weakly* 1-1-*toxic, and weakly* 2-2-*toxic.*
2. Nash-Stability-Existence *and* Individual-Stability-Existence *are* NP-*complete if for every n and every profile $\trianglerighteq = (\trianglerighteq_1, \ldots, \trianglerighteq_n)$ of strict, mutual preferences, where each player has at most three friends, the class contains an induced hedonic game that is consistent on pairs, strictly* 0-1-*toxic, strictly* 1-1-*toxic, and strictly* 2-5-*toxic.*

Lemma 11, again to be proven in the appendix, will be applied later on.

Lemma 11. *Every hedonic game with preferences derived from a FEN-hedonic game is consistent on pairs and strictly* 0-1-*toxic.*

4. The model of Borda-induced FEN-hedonic games

We have seen that in FEN-hedonic games the preference relation $\succeq_i^{+0-}$ can be incomplete in the sense that there might be pairs of coalitions that are incomparable. We now propose an approach of handling these incomparabilities by introducing a class of preference extensions of $\succeq_i^{+0-}$ in the sense of Definition 6. That is, the relations we want to define have to be complete (all coalitions have to be comparable) and, furthermore, those relations already defined by $\succeq_i^{+0-}$ have to be preserved. To achieve the former we introduce so-called *comparability functions* that are inspired by voting theory: Based on player i's preferences over the other players given in $\trianglerighteq_i^{+0-}$, we determine values that i assigns to the other players and aggregate these values to compute the values of coalitions in $\mathscr{A}_i$.

Proposition 5 gives a characterization of how such comparability functions can be defined such that those relations that are already determined by $\succeq_i^{+0-}$ are preserved. Based on this characterization, we define our comparability function as a function $w_i : A \to \mathbb{Z}$ with $w_i(i) = 0$. Clearly, $w_i(j) = 0$ has to hold for all $j \in A_i^0$. Using terminology from voting theory, we define so-called *scoring vectors*

$$\mathbf{f}_i \in \mathbb{Z}_{>0}^{\|A_i^+\|}, \quad \mathbf{e}_i \in \mathbb{Z}_{<0}^{\|A_i^-\|}$$

Table 1
Values that are derived from different choices for $\mathbf{f}_i$ and $\mathbf{e}_i$ when there are only indifferences within $\trianglerighteq_i^+$ and $\trianglerighteq_i^-$.

	sfo	**fo**	**sfp**	**fp**	**seo**	**eo**	**sep**	**ep**
Value	n	$\|A_i^+\|$	1	$n-\|A_i^+\|+1$	-1	$-(n-\|A_i^-\|+1)$	$-n$	$-\|A_i^-\|$

assigning positive integer values to i's friends and negative integer values to i's enemies, and we will focus here on using Borda-like scoring vectors. Inspired by the work of Baumeister et al. (2012) regarding modified Borda voting, we introduce several variants capturing the notions of "optimistic" and "pessimistic" assessments of friend or enemy relations.

Let $\trianglerighteq_i^{+0-}$ be the weak ranking with double threshold of player $i \in A$ with the following ordering of i's friends and enemies:

- $\trianglerighteq_i^+ = A_{i,1}^+ \rhd_i^+ A_{i,2}^+ \rhd_i^+ \cdots \rhd_i^+ A_{i,\ell}^+$, where each $A_{i,j}^+$, $1 \le j \le \ell$, contains friends player i is indifferent about, and
- $\trianglerighteq_i^- = A_{i,1}^- \rhd_i^- A_{i,2}^- \rhd_i^- \cdots \rhd_i^- A_{i,m}^-$, where each $A_{i,j}^-$, $1 \le j \le m$, contains enemies i is indifferent about.

Using this notation, we define the following variants of our Borda-like scoring vectors.

1. $\mathbf{f}_i$ can be one of the following four vectors:
   (a) *Strongly friend-optimistic* (**sfo**): Each player in $A_{i,1}^+$ gets n points, each player in $A_{i,2}^+$ gets $n-1$ points, ..., and each player in $A_{i,\ell}^+$ gets $n-\ell+1$ points.
   (b) *Friend-optimistic* (**fo**): Each player in $A_{i,1}^+$ gets $\|A_i^+\|$ points, each player in $A_{i,2}^+$ gets $\|A_i^+\|-1$ points, ..., and each player in $A_{i,\ell}^+$ gets $\|A_i^+\|-\ell+1$ points.
   (c) *Strongly friend-pessimistic* (**sfp**): Each player in $A_{i,\ell}^+$ gets 1 point, each player in $A_{i,\ell-1}^+$ gets 2 points, ..., and each player in $A_{i,1}^+$ gets ℓ points.
   (d) *Friend-pessimistic* (**fp**): Each player in $A_{i,\ell}^+$ gets $n-\|A_i^+\|+1$ points, each player in $A_{i,\ell-1}^+$ gets $n-\|A_i^+\|+2$ points, ..., and each player in $A_{i,1}^+$ gets $n-\|A_i^+\|+\ell$ points.
2. $\mathbf{e}_i$ can be one of the following four vectors:
   (a) *Strongly enemy-optimistic* (**seo**): Each player in $A_{i,1}^-$ gets -1 point, each player in $A_{i,2}^-$ gets -2 points, ..., and each player in $A_{i,m}^-$ gets $-m$ points.
   (b) *Enemy-optimistic* (**eo**): Each player in $A_{i,1}^-$ gets $-(n-\|A_i^-\|+1)$ points, each player in $A_{i,2}^-$ gets $-(n-\|A_i^-\|+2)$ points, ..., and each player in $A_{i,m}^-$ gets $-(n-\|A_i^-\|+m)$ points.
   (c) *Strongly Enemy-pessimistic* (**sep**): Each player in $A_{i,m}^-$ gets $-n$ points, each player in $A_{i,m-1}^-$ gets $-n+1$ points, ..., and each player in $A_{i,1}^-$ gets $-(n-m+1)$ points.
   (d) *Enemy-pessimistic* (**ep**): Each player in $A_{i,m}^-$ gets $-\|A_i^-\|$ points, each player in $A_{i,m-1}^-$ gets $-\|A_i^-\|+1$ points, ..., and each player in $A_{i,1}^-$ gets $-(\|A_i^-\|-m+1)$ points.

Each pair of scoring vectors $(\mathbf{f}_i, \mathbf{e}_i) \in \{\mathbf{sfo}, \mathbf{fo}, \mathbf{sfp}, \mathbf{fp}\} \times \{\mathbf{seo}, \mathbf{eo}, \mathbf{sep}, \mathbf{ep}\}$ defines a particular way of how the scores a player i assigns to the other players are derived from $\trianglerighteq_i^{+0-}$. The intuition behind these definitions and why it is reasonable to distinguish each of the four cases can be best seen assuming that player i is indifferent between all of her friends and all of her enemies. With the above notation, it holds that $\ell = 1$ and $m = 1$ and the values shown in Table 1 are assigned to i's friends and enemies depending on the choice of $\mathbf{f}_i$ and $\mathbf{e}_i$, respectively.

We see that in the friend-optimistic case a larger friend set implies higher values for the friends contained in it, while the opposite is the case in the friend-pessimistic case. The same holds for the comparison between the enemy-pessimistic and enemy-optimistic case with the difference that in the former case a larger enemy set reduces the enemies' scores and in the latter case a larger enemy set implies higher values.

Table 2
Values player 2 assigns to the players 1, 3, and 4 and the coalitions {2, 3} and {1, 2, 3, 4} for different choices of $\mathbf{f}_i$ and $\mathbf{e}_i$.

$\mathbf{f}_i$	$\mathbf{e}_i$	$w_2(j)$			$f^2_{\text{Borda}}(C)$	
		$j=1$	$j=3$	$j=4$	$C=\{2,3\}$	$C=\{1,2,3,4\}$
fo	**seo**	1	-1	-1	-1	-1
fo	**eo**	1	-3	-3	-3	-5
sfo	**seo**	4	-1	-1	-1	2

On the other hand, when there are no indifferences within $\trianglerighteq_i^+$, for $\mathbf{f}_i \in \{\mathbf{sfo}, \mathbf{fp}\}$ and $\mathbf{f}_i \in \{\mathbf{sfp}, \mathbf{fo}\}$, the two scoring vectors from one set yield the same scores for player i's friends. The same holds for $\mathbf{e}_i \in \{\mathbf{seo}, \mathbf{ep}\}$ and $\mathbf{e}_i \in \{\mathbf{eo}, \mathbf{sep}\}$, whenever there are no indifferences in $\trianglerighteq_i^-$. Concluding, it can be seen that each of the 16 different variations of scoring vectors is a legitimate option and the choice only depends on the weight each friend- or enemy-relation is supposed to have. A central organizer would have to weigh the friend- and enemy-relations against each other in the given scenario so as to make the appropriate choice.

Analogously to the definition of positional scoring rules and having Proposition 5 in mind, we define the value of a coalition from player i's view as the sum of the values she assigns to the players in the coalition.

Definition 12. Let $i \in A$ be a player. For a fixed choice of scoring vectors $\mathbf{f}_i$ and $\mathbf{e}_i$ defining the scoring function w_i, we define the *Borda-like comparability function (CF)*

$$f^i_{\text{Borda}} : \mathscr{A}_i \to \mathbb{Z}, \quad C \mapsto \sum_{j \in C \setminus \{i\}} w_i(j),$$

to be a function mapping every coalition C containing i to the sum of the scores the players in $C \setminus \{i\}$ obtained from w_i.

With this notion of comparability functions, we can derive a complete preference relation from given weak rankings with double threshold; we call this relation *Borda-induced* and define it in Definition 13 formally.

Definition 13. For a FEN-hedonic game $(A, (\trianglerighteq_1^{+0-}, \dots, \trianglerighteq_n^{+0-}))$ with n players and a fixed choice of $\mathbf{f}_i$ and $\mathbf{e}_i$, let f^i_{Borda} be the Borda-like CF.

For two coalitions $C, D \in \mathscr{A}_i$ it holds that

- $C \succeq_i^B D$ if and only if $f^i_{\text{Borda}}(C) \ge f^i_{\text{Borda}}(D)$,
- $C \succ_i^B D$ if and only if $f^i_{\text{Borda}}(C) > f^i_{\text{Borda}}(D)$, and
- $C \sim_i^B D$ if and only if $f^i_{\text{Borda}}(C) = f^i_{\text{Borda}}(D)$.

Example 14. Recall the FEN-hedonic game from Example 7 with $A = \{1,2,3,4\}$ and $\trianglerighteq_1^{+0-} = (\emptyset \mid \emptyset \mid 2 \rhd_1 3 \sim_1 4)$, $\trianglerighteq_2^{+0-} = (1 \mid \emptyset \mid 3 \sim_2 4)$, $\trianglerighteq_3^{+0-} = (2 \mid \{1,4\} \mid \emptyset)$, and $\trianglerighteq_4^{+0-} = (1 \rhd_4 2 \mid \{3\} \mid \emptyset)$. From player 2's view, the coalitions {2, 3} and {1, 2, 3, 4} are incomparable with respect to $\succeq_2^{+0-}$. Table 2 shows the values player 2 assigns to her co-players 1, 3, and 4 for different choices of scoring vectors $\mathbf{f}_i$ and $\mathbf{e}_i$ and the resulting values of the two mentioned coalitions.

We see that each of the three choices of $\mathbf{f}_i$ and $\mathbf{e}_i$ results in a different relation: While 2 is indifferent for $\mathbf{f}_i = \mathbf{fo}$ and

$\mathbf{e}_i = \mathbf{seo}$ $(\{2,3\} \sim_2^B \{1,2,3,4\})$, she weakly prefers being solely with 3 to being in $\{1,2,3,4\}$ when $\mathbf{f}_i = \mathbf{fo}$ and $\mathbf{e}_i = \mathbf{eo}$ $(\{2,3\} \succ_2^B \{1,2,3,4\})$, but for $\mathbf{f}_i = \mathbf{sfo}$ and $\mathbf{e}_i = \mathbf{seo}$ it holds that $\{1,2,3,4\} \succ_2^B \{2,3\}$.

From the definition of f_{Borda} and Proposition 5 it follows that $\succeq_i^B$ is indeed a preference extension of $\succeq_i^{+0-}$. We state this fact in Proposition 15 without proof.

Proposition 15. *Let $(A, (\trianglerighteq_1^{+0-}, \dots, \trianglerighteq_n^{+0-}))$ be a FEN-hedonic game with n players. It holds that $\succeq_i^B \in \text{Ext}\left(\succeq_i^{+0-}\right)$ for each fixed choice of $\mathbf{f}_i$ and $\mathbf{e}_i$, $i \in \{1,\dots,n\}$.*

Finally, we can define Borda-induced FEN-hedonic games.

Definition 16. Let $(A, (\trianglerighteq_1^{+0-}, \dots, \trianglerighteq_n^{+0-}))$ be a FEN-hedonic game with n players. For a fixed choice of scoring vectors $\mathbf{f}_i$ and $\mathbf{e}_i$ for $i \in \{1,\dots,n\}$, we define with $H = (A, (\succeq_1^B, \dots, \succeq_n^B))$ the *Borda-induced FEN-hedonic game*, where $\succeq_i^B$ are the Borda-induced preference extensions of $\trianglerighteq_i^{+0-}$ for $i \in A$.

Thus Borda-induced FEN-hedonic games are a class of FEN-hedonic games with preference extensions defined by the scoring vectors $\mathbf{f}_i$ and $\mathbf{e}_i$, and each fixed pair of $(\mathbf{f}_i, \mathbf{e}_i)$ defines a subclass thereof.

5. Properties of Borda-induced FEN-hedonic games

We now give an overview of some useful properties that the class of Borda-induced preference extensions fulfills. These properties will allow us to derive several of the complexity results that will be stated in Section 6.

First we analyze the connection of Borda-induced FEN-hedonic games to other classes of hedonic games. By definition, the preferences $\succeq^B$ are additively separable, thus by setting $w_i = f_{\text{Borda}}^i$ for each player $i \in A$, we can represent every Borda-induced FEN-hedonic game as an additively separable hedonic game (recall the formal definition from Section 2).

Observation 17. *Every Borda-induced FEN-hedonic game is an additively separable hedonic game.*

Note that this inclusion is strict: While for each Borda-induced FEN-hedonic game, there is an additively separable hedonic game with the same values, not every set of values can be derived from given weak rankings with double threshold.

When analyzing the complexity of stability for Borda-induced FEN-hedonic games, a first step is to check whether the results due to Peters and Elkind (2015), which we presented in Section 3.2, are applicable. We already noted that Borda-induced FEN-hedonic games allow an arbitrary ordering of players. From Lemma 11 we also know that they are consistent on pairs and strictly 0-1-toxic.

We start with three negative results presented in Proposition 18 through 20, which show that for certain choices of scoring vectors the resulting classes of Borda-induced FEN-hedonic games do not satisfy certain variants of *a*-*b*-toxicity.

Proposition 18. *When scoring vectors $(\mathbf{f}_i, \mathbf{e}_i)$ can be chosen from $\{\mathbf{sfo}, \mathbf{fo}, \mathbf{sfp}, \mathbf{fp}\} \times \{\mathbf{seo}, \mathbf{ep}\}$ or $\{\mathbf{sfo}, \mathbf{fp}\} \times \{\mathbf{sep}, \mathbf{eo}\}$, the resulting subclass of Borda-induced FEN-hedonic games is not* 1-1-*toxic (and thus not strictly* 1-1-*toxic).*

Proof. To show the above claim for each combination of the given scoring vectors, we have to find a profile of weak rankings with double threshold for which there is no derived hedonic game that fulfills the properties. Let $A = \{1,2,3,4\}$ be the set of players

Table 3
Values that players 1 and 2 assign to their co-players in the proof of Proposition 18.

$\mathbf{f}_i$	Player 1			Player 2			$\mathbf{e}_i$
	2	**3**	**4**	**1**	**3**	**4**	
sfo*	4*	−1	3*	4	−1	−2	**seo**
fo	2	−4	1	1	−3	−4	**eo**
sfp	2	−4*	1	1	−3	−4	**sep***
fp	4	−1	3	4	−1	−2	**ep**

and suppose we have the following weak rankings with double threshold:

$$\trianglerighteq_1^{+0-} = (2 \rhd 4 \mid \emptyset \mid 3), \quad \trianglerighteq_2^{+0-} = (1 \mid \emptyset \mid 3 \rhd 4),$$
$$\trianglerighteq_3^{+0-} = \trianglerighteq_4^{+0-} = (\emptyset \mid A \setminus \{i\} \mid \emptyset).$$

The values players 1 and 2 assign to their co-players for different choices of scoring vectors are given in Table 3: The first row determines player 1's and player 2's view, respectively. That is, the values player 1 assigns can be found in the twelve entries in the left part of the table and the values player 2 assigns in the twelve entries in the right part of the table. The entries in boldface in the second row denote the co-players of player 1 and player 2, respectively.

Let us, exemplarily, focus on the left part of the table, that is, player 1's view. She can assign values to players 2, 3, and 4 and these are given in boldface in the left part of the second row. The respective column below each of these players gives the value that player 1 assigns to her. Since each co-player of 1 is either a friend (thus the value is determined by the choice of $\mathbf{f}_i$) or an enemy (the value is determined by the choice of $\mathbf{e}_i$), the four rows suffice to display all possible values player 1 can assign to each of her co-players.

For example, when $\mathbf{f}_i = \mathbf{sfo}$ and $\mathbf{e}_i = \mathbf{sep}$, player 1 assigns player 2 a value of 4, player 3 a value of −4, and player 4 a value of 3 (these values and the choice of scoring vectors are marked with an asterisk in the table).

For the coalition $S = \{1,2,3\}$ and an arbitrary choice of $(\mathbf{f}_i, \mathbf{e}_i) \in \{\mathbf{seo}, \mathbf{ep}\} \times \{\mathbf{sfo}, \mathbf{fo}, \mathbf{sfp}, \mathbf{fp}\}$, we have that $f_{\text{Borda}}^1(S) > 0 = f_{\text{Borda}}^1(\{1\})$, which is equivalent to $S \succ_1 \{1\}$. For the same coalition and the scoring vectors from $\{\mathbf{sep}, \mathbf{eo}\} \times \{\mathbf{sfo}, \mathbf{fp}\}$, we obtain the same contradiction from player 2's view and we have shown that for these pairs of scoring vectors, (strict) 1-1-toxicity is not fulfilled. □

Proposition 19. *The subclass of Borda-induced FEN-hedonic games when scoring vectors $(\mathbf{f}_i, \mathbf{e}_i)$ can be chosen from $\{\mathbf{fp}, \mathbf{fo}, \mathbf{sfo}\} \times \{\mathbf{seo}, \mathbf{ep}\}$ is not weakly* 1-1-*toxic.*

Proof. Recall the game defined in the proof of Proposition 18. It holds for each of the above specified choices of scoring vectors that $f_{\text{Borda}}^1(\{1,4\}) = 1 = f_{\text{Borda}}^1(\{1,2,3\})$, which contradicts the condition for weak 1-1-toxicity. □

Recall from Section 3.2 that a profile of preference orderings on A is said to be *mutual* if $j \in A_i^+ \cup A_i^0$ if and only if $i \in A_j^+ \cup A_j^0$.

Proposition 20. *When scoring vectors $(\mathbf{f}_i, \mathbf{e}_i)$ can be chosen from $\{\mathbf{sfo}, \mathbf{fp}\} \times \{\mathbf{eo}\}$, the subclass of Borda-induced FEN-hedonic games is neither weakly* 2-2-*toxic, nor* 2-2-*toxic, nor strictly* 2-2-*toxic, not even when the profile of orderings is mutual and every player has at most three players that are no enemies.*

Proof. We have to find a Borda-induced FEN-hedonic game with mutual rankings for at most three players being no enemies, such that every derived hedonic game is not weakly 2-2-toxic, nor 2-2-toxic, nor strictly 2-2-toxic. In particular, it is enough to find one such ranking that violates these properties. Due to the structure of our following counterexamples, two slightly different examples are sufficient to disprove the three properties for each combination of scoring vectors.

For $n = 8$ players, say $A = \{a, b, c, d, e, f, g, i\}$, let $\trianglerighteq_i = a \sim_i b \triangleright_i \emptyset \triangleright_i c \sim_i d \sim_i e \sim_i f \sim_i g$ be player i's weak order induced by her weak ranking with double threshold (where the other players' preferences are arbitrary as long as they each are not enemies with at most three and the resulting preference profile is mutual). Consider coalition $S = \{a, b, c, d, i\}$. Then $\|S \cap (A_i^+ \cup A_i^0)\| = 2$ and $\|S \cap A_i^-\| = 2$. We have to show that $f^i_{\text{Borda}}(S) \geq f^i_{\text{Borda}}(\{i, j\})$ for all $j \in A_i^+ \cup A_i^0$, which directly disproves weak 2-2-toxicity. This is also enough to disprove 2-2-toxicity and strict 2-2-toxicity, as $f^i_{\text{Borda}}(\{i, j\}) > f^i_{\text{Borda}}(\{i\})$ holds for all $j \in A_i^+ \cup A_i^0$. For the first combination, **sfo** with **eo**, we have

$$\begin{aligned} f^i_{\text{Borda}}(S) &= w_i(a) + w_i(b) + w_i(c) + w_i(d) \\ &= 2n - 2(n - \|A_i^-\| + 1) \\ &= 16 - 8 \\ &= 8 = f^i_{\text{Borda}}(\{i, j\}), \end{aligned}$$

which is exactly what we wanted to show. For **fp** with **eo**, we just need to add one more enemy to $\trianglerighteq_i^{+0-}$, which is tied with all the other enemies of i, such that the resulting scores are the same as above. Hence, for both combinations, none of the three properties hold. □

These results imply that for these choices of scoring vectors we cannot apply the results due to Peters and Elkind (2015) and we have to provide specific hardness proofs in Section 6. The following results, on the other hand, will be very useful in Section 6.

Proposition 21. *Let $(A, (\trianglerighteq_1^{+0-}, \dots, \trianglerighteq_n^{+0-}))$ be a FEN-hedonic game with n players in which every player is enemies with all but at most three other players, and $(\mathbf{f}_i, \mathbf{e}_i) \in \{\mathbf{fo}, \mathbf{sfp}\} \times \{\mathbf{eo}\}$ for all $i \in \{1, \dots, n\}$. For each $x \in \{1, 2, 3\}$, every Borda-induced FEN-hedonic game $(A, (\succeq_1^B, \dots, \succeq_n^B))$ is strictly x-x-toxic (and therefore x-x-toxic and weakly x-x-toxic as well) and strictly 2-5-toxic.*

Proof. Let $(A, (\trianglerighteq_1^{+0-}, \dots, \trianglerighteq_n^{+0-}))$ be a FEN-hedonic game with $\|A_i^+ \cup A_i^0\| \leq 3$ for all players i, let $i \in A$ be a player, and let $S \subseteq A$ be a subset of the players with $i \in S$. We have to show that if

$$\|S \cap (A_i^+ \cup A_i^0)\| = x \tag{2}$$

and $\|S \cap A_i^-\| \geq x$, then $\{i\} \succ_i^B S$. First, we can safely assume, that

$$\|S \cap A_i^-\| = x, \tag{3}$$

as adding more enemies to S makes S strictly less attractive for i. Second, we can again assume that

$$S \cap A_i^- \subseteq A_{i,1}^-, \tag{4}$$

as for **eo** (and all other scoring vectors), the score only gets lower if $S \cap A_i^- \subseteq A_{i,t}^-$ for any t with $1 < t \leq m$, resulting in S being less preferred by i. Last, for any $p \in A_i^+ \cup A_i^0$,

$$w_i(p) \leq \|A_i^+\| \tag{5}$$

is another safe assumption that can be made, as this is the single highest weight a friend can contribute to S for both **fo** and **sfp**.

To show $\{i\} \succ_i^B S$, we need to show $f^i_{\text{Borda}}(S) < f^i_{\text{Borda}}(\{i\})$. The following equations are correct for both combinations, i.e., for **eo** with **fo** as well as for **eo** with **sfp**. It holds that

$$\begin{aligned} f^i_{\text{Borda}}(S) &= \sum_{j \in S \cap A_i^+} w_i(j) + \sum_{j \in S \cap A_i^-} w_i(j) \\ &\leq \sum_{j \in S \cap A_i^+} \|A_i^+\| + \sum_{j \in S \cap A_i^-} -(n - \|A_i^-\| + 1) \\ &\quad \text{due to (4) and (5)} \\ &= x\|A_i^+\| - x(n - \|A_i^-\| + 1) \\ &\quad \text{due to (2) and (3)} \\ &= x(-n + \|A_i^+\| + \|A_i^-\| - 1) \\ &= x(-\|A_i^0\| - 1 - 1) \\ &\quad \text{due to } n = \|A_i^+\| + \|A_i^0\| + \|A_i^-\| + 1 \\ &\leq -2x < 0 = f^i_{\text{Borda}}(\{i\}). \end{aligned}$$

Together with Lemmas 9(a), 9(b), and 11, this implies the desired properties. □

6. Complexity results for stability in Borda-induced FEN-hedonic games

In this section we present the results we obtained regarding the complexity of those verification and existence problems we defined in Section 2 when the considered game is from the class of Borda-induced FEN-hedonic games. Table 4 gives an overview of our results. Unless it is mentioned otherwise in the table, all results hold for each choice of scoring vectors.

We start with the complexity results for the verification problems. Recalling from Observation 17 that every Borda-induced FEN-hedonic game is also additively separable, we can transfer known upper bounds for these games to our new subclass. For the verification problem, these results are summarized in the following corollary.

Corollary 22. *For Borda-induced FEN-hedonic games the problem γ-Verification is in P for each of the stability concepts $\gamma \in$ {perfectness, individual stability, contractually individual stability, Nash stability}.*

While the verification problems regarding individual deviations are tractable, we will see that verifying whether a given coalition structure in a Borda-induced FEN-hedonic game is core stable or strictly core stable are far more complicated tasks. The proof for Theorem 23 is inspired by the result for games with enemy-oriented preferences presented by Sung and Dimitrov (2007).

Theorem 23. *For Borda-induced FEN-hedonic games with each choice of $\mathbf{f}_i$ and $\mathbf{e}_i$, the problems Core-Stability-Verification and Strict-Core-Stability-Verification are coNP-complete.*

Proof. The upper bound follows from the result for additively separable hedonic games due to Sung and Dimitrov (2007) and Aziz et al. (2013b) and Observation 17.

To prove coNP-hardness, we reduce from the complement of the Clique problem, denoted by $\overline{\text{Clique}}$. To do so, let (G, k) be a $\overline{\text{Clique}}$ instance, where $G = (V, H)$ is an undirected graph with vertex set $V = \{v_1, v_2, \dots, v_n\}$ and edge set $H = \{h_1, h_2, \dots, h_m\}$. We construct the Borda-induced FEN-hedonic game $(A, \succeq^B)$ with $n + n(k - 2)$ players in $A = \{v_1, v_2, \dots, v_n\} \cup Q$, where Q is a profile of $n(k - 2)$ players $Q = \bigcup_{i=1}^n Q_i$ with the sets $Q_i = \{q_{i,1}, q_{i,2}, \dots, q_{i,(k-2)}\}$. Let $\mathrm{N}(v)$ denote the neighborhood of vertex $v \in V$.

The profile of extensions $\succeq^B$ can be derived from the players' profile of weak rankings with double threshold $\trianglerighteq^{+0-}$ as displayed in Table 5 (note that when a set of players appears in a preference, the players in the set are unranked and the subscripts $\sim$ are dropped).

The players corresponding to the vertices in G are mutual friends if and only if they are connected by an edge, and each of

Table 4
Overview of complexity results regarding stability for Borda-induced FEN-hedonic games.

Stability	VERIFICATION	Reference	EXISTENCE	Reference
Perfectness	P	Corollary 22	P	Corollary 24
Individual stability	P	Corollary 22	NP-complete[a]	Theorem 29
Contractual individual stability	P	Corollary 22	P	Corollary 24
Nash stability	P	Corollary 22	NP-complete[a]	Theorem 28
Core stability	coNP-complete	Theorem 23	Σ_2^p-complete[b]	Theorem 31
Strict core stability	coNP-complete	Theorem 23	coNP-hard, $\in \Sigma_2^p$	Theorem 30

[a] For {**sfp**} × {**seo**, **eo**, **sep**, **ep**}, (**fo**, **eo**).
[b] For {**sfp**} × {**seo**, **ep**}.

Table 5
Weak rankings with double threshold of the players in the proof of Theorem 23.

For each ...	player	$\trianglerighteq^+$	A^0	$\trianglerighteq^-$
$i \in \{1, \ldots, n\}$	v_i	$N(v_i) \cup Q_i$	$A \setminus (N(v_i) \cup \{v_i\} \cup Q_i)$	$\emptyset$
$i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, k-2\}$	$q_{i,j}$	$\emptyset$	$\{v_i\} \cup (Q_i \setminus \{q_{i,j}\})$	$A \setminus (\{v_i\} \cup Q_i)$

these players has $k-1$ friends in Q_i that are no friends of the other v_i-players. For each $i \in \{1, \ldots, n\}$, the players in Q_i are indifferent regarding their corresponding player v_i and the players that are in the same Q_i. The remaining players in the game are their enemies, so these players do not consider anyone to be their friend.

Let $\Gamma = (\Gamma_1, \Gamma_2, \ldots, \Gamma_n)$ with $\Gamma_i = \{v_i\} \cup Q_i$ be the coalition structure. The v_i-players give their coalition in Γ $\ell(k-2)$ points, while $\ell \geq 1$ depends on the scoring vector $\mathbf{f}_i$ used for the set of friends. All players $q_{i,j} \in Q$ give their coalition a score of zero (and this is independent of the choice of $\mathbf{f}_i$ and $\mathbf{e}_i$). Note that adding any other player to $\Gamma(q_{i,j})$ turns the score of the coalition from player $q_{i,j}$'s view to a negative value.

We claim that G has a clique of size at most $k-1$ if and only if Γ is (strictly) core stable.

Only if: Assume that the largest clique in G is of size $k-1$. Since the players in Q do not have friends, they already reach a best possible score with their given coalition. For a weakly blocking coalition $P \subseteq A$ to exist, it has to contain at least one player from V preferring P to her original coalition. This can only happen if P consists of a set of players from V forming a clique. Since the largest clique in V is of size at most $k-1$, the players in the clique would assign this coalition a score of $\ell(k-2)$, which is exactly the score each v_i assigns the coalition $\Gamma(v_i)$. Thus there is no weakly blocking coalition, which directly implies that there is neither a blocking one.

If: We show the contraposition. Assume that there was a clique V' of size k in G. Then the players corresponding to the vertices in this clique form a blocking coalition (and thus a weakly blocking one) since every player in the clique gives the coalition V' a score of $\ell(k-1)$, which is larger than the score of the coalition they are assigned to in Γ. □

Now we turn to the existence problems and start with the upper bounds for all problems for which Observation 17 can be applied. For perfectness and contractually individual stability, this results in the following corollary.

Corollary 24. *For Borda-induced FEN-hedonic games with each choice of* $\mathbf{f}_i$ *and* $\mathbf{e}_i$*, the problems* PERFECTNESS-EXISTENCE *and* CONTRACTUALLY-INDIVIDUAL-STABILITY-EXISTENCE *are in* P.

For the remaining stability problems, we have a higher computational complexity and will now further analyze their lower bounds. To do so, we will make use of known hardness proofs for the class of additively separable hedonic games and show that these can be transferred to proofs suitable for Borda-induced FEN-hedonic games if the following two properties are fulfilled by the game constructed in the original hardness proof: The values that the players assign to each other have to be integers and they are not allowed to be symmetric.

Whenever these conditions are met, we can construct an equivalent Borda-induced FEN-hedonic game when the scoring vectors $\mathbf{f}_i = \mathbf{sfp}$ and $\mathbf{e}_i = \mathbf{seo}$ are used. We will further specify the notion of equivalence of two games in Lemma 27.

Construction 25 illustrates how a Borda-induced FEN-hedonic game can be derived from an arbitrary additively separable hedonic game fulfilling the two conditions above.

Construction 25. *Let $H = (A, w)$ be an additively separable hedonic game, where the integer values $w_{p_i} : A \setminus \{p_i\} \to R_{p_i}$ that the players $p_i \in A$ assign to the other players are not symmetric and where $R_{p_i} \subseteq \mathbb{Z}$ denotes the range of values that p_i assigns. We construct a Borda-induced FEN-hedonic game $H' = (A', \succeq^B)$ with* $\mathbf{f}_i = \mathbf{sfp}$ *and* $\mathbf{e}_i = \mathbf{seo}$*. Let $A' = A \cup D$ be the set of players in H', where A are the players in the original game H and we have a set of $z = \max\{\bigcup_{p_i \in A} R_{p_i}\} + |\min\{\bigcup_{p_i \in A} R_{p_i}\}| - 2$ padding players in $D = \{d_1, d_2, \ldots, d_z\}$.*

We first explain how the weak rankings with double threshold have to be constructed for the players in A. To this end, let player $p_i \in A$ be a player in the original game, and define the sets $A_{p_i}^k = \{p_j \in A \setminus \{p_i\} \mid w_{p_i}(p_j) = k\}$ for $k \in R_{p_i}$. We know that $\bigcup_{s \in R} A_{p_i}^s = A \setminus \{p_i\}$. We separate the strictly negative values in R_{p_i} (denoted by R^+) from the strictly positive ones (denoted by R^-), where we omit the index p_i for R^+ and R^- for the sake of readability. Thus $R_{p_i} = R^+ \cup R^- \cup \{0\}$. For each $p_i \in A$, we define the set of p_i's friends by $A_{p_i}^+ = \bigcup_{s \in R^+} A_{p_i}^s$, the set of p_i's enemies by $A_{p_i}^- = \bigcup_{s \in R^-} A_{p_i}^s$, and the set of neutral players for p_i is $A_{p_i}^0$.

Assuming that the elements in $R^+ = \{r_1, r_2, \ldots, r_{\|R^+\|}\}$ and $R^- = \{r'_1, r'_2, \ldots, r'_{\|R^-\|}\}$ are ordered descendingly, we can define $\trianglerighteq_{p_i}^+$ and $\trianglerighteq_{p_i}^-$ as follows (note again that we omit the index p_i in both $\trianglerighteq_{p_i}^+$ and $\trianglerighteq_{p_i}^-$ when it is clear from the context). Let $D^1, \ldots, D^{r_{\|R^+\|}}, \hat{D}^1, \ldots, \hat{D}^{r'_{\|R^-\|}} \subseteq D$ be $r_{\|R^+\|} + r'_{\|R^-\|}$ suitably sized, pairwise disjoint subsets of the padding players such that:

$$
\begin{aligned}
&\|D^1\| = r_1 - r_2 - 1, \quad \|D^2\| = r_2 - r_3 - 1, \quad \ldots, \\
&\|D^{r_{\|R^+\|}}\| = r_{\|R^+\|} - 1, \\
&\|\hat{D}^1\| = -r'_1 - 1, \quad \|\hat{D}^2\| = -r'_1 + r'_2 - 1, \quad \ldots, \\
&\|\hat{D}^{r'_{\|R^-\|}}\| = -r'_{\|R^-\|} - 1,
\end{aligned}
$$

and for each such subset $D^j = \{d_s, d_{s+1}, \ldots, d_t\}$ let $D_\triangleright^j$ be a shorthand for the ranking $d_s \triangleright d_{s+1} \triangleright \cdots \triangleright d_t$ (and analogously so for $\hat{D}_\triangleright^j$), where again the subscript p_i is omitted on $\triangleright$. Let D' be the set of

Table 6
Values of the players in H.

p_i	$w_{p_i}(p_j)$					A^2	A^1	A^{-4}
	p_j							
	p_1	p_2	p_3	p_4	p_5			
p_1	*	2	2	−4	1	$\{p_2, p_3\}$	$\{p_5\}$	$\{p_4\}$
p_2	2	*	−4	−4	−4	$\{p_1\}$	$\emptyset$	$\{p_3, p_4, p_5\}$
p_3	−4	−4	*	−4	−4	$\emptyset$	$\emptyset$	$\{p_1, p_2, p_4, p_5\}$
p_4	1	1	1	*	1	$\emptyset$	$\{p_1, p_2, p_3, p_5\}$	$\emptyset$
p_5	2	2	2	2	*	$\{p_1, p_2, p_3, p_4\}$	$\emptyset$	$\emptyset$

Table 7
Constructed preferences from Example 26.

Player	$\trianglerighteq^+$	A^0	$\trianglerighteq^-$
p_1	$p_2 \sim p_3 \triangleright p_5$	$\emptyset$	$d_1 \triangleright d_2 \triangleright d_3 \triangleright p_4 \triangleright d_4$
p_2	$p_1 \triangleright d_1$	$\emptyset$	$d_2 \triangleright d_3 \triangleright d_4 \triangleright p_3 \sim p_4 \sim p_5$
p_3	$\emptyset$	$\emptyset$	$d_1 \triangleright d_2 \triangleright d_3 \triangleright p_1 \sim p_2 \sim p_4 \sim p_5 \triangleright d_4$
p_4	$p_1 \sim p_2 \sim p_3 \sim p_5$	$\emptyset$	$d_1 \sim d_2 \sim d_3 \sim d_4$
p_5	$p_1 \sim p_2 \sim p_3 \sim p_4 \triangleright d_1$	$\emptyset$	$d_2 \sim d_3 \sim d_4$

remaining padding players for p_i. Then we define

$$\trianglerighteq^+_{p_i} : A^{r_1}_{\sim} \triangleright D^1_{\triangleright} \triangleright A^{r_2}_{\sim} \triangleright D^2_{\triangleright} \triangleright \cdots \triangleright A^{r_{\|R^+\|}}_{\sim} \triangleright D^{r_{\|R^+\|}}_{\triangleright} \quad \textit{and}$$

$$\trianglerighteq^-_{p_i} : \hat{D}^1_{\triangleright} \triangleright A^{r'_1}_{\sim} \triangleright \hat{D}^2_{\triangleright} \triangleright A^{r'_2}_{\sim} \triangleright \cdots \triangleright \hat{D}^{r'_{\|R^-\|}}_{\triangleright} \triangleright A^{r'_{\|R^-\|}}_{\sim} \triangleright D'_{\sim}.$$

Note that whenever $\|R^+\| = 1$ or $\|R^-\| = 1$ holds for a player p_i, we have $R^+ = \{r_1\}$ and $R^- = \{r'_1\}$, so $A^{r_{\|R^+\|}} = A^{r_1} = A^+_{p_i}$ and $A^{r'_{\|R^-\|}} = A^{r'_1} = A^-_{p_i}$, and $\trianglerighteq^+_{p_i}$ and $\trianglerighteq^-_{p_i}$ are defined by the last part of the above description, namely, for $D^{r_{\|R^+\|}} = D^{r_1}$ with $r_1 - 1$ padding players and for $\hat{D}^{r'_{\|R^-\|}} = \hat{D}^{r'_1}$ with $-r'_1 - 1$ padding players, we have

$$\trianglerighteq^+_{p_i} : A^{r_1}_{\sim} \triangleright D^{r_1}_{\triangleright} \quad \textit{and}$$

$$\trianglerighteq^-_{p_i} : \hat{D}^{r'_1}_{\triangleright} \triangleright A^{r'_1}_{\sim} \triangleright D'_{\sim}.$$

The padding players have no friends and no neutral players but only enemies, that is, for $d \in D$, we define $A^+_d = \emptyset = A^0_d$ and $A^- = A' \setminus \{d\}$, and we let them be indifferent between their enemies:

$$\trianglerighteq^{+0-}_d = (\emptyset \mid \emptyset \mid (A' \setminus \{d\})_{\sim}).$$

We illustrate the construction in the following example.

Example 26. Let $H = (A, w)$ be a hedonic game with the set of players $A = \{p_1, p_2, p_3, p_4, p_5\}$ and the values $w_{p_i}(p_j)$ for all players $p_i, p_j \in A$ given in Table 6, together with the resulting sets A^{-4}, A^1, and A^2.

We need $2 + 4 - 2 = 4$ padding players, d_1, d_2, d_3, and d_4, to construct the weak rankings with double threshold, which we present in Table 7.

Lemma 27. *Let $H = (A, w)$ be an additively separable hedonic game, where the values the players assign to each other are integers and the preferences are not symmetric. Let furthermore $H' = (A', \succeq^B)$ with $A' = A \cup D$ be a Borda-induced FEN-hedonic game with $\mathbf{f}_i = \mathbf{sfp}$ and $\mathbf{e}_i = \mathbf{seo}$ constructed from H according to Construction 25 and let Γ be a coalition structure in H and $\Gamma' = \Gamma \cup \bigcup_{i=1}^{\|D\|}\{d_i\}$ be a coalition structure in H'. For each stability concept γ defined in Section 2, it holds that Γ is stable in the sense of γ in H if and only if Γ' is stable in the sense of γ in H'.*

Proof. Each padding player $d_i \in D$ assigns a negative value to all players in $A' \setminus \{d_i\}$, so there are no acceptable coalitions for $d_i \in D$ except the singleton coalition $\{d_i\}$. Clearly, for each stability concept γ defined in Section 2, a given coalition structure Γ' can only be stable in the sense of γ if it assigns each $d_i \in D$ to the coalition $\{d_i\}$. With this, the above equivalence directly follows. □

Sung and Dimitrov (2010, Lemma 2 and Theorem 3) show that in additively separable hedonic games the problem Nash-Stability-Existence is NP-complete. We will show that for certain choices of scoring vectors, we can obtain the same hardness result in Borda-induced FEN-hedonic games. Note that Theorem 28 (and the same comment applies, e.g., to Theorem 31) crucially follows from Observation 17 stating that Borda-induced FEN-hedonic games are additively separable. In general, however, Burani and Zwicker (2003) have shown that responsive preferences do not imply additive separability of preferences in hedonic games.[6] Indeed, it is due to our particular approach of completing preferences over coalitions via Borda counts that gives Observation 17 and thus makes the findings by Sung and Dimitrov (2010) (or, in the case of Theorem 31, the findings by Woeginger Woeginger, 2013b) applicable to Borda-induced FEN-hedonic games. Interestingly, the specific structure of Borda-induced FEN-hedonic games does not lower the computational complexity of the related stability problems.

Theorem 28. *In Borda-induced FEN-hedonic games with each choice of scoring vectors $(\mathbf{f}_i, \mathbf{e}_i)$ from $\{\mathbf{sfp}\} \times \{\mathbf{seo}, \mathbf{eo}, \mathbf{sep}, \mathbf{ep}\}$ or $(\mathbf{f}_i, \mathbf{e}_i) = (\mathbf{fo}, \mathbf{eo})$, the problem* Nash-Stability-Existence *is* NP*-complete.*

Proof. With Observation 17 and Lemma 2 of Sung and Dimitrov (2010), the problem is in NP.

NP-hardness in the setting of additively separable hedonic games is shown by a reduction from X3C and the players in the constructed game assign values from $\{-68, 1, 2, 13, 16\}$ to each other.

For the choice of $\mathbf{f}_i = \mathbf{sfp}$ and $\mathbf{e}_i = \mathbf{seo}$, we can use Construction 25 and Lemma 27 to apply the argumentation in the proof of Theorem 3 of Sung and Dimitrov (2010).

The value -68 is the only negative value that is assigned in the additively separable hedonic game from the original proof and the argumentation remains unchanged if this value was smaller than -68. We show that for the other possible choices of $\mathbf{e}_i$ this negative value, let us call it K, is always at most -68.

Recalling the notation from Construction 25, we have that for each player $p_i \in A \subseteq A'$ with $A^-_{p_i} \neq D$, the ordering of the enemies is

$$\trianglerighteq^-_{p_i} : \hat{D}^{-68}_{\triangleright} \triangleright A''_{\sim} \triangleright D'_{\sim},$$

where we first have $\|\hat{D}^{-68}\| = 67$ padding players, then the set A'' of players p_i assigns value -68 to in the original game, followed

[6] We consider it a challenging question for future research to further explore the relation of responsive and additively separable preferences in FEN-hedonic games.

by up to 15 padding players in D' that are not contained in $A^+_{p_i}$. The set A'' corresponds to the set $A^-_{p_i,68}$ in the definition of the scoring vectors $\mathbf{e}_i$ in Section 4, and it is easy to see that $K \leq -68$ for each fixed choice of $\mathbf{e}_i \in \{\mathbf{eo}, \mathbf{sep}, \mathbf{ep}\}$.

This leaves the case of $(\mathbf{f}_i, \mathbf{e}_i) = (\mathbf{fo}, \mathbf{eo})$. From Lemma 11 and Proposition 21 we know that for every Borda-induced FEN-hedonic game $(A, (\succeq^B_1, \ldots, \succeq^B_n))$ with scoring vectors as above, there is an induced hedonic game that is consistent on pairs, 0-1-toxic, weakly 1-1-toxic, and weakly 2-2-toxic. Thus Theorem 10.2. is applicable and we obtain NP-hardness of Nash-Stability-Existence. □

With an analogous argumentation we can show the following result.

Theorem 29. *In Borda-induced FEN-hedonic games with* $(\mathbf{f}_i, \mathbf{e}_i) \in \{\mathbf{sfp}\} \times \{\mathbf{seo}, \mathbf{eo}, \mathbf{sep}, \mathbf{ep}\}$ *or* $(\mathbf{f}_i, \mathbf{e}_i) = (\mathbf{fo}, \mathbf{eo})$*, the problem* Individual-Stability-Existence *is* NP*-complete.*

Proof. NP membership follows straightforwardly with Observation 17 and Lemma 2 of Sung and Dimitrov (2010). In their NP-hardness proof, they construct an additively separable hedonic game from an X3C instance in which the players' values are from $\{-4, 2, 1\}$. We can adapt this proof to our setting by constructing a Borda-induced FEN-hedonic game with Construction 25 and applying Lemma 27.

For the other choices of $\mathbf{e}_i$, we can argue that assigning a value K that is smaller than -4, the original argumentation still applies. For the players $p_i \in A \subseteq A'$ with $A^-_{p_i} \neq D$, Construction 25 defines $\trianglerighteq^-_{p_i}$ to be:

$$\trianglerighteq^-_{p_i}: d_1 \triangleright d_2 \triangleright d_3 \triangleright A''_{\sim} \triangleright D'_{\sim},$$

where D' can have up to two elements. Here we have that A'' corresponds to $A^-_{p_i,4}$ in the definition of the scoring vectors $\mathbf{e}_i$ in Section 4 and it is, again, easy to see that for each fixed choice of $\mathbf{e}_i \in \{\mathbf{eo}, \mathbf{sep}, \mathbf{ep}\}$, it holds that $K \leq -4$.

Similarly to the proof of Theorem 28, we obtain NP-hardness for $(\mathbf{f}_i, \mathbf{e}_i) = (\mathbf{fo}, \mathbf{eo})$ with the results shown in Lemma 11, Proposition 21, and Theorem 10.2. □

Theorem 30. *For Borda-induced FEN-hedonic games with each choice of* $\mathbf{f}_i$ *and* $\mathbf{e}_i$*,* Strict-Core-Stability-Existence *is* coNP*-hard.*

Proof. We show coNP-hardness by a reduction from $\overline{\text{CLIQUE}}$ with a similar construction as the one used in the proof of Theorem 23.

To this end, let $G = (V, H)$ be an undirected graph with $V = \{v_1, v_2, \ldots, v_n\}$ and $H = \{h_1, h_2, \ldots, h_m\}$ and let $k \geq 2$ be a positive integer. Recall that $\mathrm{N}(v)$ denotes the neighborhood of vertex $v \in V$, and let $\mathrm{N}[v] = \mathrm{N}(v) \cup \{v\}$ be the closed neighborhood of v.

Construct the Borda-induced FEN-hedonic game $(A, \succeq^B)$ with the set of players $A = V \cup Q \cup R \cup T$, where the players $v_i \in V$ correspond to the vertices in the graph, $Q = \bigcup_{i=1}^n Q_i$ with $Q_i = \{q_{i,1}, q_{i,2}, \ldots, q_{i,k-2}\}$, $R = \{r_1, r_2, \ldots, r_n\}$, and $T = \{t_1, t_2, \ldots, t_n\}$. The players' weak rankings with double threshold are shown in Table 8 (note again that when a set of players appears in a preference, the players in the set are unranked and the subscripts $\sim$ are dropped).

For each $v_i \in V$, all players in Q_i are v_i's friends and so are all other players in V that are connected to v_i by an edge in G. The players in each Q_i only consider v_i to be a friend, do not care about the other players in Q_i or r_i or t_i, while the remaining players are enemies. For the players in R and T, both r_i and t_i consider $q_{i,1}$ to be their only friend for each $i \in \{1, \ldots, n\}$, they both do not care about the other players in Q_i, while considering each other to be enemies (and the remaining players are their enemies as well).

We claim that $(G, k) \notin$ Clique if and only if there exists a strictly core stable coalition structure for $(A, \succeq^B)$ for each choice of $\mathbf{f}_i$ and $\mathbf{e}_i$.

Only if: Assume there is no clique of size k in G. Then

$$\Gamma = (P^v_1, P^v_2, \ldots, P^v_n, P^r_1, P^r_2, \ldots, P^r_n, P^t_1, P^t_2, \ldots, P^t_n)$$

with $P^v_i = \{v_i\} \cup Q_i$, $P^r_i = \{r_i\}$, and $P^t_i = \{t_i\}$ is a strictly core stable coalition structure for $(A, \succeq^B)$: The players in the coalitions P^v_i are in their best valued coalitions, thus every coalition containing them would not be a weakly blocking coalition. This only leaves the players in R and T, which all are enemies, so these cannot form a weakly blocking coalition either. Thus the coalition structure is strictly core stable.

If: We show the contraposition. Assume that there is a clique of size k in G, say V'. To construct a contradiction, let Γ be a strictly core stable coalition structure. For Γ to be strictly core stable, the players corresponding to the vertices in the clique V' have to be together in a coalition in Γ and no other players can be contained in this coalition. Let the set $J = \{i \mid v_i \in V'\}$ denote the indices of the vertices that are contained in the clique V'. For these $j \in J$, we have that the players in Q_j (and, in particular, $q_{j,1}$) cannot form a coalition with their friend v_j, so the players r_j and t_j are both interested in forming a coalition with player $q_{j,1}$. The players in each Q_i can be assigned to coalitions in four different ways:

1. $\{r_j, Q_j\}$; then $\{t_j, q_{j,1}\}$ would be a weakly blocking coalition.
2. $\{t_j, Q_j\}$; then $\{r_j, q_{j,1}\}$ would be a weakly blocking coalition.
3. $\{t_j, r_j, Q_j\}$; then both $\{r_j, q_{j,1}\}$ and $\{t_j, q_{j,1}\}$ would be weakly blocking coalitions.
4. $\{Q_j\}$; then $\{r_j, q_{j,1}\}$ and $\{t_j, q_{j,1}\}$ would be weakly blocking coalitions.

We see that in all cases there exists a weakly blocking coalition, so Γ cannot be strictly core stable. □

We now turn to the Σ^p_2 result for the existence of core stable coalition structures. The proof is an adaption of the corresponding result for additively separable hedonic games, which was shown by Woeginger (2013b). Since the proof is very technical and for the sake of comparability, we will refrain from altering the proof's structure and maintain the structure presented by Woeginger (2013b). We state the result in Theorem 31 and prove it in several steps via Construction 32 and Lemmas 33, 34, and 35.

Theorem 31. *In Borda-induced FEN-hedonic games the problem* Core-Stability-Existence *is* Σ^p_2*-complete for the choice of scoring vectors* $\mathbf{f}_i = \mathbf{sfp}$ *and* $\mathbf{e}_i \in \{\mathbf{seo}, \mathbf{ep}\}$.

Proof. Woeginger (2013b) shows Σ^p_2-completeness of Core-Stability-Existence for additively separable hedonic games with a reduction from 2-Quantified-3-DNF-SAT defined in Section 2. Our approach defined in Construction 25 cannot be applied directly, but with careful adaptions we can define a Borda-induced FEN-hedonic game for which Woeginger's argumentation still works:

Let m be the number of clauses and n the number of variables in a given instance of the problem 2-Quantified-3-DNF-SAT. The values in the original game are from the set

$$\{-\infty, -2, 0, \epsilon, 1, 2, 3, 4, 5, n+2, m+n+1, 4n+m-1\},$$

where $-\infty$ denotes a "small enough number" and $\epsilon = 1/n+1$. To define a Borda-induced FEN-hedonic game, we have to define the exact value for $-\infty$ and change ϵ to a positive integer while preserving the central argumentation. We present the definition of our Borda-induced FEN-hedonic game in Construction 32 and show in Lemmas 33 through 35 where and how Woeginger's argumentation has to be adapted.

Table 8
Weak rankings with double threshold of the players in the proof of Theorem 30.

For each …	Player	$\trianglerighteq^+$	A^0	$\trianglerighteq^-$
$i \in \{1, \dots, n\}$	v_i	$\mathrm{N}(v_i) \cup Q_i$	$A \setminus (\{\mathrm{N}[v_i] \cup Q_i\})$	$\emptyset$
$i \in \{1, \dots, n\}$, $j \in \{1, \dots, k-2\}$	$q_{i,j}$	v_i	$(Q_i \setminus \{q_{i,j}\}) \cup \{r_i, t_i\}$	$A \setminus (A^+_{q_{i,j}} \cup A^0_{q_{i,j}})$
$i \in \{1, \dots, n\}$	r_i	$q_{i,1}$	$Q_i \setminus \{q_{i,1}\}$	$A \setminus (A^+_{r_i} \cup A^0_{r_i})$
$i \in \{1, \dots, n\}$	t_i	$q_{i,1}$	$Q_i \setminus \{q_{i,1}\}$	$A \setminus (A^+_{t_i} \cup A^0_{t_i})$

Table 9
$\trianglerighteq^+$ of the players in the proof of Theorem 31 for (**sfp**, **seo**).

Value:	$m+n+1$		$\cdots$		$n+2$		$\cdots$		1
$\trianglerighteq^+_{q_t}$:	q'_t	$\triangleright$	$\cdots$	$\triangleright$	q''_t	$\triangleright$	$\cdots$	$\triangleright$	$P_X \cup P_C$
Value:			$4n+2m-1$		$\cdots$		2		1
$\trianglerighteq^+_r$:			r'	$\triangleright$	$\cdots$	$\triangleright$	$P_C \cup P_X \cup P_Y$	$\triangleright$	d
Value:			4		3		2		1
$\trianglerighteq^+_{p(x)}$:			q_f	$\triangleright$	d	$\triangleright$	r	$\triangleright$	$(P_X \setminus \{\overline{x}\}) \cup \{q_t\}$
Value:					$n+1$		$\cdots$		1
$\trianglerighteq^+_{q'_t}$:					q_t	$\triangleright$	$\cdots$	$\triangleright$	P_X
Value:					6		5		
$\trianglerighteq^+_{p(c)}$:					q_t	$\triangleright$	r	$\triangleright$	$\cdots$
Value:							2		1
$\trianglerighteq^+_{p(y)}$:							r	$\triangleright$	d
Value:									1
$\trianglerighteq^+_{q''_t}$:									q_t
Value:									1
$\trianglerighteq^+_{q_f}$:									P_X
Value:									1
$\trianglerighteq^+_{q_{r'}}$:									r
Value:									–
$\trianglerighteq^+_D$:									–

Table 10
$\trianglerighteq^-$ of the players in the proof of Theorem 31 for (**sfp**, **seo**).

Value:	-1						
$\trianglerighteq^-_d$:	$A \setminus \{d\}$						
Value:	$\cdots$		$-K_{r'}$		$-K_{r'} - 1$		
$\trianglerighteq^-_{r'}$:	$\cdots$	$\triangleright$	$A \setminus \{r\}$	$\triangleright$	D'		
Value:	$\cdots$		$-K_{q_f}$		$-K_{q_f} - 1$		
$\trianglerighteq^-_{q_f}$:	$\cdots$	$\triangleright$	$A \setminus P_X$	$\triangleright$	D'		
Value:	$\cdots$		$-K_{q''_t}$		$-K_{q_f} - 1$		
$\trianglerighteq^-_{q''_t}$:	$\cdots$	$\triangleright$	$P_X \cup P_Y \cup \{q'_t, q_f, r, r'\}$	$\triangleright$	D'		
Value:	$\cdots$		$-K_{p(y)}$		$-K_{p(y)} - 1$		
$\trianglerighteq^-_{p(y)}$:	$\cdots$	$\triangleright$	$\{\overline{y}, q_t, q'_t, q''_t, q_f, r'\}$	$\triangleright$	D'		
Value:	$\cdots$		$-K_{q'_t}$		$-K_{q'_t} - 1$		
$\trianglerighteq^-_{q'_t}$:	$\cdots$	$\triangleright$	$P_C \cup P_Y \cup \{q''_t, r, r'\}$	$\triangleright$	D'		
Value:	$\cdots$		$-K_{p(x)}$		$-K_{p(x)} - 1$		
$\trianglerighteq^-_{p(x)}$:	$\cdots$	$\triangleright$	$\{q''_t, r', \overline{x}\}$	$\triangleright$	D'		
Value:	$\cdots$		$-K_r$		$-K_r - 1$		
$\trianglerighteq^-_r$:	$\cdots$	$\triangleright$	$\{q_t, q'_t, q''_t, q_f\}$	$\triangleright$	D'		
Value:	$\cdots$		$-K_{q_t}$		$-K_{q_t} - 1$		
$\trianglerighteq^-_{q_t}$:	$\cdots$	$\triangleright$	$P_Y \cup \{q_f, r, r'\}$	$\triangleright$	D'		
Value:	-1		-2		$\cdots$		$-K_{p(c)}$
$\trianglerighteq^-_D$:	d	$\triangleright$	$\{p(\ell_1), p(\ell_2), p(\ell_3)\}$	$\triangleright$	$\cdots$	$\triangleright$	$\{q'_t, q_f, r'\}$

Construction 32. *Given a* 2-Quantified-3-DNF-SAT *instance* $(X, Y, \phi(X, Y))$*, we denote the set of clauses in* ϕ *by* C *and we define*

$$A = P_X \cup P_Y \cup P_C \cup \{q_t, q'_t, q''_t, q_f, r, r'\} \cup D$$

to be the set of players such that the following hold:

- *For every literal* ℓ *over* X*, we construct a corresponding* X*-player* $p(\ell)$ *(*$2n$ *in total). We denote this set with* P_X*.*
- *For every literal* ℓ *over* Y*, we construct a corresponding* Y*-player* $p(\ell)$ *(*$2n$ *in total). We denote this set with* P_Y*.*
- *For every clause* $c \in C$*, we construct a corresponding* C*-player* $p(c)$ *(*m *in total). We denote this set with* P_C*.*
- *We have six structure players:* q_t, q'_t, q''_t, q_f, r*, and* r'*.*
- *We have a set of padding players* D*, which we will use to generate the preferences providing the needed values.*

The number of padding players is bounded by $\mathcal{O}((n+m)(n^2+nm+m^2+1))$*.*

The scoring vector for the set of friends is fixed to $\mathbf{f}_i = \mathbf{sfp}$ *and we first construct* $\trianglerighteq^+$ *for the players in* A*. Note that we change the value of* ϵ *from* $1/n+1$ *to* 1 *and adjust the score the player* q'_t *assigns player* q_t *to* $n+1$ *(instead of* 1*). Table 9 shows* $\trianglerighteq^+$ *of the players in* A *and furthermore displays the values that are assigned based on the choice of* $\mathbf{f}_i = \mathbf{sfp}$ *and* $\mathbf{e}_i = \mathbf{seo}$*. Whenever set of players are given in a preference, say of player* p*, we assume that* p *is indifferent between the players in the set. Furthermore, if a single padding player* d *is given, she can be replaced by an arbitrarily picked player from* D*. Parts of the preferences that are denoted by "*$\cdots$*" have to be filled with an appropriate number of padding players from* D*.*

The set of neutral players is $A^0_d = P_C \cup P_X \cup P_Y \setminus \{p(\ell_1), p(\ell_2), p(\ell_3)\}$ *for each* $d \in D$*,* $A^0_{q''_t} = P_C$*,* $A^0_{p(y)} = P_C \cup P_X \cup (P_Y \setminus \{\overline{y}\})$*,* $A^0_{p(x)} = P_C \cup P_Y \cup \{q'_t\}$*, and* $A^0_p = \emptyset$ *for all remaining players* $p \in A$*.*

For each player $p \in A$ *assigning the symbolic value "*$-\infty$*" to some of her enemies in the original game, we define* K_p *to be the sum of all positive values* p *assigns to other players in* $A \setminus \{p\}$*. Table 10 shows the neutral sets and* $\trianglerighteq^-$ *of the players in* A*, where* D' *denotes those padding players not contained in* $\trianglerighteq^+$ *and not contained in* $\trianglerighteq^-$ *so far. This completes the construction of the Borda-induced FEN-hedonic game for* $\mathbf{f}_i = \mathbf{sfp}$ *and* $\mathbf{e}_i = \mathbf{seo}$*.*

For the scoring vectors $\mathbf{f}_i = \mathbf{sfp}$ *and* $\mathbf{e}_i = \mathbf{ep}$*, a similar approach can be used to achieve almost the same values as in the original construction. Only the preferences of the* C*-players have to be constructed carefully. These players are the only players assigning a different value than* $-\infty$ *to a subset of their enemies, namely the value* -2 *to those literal-players that are contained in the clause the clause-player corresponds to. With* $\mathbf{e}_i = \mathbf{ep}$ *we cannot achieve the assignment of value* -2*, but the assignment of value* -3 *by adding* 12 *padding players to the enemy set and due to this change, the players* r *and* q_t *in* $\trianglerighteq^+_{p(c)}$ *each have to gain one point more, so we have the adapted preferences shown in Table 11.*

The remaining padding players in A *that have not been assigned to* $\trianglerighteq^+_{p(c)}$ *or* $\trianglerighteq^-_{p(c)}$ *have to be in* $A^0_{p(c)}$*. This ensures that Woeginger's argumentation can be adapted straightforwardly.*

We will present the argumentation for $\mathbf{f}_i = \mathbf{sfp}$ *and* $\mathbf{e}_i = \mathbf{seo}$ *in detail. Consider the following coalition structure* Γ^* *that will be used throughout the rest of the argumentation. Let* $X = X_1 \cup X_2$ *be a partition of* X *into two sets such that for each* $x \in X_1$ *we have that* $\overline{x} \in X_2$*. Define:*

$$\Gamma^* = \Big\{\{q_f, \{p(x) \mid x \in X_1\}\}, \{p(y)\}_{y \in Y}, \{r, r'\}, \{q''_t\}, \{p(c)\}_{c \in C}, \{d\}_{d \in D}, \{q_t, \{p(x) \mid x \in X_2\}, q'_t\}\Big\}. \quad (6)$$

Table 12 shows the values each player assigns to her coalition in Γ^**.*

Table 11
$\trianglerighteq^+$ and $\trianglerighteq^-$ of the C-players in the proof of Theorem 31 for (**sfp**, **ep**).

Value:	-3				-16		-17		-18
$\trianglerighteq^-_{p(c)}$	$p(\ell_1) \sim p(\ell_2) \sim p(\ell_3)$	$\rhd$	$\cdots$	$\rhd$	q'_t	$\rhd$	q_f	$\rhd$	r'
Value:	8		7						
$\trianglerighteq^+_{p(c)}$:	q_t	$\rhd$	r	$\rhd$	$\cdots$				

Table 12
Values the players assign to their coalition in Γ^* for (**sfp**, **seo**).

q_f	$\{p(x)\|x \in X_1\}$	r	r'	$\{p(x)\|x \in X_2\}$	q_t	$q't$	P_Y, P_C, q''_t, D
n	$n+3$	$4n+2m-1$	1	n	$2n+m+1$	$2n+1$	0

Based on the constructed game, we will show Theorem 31 step by step, just as Woeginger did, and we start with the following lemma.

Lemma 33. *Let $(A, \succeq^{+0-})$ be a game constructed from a 2-*Quantified*-3-DNF-SAT instance $(X, Y, \phi(X, Y))$ as in Construction 32 and assume that Γ^* is a core stable coalition structure. Then the following hold for Γ^*.*

1. *Coalition $\Gamma^*(q_f)$ consists of q_f and n of the X-players. For each $x \in X$ either $p(x)$ or $p(\bar{x})$ is in $\Gamma^*(q_f)$.*
2. *Coalition $\Gamma^*(r)$ cannot consist of r together with n X-players, n Y-players, and all m C-players.*
3. *$\Gamma^*(r) = \{r, r'\}$.*
4. *$q''_t \notin \Gamma^*(q_t)$.*
5. *$q'_t \in \Gamma^*(q_t)$.*
6. *$\Gamma^*(q_t) = \{q_t, q'_t, \{p(x)|p(x) \notin \Gamma^*(q_f)\}\}$.*
7. *Γ^* yields a value of 0 for q''_t, all Y-players, and all C-players.*

Proof of Lemma 33. Claim 1 directly follows from Lemma 4.1 of Woeginger (2013b), except that for the X-players all coalitions not containing q_f yield fewer than $n+3$ points. The remaining argumentation remains unchanged.

Claims 2 and 3 can be shown by exactly the argumentation in the proofs of Lemmas 4.2 and 4.3 of Woeginger (2013b).

Claim 4 can be shown with a similar argumentation as presented in the proof of Lemma 4.4 of Woeginger (2013b): Assume that $q''_t \in \Gamma^*(q_t)$. That implies that $\Gamma^*(q_t) \subseteq \{q_t, q''_t\} \cup P_C$ and q_t assigns a value of at most $m+n+2$, q'_t assigns a value of at most n (because she is not in a coalition with q_t), and with Claims 1 and 3 we know that each $p(x)$ assigns a value of at most $n-1$. Now consider the coalition $\{q_t, \{p(x)|p(x) \notin \Gamma^*(q_f)\}, q'_t\}$ that ensures q_t a value of $m+2n+1$, q'_t a value of $2n+1$, and the X-players each a value of n and would thus be a blocking coalition.

Claims 5, 6, and 7 can be shown by exactly the same argumentation as in the proofs of Lemmas 4.5, 4.6, and 4.7 of Woeginger (2013b). □

Lemma 34. *Let $(A, \succeq^{+0-})$ be a game constructed from a 2-*Quantified*-3-DNF-SAT instance $(X, Y, \phi(X, Y))$ as in Construction 32. If there exists a core stable coalition structure Γ^* in this game, then $(X, Y, \phi(X, Y))$ is a yes instance of 2-*Quantified*-3-DNF-SAT.*

Proof of Lemma 34. This claim can be shown by exactly the same argumentation that Woeginger (2013b) provides in Section 4 of his paper. □

Lemma 35. *Let $(A, \succeq^{+0-})$ be a game constructed from a 2-*Quantified*-3-DNF-SAT instance $(X, Y, \phi(X, Y))$ as in Construction 32. If $(X, Y, \phi(X, Y))$ is a yes instance of 2-*Quantified*-3-DNF-SAT then a core stable coalition structure Γ exists in this game.*

Proof of Lemma 35. Assume that $(X, Y, \phi(X, Y))$ is a yes instance of 2-Quantified-3-DNF-SAT with the truth-assignment τ_X for the variables in X. Define a coalition structure Γ as the one in (6) and let $p(x) \in \Gamma(q_f)$ if and only if x is set to false.

For the sake of contradiction we assume that there is a coalition S^* that blocks the coalition structure Γ. With Lemmas 5.1, 5.2, and 5.3 of Woeginger (2013b) and some further argumentation he provides, we can show that

1. $\Gamma(q_f) \not\subseteq S^*$.
2. $r, r' \notin S^*$.
3. $q_t \notin S^*$.
4. For all $c \in C$, $p(c) \notin S^*$.
5. For all $y \in Y$, $p(y) \notin S^*$.
6. $q''_t \notin S^*$.

Furthermore, we have that $p(d) \notin S^*$ for all $d \in D$, which simply follows from the fact that being in a singleton coalition already maximizes the values of the players in D. Together with Claims 1 through 6 of Lemma 33, this implies that any possibly blocking coalition S^* is the empty set, so Γ is a core stable coalition structure. □

Now we can easily conclude the proof of Theorem 31: The claim follows immediately with Construction 32 and Lemmas 34 and 35. □

7. Conclusions and future work

We have studied FEN-hedonic games where players partition the other players into friends, enemies, and neutral players and rank their friends and their enemies. To extend the players' preferences over players to preferences over coalitions, we have used bipolar responsive extensions. Since pairs of coalitions may remain incomparable under these extensions, we have proposed comparability functions based on Borda-like scoring vectors in order to resolve these incomparabilities. Then we have analyzed the computational complexity of the existence and the verification problem of some well-known stability concepts for the induced hedonic games. Table 4 at the beginning of Section 6 gives an overview of our results. Some questions remain open for the existence problem: First, for strict core stability in Borda-induced FEN-hedonic games, we have a complexity gap between coNP-hardness and Σ_2^p membership; second, our NP-completeness results for individual and Nash stability as well as our Σ_2^p-completeness result for core stability hold only for certain combinations of comparability functions. Solving these open problems would be interesting tasks for future research.

It would also be interesting to study critical restrictions of the model that may lead to a drop in complexity. For example, our model allows ties in the players' preferences, and as we have seen for the related stable matching and stable roommates problems in the Introduction, the complexity of the existence and the verification problem for various stability concepts in Borda-induced

FEN-hedonic games might change when all players are required to present strict preferences only. Finally and more generally, as noted in Footnote 6, exploring the connection between responsiveness and additive separability of preferences in FEN-hedonic games is another challenging question for future research.

Acknowledgments

For many discussions and crucial insights, we are deeply indebted to Jérôme Lang and Anja Rey who co-authored the conference predecessor (Lang et al., 2015) of this paper that in particular introduces the model of FEN-hedonic game and the polarized responsive extension principle (presented here in Section 3.1) and that in its Section 5.2 also roughly sketches the notion of Borda-induced FEN-hedonic game that is studied here in depth. This work was supported in part by DFG, Germany grants RO-1202/14-1, RO-1202/14-2, and RO-1202/15-1.

Appendix. Deferred proofs from Section 3.2

Proof of Lemma 9. Let $(A, \succeq)$ be a hedonic game with n players and preferences induced by a profile $\trianglerighteq = (\trianglerighteq_1, \ldots, \trianglerighteq_n)$ of rankings over the players. The first and the third statement follow immediately from the definitions. For the second statement, consider consistency on pairs with $k = i$. Then, $j \in A_i^+ \cup A_i^0$ implies $j \trianglerighteq_i i$, due to the definition of $A_i^+ \cup A_i^0$. Consistency on pairs, together with strict a-b-toxicity gives us $\{i, j\} \succeq_i \{i\} \succ_i S$ for all $i \in A$, for all $j \in A_i^+ \cup A_i^0$, and for all $S \subseteq A$ with $\|S \cap (A_i^+ \cup A_i^0)\| = a$ and $\|S \cap A_i^-\| \geq b$. □

Proof of Lemma 11. For consistency on pairs, let us consider a FEN-hedonic game with three players, i, j, and k, and assume that $j, k \in A_i^+ \cup A_i^0$. If $j, k \in A_i^0$, then clearly $j \sim_i^{+0-} k$ and, by definition, we also have that $\{i, j\} \sim_i^{+0-} \{i, k\}$. If, without loss of generality, $k \in A_i^0$ and $j \in A_i^+$, it holds by definition of $\trianglerighteq_i^{+0-}$ that $j \vartriangleright_i^{+0-} k$, which in turn is equivalent to $\{i, j\} \succ_i^{+0-} \{i, k\}$. For $j, k \in A_i^+$, we know from Definition 3 that $\{i, j\} \succeq_i^{+0-} \{i, k\}$ is equivalent to the existence of an injective function $\sigma : \{k\} \cap A_i^+ \to \{j\} \cap A_i^+$, with $\sigma(k) \trianglerighteq_i^{+0-} k$. Thus, with σ mapping k to j, we have the desired equivalence. To prove strict 0-1-toxicity, let $S \subseteq A$ be an arbitrary coalition and let $i, j \in A$ be two players. Assuming that $\|S \cap (A_i^0 \cup A_i^+)\| = 0$ and $\|S \cap A_i^-\| \geq 1$ holds, we know that S contains at least one enemy of player i and neither friends of i nor any players who are neutral for i. By definition, player i would rather be alone than being part of coalition S, so $\{i\} \succ_i^{+0-} S$ indeed holds. □

References

Aziz, H., Biró, P., Lang, J., Lesca, J., Monnot, J., 2016a. Optimal reallocation under additive and ordinal preferences. In: Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems. IFAAMAS, pp. 402–410.

Aziz, H., Brandt, F., Harrenstein, P., 2013a. Pareto optimality in coalition formation. Games Econom. Behav. 82, 562–581.

Aziz, H., Brandt, F., Harrenstein, P., 2014. Fractional hedonic games. In: Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems. IFAAMAS, pp. 5–12.

Aziz, H., Brandt, F., Seedig, H., 2013b. Computing desirable partitions in additively separable hedonic games. Artificial Intelligence 195, 316–334.

Aziz, H., Gaspers, S., Mackenzie, S., Walsh, T., 2015. Fair assignment of indivisible objects under ordinal preferences. Artificial Intelligence 227, 71–92.

Aziz, H., Harrenstein, P., Lang, J., Wooldridge, M., 2016b. Boolean hedonic games. In: Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning. AAAI Press, pp. 166–175.

Aziz, H., Savani, R., 2016. Hedonic games. In: Brandt, F., Conitzer, V., Endriss, U., Lang, J., Procaccia, A. (Eds.), Handbook of Computational Social Choice. Cambridge University Press, pp. 356–376, chapter 15.

Ballester, C., 2004. NP-completeness in hedonic games. Games Econom. Behav. 49 (1), 1–30.

Banerjee, S., Konishi, H., Sönmez, T., 2001. Core in a simple coalition formation game. Soc. Choice Welf. 18, 135–153.

Baumeister, D., Bouveret, S., Lang, J., Nguyen, N., Nguyen, T., Rothe, J., Saffidine, A., 2017. Positional scoring-based allocation of indivisible goods. J. Auton. Agents Multi-Agent Syst. 31 (3), 628–655.

Baumeister, D., Faliszewski, P., Lang, J., Rothe, J., 2012. Campaigns for lazy voters: Truncated ballots. In: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems. IFAAMAS, pp. 577–584.

Bilò, V., Fanelli, A., Flammini, M., Monaco, G., Moscardelli, L., 2014. Nash stability in fractional hedonic games. In: Proceedings of the 10th International Workshop on Internet & Network Economics. In: Lecture Notes in Computer Science #8877, Springer-Verlag, pp. 486–491.

Bilò, V., Fanelli, A., Flammini, M., Monaco, G., Moscardelli, L., 2015. On the price of stability of fractional hedonic games. In: Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems. IFAAMAS, pp. 1239–1247.

Bogomolnaia, A., Jackson, M., 2002. The stability of hedonic coalition structures. Games Econom. Behav. 38 (2), 201–230.

Borda, J., 1781. Mémoire sur les élections au scrutin. Histoire de L'Académie Royale des Sciences, Paris, English translation appears in the paper by de Grazia (de Grazia, 1953).

Bossong, U., Schweigert, D., 2006. Minimal paths on ordered graphs. Math. Slovaca 56 (1), 23–31.

Bouveret, S., Endriss, U., Lang, J., 2010. Fair division under ordinal preferences: Computing envy-free allocations of indivisible goods. In: Proceedings of the 19th European Conference on Artificial Intelligence. IOS Press, pp. 387–392.

Brams, S., Edelman, P., Fishburn, P., 2003. Fair division of indivisible items. Theory and Decision 55 (2), 147–180.

Brams, S., King, D., 2005. Efficient fair division: Help the worst off or avoid envy? Ration. Soc. 17 (4), 387–421.

Burani, N., Zwicker, W., 2003. Coalition formation games with separable preferences. Math. Social Sci. 45 (1), 27–52.

Caragiannis, I., Procaccia, A., 2011. Voting almost maximizes social welfare despite limited communication. Artificial Intelligence 175 (9–10), 1655–1671.

Cechlárová, K., Hajduková, J., 2003. Computational complexity of stable partitions with B-preferences. Internat. J. Game Theory 31 (3), 353–364.

Cechlárová, K., Hajduková, J., 2004. Stable partitions with $\mathscr{W}$-preferences. Discrete Appl. Math. 138 (3), 333–347.

Cechlárová, K., Romero-Medina, A., 2001. Stability in coalition formation games. Internat. J. Game Theory 29 (4), 487–494.

Darmann, A., Elkind, E., Kurz, S., Lang, J., Schauer, J., Woeginger, G., 2012. Group activity selection problem. In: Proceedings of the 8th International Workshop on Internet & Network Economics. In: Lecture Notes in Computer Science #7695, Springer-Verlag, pp. 156–169.

Delort, C., Spanjaard, O., Weng, P., 2011. Committee selection with a weight constraint based on a pairwise dominance relation. In: Proceedings of the 2nd International Conference on Algorithmic Decision Theory. In: Lecture Notes in Artificial Intelligence #6992, Springer-Verlag, pp. 28–41.

Dimitrov, D., Borm, P., Hendrickx, R., Sung, S., 2006. Simple priorities and core stability in hedonic games. Soc. Choice Welf. 26 (2), 421–433.

Ehlers, L., Klaus, B., 2003. Coalitional strategy-proofness, resource-monotonicity, and separability for multiple assignment problems. Soc. Choice Welf. 21 (2), 265–280.

Elkind, E., Rothe, J., 2015. Cooperative game theory. In: Rothe, J. (Ed.), Economics and Computation. An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division. Springer-Verlag, pp. 135–196, chapter 3.

Elkind, E., Wooldridge, M., 2009. Hedonic coalition nets. In: Proceedings of the 8th International Joint Conference on Autonomous Agents and Multiagent Systems. IFAAMAS, pp. 417–424.

Gale, D., Shapley, L., 1962. College admissions and the stability of marriage. Amer. Math. Monthly 69 (1), 9–15.

Gale, D., Sotomayor, M., 1985. Some remarks on the stable matching problem. Discrete Appl. Math. 11 (3), 223–232.

Garey, M., Johnson, D., 1979. Computers and Intractibility: A Guide to the Theory of NP-Completeness. W. H. Freeman and Company.

de Grazia, A., 1953. Mathematical deviation of an election system. Isis 44 (1–2), 41–51.

Hatfield, J., 2009. Strategy-proof, efficient, and nonbossy quota allocations. Soc. Choice Welf. 33 (3), 505–515.

Irving, R., 1985. An efficient algorithm for the stable roommates problem. J. Algorithms 6 (4), 577–595.

Irving, R., 1994. Stable marriage and indifference. Discrete Appl. Math. 48 (3), 261–272.

Irving, R., Manlove, D., 2002. The stable roommates problem with ties. J. Algorithms 43 (1), 85–105.

Irving, R., Manlove, D., O'Malley, G., 2009. Stable marriage with ties and bounded length preference lists. J. Discrete Algorithms 7 (2), 213–219.

Kuckuck, B., Rothe, J., 2018. Sequential allocation rules are separable: Refuting a conjecture on scoring-based allocation of indivisible goods. In: Proceedings

of the 17th International Conference on Autonomous Agents and Multiagent Systems. IFAAMAS, pp. 650–658.
Lang, J., Rey, A., Rothe, J., Schadrack, H., Schend, L., 2015. Representing and solving hedonic games with ordinal preferences and thresholds. In: Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems. IFAAMAS, pp. 1229–1237.
Manlove, D., Irving, R., Iwama, K., Miyazaki, S., Morita, Y., 2002. Hard variants of stable marriage. Theoret. Comput. Sci. 276 (1–2), 261–279.
Meyer, A., Stockmeyer, L., 1972. The equivalence problem for regular expressions with squaring requires exponential space. In: Proceedings of the 13th IEEE Symposium on Switching and Automata Theory. IEEE Computer Society Press, pp. 125–129.
Nguyen, N., Baumeister, D., Rothe, J., 2018. Strategy-proofness of scoring allocation correspondences for indivisible goods. Soc. Choice Welf. 50 (1), 101–122.
Nguyen, N., Nguyen, T., Rothe, J., 2017. Approximate solutions to max–min fair and proportionally fair allocations of indivisible goods. In: Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems. IFAAMAS, pp. 262–271.
Nguyen, N., Rey, A., Rey, L., Rothe, J., Schend, L., 2016. Altruistic hedonic games. In: Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems. IFAAMAS, pp. 251–259, Also presented at the 7th International Workshop on Cooperative Games in Multiagent Systems (CoopMAS 2016), co-located with AAMAS 2016, and at the 6th International Workshop on Computational Social Choice (COMSOC 2016), Toulouse, France, June 2016, both with nonarchival proceedings.
Peters, D., 2016. Complexity of hedonic games with dichotomous preferences. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence. AAAI Press, pp. 579–585.
Peters, D., Elkind, E., 2015. Simple causes of complexity in hedonic games. In: Proceedings of the 24th International Joint Conference on Artificial Intelligence. AAAI Press/IJCAI, pp. 617–623.
Rey, A., Rothe, J., Schadrack, H., Schend, L., 2016. Toward the complexity of the existence of wonderfully stable partitions and strictly core stable coalition structures in enemy-oriented hedonic games. Ann. Math. Artif. Intell. 77 (3), 317–333.
Ronn, E., 1990. NP-complete stable matching problems. J. Algorithms 11 (2), 285–304.
Roth, A., 1985. Common and conflicting interests in two-sided matching markets. Eur. Econ. Rev. 27 (1), 75–96.
Roth, A., Sotomayor, M., 1992. Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis. Cambridge University Press.
Stockmeyer, L., 1976. The polynomial-time hierarchy. Theoret. Comput. Sci. 3 (1), 1–22.
Sung, S., Dimitrov, D., 2007. On core membership testing for hedonic coalition formation games. Oper. Res. Lett. 35 (2), 155–158.
Sung, S., Dimitrov, D., 2010. Computational complexity in additive hedonic games. European J. Oper. Res. 203 (3), 635–639.
Woeginger, G., 2012. Core stability in hedonic coalition formation. Technical Report arXiv:1212.2236v1 [cs.GT], ACM Computing Research Repository (CoRR).
Woeginger, G., 2013a. Core stability in hedonic coalition formation. In: Proceedings of the 39th Conference on Current Trends in Theory and Practice of Computer Science. In: Lecture Notes in Computer Science #7741, Springer-Verlag, pp. 33–50.
Woeginger, G., 2013b. A hardness result for core stability in additive hedonic games. Math. Social Sci. 65 (2), 101–104.

CHAPTER 6

Verification in Incomplete Argumentation Frameworks

Summary

To answer the question of what parts of a discussion are to be accepted or are the 'good' arguments of the discussion, one fast encounters the question, whether the abstract models that have been used to express real world discussions are close enough to produce a good outcome. Most debates of the real world are highly complex and it is not easily motivatable why a simple model as the model of argumentation frameworks by Dung [29] is enough. Many authors already discussed specialized or extended models of these simple models, but not many tackled the question of what can be done when assuming the agents to have incomplete knowledge. In the following paper, we extend those existing models regarding incomplete knowledge and define a general version that allows for a lack of knowledge in both, the set of arguments and the set of attacks. The model is, for example, motivated through the common approach of merging different believes into one, and then extracting an aggregated result. We also answer the question on how to extend the ideas of extensions as outcome of an argumentation framework to fit to the new model of incomplete argumentation frameworks. Additionally, we define a possible and necessary version of the verification problem, which was originally defined for Dung's argumentation frameworks, for each of the semantics conflict-freeness, admissibility, completeness, preferredness, groundedness, and stability, and analyze these in terms of their computational complexity.

Contribution and Preceding Versions

The idea, model, and writing was done jointly with my coauthors, as well as Theorems 26 and 28, and Corollaries 27, 29, 42 and 50. Additionally, Theorems 43, 44, Corollary 45, and not published versions of the proofs of Theorems 46 and 48 have to be attributed to my contribution. This paper merges and extends the preliminary papers [8], [9], [11], and [12].

Publication - Baumeister, Neugebauer, Rothe, and Schadrack [13]

D. Baumeister, D. Neugebauer, J. Rothe, and H. Schadrack. Verification in incomplete argumentation frameworks. *Artificial Intelligence*, 264:1–26, 2018

Artificial Intelligence 264 (2018) 1–26

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Verification in incomplete argumentation frameworks ☆

Dorothea Baumeister, Daniel Neugebauer, Jörg Rothe *, Hilmar Schadrack

Institut für Informatik, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany

ARTICLE INFO

Article history:
Received 24 November 2017
Received in revised form 31 July 2018
Accepted 8 August 2018
Available online 10 August 2018

Keywords:
Abstract argumentation
Argumentation framework
Incomplete knowledge
Verification
Computational complexity

ABSTRACT

We tackle the problem of expressing incomplete knowledge in abstract argumentation frameworks originally introduced by Dung [26]. In applications, incomplete argumentation frameworks may arise as intermediate states in an elicitation process, or when merging different beliefs about an argumentation framework's state, or in cases where complete information cannot be obtained. We consider two specific models of incomplete argumentation frameworks, one focusing on attack incompleteness and the other on argument incompleteness, and we also provide a general model of incomplete argumentation framework that subsumes both specific models. In these three models, we study the computational complexity of variants of the verification problem with respect to six common semantics of argumentation frameworks: the conflict-free, admissible, stable, complete, grounded, and preferred semantics. We provide a full complexity map covering all three models and these six semantics. Our main result shows that the complexity of verifying the preferred semantics rises from coNP- to Σ_2^p-completeness when allowing uncertainty about either attacks or arguments, or both.

1. Introduction

Abstract argumentation frameworks are a simple, yet powerful tool for nonmonotonic reasoning that were originally introduced by Dung [26]. In this model, individual arguments are considered to be abstract entities, disregarding their internal structure and focusing only on the attack relation between them. Various semantics defined by Dung and others allow to investigate the acceptability status of sets of arguments based on the attack relation. However, abstract argumentation frameworks are suitable to describe an argumentation's state only in an optimal situation—they require that *all* relevant arguments are included and that there is no uncertainty regarding the attacks between them. If these conditions are not met, the existing methods for semantic analysis cannot be applied.

To capture uncertainty in various real-world settings like intermediate states of an evolving argumentation, partial-information settings (and, in particular, permanently unavailable information), and the task of merging individual (subjec-

☆ This paper merges and extends preliminary versions presented at the *32nd AAAI Conference on Artificial Intelligence* (AAAI'18, [8]), at the *4th International Conference on Algorithmic Decision Theory* (ADT'15, [6,11]), and at the *6th* and the *7th International Workshop on Computational Social Choice* (COMSOC'16 and COMSOC'18), both with nonarchival proceedings. Extending these preliminary conference versions, this paper describes the model of incomplete argumentation framework in more detail, contains all proofs, establishes links between the previous versions, unifies notation and all results, and provides more examples, discussion, and motivation.

* Corresponding author.
E-mail address: rothe@cs.uni-duesseldorf.de (J. Rothe).
URLs: http://ccc.cs.uni-duesseldorf.de/~baumeister (D. Baumeister), http://ccc.cs.uni-duesseldorf.de/~neugebauer (D. Neugebauer), http://ccc.cs.uni-duesseldorf.de/~rothe (J. Rothe), http://ccc.cs.uni-duesseldorf.de/~schadrack (H. Schadrack).

https://doi.org/10.1016/j.artint.2018.08.001
0004-3702/

tive) views on an argumentation, we introduce two specific models formalizing attack- and argument-incomplete argumentation frameworks, and we then combine them to express simultaneous uncertainty about attacks and arguments. Our objective in each model is to analyze how the computational complexity of certain variants of the verification problem (to be formally defined in Section 2) is affected by introducing uncertainty.

Why do we study variants of the verification problem and their complexity? Well, when one encounters an interesting semantic property, the first question that arises is: Can it be verified? And the second: How hard is it to verify it? In particular, since we extend the standard model of abstract argumentation by allowing uncertainty about the attacks and/or the arguments, it is natural to wonder whether the complexity of the related more general problem variants significantly increases. Our results show that in many cases the complexity of verification does not rise; however, we also pinpoint cases where it does (see Table 1 in Section 5 for an overview). For a bigger picture, we will compare the verification complexity with that of other computational tasks, namely checking credulous and skeptical acceptance of arguments in incomplete argumentation frameworks in Section 5.

The standard verification problem is defined in Dung's original model of argumentation framework, so we first need to adjust it to our extended models. A natural way to adapt a decision problem in the face of incomplete knowledge is to ask whether the answer is *possibly* (respectively, *necessarily*) "yes"—i.e., given all possible completions of the incomplete state, to ask whether *at least one* such completion (respectively, whether *all* these completions) are yes-instances of the original problem. This approach has already been taken in various areas of computational social choice: in voting by, e.g., Konczak and Lang [34], Xia and Conitzer [49], Chevaleyre et al. [18], and Baumeister et al. [9,10]; in fair division by Bouveret et al. [14] and Baumeister et al. [4]; in algorithmic game theory by Lang et al. [35]; and in judgment aggregation by Baumeister et al. [5]. However, this approach is new to argumentation theory: In two of this paper's predecessors, Baumeister et al. [6,11] were the first to define and study possible and necessary verification for certain semantics in incomplete argumentation frameworks, and they continued this line of research in their recent work [8]. The present paper merges and extends these preliminary versions. A general overview on the use of abstract argumentation in artificial intelligence is given by Rahwan and Simari [41] and Bench-Capon and Dunne [12].

In related work, incomplete knowledge about the attack relation has first been introduced by Coste-Marquis et al. [19] and has been analyzed with respect to argument acceptability by Cayrol et al. [16]. Unlike us, however, they develop new semantics for attack-incomplete argumentation frameworks and thus put a lot of focus on the incomplete framework itself, rather than on its completions. Other work on incomplete knowledge in abstract argumentation includes *probabilistic argumentation frameworks* (see, for example, the work of Li et al. [36], Rienstra [42], Fazzinga et al. [31,30], Hunter [33], and Doder and Woltran [25]) where arguments and/or attacks have an associated probability as a quantified notion of uncertainty.

A related concept to incomplete knowledge is that of dynamic change. Cayrol et al. [17] study how the addition or deletion of one single argument or several arguments, together with a respective change in the attack relation, can change the set of extensions of an argumentation framework. Liao et al. [37] investigate the complexity of computing the status of an argument (i.e., whether it is accepted, rejected, or undecided) upon changing the arguments and attacks. Coste-Marquis et al. [21] study how belief revision postulates can be applied to argumentation systems. Boella et al. [13] address the question of which arguments or attacks can be removed without changing the set of extensions. Another dynamic setting is that of merging or aggregating different argumentation frameworks. Coste-Marquis et al. [19] study incomplete argumentation frameworks as a possible result of merging individual views. Tohmé et al. [46] discuss criteria for methods that aggregate several attack relations into a single attack relation (without uncertainty). Delobelle et al. [23] study merging operators for abstract argumentation frameworks axiomatically. Maher [38] studies resistance to corruption in strategic argumentation. While instances in his model and in our argument-incomplete argumentation frameworks are technically similar, his results do not carry over to our problems. One difference is that he focuses on credulous or skeptical acceptance of specific arguments, whereas we consider verification of entire extensions.

Extension enforcement as defined by Baumann and Brewka [3,2] has some connections to our work; for example, expansions can be viewed as making an argumentation framework argument- and attack-incomplete. On the other hand, extension enforcement in the argument-fixed variant due to Coste-Marquis et al. [22] is obviously related to attack incompleteness. Wallner et al. [48] studied extension enforcement from an algorithmic point of view and provided algorithms and complexity results, just as we do here. However, it is clear that our models and results differ from these works.

This paper is structured as follows. In Section 2, we provide the formal model of standard argumentation framework. Sections 3.1 and 3.2 introduce, respectively, the attack-incomplete and argument-incomplete model extensions, which are then combined into a universal incompleteness model in Section 3.3. We provide a full study of the computational complexity of the possible and necessary variants of the verification problem for Dung's original semantics in Section 4, divided into upper bounds in Section 4.1 and lower bounds in Section 4.2. In Section 5, we summarize our results and point out some interesting tasks that could be tackled in future work.

2. Preliminaries

In this section, we give formalizations of the basic notions of abstract argumentation. While we adapt some notation from the book chapter by Dunne and Wooldridge [29], the underlying concepts are due to Dung [26].

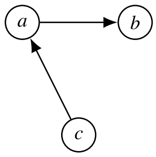


Fig. 1. A simple argumentation framework.

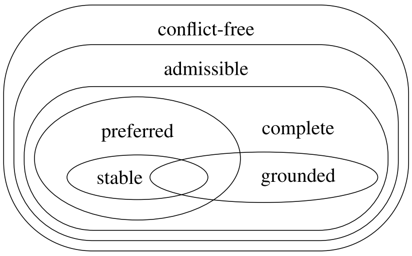


Fig. 2. Relations among various semantics for sets of arguments.

Definition 1. An *argumentation framework AF* is a pair $\langle \mathscr{A}, \mathscr{R} \rangle$, where $\mathscr{A}$ is a finite set of *arguments*, and $\mathscr{R} \subseteq \mathscr{A} \times \mathscr{A}$ is a binary relation. We say that *a attacks b* if $(a,b) \in \mathscr{R}$.

We will use the common representation of argumentation frameworks by graphs: Every argumentation framework $AF = \langle \mathscr{A}, \mathscr{R} \rangle$ can be seen as a directed graph $G_{AF} = (V, E)$ by identifying arguments with vertices and attacks with directed edges, i.e., $V = \mathscr{A}$ and $E = \mathscr{R}$.

Example 2. Fig. 1 displays the graph representation of the argumentation framework $AF = \langle \mathscr{A}, \mathscr{R} \rangle$ with $\mathscr{A} = \{a, b, c\}$ and $\mathscr{R} = \{(a,b), (c,a)\}$. It will be used—and extended along the way—as a running example throughout the paper.

In the literature, many *semantics* have been defined which allow to evaluate the acceptability status of sets of arguments. We use the semantics introduced by Dung [26] in his seminal paper:

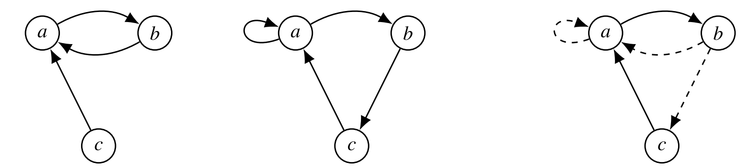
Definition 3. Let $AF = \langle \mathscr{A}, \mathscr{R} \rangle$ be an argumentation framework. A set $S \subseteq \mathscr{A}$ is

- *conflict-free* if there are no $a, b \in S$ such that $(a,b) \in \mathscr{R}$,
- *admissible* if S is conflict-free and every $a \in S$ is acceptable with respect to S, where an argument $a \in \mathscr{A}$ is *acceptable with respect to* $S \subseteq \mathscr{A}$ if, for each $b \in \mathscr{A}$ with $(b,a) \in \mathscr{R}$, there is a $c \in S$ such that $(c,b) \in \mathscr{R}$ (if an argument $a \in \mathscr{A}$ is acceptable with respect to a set $S \subseteq \mathscr{A}$, we may also say that S *defends* a),
- *preferred* if S is a maximal (with respect to set inclusion) admissible set,
- *stable* if S is conflict-free and for every $b \in \mathscr{A} \setminus S$ there is an $a \in S$ with $(a,b) \in \mathscr{R}$,
- *complete* if S is admissible and contains all $a \in \mathscr{A}$ that are acceptable with respect to S, and
- *grounded* if S is the least (with respect to set inclusion) fixed point of the characteristic function of *AF*, where the *characteristic function* $F_{AF} : 2^{\mathscr{A}} \to 2^{\mathscr{A}}$ of *AF* is defined by

$$F_{AF}(S) = \{a \in \mathscr{A} \mid a \text{ is acceptable with respect to } S\}.$$

The characteristic function always has a least fixed point, since it is monotonic with respect to set inclusion, so the existence of the (unique) grounded set is guaranteed. The complete sets of an argumentation framework can be characterized as the fixed points of F_{AF}—in particular, the grounded set is complete. Dung [26] also proved several other correlations between his semantics. In particular, he showed that every admissible set is a subset of a preferred set, and that there always is at least one preferred set (which may be the empty set). Also, every stable set is preferred, and every preferred set is complete. It is easy to find examples that a preferred or grounded set does not have to be stable, and it is easy to show that each of the above defined semantics entails conflict-freeness and admissibility. Fig. 2 displays all relations among the various semantics that we use. If an area labeled with semantics **s** is fully included in an area labeled with semantics **s**′, this indicates that in all argumentation frameworks all sets of arguments that fulfill **s** also fulfill **s**′. The converse is not necessarily true, i.e., all displayed set inclusions are strict. Further, none of the areas are disjoint, so one and the same set of arguments might fulfill all semantics simultaneously.

Dung [26] uses the notion of *extensions* of an argumentation framework as a term for those subsets that fulfill the criteria of a given semantics. For example, a set of arguments is called a *preferred extension of the argumentation framework*



(a) Agent 1's (left) and agent 2's (right) individual views

(b) Merging the views of agents 1 and 2 (dashed attacks are uncertain, i.e., in $\mathscr{R}^?$)

Fig. 3. Attack incompleteness.

if it is a preferred set of the given argumentation framework. Dung considers conflict-freeness and admissibility to be basic requirements rather than semantics, and therefore did not call conflict-free or admissible sets "extensions"—for convenience, however, we will do so sometimes.

We also need some of the basic notions from complexity theory. We assume the reader to be familiar with the complexity classes P, NP, and coNP, as well as hardness, completeness, polynomial-time-reducibility, $\le_m^p$, and (oracle) Turing machines. Problems that are solvable by a nondeterministic oracle Turing machine with access to an NP oracle belong to $\Sigma_2^p = \mathrm{NP}^{\mathrm{NP}}$; this class constitutes, together with $\Pi_2^p = \mathrm{coNP}^{\mathrm{NP}}$, the second level of the polynomial hierarchy, and was introduced by Meyer and Stockmeyer [39,44]. It is known that $\mathrm{P} \subseteq \mathrm{NP} \subseteq \Sigma_2^p$, but it is still unknown whether any of these inclusions is strict. For further details, see, e.g., the books by Papadimitriou [40] and Rothe [43].

Dunne and Wooldridge [29] surveyed several decision problems defined for argumentation frameworks, many of which are hard to decide, as they are complete for NP, coNP, or even Π_2^p. Here, we will focus on the verification problem, which—as shown by Dimopoulos and Torres [24]—is coNP-complete for the preferred semantics, but can be decided in polynomial time for all other semantics defined above, which follows immediately from the work of Dung [26]. This problem is defined as follows:

	s-Verification
Given:	An argumentation framework $\langle \mathscr{A}, \mathscr{R} \rangle$ and a subset $S \subseteq \mathscr{A}$.
Question:	Is S an **s** extension of $\langle \mathscr{A}, \mathscr{R} \rangle$?

In our notation, the boldfaced letter **s** is a placeholder for any of the six semantics defined earlier. For better readability, we will sometimes shorten their names and write cf for *conflict-freeness*, ad for *admissibility*, pr for *preferredness*, st for *stability*, cp for *completeness*, and gr for *groundedness*.

3. Three models of incomplete argumentation framework

In this section, we introduce three different notions of incompleteness for argumentation frameworks. We start with attack incompleteness in Section 3.1, followed by argument incompleteness in Section 3.2. In Section 3.3, both approaches are combined to provide a general model of incompleteness in argumentation frameworks.

3.1. Attack incompleteness

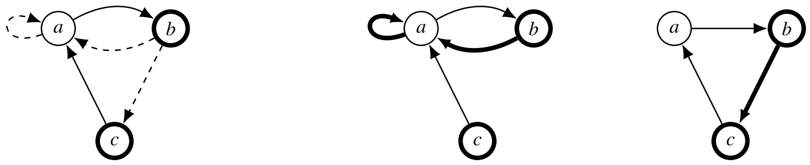
The first notion of incompleteness we consider concerns the attack relation between arguments. While Dung's original model only allows to express whether an attack (a,b) exists ($(a,b) \in \mathscr{R}$) or doesn't exist ($(a,b) \notin \mathscr{R}$), the extended model also allows to explicitly express lack of information about an attack. *Attack-incomplete argumentation frameworks* were originally proposed by Coste-Marquis et al. [19]—we employ their model, but use a slightly modified notation.

Definition 4. An *attack-incomplete argumentation framework* is a triple $\langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$, where $\mathscr{A}$ is a nonempty set of arguments and $\mathscr{R}$ and $\mathscr{R}^?$ are disjoint subsets of $\mathscr{A} \times \mathscr{A}$. $\mathscr{R}$ denotes the set of all ordered pairs of arguments between which an attack is known to definitely exist, while $\mathscr{R}^?$ contains all possible additional attacks not (yet) known to exist. The set of attacks that are known to never exist is denoted by $\mathscr{R}^- = (\mathscr{A} \times \mathscr{A}) \setminus (\mathscr{R} \cup \mathscr{R}^?)$.

Example 5. Extending the argumentation framework from Example 2 by three possible attacks,

$$\mathscr{R}^? = \{(a,a), (b,a), (b,c)\},$$

yields the attack-incomplete argumentation framework $\langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$ the graph representation of which is given in Fig. 3(b). This incomplete framework might be the result of merging several *individual (subjective)* views that share a common set of arguments but may have different attacks. Fig. 3(a) shows two such individual argumentation frameworks, which are merged into the attack-incomplete argumentation framework of Fig. 3(b) by including those attacks that exist in all individual views as definite attacks ($\mathscr{R}$), and including attacks that exist in some but not all individual views as possible attacks ($\mathscr{R}^?$).



(a) The attack-incomplete argumentation framework from Figure 3(b) with the set $S = \{b, c\}$ highlighted

(b) Optimistic completion (left) and pessimistic completion (right) for the set $S = \{b, c\}$ in the attack-incomplete argumentation framework from Figure 4(a)

Fig. 4. Optimistic completion and pessimistic completion.

In an attack-incomplete argumentation framework $\langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$, for each possible but as yet unknown attack in $\mathscr{R}^?$, when deciding whether or not the attack will be included, one obtains a standard argumentation framework that can be seen as a *completion* of $\langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$.

Definition 6. Let $AtIAF = \langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$ be a given attack-incomplete argumentation framework. An argumentation framework $AtIAF^* = \langle \mathscr{A}, \mathscr{R}^* \rangle$ with $\mathscr{R} \subseteq \mathscr{R}^* \subseteq \mathscr{R} \cup \mathscr{R}^?$ is called a *completion of AtIAF*.

The number of possible completions for a given attack-incomplete argumentation framework is clearly $2^{|\mathscr{R}^?|}$. For $\mathscr{R}^? = \emptyset$, there is no uncertainty and only one completion exists, which coincides with the attack-incomplete framework itself. In general, however, the number of completions may be exponential in relation to the instance's size.

In an attack-incomplete argumentation framework *AtIAF*, we say that a property defined for standard argumentation frameworks (e.g., a semantics) holds *possibly* if there exists a completion $AtIAF^*$ of *AtIAF* for which the property holds, and a property holds *necessarily* if it holds for all completions of *AtIAF*. Accordingly, we can define two variants of the verification problem in the attack-incomplete case for each given semantics **s**:

	s-ATT-INC-POSSIBLE-VERIFICATION (**s**-ATTINCPV)
Given:	An attack-incomplete argumentation framework $AtIAF = \langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$ and a set $S \subseteq \mathscr{A}$.
Question:	Is there a completion $AtIAF^*$ of *AtIAF* such that S is an **s** extension of $AtIAF^*$?

	s-ATT-INC-NECESSARY-VERIFICATION (**s**-ATTINCNV)
Given:	An attack-incomplete argumentation framework $AtIAF = \langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$ and a set $S \subseteq \mathscr{A}$.
Question:	For all completions $AtIAF^*$ of *AtIAF*, is S an **s** extension of $AtIAF^*$?

Both problems are potentially harder than standard verification, since they add an existential (respectively, universal) quantifier over a potentially exponential space of solutions. In Section 4.1, however, we prove that, for all cases except possible verification using the preferred semantics, their complexity in fact does not increase.

3.1.1. Optimistic and pessimistic completions

In the remainder of this section, we provide efficient algorithms that, given an attack-incomplete argumentation framework AF and a set S of arguments in AF, create a single *critical* completion—in that completion, S is most likely (or, most unlikely) among all possible completions to be an extension for some given semantics. In Section 4.1, we will use some of these critical completions to prove P membership of possible verification for the associated semantics.

We start with the *optimistic* and the *pessimistic* completion, which are critical for conflict-freeness, admissibility, and the stable semantics. These completions simply exclude (respectively, include) all possible attacks against S and include (respectively, exclude) the remaining possible attacks.

Definition 7. Let $AtIAF = \langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$ be an attack-incomplete argumentation framework and let $S \subseteq \mathscr{A}$.

- The *optimistic completion of AtIAF for S* is $AtIAF_S^{opt} = \langle \mathscr{A}, \mathscr{R}_S^{opt} \rangle$ with $\mathscr{R}_S^{opt} = \mathscr{R} \cup \{(a,b) \in \mathscr{R}^? \mid b \notin S\}$.
- The *pessimistic completion of AtIAF for S* is $AtIAF_S^{pes} = \langle \mathscr{A}, \mathscr{R}_S^{pes} \rangle$ with $\mathscr{R}_S^{pes} = \mathscr{R} \cup \{(a,b) \in \mathscr{R}^? \mid b \in S\}$.

Example 8. Consider again the attack-incomplete argumentation framework from Fig. 3(b); Fig. 4(a) shows it with the arguments from the set $S = \{b, c\}$ highlighted by boldfaced circles. Its optimistic and pessimistic completions for S are given in Fig. 4(b). The possible attacks added to the set of attacks in the optimistic (respectively, pessimistic) completion are displayed as boldfaced arcs.

Propositions 9 and 10 establish that the optimistic and pessimistic completions are indeed critical for the given properties.

Proposition 9. *Let $AtIAF = \langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$ be an attack-incomplete argumentation framework, let $S \subseteq \mathscr{A}$, and let $AtIAF_S^{opt}$ be the optimistic completion of AtIAF for S.*

1. *S is possibly conflict-free in AtIAF if and only if S is a conflict-free extension of $AtIAF_S^{opt}$.*
2. *If S is conflict-free in $AtIAF_S^{opt}$, then $a \in S$ is possibly acceptable with respect to S in AtIAF if and only if a is acceptable with respect to S in $AtIAF_S^{opt}$.*
3. *S is possibly admissible in AtIAF if and only if S is an admissible extension of $AtIAF_S^{opt}$.*
4. *S is possibly stable in AtIAF if and only if S is a stable extension of $AtIAF_S^{opt}$.*

Proof. The converse is trivial in all cases: If S fulfills a given criterion in $AtIAF_S^{opt}$, this immediately yields that S possibly fulfills the criterion in *AtIAF*. We now prove the other direction of the equivalence individually for each criterion:

1. If a set S of arguments is not conflict-free in $AtIAF_S^{opt}$, then there must be an attack between elements of S in $\mathscr{R}_S^{opt}$, which must be already in $\mathscr{R}$ due to how $\mathscr{R}_S^{opt}$ is constructed, and which therefore exists in every completion of *AtIAF*. Thus S is not a possibly conflict-free set in *AtIAF*.
2. Assume that S is conflict-free in $AtIAF_S^{opt}$. Then, if there is some $a \in S$ that is not acceptable with respect to S in $AtIAF_S^{opt}$, it must be attacked by some $b \in \mathscr{A}$ in $\mathscr{R}_S^{opt}$ and there is no attack from an element of S against b in $\mathscr{R}_S^{opt}$. By construction, $\mathscr{R}_S^{opt}$ does not contain any possible attacks (members of $\mathscr{R}^?$) that attack elements of S, and it contains all possible attacks that can defend S, i.e., that target attackers of S. Therefore, all attacks in $\mathscr{R}_S^{opt}$ against elements of S are already in $\mathscr{R}$, so the undefended attack from b against a is in every completion of *AtIAF*. Since a cannot be acceptable with respect to S in any completion of *AtIAF*, a is not possibly acceptable with respect to S in *AtIAF*.
3. Assume that S is not an admissible extension in $AtIAF_S^{opt}$, i.e., S is not conflict-free in $AtIAF_S^{opt}$ or there is some $a \in S$ that is not acceptable with respect to S in $AtIAF_S^{opt}$. In either case, the previous results imply that S is not conflict-free in any completion of *AtIAF* or a is not acceptable with respect to S in any completion of *AtIAF*. Thus S is not a possibly admissible extension in *AtIAF*.
4. If a set S of arguments is not stable in $AtIAF_S^{opt}$, S is necessarily not conflict-free in *AtIAF* or there is an $a \in \mathscr{A} \setminus S$ that is not attacked by S in $AtIAF_S^{opt}$, and therefore—by construction of $AtIAF_S^{opt}$—a cannot be attacked by S in any completion of *AtIAF*. In both cases, there is no completion of *AtIAF* for which S is stable, so S is not a possibly stable extension of *AtIAF*.

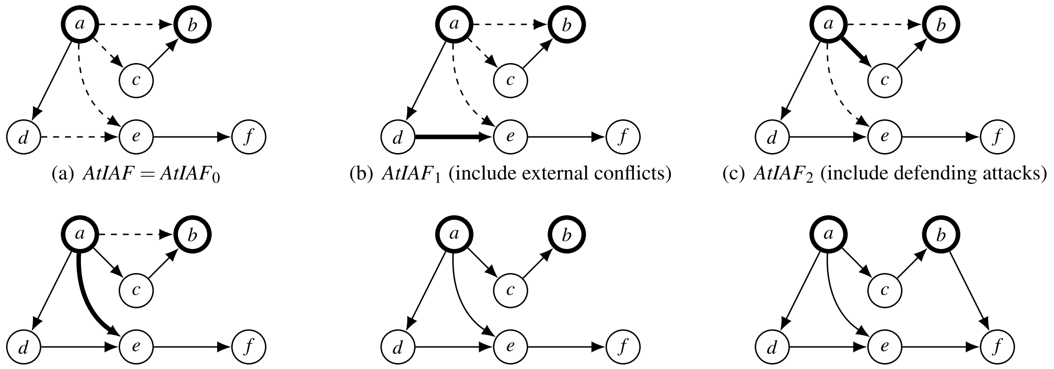
This completes the proof. □

Proposition 10. *Let $AtIAF = \langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$ be an attack-incomplete argumentation framework, $S \subseteq \mathscr{A}$, and let $AtIAF_S^{pes}$ be the pessimistic completion of AtIAF for S.*

1. *S is necessarily conflict-free in AtIAF if and only if S is a conflict-free extension of $AtIAF_S^{pes}$.*
2. *If S is conflict-free in $AtIAF_S^{pes}$, then $a \in S$ is necessarily acceptable with respect to S in AtIAF if and only if a is acceptable with respect to S in $AtIAF_S^{pes}$.*
3. *S is necessarily admissible in AtIAF if and only if S is an admissible extension of $AtIAF_S^{pes}$.*
4. *S is necessarily stable in AtIAF if and only if S is a stable extension of $AtIAF_S^{pes}$.*

Proof. Here, the left-to-right implications are trivial: If S necessarily fulfills a criterion in *AtIAF*, it must fulfill it in particular in the pessimistic completion. We prove the other direction of the implications individually:

1. If S is conflict-free in $AtIAF_S^{pes}$, then all interior attacks among elements of S are in $\mathscr{R}^-$, because if such an attack were in $\mathscr{R}$, S would not be conflict-free in any completion of *AtIAF*, and if such an attack were in $\mathscr{R}^?$, it would have been included in $\mathscr{R}_S^{pes}$, which contradicts our assumption that S is conflict-free in $AtIAF_S^{pes}$. Since all interior attacks among elements of S are in $\mathscr{R}^-$, S is necessarily conflict-free in *AtIAF*.
2. Assume that S is conflict-free in $AtIAF_S^{pes}$. Then, if each $a \in S$ is acceptable with respect to S in $AtIAF_S^{pes}$, this means that S defends each of these arguments against all their attackers. By construction, $\mathscr{R}_S^{pes}$ contains all possible attacks from $\mathscr{R}^?$ that attack elements of S, and no possible attacks that can defend S. Therefore, all attacks in $\mathscr{R}_S^{pes}$ that defend elements of S against possible or definite attacks are already in $\mathscr{R}$, otherwise they could not be in $\mathscr{R}_S^{pes}$, and are therefore in $\mathscr{R}^*$ for any completion $AtIAF^*$. This implies that each $a \in S$ is necessarily acceptable with respect to S in *AtIAF*.
3. Assume that S is an admissible extension of $AtIAF_S^{pes}$, i.e., S is conflict-free in $AtIAF_S^{pes}$ and each $a \in S$ is acceptable with respect to S in $AtIAF_S^{pes}$. The previous results then imply that S is necessarily conflict-free in *AtIAF* and each $a \in S$ is necessarily acceptable with respect to S in *AtIAF*, which immediately yields that S is necessarily admissible in *AtIAF*.



(a) $AtIAF = AtIAF_0$
(b) $AtIAF_1$ (include external conflicts)
(c) $AtIAF_2$ (include defending attacks)
(d) $AtIAF_3$ (include attacks against external arguments currently defended by S)
(e) Fixed completion is the minimal completion of $AtIAF_4 = AtIAF_3$
(f) Alternative final result for modified instance where additionally $(b, f) \in \mathcal{R}^?$

Fig. 5. Graph representations of initial, intermediate, and final argumentation frameworks in the execution of the fixed completion algorithm of Definition 11 in Example 12 using $S = \{a, b\}$. Dashed attacks again are uncertain.

4. Assume that S is a stable extension of $AtIAF_S^{pes}$, i.e., S is conflict-free in $AtIAF_S^{pes}$ and S attacks each element $b \notin S$ in $AtIAF_S^{pes}$. Again, this implies that S is necessarily conflict-free in *AtIAF*. Further, since $\mathcal{R}_S^{pes}$ only contains attacks by S that are already in $\mathcal{R}$, S necessarily attacks each element $b \notin S$ in *AtIAF*. Combined, we have that S is necessarily stable in *AtIAF*.

This completes the proof. □

3.1.2. Fixed and unfixed completions

Turning now to the complete and the grounded semantics, we define the fixed and the unfixed completion. These completions make it most likely (respectively, unlikely) for S to be a fixed point of the completion's characteristic function.

Definition 11. Let $AtIAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{R}^? \rangle$ be an attack-incomplete argumentation framework and $S \subseteq \mathcal{A}$. The *fixed completion* $AtIAF_S^{fix}$ *of AtIAF for S* is the completion that is obtained by the following algorithm. The algorithm defines a finite sequence $(AtIAF_i)_{i \geq 0}$ of attack-incomplete argumentation frameworks, with the fixed completion being the minimal completion (that discards all remaining possible attacks) of the sequence's last element.

1. Include definite attacks: Let $AtIAF_0 = AtIAF$.
2. Include external conflicts: Let $AtIAF_1 = \langle \mathcal{A}, \mathcal{R}_1, \mathcal{R}_1^? \rangle$ with
   - $\mathcal{R}_1 = \mathcal{R} \cup \{(a, b) \in \mathcal{R}^? \mid a \notin S \text{ and } b \notin S\}$ and
   - $\mathcal{R}_1^? = \mathcal{R}^? \setminus \mathcal{R}_1$.
3. Include defending attacks: Let $T = \{t \in \mathcal{A} \setminus S \mid \exists s \in S : (t, s) \in \mathcal{R}_1\}$ (i.e., each argument in T necessarily attacks S) and let $AtIAF_2 = \langle \mathcal{A}, \mathcal{R}_2, \mathcal{R}_2^? \rangle$ with
   - $\mathcal{R}_2 = \mathcal{R}_1 \cup \{(a, b) \in \mathcal{R}_1^? \mid a \in S \text{ and } b \in T\}$ and
   - $\mathcal{R}_2^? = \mathcal{R}_1^? \setminus \mathcal{R}_2$.
4. Avoid S defending arguments outside of S: For the current i (initially, $i = 2$), let $AtIAF_i^{\min}$ be the minimal completion of $AtIAF_i$ and let $D_i = F_{AtIAF_i^{\min}}(S) \setminus S$ (i.e., D_i is the set of arguments that are not in S, but that are defended by S in the current minimal completion). Let $AtIAF_{i+1} = \langle \mathcal{A}, \mathcal{R}_{i+1}, \mathcal{R}_{i+1}^? \rangle$ with
   - $\mathcal{R}_{i+1} = \mathcal{R}_i \cup \{(a, b) \in \mathcal{R}_i^? \mid a \in S \text{ and } b \in D_i\}$ and
   - $\mathcal{R}_{i+1}^? = \mathcal{R}_i^? \setminus \mathcal{R}_{i+1}$,

   and set $i \leftarrow i + 1$.
5. Repeat Step 4 until no more attacks are added.
6. The fixed completion of *AtIAF* for S is $AtIAF_S^{fix} = \langle \mathcal{A}, \mathcal{R}_S^{fix} \rangle$ with $\mathcal{R}_S^{fix} = \mathcal{R}_i$.

Example 12. Consider an instance $(AtIAF, S)$ of CP-ATTINCPV or GR-ATTINCPV consisting of an attack-incomplete argumentation framework $AtIAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{R}^? \rangle$ with $\mathcal{A} = \{a, b, c, d, e, f\}$, $\mathcal{R} = \{(a, d), (c, b), (e, f)\}$, $\mathcal{R}^? = \{(a, b), (a, c), (a, e), (d, e)\}$, and a set $S = \{a, b\}$. The algorithm for the fixed completion from Definition 11 generates the following sequence $(AtIAF_i)_{i \geq 0}$ of attack-incomplete argumentation frameworks. Each of them is illustrated by its graph representation in Fig. 5.

- In Step 1, $AtIAF_0 = AtIAF$.
- In Step 2, the attack (d, e) is included in $\mathcal{R}_1$ because both d and e are not members of S.
- In Step 3, All attacks by S against arguments in $T = \{c\}$—namely, the attack (a, c)—are included in $\mathcal{R}_2$.
- Step 4, first iteration: $D_2 = \{e\}$, because $a \in S$ defends e against its only attacker d. All attacks by S against arguments in D_2—namely, the attack (a, e)—are included in $\mathcal{R}_3$.
- Step 4, second iteration: $D_3 = \{f\}$. However, there are no possible attacks by S against f, so $\mathcal{R}_4 = \mathcal{R}_3$ and the break condition in Step 5 is met.
- The remaining possible attack (a, b) is discarded by the minimal completion of $AtIAF_4$ in Step 6. $S = \{a, b\}$ is not complete or grounded in it, since S defends f. Proposition 14 will establish that this implies that S is neither possibly complete nor possibly grounded in *AtIAF*.

Now consider a slight variation of this instance where, additionally, the possible attack $(b, f) \in \mathcal{R}^?$ exists. All steps before the second iteration of Step 4 remain the same. Again, $D_3 = \{f\}$, but now the attack (b, f) is added to $\mathcal{R}_4$. In the third iteration, $D_4 = \emptyset$, so the loop terminates. Again, (a, b) is discarded by the minimal completion of the final intermediate argumentation framework $AtIAF_5$. The fixed completion for this instance is given in Fig. 5(f). Here, S is both complete and grounded in the fixed completion and therefore possibly complete and possibly grounded in *AtIAF*.

Proposition 13. *For an attack-incomplete argumentation framework* $AtIAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{R}^? \rangle$ *and a set* $S \subseteq \mathcal{A}$ *of arguments, the fixed completion* $AtIAF_S^{fix}$ *for S can be constructed in polynomial time.*

Proof. All individual steps in the construction can obviously be carried out in time polynomial in the number of arguments. It remains to prove that Step 4 is executed at most a polynomial number of times. In each execution there is either (at least) one possible attack that is added to $\mathcal{R}_{i+1}$, or no action is taken in which case the loop terminates. Therefore, the number of times Step 4 is executed is bounded by the number of possible attacks in the attack-incomplete argumentation framework *AtIAF*, which is at most n^2, where n is the number of arguments. This completes the proof. □

Proposition 14 establishes that the fixed completion is critical for possible verification using the complete and grounded semantics.

Proposition 14. *Let* $AtIAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{R}^? \rangle$ *be an attack-incomplete argumentation framework,* $S \subseteq \mathcal{A}$*, and let* $AtIAF_S^{fix}$ *be the fixed completion of AtIAF for S.*

1. *S is a possibly complete extension of AtIAF if and only if S is a complete extension of* $AtIAF_S^{fix}$.
2. *S is a possibly grounded extension of AtIAF if and only if S is the grounded extension of* $AtIAF_S^{fix}$.

Proof. The converse is trivial in both cases. Further, if S is not an admissible extension in $AtIAF_S^{fix}$, then S is not admissible in any completion of *AtIAF*, due to the same arguments that we used for the optimistic completion and, therefore, neither possibly complete nor possibly grounded in *AtIAF*. So, we may assume that S is admissible in $AtIAF_S^{fix}$.

1. Assume that S is not a complete extension of $AtIAF_S^{fix}$, i.e., S is not a fixed point of $F_{AtIAF_S^{fix}}$. We will show that this implies that S is not possibly complete in *AtIAF*. Let $AtIAF^*$ be any completion of *AtIAF* in which S is admissible. Since S is not a fixed point of $F_{AtIAF_S^{fix}}$, there is an argument $b \notin S$ which is acceptable with respect to S in $AtIAF_S^{fix}$. We prove that, then, there must be some $c \notin S$ for which all attackers of c are attacked by S in $AtIAF^*$ ($c = b$ may or may not be the case) by individually covering all cases in which attacks are added to $\mathcal{R}_S^{fix}$:
   - All attacks from $\mathcal{R}^?$ between arguments outside of S, which are added to $\mathcal{R}_S^{fix}$ in Step 2, cannot make an argument $b \notin S$ acceptable with respect to S: If S did not attack all attackers of an argument before, it cannot do so after *more* attackers are added.
   - All attacks that are added in Step 3 are crucial for S to be admissible, and must therefore also be included in $\mathcal{R}^*$. In a case where multiple arguments in S attack a single attacker of S, it would be sufficient to include one of these defending attacks, but including all of them does not make a difference, since the criterion of being acceptable with respect to S does not distinguish between different elements of S.
   - All attacks that are added in Step 4 are attacks by S against arguments that are currently acceptable with respect to S. Since all possible attacks among arguments outside of S were already included in Step 2, the only way to destroy acceptability of these arguments is by S directly attacking them. Therefore, none of the attacks added in Step 4 can be omitted without making the respective argument acceptable with respect to S (again, it is not necessary to distinguish between multiple attacks by different arguments in S against the same argument). It is possible for a given $b \notin S$ to be acceptable with respect to S in $AtIAF_S^{fix}$ and not in $AtIAF^*$, but this happens only if S attacks an attacker

(or several attackers) of b in $AtIAF_S^{fix}$ that would otherwise be acceptable with respect to S, and which therefore must be acceptable with respect to S in $AtIAF^*$. In either case, if an argument outside of S is acceptable with respect to S in $AtIAF_S^{fix}$, then some argument outside of S must be acceptable with respect to S in each completion $AtIAF^*$ of $AtIAF$ in which S is admissible. Therefore, if S is not a complete extension of $AtIAF_S^{fix}$, it is not a complete extension of any completion $AtIAF^*$ of $AtIAF$, and therefore not a possibly complete extension of $AtIAF$.

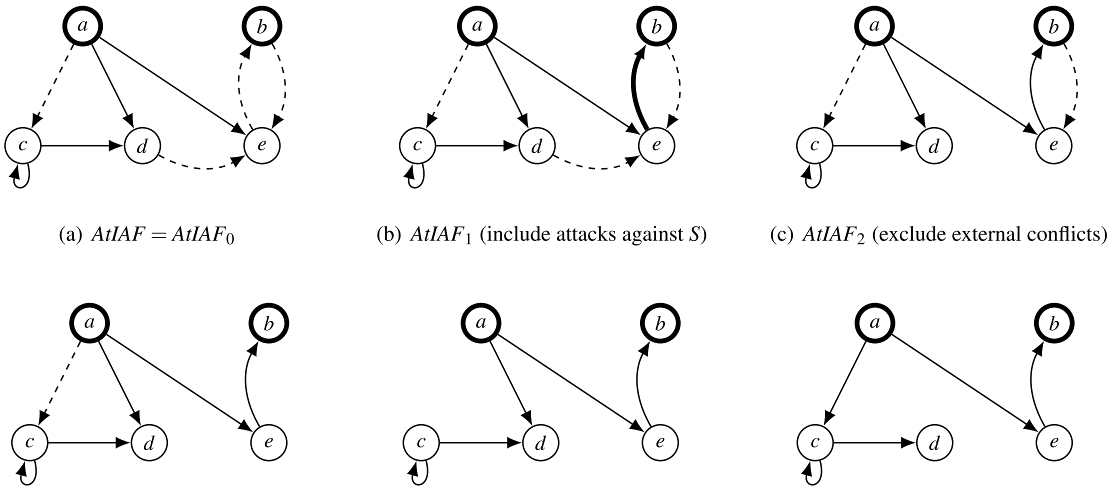
2. Let $AtIAF^*$ be an arbitrary completion of $AtIAF$ and assume that S is its grounded extension. We prove that, then, S is also the grounded extension of $AtIAF_S^{fix}$. Let $A_i = F^i_{AtIAF^*}(\emptyset)$ and $B_i = F^i_{AtIAF_S^{fix}}(\emptyset)$, where F^i is the i-fold composition of the respective characteristic function F. Since S is grounded in $AtIAF^*$, it is complete in $AtIAF_S^{fix}$ due to our previous result, and it holds that $A_i \subseteq S$ for all $i \geq 0$ and there exists a $j \geq 0$ such that for all $i \geq j$, it holds that $A_i = S$. We will prove that $A_i \subseteq B_i \subseteq S$ for all $i \geq 0$. Combined, these statements show that there exists some j such that $B_i = S$ for all $i \geq j$, which is equivalent to S being the grounded extension of $AtIAF_S^{fix}$.
First, we prove that $A_i \subseteq B_i$ for all $i \geq 0$. For $i = 0$, we have $A_i = B_i = \emptyset$. For $i = 1$, A_i (respectively, B_i) is the set of all unattacked arguments in $AtIAF^*$ (respectively, in $AtIAF_S^{fix}$). We know that $A_1 \subseteq S$. Since the fixed completion does not include any possible attacks against elements of S, all $a \in S$ that are unattacked in $AtIAF^*$ are unattacked in $AtIAF_S^{fix}$, too, which proves $A_1 \subseteq B_1$. If we now have $A_k \subseteq B_k$ for some $k \geq 1$, this implies $A_{k+1} \subseteq B_{k+1}$: Assume that this were not true, i.e., that $A_k \subseteq B_k$, but there is an argument $a \in A_{k+1}$ with $a \notin B_{k+1}$. a is acceptable with respect to A_k in $AtIAF^*$ but not acceptable with respect to B_k in $AtIAF_S^{fix}$. We know that—since $A_{k+1} \subseteq S$—no possible attacks against A_{k+1} (and in particular, against a) are included in $AtIAF_S^{fix}$ and all possible defending attacks by arguments in A_{k+1} against arguments outside of S are included in $AtIAF_S^{fix}$. Further, no element of S attacks a in $AtIAF_S^{fix}$, since $a \in S$ and S is complete in $AtIAF_S^{fix}$. Therefore, a is acceptable with respect to A_k in $AtIAF_S^{fix}$; otherwise it could not be acceptable with respect to A_k in $AtIAF^*$. Now, the only way for a to not be acceptable with respect to B_k in $AtIAF_S^{fix}$ is if there were some $b \in B_k \setminus A_k$ that necessarily attacks a. Then there would have to be a defending attack by an argument $d \in A_k$ against b in $AtIAF^*$, since a is acceptable with respect to A_k in $AtIAF^*$. This implies that $b \notin S$, since S is conflict-free in $AtIAF^*$. Finally, since (d, b) is a possible (or even a necessary) defending attack by an element of S against $b \notin S$, $(d, b) \in \mathscr{R}_S^{fix}$ holds by construction of the fixed completion, which contradicts that B_k is admissible in $AtIAF_S^{fix}$. Therefore, a must be acceptable with respect to B_k in $AtIAF_S^{fix}$, which proves that $A_{k+1} \subseteq B_{k+1}$.
Now we prove that $B_i \subseteq S$ for all $i \geq 0$: Assume that $B_i \nsubseteq S$ for some $i \geq 0$. Then it also holds that $G_S^{fix} \nsubseteq S$ for the grounded extension G_S^{fix} of $AtIAF_S^{fix}$. It further holds that $S \subset G_S^{fix}$, since there exists a $j \geq 0$ such that $S \subseteq B_i$ for all $i \geq j$, as established before. However, this contradicts the fact that S is complete in $AtIAF_S^{fix}$, since the grounded extension G_S^{fix} of $AtIAF_S^{fix}$ is its least complete extension with respect to set inclusion, as was shown by Dung [26], and the complete set S cannot be a strict subset of G_S^{fix}.

This completes the proof. □

We now turn to the unfixed completion, which can serve as a critical completion for necessary verification for the complete and grounded semantics.

Definition 15. Let $AtIAF = \langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$ be an attack-incomplete argumentation framework, let $S \subseteq \mathscr{A}$, and fix any ordering $\sigma : \mathscr{A} \to \{1, \ldots, |\mathscr{A}|\}$ on the arguments in $\mathscr{A}$. The *unfixed completion $AtIAF_{S,\sigma}^{unf}$ of AtIAF for S with respect to σ* is the completion that is obtained by the following algorithm. The algorithm defines a finite sequence $(AtIAF_i)_{i \geq 0}$ of attack-incomplete argumentation frameworks, with the unfixed completion being the minimal completion of the sequence's last element.

1. Include definite attacks: Let $AtIAF_0 = AtIAF$.
2. Include attacks against S: Let $AtIAF_1 = \langle \mathscr{A}, \mathscr{R}_1, \mathscr{R}_1^? \rangle$ with
   - $\mathscr{R}_1 = \mathscr{R} \cup \{(a, b) \in \mathscr{R}^? \mid b \in S\}$ and
   - $\mathscr{R}_1^? = \mathscr{R}^? \setminus \mathscr{R}_1$.
3. Exclude external conflicts: Let $AtIAF_2 = \langle \mathscr{A}, \mathscr{R}_2, \mathscr{R}_2^? \rangle$ with
   - $\mathscr{R}_2 = \mathscr{R}_1$ and
   - $\mathscr{R}_2^? = \mathscr{R}_1^? \setminus \{(a, b) \in \mathscr{R}_1^? \mid a \notin S \text{ and } b \notin S\}$.
4. Exclude defending attacks: Let $T = \{t \in \mathscr{A} \setminus S \mid \exists s \in S : (t, s) \in \mathscr{R}_2\}$ (i.e., each argument in T necessarily attacks S) and let $AtIAF_3 = \langle \mathscr{A}, \mathscr{R}_3, \mathscr{R}_3^? \rangle$ with
   - $\mathscr{R}_3 = \mathscr{R}_2$ and
   - $\mathscr{R}_3^? = \mathscr{R}_2^? \setminus \{(a, b) \in \mathscr{R}_2^? \mid a \in S \text{ and } b \in T\}$.
5. Try to make S defend arguments outside of S: Let $D = \mathscr{A} \setminus S = \{d_1, \ldots, d_k\}$. For the current i (initially, $i = 3$) and successively for each $d_j \in D$ (in order according to σ), do:



(a) $AtIAF = AtIAF_0$ (b) $AtIAF_1$ (include attacks against S) (c) $AtIAF_2$ (exclude external conflicts)

(d) $AtIAF_3$ (exclude defending attacks) (e) Fixed completion is the minimal completion of $AtIAF_4 = AtIAF_3$ (f) Alternative final result for modified instance where $(a,d) \notin \mathscr{R}$

Fig. 6. Graph representations of initial, intermediate, and final argumentation frameworks in the execution of the unfixed completion algorithm of Definition 15 in Example 16 using $S = \{a, b\}$. Dashed attacks again are uncertain.

(a) For $S_{d_j} = S \cup \{d_j\}$, let $AtIAF^{opt}_{i,S_{d_j}}$ be the optimistic completion of $AtIAF_i$ for S_{d_j} and let $AtIAF^{\min}_i$ be the minimal completion of $AtIAF_i$.

(b) If d_j is defended by S in $AtIAF^{opt}_{i,S_{d_j}}$, but not defended by S in $AtIAF^{\min}_i$, let $AtIAF_{i+1} = \langle \mathscr{A}, \mathscr{R}_{i+1}, \mathscr{R}^?_{i+1} \rangle$ with

- $\mathscr{R}_{i+1} = \mathscr{R}_i \cup \{(a,b) \in \mathscr{R}^?_i \mid a \in S \text{ and } (b, d_j) \in \mathscr{R}_i\}$ and
- $\mathscr{R}^?_{i+1} = \mathscr{R}^?_i \setminus \mathscr{R}_{i+1}$,

and set $i \leftarrow i+1$. (To accept an argument d_j that is not currently defended by S but possibly defended by S, include all possible attacks by S against d_j's attackers.)

6. The unfixed completion of $AtIAF$ for S with respect to σ is $AtIAF^{unf}_{S,\sigma} = \langle \mathscr{A}, \mathscr{R}^{unf}_{S,\sigma} \rangle$ with $\mathscr{R}^{unf}_{S,\sigma} = \mathscr{R}_i$.

Example 16. Consider an instance $(AtIAF, S)$ of cp-AttIncNV or gr-AttIncNV consisting of an attack-incomplete argumentation framework $AtIAF = \langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$ with $\mathscr{A} = \{a,b,c,d,e\}$, $\mathscr{R} = \{(a,d), (a,e), (c,c), (c,d)\}$, $\mathscr{R}^? = \{(a,c), (b,e), (d,e), (e,b)\}$, and a set $S = \{a,b\}$. σ orders arguments lexicographically. The algorithm for the unfixed completion from Definition 15 generates the following sequence $(AtIAF_i)_{i \geq 0}$ of attack-incomplete argumentation frameworks (each of them is illustrated by its graph representation in Fig. 6):

- In Step 1, $AtIAF_0 = AtIAF$.
- In Step 2, the attack (e,b) is included in $\mathscr{R}_1$ because b is a member of S.
- In Step 3, the attack (d,e) is excluded from $\mathscr{R}^?_2$ because both d and e are not members of S.
- In Step 4, all attacks by arguments in S against arguments in $T = \{e\}$—namely, the attack (b,e)—are excluded from $\mathscr{R}^?_3$.
- In Step 5, we have $D = \{c,d,e\}$. The iteration order due to σ is c, then d, and then e. There is only one possible attack remaining, namely, (a,c).
  - $S_c = \{a,b,c\}$. Both the optimistic completion $AtIAF^{opt}_{4,S_c}$ of $AtIAF_4$ for S_c and the minimal completion $AtIAF^{\min}_4$ of $AtIAF_4$ discard the possible attack (a,c). c is defended by S in both completions, so the condition in Step 5b is not met and no new intermediate argumentation framework is created.
  - $S_d = \{a,b,d\}$. Here, the optimistic completion $AtIAF^{opt}_{4,S_d}$ of $AtIAF_4$ for S_d includes the possible attack (a,c) and the minimal completion $AtIAF^{\min}_4$ of $AtIAF_4$ discards the possible attack (a,c). However, d is not defended by S in either of the two completions, so again, there is no new intermediate argumentation framework.
  - $S_e = \{a,b,e\}$. Again, the optimistic completion $AtIAF^{opt}_{4,S_e}$ of $AtIAF_4$ for S_e includes the possible attack (a,c) and the minimal completion $AtIAF^{\min}_4$ of $AtIAF_4$ discards the possible attack (a,c). Again, e is not defended by S in either completion and there is no new intermediate argumentation framework.
- In Step 6, the remaining possible attack (a,c) in $AtIAF_4$ is discarded by the unfixed completion. $S = \{a,b\}$ is complete and grounded in it, and therefore, as will be shown in Proposition 18, S is necessarily complete and grounded in $AtIAF$.

Consider again a slight variation $(AtIAF', S)$ of this instance where the attack (a, d) does not exist. All steps before the second iteration of Step 5 are the same as before. Again, the optimistic completion $AtIAF'^{opt}_{4,S_d}$ of $AtIAF'_4$ for S_d includes the possible attack (a, c) and the minimal completion $AtIAF'^{\min}_4$ of $AtIAF'_4$ discards the possible attack (a, c). This time, d is defended by S in $AtIAF'^{opt}_{4,S_d}$, but not in $AtIAF'^{\min}_4$. Thus, a new intermediate argumentation framework $AtIAF'_5$ is created which includes the possible attack (a, c) and which also is the unfixed completion of $AtIAF'$ displayed in Fig. 6(f). Here, S defends $d \notin S$ and is neither complete nor grounded in the unfixed completion, and therefore clearly neither necessarily complete nor necessarily grounded in $AtIAF'$.

Proposition 17. *For an attack-incomplete argumentation framework $AtIAF = \langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$, a set $S \subseteq \mathscr{A}$ of arguments, and an ordering σ on $\mathscr{A}$, the unfixed completion $AtIAF^{unf}_{S,\sigma}$ of AtIAF for S with respect to σ can be constructed in polynomial time.*

Proof. Again, all individual steps can be carried out in time polynomial in the number of arguments. The sub-loop in Step 5 has a predefined number of iterations that is bounded by the number n of arguments. The construction of the minimal and the optimistic completion in each iteration is possible in polynomial time. This completes the proof. □

Proposition 18 establishes that the unfixed completion is critical for necessary verification using the complete and grounded semantics.

Proposition 18. *Let $AtIAF = \langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$ be an attack-incomplete argumentation framework, $S \subseteq \mathscr{A}$, σ be an ordering on $\mathscr{A}$, and let $AtIAF^{unf}_{S,\sigma}$ be the unfixed completion of AtIAF for S with respect to σ.*

1. *S is a necessarily complete extension of AtIAF if and only if S is a complete extension of $AtIAF^{unf}_{S,\sigma}$.*
2. *S is a necessarily grounded extension of AtIAF if and only if S is the grounded extension of $AtIAF^{unf}_{S,\sigma}$.*

Proof. Here, the left-to-right implication is trivial in both cases. We prove the other direction of the implications individually. First, if S is not necessarily admissible in $AtIAF$, S is not admissible either (and, therefore, neither complete nor grounded) in $AtIAF^{unf}_{S,\sigma}$, because $AtIAF^{unf}_{S,\sigma}$ includes all possible attacks against arguments in S and excludes all defending attacks by arguments in S. We may therefore assume that S is necessarily admissible in $AtIAF$.

1. Assume that S is not necessarily complete in $AtIAF$. We prove that S is not complete in $AtIAF^{unf}_{S,\sigma}$: Since S is necessarily admissible but not necessarily complete in $AtIAF$, there is a completion $AtIAF^*$ of $AtIAF$ in which there exists some $b' \in \mathscr{A} \setminus S$ that is acceptable with respect to S in $AtIAF^*$. Obviously, this means that b' is possibly acceptable with respect to S in $AtIAF$. We will prove that, after each step of the algorithm, if there is some $b \in \mathscr{A} \setminus S$ that is acceptable with respect to S in $AtIAF_i$, then there also is some $c \in \mathscr{A} \setminus S$ that is acceptable with respect to S in $AtIAF_{i+1}$ ($c = b$ may or may not be the case).
   - After Step 1, b' is possibly acceptable with respect to S in $AtIAF_0$, since $AtIAF_0 = AtIAF$.
   - After Step 2, b' is possibly acceptable with respect to S in $AtIAF_1$, because including attacks against S has no influence on whether S possibly attacks all attackers of b'.
   - After Step 3, b' is possibly acceptable with respect to S in $AtIAF_2$, because excluding attacks between arguments in $\mathscr{A} \setminus S$ can only make it more likely for S to attack all attackers of b'.
   - Step 4 has no effect on instances where S is necessarily admissible, because there are no possible defending attacks by S against $\mathscr{A} \setminus S$ that could be excluded, since in such an instance S necessarily defends itself against all possible attacks.
   - The only way for an argument $b \in \mathscr{A} \setminus S$ to no longer be possibly acceptable with respect to S in $AtIAF_{i+1}$ after an iteration of Step 5 is if an attack by some $a \in S$ against b is included. If this is the case, the defended argument d_j is possibly acceptable with respect to S in $AtIAF_{i+1}$. Either way, the previously possibly acceptable argument b or the new argument d_j is possibly acceptable with respect to S in $AtIAF_{i+1}$.

   After Step 4, the only attacks that are not yet definite are attacks by arguments in S against arguments in $\mathscr{A} \setminus S$. Therefore, the only way for the condition in Step 5b to be met—i.e., d_j is possibly, but not currently accepted by S—is if there is an attack $(a, b) \in \mathscr{R}^?_i$ with $a \in S$ and $(b, d_j) \in \mathscr{R}_i$, which proves that $AtIAF_{i+1} \neq AtIAF_i$. So, when the algorithm terminates in Step 6, we know that there is an argument $b \in \mathscr{A} \setminus S$ that is possibly acceptable with respect to S in $AtIAF_i$ (as proven earlier) and that is also acceptable with respect to S in $AtIAF_i$'s minimal completion, because otherwise the condition in Step 5b would have been met. Since the unfixed completion is $AtIAF_i$'s minimal completion, this establishes that there is an argument in $\mathscr{A} \setminus S$ that is acceptable with respect to S in $AtIAF^{unf}_{S,\sigma}$, which implies that S is not complete in $AtIAF^{unf}_{S,\sigma}$, and concludes the proof of the first item.
2. Assume that S is the grounded extension of $AtIAF^{unf}_{S,\sigma}$. We prove that, then, S is the grounded extension of all completions of $AtIAF$. Let $AtIAF^*$ be an arbitrary completion of $AtIAF$ and let $A_i = F^i_{AtIAF^*}(\emptyset)$ and $B_i = F^i_{AtIAF^{unf}_{S,\sigma}}(\emptyset)$, where F^i

is the i-fold composition of the respective characteristic function F. Since S is grounded in $AtIAF^{unf}_{S,\sigma}$, it is complete in $AtIAF^*$ due to our previous result, and it holds that $B_i \subseteq S$ for all $i \geq 0$ and there exists a $j \geq 0$ such that for all $i \geq j$, it holds that $B_i = S$. We will prove that $B_i \subseteq A_i \subseteq S$ for all $i \geq 0$. Combined, these statements show that there exists some j such that $A_i = S$ for all $i \geq j$, which is equivalent to S being the grounded extension of $AtIAF^*$.
First, we prove that $B_i \subseteq A_i$ for all $i \geq 0$: For $i = 0$, we have $A_i = B_i = \emptyset$. For $i = 1$, A_i (respectively, B_i) is the set of all unattacked arguments in $AtIAF^*$ (respectively, in $AtIAF^{unf}_{S,\sigma}$). We know that $B_1 \subseteq S$. Since the unfixed completion includes all possible attacks against elements of S, all $a \in S$ that are unattacked in $AtIAF^{unf}_{S,\sigma}$ are necessarily unattacked, and therefore unattacked in $AtIAF^*$, too, which proves $B_1 \subseteq A_1$. If we now have $B_k \subseteq A_k$ for some $k \geq 1$, this implies $B_{k+1} \subseteq A_{k+1}$: Assume that this is not true, i.e., that $B_k \subseteq A_k$, but there is an argument $b \in B_{k+1}$ with $b \notin A_{k+1}$. b is acceptable with respect to B_k in $AtIAF^{unf}_{S,\sigma}$ but not acceptable with respect to A_k in $AtIAF^*$. Recall that all possible attacks against B_{k+1} (and in particular, against b) are included in $AtIAF^{unf}_{S,\sigma}$ and no possible defending attacks by arguments in B_{k+1} against arguments outside of S are included in $AtIAF^{unf}_{S,\sigma}$. Therefore, since b is acceptable with respect to $B_k \subseteq S$ in $AtIAF^{unf}_{S,\sigma}$, it is necessarily acceptable with respect to B_k and, in particular, acceptable with respect to B_k in $AtIAF^*$. Now, the only way for b to not be acceptable with respect to A_k in $AtIAF^*$ is if there were some $a \in A_k \setminus B_k$ that possibly attacks b. Then there would have to be a defending attack by an argument $d \in B_k$ against a in $AtIAF^{unf}_{S,\sigma}$, since b is acceptable with respect to B_k in $AtIAF^{unf}_{S,\sigma}$. This implies that $a \notin S$, since S is conflict-free in $AtIAF^{unf}_{S,\sigma}$. Finally, since (d, a) is a necessary attack, it holds in particular that $(d, a) \in \mathscr{R}^*$, which contradicts that A_k is admissible in $AtIAF^*$. Therefore, b must be acceptable with respect to A_k in $AtIAF^*$, which proves that $B_{k+1} \subseteq A_{k+1}$.
Now we prove that $A_i \subseteq S$ for all $i \geq 0$: Assume that $A_i \nsubseteq S$ for some $i \geq 0$. Then it also holds that $G^* \nsubseteq S$ for the grounded extension G^* of $AtIAF^*$. It further holds that $S \subset G^*$, since there exists a $j \geq 0$ such that $S \subseteq A_i$ for all $i \geq j$, as established before. However, this contradicts the fact that S is complete in $AtIAF^*$, since the grounded extension G^* of $AtIAF^*$ is its least complete extension with respect to set inclusion, and cannot be a strict subset of the complete extension S.

This completes the proof. □

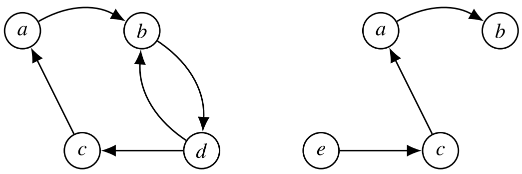
3.2. Argument incompleteness

In our second model, we allow uncertainty about the set of arguments. While the total set of arguments that may take part in the argumentation is known (and finite), there is uncertainty for some of these arguments as to whether or not they actually exist in the argumentation—they may not be constructible given a certain knowledge base, they may not be applicable, or they may simply not be brought forward by any agent. Note that this notion of possible nonexistence is different from that of (in)acceptability.

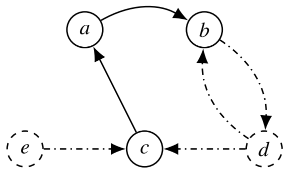
Definition 19. An *argument-incomplete argumentation framework* is a triple $\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle$, where $\mathscr{A}$ and $\mathscr{A}^?$ are disjoint sets of arguments and $\mathscr{R}$ is a subset of $(\mathscr{A} \cup \mathscr{A}^?) \times (\mathscr{A} \cup \mathscr{A}^?)$. $\mathscr{A}$ is the set of arguments that are known to definitely exist, while $\mathscr{A}^?$ contains all possible additional arguments that are not (yet) known to exist. Attacks in $\mathscr{R}$ that are incident to at least one uncertain argument (i.e., a member of $\mathscr{A}^?$) are called *conditionally definite*; they are known to definitely exist exactly if both incident arguments are known to definitely exist. All other attacks in $\mathscr{R}$ (i.e., attacks not incident to a member of $\mathscr{A}^?$) are simply called *definite*.

Note that, in this model, there is no uncertainty regarding the attack relation—even though conditionally definite attacks may be indirectly excluded by excluding an incident argument. As an example, consider a discussion where each agent has a private set of arguments that they can bring forward, but they may also choose to not introduce some of the arguments that they know of—maybe for strategic purposes. However, for the “outcome” of the argumentation, only those arguments that were explicitly stated by some agent are considered. Such a situation could be modeled using an argument-incomplete argumentation framework.

Example 20. Extending the argumentation framework from Example 2 by two possible arguments $\mathscr{A}^? = \{d, e\}$ together with an expansion of the attack relation, by including the attacks (d, b), (d, c), (b, d), and (e, c), yields the argument-incomplete argumentation framework $\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle$ the graph representation of which is given in Fig. 7(b). As already discussed in Example 5, such an incomplete framework might result from merging several individual views, which agree on all attacks over those arguments that are known to all agents but may have different argument sets. Fig. 7(a) shows two such individual argumentation frameworks, which are then merged into the argument-incomplete argumentation framework of Fig. 7(b) by including all arguments that are known in every agent’s argumentation framework as definite arguments ($\mathscr{A}$), and including arguments that exist in some but not in all agents’ argumentation frameworks as possible arguments ($\mathscr{A}^?$). Note that there is no choice of whether or not we include attacks: Attacks must be identical in all agents’ individual views that contain the corresponding arguments, and an attack is included in the argument-incomplete argumentation framework if and only if both adjacent arguments are included.



(a) Agent 1's (left) and agent 2's (right) individual views



(b) Merging the views of agents 1 and 2 (dashed arguments are uncertain, i.e., in $\mathscr{A}^?$, and dash-dotted attacks are conditionally definite, i.e., in $\mathscr{R}$ and incident to an argument in $\mathscr{A}^?$)

Fig. 7. Argument incompleteness.

Also for argument-incomplete argumentation frameworks, we can define completions quite similar to those of Definition 6:

Definition 21. Let $ArIAF = \langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle$ be an argument-incomplete argumentation framework. For a set $\mathscr{A}^*$ of arguments with $\mathscr{A} \subseteq \mathscr{A}^* \subseteq \mathscr{A} \cup \mathscr{A}^?$, define the *restriction of* $\mathscr{R}$ *to* $\mathscr{A}^*$ by $\mathscr{R}|_{\mathscr{A}^*} = \{(a,b) \in \mathscr{R} \mid a, b \in \mathscr{A}^*\}$. Then an argumentation framework $ArIAF^* = \langle \mathscr{A}^*, \mathscr{R}|_{\mathscr{A}^*} \rangle$ is called a *completion of ArIAF*.

Note that a conditionally definite attack can be contained in a completion $ArIAF^*$ only if $ArIAF^*$ includes both arguments incident to this attack. Obviously, the total number of possible completions is again exponential—this time in the number of possible new arguments, i.e., there can be up to $2^{|\mathscr{A}^?|}$ possible completions.

Let us now define the two variants of the verification problem in argument-incomplete argumentation frameworks for each given semantics **s**:

	s-Arg-Inc-Possible-Verification (**s**-ArgIncPV)
Given:	An argument-incomplete argumentation framework $ArIAF = \langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle$ and a set $S \subseteq \mathscr{A} \cup \mathscr{A}^?$.
Question:	Is there a completion $ArIAF^* = \langle \mathscr{A}^*, \mathscr{R}|_{\mathscr{A}^*} \rangle$ of $ArIAF$ such that $S|_{\mathscr{A}^*} = S \cap \mathscr{A}^*$ is an **s** extension of $ArIAF^*$?

	s-Arg-Inc-Necessary-Verification (**s**-ArgIncNV)
Given:	An argument-incomplete argumentation framework $ArIAF = \langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle$ and a set $S \subseteq \mathscr{A} \cup \mathscr{A}^?$.
Question:	For all completions $ArIAF^* = \langle \mathscr{A}^*, \mathscr{R}|_{\mathscr{A}^*} \rangle$ of $ArIAF$, is $S|_{\mathscr{A}^*} = S \cap \mathscr{A}^*$ an **s** extension of $ArIAF^*$?

3.3. General incompleteness

We now combine the two given models by allowing incomplete knowledge about both the attack relation and the set of arguments at the same time.

Definition 22. An *incomplete argumentation framework* is a quadruple $\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle$, where $\mathscr{A}$ and $\mathscr{A}^?$ are disjoint sets of arguments and $\mathscr{R}$ and $\mathscr{R}^?$ are disjoint subsets of $(\mathscr{A} \cup \mathscr{A}^?) \times (\mathscr{A} \cup \mathscr{A}^?)$. $\mathscr{A}$ (respectively, $\mathscr{R}$) is the set of arguments (respectively, the set of attacks) that are known to definitely exist, while $\mathscr{A}^?$ (respectively, $\mathscr{R}^?$) contains all possible additional arguments (respectively, all possible additional attacks) not (yet) known to exist. The set of attacks that are known to never exist is denoted by $\mathscr{R}^- = (\mathscr{A} \times \mathscr{A}) \setminus (\mathscr{R} \cup \mathscr{R}^?)$.

The difference between a conditionally definite attack (which, recall Definition 19, belongs to $\mathscr{R}$ and is incident to at least one argument in $\mathscr{A}^?$) and an uncertain attack (a member of $\mathscr{R}^?$) is that the former must occur in all completions containing both of its incident arguments, whereas the latter may vanish in a completion containing both incident arguments.

Again, an incomplete argumentation framework can be the result of merging a number of individual argumentation frameworks. Recall that, in Section 3.1, we only allowed those argumentation frameworks to be merged that share a common

(a) Another agent's individual view (b) Merging the argumentation frameworks of Figures 3(a), 7(a), and 8(a) (dashed attacks and arguments again are uncertain, while dash-dotted attacks are conditionally definite)

Fig. 8. General incompleteness.

set of arguments, i.e., we could aggregate only those argumentation frameworks $AF_1 = \langle \mathscr{A}_1, \mathscr{R}_1 \rangle, \dots, AF_n = \langle \mathscr{A}_n, \mathscr{R}_n \rangle$ for which $\mathscr{A}_i = \mathscr{A}_j$ holds for each $i, j \in \{1, \dots, n\}$. And in Section 3.2 we restricted ourselves to those argumentation frameworks that agree on all attacks between common arguments. Formally, this can be expressed by requiring $\mathscr{R}_i|_{\mathscr{A}_i \cap \mathscr{A}_j} = \mathscr{R}_j|_{\mathscr{A}_i \cap \mathscr{A}_j}$ for all $i, j \in \{1, \dots, n\}$.

In the general model, however, we do not restrict the input anymore. Hence, we need to specify how we can merge argumentation frameworks that were not mergeable before, namely those over possibly different sets of arguments regarding attack incompleteness, and those over possibly different attack relations in the case of argument incompleteness.

Definition 23. The merging operation for n individual argumentation frameworks $AF_1, \dots, AF_n$ produces the following incomplete argumentation framework $\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle$: $\mathscr{A}$ consists of all arguments that belong to *all* $AF \in \{AF_1, \dots, AF_n\}$. $\mathscr{A}^?$ consists of all arguments that belong to *at least one* (but not to all) $AF \in \{AF_1, \dots, AF_n\}$. $\mathscr{R}$ consists of all attacks (a, b) that belong to *all* $AF \in \{AF_1, \dots, AF_n\}$ containing both a and b. $\mathscr{R}^?$ consists of all attacks (a, b) that belong to *at least one* (but not to all) $AF \in \{AF_1, \dots, AF_n\}$ that contain both a and b.

Example 24. Extending the argumentation framework from Example 2 the same way we did in Examples 5 and 20, we obtain the incomplete argumentation framework $\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle$ the graph representation of which is given in Fig. 8(b). This incomplete argumentation framework is the result of merging the individual argumentation frameworks from Figs. 3(a), 7(a), and 8(a) according to Definition 23.

The given merge operation is a strict generalization of those in Sections 3.1 and 3.2. If we restrict the input of the merging operation the same way we restricted the input in Section 3.1 (that is, requiring $\mathscr{A}_i = \mathscr{A}_j$ for all $i, j \in \{1, \dots, n\}$), we have $\mathscr{A}^? = \emptyset$ and the same merging operation as defined there. On the other hand, if we restrict the input the same way we did in Section 3.2 (that is, requiring $\mathscr{R}_i|_{\mathscr{A}_i \cap \mathscr{A}_j} = \mathscr{R}_j|_{\mathscr{A}_i \cap \mathscr{A}_j}$ for all $i, j \in \{1, \dots, n\}$), we have $\mathscr{R}^? = \emptyset$ and the same merging operation as defined there. Accordingly, incomplete argumentation frameworks are a true generalization of both individual models of incomplete argumentation frameworks. Fixing $\mathscr{A}^? = \emptyset$ yields exactly the class of attack-incomplete argumentation frameworks, and fixing $\mathscr{R}^? = \emptyset$ yields exactly the class of argument-incomplete argumentation frameworks.

The merging operation we defined above regarding the argument sets can be seen as a global merging: If an argument is contained in all input argumentation frameworks, put it into $\mathscr{A}$, otherwise into $\mathscr{A}^?$. In contrast, the merging operation regarding the attack relation is a local merging: If an attack (a, b) is contained in all those inputs that actually have an opinion over both a and b, put it into $\mathscr{R}$, otherwise into $\mathscr{R}^?$. This conforms to the way in which *consensual expansion*, as defined by Coste-Marquis et al. [19], handles the merging of attacks.

In the general model of incomplete argumentation framework, a notion of *completion* can now be defined as follows.

Definition 25. Let $IAF = \langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle$ be a given incomplete argumentation framework. An argumentation framework $IAF^* = \langle \mathscr{A}^*, \mathscr{R}^* \rangle$ with $\mathscr{A} \subseteq \mathscr{A}^* \subseteq \mathscr{A} \cup \mathscr{A}^?$ and $\mathscr{R}|_{\mathscr{A}^*} \subseteq \mathscr{R}^* \subseteq \left(\mathscr{R} \cup \mathscr{R}^?\right)|_{\mathscr{A}^*}$ is called a *completion of IAF*.

Finally, for each given semantics **s**, the variants of the verification problem adapted to incomplete argumentation frameworks are defined analogously to those in the purely attack-incomplete and the purely argument-incomplete setting.

	s-Inc-Possible-Verification (**s**-IncPV)
Given:	An incomplete argumentation framework $IAF = \langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle$ and a set $S \subseteq \mathscr{A} \cup \mathscr{A}^?$.
Question:	Is there a completion $IAF^* = \langle \mathscr{A}^*, \mathscr{R}^* \rangle$ of IAF such that $S|_{\mathscr{A}^*} = S \cap \mathscr{A}^*$ is an **s** extension of IAF^*?

	s-Inc-Necessary-Verification (**s**-IncNV)
Given:	An incomplete argumentation framework $IAF = \langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle$ and a set $S \subseteq \mathscr{A} \cup \mathscr{A}^?$.
Question:	For all completions $IAF^* = \langle \mathscr{A}^*, \mathscr{R}^* \rangle$ of IAF, is $S|_{\mathscr{A}^*} = S \cap \mathscr{A}^*$ an **s** extension of IAF^*?

4. Complexity of verification in incomplete argumentation frameworks

In this section, we provide a full map of the complexity of possible and necessary verification in all three presented models of incompleteness and for the conflict-free, admissible, stable, complete, grounded, and preferred semantics. All results are summarized in Table 1 in Section 5.

Since general incomplete argumentation frameworks are a generalization of both individual models of incompleteness, all upper complexity bounds for the general model carry over to both individual models, and all lower complexity bounds for any of the individual models carry over to the general model.

4.1. Upper bounds

We start by providing some simple upper bounds for the general incompleteness model (omitting results that are replaced by tighter results later).

Theorem 26.

1. *For* $\mathbf{s} \in \{\text{AD}, \text{ST}, \text{CP}, \text{GR}\}$, $\mathbf{s}$-IncPV *is in* NP.
2. PR-IncPV *is in* Σ_2^p.
3. PR-IncNV *is in* coNP.

Proof. The results follow directly from the quantifier representations of the given problems: In the possible case, we start with an existential quantifier, and in the necessary case with a universal quantifier. For $\mathbf{s} \in \{\text{AD}, \text{ST}, \text{CP}, \text{GR}\}$, it can be checked in polynomial time whether the given subset is an $\mathbf{s}$ extension, which provides the results of Item 1. The standard verification problem for the preferred semantics belongs to coNP, hence it can be written as a universal quantifier followed by a statement checkable in polynomial time. Therefore, we have two alternating quantifiers in the case of PR-IncPV (Item 2), and two universal quantifiers collapsing into one in the case of PR-IncNV (Item 3). This completes the proof. □

In Corollary 27, we derive the same upper bounds for the attack- and argument-incomplete models (again omitting results that are replaced by tighter results later).

Corollary 27.

1. *For* $\mathbf{s} \in \{\text{AD}, \text{ST}, \text{CP}, \text{GR}\}$, $\mathbf{s}$-ArgIncPV *is in* NP.
2. PR-AttIncPV *and* PR-ArgIncPV *are in* Σ_2^p.
3. PR-AttIncNV *and* PR-ArgIncNV *are in* coNP.

Next, we provide proofs for the cases where we were able to establish P membership. First, verification for conflict-freeness remains easy in all cases.

Theorem 28. CF-IncPV *and* CF-IncNV *are in* P.

Proof. Given an incomplete argumentation framework $IAF = \langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle$ and a set $S \subseteq \mathscr{A} \cup \mathscr{A}^?$ of arguments, S is possibly conflict-free in *IAF* if and only if $S|_{\mathscr{A}}$ is conflict-free in the *minimal completion* $\langle \mathscr{A}, \mathscr{R}|_{\mathscr{A}} \rangle$ of *IAF*, which discards all additional arguments and attacks. Similarly, S is necessarily conflict-free in *IAF* if and only if S is conflict-free in the *maximal completion* $\langle \mathscr{A} \cup \mathscr{A}^?, \mathscr{R} \cup \mathscr{R}^? \rangle$ of *IAF*, which includes all additional arguments and attacks. Since both the minimal and maximal completion can clearly be constructed in polynomial time, we have P membership for both problems. □

The following upper bounds then follow immediately; note that membership of CF-AttIncPV and CF-AttIncNV in P has previously been proven by Coste-Marquis et al. [19].

Corollary 29. CF-AttIncPV*,* CF-AttIncNV*,* CF-ArgIncPV*, and* CF-ArgIncNV *are in* P.

Next, we extend P membership of CF-ArgIncNV to the admissible and stable semantics.

Theorem 30. AD-ArgIncNV *and* ST-ArgIncNV *are in* P.

Proof. Let $I = (\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle, S)$ be an instance of AD-ArgIncNV. If S is not necessarily conflict-free in $\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle$, it is not necessarily admissible in $\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle$, either. Since CF-ArgIncNV is in P, this can be checked in polynomial time. In the following, we may assume that S is necessarily conflict-free.

Let $\mathscr{A}_0 = \mathscr{A} \cup (\mathscr{A}^? \setminus S)$ and $C_0 = \langle \mathscr{A}_0, \mathscr{R}|_{\mathscr{A}_0} \rangle$, and for each argument $a \in \mathscr{A}^? \cap S$, let $\mathscr{A}_a = \mathscr{A}_0 \cup \{a\}$ and $C_a = \langle \mathscr{A}_a, \mathscr{R}|_{\mathscr{A}_a} \rangle$. If, for some $x \in \{0\} \cup (\mathscr{A}^? \cap S)$, $S|_{\mathscr{A}_x}$ is not admissible in the completion C_x, we clearly have $I \notin$ AD-ARGINCNV. Since the number of these completions is bounded by the number of arguments (plus one), this can again be verified in polynomial time. We may now assume that, in each completion C_x, $S|_{\mathscr{A}_x}$ is admissible.

Note that each of these completions includes *all* possible attacks against the respective set $S|_{\mathscr{A}_x}$, because the completions include all possibly harmful arguments (members of $\mathscr{A}_0$) and because there cannot be any attacks among members of S. This yields that $S|_{\mathscr{A}_0}$ defends all attacks against its elements in *any* completion, and, for all $a \in \mathscr{A}^? \cap S$, $S|_{\mathscr{A}_a}$ defends all attacks against a in *any* completion. Finally, since in any completion $C^* = \langle \mathscr{A}^*, \mathscr{R}|_{\mathscr{A}^*} \rangle$, it holds that $S|_{\mathscr{A}^*} \subseteq \bigcup_x S|_{\mathscr{A}_x}$, we can conclude that each element of $S|_{\mathscr{A}^*}$ is acceptable with respect to $S|_{\mathscr{A}^*}$ in C^*, so S is necessarily admissible in $\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle$ and $I \in$ AD-ARGINCNV.

For ST-ARGINCNV, the same construction as above can be used. We can again conclude that $I \notin$ ST-ARGINCNV in all cases where we had $I \notin$ AD-ARGINCNV, since each stable set needs to be admissible. In addition, it is easy to see that, in order for S to be necessarily stable, the set $S|_{\mathscr{A}_0}$ in the completion C_0 as defined above needs to attack all arguments in $\mathscr{A}_0 \setminus S$. However, since $\mathscr{A}_0 \setminus S = \mathscr{A} \setminus S$ ($\mathscr{A}_0$ contains all arguments that are not in S) and further $S|_{\mathscr{A}_0}$ is a subset of $S|_{\mathscr{A}^*}$ for any completion with argument set $\mathscr{A}^*$, this already yields that $S|_{\mathscr{A}^*}$ necessarily attacks all arguments outside of $S|_{\mathscr{A}^*}$ in any completion, and we have $I \in$ ST-ARGINCNV. □

We further lift the previous result to the general incompleteness model.

Theorem 31. AD-INCNV *and* ST-INCNV *are in* P.

Proof. Let (IAF, S) with $IAF = \langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle$ be an instance of AD-INCNV. Let $IAF_S^{pes} = \langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}_S^{pes} \rangle$ with $\mathscr{R}_S^{pes} = \mathscr{R} \cup \{(a, b) \in \mathscr{R}^? \mid b \in S\}$ be the pessimistic argument-incomplete argumentation framework obtained when eliminating attack incompleteness by including each and only those attacks that target S (which can clearly be done in polynomial time).

We will prove that $(IAF, S) \in$ AD-INCNV $\iff (IAF_S^{pes}, S) \in$ AD-ARGINCNV. Since AD-ARGINCNV $\in$ P and IAF_S^{pes} can be created from IAF in polynomial time, this yields that AD-INCNV $\in$ P. A completely analogous argument applies to the stable semantics and the problem ST-INCNV.

If $(IAF, S) \in$ AD-INCNV, then $(IAF_S^{pes}, S) \in$ AD-ARGINCNV follows trivially, since the set of completions of IAF_S^{pes} is a subset of the completions of IAF. We prove the other direction of the equivalence by contraposition. Assume that $(IAF, S) \notin$ AD-INCNV. Then there is a completion IAF^* of IAF in which S is not admissible. Create a completion IAF_S^{pes*} from the argument-incomplete argumentation framework IAF_S^{pes} by adding exactly those elements of $\mathscr{A}^?$ to the set of arguments that are also added in IAF^*. By construction, in IAF_S^{pes*} all attacks against arguments in S that exist in IAF^* are included, too, and any attacks against arguments outside of S that are not in IAF^* are not included, either. Since S is not admissible in IAF^*, it can clearly not be admissible in IAF_S^{pes*}. Therefore, we have $(IAF_S^{pes}, S) \notin$ AD-ARGINCNV. This completes the proof. □

The following upper bounds then follow immediately; note that membership of AD-ATTINCNV in P has previously been proven by Coste-Marquis et al. [19].

Corollary 32. AD-ATTINCNV *and* ST-ATTINCNV *are in* P.

Turning to the complete and grounded semantics, we can successively prove P membership of CP-INCNV and GR-INCNV in Theorems 33 and 38, respectively.

Theorem 33. CP-INCNV *is in* P.

Proof. Let (IAF, S) with $IAF = \langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle$ be an instance of CP-INCNV. Since AD-INCNV $\in$ P, we may assume that S is necessarily admissible in IAF. Then, we clearly have $(IAF, S) \notin$ CP-INCNV if and only if there is at least one argument outside of S that is acceptable with respect to S in some completion of IAF. It remains to show how to check this criterion.

If all arguments $a \in (\mathscr{A} \cup \mathscr{A}^?) \setminus S$ are definitely attacked by S, i.e., $(b_a, a) \in \mathscr{R}$ for each such argument a and some corresponding $b_a \in S$, then S is necessarily stable and therefore necessarily complete, and we are done. Now assume this is not the case and let $a \in (\mathscr{A} \cup \mathscr{A}^?) \setminus S$ be any argument outside of S that is not definitely attacked by S, i.e., $(b, a) \notin \mathscr{R}$ for all $b \in S \cap \mathscr{A}$ (if a were attacked by S, it clearly could not be acceptable with respect to S in any completion). Let $Att(a) = \{b \in \mathscr{A} \cup \mathscr{A}^? \mid (b, a) \in \mathscr{R}\}$ be the set of all arguments with a definite attack against a. Further, let $\mathscr{R}_a = \mathscr{R} \cup \{(b, c) \in \mathscr{R}^? \mid b \in S \text{ and } c \in Att(a) \setminus \{a\}\}$ be the set of attacks that includes all and only those possible attacks for which the attacker is in S and the target is an attacker of a.

Consider now the completion $C_a = \langle \mathscr{A}_a, \mathscr{R}_a|_{\mathscr{A}_a} \rangle$ where $\mathscr{A}_a = \mathscr{A} \cup \{a\} \cup \{b \in \mathscr{A}^? \mid (b, a) \notin \mathscr{R}_a\}$, i.e., C_a uses the attack relation $\mathscr{R}_a$ and includes a and exactly those possible arguments that do not attack a (in $\mathscr{R}_a$).

If, for any of these completions, a is acceptable with respect to S in C_a, then S is not complete in C_a and therefore not necessarily complete. If, on the other hand, each argument a is not acceptable with respect to S in the respective completion C_a, then none of these arguments are possibly acceptable with respect to S, and therefore, S is necessarily complete: Assume that a is not acceptable with respect to S in C_a, i.e., there is some $b \in \mathscr{A}_a$ with $(b,a) \in \mathscr{R}_a|_{\mathscr{A}_a}$ and S does not attack b in C_a. By construction of C_a, we know that b is a definite argument, i.e., $b \in \mathscr{A}$, and (b,a) is a definite attack, i.e., $(b,a) \in \mathscr{R}$, so b attacks a in any completion that contains a. Also, in all completions S either does not defend a against b, or S attacks a, since all possible arguments in S either attack a or are already included in C_a. So, a is not possibly acceptable with respect to S.

All steps taken can clearly be performed in polynomial time. This completes the proof. □

The following upper bounds then follow immediately.

Corollary 34. CP-ATTINCNV *and* CP-ARGINCNV *are in* P.

Next, we introduce the notion of ungrounded completion of an incomplete argumentation framework as a tool to prove P membership of GR-INCNV.

Definition 35. Let $IAF = \langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle$ be an incomplete argumentation framework and $S \subseteq \mathscr{A} \cup \mathscr{A}^?$ be a set of arguments in *IAF*. The *ungrounded completion* IAF_S^{ungr} *of IAF for S* is the completion that is obtained by the following algorithm. The algorithm first eliminates attack incompleteness and then defines a finite sequence $(IAF_i)_{i \geq 0}$ of argument incomplete argumentation frameworks, with the ungrounded completion being the maximal completion (that includes all remaining possible arguments) of the sequence's last element.

1. Eliminate attack incompleteness: Let $\mathscr{R}_0 = \mathscr{R} \cup \{(a,b) \in \mathscr{R}^? \mid b \in S\}$, i.e., include only those possible attacks that attack S.
2. Let initially $G_0 = \emptyset$, $\mathscr{A}_0^? = \mathscr{A}^?$, $IAF_0 = \langle \mathscr{A}, \mathscr{A}_0^?, \mathscr{R}_0 \rangle$ and $i = 0$.
3. Let Max_i be the maximal completion of IAF_i and let $X_i \subseteq S$ be the set of arguments in S that are acceptable with respect to G_i in Max_i, i.e., $X_i = F_{Max_i}(G_i) \cap S$. Add the definite arguments in X_i to G and exclude the possible arguments in X_i from the framework, i.e.,
   - $G_{i+1} = G_i \cup (X_i \setminus \mathscr{A}^?)$,
   - $\mathscr{A}_{i+1}^? = \mathscr{A}_i^? \setminus X_i$, and
   - $\mathscr{R}_{i+1} = \mathscr{R}_i|_{\mathscr{A} \cup \mathscr{A}_{i+1}^?}$.

   Set $i \leftarrow i + 1$.
4. Repeat the previous step until $G_i = G_{i-1}$.
5. The ungrounded completion of *IAF* for S is $IAF_S^{ungr} = \langle \mathscr{A}_S^{ungr}, \mathscr{R}_i \rangle$ with $\mathscr{A}_S^{ungr} = \mathscr{A} \cup \mathscr{A}_i^?$.

Intuitively, the ungrounded completion removes all and only those arguments that are in S and that are possible candidates for membership in the grounded extension (elements of X_i in each iteration i)—all other arguments are included. The purpose of that is to make it as unlikely as possible for S to be grounded in this completion.

Lemma 36 establishes that the ungrounded completion is polynomial-time computable.

Lemma 36. *For an incomplete argumentation framework* $IAF = \langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle$ *and a set* $S \subseteq \mathscr{A} \cup \mathscr{A}^?$ *of arguments, the ungrounded completion* IAF_S^{ungr} *of IAF for S can be constructed in polynomial time.*

Proof. All individual steps can obviously be carried out in time polynomial in the number of arguments. Also, the loop in Step 4 runs at most a polynomial number of times, since in each execution of the loop there is either (at least) one definite argument that is added to G_{i+1}, or no action is taken in which case the loop terminates. Therefore, the number of times the loop is executed is bounded by the number of definite arguments in the incomplete argumentation framework *AtIAF*. This completes the proof. □

The ungrounded completion is *critical* in the following sense: If a necessarily complete set S is grounded even in the ungrounded completion, then it must be grounded in all completions. This is formalized in Lemma 37.

Lemma 37. *Let* $IAF = \langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle$ *be an incomplete argumentation framework,* $S \subseteq \mathscr{A} \cup \mathscr{A}^?$ *be a necessarily complete set of arguments in IAF, and let* IAF_S^{ungr} *be the ungrounded completion of IAF for S. S is the necessarily grounded extension of IAF if and only if* $S|_{\mathscr{A}_S^{ungr}}$ *is the grounded extension of* IAF_S^{ungr}.

Proof. If $S|_{\mathscr{A}_S^{ungr}}$ is not the grounded extension of IAF_S^{ungr}, it immediately follows that S is not necessarily grounded in *IAF*. We now prove the other direction of the equivalence: Let $S|_{\mathscr{A}_S^{ungr}}$ be the grounded extension of IAF_S^{ungr}. We prove that, then, S is necessarily grounded in *IAF*.

First, we observe that whenever $S|_{\mathscr{A}_S^{ungr}}$ is the grounded extension of IAF_S^{ungr} (which we know by assumption), then $S|_{\mathscr{A}_S^{ungr}} = G_{i'}$ for the set $G_{i'}$ in the last iteration i' of the algorithm: $G_{i'} \subseteq S|_{\mathscr{A}_S^{ungr}}$ holds because, by construction, $G_{i'}$ consists only of definite arguments. For the other inclusion $S|_{\mathscr{A}_S^{ungr}} \subseteq G_{i'}$, we can utilize the fact that $G_{i'}$ is a complete extension of IAF_S^{ungr}: $G_{i'}$ is conflict-free since it is a subset of the grounded extension $S|_{\mathscr{A}_S^{ungr}}$, and it is a fixed point of the characteristic function due to the condition in Step 4 of the algorithm. Since the grounded extension is a subset of all complete extensions, this directly infers the desired inclusion $S|_{\mathscr{A}_S^{ungr}} \subseteq G_{i'}$. Since $G_{i'}$ consists only of definite arguments, we know that $S|_{\mathscr{A}_S^{ungr}}$ consists only of definite arguments under the given assumptions.

Now, let $IAF^* = \langle \mathscr{A}^*, \mathscr{R}|_{\mathscr{A}^*} \rangle$ be any completion of $\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle$ (different from the ungrounded completion) and let G^* be its grounded extension. Since we know by assumption that $S|_{\mathscr{A}^*}$ is complete in IAF^*, with the fact (proven by Dung [26]) that the grounded extension is contained in all complete extensions of the same argumentation framework, we can conclude that $G^* \subseteq S|_{\mathscr{A}^*}$.

However, we also have $S|_{\mathscr{A}^*} \subseteq G^*$: Since $S|_{\mathscr{A}_S^{ungr}}$ contains only definite arguments, these must be in G^*, too. Now assume that $S|_{\mathscr{A}^*} \nsubseteq G^*$. Then there is a possible (nondefinite) argument $a \in (S|_{\mathscr{A}^*} \setminus G^*)$. We know that a is not included in the ungrounded completion. We also know that a is not acceptable with respect to G^* in IAF^*, because otherwise it would need to be included in the grounded set G^*. Also, since $S|_{\mathscr{A}_S^{ungr}} \subseteq G^*$, a is not acceptable with respect to $S|_{\mathscr{A}_S^{ungr}}$ either (remember that S is necessarily complete and, in particular, necessarily conflict-free in IAF, so any attackers must be outside of S). So, there must be an attacker $b \notin S$ of a which is not attacked by G^* (and, therefore, not attacked by $S|_{\mathscr{A}_S^{ungr}}$) in IAF^*. Since the ungrounded completion includes all arguments that are not in S, b is also included in $\mathscr{A}_S^{ungr}$. Further, since the ungrounded completion includes all and only those possible attacks that target S, the attack (b, a) is included and any possible defending attacks are not included in the ungrounded completion. However, this means that the attack (b, a) is not defended by $S|_{\mathscr{A}_S^{ungr}}$ in the ungrounded completion, which, by its construction, would mean that a would be included in $\mathscr{A}_S^{ungr}$ (a could only be excluded in Step 3 if it is acceptable with respect to a subset of $S|_{\mathscr{A}_S^{ungr}}$, which a is not, due to the attack by b). This contradicts the fact that a is not included in the ungrounded completion. Therefore, such an argument a cannot exist and we can conclude $S|_{\mathscr{A}^*} \subseteq G^*$ and, in total, $S|_{\mathscr{A}^*} = G^*$. So, $S|_{\mathscr{A}^*}$ is grounded in IAF^* and, since IAF^* was kept generic, S is necessarily grounded in IAF. □

Using the above lemmas, we are now ready to show that for the grounded semantics, necessary verification in incomplete argumentation frameworks remains efficient.

Theorem 38. GR-INCNV *is in* P.

Proof. Let $(\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle, S)$ be an instance of GR-INCNV. If the set S is not necessarily complete in $\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle$, it is not necessarily grounded in $\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}, \mathscr{R}^? \rangle$, either. By Theorem 33, the former can be checked in polynomial time. Therefore, we may assume that S is necessarily complete.

Lemma 36 provides polynomial-time constructability for the ungrounded completion. Given a completion, GR-VERIFICATION can be solved in polynomial time, and Lemma 37 yields that the answer to GR-INCNV is the same as that to GR-VERIFICATION for the ungrounded completion. □

The following upper bounds then follow immediately.

Corollary 39. GR-ATTINCNV *and* GR-ARGINCNV *are in* P.

We have completed our proofs for P membership of necessary verification in all three incompleteness models for the admissible, stable, complete, and grounded semantics. In Theorems 40 and 41, using the notions of optimistic completion (Definition 7) and fixed completion (Definition 11), respectively, we prove that possible verification can also be efficiently decided for these four semantics in the attack-incomplete model.

Theorem 40. *For* **s** ∈ {AD, ST}, **s**-ATTINCPV *is in* P.

Proof. The optimistic completion can obviously be constructed in polynomial time. As already mentioned, the problem **s**-VERIFICATION can be solved in polynomial time for a given completion. Proposition 9 then provides that the answer to **s**-ATTINCPV is the same as that to **s**-VERIFICATION for the optimistic completion. □

Theorem 41. *For* **s** ∈ {CP, GR}, **s**-ATTINCPV *is in* P.

Proof. Propositions 13 provides polynomial-time constructability for the fixed completion. Given a completion, **s**-VERIFICATION can be solved in polynomial time, and Proposition 14 implies that the answer to **s**-ATTINCPV is the same as that to **s**-VERIFICATION for the fixed completion. □

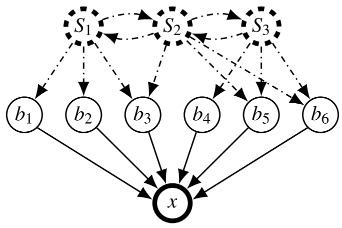


Fig. 9. A yes-instance of AD-ARGINCPV created from a yes-instance of X3C.

4.2. Lower bounds

In this section, we prove tight lower bounds for all remaining cases.

First, by a straightforward reduction from the VERIFICATION problem for standard argumentation frameworks, we observe in Corollary 42 that the upper bounds from the previous section coincide with the lower bounds for PR-ATTINCNV, PR-ARGINCNV, and PR-INCNV.

Corollary 42. *PR-ATTINCNV, PR-ARGINCNV, and PR-INCNV are coNP-hard.*

Next, we present results for possible verification, where introducing argument incompleteness raises the complexity from P to NP-completeness for the admissible, stable, complete, and grounded semantics.

Theorem 43. *AD-ARGINCPV is NP-hard.*

Proof. To show NP-hardness, we reduce from the following NP-complete problem (see, e.g., the book by Garey and Johnson [32]):

	EXACT-COVER-BY-3-SETS (X3C)
Given:	A set $B = \{b_1, \dots, b_{3k}\}$ and a family $\mathscr{S}$ of subsets of B, with $\|S_j\| = 3$ for all $S_j \in \mathscr{S}$.
Question:	Does there exist a subfamily $\mathscr{S}' \subseteq \mathscr{S}$ of size k that exactly covers B, i.e., $\bigcup_{S_j \in \mathscr{S}'} S_j = B$?

Given an instance $(B, \mathscr{S}) = (\{b_1, \dots, b_{3k}\}, \{S_1, \dots, S_m\})$ of X3C, we construct an instance $(\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}\rangle, S)$ of AD-ARGINCPV as follows (where we slightly abuse notation and use the same identifiers for both instances; it will always be clear from the context, though, which instance an element belongs to):

$$\begin{aligned} \mathscr{A} &= \{x\} \cup B, \\ \mathscr{A}^? &= \mathscr{S}, \\ \mathscr{R} &= \{(b_i, x) \mid b_i \in B\} \cup \{(S_j, b_{j_1}), (S_j, b_{j_2}), (S_j, b_{j_3}) \mid S_j = \{b_{j_1}, b_{j_2}, b_{j_3}\} \in \mathscr{S}\} \cup \\ &\quad \{(S_i, S_j), (S_j, S_i) \mid S_i, S_j \in \mathscr{S} \text{ and } S_i \cap S_j \neq \emptyset\}, \\ S &= \{x\} \cup \mathscr{S}. \end{aligned}$$

In particular, $\mathscr{A} \cup \mathscr{A}^?$ contains one argument b_i for every element $b_i \in B$, $1 \le i \le 3k$, one argument S_j for every set S_j in $\mathscr{S}$, $1 \le j \le m$, and one additional argument x. All arguments corresponding to elements of B attack x, and each argument S_j attacks the three arguments corresponding to those elements of B that belong to S_j in $\mathscr{S}$. Additionally, there are attacks between S_i and S_j if the corresponding sets in $\mathscr{S}$ are not disjoint. Finally, $\mathscr{A}$ and S act as opponents: x belongs to both, but the arguments corresponding to elements in B belong to $\mathscr{A}$ only, whereas the arguments corresponding to the sets in $\mathscr{S}$ belong to S only.

Let us give two examples of this construction resulting from two distinct X3C instances, $(B, \mathscr{S}_1)$ and $(B, \mathscr{S}_2)$, with $B = \{b_1, \dots, b_6\}$. On the one hand, Fig. 9 shows a yes-instance of AD-ARGINCPV created from a yes-instance of X3C: $(B, \mathscr{S}_1)$ with $\mathscr{S}_1 = \{\{b_1, b_2, b_3\}, \{b_3, b_5, b_6\}, \{b_4, b_5, b_6\}\}$. On the other hand, Fig. 10 shows a no-instance of AD-ARGINCPV created from a no-instance of X3C: $(B, \mathscr{S}_2)$ with $\mathscr{S}_2 = \{\{b_1, b_2, b_3\}, \{b_3, b_5, b_6\}, \{b_2, b_4, b_6\}\}$. In both figures, $\mathscr{A}$ contains the solid arguments, the dashed arguments belong to $\mathscr{A}^?$, and the boldfaced arguments are part of S.

We claim that $(B, \mathscr{S}) \in$ X3C if and only if $(\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}\rangle, S) \in$ AD-ARGINCPV.

($\Longrightarrow$) Clearly, if $(B, \mathscr{S})$ is a yes-instance of X3C, we can add exactly those arguments S_i to $\mathscr{A}$ that correspond to an exact cover of B. Let $\mathscr{A}^*$ be the argument set of this completion. In $\mathscr{A}^*$, every b_i, $1 \le i \le 3k$, is attacked by exactly one argument S_j, $1 \le j \le m$, due to the exact cover. Hence, $x \in S|_{\mathscr{A}^*}$ is defended against every attack. Additionally, the

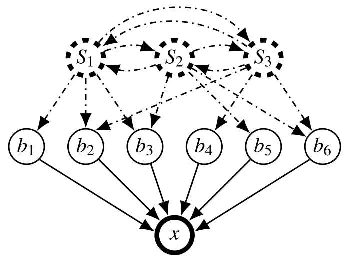


Fig. 10. A no-instance of AD-ARGINCPV created from a no-instance of X3C.

arguments S_j in $\mathscr{A}^*$ have no attacks between them, because the corresponding sets are pairwise disjoint, which implies that no new attacks on the elements of $S|_{\mathscr{A}^*}$ are introduced. But this means that $S|_{\mathscr{A}^*}$ is admissible in $\langle \mathscr{A}^*, \mathscr{R}|_{\mathscr{A}^*}\rangle$.

($\Longleftarrow$) If there is a completion with the argument set $\mathscr{A}^*$, this completion must defend x against every b_i, $1 \le i \le 3k$. This means that there must exist a cover of the elements of B by the sets of $\mathscr{S}$. But because the arguments S_j attack each other whenever they are not disjoint, this cover must be exact; otherwise, the set $S|_{\mathscr{A}^*}$ would not be conflict-free. Hence, there exists an exact cover of B. $\square$

Theorem 44. *For* **s** $\in$ {ST, CP, GR}, **s**-ARGINCPV *is* NP-*hard.*

Proof. We show NP-hardness for all three problems by showing that the reduction used in Theorem 43 also works for the stable, complete, and grounded semantics. To this end, we will prove that the following four statements are pairwise equivalent for the instance $(\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}\rangle, S)$ constructed in the proof of Theorem 43:

- $(\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}\rangle, S) \in$ AD-ARGINCPV,
- $(\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}\rangle, S) \in$ ST-ARGINCPV,
- $(\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}\rangle, S) \in$ GR-ARGINCPV, and
- $(\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}\rangle, S) \in$ CP-ARGINCPV.

$(\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}\rangle, S) \in$ AD-ARGINCPV implies $(\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}\rangle, S) \in$ ST-ARGINCPV: If $S|_{\mathscr{A}^*}$ is admissible for a completion $\langle \mathscr{A}^*, \mathscr{R}|_{\mathscr{A}^*}\rangle$, it is, in particular, conflict-free. We know from the reduction that $\langle \mathscr{A}^*, \mathscr{R}|_{\mathscr{A}^*}\rangle$ only contains arguments S_j that do not attack each other, and all these arguments belong to $S|_{\mathscr{A}^*}$. Hence, the only arguments outside of $S|_{\mathscr{A}^*}$ are the b_is. But all of them are attacked, as explained in the proof of Theorem 43. Therefore, $S|_{\mathscr{A}^*}$ is a stable extension of $\langle \mathscr{A}^*, \mathscr{R}|_{\mathscr{A}^*}\rangle$.

$(\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}\rangle, S) \in$ ST-ARGINCPV implies $(\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R}\rangle, S) \in$ GR-ARGINCPV: If $S|_{\mathscr{A}^*}$ is stable for a completion $\langle \mathscr{A}^*, \mathscr{R}|_{\mathscr{A}^*}\rangle$, it must contain all arguments of $\mathscr{A}^*$ except for the b_is. As every stable extension is conflict-free, there are no attacks between arguments that correspond to an S_j. This means for the characteristic function of this completion $\langle \mathscr{A}^*, \mathscr{R}|_{\mathscr{A}^*}\rangle$ that the output of the first step is the set that contains exactly those S_j. In the second step, we add argument x, because all those S_j defend x against all attacks from the arguments b_i. No new arguments are added in step three. Therefore, this set is the grounded extension of the argumentation framework $\langle \mathscr{A}^*, \mathscr{R}|_{\mathscr{A}^*}\rangle$. But this set is exactly the set $S|_{\mathscr{A}^*}$. Hence, $S|_{\mathscr{A}^*}$ is the grounded extension of $\langle \mathscr{A}^*, \mathscr{R}|_{\mathscr{A}^*}\rangle$.

It is easy to see the two remaining implications needed to prove the equivalences: Every grounded set is complete, and every complete set is admissible. This completes the proof. $\square$

The previous hardness results carry over to the general model and coincide with the respective upper bounds from Theorem 26.

Corollary 45. *For* **s** $\in$ {AD, ST, CP, GR}, **s**-INCPV *is* NP-*complete.*

Our final results show that the complexity of possible verification for the preferred semantics raises from coNP-hardness to Σ_2^p-completeness in all three models.

Theorem 46. PR-ATTINCPV *is* Σ_2^p-*hard.*

Proof. First, we quickly recall some notation from propositional logic. A boolean variable x has two literals, x and $\neg x$. A boolean formula is in conjunctive normal form (CNF) if it is a conjunction of disjunctions of literals (clauses), and in disjunctive normal form (DNF) if it is a disjunction of conjunctive clauses of literals. 3-CNF (respectively, 3-DNF) denotes CNF (respectively, DNF) with at most three literals per clause. A truth assignment τ on a set X of variables is a function

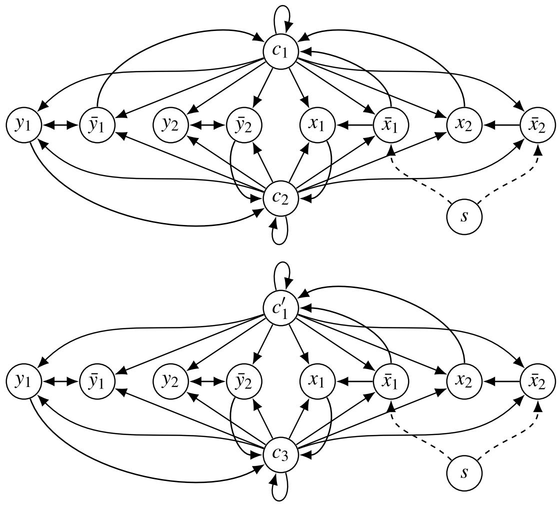


Fig. 11. Graph representations of the attack-incomplete argumentation frameworks created from clauses $c_2 = (x_1 \vee y_1 \vee \neg y_2)$ and either $c_1 = (\neg x_1 \vee x_2 \vee \neg y_1)$ (top) or $c'_1 = (\neg x_1 \vee x_2)$ (bottom) following the construction in the proof of Theorem 46. Dashed attacks indicate uncertainty as usual. The first instance is a no-instance of PR-ATTINCPV, the second is a yes-instance.

$\tau : X \to \{\texttt{true}, \texttt{false}\}$. For a formula φ and truth assignments $\tau_1, \tau_2, \dots, \tau_k$ on disjoint sets of variables, $\varphi[\tau_1, \tau_2, \dots, \tau_k]$ denotes the formula obtained by replacing variables in φ with their truth values in $\tau_1, \tau_2, \dots, \tau_k$.

To prove Σ_2^p-hardness, we reduce from the quantified satisfiability problem Σ_2SAT, which is well known to be complete for Σ_2^p (see [44]):

	Σ_2SAT
Given:	A 3-DNF formula φ on two disjoint sets of variables, X and Y.
Question:	Does $\exists \tau_X \ \forall \tau_Y : \varphi[\tau_X, \tau_Y]$ evaluate to $\texttt{true}$ (where τ_X and τ_Y are truth assignments on X and Y, respectively)?

Let (φ, X, Y) be an instance of Σ_2SAT, where $X = \{x_1, \dots, x_{|X|}\}$ and $Y = \{y_1, \dots, y_{|Y|}\}$ are two disjoint sets of propositional variables and φ is a 3-DNF formula over $X \cup Y$. For $\bar{\varphi} = \neg\varphi$, the question in Σ_2SAT is equivalent to asking whether $\exists \tau_X \forall \tau_Y : \bar{\varphi}[\tau_X, \tau_Y] = \texttt{false}$, where $\bar{\varphi} = c_1 \wedge \cdots \wedge c_m$ is a formula in 3-CNF with clauses c_1 through c_m. From now on, we will mostly use this CNF formulation of the problem.

We create an instance $(\langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle, S)$ of PR-ATTINCPV from (φ, X, Y) as follows (see Fig. 11 for an example):

$$\mathscr{A} = \left\{ \begin{array}{ll} y_i, \bar{y}_i, & \text{for } y_i \in Y \\ x_i, \bar{x}_i, & \text{for } x_i \in X \\ c_i, & \text{for } c_i \text{ in } \bar{\varphi} \\ s & \end{array} \right\},$$

$$\mathscr{R} = \left\{ \begin{array}{ll} (\bar{y}_i, y_i), (y_i, \bar{y}_i), & \text{for } y_i \in Y \\ (\bar{x}_i, x_i), & \text{for } x_i \in X \\ (c_i, c_i), & \text{for } c_i \text{ in } \bar{\varphi} \\ (c_i, y_j), (c_i, \bar{y}_j), & \text{for } c_i \text{ in } \bar{\varphi}, y_j \in Y \\ (c_i, x_k), (c_i, \bar{x}_k), & \text{for } c_i \text{ in } \bar{\varphi}, x_k \in X \\ (y_j, c_i), & \text{if } y_j \text{ in } c_i \\ (\bar{y}_j, c_i), & \text{if } \neg y_j \text{ in } c_i \\ (x_k, c_i), & \text{if } x_k \text{ in } c_i \\ (\bar{x}_k, c_i), & \text{if } \neg x_k \text{ in } c_i \end{array} \right\},$$

$$\mathscr{R}^? = \left\{ \ (s, \bar{x}_i), \ \text{for } x_i \in X \ \right\}.$$

Finally, let $S = \{s\}$. We call all arguments x_i, $\bar{x}_i$, y_i, and $\bar{y}_i$ *literal arguments* and arguments c_i *clause arguments*. Note that S is necessarily admissible in $\langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$, so the verification of possible preferredness boils down to checking whether all supersets of S are nonadmissible in some completion of $\langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$.

We prove that

$$(\varphi, X, Y) \in \Sigma_2\text{SAT} \iff (\langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle, S) \in \text{PR-ATTINCPV}.$$

Assume that $(\varphi, X, Y) \in \Sigma_2$SAT, i.e., $\exists \tau_X \forall \tau_Y : \bar{\varphi}[\tau_X, \tau_Y] = \texttt{false}$. Let τ_X be an assignment of truth values to the variables in X that satisfies $\forall \tau_Y : \bar{\varphi}[\tau_X, \tau_Y] = \texttt{false}$. Let $\langle \mathscr{A}, \mathscr{R}^{\tau_X} \rangle$ be the completion of $\langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$ obtained by letting $\mathscr{R}^{\tau_X} = \mathscr{R} \cup \{(s, \bar{x}_i) \in \mathscr{R}^? \mid \tau_X(x_i) = \texttt{true}\}$. In $\langle \mathscr{A}, \mathscr{R}^{\tau_X} \rangle$, the assignment τ_X to the variables in X is translated to a commitment on literal arguments: If, for $x_i \in X$, $\tau_X(x_i) = \texttt{true}$, then the attack by s against argument $\bar{x}_i$ is included and $\bar{x}_i$ can no longer be a member of admissible supersets of S, while argument x_i is defended by s and potentially can be such a member. On the other hand, if $\tau_X(x_i) = \texttt{false}$, the attack is excluded and the roles are switched: Argument x_i cannot be defended against argument $\bar{x}_i$ by S (or any conflict-free superset of S), so x_i cannot be contained in admissible supersets of S, whereas $\bar{x}_i$ can.

Now let τ_Y be any truth assignment for Y. We know that $\bar{\varphi}[\tau_X, \tau_Y] = \texttt{false}$. Transform τ_X and τ_Y to a set $S_{(\tau_X, \tau_Y)} \supset S$ of arguments by letting

$$S_{(\tau_X, \tau_Y)} = S \cup \{x_i \mid \tau_X(x_i) = \texttt{true}\} \cup \{\bar{x}_i \mid \tau_X(x_i) = \texttt{false}\}$$
$$\cup \{y_i \mid \tau_Y(y_i) = \texttt{true}\} \cup \{\bar{y}_i \mid \tau_Y(y_i) = \texttt{false}\}.$$

It is easy to see that $S_{(\tau_X, \tau_Y)}$ is conflict-free in $\langle \mathscr{A}, \mathscr{R}^{\tau_X} \rangle$. However, $S_{(\tau_X, \tau_Y)}$ cannot defend itself against all clause arguments $c_1, \dots, c_m$ in $\langle \mathscr{A}, \mathscr{R}^{\tau_X} \rangle$, and therefore is not admissible: Since $\bar{\varphi}$ is in CNF and $\bar{\varphi}[\tau_X, \tau_Y] = \texttt{false}$, at least one clause in $\bar{\varphi}$ is unfulfilled. Let c_j be any such clause. Since the clauses of $\bar{\varphi}$ are disjunctions of literals, all literals in c_j are unfulfilled. The only arguments in $\mathscr{A}$ that attack the clause argument c_j are the literal arguments whose corresponding literals appear in clause c_j. However, by construction, none of these arguments are in $S_{(\tau_X, \tau_Y)}$, since all these literals are $\texttt{false}$ in τ_X and τ_Y. Therefore, no argument in $S_{(\tau_X, \tau_Y)}$ attacks argument c_j. On the other hand, c_j attacks all literal arguments and therefore it attacks $S_{(\tau_X, \tau_Y)}$, which proves that $S_{(\tau_X, \tau_Y)}$ is not admissible in $\langle \mathscr{A}, \mathscr{R}^{\tau_X} \rangle$.

All conflict-free supersets of S are either the set $S_{(\tau_X, \tau_Y)}$ for some τ_Y or a subset of one of these. We proved that none of these can be admissible, and in consequence, that S is preferred in $\langle \mathscr{A}, \mathscr{R}^{\tau_X} \rangle$, so we have $(\langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle, S) \in$ PR-ATTINCPV.

For the other direction, assume that $(\varphi, X, Y) \notin \Sigma_2$SAT, i.e., $\forall \tau_X \exists \tau_Y : \bar{\varphi}[\tau_X, \tau_Y] = \texttt{true}$. Let τ_X be any assignment on X and let τ_Y be an assignment on Y that satisfies $\bar{\varphi}[\tau_X, \tau_Y] = \texttt{true}$. Create the completion $\langle \mathscr{A}, \mathscr{R}^{\tau_X} \rangle$ and the set $S_{(\tau_X, \tau_Y)}$ as before. Since $\bar{\varphi}[\tau_X, \tau_Y] = \texttt{true}$, all clauses in $\bar{\varphi}$ are fulfilled, which means that in each clause at least one literal must be fulfilled. Each such literal corresponds to a literal argument in $S_{(\tau_X, \tau_Y)}$, which attacks the corresponding clause argument. So, $S_{(\tau_X, \tau_Y)}$ is admissible, which shows that S is not preferred in $\langle \mathscr{A}, \mathscr{R}^{\tau_X} \rangle$, and since τ_X was generic, S is not preferred in any completion of $\langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle$, which proves $(\langle \mathscr{A}, \mathscr{R}, \mathscr{R}^? \rangle, S) \notin$ PR-ATTINCPV. □

Example 47. Consider a Σ_2SAT instance (φ, X, Y) with $X = \{x_1, x_2\}$, $Y = \{y_1, y_2\}$ and $\varphi = (x_1 \wedge \neg x_2 \wedge y_1) \vee (\neg x_1 \wedge \neg y_1 \wedge y_2)$. We have $\bar{\varphi} = \neg\varphi = c_1 \wedge c_2$ with $c_1 = (\neg x_1 \vee x_2 \vee \neg y_1)$ and $c_2 = (x_1 \vee y_1 \vee \neg y_2)$. We have $(\varphi, X, Y) \notin \Sigma_2$SAT, because for all assignments τ_X on X and the assignment τ_Y with $\tau_Y(y_1) = \texttt{false}$, $\tau_Y(y_2) = \texttt{false}$ we have $\varphi[\tau_X, \tau_Y] = \texttt{false}$, or, equivalently, $\bar{\varphi}[\tau_X, \tau_Y] = \texttt{true}$.

To create a yes-instance, we slightly modify this Σ_2SAT instance by setting $\varphi' = (x_1 \wedge \neg x_2) \vee (\neg x_1 \wedge \neg y_1 \wedge y_2)$, i.e., $\neg y_1$ is omitted in the first clause. We now have $\bar{\varphi}' = \neg\varphi' = c_1' \wedge c_2$, where $c_1' = (\neg x_1 \vee x_2)$, and $c_2 = (x_1 \vee y_1 \vee \neg y_2)$ is unchanged. (φ', X, Y) is a yes-instance of Σ_2SAT, because for the assignment τ_X' on X with $\tau_X'(x_1) = \texttt{true}$, $\tau_X'(x_2) = \texttt{false}$ and for all assignments τ_Y' on Y, we have $\varphi'[\tau_X', \tau_Y'] = \texttt{true}$, or, equivalently, $\bar{\varphi}'[\tau_X', \tau_Y'] = \texttt{false}$.

Fig. 11 shows the graph representations of two attack-incomplete argumentation frameworks that are created from (φ, X, Y) (top) and (φ', X, Y) (bottom) according to the construction in the proof of Theorem 46. The top graph together with the set $\{s\}$ constitutes a no-instance for PR-ATTINCPV. The set $\{s, \bar{y}_1, \bar{y}_2\}$ (corresponding to τ_Y from above) is an admissible superset of $\{s\}$ in all completions of the incomplete argumentation framework. The bottom graph together with the set $\{s\}$ constitutes a yes-instance for PR-ATTINCPV. The completion that corresponds to the assignment τ_X' as defined above includes the possible attack $(s, \bar{x}_1)$ and excludes the possible attack $(s, \bar{x}_2)$ In this completion, there are no admissible supersets of $\{s\}$ that counterattack c_1', so $\{s\}$ is preferred.

The same hardness can be proven for the argument-incomplete model.

Theorem 48. PR-ARGINCPV *is* Σ_2^p*-hard.*

Proof. Again, we reduce from Σ_2SAT using a very similar construction. Given an instance (φ, X, Y) of Σ_2SAT, we create an instance $(\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle, S)$ of PR-ARGINCPV by setting $S = \emptyset$ and:

$$\mathscr{A} = \left\{ \begin{array}{ll} y_i, \bar{y}_i, & \text{for } y_i \in Y \\ \bar{x}_i, & \text{for } x_i \in X \\ c_i, & \text{for } c_i \text{ in } \bar{\varphi} \end{array} \right\},$$

$$\mathscr{A}^? = \left\{ x_i, \ \text{for } x_i \in X \right\},$$

$$\mathscr{R} = \left\{ \begin{array}{ll} (\bar{y}_i, y_i), (y_i, \bar{y}_i), & \text{for } y_i \in Y \\ (x_i, \bar{x}_i), & \text{for } x_i \in X \\ (c_i, c_i), & \text{for } c_i \text{ in } \bar{\varphi} \\ (c_i, y_j), (c_i, \bar{y}_j), & \text{for } c_i \text{ in } \bar{\varphi}, y_j \in Y \\ (c_i, x_k), (c_i, \bar{x}_k), & \text{for } c_i \text{ in } \bar{\varphi}, x_k \in X \\ (y_j, c_i), & \text{if } y_j \text{ in } c_i \\ (\bar{y}_j, c_i), & \text{if } \neg y_j \text{ in } c_i \\ (x_k, c_i), & \text{if } x_k \text{ in } c_i \\ (\bar{x}_k, c_i), & \text{if } \neg x_k \text{ in } c_i \end{array} \right\}.$$

Again, S is necessarily admissible in $\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle$, so for the verification of possible preferredness it is enough to check whether there is a completion of $\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle$ where all supersets of S are nonadmissible.

We prove that

$$(\varphi, X, Y) \in \Sigma_2\text{SAT} \Longleftrightarrow (\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle, S) \in \text{PR-ARGINCPV}.$$

Assume that $(\varphi, X, Y) \in \Sigma_2\text{SAT}$, i.e., $\exists \tau_X \forall \tau_Y : \bar{\varphi}[\tau_X, \tau_Y] = \texttt{false}$. Let τ_X be an assignment of truth values to the variables in X that satisfies $\forall \tau_Y : \bar{\varphi}[\tau_X, \tau_Y] = \texttt{false}$. Let $\langle \mathscr{A}^{\tau_X}, \mathscr{R}|_{\mathscr{A}^{\tau_X}} \rangle$ be the completion of $\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle$ obtained by letting $\mathscr{A}^{\tau_X} = \mathscr{A} \cup \{x_i \in \mathscr{A}^? \mid \tau_X(x_i) = \texttt{true}\}$. In $\langle \mathscr{A}^{\tau_X}, \mathscr{R}|_{\mathscr{A}^{\tau_X}} \rangle$, the assignment τ_X to the variables in X is translated to a commitment on literal arguments: If, for $x_i \in X$, $\tau_X(x_i) = \texttt{true}$, then argument x_i is included in $\mathscr{A}^{\tau_X}$ and has an attack against argument $\bar{x}_i$ which S cannot defend, so x_i is a candidate for membership in admissible supersets of S and $\bar{x}_i$ is not. If $\tau_X(x_i) = \texttt{false}$, then x_i is excluded and does not attack $\bar{x}_i$, so $\bar{x}_i$ could be in admissible supersets of S.

Now let τ_Y be any truth assignment for Y. We know that $\bar{\varphi}[\tau_X, \tau_Y] = \texttt{false}$. Transform τ_X and τ_Y to a set $S_{(\tau_X, \tau_Y)} \supset S$ of arguments by letting

$$\begin{aligned} S_{(\tau_X, \tau_Y)} = S &\cup \{x_i \mid \tau_X(x_i) = \texttt{true}\} \cup \{\bar{x}_i \mid \tau_X(x_i) = \texttt{false}\} \\ &\cup \{y_i \mid \tau_Y(y_i) = \texttt{true}\} \cup \{\bar{y}_i \mid \tau_Y(y_i) = \texttt{false}\}. \end{aligned}$$

It is easy to see that $S_{(\tau_X, \tau_Y)}$ is conflict-free in $\langle \mathscr{A}^{\tau_X}, \mathscr{R}|_{\mathscr{A}^{\tau_X}} \rangle$. However, $S_{(\tau_X, \tau_Y)}$ cannot defend itself against all clause arguments $c_1, \ldots, c_m$ in $\langle \mathscr{A}^{\tau_X}, \mathscr{R}|_{\mathscr{A}^{\tau_X}} \rangle$, and therefore is not admissible: Since $\bar{\varphi}$ is in CNF and $\bar{\varphi}[\tau_X, \tau_Y] = \texttt{false}$, at least one clause in $\bar{\varphi}$ is unfulfilled. Let c_j be any such clause. Since the clauses of $\bar{\varphi}$ are disjunctions of literals, all literals in c_j are unfulfilled. The only arguments in $\mathscr{A}^{\tau_X}$ that attack the clause argument c_j are the literal arguments whose corresponding literals appear in clause c_j. However, by construction, none of these arguments are in $S_{(\tau_X, \tau_Y)}$, since all these literals are $\texttt{false}$ in τ_X and τ_Y. Therefore, no argument in $S_{(\tau_X, \tau_Y)}$ attacks argument c_j. On the other hand, c_j attacks all literal arguments and therefore it attacks $S_{(\tau_X, \tau_Y)}$, which proves that $S_{(\tau_X, \tau_Y)}$ is not admissible in $\langle \mathscr{A}^{\tau_X}, \mathscr{R}|_{\mathscr{A}^{\tau_X}} \rangle$.

All conflict-free supersets of S are either the set $S_{(\tau_X, \tau_Y)}$ for some τ_Y or a subset of one of these. We proved that none of these can be admissible, and in consequence, that S is preferred in $\langle \mathscr{A}^{\tau_X}, \mathscr{R}|_{\mathscr{A}^{\tau_X}} \rangle$, so we have $(\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle, S) \in$ PR-ARGINCPV.

For the other direction, assume that $(\varphi, X, Y) \notin \Sigma_2\text{SAT}$, i.e., $\forall \tau_X \exists \tau_Y : \bar{\varphi}[\tau_X, \tau_Y] = \texttt{true}$. Let τ_X be any assignment on X and let τ_Y be an assignment on Y that satisfies $\bar{\varphi}[\tau_X, \tau_Y] = \texttt{true}$. Create the completion $\langle \mathscr{A}^{\tau_X}, \mathscr{R}|_{\mathscr{A}^{\tau_X}} \rangle$ and the set $S_{(\tau_X, \tau_Y)}$ as before. Since $\bar{\varphi}[\tau_X, \tau_Y] = \texttt{true}$, all clauses in $\bar{\varphi}$ are fulfilled, which means that in each clause at least one literal must be fulfilled. Each such literal corresponds to a literal argument in $S_{(\tau_X, \tau_Y)}$, which attacks the corresponding clause argument. So, $S_{(\tau_X, \tau_Y)}$ is admissible, which shows that S is not preferred in $\langle \mathscr{A}^{\tau_X}, \mathscr{R}|_{\mathscr{A}^{\tau_X}} \rangle$, and since τ_X was generic, S is not preferred in any completion of $\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle$, which proves $(\langle \mathscr{A}, \mathscr{A}^?, \mathscr{R} \rangle, S) \notin$ PR-ARGINCPV. □

Example 49. Fig. 12 shows the graph representations of two argument-incomplete argumentation frameworks that are created from the same Σ_2SAT instances (φ, X, Y) (top) and (φ', X, Y) (bottom) used in Example 47. Here, the set $\emptyset$ constitutes a no-instance for PR-ARGINCPV together with the top graph and a yes-instance together with the bottom graph. In the no-instance, the set $\{\bar{y}_1, \bar{y}_2\}$ is an admissible superset of $\emptyset$ in all completions of the incomplete argumentation framework. In the yes-instance, the completion that includes the possible argument x_1 and excludes the possible argument x_2 has no admissible supersets of $\emptyset$ that counterattack c'_1, so $\emptyset$ is preferred.

Both previous results also provide Σ_2^p-hardness for the problem PR-INCPV in the general model, which completes our complexity analysis.

Corollary 50. PR-INCPV *is* Σ_2^p*-hard.*

5. Conclusion and future work

We introduced three specific models of incompleteness in argumentation frameworks, one focusing on *attack incompleteness* alone, one on *argument incompleteness* alone, and one that combines these two models so as to provide a *general model*

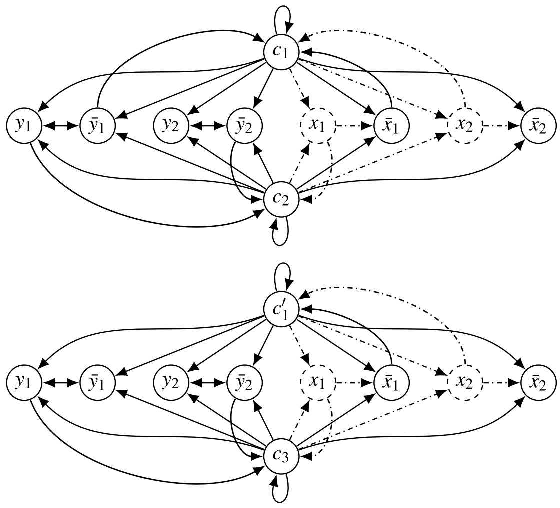


Fig. 12. Graph representations of the argument-incomplete argumentation frameworks created from clauses $c_2 = (x_1 \vee y_1 \vee \neg y_2)$ and either $c_1 = (\neg x_1 \vee x_2 \vee \neg y_1)$ (top) or $c'_1 = (\neg x_1 \vee x_2)$ (bottom) following the construction in the proof of Theorem 48. Dashed arguments indicate uncertainty as usual, and conditionally definite attacks are dash-dotted as usual. The first instance is a no-instance of PR-ARGINCPV, the second is a yes-instance.

Table 1
Overview of complexity results for various semantics (first column) in the argumentation framework model without uncertainty (second column), with results marked by ♠ due to Dung [26] and the result marked by ♣ due to Dimopoulos and Torres [24]; in the attack-incomplete model (third and sixth column) from Section 3.1, with results marked by ★ due to Coste-Marquis et al. [19]; in the argument-incomplete model (fourth and seventh column) from Section 3.2; and in the combined model (fifth and eighth column) from Section 3.3. Key: For a complexity class $\mathscr{C}$, $\mathscr{C}$-c. stands for $\mathscr{C}$-completeness and VER is a shorthand for VERIFICATION.

s	VER	ATTINCNV	ARGINCNV	INCNV	ATTINCPV	ARGINCPV	INCPV
CF	in P♠	in P★	in P	in P	in P★	in P	in P
AD	in P♠	in P★	in P	in P	in P	NP-c.	NP-c.
ST	in P♠	in P	in P	in P	in P	NP-c.	NP-c.
CP	in P♠	in P	in P	in P	in P	NP-c.	NP-c.
GR	in P♠	in P	in P	in P	in P	NP-c.	NP-c.
PR	coNP-c.♣	coNP-c.	coNP-c.	coNP-c.	Σ_2^p-c.	Σ_2^p-c.	Σ_2^p-c.

of incompleteness. We then have studied the computational complexity of two variants of the verification problem, one formalizing the *possibility* of completing the given incomplete state and the other formalizing the *necessity* of completion, both with respect to six common semantics of argumentation frameworks.

Table 1 gives an overview of the complexity results for the verification problem in the standard model and in the three incompleteness models considered in this paper. The complexity results show a pattern in how introducing incomplete information affects the complexity of the verification problem in abstract argumentation frameworks. We observe that there are only two triggers for an increase of complexity: the preferred semantics for possible verification in all three models, and the admissible semantics (along with all other semantics that entail admissibility) for possible verification in the model of argument incompleteness (and, therefore, also in the general incompleteness model). In all other cases—in particular, for all variants of necessary verification—introducing incomplete information does not make the verification problem computationally harder. Note that each of our hardness results for verification problems carries over to any more general model; so our approach is potentially useful in other frameworks as well. We further note that the Σ_2^p-completeness results for possible verification in the preferred semantics are significantly more severe than the NP- or coNP-completeness results for possible verification in the other semantics entailing admissibility and for standard or necessary verification in the preferred semantics: The known methods to circumvent NP- or coNP-hardness in practice (e.g., by using fast SAT-solvers) are much more efficient than those known to tame Σ_2^p-hardness (e.g., by using QBF solvers). Nevertheless, there are some approaches to tackle problems on the second level of the polynomial-time hierarchy—especially in the field of argumentation (see, e.g., the work of Thimm and Villata [45] on the first competition on computational models of argumentation)—that can be adapted to our problem.

To put the above results into a bigger picture, let us briefly compare the complexity of the verification problem in incomplete argumentation frameworks with the complexity of other computational tasks, namely that of credulous and skeptical acceptance of arguments in incomplete argumentation frameworks. Extending previous work by Coste-Marquis [20], Dimopoulos and Torres [24], and Dunne and Bench-Capon [28], Baumeister et al. [7] have recently settled, for the same six semantics considered in this paper, the complexity of the problems related to credulous and skeptical acceptance of arguments in standard argumentation frameworks as well as for their possible and necessary variants in incomplete argumentation frameworks. While most entries in Table 1 for verification are P results (with the exception of the coNP- and Σ_2^p-completeness results for the preferred semantics and of the NP-completeness results for $\mathbf{s} \in \{\text{AD}, \text{ST}, \text{CP}, \text{GR}\}$ in (argument-)incomplete argumentation frameworks), the complexity results for possible/necessary credulous/skeptical acceptance are much more varied, ranging from P membership to completeness in the third level of the polynomial hierarchy and all its intermediate levels: completeness for NP, coNP, Σ_2^p, Π_2^p, and even $\Sigma_3^p = \mathrm{NP}^{\mathrm{NP}^{\mathrm{NP}}}$.

A task for future work is to analyze the complexity of possible and necessary variants of other decision problems than verification or credulous or skeptical acceptance of individual arguments. Also, the range of classical semantics considered here could be extended by including other, more recently proposed semantics like the stage semantics [47], the semi-stable semantics [15], the ideal semantics [27], or the CF2 semantics [1].

Acknowledgements

This paper merges and extends two preliminary ADT'15 papers [6,11], a COMSOC'16 workshop paper, and an AAAI'18 paper [8], and we thank these reviewers as well as the AIJ reviewers for their helpful comments. This work was supported in part by an NRW grant for gender-sensitive universities and the project "Online Participation," both funded by the NRW Ministry for Innovation, Science, and Research.

References

[1] P. Baroni, M. Giacomin, G. Guida, SCC-recursiveness: a general schema for argumentation semantics, Artif. Intell. 168 (1–2) (2005) 162–210.
[2] R. Baumann, What does it take to enforce an argument? Minimal change in abstract argumentation, in: Proceedings of the 20th European Conference on Artificial Intelligence, in: Front. Artif. Intell. Appl., vol. 242, IOS Press, 2012, pp. 127–132.
[3] R. Baumann, G. Brewka, Expanding argumentation frameworks: enforcing and monotonicity results, in: Proceedings of the 3rd International Conference on Computational Models of Argument, in: Front. Artif. Intell. Appl., vol. 216, IOS Press, September 2010, pp. 75–86.
[4] D. Baumeister, S. Bouveret, J. Lang, N. Nguyen, T. Nguyen, J. Rothe, A. Saffidine, Positional scoring-based allocation of indivisible goods, Auton. Agents Multi-Agent Syst. 31 (3) (2017) 628–655.
[5] D. Baumeister, G. Erdélyi, O. Erdélyi, J. Rothe, Complexity of manipulation and bribery in judgment aggregation for uniform premise-based quota rules, Math. Soc. Sci. 76 (2015) 19–30.
[6] D. Baumeister, D. Neugebauer, J. Rothe, Verification in attack-incomplete argumentation frameworks, in: Proceedings of the 4th International Conference on Algorithmic Decision Theory, in: Lect. Notes Artif. Intell., vol. 9346, Springer-Verlag, September 2015, pp. 341–358.
[7] D. Baumeister, D. Neugebauer, J. Rothe, Credulous and skeptical acceptance in incomplete argumentation frameworks, in: Proceedings of the 7th International Conference on Computational Models of Argument, in: Frontiers in Artificial Intelligence and Applications, IOS Press, September 2018, in press.
[8] D. Baumeister, D. Neugebauer, J. Rothe, H. Schadrack, Complexity of verification in incomplete argumentation frameworks, in: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, AAAI Press, February 2018.
[9] D. Baumeister, M. Roos, J. Rothe, Computational complexity of two variants of the possible winner problem, in: Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems, IFAAMAS, May 2011, pp. 853–860.
[10] D. Baumeister, M. Roos, J. Rothe, L. Schend, L. Xia, The possible winner problem with uncertain weights, in: Proceedings of the 20th European Conference on Artificial Intelligence, in: Front. Artif. Intell. Appl., vol. 242, IOS Press, August 2012, pp. 133–138.
[11] D. Baumeister, J. Rothe, H. Schadrack, Verification in argument-incomplete argumentation frameworks, in: Proceedings of the 4th International Conference on Algorithmic Decision Theory, in: Lect. Notes Artif. Intell., vol. 9346, Springer-Verlag, September 2015, pp. 359–376.
[12] T. Bench-Capon, P. Dunne, Argumentation in artificial intelligence, Artif. Intell. 171 (10–15) (2007) 619–641.
[13] G. Boella, S. Kaci, L. van der Torre, Dynamics in argumentation with single extensions: abstraction principles and the grounded extension, in: Proceedings of the 10th European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty, in: Lect. Notes Artif. Intell., vol. 5590, Springer-Verlag, July 2009, pp. 107–118.
[14] S. Bouveret, U. Endriss, J. Lang, Fair division under ordinal preferences: computing envy-free allocations of indivisible goods, in: Proceedings of the 19th European Conference on Artificial Intelligence, in: Front. Artif. Intell. Appl., vol. 215, IOS Press, August 2010, pp. 387–392.
[15] M. Caminada, Semi-stable semantics, in: Proceedings of the 1st International Conference on Computational Models of Argument, in: Front. Artif. Intell. Appl., vol. 144, IOS Press, September 2006, pp. 121–130.
[16] C. Cayrol, C. Devred, M. Lagasquie-Schiex, Handling ignorance in argumentation: semantics of partial argumentation frameworks, in: Proceedings of the 9th European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty, in: Lect. Notes Artif. Intell., vol. 4724, Springer-Verlag, November 2007, pp. 259–270.
[17] C. Cayrol, F. de Saint-Cyr, M. Lagasquie-Schiex, Change in abstract argumentation frameworks: adding an argument, J. Artif. Intell. Res. 38 (2010) 49–84.
[18] Y. Chevaleyre, J. Lang, N. Maudet, J. Monnot, L. Xia, New candidates welcome! Possible winners with respect to the addition of new candidates, Math. Soc. Sci. 64 (1) (2012) 74–88.
[19] S. Coste-Marquis, C. Devred, S. Konieczny, M. Lagasquie-Schiex, P. Marquis, On the merging of Dung's argumentation systems, Artif. Intell. 171 (10) (2007) 730–753.
[20] S. Coste-Marquis, C. Devred, P. Marquis, Symmetric argumentation frameworks, in: Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty, in: Lect. Notes Artif. Intell., vol. 3571, Springer-Verlag, July 2005, pp. 317–328.
[21] S. Coste-Marquis, S. Konieczny, J. Mailly, P. Marquis, On the revision of argumentation systems: minimal change of arguments statuses, in: Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning, AAAI Press, July 2014, pp. 52–61.
[22] S. Coste-Marquis, S. Konieczny, J. Mailly, P. Marquis, Extension enforcement in abstract argumentation as an optimization problem, in: Proceedings of the 24th International Joint Conference on Artificial Intelligence, AAAI Press/IJCAI, July 2015, pp. 2876–2882.
[23] J. Delobelle, A. Haret, S. Konieczny, J. Mailly, J. Rossit, S. Woltran, Merging of abstract argumentation frameworks, in: Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning, AAAI Press, April 2016, pp. 33–42.
[24] Y. Dimopoulos, A. Torres, Graph theoretical structures in logic programs and default theories, Theor. Comput. Sci. 170 (1) (1996) 209–244.
[25] D. Doder, S. Woltran, Probabilistic argumentation frameworks – a logical approach, in: Proceedings of the 8th International Conference on Scalable Uncertainty Management, in: Lect. Notes Artif. Intell., vol. 8720, Springer-Verlag, September 2014, pp. 134–147.

[26] P. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artif. Intell. 77 (2) (1995) 321–357.

[27] P. Dung, P. Mancarella, F. Toni, A dialectic procedure for sceptical, assumption-based argumentation, in: Proceedings of the 1st International Conference on Computational Models of Argument, in: Front. Artif. Intell. Appl., vol. 144, IOS Press, September 2006, pp. 145–156.

[28] P. Dunne, T. Bench-Capon, Coherence in finite argument systems, Artif. Intell. 141 (1) (2002) 187–203.

[29] P. Dunne, M. Wooldridge, Complexity of abstract argumentation, in: I. Rahwan, G. Simari (Eds.), Argumentation in Artificial Intelligence, Springer, 2009, pp. 85–104, chapter 5.

[30] B. Fazzinga, S. Flesca, F. Parisi, Efficiently estimating the probability of extensions in abstract argumentation, in: Proceedings of the 7th International Conference on Scalable Uncertainty Management, in: Lect. Notes Artif. Intell., vol. 8078, Springer-Verlag, 2013, pp. 106–119.

[31] B. Fazzinga, S. Flesca, F. Parisi, On the complexity of probabilistic abstract argumentation, in: Proceedings of the 23rd International Joint Conference on Artificial Intelligence, AAAI Press/IJCAI, August 2013, pp. 898–904.

[32] M. Garey, D. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman and Company, 1979.

[33] A. Hunter, Probabilistic qualification of attack in abstract argumentation, Int. J. Approx. Reason. 55 (2) (2014) 607–638.

[34] K. Konczak, J. Lang, Voting procedures with incomplete preferences, in: Proceedings of the Multidisciplinary IJCAI-05 Workshop on Advances in Preference Handling, July/August 2005, pp. 124–129.

[35] J. Lang, A. Rey, J. Rothe, H. Schadrack, L. Schend, Representing and solving hedonic games with ordinal preferences and thresholds, in: Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems, IFAAMAS, May 2015, pp. 1229–1237.

[36] H. Li, N. Oren, T. Norman, Probabilistic argumentation frameworks, in: Proceedings of the 1st International Workshop on Theory and Applications of Formal Argumentation, in: Lect. Notes Artif. Intell., vol. 7132, Springer-Verlag, July 2011, pp. 1–16.

[37] B. Liao, L. Jin, R. Koons, Dynamics of argumentation systems: a division-based method, Artif. Intell. 175 (11) (2011) 1790–1814.

[38] M. Maher, Resistance to corruption of strategic argumentation, in: Proceedings of the 30th AAAI Conference on Artificial Intelligence, AAAI Press, February 2016, pp. 1030–1036.

[39] A. Meyer, L. Stockmeyer, The equivalence problem for regular expressions with squaring requires exponential space, in: Proceedings of the 13th IEEE Symposium on Switching and Automata Theory, IEEE Computer Society Press, 1972, pp. 125–129.

[40] C. Papadimitriou, Computational Complexity, 2nd edition, Addison-Wesley, 1995, Reprinted with corrections.

[41] I. Rahwan, G. Simari (Eds.), Argumentation in Artificial Intelligence, Springer, 2009.

[42] T. Rienstra, Towards a probabilistic Dung-style argumentation system, in: Proceedings of the 1st International Conference on Agreement Technologies, in: CEUR Workshop Proc., October 2012, pp. 138–152.

[43] J. Rothe, Complexity Theory and Cryptology. An Introduction to Cryptocomplexity, Texts Theoret. Comput. Sci. EATCS Ser., Springer-Verlag, 2005.

[44] L. Stockmeyer, The polynomial-time hierarchy, Theor. Comput. Sci. 3 (1) (1976) 1–22.

[45] M. Thimm, S. Villata, The first international competition on computational models of argumentation: results and analysis, Artif. Intell. 252 (2017) 267–294.

[46] F. Tohmé, G. Bodanza, G. Simari, Aggregation of attack relations: a social-choice theoretical analysis of defeasibility criteria, in: Proceedings of the 5th International Symposium on Foundations of Information and Knowledge Systems, in: Lect. Notes Artif. Intell., vol. 4932, Springer-Verlag, February 2008, pp. 8–23.

[47] B. Verheij, Two approaches to dialectical argumentation: admissible sets and argumentation stages, in: Proceedings of the 8th Dutch Conference on Artificial Intelligence, 1996, pp. 357–368.

[48] J. Wallner, A. Niskanen, M. Järvisalo, Complexity results and algorithms for extension enforcement in abstract argumentation, J. Artif. Intell. Res. 60 (2017) 1–40.

[49] L. Xia, V. Conitzer, Determining possible and necessary winners given partial orders, J. Artif. Intell. Res. 41 (2011) 25–67.

CHAPTER 7

Conclusion

We have studied the behavior of agents from two different perspectives, hedonic games and abstract argumentation. In Chapter 3 we tried to close the open gap for a specific question in hedonic games: How hard is it to decide, whether for a given enemy-oriented hedonic game the strict core is nonempty (SCSCS)? As hedonic games are often represented by graphs, a second, very closely related question, crossed our way: How hard is it to decide for a given undirected graph, whether there exists a wonderfully stable partition of the vertices (WSPE)? It turned out, that both problems are closely related, yet not easily reducible to each other. However, it is possible to use similar techniques to prove similar results in both settings. This leads to hardness results for NP, coNP and finally DP for both problems, SCSCS and WSPE. As side effects, we were able to prove completeness results for WSPV and k-WSPE. The former is the verification version of WSPE, in which we ask whether a given partition is wonderfully stable in a given undirected graph. The latter is WSPE restricted to graphs in which all vertices have the same fixed clique number k. As both problems, WSPE and SCSCS, are essentially equivalent for each such class of restricted graphs, we could directly derive NP-completeness for k-SCSCS, the corresponding version of k-WSPE in the setting of hedonic games. As upper bounds have already been known (Θ_2^p for WSPE and Σ_2^p for SCSCS) the goal was to close the gap between DP and the respective upper bound even further. We were able to establish a shortcut and have shown, that the proof of coDP-hardness for WSPE is sufficient to prove Θ_2^p-hardness. The same argument also works for SCSCS, however, it still remains unclear whether the upper bound of Σ_2^p for SCSCS can be lowered to Θ_2^p or not.

In Chapters 4 and 5, we proposed a new type of encoding, called *weak ranking with double threshold*, that combines the singleton approach with the friend- and enemy-oriented encoding. An advantage of this idea is a huge increase in expressivity without increasing the size of the input. However, a major drawback is the need for a comprehensive procedure to extend these rankings to a preference order over coalitions. We have chosen to use the polarized responsive extension principle, which yields a partial order over coalitions containing the player. Adversely, we need total orders, which led us to two ideas: In Chapter 4, we dealt with this problem by leaving these incomparabilities open and using notions such as *possible* and *necessary*. Similar to the problems of Chapter 3, we focused on the verification and existence problem and the stability concepts perfectness, individual rationality, (contractual) individual stability, Nash stability, (strict) core stability, Pareto optimality, and (strict) popularity. We investigated all possible combinations of the ten different stability concepts and the four cases of decision problems, and established a wide range of results ranging from feasibility results to hardness results for NP and coNP, but also left open important gaps, especially for the stability concepts (strict) core stability, Pareto optimality, and (strict) popularity.

In Chapter 5 we used Borda-like comparability functions to break incomparabilities. We recommended four different versions of these functions for the friends, and four analogous versions for the enemies. Each combination of one function for the friends and one for the enemies defines a way of how the scores are derived from the rankings, and therefore results in a possibly different hedonic game. We established four feasibility results in the case of the verification problem (for perfectness, individual stability, contractual individual stability, and Nash stability) and two in the case of the existence problem (perfectness and contractual individual stability). For the two remaining stability concepts studied in Chapter 5, core stability and strict core stability, we were able to prove coNP-completeness in the verification case, with all results so far being independent from the choice of the comparability function. However, the last four remaining cases (individual stability, Nash stability, core stability, and strict core stability in combination with the existence problem) are partly open: While strict core stability seems to be a hard case in general, as we have not been able to tighten the gap between coNP-hardness and membership in Σ_2^p for any choice of comparability functions, the three other cases depend highly on the choice of the comparability functions.

Abstract argumentation was introduced in Chapter 6. We proposed a new extended model of argumentation frameworks that allows us to model situations with incomplete information. This includes incomplete information in both the set of arguments and attacks. As in the other chapters, we investigated the verification problem from computational complexity applied to the solution concepts of argumentation frameworks proposed by Dung [29], which are called *semantics*. To deal with the incomplete information, we again use the notions of possibility and necessity. We established many feasibility results for the semantics conflict-freeness, admissibility, stability, completeness, groundedness, and preferredness, especially in the case of necessary verification, while hardness almost solely occurs for possible verification. The only exception is the preferred semantics, for which the standard verification problem already was coNP-complete. Even though this complexity does not increase in the necessary case, it increases to Σ_2^p-completeness in the case of possible verification.

For future work it seems to be a good idea to continue with a complexity analysis of the investigated models to close the open gaps, but also to investigate decision problems such as credulous or skeptical acceptance, as already started by Baumeister et al. [10]. For credulous acceptance we ask whether there is a coalition structure (respectively argument set) that contains an a priorly fixed coalition (respectively argument) and that satisfies a given stability concept (respectively semantics). For skeptical acceptance we ask whether there is one coalition (respectively argument) that is contained in any coalition structure (respectively argument set) that satisfies a given stability concept (respectively semantics). Additionally, it seems to be important for any application of these formal analysis to find suitable ideas that refine the contrasting concepts of possibility and necessity but also credulous and skeptical acceptance. A goal could be to search for intermediate states that represent the *some*-part, instead of only allowing *one* or *all*. In general, it could be very fruitful to continue with close interdisciplinary work and capture exactly those problems from other disciplines that seem to profit from a formal analysis most, and that can sustainably influence and shape the future of our research fields.

Bibliography

[1] J. Alcalde and P. Revilla. Researching with whom? Stability and manipulation. *Journal of Mathematical Economics*, 40(8):869–887, 2004.

[2] L. Amgoud and C. Cayrol. On the acceptability of arguments in preference-based argumentation. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 1–7. Morgan Kaufmann Publishers Inc., 1998.

[3] L. Amgoud, C. Cayrol, M. Lagasquie-Schiex, and P. Livet. On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems*, 23(10):1062–1093, 2008.

[4] S. Arora and B. Barak. *Computational Complexity: A Modern Approach.* Cambridge University Press, 2009.

[5] H. Aziz and R. Savani. Hedonic games. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 15, pages 356–376. Cambridge University Press, 2016.

[6] S. Banerjee, H. Konishi, and T. Sönmez. Core in a simple coalition formation game. *Social Choice and Welfare*, 18:135–153, 2001.

[7] P. Baroni, F. Cerutti, M. Giacomin, and G. Guida. Afra: Argumentation framework with recursive attacks. *International Journal of Approximate Reasoning*, 52(1):19–37, 2011.

[8] D. Baumeister, J. Rothe, and H. Schadrack. Verification in argument-incomplete argumentation frameworks. In *Proceedings of the 4th International Conference on Algorithmic Decision Theory*, pages 359–376. Springer-Verlag *Lecture Notes in Artificial Intelligence #9346*, September 2015.

[9] D. Baumeister, D. Neugebauer, J. Rothe, and H. Schadrack. Verification in incomplete argumentation frameworks. In *Proceedings of the 6th International Workshop on Computational Social Choice*, June 2016.

[10] D. Baumeister, D. Neugebauer, and J. Rothe. Credulous and skeptical acceptance in incomplete argumentation frameworks. In *Proceedings of the 7th International Conference on Computational Models of Argument*, volume 305 of *Frontiers in Artificial Intelligence and Applications*, pages 181–192. IOS Press, 2018.

[11] D. Baumeister, D. Neugebauer, J. Rothe, and H. Schadrack. Complexity of verification in incomplete argumentation frameworks. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. AAAI Press, February 2018.

[12] D. Baumeister, D. Neugebauer, J. Rothe, and H. Schadrack. Complexity of verification in incomplete argumentation frameworks. In *Proceedings of the 7th International Workshop on Computational Social Choice*, June 2018.

[13] D. Baumeister, D. Neugebauer, J. Rothe, and H. Schadrack. Verification in incomplete argumentation frameworks. *Artificial Intelligence*, 264:1–26, 2018.

[14] R. Beigel. Bounded queries to SAT and the boolean hierarchy. *Theoretical Computer Science*, 84(2):199–223, 1991.

[15] R. Beigel, L. Hemachandra, and G. Wechsung. On the power of probabilistic polynomial time: $\mathrm{P}^{\mathrm{NP}[\log]} \subseteq \mathrm{PP}$. In *Proceedings of the 4th Structure in Complexity Theory Conference*, pages 225–227. IEEE Computer Society Press, June 1989.

[16] T. Bench-Capon. Value-based argumentation frameworks. In *Proceedings of the 9th International Workshop on Non-Monotonic Reasoning*, pages 444–453, 2002.

[17] A. Bogomolnaia and M. Jackson. The stability of hedonic coalition structures. *Games and Economic Behavior*, 38(2):201–230, 2002.

[18] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

[19] G. Brewka and S. Woltran. Abstract dialectical frameworks. In *Principles of Knowledge Representation and Reasoning: Proceedings of the 12th International Conference*. AAAI Press, 2010.

[20] G. Brewka, S. Ellmauthaler, H. Strass, J. Wallner, and S. Woltran. Abstract dialectical frameworks revisited. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, pages 803–809. AAAI Press/IJCAI, 2013.

[21] J. Cai, T. Gundermann, J. Hartmanis, L. Hemachandra, V. Sewelson, K. Wagner, and G. Wechsung. The boolean hierarchy I: Structural properties. *SIAM Journal on Computing*, 17(6):1232–1252, 1988.

[22] J. Cai, T. Gundermann, J. Hartmanis, L. Hemachandra, V. Sewelson, K. Wagner, and G. Wechsung. The boolean hierarchy II: Applications. *SIAM Journal on Computing*, 18(1):95–111, 1989.

[23] K. Cechlárová and J. Hajduková. Computational complexity of stable partitions with B-preferences. *International Journal of Game Theory*, 31(3):353–364, 2003.

[24] K. Cechlárová and J. Hajduková. Stable partitions with $\mathcal{W}$-preferences. *Discrete Applied Mathematics*, 138(3):333–347, 2004.

[25] K. Cechlárová and A. Romero-Medina. Stability in coalition formation games. *International Journal of Game Theory*, 29(4):487–494, 2001.

[26] D. Dimitrov, P. Borm, R. Hendrickx, and S. Sung. Simple priorities and core stability in hedonic games. *Social Choice and Welfare*, 26(2): 421–433, 2006.

[27] Y. Dimopoulos and A. Torres. Graph theoretical structures in logic programs and default theories. *Theoretical Computer Science*, 170(1): 209–244, 1996.

[28] P. Dondio. Multi-valued argumentation frameworks. In *International Workshop on Rules and Rule Markup Languages for the Semantic Web*, pages 142–156. Springer, 2014.

[29] P. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.

[30] P. Dunne and M. Wooldridge. Complexity of abstract argumentation. In I. Rahwan and G. Simari, editors, *Argumentation in Artificial Intelligence*, chapter 5, pages 85–104. Springer, 2009.

[31] U. Endriss, editor. *Trends in Computational Social Choice*. AI Access, 2017.

[32] D. Gale and L. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

[33] L. Hemachandra. The strong exponential hierarchy collapses. *Journal of Computer and System Sciences*, 39(3):299–322, 1989.

[34] E. Hemaspaandra, L. Hemaspaandra, and J. Rothe. Exact analysis of Dodgson elections: Lewis Carroll's 1876 voting system is complete for parallel access to NP. *Journal of the ACM*, 44(6):806–825, 1997.

[35] E. Hemaspaandra, L. Hemaspaandra, and J. Rothe. Raising NP lower bounds to parallel NP lower bounds. *SIGACT News*, 28(2):2–13, 1997.

[36] E. Hemaspaandra, H. Spakowski, and J. Vogel. The complexity of Kemeny elections. *Theoretical Computer Science*, 349(3):382–391, 2005.

[37] H. Jakobovits and D. Vermeir. Dialectic semantics for argumentation frameworks. In *Proceedings of the 7th international conference on Artificial intelligence and law*, pages 53–62. ACM, 1999.

[38] J. Köbler, U. Schöning, and K. Wagner. The difference and truth-table hierarchies for NP. *R.A.I.R.O. Informatique théorique et Applications*, 21:419–435, 1987.

[39] T. Krauthoff, M. Baurmann, G. Betz, and M. Mauve. Dialog-based online argumentation. In *Proceedings of the 6th International Conference on Computational Models of Argument*, pages 33–40, September 2016.

[40] J. Lang, A. Rey, J. Rothe, H. Schadrack, and L. Schend. Representing and solving hedonic games with ordinal preferences and thresholds. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems*, pages 1229–1237. IFAAMAS, May 2015.

[41] H. Li, N. Oren, and T. Norman. Probabilistic argumentation frameworks. In *Proceedings of the 1st International Workshop on Theory and Applications of Formal Argumentation*, pages 1–16. Springer-Verlag *Lecture Notes in Artificial Intelligence #7132*, July 2011.

[42] A. Meyer and L. Stockmeyer. The equivalence problem for regular expressions with squaring requires exponential space. In *Proceedings of the 13th IEEE Symposium on Switching and Automata Theory*, pages 125–129. IEEE Computer Society Press, 1972.

[43] S. Modgil. An abstract theory of argumentation that accommodates defeasible reasoning about preferences. In *Proceedings of the 9th European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 648–659. Springer-Verlag *Lecture Notes in Artificial Intelligence #4724*, November 2007.

[44] S. Modgil and T. Bench-Capon. Integrating object and meta-level value based argumentation. In *Proceedings of the 2nd International Conference on Computational Models of Argument*, volume 172, pages 240–251. IOS Press, September 2008.

[45] J. Nash. The bargaining problem. *Econometrica: Journal of the Econometric Society*, 18(2):155–162, 1950.

[46] D. Neugebauer. Generating defeasible knowledge bases from real-world argumentations using D-BAS. In *Proceedings of the 1st Workshop on Advances In Argumentation in Artificial Intelligence*, volume 2012, pages 105–110. CEUR Workshop Proceedings, November 2017.

[47] N. Nguyen, T. Nguyen, M. Roos, and J. Rothe. Computational complexity and approximability of social welfare optimization in multiagent resource allocation. *Journal of Autonomous Agents and Multi-Agent Systems*, 28(2):256–289, 2014.

[48] C. Papadimitriou. *Computational Complexity.* Addison-Wesley, 2nd edition, 1995. Reprinted with corrections.

[49] C. Papadimitriou and M. Yannakakis. The complexity of facets (and some facets of complexity). *Journal of Computer and System Sciences*, 28(2):244–259, 1984.

[50] C. Papadimitriou and S. Zachos. Two remarks on the power of counting. In *Proceedings of the 6th GI Conference on Theoretical Computer Science*, pages 269–276. Springer-Verlag *Lecture Notes in Computer Science #145*, 1983.

[51] D. Peters and E. Elkind. Simple causes of complexity in hedonic games. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, pages 617–623. AAAI Press/IJCAI, 2015.

[52] I. Rahwan and G. Simari, editors. *Argumentation in Artificial Intelligence*. Springer, 2009.

[53] Y. Reisch, J. Rothe, and L. Schend. The margin of victory in Schulze, cup, and Copeland elections: Complexity of the regular and exact variants. In *Proceedings of the 7th European Starting AI Researcher Symposium*, pages 250–259. IOS Press, August 2014.

[54] A. Rey, J. Rothe, H. Schadrack, and L. Schend. Toward the complexity of the existence of wonderfully stable partitions and strictly core stable coalition structures in hedonic games. In *Website Proceedings of the Special Session on Computational Social Choice at the 13th International Symposium on Artificial Intelligence and Mathematics*, January 2014.

[55] A. Rey, J. Rothe, H. Schadrack, and L. Schend. Toward the complexity of the existence of wonderfully stable partitions and strictly core stable coalition structures in hedonic games. In *Proceedings of the 11th Conference on Logic and the Foundations of Game and Decision Theory*, July 2014.

[56] A. Rey, J. Rothe, H. Schadrack, and L. Schend. Toward the complexity of the existence of wonderfully stable partitions and strictly core stable coalition structures in enemy-oriented hedonic games. *Annals of Mathematics and Artificial Intelligence*, 77(3-4):317–333, 2016.

[57] T. Riege and J. Rothe. Completeness in the boolean hierarchy: Exact-Four-Colorability, minimal graph uncolorability, and exact domatic number problems – a survey. *Journal of Universal Computer Science*, 12(5):551–578, 2006.

[58] C. Rinner. Argumentation maps: Gis-based discussion support for online planning. *Environment and Planning B: Planning and Design*, 28 (6):847–863, 2001.

[59] A. Roth and E. Peranson. The redesign of the matching market for american physicians: Some engineering aspects of economic design. *American economic review*, 89(4):748–780, 1999.

[60] J. Rothe. *Complexity Theory and Cryptology. An Introduction to Cryptocomplexity*. EATCS Texts in Theoretical Computer Science. Springer-Verlag, 2005.

[61] J. Rothe, H. Spakowski, and J. Vogel. Exact complexity of the winner problem for Young elections. *Theory of Computing Systems*, 36(4):375–386, 2003.

[62] J. Rothe, H. Schadrack, and L. Schend. Borda-induced hedonic games with friends, enemies, and neutral players. *Mathematical Social Sciences*, 96:21–36, 2018.

[63] W. Saad, Z. Han, T. Basar, M. Debbah, and A. Hjorungnes. Hedonic coalition formation for distributed task allocation among wireless agents. *IEEE Transactions on Mobile Computing*, 10(9):1327–1344, 2011.

[64] M. Schaefer and C. Umans. Completeness in the polynomial-time hierarchy: Part I: A compendium. *SIGACT News*, 33(3):32–49, September 2002.

[65] M. Schaefer and C. Umans. Completeness in the polynomial-time hierarchy: Part II. *SIGACT News*, 33(4):22–36, December 2002.

[66] L. Stockmeyer. The polynomial-time hierarchy. *Theoretical Computer Science*, 3(1):1–22, 1976.

[67] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[68] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior (commemorative edition)*. Princeton University Press, 4th edition, 2007.

[69] G. Vreeswijk and H. Prakken. Credulous and sceptical argument games for preferred semantics. In *European Workshop on Logics in Artificial Intelligence*, pages 239–253. Springer, 2000.

[70] K. Wagner. More complicated questions about maxima and minima, and some closures of NP. *Theoretical Computer Science*, 51(1–2):53–80, 1987.

[71] K. Wagner. Bounded query classes. *SIAM Journal on Computing*, 19 (5):833–846, 1990.

[72] D. Walton. Burden of proof. *Argumentation*, 2(2):233–254, 1988.

[73] R. Whately. *Elements of rhetoric*. Longmans, Green, Reader, and Dyer, reprinted from 7th edition, 1867.

[74] G. Woeginger. Core stability in hedonic coalition formation. In *Proceedings of the 39th Conference on Current Trends in Theory and Practice of Computer Science*, pages 33–50. Springer-Verlag *Lecture Notes in Computer Science #7741*, January 2013.

[75] G. Woeginger. A hardness result for core stability in additive hedonic games. *Mathematical Social Sciences*, 65(2):101–104, 2013.