Causes of Stability in Dynamic Coalition Formation

Niclas Boehmer[1], **Martin Bullinger**[2], **Anna Maria Kerkmann**[3]

[1] Algorithmics and Computational Complexity, Technische Universität Berlin
[2] School of Computation, Information and Technology, Technische Universität München
[3] Institut für Informatik, Heinrich-Heine-Universität Düsseldorf
niclas.boehmer@tu-berlin.de, martin.bullinger@in.tum.de, anna.kerkmann@hhu.de

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 (*resent*) or increase their utility for a joining agent (*appreciation*). We show that in contrast to classical dynamics, for resentful or appreciative agents, dynamics are guaranteed to converge under mild conditions for various stability concepts. Thereby, we establish that both resent and appreciation are strong stability-driving forces.

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. 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. 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. The goal here is to reach stable coalition structures through decentralized individual decisions (Brandt, Bullinger, and Wilczynski 2021; Brandt, Bullinger, and Tappe 2022). 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, either Alice or Bob are 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.

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 aban-

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

	SCS	CS	IS	CNS	NS
ASHG					
resent	✓ (3.1)	✓ (3.1)	✓ (3.1)	?	?
resent+IR	✓ (3.1)	✓ (3.1)	✓ (3.1)	✓ (3.3)	✓ (3.3)
appreciation	✗ (4.4)	✗ (4.4)	?	✓ (4.5)	?
MFHG					
resent	?	?	?	?	✗ (3.4)
resent+IR	✓ (3.2)	✓ (3.2)	✓ (3.2)	✓ (3.3)	✓ (3.3)
appreciation	✗ (4.4)	✗ (4.4)	?	?	✗ (4.6)

Table 1: Overview of results. "✓" means that the corresponding dynamics is guaranteed to converge; "✗" means that we have an example for an infinite sequence. See Section 2 for definitions. For each result, we include the number of the respective statement.

doning her. A deviator abandoning a resentful agent again and again eventually looses all of her attraction to the resentful agent. 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 Corollary 4.3 four of our open questions concerning both resentful and appreciative agents are in some sense 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 (Drèze and Greenberg 1980), but their constant and broad consideration only started with key publications by Banerjee, Konishi, and Sönmez (2001), Cechlárová and Romero-Medina (2001), and Bogomolnaia and Jackson (2002). An overview of hedonic games is provided in the survey by Aziz and Savani (2016). The search for suitable representations of reasonable classes of hedonic games has led to various proposals (see, e.g., Cechlárová and Romero-Medina 2001; Bogomolnaia and Jackson 2002; Ballester 2004; Elkind and Wooldridge 2009; Olsen 2012; Aziz et al. 2019).

Various stability concepts and their computational boundaries have been previously studied. We focus on results concerning ASHGs (Bogomolnaia and Jackson 2002) and MFHGs (Olsen 2012). Sung and Dimitrov (2010) show prototype NP-hardness reductions for single-agent stability concepts in ASHGs, paving the way for many similar results for single-agent and group stability (see, e.g., Aziz, Brandt, and Seedig 2013; Brandt, Bullinger, and Tappe 2022; Bullinger 2022). Gairing and Savani (2019) consider ASHGs under symmetric utilities and show PLS-completeness of computing stable states, while Woeginger (2013) and Peters (2017) show Σ_2^{P}-completeness of the (strict) core in ASHGs. Peters and Elkind (2015) provide a meta view on computational hardness. For MFHGs, 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 (Monaco, Moscardelli, and Velaj 2018). 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 ASHGs and MFHGs (Aziz, Brandt, and Seedig 2013; Elkind, Fanelli, and Flammini 2020; Bullinger 2020; Brandt and Bullinger 2022). These papers provide more evidence that MFHGs seem to be less complex than ASHGs.

The dynamical, distributed approach to coalition formation received increased attention very recently (Hoefer, Vaz, and Wagner 2018; Bilò et al. 2018; Carosi, Monaco, and Moscardelli 2019; Brandt, Bullinger, and Wilczynski 2021; Fanelli, Monaco, and Moscardelli 2021; Brandt, Bullinger, and Tappe 2022; Bilò, Monaco, and Moscardelli 2022). There, Bilò et al. (2018); Brandt, Bullinger, and Wilczynski (2021); Brandt, Bullinger, and Tappe (2022) consider stability based on single-agent deviations, whereas Carosi, Monaco, and Moscardelli (2019); Fanelli, Monaco, and Moscardelli (2021) consider group stability.

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 $[i] = \{1, \ldots, i\}$.

2.1 Cardinal Hedonic Games

Let $N = [n]$ be a finite set of *agents*. A *coalition* is any 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 $u_i : N \to \mathbb{Q}$ 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 $A_i : \mathcal{N}_i \times \mathbb{Q}^n \to \mathbb{Q}$ 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 (ASHGs) (Bogomolnaia and Jackson 2002) 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 (MFHGs) (Olsen 2012) 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.

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 $j \in \pi(i) \setminus \{i\}$.

$$\begin{tikzcd} & \text{NS} \arrow[dl] \arrow[dr] & & \text{SCS} \arrow[dl] \arrow[dr] & \\ \text{CNS} & & \text{IS} & & \text{CS} \end{tikzcd}$$

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

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 Aziz and Savani 2016).

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 *resent* 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 *resent*, agents decrease their utilities for all deviators that leave them (by one), while under *appreciation*, agents increase their utilities for all deviators that join them (by one).[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
$$u_i^t(j) = \begin{cases} u_i^{t-1}(j) + 1 & i \neq k, j = k, j \in \pi^t(i), \\ u_i^{t-1}(j) & \text{else.} \end{cases}$$

[1]Note that our choice of decreasing, resp., increasing the utilities by one is somewhat arbitrary, as our theoretical results hold for any fixed increase or decrease of utilities. However, note that 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 our full version (Boehmer, Bullinger, and Kerkmann 2022). 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.

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 (S)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 resent 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 resent 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.*

[2] All missing proofs can be found in our full version (Boehmer, Bullinger, and Kerkmann 2022).

Proof sketch. We claim that in an infinite sequence of SCS deviations, there are infinitely many deviations due to which the previously non-negative utility of some agent for another is decreased, directly leading to a contradiction. Assume for the sake of contradiction that, from some time step t_0 on, no deviation decreases the non-negative utility of some agent for another. So, if agent a leaves agent b and thus b's utility for a decreases, then b already had a negative utility for a. Consequently, by aversion to enemies, each deviation from t_0 onward is a Pareto improvement. Using enemy monotonicity, it can be shown that every sequence of Pareto improvements has bounded length, and we thus reach a contradiction. □

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 ASHGs 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, resent is clearly irrelevant and there exist anonymous hedonic games where IS dynamics cycle (Brandt, Bullinger, and Wilczynski 2021). 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 MFHGs 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 MFHGs (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, resent 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 sketch. Assume that there is a game with an infinite sequence of individually rational NS deviations. Then, by Lemma 2.1 there is a time step t_0 after which each deviation is performed infinitely often. This implies that if agent a leaves agent b at some point after t_0, then b has an arbitrarily negative utility for a at some point after t_0. Applying enemy domination, this implies that there is a point in time after which agent b can never perform an individually rational NS deviation to a coalition containing a. Thus, after t_0, no agent can ever be abandoned by an agent that she joined after t_0. Using a potential function argument, one can show that this implies that the dynamics always converges. □

As AS and MF satisfy enemy domination (Proposition 2.3), Theorem 3.3 implies that the individually rational NS dynamics converges in ASHGs and MFHGs for resentful agents. However, we do not know under which conditions resent 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 counterintuitive, 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 MFHGs 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 ASHGs). 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 establishing a close connection between cycling dynamics for resentful and appreciative agents in ASHGs, highlighting a close connection between the two studied models. Subsequently, we analyze CS and (C)NS dynamics.

4.1 From Resent to Appreciation

We describe how we can transform certain types of infinite sequences of deviations for resentful agents to sequences for appreciative agents and vice versa. We focus on ASHGs, yet believe that similar statements can hold for other classes of hedonic games. We start with Nash stability.

Theorem 4.1. *The following statements are equivalent:*

1. *There exists an ASHG admitting an infinite and periodic sequence of NS deviations for resentful agents.*
2. *There exists an ASHG admitting an infinite and periodic sequence of NS deviations for appreciative agents.*

The idea to prove Theorem 4.1 is to reverse a periodic fragment of an infinite sequence and to appropriately adjust the initial utilities. This essentially reverses the roles of resent and appreciation, as the agents that an agent a leaves in the sequence for resentful agents correspond to the agents a joins in the sequence for appreciative agents.

By Proposition 3.5, Theorem 4.1 can be extended to also include infinite and periodic sequences of CNS deviations for resentful agents. In fact, the equivalence can be extended even further, as we show that in ASHGs with appreciative agents, 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 ASHGs 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).

To sum up, combining Theorem 4.1 and Propositions 4.2 and 3.5, we get the following equivalences.

Corollary 4.3. *The following statements are equivalent:*

1. *There exists an ASHG admitting an infinite and periodic sequence of CNS deviations for resentful agents.*
2. *There exists an ASHG admitting an infinite and periodic sequence of NS deviations for resentful agents.*
3. *There exists an ASHG admitting an infinite and periodic sequence of NS deviations for appreciative agents.*
4. *There exists an ASHG admitting an infinite and periodic sequence of IS deviations for appreciative agents.*

4.2 Convergence for Appreciative Agents

We now give an overview under which circumstances appreciation is (not) sufficient to guarantee convergence in MFHGs and ASHGs. In contrast to resent, appreciation is not sufficient to guarantee convergence of CS dynamics.[3]

Theorem 4.4. *The individually rational CS dynamics may cycle in ASHGs and MFHGs for appreciative agents.*

However, in the games considered in Theorem 4.4, 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 ASHGs and MFHGs for appreciative agents.

Lastly, we consider IS and (C)NS dynamics. In ASHGs for appreciative agents, it remains open whether IS and NS dynamics may cycle. In fact, we have seen in Proposition 4.2 that these two questions are equivalent and in Corollary 4.3 that they are very closely related to our open questions concerning resentful agents. On the other hand, for CNS, appreciation is sufficient to guarantee convergence.

Theorem 4.5. *The CNS dynamics converges in ASHGs for appreciative agents.*

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

[3]For ASHGs, the next statement can be extended to an ASHG where initial valuations are symmetric by slightly modifying the game presented by Aziz, Brandt, and Seedig (2013, Figure 2).

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

Table 2: Some of our experimental results on ASHGs. 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.

Theorem 4.6. *The individually rational NS dynamics may cycle in MFHGs for appreciative agents.*

It remains open whether IS dynamics may cycle in MFHGs for appreciative agents. Note that the arguments from Proposition 4.2 for showing the "equivalence" for IS and NS dynamics under appreciation do not work for MFHGs.

5 Simulations

We analyze by means of simulations how resent and appreciation influence NS dynamics in ASHGs 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 our full version for details (Boehmer, Bullinger, and Kerkmann 2022). We focus on ASHGs 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 resent 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 resent and appreciation are helpful to establish convergence, appreciation is more powerful than resent here, and in fact adding resent 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 resent or appreciation. On average, significantly fewer agents have an NS deviation with respect to their initial utilities.

Gaussian Utilities For Gaussian utilities, the original NS dynamics without resent 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 ASHGs 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 Corollary 4.3). For future work, complementing our simulations, it would be interesting to theoretically analyze the effects of combining resent 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 $C \setminus C'$. An intuitive reason why deviator-resent 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-resent 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-resent guarantees convergence in ASHGs and MFHGs 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 ASHGs and MFHGs with infinite sequences of individually rational CNS deviations. Deviator-resent is nevertheless a powerful force for stability, as CS dynamics with individual rational deviations in MFHGs and ASHGs and IS dynamics in ASHGs always converge. However, deviator-resent is not sufficient to guarantee convergence of IS and general CS dynamics in MFHGs, yielding a different behavior of ASHGs and MFHGs for IS dynamics. Notably, we did not prove any such contrasts in our analysis of resent and appreciation. Overall, our results indicate that deviator-resent has clear ramifications on convergence guarantees, yet the general picture seems to be slightly more nuanced than for resent 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 resent 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 (Brandt, Bullinger, and Tappe 2022). In our full version (Boehmer, Bullinger, and Kerkmann 2022), we present three reductions showing that for ASHGs 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).

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

Aziz, H.; Brandl, F.; Brandt, F.; Harrenstein, P.; Olsen, M.; and Peters, D. 2019. Fractional Hedonic Games. *ACM Transactions on Economics and Computation*, 7(2): 1–29.

Aziz, H.; Brandt, F.; and Seedig, H. G. 2013. Computing Desirable Partitions in Additively Separable Hedonic Games. *Artificial Intelligence*, 195: 316–334.

Aziz, H.; and Savani, R. 2016. Hedonic Games. In Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; and Procaccia, A. D., eds., *Handbook of Computational Social Choice*, chapter 15. Cambridge University Press.

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

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

Bilò, V.; Fanelli, A.; Flammini, M.; Monaco, G.; and Moscardelli, L. 2018. Nash Stable Outcomes in Fractional Hedonic Games: Existence, Efficiency and Computation. *Journal of Artificial Intelligence Research*, 62: 315–371.

Bilò, V.; Monaco, G.; and Moscardelli, L. 2022. Hedonic Games with Fixed-Size Coalitions. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*.

Boehmer, N.; Bullinger, M.; and Kerkmann, A. M. 2022. Causes of Stability in Dynamic Coalition Formation. Technical report, https://arxiv.org/abs/2211.17169.

Bogomolnaia, A.; and Jackson, M. O. 2002. The Stability of Hedonic Coalition Structures. *Games and Economic Behavior*, 38(2): 201–230.

Brandt, F.; and Bullinger, M. 2022. Finding and Recognizing Popular Coalition Structures. *Journal of Artificial Intelligence Research*, 74: 569–626.

Brandt, F.; Bullinger, M.; and Tappe, L. 2022. Single-Agent Dynamics in Additively Separable Hedonic Games. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*, 4867–4874.

Brandt, F.; Bullinger, M.; and Wilczynski, A. 2021. Reaching Individually Stable Coalition Structures in Hedonic Games. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, 5211–5218.

Bullinger, M. 2020. Pareto-Optimality in Cardinal Hedonic Games. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 213–221.

Bullinger, M. 2022. Boundaries to Single-Agent Stability in Additively Separable Hedonic Games. In *Proceedings of the 47th International Symposium on Mathematical Foundations of Computer Science (MFCS)*. Forthcoming.

Carosi, R.; Monaco, G.; and Moscardelli, L. 2019. Local Core Stability in Simple Symmetric Fractional Hedonic Games. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 574–582.

Cechlárová, K.; and Romero-Medina, A. 2001. Stability in Coalition Formation games. *International Journal of Game Theory*, 29: 487–494.

Drèze, J. H.; and Greenberg, J. 1980. Hedonic Coalitions: Optimality and Stability. *Econometrica*, 48(4): 987–1003.

Elkind, E.; Fanelli, A.; and Flammini, M. 2020. Price of Pareto Optimality in hedonic games. *Artificial Intelligence*, 288: 103357.

Elkind, E.; and Wooldridge, M. 2009. Hedonic coalition nets. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 417–424.

Fanelli, A.; Monaco, G.; and Moscardelli, L. 2021. Relaxed core stability in fractional hedonic games. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, 182–188.

Gairing, M.; and Savani, R. 2019. Computing Stable Outcomes in Symmetric Additively Separable Hedonic Games. *Mathematics of Operations Research*, 44(3): 1101–1121.

Hoefer, M.; Vaz, D.; and Wagner, L. 2018. Dynamics in matching and coalition formation games with structural constraints. *Artificial Intelligence*, 262: 222–247.

Monaco, G.; Moscardelli, L.; and Velaj, Y. 2018. Stable Outcomes in Modified Fractional Hedonic Games. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 937–945.

Olsen, M. 2012. On defining and computing communities. In *Proceedings of the 18th Computing: Australasian Theory Symposium (CATS)*, volume 128 of *Conferences in Research and Practice in Information Technology (CRPIT)*, 97–102.

Peters, D. 2017. Precise Complexity of the Core in Dichotomous and Additive Hedonic Games. In *Proceedings of the 5th International Conference on Algorithmic Decision Theory (ADT)*, 214–227.

Peters, D.; and Elkind, E. 2015. Simple causes of complexity in hedonic games. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, 617–623.

Sung, S. C.; and Dimitrov, D. 2010. Computational Complexity in Additive Hedonic Games. *European Journal of Operational Research*, 203(3): 635–639.

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