

Causes of Stability in Dynamic Coalition Formation

Niclas Boehmer, Martin Bullinger, and Anna Maria Kerkmann

Abstract

We study the formation of stable outcomes via simple dynamics in cardinal hedonic games, where the utilities of agents change over time depending on the history of the coalition formation process. Specifically, we analyze situations where members of a coalition decrease their utility for a leaving agent (*resentment*) or increase their utility for a joining agent (*appreciation*). We show a series of convergence results for dynamics for resentful or appreciative agents which do not hold for classical dynamics. In particular, resentment turns out to be a strong stability-driving force. If resentment is expressed by the deviator, we find a more nuanced picture, but still a broad possibility of convergence. We complement our theoretical analysis with simulations that allow us to get insight on the average running time of the dynamics and on the structure of the produced outcomes. On the theoretical side, we obtain broad hardness of determining the fastest convergence time, a result which also carries over to classical dynamics under static utility functions.

1 Introduction

Coalition formation is a vibrant topic in multi-agent systems that has been continuously researched during the last decades. It concerns the question of dividing a set of agents, for example, humans or machines, into disjoint coalitions such as research teams. Agents carry preferences over these coalition structures. A common assumption is that externalities, that is, the coalition structure outside one's own coalition, play no role. This is captured in the prominent framework of hedonic games. Moreover, the desirability of a coalition structure is usually measured with respect to stability. Abstractly speaking, a coalition structure is stable if there is no agent or set of agents that can perform a beneficial deviation by joining existing coalitions or by forming new coalitions.

There are two specific properties of hedonic games crucially influencing past research. First, the number of possible coalitions an agent can be part of is exponentially large. Therefore, a repeatedly considered challenge is to come up with reasonable succinctly representable settings. It is very prominent in this context to aggregate utilities from cardinal valuations of other agents [8, 3, 1]. Second, most established stability concepts suffer from non-existence under strong restrictions which often leads to computational boundaries such as hardness of the decision problem whether a stable state exists [28, 2, 1]. Much of the research has therefore focused on identifying suitable conditions guaranteeing stable states.

The dominant coalition formation framework is static in two dimensions. First, stability is usually a static concept in the sense that, since a coalition structure is either stable or not, we are only interested in *finding* these stable structures. The underlying assumption here is that we operate in a centralized system where a (desirable) coalition structure can be created by a central authority. This paradigm has only recently been complemented by interpreting deviations of agents as a dynamic process, and analyzing how stable coalition structures can be reached through decentralized individual decisions [10, 11]. Second, utility functions are static. To demonstrate the implications of this assumption, we describe a run-and-chase example, which is present in many classes of hedonic games. Consider a situation where there are only the two agents Alice and Bob. Alice wants to be alone in her coalition, whereas Bob wants to be in a joint coalition with Alice. It is clear that in the two possible coalition structures, there is always an agent who wants to change their situation. From a centralized perspective, this simply means that no coalition structure has the prospect of stability. In a distributed, dynamic setting where utilities are static, the following occurs indefinitely: Whenever Alice and Bob are in a joint coalition, then Alice leaves the coalition to be alone. However, whenever Alice and Bob are in two separate coalitions, then Bob joins Alice. In practice, such an infinite situation is unreasonable: After playing run-and-chase for a while, one of the agents is likely to change their behavior, and therefore their preferences. On the one hand, Bob might get frustrated because he is constantly left by Alice and therefore stops his efforts to join her. On the other hand, Alice could realize the high effort that Bob makes to be in a coalition with her and feels sufficient appreciation to eventually accept Bob in her coalition. In both scenarios, we reach a state that is stable because of the *history* of the coalition formation process.

Table 1: Overview of results. Each column corresponds to one stability criterion (see Section 2 for definitions). “✓” means that the corresponding dynamics is guaranteed to converge; “✗” means that we have an example for an infinite sequence. For each result, we include the number of the respective statement. For appreciation, all non-convergence results also hold if deviations are restricted to be individually rational.

	SCS	CS	IS	CNS	NS
ASHG					
resentment	✓ (3.1)	✓ (3.1)	✓ (3.1)	?	?
resentment+IR	✓ (3.1)	✓ (3.1)	✓ (3.1)	✓ (3.3)	✓ (3.3)
appreciation	✗ (4.3)	✗ (4.3)	?	✓ (4.4)	?
MFHG					
resentment	?	?	?	?	✗ (3.4)
resentment+IR	✓ (3.2)	✓ (3.2)	✓ (3.2)	✓ (3.3)	✓ (3.3)
appreciation	✗ (4.3)	✗ (4.3)	?	?	✗ (4.1)

In this paper, we model situations where the history influences the agents’ utilities, offering a new perspective on the reachability of stable coalition structures. We study a dynamic coalition formation process where agents perform deviations based on stability concepts. However, in contrast to previous work on dynamics, we assume that a deviation has an effect on the *perception* of the deviator, resulting in agents changing their utility for the deviator. We distinguish two approaches. First, an agent might act *resentfully* in the sense that, like Bob, she lowers her utility for an agent abandoning her. If a resentful agent is repeatedly abandoned by a deviator, the resentful agent eventually loses interest in being together with the deviator. On the other hand, an agent could *appreciate* the effort of another agent to be part of her coalition, and therefore, like Alice, increase her utility for an agent whenever the agent joins her. After sufficient effort, the urge to leave the deviator ceases.

1.1 Contribution

We initiate the study of cardinal hedonic games under utility functions changing over time. In particular, we consider utility modifications based on the resentful and appreciative perception of other agents. We investigate whether decentralized dynamics based on various types of deviations are guaranteed to converge. Deviations might be constrained to be individually rational (IR), that is, a deviating agent needs to prefer her new coalition to being alone. We showcase our results by considering additively separable hedonic games (ASHGs) and modified fractional hedonic games (MFHGs), where an agent’s utility for a coalition is the sum or average utility for the other agents in the coalition, respectively. Table 1 provides an overview of these results. First, for resentful agents performing individually rational deviations, convergence is guaranteed in all considered cases. If deviations may also violate individual rationality, the situation becomes more complicated and elusive to a complete understanding; nevertheless, we establish several convergence guarantees while also having an involved example of a cycling dynamics in MFHGs. In contrast, appreciation is usually not sufficient to guarantee convergence. Notably, as proved in Propositions 3.5 and 4.2, some of our open questions are equivalent.

In fact, most of our results do not only apply to ASHGs and/or MFHGs but to larger classes of hedonic games. For this, we develop an axiomatic framework for utility aggregation based on the perception of friends and enemies, that is, agents yielding positive and negative utility, respectively.

In our simulations, we observe that our model of dynamic utilities leads to the (quick) convergence of Nash dynamics. Moreover, we analyze the structure and expressiveness of the produced outcomes. Finally, we outline results for other perception models and for computational questions concerned with finding shortest converging sequences.

1.2 Related Work

Hedonic games originate from economic theory [17], but their constant and broad consideration only started with key publications by Banerjee et al. [5], Cechlárová and Romero-Medina [15], and Bogomolnaia and Jackson [8]. An overview of hedonic games is provided in the survey by Aziz and Savani [1]. The search for suitable representations of reasonable classes of hedonic games has led to various

proposals [see, e.g., 15, 8, 4, 18, 24, 3].

Various stability concepts and their computational boundaries have been previously studied. We focus on results concerning ASHG [8] and MFHG [24]. Sung and Dimitrov [28] show prototype NP-hardness reductions for single-agent stability concepts in ASHG, paving the way for many similar results for single-agent and group stability [see, e.g., 2, 11, 13]. Gairing and Savani [21] consider ASHG under symmetric utilities and show PLS-completeness of computing stable states, while Woeginger [29] and Peters [25] show Σ_2^P -completeness of the (strict) core in ASHG. Peters and Elkind [26] provide a meta view on computational hardness. For MFHG, there seem to be less computational boundaries. Indeed, for symmetric and binary utilities, stable states exist and can be efficiently computed. Core stability is even tractable for symmetric and arbitrarily weighted utilities [23]. Apart from the consideration of stability, other desirable notions of efficiency or fairness such as Pareto optimality, envy-freeness, or popularity have been studied for ASHG and MFHG [2, 19, 12, 9]. These papers provide more evidence that MFHG seem to be less complex than ASHG.

The dynamical, distributed approach to coalition formation received increased attention very recently [22, 6, 14, 10, 20, 11, 7]. In this line of research, the goal is typically to establish under which circumstances dynamics are (not) guaranteed to converge [6, 14, 10, 20, 11]. In addition, computational questions such as deciding whether all (or at least one) execution of some dynamic from a given initial state converge have been examined [10, 11]. Such questions have been explored both for single-agent deviations Bilò et al. [6], Brandt et al. [10, 11] and group stability Carosi et al. [14], Fanelli et al. [20] for a variety of hedonic games, including additively separable and fractional hedonic games. Our work differs from these papers in that we analyze the convergence of dynamics in cases where the utility of agents depends on the history of the process.

2 Preliminaries and Model

In this section, we define the basic coalition formation setting, our specific model, and provide some first observations. For an integer $i \in \mathbb{N}$, we define $[i] = \{1, \dots, i\}$.

2.1 Cardinal Hedonic Games

Let $N = [n]$ be a finite set of *agents*. A *coalition* is any non-empty subset of N . We denote the set of all possible coalitions containing agent $i \in N$ by $\mathcal{N}_i = \{C \subseteq N : i \in C\}$. Any *partition* of the agents N is also called *coalition structure* and we denote the set of all partitions of N by Π_N . Given an agent $i \in N$ and a partition $\pi \in \Pi_N$, let $\pi(i)$ denote the coalition of i , i.e., the unique coalition $C \in \pi$ with $i \in C$. A (*cardinal*) *hedonic game* is a pair (N, u) consisting of a set N of agents and a utility profile $u = (u_i)_{i \in N}$ where $u_i : N \rightarrow \mathbb{Q}$ is the *utility function* of agent i . Thus, for $i, j \in N$, $u_i(j)$ is i 's utility for agent j . We sometimes equivalently view a utility function as a vector $u_i \in \mathbb{Q}^n$. An agent $j \in N$ is a *friend* (or *enemy*) of an agent $i \in N$ if $u_i(j) > 0$ (or $u_i(j) < 0$).

To move from utilities for single agents to utilities over coalitions, we use *cardinal aggregation functions* (CAFs). For every agent $i \in N$, the CAF $A_i : \mathcal{N}_i \times \mathbb{Q}^n \rightarrow \mathbb{Q}$ specifies i 's utility for a given coalition for her given utility vector. Then, the utility of an agent for a partition π with respect to aggregation function A_i is $u_i^{A_i}(\pi) = A_i(\pi(i), u_i)$. To keep notation concise, we sometimes omit the CAF as a superscript when it is clear from the context. For an agent $i \in N$ with utility function u_i , a coalition $C \in \mathcal{N}_i$ is *individually rational* (IR) if $A_i(C, u_i) \geq A_i(\{i\}, u_i)$. Further, a partition π is *individually rational* (IR) for agent i if $\pi(i)$ is an individually rational coalition.

Common classes of cardinal hedonic games such as the two specific classes studied in this paper have a straightforward representation with respect to CAFs. For each agent $i \in N$ with utility function u_i ,

- additively separable hedonic games (ASHG) [8] use the aggregation function AS defined by $AS_i(C, u_i) = \sum_{j \in C \setminus \{i\}} u_i(j)$ and
- modified fractional hedonic games (MFHG) [24] use the aggregation function MF defined by $MF_i(C, u_i) = \frac{\sum_{j \in C \setminus \{i\}} u_i(j)}{|C|-1}$ if $|C| \geq 2$ and $MF_i(C, u_i) = 0$, otherwise.

Note that both in ASHG and MFHG agents derive zero utility from being in a singleton coalition.

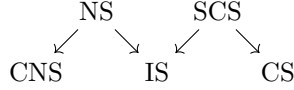


Figure 1: Relations among our stability concepts. Arrows indicate implications. For example, strict core stability (SCS) implies core stability (CS) and individual stability (IS).

2.2 Deviations and Stability

As indicated in the introduction, we distinguish different stability notions based on single-agent deviations and group deviations. Given a partition $\pi \in \Pi_N$, agent $i \in N$ might perform a *single-agent deviation* from $\pi(i)$ to any coalition $C \in \pi \cup \{\emptyset\}$, resulting in the partition $\pi' = (\pi \setminus \{\pi(i), C\}) \cup \{\pi(i) \setminus \{i\}, C \cup \{i\}\}$; and a group of agents $C \subseteq N$ might perform a *group deviation*, leading to the partition $\pi' = (\pi \setminus \{\pi(j) \mid j \in C\}) \cup \{\pi(j) \setminus C \mid j \in C\} \cup \{C\}$. Depending on which agents improve as a result of a deviation, we distinguish the following *types of deviations*. Agent i 's single-agent deviation from $\pi(i)$ to $C \in \pi \cup \{\emptyset\}$, resulting in partition π' , is a *Nash* (NS) deviation if $u_i(\pi') > u_i(\pi)$. An NS deviation of i from π to π' is called

- an *individual* (IS) deviation if $u_j(\pi') \geq u_j(\pi)$ for all $j \in C$, where C is the coalition to which i deviated; and
- a *contractual Nash* (CNS) deviation if $u_j(\pi') \geq u_j(\pi)$ for all $j \in \pi(i) \setminus \{i\}$.

A group deviation of coalition C from π to π' is

- a *core* (CS) deviation if $u_i(\pi') > u_i(\pi)$ for all $i \in C$; and
- a *strict core* (SCS) deviation if $u_i(\pi') \geq u_i(\pi)$ for all $i \in C$ and $u_j(\pi') > u_j(\pi)$ for some $j \in C$.

Finally, for all types of deviations introduced above, we define the respective stability notion of a partition by the absence of a corresponding deviation. For example, a partition π is said to be Nash-stable (NS) if there is no NS deviation from π to another partition. The logical relations among the resulting stability concepts are illustrated in Figure 1 (see also the survey of Aziz and Savani [1]).

For a given partition, several single-agent or group deviations might be possible. Yet, some deviations seem to be more reasonable than others. We say that a deviation is IR if the resulting partition is IR for all deviating agents. For all our considered stability concepts it holds that if an agent has a deviation (that is potentially not IR), then she also has an IR deviation where she forms a singleton coalition.

2.3 Dynamic Coalition Formation

We now introduce our model of dynamic coalition formation over time, and the concepts of *resentment* and *appreciation*. Throughout the paper, we consider sequences of partitions $(\pi^t)_{t \geq 0}$, where for every $t \geq 1$, π^t evolves from π^{t-1} by means of some single-agent or group deviation. We assume that both the initial coalition structure π^0 and the initial utility vectors u_i^0 for each agent $i \in N$ are given. However, utilities change over time as follows. Under *resentment*, agents decrease their utilities by one for all deviators that leave them, while under *appreciation*, agents increase their utilities by one for all deviators that join them.¹ More formally, if for some $t \geq 1$, π^t evolves from π^{t-1} via a single-agent deviation of agent $k \in N$, then, for $i, j \in N$,

- for *resentful* agents, $u_i^t(j)$ arises from $u_i^{t-1}(j)$ as $u_i^t(j) = \begin{cases} u_i^{t-1}(j) - 1 & i \neq k, j = k, j \in \pi^{t-1}(i), \\ u_i^{t-1}(j) & \text{else.} \end{cases}$
- for *appreciative* agents, $u_i^t(j)$ arises from $u_i^{t-1}(j)$ as $u_i^t(j) = \begin{cases} u_i^{t-1}(j) + 1 & i \neq k, j = k, j \in \pi^t(i), \\ u_i^{t-1}(j) & \text{else.} \end{cases}$

¹Note that our choice of decreasing or increasing the utilities by one is somewhat arbitrary, as our theoretical results hold for any fixed increase or decrease of utilities. Moreover, our results still work if every agent has an individual constant to modify utilities. However, in case the utility change in each round is not constant, our convergence guarantees are no longer applicable, as, for instance, run-and-chase situations can occur.

If for $t \geq 1$, π^t evolves from π^{t-1} via a group deviation of $C \subseteq N$, then, for $i, j \in N$,

- for *resentful* agents, $u_i^t(j)$ arises from $u_i^{t-1}(j)$ as $u_i^t(j) = \begin{cases} u_i^{t-1}(j) - 1 & i \notin C, j \in C, j \in \pi^{t-1}(i), \\ u_i^{t-1}(j) & \text{else.} \end{cases}$
- for *appreciative* agents, $u_i^t(j)$ arises from $u_i^{t-1}(j)$ as $u_i^t(j) = \begin{cases} u_i^{t-1}(j) + 1 & i \neq j, i \in C, j \in C, \\ u_i^{t-1}(j) & \text{else.} \end{cases}$

We are concerned about sequences of partitions that evolve by deviations with respect to the current utilities of the agents. For any stability concept $\alpha \in \{NS, IS, CNS, CS, SCS\}$, a sequence of partitions $(\pi^t)_{t \geq 0}$ is called an execution of an α dynamics if π^t evolves from π^{t-1} through an α deviation with respect to the utility functions $(u_i^{t-1})_{i \in N}$. If all deviations are individually rational, we call the dynamics *individually rational*, e.g., individually rational NS dynamics in the case of Nash stability.

An execution of an α dynamics *converges* if it terminates after a finite number of T steps in a partition π^T that is stable with respect to $(u_i^T)_{i \in N}$ under the stability notion α . We say that the α dynamics *converges* if every execution of the α dynamics converges for every initial utility profile and partition. By contrast, the dynamics *cycles* if there exists an infinite execution of the dynamics (for some initial utilities and partition). The central question of this paper is when dynamics converge for resentful or appreciative agents.

It is convenient to use a compact notation for utilities. We write $u_i^{t, A_i}(\pi) = A_i(\pi(i), u_i^t)$ and $u_i^{t, A_i}(C) = A_i(C, u_i^t)$ for the utility of agent i at time t for a partition $\pi \in \Pi_N$ or for coalition $C \in \mathcal{N}_i$, respectively. If the CAF A_i is clear from context, we usually omit it as superscript.

Before our main analysis, we present a useful lemma that holds for arbitrary dynamics. The lemma can be applied to show that, from a certain point onwards, every deviation occurs infinitely often in an infinite execution of a dynamic.²

Lemma 2.1. *Let $(\pi^t)_{t \geq 0}$ be an infinite sequence of partitions induced by single-agent (or group) deviations. Then, there exists a $t_0 \geq 0$ such that every single-agent (or group) deviation performed at some time $t \geq t_0$ occurs infinitely often.*

Lastly, we call an infinite sequence of partitions $\pi = (\pi^t)_{t \geq 0}$ *periodic* if there exist $t_0 \in \mathbb{N}$ and $p \in \mathbb{N}$ such that, for all $k \in \mathbb{N}_0$ and $l \in \{0, \dots, p-1\}$, it holds that $\pi^{t_0+kp+l} = \pi^{t_0+l}$.

2.4 Properties of Aggregation Functions

We now introduce some useful properties of CAFs. For a simplified exposition, we give intuitive, informal definitions. A formal treatment can be found in Appendix A. The CAF of an agent $i \in N$ satisfies

- *aversion to enemies* (ATE) if i 's aggregated utility does not decrease when an enemy leaves i 's coalition.
- *individually rational aversion to enemies* (IR ATE) if i 's aggregated utility does not decrease when an enemy leaves one of i 's individually rational coalitions.
- *enemy monotonicity* (EM) if decreasing the utility for an enemy cannot increase i 's aggregated utility for a coalition containing the enemy.
- *enemy domination* (ED) if in case i 's utility for some agent j is sufficiently negative and i 's utility for every other agent is bounded, then no coalition containing j is individually rational for i .

All of these axioms capture the treatment of enemies. The first two axioms deal with situations where an enemy leaves the agent's coalition, where ATE is stronger than IR ATE. On the other hand, EM and ED are variable utility conditions describing situations where the utility for an enemy decreases or some agent turns into a very bad enemy, respectively. Apart from the implication between ATE and IR ATE, there are no other logical relationships between any pair of axioms.

²All missing proofs can be found in the appendix.

Example 2.2. In this example, we consider a game (N, u) for which the CAF MF violates ATE. Let $N = \{a, b, c\}$ and let the single-agent utilities be $u_a(b) = -1$, $u_a(c) = -3$, $u_b(a) = 1$, and $u_b(c) = -1$. (The utilities $u_c(a)$ and $u_c(b)$ are irrelevant.)

Then, removing an enemy can make an agent worse. Indeed, $MF_a(N, u_a) = -2 > -3 = MF_a(\{a, c\}, u_a)$. Hence, MF violates ATE. On the other hand, as we will see in Proposition 2.3, removing an enemy from an individually rational coalition cannot decrease the utility in an MFHG. For instance, $MF_b(N, u_b) = 0 < 1 = MF_b(\{a, b\}, u_b)$. \triangleleft

Still, classical aggregation functions usually satisfy (most of) our introduced axioms.

Proposition 2.3. *The additively separable CAF AS_i satisfies ATE, IR ATE, EM, and ED. The modified fractional CAF MF_i satisfies IR ATE, EM, and ED but violates ATE.*

3 Dynamics for Resentful Agents

In this section, we study the convergence of different types of dynamics for resentful agents. We start by considering SCS, CS, and IS dynamics, before turning to CNS and NS dynamics.

3.1 Core Stability and Individual Stability

If deviating agents need consensus from their new coalition, it turns out that resentment is a strong force to establish convergence. The intuitive reason for this is that an agent a can only leave an agent b for a limited number of times until resentment prevents that they form a joint coalition again. In fact, otherwise b 's utility for a becomes arbitrarily negative and b no longer gives a her consent to join. We will prove that SCS dynamics, and thereby also CS and IS dynamics, always converge for a wide class of CAFs.

Theorem 3.1. *The SCS, CS, and IS dynamics converge for resentful agents whose CAFs satisfy aversion to enemies and enemy monotonicity.*

Proof. It suffices to consider SCS dynamics because every CS dynamics and every IS dynamics is also an SCS dynamics.

Let a hedonic game (N, u^0) with resentful agents be given where every agent $i \in N$ has a CAF A_i that satisfies aversion to enemies and enemy monotonicity. The key insight to show convergence of the SCS dynamics is to prove that in every infinite sequence of SCS deviations, it happens infinitely often that the *non-negative* single-agent valuation of some agent for another agent is decreased (due to resentment). This is a contradiction, as the number of such deviations is bounded by $\sum_{i,j \in N: u_i^0(j) \geq 0} (\lfloor u_i^0(j) \rfloor + 1)$.

Consider an infinite execution $(\pi^t)_{t \geq 0}$ of the SCS dynamics. For the sake of contradiction, assume that there is a step $t_0 \geq 0$ in this execution such that, starting with t_0 , it never happens again that the non-negative valuation of some agent for another agent is decreased. The first step towards a contradiction is to show that every SCS deviation is a Pareto improvement in this situation.³ Indeed, every deviation is weakly improving for every agent in the coalitions that are abandoned: As no agent decreases her non-negative utility for another agent, each abandoned agent is solely abandoned by agents for which she momentarily has a negative utility. We can then iteratively apply aversion to enemies to conclude that the deviation (weakly) increases the utility of each abandoned agent. Further, since by the definition of an SCS deviation, every agent of the newly formed coalition is weakly improving, and some of them are strictly improving, each SCS deviation is a Pareto improvement.

It remains to prove that there cannot be sequences of Pareto improvements of infinite length (due to the changes of the utility functions implied by resentment). To prove this, we will make use of the enemy monotonicity property. To this end, we will show that every agent i can deviate at most once to each coalition $C \in \mathcal{N}_i$ while *improving her utility* by doing so. This provides an upper bound for the length of any sequence of Pareto improving deviations.

Consider a time $t \geq t_0$ and define for some agent $i \in N$ the set $\mathcal{C}_i^t = \{C \in \mathcal{N}_i: u_i^t(C) > u_i^t(\pi^t)\}$, i.e., the set of coalitions that would lead to an improvement of agent i at time t . Consider the potential

³An outcome is a *Pareto improvement* over another outcome if it is weakly better for all agents and strictly better for some agents.

function $V^t = \sum_{i \in N} |\mathcal{C}_i^t|$. Note that $V^{t_0} \leq \sum_{i \in N} |\mathcal{N}_i|$ is initially bounded and that the potential attains only non-negative integer values. We will show that the potential decreases in every time step after t_0 .

To this end, let $i \in N$ be some fixed agent and $t \geq t_0$ some fixed time. If $\pi^{t+1}(i) = \pi^t(i)$, then agent i does not change her utilities, and $\mathcal{C}_i^{t+1} = \mathcal{C}_i^t$. Further, if $\pi^{t+1}(i) \neq \pi^t(i)$ but agent i was not part of the deviating coalition, then, by our above assumption that no non-negative utility gets decreased, $u_i^t(j) < 0$ for all agents $j \in \pi^t(i) \setminus \pi^{t+1}(i)$. Hence, we can repeatedly apply aversion to enemies for all agents in $\pi^t(i) \setminus \pi^{t+1}(i)$ to conclude that $u_i^{t+1}(\pi^{t+1}) = u_i^t(\pi^{t+1}) \geq u_i^t(\pi^t)$. Further, as only i 's valuations of agents from $\pi^t(i) \setminus \pi^{t+1}(i)$ got modified (decreased), the only coalitions $C \in \mathcal{N}_i$ that might have changed their value for agent i contain some agent in $\pi^t(i) \setminus \pi^{t+1}(i)$. Hence, we can (repeatedly) apply enemy monotonicity to conclude that $u_i^{t+1}(C) \leq u_i^t(C)$ for all such coalitions. Hence, $\mathcal{C}_i^{t+1} \subseteq \mathcal{C}_i^t$.

Finally, if agent i is part of the deviating coalition at step t , then she does not change her utilities for any agent and weakly improves the valuation of her coalition. Hence, $\mathcal{C}_i^{t+1} \subseteq \mathcal{C}_i^t$ and we can conclude that $V^{t+1} \leq V^t$. Further, if i strictly improves her utility, then $u_i^{t+1}(\pi^{t+1}) > u_i^{t+1}(\pi^t)$ and it follows that $\pi^{t+1}(i) \in \mathcal{C}_i^t \setminus \mathcal{C}_i^{t+1}$. Hence, $\mathcal{C}_i^{t+1} \subsetneq \mathcal{C}_i^t$. Since this has to be the case for at least one agent in every time step, we conclude that $V^{t+1} < V^t$. \square

As the cardinal aggregation function AS satisfies aversion to enemies and enemy monotonicity (Proposition 2.3), Theorem 3.1 in particular implies that the SCS, CS, and IS dynamics always converge in ASHG for resentful agents.

Notably, Theorem 3.1 breaks down if we consider a CAF violating aversion to enemies, even if enemy monotonicity is still satisfied. Indeed, we can then “ignore” individual utilities. For instance, anonymous hedonic games where agents only care about the size of their coalitions satisfy enemy monotonicity. In such games, resentment is clearly irrelevant and there exist anonymous hedonic games where IS dynamics cycle [10]. Consequently, a result similar to Theorem 3.1 for aggregation functions that only satisfy enemy monotonicity cannot be obtained. On the other hand, it remains an open question whether enemy monotonicity is necessary for Theorem 3.1.

Unfortunately, MF violates aversion to enemies (Proposition 2.3), implying that Theorem 3.1 cannot be directly applied to MFHG for resentful agents. Nevertheless, if we require the performed SCS deviations to be individually rational, then we can achieve convergence for a class of games containing MFHG (see Proposition 2.3).

Theorem 3.2. *The individually rational SCS, CS, and IS dynamics converge for resentful agents whose CAFs satisfy individually rational aversion to enemies and enemy monotonicity.*

It remains open whether general SCS, CS, or IS dynamics for resentful agents may cycle in an MFHG.

3.2 Contractual Nash Stability and Nash Stability

For individually rational NS dynamics, resentment helps to establish convergence for a wide class of games.

Theorem 3.3. *The individually rational NS dynamics converges for resentful agents whose CAFs satisfy enemy domination.*

Proof. Let a hedonic game (N, u^0) with resentful agents be given where every agent $i \in N$ has a CAF A_i that satisfies enemy domination. Assume for contradiction that there is an infinite execution of the individually rational NS dynamics $(\pi^t)_{t \geq 0}$. Suppose that, for every $t \geq 1$, π^t evolves from π^{t-1} by an individually rational NS deviation of agent d^t .

By Lemma 2.1, there exists $t_0 \geq 0$ such that every deviation performed after t_0 is performed infinitely often. We will reach a contradiction in two steps. First, we use enemy domination to show that no agent can ever be abandoned by an agent that she joined after t_0 . Then, as a second step, we use this insight to show the existence of a non-negative potential function decreasing in every time step after t_0 .

For the first step, let $i \in N$ be an agent and let $C \in \mathcal{N}_i$ be a coalition such that agent i performs a deviation at time $t_1 \geq t_0$ to form coalition C . Then, no agent from $C \setminus \{i\}$ can abandon agent i in any of their deviations after time t_0 . Assume for contradiction that there exists an agent $j \in C \setminus \{i\}$ who abandons agent i after time t_0 . Since the aggregation function satisfies enemy domination, there is a constant $c(u_i, j)$ such that for all utility vectors u'_i with $u'_i(k) \leq u_i^{t_1}(k)$ for all $k \in N$ and $u'_i(j) \leq c(u_i, j)$, it holds that $u'_i(C) < u'_i(\{i\})$.

Since every deviation occurs infinitely often, there exists a time $t_2 \geq t_1$ such that agent j has abandoned agent i for at least $u_i^{t_1}(j) - c(u_i, j)$ times between time t_1 and time t_2 . Since agent i is resentful, this implies that $u_i^{t_2}(j) \leq u_i^{t_1}(j) - (u_i^{t_1}(j) - c(u_i, j)) = c(u_i, j)$. Additionally, resentful agents can only decrease utilities for other agents. Therefore, it holds for all $t \geq t_2$ and $k \in N$ that $u_i^t(k) \leq u_i^{t_1}(k)$ and $u_i^t(j) \leq c(u_i, j)$. Consequently, it follows from enemy domination for all times $t \geq t_2$ that agent i cannot deviate to form coalition C again because this deviation would not be individually rational. However, this contradicts the fact that every deviation after time t_1 has to be performed infinitely often. This establishes the second step, i.e., that an agent can never be left by an agent that is part of a coalition which she joins after time t_0 .

For the second step, we will now define a potential function that is bounded from below by 0, integer-valued, and strictly decreasing in every step $t \geq t_1$. Therefore, fix an agent $i \in N$ and define $A_i = \{j \in N : j \text{ abandons } i \text{ after time } t_0\}$. Given $t \geq t_0$, we have to distinguish two cases. If i already performed a deviation after time t_0 , then let $d_i(t) = \max\{t_0 \leq t' \leq t : d^{t'} = i\}$ be the last time that i performed a deviation between t_0 and t . In this case, define $P_i(t) = \{C \subseteq N \setminus A_i : i \in C, u_i^t(C) > u_i^t(\pi^{d_i(t)}(i))\}$. Otherwise, set $P_i(t) = 2^{N \setminus A_i}$. Note that $P_i(t)$ contains all coalitions which i could *potentially* form through a deviation after time t . Define the potential $\Lambda(t) = \sum_{j \in N} |P_j(t)|$.

We observe that $P_i(t+1) = P_i(t)$ if $i \neq d^{t+1}$. Assume now that $i = d^{t+1}$. If the first deviation of i occurs at time t , then $P_i(t+1) \subseteq P_i(t) \setminus \{\pi^t(i)\}$. Otherwise, it must hold that $\pi^t(i) = \pi^{d_i(t)}(i)$, i.e., i is part of the same coalition in step t to which it deviated in step $d_i(t)$, which is the last deviation performed by i . Indeed, due to the second step, no agent that was present when i joined $\pi^{d_i(t)}(i) \setminus \{i\}$ can have left and i is not allowed to leave if any agent from $N \setminus \pi^{d_i(t)}(i)$ is still present. Also, agent i 's utilities for agents in $N \setminus A_i$ have not changed since her last deviation and therefore $u_i^t(j) = u_i^{d_i(t)}(j)$ for all $j \in N \setminus A_i$. Thus, i derives the same utility from $\pi^t(i)$ and $\pi^{d_i(t)}(i)$. It follows that $P_i(t+1) \subseteq P_i(t) \setminus \{\pi^t(i)\}$. Consequently, in each case, $\Lambda(t+1) < \Lambda(t)$. As $\Lambda(t) \geq 0$ for all $t \geq t_1$, the dynamics can run for at most $\Lambda(t_1)$ steps after time t_1 . This completes the proof. \square

As *AS* and *MF* satisfy enemy domination (Proposition 2.3), Theorem 3.3 implies that the individually rational NS dynamics converges in ASHG and MFHG for resentful agents. However, we do not know under which conditions resentment is sufficient to guarantee convergence for arbitrary (not necessarily individually rational) NS dynamics. In this case, our proof for Theorem 3.3 no longer works because it is possible that agents join coalitions for which they have an arbitrarily low utility (if the utility for their abandoned coalition was even worse). In fact, slightly counterintuitively, there is a non-trivial example of a cycling NS dynamics in an MFHG for resentful agents.

Theorem 3.4. *The NS dynamics may cycle in MFHG for resentful agents.*

This result indicates that some condition like aversion to enemies is probably needed for establishing a convergence guarantee for general NS dynamics; however, it remains open whether such a result is possible (even for ASHG). Notably, this question for CAFs satisfying aversion to enemies is the same as asking whether a CNS dynamics may cycle: For resentful agents in case of a cycling NS dynamics, there is also a cycling CNS dynamics.

Proposition 3.5. *For resentful agents with CAFs satisfying aversion to enemies, every sequence of NS deviations contains only finitely many deviations that are not CNS deviations.*

4 Dynamics for Appreciative Agents

We now turn to analyzing the effects of appreciation on the convergence of different types of dynamics. Here, as statements for general CAFs would require the introduction of (even) further axioms, we focus on *AS* and *MF* instead. We start by considering the possibility of cyclic NS and IS dynamics. Subsequently, we analyze CS and CNS dynamics.

4.1 From Resentment to Appreciation

We proved in Theorem 3.4 that NS dynamics may cycle in MFHG for resentful agents. ‘‘Reversing’’ this sequence and appropriately adjusting the initial utilities leads to a cycling NS dynamics for appreciative agents.

Theorem 4.1. *The individually rational NS dynamics may cycle in MFHGs for appreciative agents.*

For ASHG, we are not able to prove or disprove the convergence of NS dynamics. Still, we obtain an analogous result to Proposition 3.5 showing that the question whether there is a cycling NS dynamics is equivalent to asking for a cycling IS dynamics.

Proposition 4.2. *For appreciative agents in ASHG every sequence of NS deviations contains only finitely many deviations that are not IS deviations.*

The idea is that in every infinite sequence of NS deviations, there exists a certain time step from which on agent a has a positive utility for each agent b that joins a (because b has already joined a sufficiently often). However, note that the proof of Proposition 4.2 does not work for MFHGs, as in MFHGs (in contrast to ASHG) an agent a might veto the join of an agent b even if a has a positive utility for b . Thus, despite of Theorem 4.1, it remains open whether IS dynamics may cycle in MFHGs for appreciative agents.

Finally, we conjecture that it is possible to revert cyclic dynamics similar to the counterexamples for MFHGs to obtain the equivalence of cycling NS dynamics under resentment and appreciation.⁴ This would even show the equivalence of all four problems considered in Propositions 3.5 and 4.2.

4.2 Convergence for Appreciative Agents

We now consider stability concepts beyond Nash stability and give an overview under which circumstances appreciation is (not) sufficient to guarantee convergence in MFHGs and ASHG. In contrast to resentment, appreciation is not sufficient to guarantee convergence of CS dynamics.⁵

Theorem 4.3. *The individually rational CS dynamics may cycle in ASHG and MFHGs for appreciative agents.*

However, in the games considered in Theorem 4.3, there exists an execution of the CS dynamics that converges. This raises the (open) question whether a converging execution of the CS dynamics exists for every initial state in ASHG and MFHGs for appreciative agents.

Lastly, we consider CNS dynamics. While the convergence of CNS dynamics remains open under resentment, appreciation is sufficient to guarantee convergence. The proof idea is to consider an infinite execution of the dynamics and identify a pair of agents i and j such that i infinitely often joins a coalition containing j , whereas j infinitely often leaves a coalition containing i . We then show that it cannot be beneficial for i to join j at some point.

Theorem 4.4. *The CNS dynamics converges in ASHG for appreciative agents.*

Proof. Let an ASHG (N, u^0) be given and consider an execution $(\pi^t)_{t \geq 0}$ of the CNS dynamics. Assume for contradiction that the dynamics is infinite. By Lemma 2.1, there exists $t_0 \geq 0$ such that every deviation performed at time $t \geq t_0$ is performed infinitely often.

Now, consider an agent i that joins a coalition C containing agent j at some time $t \geq t_0$. We claim that it is impossible that j ever joins a coalition containing agent i after time t_0 . For the contrary, assume that this happens, and therefore happens infinitely often. Then, since agent i and agent j join each other infinitely often, there exists a time $t' \geq t_0$ such that $u_j^{t'}(i) > 0$ and $u_i^{t'}(j) > 0$ for all $t \geq t'$. Some time after t' , the agents i and j will be in a joint coalition, again. Then, to perform the deviation again where i joins C , the two agents have to be dissolved, i.e., one of them has to leave the other. However, since they by now have positive utility for each other, each of them would block the other agent from leaving according to the additively separable CAF. Hence, this cannot happen, a contradiction. Consequently, agent j cannot join a coalition containing i after time t_0 .

Thus, the utilities of an agent i for agents in coalitions which she joins by a deviation are not affected after time t_0 by appreciation. Therefore, there exists a global constant U such that the maximum utility of any agent obtained after any deviation is U . We derive a contradiction by showing that some agent has to abandon a coalition of unbounded utility.

To this end, we prove the following claim.

⁴The conference version of our paper contains a proof of this conjecture which unfortunately turned out to be erroneous.

⁵For ASHG, the next statement can be extended to an ASHG where initial valuations are symmetric by slightly modifying the game presented by Aziz et al. [2, Figure 2].

Table 2: Some of our experimental results on ASHG. The columns contain the name of the dynamics, the average number of steps until convergence, the average number of coalitions in the produced outcome, and the average maximum size of a coalition in the produced outcome.

	steps	#coalitions	max coalition size
Uniform utilities			
resentment	60 055	50	1
appreciation	4309	2.74	42.57
res+apprec	15 261	5.26	18.43
Gaussian utilities			
resentment	968	25.74	25.2
appreciation	1226	21.69	25.19
res+apprec	694	24.64	25.28

Claim 4.5. *There exists an agent $i \in N$ that abandons a coalition C containing an agent j that joins i at some point during the dynamics.*

Proof. Assume for contradiction that no such agent exists. To derive a contradiction, we will construct an infinite sequence of coalitions increasing in utility with respect to the utility at time t_0 . This cannot happen, because the number of coalitions an agent can be part of (and therefore the number of different utility values that an agent can achieve at a fixed time) is bounded.

We claim that there exists an agent $d \in N$ and a sequence of coalitions $(C_k)_{k \geq 0}$ such that for every $k \geq 1$, the following two conditions hold: (i) d 's utility with respect to time t_0 is strictly increasing, i.e., $u_d^{t_0}(C_k) > u_d^{t_0}(C_{k-1})$ and (ii) the coalition C_k is formed by a deviation of agent d .

Consider the first agent d performing a deviation after time t_0 , where d abandons coalition C_0 to form coalition C_1 . Then, $u_d^{t_0}(C_1) > u_d^{t_0}(C_0)$. Hence, we have found the first step of the sequence. Now, assume that we have constructed a sequence $(C_k)_{k=0}^m$ which satisfies the two conditions. We know that C_m is formed by a deviation of agent d . Consider the next time \hat{t} where some agent $e \in C_m$ leaves the coalition C containing C_m (which must happen, because the dynamics is infinite). If $e \neq d$, then we derive a contradiction to our initial assumption because e then leaves a coalition containing an agent, namely agent d , that joined her by a deviation. Hence, $e = d$. If there exists an agent $f \in C \setminus C_m$, then we again derive a contradiction because f must have joined d at some point. This implies that $C = C_m$. Set C_{m+1} to the coalition joined by agent d . Then, C_{m+1} clearly fulfills the second condition. Also, $u_d^{t_0}(C_{m+1}) = u_d^{\hat{t}}(C_{m+1}) > u_d^{\hat{t}}(C) = u_d^{\hat{t}}(C_m) = u_d^{t_0}(C_m)$. In the first and last equality, we use that appreciation does not affect the utilities for agents joined by agent d after time t_0 . Hence, also the first condition is fulfilled. As such a sequence of coalitions cannot exist, we derive a contradiction, and the claim must hold. \triangleleft

Now, consider an agent i that abandons a coalition C containing an agent j that joins i at some point during the dynamics. Since the deviation where agent j joins i happens infinitely often, there exists a time $T \geq t_0$ such that $u_i^t(j) \geq U - \sum_{l \in C \setminus \{j\}} u_i^{t_0}(l)$ for all $t \geq T$. Consider the next time $T' \geq T$ where i abandons C to form some coalition D . Then, $u_i^{T'}(D) \leq U \leq u_i^{T'}(j) + \sum_{l \in C \setminus \{j\}} u_i^{t_0}(l) \leq u_i^{T'}(C)$. In the last inequality, we use that the utility of i for other agents can only have increased since time t_0 (because of appreciation). Hence, this deviation was not beneficial for i . This is a final contradiction showing that an infinite dynamics cannot exist. \square

5 Simulations

We analyze by means of simulations how resentment and appreciation influence NS dynamics in ASHG by examining the speed of convergence and the composition of the reached stable states. This gives insights in the actual process that leads to convergence beyond the convergence guarantees and counterexamples presented before. We only provide a brief overview of some of our results; see Appendix D for details. We focus on ASHG with $n = 50$ agents and sample 100 games for each of the following utility models:

Uniform For two agents $a, b \in N$ with $a \neq b$, we sample $u_a(b)$ by sampling an integer from $[-100, 100]$.

Gaussian For each agent $a \in N$, we sample her *base qualification* μ_a by sampling an integer from $[-100, 100]$. For two agents $a, b \in N$ with $a \neq b$, we sample $u_a(b)$ by drawing an integer from the Gaussian distribution with mean μ_b and standard deviation 10.

Our dynamics start with the singleton partition. Subsequently, we perform an NS deviation selected uniformly at random until the dynamics converges. In addition to the concepts considered in our theoretical analysis, we also consider *resentful-appreciative agents*, i.e., agents that are both resentful and appreciative. Table 2 shows parts of our results.

Uniform Utilities The original NS dynamics without resentment or appreciation did not converge in any of our sampled games within a limit of 100 000 steps. In contrast to this, for resentful, appreciative and resentful-appreciative agents, NS dynamics always converged within this limit. However, resentful agents needed much longer (i.e., on average 60 005 steps) than resentful-appreciative agents (15 261 steps) and appreciative agents (4309 steps). Thus, while both resentment and appreciation are helpful to establish convergence, appreciation is more powerful than resentment here, and in fact adding resentment to appreciation rather hurts than helps (as the two can “cancel out”). Moreover, generally speaking, a lot of steps until convergence are still needed. In fact the produced outcomes for all three dynamics are quite “degenerated”: For resentful agents, in all games, all pairwise utilities have become non-positive resulting in a final outcome only consisting of singletons. For appreciative agents, there is typically one large coalition containing 40 or more agents together with one or two small coalitions. Notably, it does not happen here that eventually all utilities between pairs of agents are positive but only that certain pairwise utilities become large enough so that enough agents favor larger coalitions (even if their utility for some coalition members is negative). For resentful-appreciative agents, we usually have several medium-size coalitions in the produced outcomes which are thus in some sense in between outcomes for resentful agents and outcomes for appreciative agents. In fact, outcomes for resentful-appreciative agents also have a stronger connection to the initial utilities than for resentment or appreciation. On average, significantly a smaller number of agents has an NS deviation in the produced outcome with respect to their initial utilities.

Gaussian Utilities For Gaussian utilities, the original NS dynamics without resentment or appreciation converged for 3 of our 100 games within 100 000 steps. In contrast, for resentful, appreciative, and resentful-appreciative agents, NS dynamics converged in all games. In particular, convergence was much quicker (at most 2000 steps) than under uniform utilities, indicating that Gaussian utilities seem to facilitate reaching stable states in ASHG compared to uniform utilities. The difference between resentful agents (converging in on average 968 steps), appreciative agents (1226 steps), and resentful-appreciative agents (694 steps) is less profound here with resentful-appreciative agents converging fastest. Moreover, the outcomes produced by our three dynamics are quite similar and are in fact all quite close to being stable in the initial game (only around 10% of the agents have an NS deviation with respect to their initial utilities). The outcomes typically consist of one large coalition containing roughly half of the agents (these are usually the agents with a positive ground qualification), while other agents are placed into coalitions of size one or two.

6 Discussion

We initiated the study of hedonic games with time-dependent utility functions being influenced by previous deviations. In our theoretical analysis, we have investigated whether the resentful or appreciative perception of other agents is sufficient to guarantee convergence for dynamics based on various deviation types. We have posed several open questions throughout the paper (even showing the equivalence of some of them in Propositions 3.5 and 4.2). For future work, complementing our simulations, it would be interesting to theoretically analyze the effects of combining resentment and appreciation. A concrete open question here is whether CS dynamics are guaranteed to converge, which is the case for resentful agents but not for appreciative agents. More generally, it is also possible to consider other effects that could affect agents’ valuations over time and potentially contribute to additional convergence results.

Deviator-Resent One specific idea is deviator-resent, which models the restraint of a deviator to revert her decision to abandon other agents. In this case, an agent leaving coalition C to join coalition C' decreases her utility for all agents in $C \setminus C'$. We present results concerning deviator-resentment in Appendix E. An intuitive reason why deviator-resentment can contribute to the convergence of dynamics is that, after agent a abandons a coalition C , a 's utility for C decreases and thus a is less likely to join C again. However, deviator-resentment does not resolve the run-and-chase example, implying that NS dynamics may cycle for a wide variety of hedonic games with deviator-resentful agents. In contrast, for CNS dynamics, we show that deviator-resentment guarantees convergence in ASHG and MFHG if agents only deviate to non-singleton coalitions if they strictly prefer them to being alone. While this additional constraint might look arbitrary, we remark that it is needed as there are ASHG and MFHG with infinite sequences of individually rational CNS deviations. Deviator-resentment is nevertheless a powerful force for stability, as CS dynamics with individual rational deviations in MFHG and ASHG and IS dynamics in ASHG always converge. However, deviator-resentment is not sufficient to guarantee convergence of IS and general CS dynamics in MFHG, yielding a different behavior of ASHG and MFHG for IS dynamics. Notably, we did not prove any such contrasts in our analysis of resentment and appreciation. Overall, our results indicate that deviator-resentment has clear ramifications on convergence guarantees, yet the general picture seems to be slightly more nuanced than for resentment or appreciation.

Shortest Converge Sequences In our simulations, we have analyzed how fast random executions of NS dynamics converge for resentful and/or appreciative agents. An interesting related direction to shed further light on the power of resentment and appreciation is to analyze the length of the *shortest* converging execution of dynamics. The corresponding computational problem is to decide whether in a given game (where a stable outcome is guaranteed to be reachable), there is a converging deviation sequence of a given length from some given starting partition. Notably, this problem has not been addressed in the literature for classical dynamics, yet is of no less relevance in the general case. While existing hardness results for deciding whether a game admits a stable outcome suggest the hardness of the shortest converge sequence problem for the general case, they do not directly imply hardness, as stable states might not be reachable from some initial partition via the allowed deviations. Moreover, convergence might require exponential time from some initial partition [11]. In Appendix F, we present reductions showing that for ASHG deciding whether we can converge in a given number of steps is NP-hard for CS, IS, CNS, and NS dynamics for resentful, for appreciative, and for classical agents (with constant utility functions over time).

Comparison to Static Models To conclude, we consider the effects of resentment and appreciation in view of existing static utility models, and we want to discuss two phenomena that sound counterintuitive at first glance. First, we have seen that resentment is a strong force for convergence whereas appreciation is less helpful. However, in friend-and-enemy oriented ASHG [16], a friend-oriented approach, where agents focus on which friends are in their coalition, seems to be more favorable than an enemy-oriented approach in terms of achieving stable outcomes. For instance, a partition in the strict core under friends appreciation can be found in polynomial time, whereas finding a partition in the core under enemies aversion is NP-hard [16]. This behavior appears to be in contrast to our results because reducing utilities due to resentment seems to lead to games similar to the enemy-oriented approach, whereas appreciation has the flavor of the friend-oriented approach. Our explanation why resentment is still the stronger force for convergence is that it systematically eliminates “stability-breaking” friendships while appreciation cannot completely nullify such friendships.

Second, it is interesting to see that ASHG allow for stronger convergence results than MFHG. This is interesting because MFHG are generally more stabilizable than ASHG (see our discussion in the related work). For instance, there always exist partitions in the core of symmetric MFHG where every coalition is of size at most 2 [23], whereas it is Σ_2^P -complete to decide about the existence of core partitions in ASHG [29]. Our explanation here is that the simpler combinatorial structure of MFHG actually helps to achieve cycles in variable utility models. Indeed, the preferences in MFHG are invariant under a shift of all utilities by a common constant, which is not true in ASHG due to coalitions of different sizes.

Acknowledgments

This work was partially supported by the Deutsche Forschungsgemeinschaft under grants NI 369/22, BR 2312/11-2, and BR 2312/12-1, and the NRW project *Online Participation*. We would like to thank the participants of the Dagstuhl seminar 21331 on *Coalition Formation Games*, and especially Florian Brandl, Grégory Bonnet, Edith Elkind, Bettina Klaus, Seckin Özbilen, and Sanjukta Roy, for fruitful discussions.

References

- [1] H. Aziz and R. Savani. Hedonic games. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 15. Cambridge University Press, 2016.
- [2] H. Aziz, F. Brandt, and H. G. Seedig. Computing desirable partitions in additively separable hedonic games. *Artificial Intelligence*, 195:316–334, 2013.
- [3] H. Aziz, F. Brandl, F. Brandt, P. Harrenstein, M. Olsen, and D. Peters. Fractional hedonic games. *ACM Transactions on Economics and Computation*, 7(2):1–29, 2019.
- [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] V. Bilò, A. Fanelli, M. Flammini, G. Monaco, and L. Moscardelli. Nash stable outcomes in fractional hedonic games: Existence, efficiency and computation. *Journal of Artificial Intelligence Research*, 62:315–371, 2018.
- [7] V. Bilò, G. Monaco, and L. Moscardelli. Hedonic games with fixed-size coalitions. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*, pages 9287–9295, 2022.
- [8] A. Bogomolnaia and M. O. Jackson. The stability of hedonic coalition structures. *Games and Economic Behavior*, 38(2):201–230, 2002.
- [9] F. Brandt and M. Bullinger. Finding and recognizing popular coalition structures. *Journal of Artificial Intelligence Research*, 74:569–626, 2022.
- [10] F. Brandt, M. Bullinger, and A. Wilczynski. Reaching individually stable coalition structures in hedonic games. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, pages 5211–5218, 2021.
- [11] F. Brandt, M. Bullinger, and L. Tappe. Single-agent dynamics in additively separable hedonic games. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*, pages 4867–4874, 2022.
- [12] M. Bullinger. Pareto-optimality in cardinal hedonic games. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 213–221, 2020.
- [13] M. Bullinger. Boundaries to single-agent stability in additively separable hedonic games. In *Proceedings of the 47th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, pages 26:1–26:15, 2022.
- [14] R. Carosi, G. Monaco, and L. Moscardelli. Local core stability in simple symmetric fractional hedonic games. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 574–582, 2019.
- [15] K. Cechlárová and A. Romero-Medina. Stability in coalition formation games. *International Journal of Game Theory*, 29:487–494, 2001.

- [16] D. Dimitrov, P. Borm, R. Hendrickx, and S. C. Sung. Simple priorities and core stability in hedonic games. *Social Choice and Welfare*, 26(2):421–433, 2006.
- [17] J. H. Drèze and J. Greenberg. Hedonic coalitions: Optimality and stability. *Econometrica*, 48(4): 987–1003, 1980.
- [18] E. Elkind and M. Wooldridge. Hedonic coalition nets. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 417–424, 2009.
- [19] E. Elkind, A. Fanelli, and M. Flammini. Price of pareto optimality in hedonic games. *Artificial Intelligence*, 288:103357, 2020.
- [20] A. Fanelli, G. Monaco, and L. Moscardelli. Relaxed core stability in fractional hedonic games. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 182–188, 2021.
- [21] M. Gairing and R. Savani. Computing stable outcomes in symmetric additively separable hedonic games. *Mathematics of Operations Research*, 44(3):1101–1121, 2019.
- [22] M. Hoefer, D. Vaz, and L. Wagner. Dynamics in matching and coalition formation games with structural constraints. *Artificial Intelligence*, 262:222–247, 2018.
- [23] G. Monaco, L. Moscardelli, and Y. Velaž. Stable outcomes in modified fractional hedonic games. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 937–945, 2018.
- [24] M. Olsen. On defining and computing communities. In *Proceedings of the 18th Computing: Australasian Theory Symposium (CATS)*, volume 128, pages 97–102, 2012.
- [25] D. Peters. Precise complexity of the core in dichotomous and additive hedonic games. In *Proceedings of the 5th International Conference on Algorithmic Decision Theory (ADT)*, pages 214–227, 2017.
- [26] D. Peters and E. Elkind. Simple causes of complexity in hedonic games. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 617–623, July 2015.
- [27] S. Sung and D. Dimitrov. On myopic stability concepts for hedonic games. *Theory and Decision*, 62(1):31–45, 2007.
- [28] S. C. Sung and D. Dimitrov. Computational complexity in additive hedonic games. *European Journal of Operational Research*, 203(3):635–639, 2010.
- [29] G. Woeginger. A hardness result for core stability in additive hedonic games. *Mathematical Social Sciences*, 65(2):101–104, 2013.

A Additional Material for Section 2

We start this section by providing the proof for our lemma considering infinite sequences of partitions.

Lemma 2.1. *Let $(\pi^t)_{t \geq 0}$ be an infinite sequence of partitions induced by single-agent (or group) deviations. Then, there exists a $t_0 \geq 0$ such that every single-agent (or group) deviation performed at some time $t \geq t_0$ occurs infinitely often.*

Proof. We simultaneously prove the part of the lemma about single-agent deviations and group deviations by showing a stronger lemma that implies both. To this end, we consider labeled transitions. A *labeled sequence* with respect to μ is a sequence $a = (a^t)_{t \geq 0}$ together with a (labeling) function $\mu: \mathbb{N}_{\geq 0} \rightarrow L$. Given a labeled sequence $a = (a^t)_{t \geq 0}$ with respect to μ , a *labeled transition* of a with respect to L is a tuple (l, a_1, a_2) such that there exists $t \geq 0$ with $a^t = a_1$, $a^{t+1} = a_2$, and $\mu(t) = l$. Then, Lemma 2.1 follows from the next claim by interpreting the deviating single agent or group of agents as labels of a transition between partitions.

Claim A.1. *Let N be a non-empty and finite ground set and L be a non-empty and finite set of labels. Let $a = (a^t)_{t \geq 0}$ be an infinite labeled sequence over A with respect to μ where $a^t \in N$ for all $t \geq 0$. Then, there exists a $t_0 \geq 0$ such that every labeled transition of a with respect to L performed at some time $t \geq t_0$ occurs infinitely often.*

Proof. Let $a = (a^t)_{t \geq 0}$ be an infinite labeled sequence with respect to μ over finite and non-empty ground sets and label sets. Given $a_1, a_2 \in N$ and $l \in L$, define $t(l, a_1, a_2) = \sup\{t \geq 0: a^t = a_1, a^{t+1} = a_2, \mu(t) = l\}$, where we set $\sup \emptyset = 0$. In other words, $t(l, a_1, a_2)$ defines the last time step where the labeled transition (l, a_1, a_2) occurs. We define $t_0 = \max\{t(l, a_1, a_2): l \in L, a_1, a_2 \in N, t(l, a_1, a_2) < \infty\}$. By construction, every labeled transition performed after t_0 must occur infinitely often. \triangleleft

This completes the proof of the lemma. \square

For the remainder of this section, we will have a closer look at the properties of hedonic aggregation functions. In addition to the axioms introduced in Section 2, we also define two more axioms that are relevant for the results in Appendix E. We start by formally defining all axioms. For convenience, we also give definitions in words similar to the definition in Section 2. The additional axioms *friend necessity* and *single friend desire* are weak conditions about friends.

For any $i \in N$, a CAF A_i satisfies

- *aversion to enemies* (ATE) if, for all coalitions $C \in \mathcal{N}_i$, agents $j \in C \setminus \{i\}$, and utility vectors $u_i \in \mathbb{Q}^n$ with $u_i(j) < 0$, it holds that $A_i(C, u_i) \leq A_i(C \setminus \{j\}, u_i)$. In other words, the aggregated utility is weakly better whenever an enemy leaves i 's coalition.
- *individually rational aversion to enemies* (IR ATE) if, for all coalitions $C \in \mathcal{N}_i$, agents $j \in C \setminus \{i\}$, and utility vectors $u_i \in \mathbb{Q}^n$ with $u_i(j) < 0$, it holds that $A_i(C, u_i) \leq A_i(C \setminus \{j\}, u_i)$ if $A_i(C, u_i) \geq A_i(\{i\}, u_i)$. In other words, the aggregated utility is weakly better when an enemy leaves one of i 's individually rational coalitions.
- *enemy monotonicity* (EM) if, for all coalitions $C \in \mathcal{N}_i$, agents $j \in N$, and utility vectors $u_i, u'_i \in \mathbb{Q}^n$ with $u_i(k) = u'_i(k)$ for all $k \neq j$ and $u'_i(j) < u_i(j) < 0$, it holds that $A_i(C, u_i) \geq A_i(C, u'_i)$. In other words, decreasing the utility for an enemy cannot improve a coalition value.
- *enemy domination* (ED) if, for all utility vectors $u_i \in \mathbb{Q}^n$ and agents $j \in N \setminus \{i\}$, there exists a constant $c(u_i, j)$ such that for all utility vectors $u'_i \in \mathbb{Q}^n$ with $u'_i(k) \leq u_i(k)$ for all $k \in N$ and $u'_i(j) \leq c(u_i, j)$, it holds for every $C \in \mathcal{N}_i$ with $j \in C$ that $A_i(C, u'_i) < A_i(\{i\}, u'_i)$. In other words, if the utility for some agent is sufficiently negative and the utility for every other agent is bounded, then no coalition containing this agent is individually rational.
- *friend necessity* (FN) if, for all coalitions $C \in \mathcal{N}_i$ and utility vectors $u_i \in \mathbb{Q}^n$, $A_i(C, u_i) > 0$ implies that there exists $j \in C \setminus \{i\}$ with $u_i(j) > 0$. In other words, an agent can only have a positive utility for a coalition if it contains a friend.

- *single friend desire* (SFD) if, for all coalitions $C \in \mathcal{N}_i$, agents $j \in C \setminus \{i\}$, and utility vectors $u_i \in \mathbb{Q}^n$ such $u_i(j) > 0$ and $u_i(k) \leq 0$ for all $k \in C \setminus \{j\}$, it holds that $A_i(C, u_i) > A_i(C \setminus \{j\}, u_i)$. In other words, a coalition to which exactly one friend belongs is strictly preferred to the same coalition without this friend.

The proof of Proposition 2.3 is integrated in the next more general proposition considering all axioms.

Proposition A.2. *The additively separable CAF AS_i satisfies ATE, IR ATE, FN, EM, ED, and SFD. The modified fractional CAF MF_i satisfies IR ATE, FN, EM, ED, and SFD but violates ATE.*

Proof. We first consider additively separable aggregation and each axiom separately.

- ATE holds because a sum gets smaller when removing a negative summand.
- IR ATE follows from ATE.
- FN holds because a sum can only be positive if some summand is positive.
- EM holds because a sum can only get smaller when decreasing every summand (and decreasing one of them strictly).
- ED holds because $\sum_{k \in C \setminus \{i\}} u_i(k)$ gets negative if all summands only diminish and, for $j \in C \setminus \{i\}$, $u_i(j) \leq -1 - \sum_{k \in C \setminus \{j\}} u_i(k)$. Hence, we can choose the constant $c(u_i, j) = \min_{C \in \mathcal{N}_i: j \in C} -1 - \sum_{k \in C \setminus \{j\}} u_i(k)$.
- SFD holds because a sum gets larger when adding a positive number.

Next, we consider modified fractional aggregation. The violation of ATE is considered in Example 2.2.

- IR ATE: Let $C \subseteq \mathcal{N}_i$, $j \in C \setminus \{i\}$, and $u_i \in \mathbb{Q}^n$ with $u_i(j) < 0$. Moreover, suppose that $MF_i(C, u_i) \geq 0$. Then, $|C| \geq 3$ because otherwise $MF_i(C, u_i) = u_i(j) < 0$. Consequently,

$$\begin{aligned}
& MF_i(C \setminus \{j\}, u_i) \geq MF_i(C, u_i) \\
\Leftrightarrow & \frac{\sum_{k \in C \setminus \{i, j\}} u_i(k)}{|C| - 2} \geq \frac{\sum_{k \in C \setminus \{i\}} u_i(k)}{|C| - 1} \\
\Leftrightarrow & (|C| - 1) \sum_{k \in C \setminus \{i, j\}} u_i(k) \geq (|C| - 2) \sum_{k \in C \setminus \{i\}} u_i(k) \\
\Leftrightarrow & \sum_{k \in C \setminus \{i, j\}} u_i(k) \geq (|C| - 2) u_i(j) \\
\Leftrightarrow & \frac{\sum_{k \in C \setminus \{i\}} u_i(k)}{|C| - 1} \geq u_i(j).
\end{aligned}$$

Hence, whenever $u_i(j) < 0$ and $MF_i(C, u_i) \geq 0$, the utility increases in the coalition where j has left.

- FN, EM, ED, and SFD hold for the same reasons as for ASHG's because the division by a positive number does not change the respective arguments.

□

B Additional Material for Section 3

We start this section by considering the convergence of individually rational SCS, CS, and IS dynamics for resentful agents.

Theorem 3.2. *The individually rational SCS, CS, and IS dynamics converge for resentful agents whose CAFs satisfy individually rational aversion to enemies and enemy monotonicity.*

Proof. Again, it suffices to consider SCS dynamics.

Let a hedonic game (N, u^0) with resentful agents be given where every agent $i \in N$ has a CAF A_i that satisfies individually rational aversion to enemies and enemy monotonicity. Assume that agents perform only individually rational deviations. Assume for contradiction that there exists an infinite execution of the individually rational SCS dynamics. Since the single-agent utilities for other agents can only decrease and are bounded by the initial partition of the sequence, there exists a time $t_0 \geq 0$ in the infinite dynamics after which no agent can be left by an agent for which she has non-negative utility. By Lemma 2.1, there exists a time $t_1 \geq t_0$ such that every SCS deviation performed after time t_1 is performed infinitely often.

Now, consider the first group deviation performed after time t_1 , where a coalition C is formed and let $i \in C$ be an arbitrary agent in this coalition. Note that the formation of C was an individually rational deviation, and therefore C is individually rational for agent i at time t_1 . We claim that i 's utility cannot decrease until the next time when the same deviation is performed, and that all coalitions that i is part of until then are IR. Until the next repetition of the same deviation, there are two potential cases when i 's utility is affected. First, it can happen that i is part of a deviating coalition. Clearly, this cannot decrease her utility, and therefore not affect her individual rationality. Second, it can happen that i 's coalition is left by a set of agents D . Since i is resentful and decreases her utility for all agents in D , it must be the case that her utility is negative for every agent in D at the point in time where the deviation involving D occurs.

Moreover, i 's coalition is IR when D leaves. Indeed, assume for contradiction that i 's coalition is not IR, and consider the first time after the formation of C when i is in a coalition that is not IR. By our analysis before, this can only happen after i performed a group deviation, but since this deviation originated from an IR coalition, the resulting coalition must also be IR, a contradiction.

Now, by applying individually rational aversion to enemies for each agent in D one after another, we can conclude that i cannot have decreased her utility at the point in time where she is left by D .

Hence, at the next time, when coalition C is formed, by the same deviation as our initial deviation, none of the agents in C has decreased her utility. Moreover, enemy monotonicity implies that the utility of C of any agent for C can only have decreased since the last time when C was formed. Hence, no agent can strictly improve her utility when forming C again, a contradiction. We conclude that the dynamics cannot run infinitely. \square

Theorem 3.4. *The NS dynamics may cycle in MFHG for resentful agents.*

Proof. We now describe an involved example of an MFHG together with an infinite periodic sequence of NS deviations for resentful agents. We construct this example in a way such that each agent leaves every other agent exactly once in each cycle. This establishes that the agents' preference between relevant coalitions is *maintained*: if a deviating agent prefers a joined coalition C_1 to an abandoned coalition C_2 before (and during) the *first* execution of the cycle, then it still prefers C_1 to C_2 before (and during) *each* execution of the cycle.

Consider the game with agent set $N = \{a, a', b, b', c, c'\}$ and utilities as depicted in Table 3. The initial utilities result from setting $x = 0$ in Table 3. Note that the utility values are not chosen to be minimal but simply in a way that it can easily be verified that deviations are indeed NS deviations. We now present an infinite sequence $(\pi^t)_{t \geq 0}$ of partitions, always consisting of three coalitions. For each partition, we refer to the first listed coalition as C_1 , to the second as C_2 , and the third as C_3 . For the sake of clarity, for each partition, we also specify which agent deviates to which coalition in the next step. Specifically, for $n \geq 0$, we have

- $\pi^{18n+1} = \{\{b', a'\}, \{a\}, \{c', c, b\}\}$ with agent b deviating to C_1 ,
- $\pi^{18n+2} = \{\{b', a', b\}, \{a\}, \{c', c\}\}$ with agent c deviating to C_1 ,
- $\pi^{18n+3} = \{\{b', a', b, c\}, \{a\}, \{c'\}\}$ with agent a' deviating to C_3 ,
- $\pi^{18n+4} = \{\{b', b, c\}, \{a\}, \{c', a'\}\}$ with agent a' deviating to C_2 ,
- $\pi^{18n+5} = \{\{b', b, c\}, \{a, a'\}, \{c'\}\}$ with agent c deviating to C_2 ,
- $\pi^{18n+6} = \{\{b', b\}, \{a, a', c\}, \{c'\}\}$ with agent b' deviating to C_3 ,

Table 3: Example from Theorem 3.4. Utilities after x cycles of deviations. Each row depicts the utility that one agent has for the other agents.

	a	a'	b	b'	c	c'
a	–	$20 - x$	$10 - x$	$230 - x$	$0 - x$	$230 - x$
a'	$110 - x$	–	$30 - x$	$120 - x$	$30 - x$	$100 - x$
b	$0 - x$	$230 - x$	–	$20 - x$	$10 - x$	$230 - x$
b'	$30 - x$	$100 - x$	$110 - x$	–	$30 - x$	$120 - x$
c	$10 - x$	$230 - x$	$0 - x$	$230 - x$	–	$20 - x$
c'	$30 - x$	$120 - x$	$30 - x$	$100 - x$	$110 - x$	–

- $\pi^{18n+7} = \{\{b\}, \{a, a', c\}, \{c', b'\}\}$ with agent c deviating to C_3 ,
- $\pi^{18n+8} = \{\{b\}, \{a, a'\}, \{c', b', c\}\}$ with agent a deviating to C_3 ,
- $\pi^{18n+9} = \{\{b\}, \{a'\}, \{c', b', c, a\}\}$ with agent b' deviating to C_2 ,
- $\pi^{18n+10} = \{\{b\}, \{a', b'\}, \{c', c, a\}\}$ with agent b' deviating to C_1 ,
- $\pi^{18n+11} = \{\{b, b'\}, \{a'\}, \{c', c, a\}\}$ with agent a deviating to C_1 ,
- $\pi^{18n+12} = \{\{b, b', a\}, \{a'\}, \{c', c\}\}$ with agent c' deviating to C_2 ,
- $\pi^{18n+13} = \{\{b, b', a\}, \{a', c'\}, \{c\}\}$ with agent a deviating to C_2 ,
- $\pi^{18n+14} = \{\{b, b'\}, \{a', c', a\}, \{c\}\}$ with agent b deviating to C_2 ,
- $\pi^{18n+15} = \{\{b'\}, \{a', c', a, b\}, \{c\}\}$ with agent c' deviating to C_1 ,
- $\pi^{18n+16} = \{\{b', c'\}, \{a', a, b\}, \{c\}\}$ with agent c' deviating to C_3 ,
- $\pi^{18n+17} = \{\{b'\}, \{a', a, b\}, \{c, c'\}\}$ with agent b deviating to C_3 , and
- $\pi^{18n+18} = \{\{b'\}, \{a', a\}, \{c, c', b\}\}$ with agent a' deviating to C_1 .

Then, it is possible to verify that for $k \geq 1$, π^{k-1} leads to π^k by means of an NS deviation. Hence, we have presented an MFHG with an infinite sequence of NS deviations for resentful agents. \square

We conclude this section by examining the occurrence of CNS deviations in NS dynamics.

Proposition 3.5. *For resentful agents with CAFs satisfying aversion to enemies, every sequence of NS deviations contains only finitely many deviations that are not CNS deviations.*

Proof. Let a hedonic game (N, u^0) with resentful agents be given where every agent $i \in N$ has a CAF A_i that satisfies aversion to enemies. Furthermore, assume that there exists an infinite sequence $(\pi^t)_{t \geq 0}$ of partitions resulting from NS deviations of resentful agents. By Lemma 2.1, there exists a time $t_0 \geq 0$ such that every deviation performed after t_0 must occur infinitely often.

Define $\mathcal{L} = \{(i, j) \in N^2 : i \text{ left by } j \text{ after time } t_0\}$, i.e., the set of pairs of such agents. Let $(i, j) \in \mathcal{L}$. Then, there exists a time $t(i, j) \geq t_0$ such that i was left by j for at least $\lfloor u_i^{t_0}(j) \rfloor + 1$ times after time t_0 and before time $t(i, j)$. Consider the time $t_1 = \max\{t(i, j) : (i, j) \in \mathcal{L}\}$. We claim that all deviations after time t_1 are CNS deviations. Indeed, assume that agent i is left by agent j at time $t \geq t_1$. By construction, as $t(i, j) \leq t_1$, it holds that $u_i^t(j) < 0$. Consequently, enemy monotonicity implies that $A_i(\pi^t, u_i^t) \leq A_i(\pi^{t+1}, u_i^t)$. Hence, every deviation after time t_1 is a CNS deviation. Thus, there are only finitely many deviations (at most t_1 many) that are not CNS deviations. \square

C Additional Material for Section 4

Theorem 4.1. *The individually rational NS dynamics may cycle in MFHGs for appreciative agents.*

Table 4: Example from Theorem 4.1. Utilities after x cycles of deviations. Each row depicts the utility that one agent has for the other agents.

	a	a'	b	b'	c	c'
a	-	$110 + x$	$120 + x$	$-100 + x$	$130 + x$	$-100 + x$
a'	$20 + x$	-	$100 + x$	$10 + x$	$100 + x$	$30 + x$
b	$130 + x$	$-100 + x$	-	$110 + x$	$120 + x$	$-100 + x$
b'	$100 + x$	$30 + x$	$20 + x$	-	$100 + x$	$10 + x$
c	$120 + x$	$-100 + x$	$130 + x$	$-100 + x$	-	$110 + x$
c'	$100 + x$	$10 + x$	$100 + x$	$30 + x$	$20 + x$	-

Proof. We now describe an involved example of an MFHG together with an infinite periodic sequence of NS deviations for appreciative agents. Our example is very similar to the one presented for resentful agents in Theorem 3.4. As in Theorem 3.4, each agent joins each other agent exactly once in each cycle. This establishes the same invariant: if a deviating agent prefers a joined coalition C_1 to an abandoned coalition C_2 before (and during) the *first* execution of the cycle, then it still prefers C_1 to C_2 before (and during) *each* execution of the cycle.

Consider the game with agent set $N = \{a, a', b, b', c, c'\}$ and utilities as depicted in Table 4. The initial utilities result from setting $x = 0$ in Table 4. We now present an infinite sequence $(\pi^t)_{t \geq 0}$ of partitions, always consisting of three coalitions. For each partition, we refer to the first listed coalition as C_1 , to the second as C_2 , and the third as C_3 . For the sake of clarity, for each partition, we also specify which agent deviates to which coalition in the next step. Specifically, for $n \geq 0$, we have

- $\pi^{18n+1} = \{\{b'\}, \{a, a'\}, \{b, c, c'\}\}$ with agent b deviating to C_2 ,
- $\pi^{18n+2} = \{\{b'\}, \{b, a, a'\}, \{c, c'\}\}$ with agent c' deviating to C_1 ,
- $\pi^{18n+3} = \{\{b', c'\}, \{b, a, a'\}, \{c\}\}$ with agent c' deviating to C_2 ,
- $\pi^{18n+4} = \{\{b'\}, \{b, a, a', c'\}, \{c\}\}$ with agent b deviating to C_1 ,
- $\pi^{18n+5} = \{\{b', b\}, \{a, a', c'\}, \{c\}\}$ with agent a deviating to C_1 ,
- $\pi^{18n+6} = \{\{b', b, a\}, \{a', c'\}, \{c\}\}$ with agent c' deviating to C_3 ,
- $\pi^{18n+7} = \{\{b', b, a\}, \{a'\}, \{c, c'\}\}$ with agent a deviating to C_3 ,
- $\pi^{18n+8} = \{\{b', b\}, \{a'\}, \{c, c', a\}\}$ with agent b' deviating to C_2 ,
- $\pi^{18n+9} = \{\{b\}, \{a', b'\}, \{c, c', a\}\}$ with agent b' deviating to C_3 ,
- $\pi^{18n+10} = \{\{b\}, \{a'\}, \{c, c', a, b'\}\}$ with agent a deviating to C_2 ,
- $\pi^{18n+11} = \{\{b\}, \{a', a\}, \{c, c', b'\}\}$ with agent c deviating to C_2 ,
- $\pi^{18n+12} = \{\{b\}, \{a', a, c\}, \{c', b'\}\}$ with agent b' deviating to C_1 ,
- $\pi^{18n+13} = \{\{b, b'\}, \{a', a, c\}, \{c'\}\}$ with agent c deviating to C_1 ,
- $\pi^{18n+14} = \{\{b, b', c\}, \{a', a\}, \{c'\}\}$ with agent a' deviating to C_3 ,
- $\pi^{18n+15} = \{\{b, b', c\}, \{a\}, \{c', a'\}\}$ with agent a' deviating to C_1 ,
- $\pi^{18n+16} = \{\{b, b', c, a'\}, \{a\}, \{c'\}\}$ with agent c deviating to C_3 ,
- $\pi^{18n+17} = \{\{b, b', a'\}, \{a\}, \{c', c\}\}$ with agent b deviating to C_3 ,
- $\pi^{18n+18} = \{\{b', a'\}, \{a\}, \{c', c, b\}\}$ with agent a' deviating to C_2 .

Then, it is possible to verify that for $k \geq 1$, π^{k-1} leads to π^k by means of an NS deviation. Hence, we have presented an MFHG with an infinite sequence of NS deviations for appreciative agents. \square

Proposition 4.2. *For appreciative agents in ASHG every sequence of NS deviations contains only finitely many deviations that are not IS deviations.*

Proof. Let an ASHG (N, u^0) with appreciative agents be given. Furthermore, assume that there exists an infinite sequence $(\pi^t)_{t \geq 0}$ of partitions resulting from NS deviations. By Lemma 2.1, there exists a time t_0 such that every deviation performed after t_0 occurs infinitely often.

Define $\mathcal{L} = \{(i, j) \in N^2: i \text{ is joined by } j \text{ after time } t_0\}$. Let $(i, j) \in \mathcal{L}$. Then, there exists a time $t(i, j) \geq t_0$ such that i was joined by j for at least $\lceil |u_i^{t_0}(j)| \rceil + 1$ times after time t_0 and before time $t(i, j)$. Consider the time $t_1 = \max\{t(i, j): (i, j) \in \mathcal{L}\}$. We claim that all deviations after time t_1 are IS deviations. Indeed, assume that agent i is joined by agent j at time $t \geq t_1$. By construction, as $t(i, j) \leq t_1$, it holds that $u_i^t(j) > 0$. Consequently, agent i prefers π^{t+1} to π^t . Hence, every deviation after time t_1 is an IS deviation. Thus, there are only finitely many deviations (at most t_1 many) that are not IS deviations. \square

Theorem 4.3. *The individually rational CS dynamics may cycle in ASHG and MFHG for appreciative agents.*

Proof. Let $N = \{a, b, c\}$ be the set of agents and let the agents' initial utilities be as follows:

$$u_a^0(b) = u_b^0(c) = u_c^0(a) = 4, \quad u_a^0(c) = u_b^0(a) = u_c^0(b) = 1.$$

Let $\pi^0 = \{\{a\}, \{b\}, \{c\}\}$ and for $t > 0$, let

$$\pi^t = \begin{cases} \{\{a, c\}, \{b\}\}, & \text{if } t \bmod 3 = 0 \\ \{\{a, b\}, \{c\}\}, & \text{if } t \bmod 3 = 1 \\ \{\{a\}, \{b, c\}\}, & \text{if } t \bmod 3 = 2. \end{cases}$$

We claim that $(\pi^t)_{t \geq 0}$ is an infinite sequence, where for every $t \geq 1$, π^t evolves from π^{t-1} by a core deviation of coalition C^t . Specifically, we have

$$C^t = \begin{cases} \{a, c\}, & \text{if } t \bmod 3 = 0 \\ \{a, b\}, & \text{if } t \bmod 3 = 1 \\ \{b, c\}, & \text{if } t \bmod 3 = 2. \end{cases}$$

Thus in each cycle of length three, each agent performs exactly one core deviation with any other agent. Thus, for $t \geq 0$ with $t \bmod 3 = 0$ it holds that

$$u_a^t(b) = u_b^t(c) = u_c^t(a) = \frac{2}{3}t + 4$$

and

$$u_a^t(c) = u_b^t(a) = u_c^t(b) = \frac{2}{3}t + 1.$$

Thus, for each $t > 0$, it holds that $u_a^t(b) > u_a^t(c)$, $u_b^t(c) > u_b^t(a)$, and $u_c^t(a) > u_c^t(b)$, implying that each of the deviations C^t for $t > 0$ is a core deviation if each agent $i \in N$ aggregates utilities according to AS_i or MF_i . \square

D Simulations

In this section, we analyze by means of simulations how resentment and appreciation influence dynamics in ASHG. We focus on NS dynamics, as in randomly sampled ASHG the IS dynamics typically converge quickly even without resentment or appreciation (implying that they have only a small effect). Moreover, executing core dynamics is computationally too costly, as already checking whether an outcome is core stable is computationally intractable.

D.1 Setup

We mostly focus on ASHG with an agent set N containing $n = 50$ agents,⁶ and sample their utilities using one of the following two models:

⁶We analyze the influence of the number of agents in Appendix D.5.

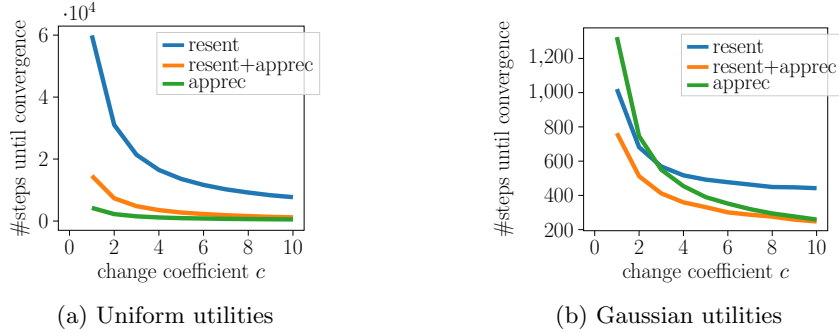


Figure 2: Influence of the change coefficient on the average convergence time

Uniform For two agents $a, b \in N$ with $a \neq b$, we sample $u_a(b)$, i.e., a 's value for b , by drawing a random integer between -100 and 100 .

Gaussian For each agent $a \in N$, we sample her *base qualification* μ_a by drawing a random integer between -100 and 100 . For two agents $a, b \in N$ with $a \neq b$, we sample $u_a(b)$ by drawing an integer from the Gaussian distribution with mean μ_b and standard deviation 10 .⁷

Our dynamics start with the singleton partition. Subsequently, in each step, we compute all possible NS deviations. If there are no NS deviations, we stop; otherwise, we sample one NS deviation uniformly at random and execute it. To be able to vary the “intensity” of the resentful/appreciative perception, we introduce a change coefficient c , which we typically set to 1: For resentful agents, if an agent a deviates from a coalition C' to a coalition C , then we reduce the utility that agents from C' have for a by c ; for appreciative agents, we increase the utility that agents from C have for a by c . We also examine what happens if agents are both resentful and appreciative and both of the above described effects are present. In this case we speak of resentful-appreciative agents. For all our simulations, we set a time-out of 100 000, i.e., after 100 000 steps we report that the dynamics did not converge.⁸

D.2 Convergence Time

We start by analyzing the influence of resentment and of appreciation on how fast NS dynamics converge. For this, we sampled 100 games with 50 agents and for each recorded the number of steps until convergence.

Uniform Utilities In Figure 2a, we visualize the results for uniform utilities. The NS dynamics for agents that are neither resentful nor appreciative did not converge in any of our sampled games (within the limit of 100 000 steps). By contrast, even for a change coefficient $c = 1$, NS dynamics converged in all games for resentful or appreciative agents. However, there is a clear difference between these two: For resentful agents, the average number of steps until convergence is 59 910 for $c = 1$, whereas for appreciative agents the NS dynamics converges much faster (for $c = 1$ the average convergence time is 4 243). Increasing c to 4, for resentful agents and for appreciative agents, the average convergence time roughly quarters, while increasing c to 10 only decreases the time by an additional factor of two. In sum, appreciation seems more helpful to establish fast convergence than resentment in case of uniform utilities. Nevertheless, for both concepts, the number of steps until convergence is quite large (compared to the number of agents). While increasing the change coefficient leads to faster convergence the decrease in convergence time is particularly strong for smaller values of the coefficient, indicating that NS dynamics need some time to find the “right” deviations somewhat independent of the value of the change coefficient.

In contrast to our theoretical analysis, we also consider the case of resentful-appreciative agents. Compared to appreciative agents, adding resentment leads to a substantial increase of the convergence

⁷We analyze the influence of the chosen standard deviation in Appendix D.4.

⁸We want to remark that this does not necessarily imply that no NS stable outcome exists in such a game or that there is no path to stability for the dynamics, but rather that selecting NS deviations randomly was not sufficient to ensure convergence (in a reasonable time).

Table 5: Properties of different dynamics and the produced outcomes for different ASHG. For an explanation of the table, we refer to the beginning of Appendix D.3.

	steps	coalition sizes			stability in original		utilities		
		number	avg. size	max size	# ag. IR viol	# ag. NS dev	avg. util	avg. change	pos. vals
Uniform utilities $n = 50$									
resentment	60055.0	50	1.0	1	0	50	0.0	-51.25	0.0
appreciation	4309.0	2.74	19.31	42.57	18.7	22.21	589.39	12.8	0.56
res+apprec	15261.0	5.26	9.73	18.43	2.33	10.15	237.08	-0.13	0.51
Gaussian utilities $n = 50$									
resentment	968.0	25.74	1.99	25.2	0	5.07	626.3	-0.42	0.49
appreciation	1226.0	21.69	2.36	25.19	3.67	5.11	638.79	0.6	0.51
res+apprec	694.0	24.64	2.07	25.28	0.61	4.66	637.34	-0.12	0.5
Gaussian utilities $n = 25$									
resentment	282.0	13.62	1.91	12.36	0	2.11	291.49	-0.49	0.48
appreciation	438.0	11.92	2.2	12.29	1.76	2.24	296.59	0.77	0.5
res+apprec	208.0	13.32	1.95	12.42	0.18	1.97	297.54	-0.12	0.49
Gaussian utilities $n = 25$ with stable outcome (13 games)									
original	100.38	12.46	2.15	13.54	0	0	344.26	0.0	0.52
resentment	90.69	12.46	2.15	13.54	0	0	343.14	-0.16	0.52
appreciation	92.77	12.31	2.24	13.54	0.15	0.15	348.78	0.29	0.52
res+apprec	100	12.46	2.15	13.54	0	0	347.82	-0.05	0.52

time by a factor of around 2.5 (while for resentful agents, adding appreciation still leads to faster convergence). While this slower convergence for resentful-appreciative agents compared to appreciative agents may be surprising at first glance, recall the intuitive justifications why resentment and appreciation contribute to a faster convergence. For appreciative agents, utilities only increase over time, whereas for resentful agents utilities only decrease over time. Thus, if we combine the two, it is in principle possible that some valuations that increase for appreciative agents stay constant for resentful-appreciative agents, and the two effects can cancel out each other. For uniform utilities, this effect seems to be stronger than the additional “stability force” established by resentment.

Gaussian Utilities In Figure 2b, we visualize the result of our first set of simulations for Gaussian utilities. In this case, the NS dynamics for agents that are neither resentful nor appreciative converged in 3 of the 100 games. In contrast, for resentful or for appreciative agents, the NS dynamics converged in all games. In particular, convergence was much quicker (in at most 2000 steps) than under uniform utilities, indicating that ASHG under Gaussian utilities seems to facilitate reaching stable states compared to uniform utilities. Examining the results in more detail, the difference between resentful and appreciative agents is less profound here than for uniform utilities. While resentment leads to faster convergence for $c < 3$, appreciation is more powerful for $c \geq 3$. Considering the influence of the change coefficient, for resentful agents, we again see the trend that increasing the change coefficient has a strong effect for smaller c but that this effect becomes less strong for larger c . On the other hand, for appreciative agents, the relation between c and the convergence time is rather linear. Moreover, in contrast to uniform utilities, resentment and appreciation seem to not “cancel out” each other. For resentful-appreciative agents, NS dynamics converge faster than for either of the two separately.

D.3 Structure of Outcomes and Comparison to Base Game

We now take a closer look at the outcomes and utility functions produced by NS dynamics for resentful and/or appreciative agents. Again, we generated 100 games with $n = 50$ agents each for uniform and Gaussian utilities. In addition, we generated 100 games with $n = 25$ agents and Gaussian utilities (as the original NS dynamics converge more often in such games). In all games, we set the change coefficient c to 1. Table 5 summarizes the results of our simulations. All values in the table are averaged from the respective 100 games. In the last part of the table, we only consider the 13 games with $n = 25$ and Gaussian utilities in which an NS dynamics in the original game converged within 100 000 steps. For reference, we depict the number of steps the dynamics needed to converge in the first column. We analyze the structure of the produced NS outcomes as follows. Columns three to six consider the produced coalitions, that is, the number of coalitions and their average and maximum size. The next two

columns concern the outcome’s degree of stability with respect to the original utilities, where we record the number of agents violating individual rationality and possessing an NS deviation, respectively. The last three columns concern the change of the utility profile, that is, the average utility in the outcome partition with respect to the final utilities, the average change of each entry of the utility function comparing the initial and final utilities, and the fraction of pairs of friends with respect to the final utilities, that is tuples $(a, b) \in N^2$ for which $u_a(b) > 0$.⁹

Uniform Utilities We first focus on uniform utilities. Recall that our dynamics need many steps until reaching convergence. Therefore, it is not surprising that the produced outcomes are quite “degenerated” for both resentful agents and appreciative agents. For resentful agents, the produced outcomes consist only of singleton coalitions in all games, which means that agents have left each other sufficiently often to ensure that all pairwise utilities are non-positive. This is also reflected by the facts that all agents have NS deviations with respect to their original utilities, and that on average the valuations of agents changed by 51.25. Overall, the produced outcomes have little connection to the original game and simply exploit that all utilities become negative at some point for resentful agents.

By contrast, for appreciative agents, there is typically one large coalition containing 40 or more agents together with one or two small coalitions. Indeed, this is also reflected by the observations that the average number of coalitions is 2.74 and the average maximum size is 42.57. The typical run of an NS dynamics for appreciative agents here can be described in two phases. First, agents increase their utility for each other by deviating between smaller coalitions (where some agents, which are negatively valued by many others, are not joined, which often leads to them being part of small coalitions in the final outcome). Subsequently, in a second phase, agents already possess a generally high utility level, and thus tend to favor large coalitions (even when having a negative utility for some of the agents in the coalition). Notably, it does not happen that eventually all utilities between pairs of agents are positive, and in this sense, the behavior of appreciative agents is not the contrary of the behavior of resentful agents. In fact, slightly counterintuitively, only 56% of agent pairs have a positive evaluation after convergence of the dynamics. Consequently, agents (from the large coalition) often dislike other coalition members: On average, an agent only values 59% of her coalition members positively. Nevertheless, the relationship of the produced outcome and the agent’s original utilities is still quite low with 18.7 agents for which individual rationality is violated and 22.21 agents having an NS deviation.

For resentful-appreciative agents, the produced outcomes are in some sense between the two extremes for resentful agents and for appreciative agents: Typically, several medium-size coalitions form (the average number of coalitions is 5.26 and the maximum size is on average 18.43). Further, the average change of the utility values is -0.13 and only 51% of agent pairs have a positive evaluation, implying that the utility changes caused by resentment and by appreciation cancel out each other from an aggregated perspective. However, on an individual level, utilities still change quite drastically, as the absolute difference between the initial utilities values and the values at the end of the dynamic is on average 12.15. Notably, the outcome produced by resentful-appreciative agents is also in some sense less “degenerated” as for the the two separately: It consists of medium-size coalitions, utilities are structurally more similar to the initial utilities, and most importantly, the outcome is closer to stability in the original game (with only 2.33 agents for which individual rationality is violated and only 10.15 agents having an NS deviation).

Gaussian Utilities We now turn to Gaussian utilities, where the dynamics converge much quicker than for uniform utilities. Moreover, the outcomes produced by our three dynamics are quite similar, which follows the intuition that the final utility profiles remain quite similar after the execution of “few” steps. In general, NS outcomes produced by our dynamics typically consist of one large coalition containing roughly half of the agents (these are usually the agents with positive ground qualification), while other agents are placed into coalitions of size one or two.

Let us first focus on the second and third part of Table 5 presenting the results of 100 games with $n = 50$ and $n = 25$ agents, respectively. First, the outcomes for appreciative agents typically contain fewer coalitions than for resentful agents. Interestingly, this is not achieved because of the size of the “large” coalition but due to a larger average size of the many small coalitions. For resentful-appreciative agents, the produced outcomes are structurally similar to the ones for resentful agents with a slightly

⁹Note that, in both utility models, this value is on average 0.5 for the initial utilities.

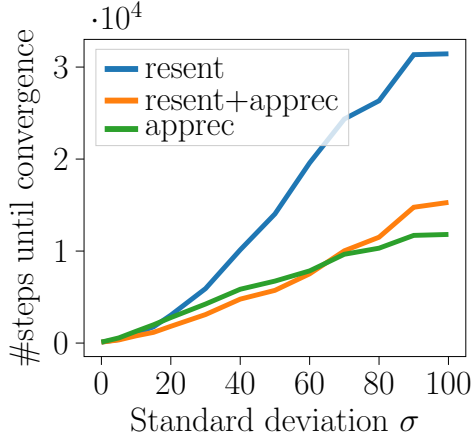


Figure 3: Influence of the standard deviation of Gaussian utilities on the average convergence time. We only take into account dynamics converging within the time limit of 100 000 steps. For resentful agents, the NS dynamics did not converge in 7/23/48/63/75 games for a standard deviation of 60/70/80/90/100, respectively (for appreciative agents and for resentful-appreciative agents the dynamics always converged).

smaller number of coalitions (achieved by an increase of the size of the “large” coalition). Second, for all three types of dynamics the produced outcomes are much closer to being stable in the original game than for uniform utilities (only around 10% of the agents have an NS deviation). In particular, for resentful-appreciative agents, the produced outcomes are closest to stability in the original game. Lastly, considering the average utility of agents in the produced outcome, the three perception types produce quite similar results: Naturally, for appreciative agents, the average utility of agents in the produced outcome is highest. However, in both other dynamics, agents are only slightly less happy.

In the fourth part of the table, we depict statistics concerning the 13 of our 100 games with $n = 25$ agents having Gaussian utilities for which an original NS dynamic converged (in the time limit). In the first line, we show properties of the outcomes produced by the original dynamics. The NS outcomes produced by the original NS dynamics are structurally quite similar to the ones shown in the third part of the table, indicating that the outcomes produced by our three dynamics on games with 25 or 50 agents having Gaussian utilities are quite “natural” and not degenerated. The similarities become even more profound when comparing the outcomes produced by the original dynamics on the selected 13 games to the outcomes produced by our three types of dynamics on the same 13 games: For resentful and resentful-appreciative agents, the produced outcomes are very similar or even identical to the outcome produced by the original NS dynamics in all 13 games (and are in particular always stable in the original games). For appreciative agents, the “large” coalition in the outcome is typically quite similar or identical to the “large” coalition produced by the original dynamics; however, sometimes fewer small coalitions are produced (in particular, some of the produced outcomes are not stable in the original game). Concerning the average utility in the final partition, resentful agents are only marginally less happy than in the original dynamics, indicating that only very few agents end up in a coalition with an agent they left at some point, whereas the average utility is slightly higher for appreciative agents and resentful-appreciative agents.

D.4 Influence of Standard Deviation for Gaussian Utilities

In this section, we focus on Gaussian utilities and analyze the influence of the standard deviation σ on convergence times. Recalling the drastic differences in the behavior of the dynamics between Gaussian utilities (with $\sigma = 10$) and uniform utilities observed in the previous subsections, it is to be expected that the convergence time increases with increasing σ (as we are in some sense getting closer to uniform utilities for larger σ). To analyze this effect in more detail, for $\sigma \in \{0, 5, 10, \dots, 95, 100\}$, we sampled 100 games with $n = 50$ agents having Gaussian utilities with standard deviation σ , and measured the average convergence times of our three types of NS dynamics for a change coefficient $c = 1$. The results are shown in Figure 3, where we only show the average convergence time of all dynamics that converged in the

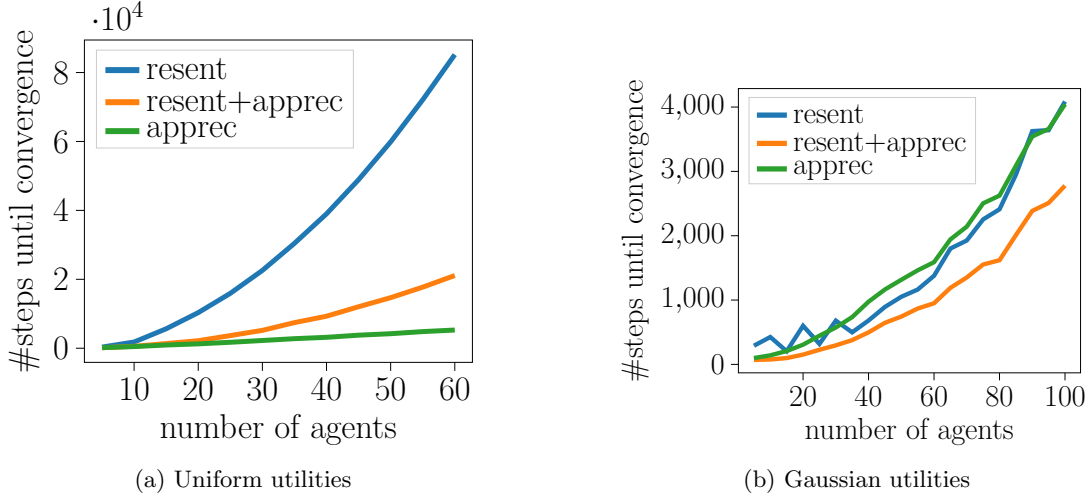


Figure 4: Influence of the number of agents on the average convergence time.

time bound of 100 000 steps. For resentful agents, the NS dynamics did not converge in 7/23/48/63/75 games for a standard deviation of 60/70/80/90/100, respectively (for appreciative agents and resentful-appreciative agents the dynamics always converged). Overall, our hypothesis that increasing the standard deviation leads to an increased convergence time is confirmed clearly. Slightly unexpected, for $\sigma \geq 80$, Gaussian utilities seem to be even “harder” than uniform utilities: For resentful agents, the dynamics did not even converge for roughly half of the games and for appreciative agents the average convergence time is about twice as large. The reason for this is that for large σ utilities often fall out of the range $[-100, 100]$ (from which we sample the agent’s base qualification for Gaussian utilities and the utilities for uniform utilities). In games which are “far away” from stability and for which resentful agents or appreciative agents produce degenerated outcomes (as described in the previous section) an increased range of the utility values clearly contributes to a higher convergence time: For instance, for resentful agents, we have observed that the dynamics typically runs until all utility values are negative, which takes more time if some utility values are initially larger.

D.5 Influence of the Number of Agents

We now briefly analyze the influence of the number of agents on the convergence time of NS dynamics with resentful and/or appreciative agents. For this, for different numbers of agents, we generated 100 games and executed NS dynamics with resentful and/or appreciative agents and a change coefficient of $c = 1$. We show the results for uniform utilities in Figure 4a and the results for Gaussian utilities in Figure 4b. Overall, unsurprisingly, the higher the number of agents, the higher is the convergence time of our dynamics.

For uniform utilities, the relation of the different types of dynamics is independent of the number of agents. For appreciative agents, the average convergence time is the lowest, while convergence is still faster for resentful agents than for resentful-appreciative agents. The increase in the convergence time between $n = 10$ and $n = 60$ is almost linear for all three types. However, the slope substantially depends on the considered type: For resentful agents, comparing $n = 10$ and $n = 60$, the convergence time increases by a factor of 47, for appreciative agents the factor is 11, and for resentful-appreciative agents the factor is 38. Thereby, the difference between the three becomes larger with a growing number of agents. This again highlights the different nature of the three types of NS dynamics for uniform utilities that we have observed above: For resentful agents, for uniform utilities, all utility values need to become negative, which requires substantially, but linearly, more deviations the higher the number of agents. In contrast, for appreciative agents, agents typically only join each other until almost everyone has a positive valuation for a very large coalition. Naturally also this requires more deviations the higher the number of agents, yet with slower scaling.

For Gaussian utilities, we also see an almost linear increase of the average convergence time. Moreover, for most considered agent numbers, dynamics for resentful-appreciative agents converge faster than the

ones for resentful agents which in turn converge faster than the ones for appreciative agents. However, in contrast to uniform utilities, here the difference between the three types of NS dynamics changes only slightly for different numbers of agents. Furthermore, comparing $n = 10$ and $n = 60$, the convergence time of all three types of NS dynamics only increases by a factor of around 12. Lastly, for resentful agents, we can observe a slightly unexpected phenomenon, as for a smaller number of agents the convergence time does not steadily increase, e.g., for $n = 10$ the average converge time is 425 steps and for $n = 15$ it is 204 steps.

E Dynamics with Resentful Deviators

Previously, we have assumed that a deviation of an agent a changes the utility other agents have for a . In contrast, one could also consider what happens if a deviation of a changes a 's utility for other agents. Under deviator-resentment, we assume that an agent decreases her utility for all agents she abandons. Intuitively, deviator-resentment contributes to the existence of stable states, as it implies that an agent a is less likely to enter a coalition containing agents that she has previously abandoned.

Formally, for deviator-resentful agents, if for some $t \geq 1$, π^t evolves from π^{t-1} via a single-agent deviation of agent $k \in N$, then, for $i, j \in N$, $u_i^t(j)$ arises from $u_i^{t-1}(j)$ as

$$u_i^t(j) = \begin{cases} u_i^{t-1}(j) - 1 & i = k, j \in \pi^{t-1}(i) \setminus \{i\}, \\ u_i^{t-1}(j) & \text{else.} \end{cases}$$

Similarly, for deviator-resentful agents, if for $t \geq 1$, π^t evolves from π^{t-1} via a group deviation of $C \subseteq N$, then, for $i, j \in N$, $u_i^t(j)$ arises from $u_i^{t-1}(j)$ as

$$u_i^t(j) = \begin{cases} u_i^{t-1}(j) - 1 & i \in C, j \in \pi^{t-1}(i) \setminus C, \\ u_i^{t-1}(j) & \text{else.} \end{cases}$$

E.1 Contractual Nash Stability and Nash Stability

We start by observing that deviator-resentment does not resolve run-and-chase examples such as the dynamics between two agents a and b where initially a has utility 1 for b and b initially has utility -1 for a . In such examples, a will still always join b in one step, while b leaves a in the next step (thereby decreasing b 's utility for a even further).

Observation E.1. *The NS dynamics may cycle for deviator-resentful agents whose CAFs satisfy friend necessity and single friend desire.*

By Proposition A.2 this implies that the NS dynamics may cycle in ASHG and MFHG.

Moreover, somewhat surprisingly also CNS dynamics may still cycle in ASHG and MFHG with deviator-resentful agents, even if deviations are restricted to be individually rational. Notably, this is in a clear contrast to our previous results for resentful agents where individual rationality was for all types of dynamics sufficient to guarantee convergence.

Proposition E.2. *The individually rational CNS dynamics may cycle in MFHG and ASHG for deviator-resentful agents.*

Proof. Consider a cardinal hedonic game with $N = \{a, b, c\}$ where the initial single-agent utilities are given as $u_a(b) = u_b(c) = u_c(a) = 0$ and $u_b(a) = u_c(b) = u_a(c) = -1$. For $k \geq 0$, let $\pi^{3k} = \{\{a, b\}, \{c\}\}$, $\pi^{3k+1} = \{\{a\}, \{b, c\}\}$, and $\pi^{3k+2} = \{\{a, c\}, \{b\}\}$.

Then, $(\pi^k)_{k \geq 0}$ represents an individually rational CNS dynamics with respect to additively separable and modified fractional utility aggregation. \square

It is possible to modify the previous examples to start a dynamics from the singleton partition.

In contrast to this, as soon as we enforce that agents only deviate to a non-singleton coalition if they strictly prefer it to being in a singleton coalition, convergence can be guaranteed.

Proposition E.3. *The CNS dynamics converges for deviator-resentful agents whose CAFs satisfy friend necessity and single friend desire if agents only deviate to non-singleton coalitions if they strictly prefer them to being in a singleton.*

Proof. Let a hedonic game (N, u^0) with deviator-resentful agents be given where every agent $i \in N$ has a CAF A_i that satisfies friend necessity and single friend desire. Assume for the sake of contradiction that there exists an infinite sequence $(\pi^t)_{t \geq 0}$ of partitions resulting from CNS deviations of deviator-resentful agents where agents only deviate to non-singleton coalitions if they strictly prefer them to being in a singleton. By Lemma 2.1, there exists a time $t_0 \geq 0$ such that every deviation performed after t_0 must occur infinitely often. This implies in particular that there is a step $t'_0 \geq t_0$ such that an agent i has a negative utility for all agents that i leaves at some point after t'_0 .

Now, let us fix a time step $t' \geq t'_0$ where some agent d deviates into a non-singleton coalition. As every deviation is repeated (infinitely often) after $t_0 \leq t'$, there has to be some time step where d deviates again. In particular, let t'' be the smallest t with $t > t'$ where d deviates again after time t' . For some time step t , let $C_t := \pi^t(d)$ be the coalition of d after step t . We now examine the coalitions $C_{t'}, \dots, C_{t''-1}$, i.e., the coalitions d is part of between her two deviations.

As we assume that an agent only deviates into a non-singleton coalition if she strictly prefers it to being in a singleton coalition, d prefers $C_{t'}$ to being in a singleton. Hence, by friend necessity, there is an agent $a \in C_{t'}$ for which d has positive utility at time $t' - 1$. By single friend desire, it follows that there is an agent $a \in C_{t''-1}$ for which d has a positive utility. Indeed, otherwise, there is some $t \in [t', t'' - 2]$ such that there is a friend of d in C_t but not in C_{t+1} . However, this cannot happen because, due to single friend desire, d strictly prefers C_t to C_{t+1} , implying that d would have vetoed the deviation taking place in time step $t + 1$. This implies that there is at least one agent for which d has positive utility in $C_{t''-1}$. However, d leaves this agent in time step t'' . This contradicts our initial assumption that each agent has a negative utility for all agents they leave in some step after $t'_0 \leq t' \leq t''$. \square

As the CAFs AS_i and MF_i satisfy friend necessity and single friend desire (Proposition A.2), Proposition E.3 implies that, for every ASHG and MFHG with deviator-resentful agents, some execution of the CNS dynamics converges.

E.2 Core Stability and Individual Stability

We now turn to the consent-based stability concepts individual stability and (strict) core stability. Here, the influence of deviator-resentment is more profound.

Individual Rationality

We start by considering individually rational SCS dynamics. Here, like for resentful agents, convergence is guaranteed for a wide class of CAFs.

Proposition E.4. *The individually rational SCS, CS, and IS dynamics converge for deviator-resentful agents whose CAFs satisfy enemy domination.*

Proof. We show the statement for SCS deviations, which implies convergence of CS and IS dynamics. Let a hedonic game (N, u^0) with deviator-resentful agents be given where every agent $i \in N$ has a CAF A_i that satisfies enemy domination. Assume for the sake of contradiction that there exists an infinite sequence $(\pi^t)_{t \geq 0}$ of partitions resulting from SCS deviations of deviator-resentful agents. By Lemma 2.1, there exists a time $t_0 \geq 0$ such that every deviation performed after t_0 must occur infinitely often.

We first show that if two agent i and j perform a core deviation together at some point after t_0 , then i cannot leave j nor can j leave i after t_0 . We prove that i cannot leave j (the other case is symmetric). By enemy domination, there exists a constant $c(u_i^{t_0}, j)$ such that for all utility vectors $u'_i \in \mathbb{Q}^n$ with $u'_i(k) \leq u_i^{t_0}(k)$ for all $k \in N$ and $u'_i(j) \leq c(u_i^{t_0}, j)$, it holds for every $C \in \mathcal{N}_i$ with $j \in C$ that $A_i(C, u'_i) < A_i(\{i\}, u'_i)$.

As each deviation after t_0 occurs infinitely often, the fact that i leaves j after t_0 infinitely often implies that there exists a time $t_1 \geq t_0$ such that agent i leaves agent j for $u_i^{t_0}(j) - c(u_i^{t_0}, j)$ times between time t_0 and time t_1 . Since agent i is deviator-resentful, this implies that $u_i^{t_1}(j) \leq c(u_i^{t_0}, j)$. Additionally, deviator-resentful agents can only decrease utilities for other agents. Therefore, it holds

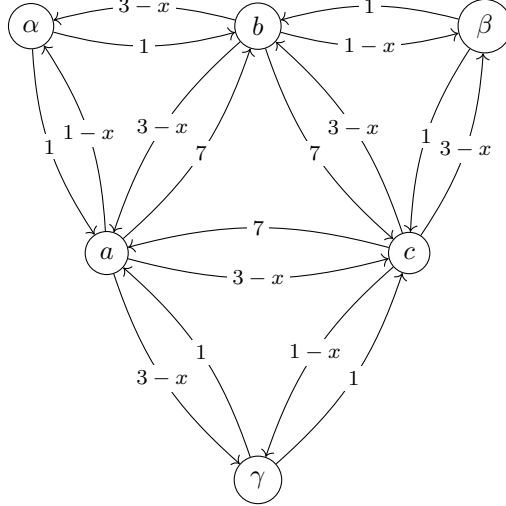


Figure 5: Example from Proposition E.5. Utilities after x cycles of deviations. All missing edges indicate that the utilities of the agents for each other are zero.

for all $t \geq t_0$ and $k \in N$ that $u_i^t(k) \leq u_i^{t_0}(k)$. Consequently, from the definition of $c(u_i^{t_0}, j)$ it follows that agent i cannot deviate to form a coalition together with j after t_1 , because this deviation would not be individually rational. However, such a deviation takes place because we have assumed that i and j deviate together at some point after t_0 and such a deviation is performed infinitely often. This establishes that i and j cannot leave each other after t_0 .

Finally, consider some time $t \geq t_0$ in which a coalition C_t performs a group deviation. Then, from our above observation we get that no agent that is part of C_t can ever leave another agent that is part of C_t after this, implying that $\pi^{t'}(j) \subseteq C_t$ for all $j \in C_t$ and $t' \geq t$. This implies that the size of C_t needs to monotonically increase over time. Notably, this holds for all coalitions that ever performed a deviation after time t_0 . Yet the overall number of agents is bounded, resulting in a contradiction. \square

As AS_i and MF_i satisfy enemy domination (Proposition A.2), Proposition E.4 implies that the individually rational SCS, IS, and CS dynamics converge in ASHG and MFHG for deviator-resentful agents.

Absence of Individual Rationality

If we drop the requirement that deviations are individual rational, the picture changes. For MFHG, IS, CS, and SCS dynamics may cycle.

Proposition E.5. *The IS, CS, and SCS dynamics may cycle in MFHG for deviator-resentful agents, even when starting from the singleton partition.*

Proof. Consider the MFHG with agent set $N = \{a, b, c, d, e, f\}$ and utilities as depicted in Figure 5, where the number of an arc from agents i to j describes the utility that i has for j . The initial utilities result from setting x to 0 in Figure 5. Consider the following infinite sequence of partitions. For $n \geq 0$,

- $\pi^{3n} = \{\{\alpha, a, b\}, \{\beta\}, \{\gamma, c\}\}$,
- $\pi^{3n+1} = \{\{\alpha, a\}, \{\beta, b, c\}, \{\gamma\}\}$, and
- $\pi^{3n+2} = \{\{\alpha\}, \{\beta, b\}, \{\gamma, a, c\}\}$.

Note that the utilities of agents after x executions of this cycle of three deviations are shown in Figure 5. Observing that for $k \geq 1$, π^{k-1} leads to π^k by means of a (S)CS deviation completes the counterexample. Note that we can also modify the dynamics to start in the singleton partition, by inserting the two partitions $\{\{\alpha\}, \{\beta\}, \{\gamma\}, \{a\}, \{b\}, \{c\}\}$ and $\{\{\alpha, a, b\}, \{\beta\}, \{\gamma\}, \{c\}\}$ in the beginning of the dynamics.

Using the same initial utilities, for IS deviations we have the following cycling sequence of partitions. For $n \geq 0$,

- $\pi^{6n} = \{\{\alpha, a, b\}, \{\beta\}, \{\gamma, c\}\}$,
- $\pi^{6n+1} = \{\{\alpha, a, b\}, \{\beta, c\}, \{\gamma\}\}$,
- $\pi^{6n+2} = \{\{\alpha, a\}, \{\beta, b, c\}, \{\gamma\}\}$,
- $\pi^{6n+3} = \{\{\alpha\}, \{\beta, b, c\}, \{\gamma, a\}\}$,
- $\pi^{6n+4} = \{\{\alpha\}, \{\beta, b\}, \{\gamma, a, c\}\}$, and
- $\pi^{6n+5} = \{\{\alpha, b\}, \{\beta\}, \{\gamma, a, c\}\}$.

Again the utilities of agents after x executions of this cycle of six partitions are shown in Figure 5. We again observe that for $k \geq 1$, π^{k-1} leads to π^k by means of a IS deviation, which completes the counterexample. Again we can also modify the example to start in a singleton partition by adding the three partitions $\{\{\alpha\}, \{\beta\}, \{\gamma\}, \{a\}, \{b\}, \{c\}\}$, $\{\{\alpha, a\}, \{\beta\}, \{\gamma\}, \{b\}, \{c\}\}$, $\{\{\alpha, a, b\}, \{\beta\}, \{\gamma\}, \{c\}\}$ in the beginning of the dynamics. \square

By contrast, IS dynamics are still always guaranteed to converge in ASHG. Note that this is the only contrast between ASHGs and MFHGs proven in this paper (in all other cases we either have the same result for both classes, or a result for one and an open question for the other).

Theorem E.6. *In ASHGs, the IS dynamics converges for deviator-resentful agents.*

Proof. Let an ASHG (N, u^0) be given. For the sake of contradiction assume that there exists an infinite sequence $(\pi^t)_{t \geq 0}$ of partitions resulting from IS deviations of deviator-resentful agents. For each $t \geq 0$, let d^t be the agent deviating in step t from $\pi^{t-1}(d^t)$ to $\pi^t(d^t)$. By Lemma 2.1, there exists $t_0 \geq 0$ such that every deviation performed after t_0 is performed infinitely often.

Note that for each $t \geq t_0$ it holds that d^t cannot be left by an agent $j \in \pi^t(d^t)$ after t_0 , as otherwise j will leave d^t infinitely often and thus, as utilities are only decreasing, at some point will no longer approve the join of d^t , as she derives negative utility from d^t . We refer to this as the first observation.

Fix some $t \geq t_0$ and let $d := d_t$ and $C := \pi^t(d^t)$ (note that d joins $C \setminus \{d^t\}$ at step t). As our second observation we now show that there is some $t' > t$ with $d^{t'} = d$ and $\pi^{t'-1}(d^{t'}) = C$.

From our choice of t_0 it follows that there is some $t' \geq t$ where d performs a deviation for the next time. Assume now, for the sake of contradiction that there is some t'' with $t' > t'' > t$ where $\pi^{t''}(d) \neq \pi^{t''-1}(d)$ and select the smallest such t'' . If $\pi^{t''}(d)$ is changed because an agent j left it, then, by our choice of t'' , it follows that $j \in C$, which contradicts our first observation. If $\pi^{t''}(d)$ is changed because an agent j joined it, then it needs to hold that j leaves the coalition of d again before t' : Otherwise, d leaves j which again contradicts the first observation as d needs to approve the join of j . Thus, the second observation follows.

However, the second observation implies that for d it holds that $u_d^t(\pi^t(d)) = u_d^{t'-1}(\pi^{t'-1}(d))$. In the next step, d performs an IS deviation and thus increases her utility, i.e., $u_d^{t'}(\pi^{t'}(d)) > u_d^{t'-1}(\pi^{t'-1}(d))$. Afterwards, the second observation can be applied again until d performs the next deviation. Thus, d 's utility is strictly increasing, however, as utilities are initially bounded and resentment can only cause their decay, this leads to a contradiction. Hence, IS dynamics have to converge. \square

Our results give broad insight into dynamics for deviator-resentful agents. Still, we were not able to settle whether SCS or CS dynamics are guaranteed to converge in ASHG for deviator-resentful agents without the individual rationality assumption, leaving this as an open question.

F Fastest Convergence

As discussed in Section 6, we now investigate the problem of deciding whether there exists an execution of some dynamics for a given hedonic game with resentful or appreciative agents, converging *within a certain time*. For $\alpha \in \{NS, IS, CNS, CS\}$, we define the following decision problem:

Given: An additively separable hedonic game with resentful agents and some integer k .

Question: Is there an execution $(\pi^t)_{t \geq 0}$ of the α dynamics that starts from the singleton partition π^0 and converges in at most k steps?

Defining APPRECIATIVE-ASHG- α -SEQUENCE analogously for appreciative agents, we can show that these problems are NP-hard for NS, IS, CNS, and CS.

Theorem F.1. *For $\alpha \in \{NS, IS, CNS, CS\}$, RESENTFUL- and APPRECIATIVE-ASHG- α -SEQUENCE are NP-hard.*

Before turning to the proof of Theorem F.1, we want to make a few remarks. First, note that all our constructions in the following proofs also work for normal agents with constant utilities over time. Hence, our constructions also imply the hardness of the respective decision problem for (time-independent) ASHG.

Second, we want to discuss membership in NP. Note that the sequence of partitions leading to a stable partition is a certificate for a Yes-instance of size $\mathcal{O}(nk)$. Hence, if k is given in unary encoding, then this is a certificate of polynomial size. Moreover, for the reduced instance in our proofs, it holds that $k \in \mathcal{O}(n)$, and therefore the hardness holds on a class of instances for which membership in NP is satisfied. Still, it is not clear whether RESENTFUL- and APPRECIATIVE-ASHG- α -SEQUENCE are in NP for a general k in binary encoding.

We are ready to continue with the proof of Theorem F.1 and split it into three parts (Lemmas F.2 to F.4).

Lemma F.2. *For $\alpha \in \{NS, IS\}$, RESENTFUL- and APPRECIATIVE-ASHG- α -SEQUENCE are NP-hard.*

Proof. We first show the statement for IS and resentful agents. We reduce from RESTRICTED EXACT COVER BY 3-SETS (RX3C) where we are given a finite universe $U = \{x_1, \dots, x_{3t}\}$ and a family $\mathcal{S} = \{S_1, \dots, S_{3t}\}$ of 3-subsets of U where each element from U appears in exactly three sets from \mathcal{S} . The question is whether there is a family $\mathcal{S}' \subseteq \mathcal{S}$ which is an exact cover of U .

Construction. Given an instance (U, \mathcal{S}) of RX3C, we set $k = 10t$ and create the following additively separable hedonic game. We add one *element agent* x for each element $x \in U$ and one *set agent* S for each set $S \in \mathcal{S}$. Moreover, we add $2t$ *filling agents* $F = \{f_1, \dots, f_{2t}\}$. Lastly, we add a penalizing gadget consisting of $10t + 1$ *penalizing agents* $\{p\} \cup Q$ with $Q = \{p_{i,j} \mid i \in [5t], j \in [2]\}$. The resulting set of agents is $N = U \cup \mathcal{S} \cup F \cup \{p\} \cup Q$. The initial utilities of the players are illustrated in Figure 6:

- Each element agent has utility zero for each other element agent and utility 1 for each set agent.
- Each filling agent has utility 1 for each set agent.
- Each set agent $S \in \mathcal{S}$ has utility $20t$ for the three element agents $x \in S$, utility $60t$ for each filling agent, utility $60t$ for p , and utility zero for each penalizing agent in Q .
- p has utility $10t$ for each set agent and utility zero for each penalizing agent in Q .
- Each penalizing agent $p_{i,j} \in Q$ has utility $30t$ for each set agent, utility $30t$ for p , utility $40t$ for her corresponding penalizing agent $p_{i,k}$ with $k \in [2], k \neq j$, and utility zero for all remaining penalizing agents in Q .
- All not explicitly mentioned utilities are $-1000t$.

We now show that there is an exact cover of U if and only if there is a sequence starting with the singleton partition and reaching an IS partition after at most k IS deviations, where agents are resentful.

(\Rightarrow) Given an exact cover $\mathcal{S}' \subseteq \mathcal{S}$, we let the agents deviate as follows. For each $S \in \mathcal{S} \setminus \mathcal{S}'$, one filling agent deviates to S (there are $2t$ of these sets and $2t$ filling agents). For each $S \in \mathcal{S}'$, the three element agents x with $x \in S$ deviate to S . As \mathcal{S}' is an exact cover, each filling and each element agent deviate exactly once, leading to $5t$ deviations so far. Furthermore, for $i \in [5t]$, $p_{i,2}$ deviates to $p_{i,1}$. Thus, we have $10t$ deviations in total. If we order these deviations arbitrarily, then all deviations are NS deviations, as each agent deviates from the singleton coalition to a coalition for which she has positive

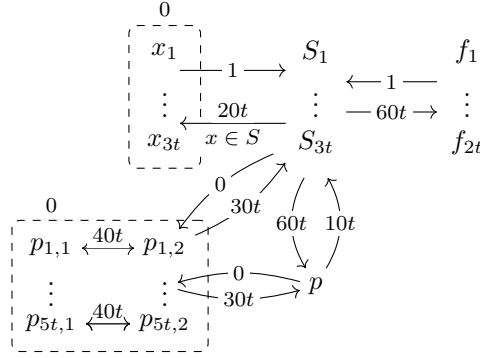


Figure 6: Scheme of the additively separable hedonic games in Lemma F.2. All omitted utilities are $-1000t$.

utility. Furthermore, these NS deviations are IS deviations as the agents in the joined coalitions have non-negative utilities for the deviators. It remains to show that the resulting partition is NS (and thus also IS): Observe that there is at most one set agent per coalition. As element and filling agents only derive a positive utility from set agents and are in a coalition with a set agent, no element or filling agent has an NS deviation. Agent p only has a positive utility for the set agents but all set agents are accompanied by either element or filling agents for which p has a large negative utility. Hence, p has no NS deviation. Agents $p_{i,1}$ and $p_{i,2}$ with $i \in [5t]$ do not want to deviate as they have utility $40t$ for coalition $\{p_{i,1}, p_{i,2}\}$ and value no other available coalition with more than $30t$. Lastly, each set agent $S \in \mathcal{S}$ is in a coalition from which she derives utility $60t$. All other coalitions are either the coalition just containing p (for which S has utility $60t$ as well), coalitions containing other penalizing agents (for which S has utility zero), or contain a different set agent (for which S has large negative utility). Thus, S has no NS deviation either and the created partition is NS.

(\Leftarrow) Assume that we have a sequence of $\ell \leq k = 10t$ IS deviations leading from the singleton partition π^0 to an IS partition π^ℓ . As a first observation, note that in this sequence no two agents who have utility $-1000t$ for each other can ever be in the same coalition because they would never join each other. Second, as we start from the singleton partition, the resentment an agent has build up for the other agents after ℓ deviations sums up to at most $\frac{\ell}{2} \leq 5t$.

Next, we show that, in π^ℓ , no set agent is in a joint coalition with p . For the sake of a contradiction, assume that $p \in \pi^\ell(S)$ for some $S \in \mathcal{S}$. By the first observation, this means that there is no other set agent, no element agent, and no filling agent in $\pi^\ell(S)$. Hence $\pi^\ell(S) \subseteq \{p, S\} \cup Q$. Let $p_{i,j} \in Q$ be a penalizing agent who has not performed any deviation, yet. This agent has to exist because, otherwise, we would have needed at least $10t + 1$ deviations to reach π^ℓ . Then, $p_{i,j}$ has a utility of at least $u_{p_{i,j}}^0(S) + u_{p_{i,j}}^0(p) - 5t = 55t$ for $\pi^\ell(S)$ and thus has a possible NS deviation to $\pi^\ell(S)$. (As $p_{i,j}$ has not deviated yet, she can have a utility of at most $40t$ for $\pi^\ell(p_{i,j})$ if $p_{i,k}$ with $k \in [2], k \neq j$ has joined her.) This NS deviation is also an IS deviation as, initially, no agent from $\pi^\ell(S)$ has a negative utility for $p_{i,j}$. This is a contradiction to π^ℓ being IS. So, $\pi^\ell(p) \subseteq \{p\} \cup Q$.

As π^ℓ is IS, no set agent S has any possible IS deviation from $\pi^\ell(S)$ to $\pi^\ell(p)$. As p and all other penalizing agents would always allow S to join her (note that they cannot have decreased their utility by more than $5t$ for S until step ℓ), S has no NS deviation to join $\pi^\ell(p)$. Since (a) the resentment that S has build up for any agent sums up to at most $-5t$, (b) S initially has utilities zero for all penalizing agents $p_{i,j} \in \pi^\ell(p)$, (c) S initially has utility $60t$ for p , and (d) S does not want to move to p , it follows that $u_S^\ell(\pi^\ell(S)) \geq 60t - 5t = 55t$.

Therefore, $\pi^\ell(S)$ contains at least one filling agent or three element agents x with $x \in S$. Since this holds for all set agents and no two set agents are in one coalition, the existence of an exact cover (namely $S' = \{S \in \mathcal{S} \mid S \text{ is together with three element agents}\}$) is implied.

By the same proof as above, there is an exact cover of U exactly if there is a sequence of at most k NS deviations leading to an NS partition for resentful agents.

For appreciative agents, we can use the same construction. The proof of correctness is very similar. \square

Lemma F.3. RESENTFUL- and APPRECIATIVE-ASHG-CNS-SEQUENCE are NP-hard.

agent do not have any CNS deviation because they have a set agent in their coalition who would veto their deviation. Third, p_1 would veto a deviation of p_0 , so p_0 has no CNS deviation. Fourth, p_1 does not want to deviate to another coalition as she has utility 1 for $\pi(p_1) = \{p_0, p_1\}$ and every other coalition in π contains an agent for which she has utility $-1000t$. Last, p_2, p_3 , and p_4 do not want to deviate as they only have a positive utility (of $200t$) for agent p_1 but p_1 is together with p_0 for which they have utility $-1000t$.

(\Leftarrow) Assume that we have a sequence of $\ell \leq k = 5t + 1$ CNS deviations leading from the singleton partition π^0 to a CNS partition π^ℓ . In this sequence, no two agents who mutually have utility $-1000t$ for each other can ever be in the same coalition because they would never join each other. Furthermore, as we start from the singleton partition, the resentment an agent has accumulated for the other agents after ℓ deviations sums up to at most $\frac{\ell}{2} \leq 3t$.

We will now show that no set agent is together with p_0 in π^ℓ . For the sake of a contradiction, assume that $S \in \pi^\ell(p_0)$ for some $S \in \mathcal{S}$. By the first observation, this implies that $\pi^\ell(p_0) \subseteq \{S, p_0, p_1\}$. It further holds that $p_1 \notin \pi^\ell(p_0)$ because, otherwise, p_1 would have a CNS deviation to a singleton coalition. Hence, $\pi^\ell(p_0) = \{S, p_0\}$. Next, consider coalition $\pi^\ell(p_1)$. By the first observation, we have $\pi^\ell(p_1) \subseteq \mathcal{S} \cup Q$. Furthermore, there is no set agent in $\pi^\ell(p_1)$: Indeed, assume for contradiction that there is some set agent S' in $\pi^\ell(p_1)$. Then, by the first observation, neither p_2, p_3 , nor p_4 are in $\pi^\ell(p_1)$. Furthermore, we already now that p_0 is not in $\pi^\ell(p_1)$. It follows that $\pi^\ell(p_1) = \{S', p_1\}$ which means that p_1 has a CNS deviation to a singleton coalition, a contradiction. Hence, $\pi^\ell(p_1) \subseteq \{p_1, \dots, p_4\}$. For $i \in \{2, 3, 4\}$, it also holds by the first observation that $\pi^\ell(p_i) \subseteq \{p_1, \dots, p_4\}$. We will now show that $\{p_1, \dots, p_4\}$ can not be divided into a CNS partition [cf. 27, Example 2]. Note that the arguments also hold when lowering any utilities among this agents by at most $3t$ and therefore π^ℓ is no CNS partition, which is a contradiction.

First, none of p_2, p_3 , and p_4 can be in the same coalition because the agent with a initial utility of $-400t$ for another agent in her coalition would have a CNS deviation to a singleton. Hence, the remaining possible partitions of $\{p_1, \dots, p_4\}$ are $\pi' = \{\{p_1\}, \{p_2\}, \{p_3\}, \{p_4\}\}$, $\pi'' = \{\{p_1, p_2\}, \{p_3\}, \{p_4\}\}$, $\pi''' = \{\{p_2\}, \{p_1, p_3\}, \{p_4\}\}$, and $\pi'''' = \{\{p_2\}, \{p_3\}, \{p_1, p_4\}\}$. But all these partitions are not CNS: π' is not stable because p_2 has a CNS deviation to $\{p_1\}$; π'' is not stable because p_4 has a CNS deviation to $\{p_1, p_2\}$; π''' is not stable because p_2 has a CNS deviation to $\{p_1, p_3\}$; π'''' is not stable because p_3 has a CNS deviation to $\{p_1, p_4\}$. Thus, we have shown that π^ℓ is not CNS. This is a contradiction and therefore the initial assumption was wrong, i.e., no set agent is together with p_0 in π^ℓ .

Since π^ℓ is CNS, no set agent has a CNS deviation to p_0 . Note that p_0 is the only agent who has positive utilities for the set agents. So, for any set agent $S \in \mathcal{S}$, there is no agent in $\pi^\ell(S)$ who would veto a deviation of S . Therefore, no set agent has an NS deviation to p_0 . Furthermore, we know that $\pi^\ell(p_0) \subseteq \{p_0, p_1\}$ because of the first observation and because no set agent is with p_0 . So, for every set agent $S \in \mathcal{S}$, we have $u_S^\ell(\pi^\ell) \geq u_S^\ell(\pi^\ell(p_0) \cup \{S\}) \geq 60t - 3t = 57t$ (where $60t$ is S 's initial utility for p_0 and $-3t$ is a bound for the resentment that S might have build for p_0 or p_1). Therefore, $\pi^\ell(S)$ contains at least one filling agent or three element agents x with $x \in S$. Since no two set agents are in the same coalition, this implies the existence of an exact cover for U .

For appreciative agents, we can use almost the same construction. We just set p_0 's initial utility for p_1 to -1 while all other utilities stay the same. Then, the proof of correctness is very similar to the proof for resentful agents. \square

Lemma F.4. RESENTFUL- and APPRECIATIVE-ASHG-CS-SEQUENCE are NP-hard.

Proof. We first show the statement for resentful agents and reduce from RX3C.

Construction. Let an instance of RX3C (U, \mathcal{S}) with a universe $U = \{x_1, \dots, x_{3t}\}$ and a family $\mathcal{S} = \{S_1, \dots, S_{3t}\}$ of 3-subsets of U be given. We set $k = 3t + 2$ and create the following additively separable hedonic game. The set of agents is $N = U \cup \mathcal{S} \cup F \cup Q$ with element agents U , set agents \mathcal{S} , filling agents $F = \{f_1, \dots, f_{2t}\}$, and penalizing agents $Q = \{p_0, \dots, p_6\}$. The initial utilities of the agents are illustrated in Figure 8:

- Each set agent S has utility $60t$ for each filling agent, utility $20t$ for element agents x with $x \in S$, and utility $60t$ for the penalizing agent p_0 .
- Each element agent has utility 1 for each set agent and utility 0 for all other element agents.
- Each filling agent has utility 1 for each set agent.

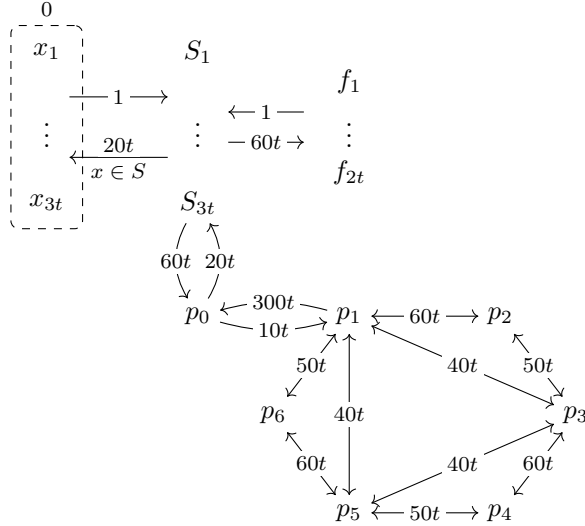


Figure 8: Additively separable hedonic games in Lemma F.4. All omitted utilities are $-1000t$.

- p_0 has utility $20t$ for all set agents and utility $10t$ for p_1 .
- Agents p_1 , p_3 , and p_5 have utility $40t$ for each other.
- For $i \in \{1, 3, 5\}$, agent p_i has utility $60t$ for agent p_{i+1} and vice versa.
- For $i \in \{2, 4, 6\}$, agent p_i has utility $50t$ for agent p_{i+1} and vice versa (where $p_7 = p_1$).
- All not explicitly mentioned utilities are $-1000t$.

Note that the penalizing gadget consisting of agents p_1, \dots, p_6 is a variant of an ASHG by Aziz et al. [2, Example 1] for which no CS partition exists (for normal ASHG).

(\Rightarrow) Given an exact cover $\mathcal{S}' \subseteq \mathcal{S}$, consider the following sequence of group deviations. For each $S \in \mathcal{S}'$, S and the three $x \in S$ deviate together (t deviations). For each $S \in \mathcal{S} \setminus \mathcal{S}'$, S deviates with one filling agent that is still in a singleton ($2t$ deviations). Now, p_0 deviates with p_1 . Lastly, p_3 , p_4 , and p_5 deviate together. Hence, we have $3t + 2$ deviations in total.

All these deviations are CS deviations as all these deviations are performed from singletons to coalitions with positive utilities for all deviators.

Next, we show that the resulting partition π is CS. First note that no two agents who have utility $-1000t$ for each other want to deviate together. Each element agent x only has a positive utility of 1 for all set agents and is in a coalition with one set agent. As no two set agents will ever deviate together, there is no coalition that x could deviate with to improve her utility. The same holds for all filling agents. Each set agent S has utility $60t$ for $\pi(S)$. As no element or filling agent wants to deviate, the only remaining agent for which S has positive utility is p_0 . Yet, S only has utility $60t$ for p_0 which is no improvement compared to $\pi(S)$. So no set agent has a CS deviation. Therefore, also p_0 has no CS deviation as she would only like to deviate with some set agents. p_1 has utility $300t$ for $\pi(p_1) = \{p_0, p_1\}$ and would thus not deviate without p_0 . Similarly, the remaining agents p_2, \dots, p_6 have no CS deviation without p_1 . Thus, the whole partition π is CS.

(\Leftarrow) Assume that we have a sequence of $\ell \leq k = 3t + 2$ CS deviations leading from the singleton partition π^0 to a CS partition π^ℓ . In this sequence, no agent is ever in a coalition with an agent for which she has utility $-1000t$ because she would never deviate with this agent. Therefore, the largest coalitions in this sequence have size at most four (containing one set agent and three element agents). It follows that the sum of the resentment that a single agent accumulates for the other agents during $\ell \leq 3t + 2$ deviations is bounded by $\frac{\ell}{2} \cdot 3 \leq \frac{3t+2}{2} \cdot 3 = \frac{9}{2}t + 3$ because she can be left by at most three agents every second deviation.

We will now show that no set agent is in a joint coalition with p_0 in π^ℓ . For the sake of a contradiction, assume that $S \in \pi^\ell(p_0)$ for some $S \in \mathcal{S}$. By the first observation, this implies that $\pi^\ell(p_0) = \{S, p_0\}$.

This means that p_1, \dots, p_6 have formed coalitions among each other. Yet, as argued by Aziz et al. [2, Example 1], there is no CS partition for these six agents. Note that resentment, which influences each player by a total utility change of at most $\frac{9}{2}t + 3$, does not facilitate stability because the differences between the utilities in this group are at least $10t$. It follows that π^ℓ is not CS which is a contradiction. So no set agent is together with p_0 in π^ℓ .

Next, since π^ℓ is CS, no coalition $\{S, p_0\}$ with $S \in \mathcal{S}$ has a CS deviation. Since p_0 is together with no set agent in π^ℓ , p_0 can have a utility of at most $10t$ for π^ℓ and wants to deviate to $\{S, p_0\}$. So, S does not want to deviate to $\{S, p_0\}$. Therefore, $u_S^\ell(\pi^\ell) \geq u_S^\ell(\{S, p_0\}) \geq 60t - (\frac{9}{2}t + 3) = \frac{111}{2}t - 3$. Hence, $\pi^\ell(S)$ contains at least one filling agent or three element agents x with $x \in S$. Since no two set agents are in the same coalition, this implies the existence of an exact cover for U .

For appreciative agents, we can use the same construction and the proof is very similar. □

Niclas Boehmer
Technische Universität Berlin
Berlin, Germany
Email: niclas.boehmer@tu-berlin.de

Martin Bullinger
Technische Universität München
München, Germany
Email: martin.bullinger@tum.de

Anna Maria Kerkmann
Heinrich-Heine-Universität Düsseldorf
Düsseldorf, Germany
Email: anna.kerkmann@hhu.de