

# Single-Agent Dynamics in Additively Separable Hedonic Games

Felix Brandt, Martin Bullinger, and Leo Tappe

Institut für Informatik, Technische Universität München  
brandtf@in.tum.de, bullinge@in.tum.de, leo.tappe@tum.de

## Abstract

The formation of stable coalitions is a central concern in multi-agent systems. A considerable stream of research defines stability via the absence of beneficial deviations by single agents. Such deviations require an agent to improve her utility by joining another coalition while possibly imposing further restrictions on the consent of the agents in the welcoming as well as the abandoned coalition. While most of the literature focuses on unanimous consent, we also study consent decided by majority vote, and introduce two new stability notions that can be seen as local variants of popularity. We investigate these notions in additively separable hedonic games by pinpointing boundaries to computational complexity depending on the type of consent and restrictions on the utility functions. The latter restrictions shed new light on well-studied classes of games based on the appreciation of friends or the aversion to enemies. Many of our positive results follow from the *Deviation Lemma*, a general combinatorial observation, which can be leveraged to prove the convergence of simple and natural single-agent dynamics under fairly general conditions.

## Introduction

Coalition formation is a central concern in multi-agent systems and considers the question of grouping a set of agents, e.g., humans or machines, into coalitions such as teams, clubs, or societies. A prominent framework for studying coalition formation is that of hedonic games, where agents' utilities are solely based on the coalition they are part of, and which thus disregards inter-coalitional relationships (Drèze and Greenberg 1980). Hedonic games have been successfully used to model various scenarios evolving from operations research or the mathematical social sciences such as research team formation (Alcalde and Revilla 2004), task allocation (Saad et al. 2011), or community detection (Newman 2004; Aziz et al. 2019). Identifying desirable coalition structures is often based on the prospect of coalitions to stay together. To this end, various notions of stability have been introduced and studied. A coalition structure (henceforth partition) is stable when no individual or group of agents benefits by joining another coalition or by forming a new coalition.

In this paper, we focus on deviations by single agents. The simplest example is a Nash deviation where some agent unilaterally decides to leave her current coalition in order to join

another coalition. While such a deviation clearly captures the incentive of single agents to perform deviations, it completely ignores the other agents' opinions about the deviation. To overcome this shortcoming, various restrictions of Nash deviations have been proposed. This has motivated stability notions such as individual stability or contractual Nash stability, which consider the unanimous consent of some or all of the coalitions directly affected by the deviation. While unanimous consent is in fact used in the formation process of international bodies like the EU or the NATO, it might be impractical and even undesirable in small- or medium-scale coalition formation scenarios. As a compromise, we also study intermediate notions of stability based on majority votes among the involved coalitions. This setting has received little attention so far (Gairing and Savani 2019), and we will also define some new majority-based stability notions.

The study of hedonic games was initiated by Drèze and Greenberg (1980), and later popularized by Banerjee, Konishi, and Sönmez (2001), Cechlárová and Romero-Medina (2001), and Bogomolnaia and Jackson (2002). Since then, a large body of research has been devoted to defining suitable game representations and solution concepts. An overview of many important aspects is provided in the survey by Aziz and Savani (2016). One prominent, natural, and arguably simple type of hedonic games is given by additively separable hedonic games (Bogomolnaia and Jackson 2002). In these games, agents entertain cardinal utilities for other agents and the utility for a coalition is defined by taking the sum of the individual utility values. This game representation allows, for instance, the modeling of settings where agents have friends and enemies, and their goal is to simultaneously maximize the number of friends and minimize the number of enemies, while one of these two goals can have higher priority than the other one (Dimitrov et al. 2006). Our work provides exact boundaries for the computational tractability of stability concepts based on single-agent deviations in additively separable hedonic games, showing a clear cut between Nash stability and stability notions under consent. We give simple and precise conditions for restricted classes of utility functions that pinpoint the boundaries of computational tractability. This includes well-studied classes of games where agents only distinguish between friends and enemies.

A more recent line of research on stability notions focuses on the dynamical aspects leading to the formation of

stable outcomes (e.g., Bilò et al. 2018; Hoefer, Vaz, and Wagner 2018; Carosi, Monaco, and Moscardelli 2019; Brandt, Bullinger, and Wilczynski 2021). This yields an important distributed perspective on the coalition formation process. The value of some positive computational results in the context of hedonic games is diminished by the fact that they implicitly assume that a central authority has the means to collect all individual preferences, compute a stable partition, and enforce this partition on the agents. In contrast, simple dynamics based on single-agent deviations provide a much more plausible explanation for the formation of stable partitions. A versatile tool to prove the convergence of dynamics are potential functions, which guide the dynamics towards stable states (e.g., Bogomolnaia and Jackson 2002; Suksompong 2015). We extend the applicability of this approach by considering *non-monotonic* potential functions, i.e., potential functions that might decrease in some rounds of the dynamic process. This is possible because the total number of rounds can be bounded by observing the potential function from a global perspective using a new general combinatorial insight that we call the Deviation Lemma. The Deviation Lemma is not restricted to additively separable utilities or the specific type of single-agent deviations. For instance, the combinatorial relationship of the lemma also arises naturally in the analysis of deviation dynamics in classes of games beyond the scope of this paper, such as anonymous hedonic games (Bogomolnaia and Jackson 2002). In fact, the lemma holds for every sequence of partitions such that each partition evolves from its predecessor by having one element move to another partition class. It establishes a relationship between the development of the sizes of coalitions involved in deviations to information solely based on the starting partition and the terminal partition of the sequence.

For the special case of symmetric utility functions, additively separable hedonic games are well understood: the standard notion of utilitarian social welfare represents an increasing potential function for the dynamics induced by Nash stability (Bogomolnaia and Jackson 2002), but finding stable states (even under unanimous consent of the welcoming coalition) leads to PLS-complete problems (Gairing and Savani 2019). As we will see, this implies worst-case exponential running time of the dynamics. By contrast, our results hold for restricted sets of non-symmetric utility functions and our computational boundaries lie between polynomial-time computability and NP-completeness. In fact, whenever we identify a potential function guaranteeing the existence of stable outcomes, we are also able to prove that, from any starting partition, the corresponding simple dynamics of single-agent deviations converges to a stable partition in a polynomial number of rounds.

## Preliminaries and Model

In this section we introduce hedonic games and our stability concepts. We use the notation  $[k] = \{1, \dots, k\}$  for any positive integer  $k$ .

### Hedonic Games

Throughout the paper, we consider settings with a set  $N = [n]$  of  $n$  agents. The goal of coalition formation is to find a

partition of the agents into different disjoint coalitions according to their preferences. Hence, we search a partition  $\pi : N \rightarrow 2^N$  such that  $i \in \pi(i)$  for every agent  $i \in N$  and either  $\pi(i) = \pi(j)$  or  $\pi(i) \cap \pi(j) = \emptyset$  holds for every pair of agents  $i$  and  $j$ , where  $\pi(i)$  denotes the coalition to which agent  $i$  belongs. We refer to the partition  $\pi$  given by  $\pi(i) = \{i\}$  for every agent  $i \in N$  as the *singleton partition*, and to  $\pi = \{N\}$  as the *grand coalition*.

Let  $\mathcal{N}_i$  denote all possible coalitions containing agent  $i$ , i.e.,  $\mathcal{N}_i = \{C \subseteq N : i \in C\}$ . A *hedonic game* is defined by a tuple  $(N, \succsim)$ , where  $N$  is an agent set and  $\succsim = (\succsim_i)_{i \in N}$  is a tuple of weak orders  $\succsim_i$  over  $\mathcal{N}_i$  which represent the preferences of the respective agent  $i$ . Hence, agents express preferences only over the coalitions which they are part of without considering externalities. The generality of the definition of hedonic games gives rise to many interesting subclasses of games that have been proposed in the literature. Many of these classes rely on cardinal utility functions  $v_i : N \rightarrow \mathbb{R}$  for every agent  $i$ , which are aggregated in various ways (Aziz et al. 2019; Bogomolnaia and Jackson 2002; Olsen 2012). One particularly natural and prominent such model considers aggregation by taking the sum of individual utilities. Formally, following Bogomolnaia and Jackson (2002), an *additively separable hedonic game* (ASHG)  $(N, v)$  consists of an agent set  $N$  and a tuple  $v = (v_i)_{i \in N}$  of utility functions  $v_i : N \rightarrow \mathbb{R}$  such that  $\pi(i) \succsim_i \pi'(i)$  iff  $\sum_{j \in \pi(i)} v_i(j) \geq \sum_{j \in \pi'(i)} v_i(j)$ . Clearly, ASHG are a subclass of hedonic games, and we can assume without loss of generality that  $v_i(i) = 0$  (or set the utility of an agent for herself to an arbitrary constant). ASHG have a natural representation by a complete directed graph  $G = (N, E)$  with weight  $v_i(j)$  on arc  $(i, j)$ . An ASHG is called *symmetric* if  $v_i(j) = v_j(i)$  for every pair of agents  $i$  and  $j$ , and it can then be represented by a complete undirected graph with weight  $v_i(j)$  on edge  $\{i, j\}$ . There are various classes of ASHG with certain restrictions for the utility functions that allow a natural interpretation in terms of friends and enemies. An agent  $j$  is called *friend* (respectively, *enemy*) of agent  $i$  if  $v_i(j) > 0$  (respectively,  $v_i(j) < 0$ ). An ASHG is called *friends-and-enemies game* (FEG) if  $v_i(j) \in \{-1, 1\}$  for every pair of agents  $i$  and  $j$ . Further, following Dimitrov et al. (2006), an ASHG is called an *appreciation of friends game* (AFG) (respectively, an *aversion to enemies game* (AEG)) if  $v_i(j) \in \{-1, n\}$  (respectively,  $v_i(j) \in \{-n, 1\}$ ). In all of these games, agents pursue the objective to maximize their number of friends while minimizing their number of enemies. In the case of an FEG, these two goals have equal priority, while there is a strict priority for one of the goals in AFGs and AEGs.

### Stability Based on Single-Agent Deviations

We want to study stability under single agents' incentives to perform deviations. A *single-agent deviation* performed by agent  $i$  transforms a partition  $\pi$  into a partition  $\pi'$  where  $\pi(i) \neq \pi'(i)$  and, for all agents  $j \neq i$ ,

$$\pi'(j) = \begin{cases} \pi(j) \setminus \{i\} & \text{if } j \in \pi(i), \\ \pi(j) \cup \{i\} & \text{if } j \in \pi'(i), \\ \pi(j) & \text{otherwise.} \end{cases}$$

We write  $\pi \xrightarrow{i} \pi'$  to denote a single-agent deviation performed by agent  $i$  transforming partition  $\pi$  to partition  $\pi'$ .

We consider the case of myopic rational agents who only engage in a deviation if it immediately makes them better off. Formally, a *Nash deviation* is a single-agent deviation performed by agent  $i$  making agent  $i$  better off, i.e.,  $\pi'(i) \succ_i \pi(i)$ . Any partition in which no Nash deviation is possible is called *Nash stable* (NS).

This concept of stability is very strong and comes with the drawback that only the preferences of the deviating agent are considered. Therefore, various refinements have been proposed which additionally require the consent of the abandoned and the welcoming coalition. For a compact representation, we introduce them via the notion of favor sets.

Let  $C \subseteq N$  be a coalition and  $i \in N$  be an agent. The *favor-in set* of  $C$  with respect to  $i$  is the set of agents in  $C$  (excluding  $i$ ) that strictly favor having  $i$  inside of  $C$  than outside, i.e.,  $F_{\text{in}}(C, i) = \{j \in C \setminus \{i\} : C \cup \{i\} \succ_j C \setminus \{i\}\}$ . Similarly, the *favor-out set* of  $C$  with respect to  $i$  is the set of agents in  $C$  (excluding  $i$ ) that strictly favor having  $i$  outside of  $C$  than inside, i.e.,  $F_{\text{out}}(C, i) = \{j \in C \setminus \{i\} : C \setminus \{i\} \succ_j C \cup \{i\}\}$ .

Following Bogomolnaia and Jackson (2002) and Dimitrov and Sung (2007), an *individual deviation* (respectively, *contractual deviation*) is a Nash deviation  $\pi \xrightarrow{i} \pi'$  such that  $F_{\text{out}}(\pi'(i), i) = \emptyset$  (respectively,  $F_{\text{in}}(\pi(i), i) = \emptyset$ ). A single-agent deviation that is both an individual and a contractual deviation is called *contractual individual deviation*. All of these deviation concepts give rise to a respective stability concept. A partition is called *individually stable* (IS), *contractually Nash stable* (CNS), or *contractually individually stable* (CIS) if it allows for no individual, contractual, or contractual individual deviations, respectively.

While these stability concepts include agents affected by the deviation, they require unanimous consent, which might be unnecessarily strong in some settings. Based on this observation, we define several hybrid stability concepts where the possibility of a deviation by some agent is decided via majority votes of the involved agents.

A Nash deviation  $\pi \xrightarrow{i} \pi'$  is called *majority-in deviation* (respectively, *majority-out deviation*) if  $|F_{\text{in}}(\pi'(i), i)| \geq |F_{\text{out}}(\pi'(i), i)|$  (respectively,  $|F_{\text{out}}(\pi(i), i)| \geq |F_{\text{in}}(\pi(i), i)|$ ). A single-agent deviation that is both a majority-in deviation and a majority-out deviation is called *separate-majorities deviation*. As before, a partition is called *majority-in stable* (MIS), *majority-out stable* (MOS), or *separate-majorities stable* (SMS) if it allows for no majority-in, majority-out, or separate-majorities deviations, respectively. The concepts MIS and MOS are a special case of voting-based stability notions by Gairing and Savani (2019) for a threshold of  $1/2$ .

Finally, it is possible to relax SMS by performing one joint vote instead of two separate votes. A Nash deviation  $\pi \xrightarrow{i} \pi'$  is called a *joint-majority deviation* if  $|F_{\text{out}}(\pi(i), i)| + |F_{\text{in}}(\pi'(i), i)| \geq |F_{\text{in}}(\pi(i), i)| + |F_{\text{out}}(\pi'(i), i)|$ . A partition is then called *joint-majority stable* (JMS) if it allows for no joint-majority deviations. JMS is particularly interesting as it is a natural local version of popularity (Pop), an axiom recently studied in the context of hedonic games (Gärdenfors

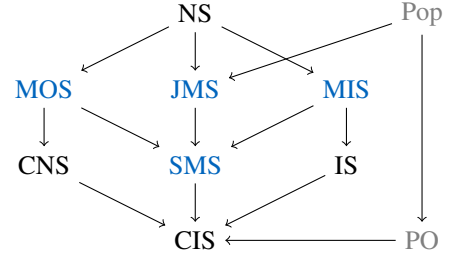


Figure 1: Logical relationships between stability notions and other solutions concepts. An arrow from concept  $\alpha$  to concept  $\beta$  indicates that if a partition satisfies  $\alpha$ , it also satisfies  $\beta$ . Majority-based stability notions are highlighted in blue, other single-agent based stability notions in black.

1975; Cseh 2017; Brandt and Bullinger 2020).<sup>1</sup>

Also note that while CIS is a refinement of Pareto optimality (PO), there is no logical relationship between other (majority-based) stability concepts and PO. In particular, we denote the stability concepts based on single-agent deviations by  $\mathcal{S}$ , i.e.,  $\mathcal{S} = \{\text{NS, IS, CNS, CIS, MIS, MOS, SMS, JMS}\}$ . A taxonomy of our related solution concepts is provided in Figure 1. For a more concise notation, we refer to deviations with respect to stability concept  $\alpha \in \mathcal{S}$  as  $\alpha$ -deviations, e.g., IS-deviations for  $\alpha = \text{IS}$ .

All these stability concepts naturally induce dynamics where we choose some starting partition and obtain a successor partition by having some agent perform a deviation from the current partition. More precisely, given a stability concept  $\alpha \in \mathcal{S}$ , an execution of  $\alpha$ -dynamics is an infinite or finite sequence  $(\pi_j)_{j \geq 0}$  of partitions and a corresponding sequence  $(i_j)_{j \geq 1}$  of (deviating) agents such that  $\pi_{j-1} \xrightarrow{i_j} \pi_j$  is an  $\alpha$ -deviation for every  $j$ . The partition  $\pi_0$  is then called the *starting partition*. Given a hedonic game  $G$ , and a stability concept  $\alpha \in \mathcal{S}$ , we say that the dynamics *converges for starting partition*  $\pi_0$  if every execution of the  $\alpha$ -dynamics on  $G$  with starting partition  $\pi_0$  is finite. Additionally, the  $\alpha$ -dynamics *converges* on  $G$  if it converges for every starting partition.

Proving convergence of dynamics is a very natural way to prove the existence of stable states and underlines the robustness of the stability concept. It complements a static solution concept with a decentralized process to reach a solution.

## Results

In this section, we present our results.

### Computational Boundaries for Nash Stability

First, we consider the notion of Nash stability. In the absence of negative utility values, the partition consisting solely of the grand coalition is Nash stable. Conversely, in the absence of

<sup>1</sup>Informally speaking, a partition is popular if there is no other partition preferred by a majority of the agents. JMS partitions can only be challenged by partitions evolving through Nash deviations.

positive utility values, the singleton partition is Nash stable. It is therefore necessary for an ASHG to have both positive and negative utility values in order to admit a non-trivial Nash stable partition (see also Gairing and Savani 2019).

Sung and Dimitrov (2010) showed that deciding whether an ASHG has an NS partition is NP-hard by a reduction from EXACT3COVER. This reduction produces an ASHG with four distinct positive utility values and one negative utility value. We improve upon this result by showing that a reduction is possible with only one positive and one negative utility value. Moreover, it is possible for *any* choice of these two utility values, as long as the absolute value of the negative utility value is at least as large as the positive utility value. We state the theorem in a general way allowing the positive and negative utility value to be dependent on the number of agents of the particular instance. In this way, we simultaneously cover several important cases. For instance, the hardness holds for fixed constant positive and negative utility values as in FEGs, or for AFGs and AEGs. Note that for all of our stability notions, a stable partition is a polynomial-time verifiable certificate: one can simply check whether any agent can perform a deviation, and if no one can, the partition is stable. Therefore, we omit the proof of membership in NP in all of our reductions. The proof of the next and some subsequent results are omitted due to space restrictions.

**Theorem 1.** *Let  $f^+ : \mathbb{N} \rightarrow \mathbb{Q}_{>0}$  and  $f^- : \mathbb{N} \rightarrow \mathbb{Q}_{<0}$  be two polynomial-time computable functions satisfying  $|f^-(m)| \geq f^+(m)$  for all  $m \in \mathbb{N}$ . Then, the problem of deciding whether an ASHG with utility values restricted to  $\{f^-(n), f^+(n)\}$  has an NS partition is NP-complete.*

Theorem 1 requires the negative utility value to be at least as large in absolute value as the positive utility value. While we leave open the computational complexity for completely arbitrary pairs of negative and positive values, we can show that the problem is also hard when the positive utility value is significantly larger than the absolute value of the negative utility value. The reduction is a variant of the reduction in Theorem 1.

**Theorem 2.** *Deciding whether an AFG has an NS partition is NP-complete.*

## Deviation Lemma and Applications

By contrast, restricting the utility values to one positive and one negative value leads to positive results for other notions of stability. These results can be shown in a unified manner using a potential function argument that crucially hinges on the following general observation.

**Lemma 1** (Deviation Lemma). *Let  $\pi_0 \xrightarrow{i_1} \pi_1 \xrightarrow{i_2} \dots \xrightarrow{i_k} \pi_k$  be a sequence of  $k$  single-agent deviations. Then, the following identity holds:*

$$\sum_{j \in [k]} |\pi_j(i_j)| - |\pi_{j-1}(i_j)| = \frac{1}{2} \sum_{i \in N} |\pi_k(i)| - |\pi_0(i)|. \quad (1)$$

*Proof.* Let  $\pi_0 \xrightarrow{i_1} \pi_1 \xrightarrow{i_2} \dots \xrightarrow{i_k} \pi_k$  be a sequence of  $k$  single-agent deviations and fix some  $j \in [k]$ . Then, the fol-

lowing facts hold:

$$\begin{aligned} |\pi_j(i_j)| &= \left( \sum_{i \in \pi_j(i_j) \setminus \{i_j\}} |\pi_j(i)| - |\pi_{j-1}(i)| \right) + 1, \\ |\pi_{j-1}(i_j)| &= \left( \sum_{i \in \pi_{j-1}(i_j) \setminus \{i_j\}} |\pi_{j-1}(i)| - |\pi_j(i)| \right) + 1, \\ \pi_j(i) &= \pi_{j-1}(i) \quad \forall i \in N \setminus (\pi_j(i_j) \cup \pi_{j-1}(i_j)). \end{aligned}$$

Combining these facts allows us to express the difference of the deviator's coalition sizes as follows:

$$\begin{aligned} |\pi_j(i_j)| - |\pi_{j-1}(i_j)| &= \left( \sum_{i \in \pi_j(i_j) \setminus \{i_j\}} |\pi_j(i)| - |\pi_{j-1}(i)| \right) \\ &\quad - \left( \sum_{i \in \pi_{j-1}(i_j) \setminus \{i_j\}} |\pi_{j-1}(i)| - |\pi_j(i)| \right) \\ &\quad + \sum_{i \in N \setminus (\pi_j(i_j) \cup \pi_{j-1}(i_j))} |\pi_j(i)| - |\pi_{j-1}(i)| \\ &= \sum_{i \in N \setminus \{i_j\}} |\pi_j(i)| - |\pi_{j-1}(i)|. \end{aligned}$$

Adding  $|\pi_j(i_j)| - |\pi_{j-1}(i_j)|$  to both sides yields

$$2(|\pi_j(i_j)| - |\pi_{j-1}(i_j)|) = \sum_{i \in N} |\pi_j(i)| - |\pi_{j-1}(i)|.$$

Summing these terms for all  $j \in [k]$ , interchanging summation order, and telescoping gives

$$\begin{aligned} \sum_{j \in [k]} 2(|\pi_j(i_j)| - |\pi_{j-1}(i_j)|) &= \sum_{j \in [k]} \sum_{i \in N} |\pi_j(i)| - |\pi_{j-1}(i)| \\ 2 \sum_{j \in [k]} |\pi_j(i_j)| - |\pi_{j-1}(i_j)| &= \sum_{i \in N} \sum_{j \in [k]} |\pi_j(i)| - |\pi_{j-1}(i)| \\ 2 \sum_{j \in [k]} |\pi_j(i_j)| - |\pi_{j-1}(i_j)| &= \sum_{i \in N} |\pi_k(i)| - |\pi_0(i)|. \end{aligned}$$

Dividing both sides by 2 completes the proof.  $\square$

The Deviation Lemma is especially useful as the right-hand side of Equation (1) does not depend on  $k$ , and we can therefore also find bounds for its left-hand side solely depending on the number of players  $n$ .

**Lemma 2.** *Consider a sequence of  $k$  successive single-agent deviations*

$$\pi_0 \xrightarrow{i_1} \pi_1 \xrightarrow{i_2} \dots \xrightarrow{i_k} \pi_k.$$

*Then, the following bounds hold:*

$$-\frac{n(n-1)}{2} \leq \sum_{j \in [k]} |\pi_j(i_j)| - |\pi_{j-1}(i_j)| \leq \frac{n(n-1)}{2}.$$

*Proof.* Observe that for all  $i \in N$  and all partitions  $\pi$ , we have

$$1 \leq |\pi(i)| \leq n.$$

Thus, we can find the bounds

$$-n(n-1) \leq \sum_{i \in N} |\pi_k(i)| - |\pi_0(i)| \leq n(n-1).$$

Applying Lemma 1 yields the desired result.  $\square$

We demonstrate the power of the Deviation Lemma by proving convergence of the dynamics for a variety of deviation types and classes of ASHG.

**Theorem 3.** *The dynamics of IS-deviations always converges in ASHG with at most one nonnegative utility value.*

*Proof.* Let  $(N, v)$  be an ASHG such that the  $v_i$  take on at most one nonnegative value. If there are no nonnegative valuations, all IS-deviations are singleton formations, so after at most  $n$  deviations, we reach a stable partition. Now, suppose that there is exactly one nonnegative utility value  $x \geq 0$ . If there are no negative valuations, then in case  $x = 0$  we terminate immediately, and in case  $x > 0$  the grand coalition will form after at most  $n^2$  deviations. The latter holds because every deviation increases the number of pairs of agents which are part of the same coalition. Thus, we will now assume that in addition to the single nonnegative utility value  $x$ , there is at least one negative utility value, and we denote the largest absolute value of a negative utility value by  $y$ . Further, define  $\Delta = \min\{v_i(C) - v_i(C') : i \in N, C, C' \in \mathcal{N}_i, v_i(C) > v_i(C')\}$ . Intuitively,  $\Delta > 0$  is the minimum improvement any agent is guaranteed to have when making a NS-deviation. Further, consider the potential function  $\Phi$  defined by the social welfare of a partition as  $\Phi(\pi) = \sum_{i \in N} v_i(\pi)$ .

Let us investigate how this potential changes for a single IS-deviation  $\pi \xrightarrow{i} \pi'$ .

$$\begin{aligned} \Phi(\pi') - \Phi(\pi) &= \underbrace{v_i(\pi') - v_i(\pi)}_{\text{deviator}} \\ &+ \underbrace{\sum_{j \in \pi'(i) \setminus \{i\}} v_j(\pi') - v_j(\pi)}_{\text{welcoming coalition}} + \underbrace{\sum_{j \in \pi(i) \setminus \{i\}} v_j(\pi') - v_j(\pi)}_{\text{abandoned coalition}} \\ &= v_i(\pi') - v_i(\pi) + \sum_{j \in \pi'(i) \setminus \{i\}} v_j(i) - \sum_{j \in \pi(i) \setminus \{i\}} v_j(i) \\ &= v_i(\pi') - v_i(\pi) + x(|\pi'(i)| - 1) - \sum_{j \in \pi(i) \setminus \{i\}} v_j(i) \\ &\geq \Delta + x(|\pi'(i)| - 1) - x(|\pi(i)| - 1) \\ &= \Delta + x(|\pi'(i)| - |\pi(i)|). \end{aligned}$$

The third equality comes from the fact that  $i$  performs an IS-deviation, so all agents  $j \in \pi'(i) \setminus \{i\}$  must accept  $i$ , which means they must have  $v_j(i) = x$ . Now, let  $\pi_0$  be any initial partition and consider any sequence of  $k$  successive IS-deviations

$$\pi_0 \xrightarrow{i_1} \pi_1 \xrightarrow{i_2} \dots \xrightarrow{i_k} \pi_k.$$

Telescoping and termwise application of the above inequality yields  $\Phi(\pi_k) - \Phi(\pi_0) = \sum_{j \in [k]} \Phi(\pi_j) - \Phi(\pi_{j-1}) \geq \sum_{j \in [k]} \Delta + x(|\pi_j(i_j)| - |\pi_{j-1}(i_j)|) = k\Delta + x \sum_{j \in [k]} |\pi_j(i_j)| - |\pi_{j-1}(i_j)|$ . We recognize the sum from

the Deviation Lemma, which can be bounded from below using Lemma 2:

$$\Phi(\pi_k) - \Phi(\pi_0) \geq k\Delta - x \frac{n(n-1)}{2}. \quad (2)$$

As the right hand side is unbounded in  $k$ , the sequence must be finite. To be precise, we can bound the potentials of the initial and final partitions by

$$\Phi(\pi_0) \geq -n(n-1)y, \quad \Phi(\pi_k) \leq n(n-1)x.$$

Substituting in these bounds and rearranging for  $k$  gives

$$k \leq \frac{(2y + 3x)n(n-1)}{2\Delta}. \quad (3)$$

$\square$

There are a few important insights gained by the previous proof. First, the bound obtained via the Deviation Lemma does not mean that the potential function  $\Phi$  is increasing in every round. In fact, since utilities are not necessarily symmetric, the deviating agent might move from a rather large coalition to a smaller coalition only improving her utility by  $\Delta$  whereas the utility of all agents in the abandoned coalition are decreased by  $x$ . In fact, the Deviation Lemma does not give us control of the potential function in a single round. Also, it does not control the utility changes caused by the deviator. We apply it to control the utility changes of agents involved in deviations except for the deviator to obtain Equation (2). Hence, we can bound their utility changes by a global constant solely depending on input data. The utility changes caused by the deviator will then eventually lead to the potential reaching a local maximum.

Second, we can easily obtain polynomial bounds on the running time of the dynamics. If  $x$  and  $y$  are polynomially bounded in  $n$  and all valuations are integer, polynomial running time is directly obtained from Equation (3). In particular, this is the case for FEGs, AFGs, and AEGs, so individually stable partitions can be found in polynomial time for these games. After showing two more applications of the Deviation Lemma for other types of deviations, we will capture this observation in Corollary 1.

Third, the previous theorem is tight in the sense that the dynamics can cycle if we have two nonnegative utility values. Indeed, in an instance with agent set  $N = [3]$  and utility values  $v_i(j) = 1, v_j(i) = 0$  for  $(i, j) \in \{(1, 2), (2, 3), (3, 1)\}$ , the dynamics can infinitely cycle among the partitions  $\{\{1, 2\}, \{3\}\}$ ,  $\{\{1\}, \{2, 3\}\}$ , and  $\{\{1, 3\}, \{2\}\}$ . However, the partition consisting of the grand coalition is individually stable and can be reached through the dynamics.

Our next application of the Deviation Lemma considers contractual Nash stability, where we obtain a similar result if we allow at most one nonpositive value. The proof is completely analogous and is therefore omitted. Note that this result also breaks down if we simultaneously allow the utility values  $-1$  and  $0$  by constructing a similar cycle as in the previous example.

**Theorem 4.** *The dynamics of CNS-deviations always converges in ASHG with at most one nonpositive utility value.*

Theorems 3 and 4 use the Deviation Lemma to derive positive results for the single-sided unanimity-based stability notions IS and CNS. In a third application of the deviation lemma, we show that this technique is also applicable to majority-based stability notions, at least when we involve both the welcoming and the abandoned coalition in the vote. The key idea is a suitable arrangement of the terms occurring in the difference of the potential with respect to the agents affected by a deviation.

**Theorem 5.** *The dynamics of JMS-deviations always converges in ASHG with at most two distinct utility values.*

Note that since every JMS-deviation is also an SMS-deviation, the previous result holds for SMS as well. As in the discussion after Theorem 3, we obtain a polynomial running time of the dynamics for appropriate restrictions of the cases. We collect important consequences in the following corollary. In particular, we extend results by Dimitrov et al. (2006) and Aziz and Brandl (2012) who proved the existence of IS partitions for AFGs and AEGs, respectively.<sup>2</sup>

**Corollary 1.** *The dynamics of IS-, CNS-, and JMS-deviations always converges in polynomial time in AFGs, AEGs, and FEGs.*

We would like to stress that convergence of the dynamics does not guarantee a polynomial running time in general. An example is the case of symmetric utility values in ASHG. For NS this can be directly inferred from the PLS-reduction by (Gairing and Savani 2019), which satisfies *tightness*, a property of reductions defined by Schäffer and Yannakakis (1991).

**Proposition 1.** *The dynamics of NS-deviations in symmetric ASHG may require exponentially many rounds before converging to an NS partition.*

*Proof.* It is easy to verify that the PLS-reduction from PARTYAFFILIATION under the Flip neighborhood by Gairing and Savani (2019, Observation 2) is tight. Schäffer and Yannakakis (1991, Lemma 3.3) showed that tight reductions preserve the existence of exponentially long running times of the standard local search algorithm, i.e., the NS-dynamics in our case. Note that the standard local search algorithm of the source problem can have an exponential running time, because PARTYAFFILIATION is a generalization of MAXCUT whose standard local search algorithm can run in exponential time with respect to the flip neighborhood (Schäffer and Yannakakis 1991, Theorem 5.15).<sup>3</sup>  $\square$

While the previous proposition uses a nonconstructive argument avoiding to construct an explicit example with an exponential running time, it is possible to construct such an example even in the more restricted case of IS-dynamics. To this end, it is possible to modify an example for MAXCUT provided by Monien and Tscheuschner (2010) by essentially

<sup>2</sup>These contributions actually show existence of partitions satisfying properties stronger than IS.

<sup>3</sup>We refer to the respective references for formal definitions of the involved combinatorial problems.

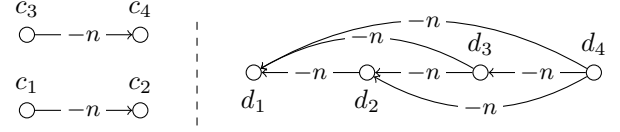


Figure 2: The *aversion to enemies* games without MIS partition (left) and MOS partition (right) from Proposition 3. Omitted edges have weight 1.

reverting the sequence of flips for MAXCUT to obtain an execution of the IS-dynamics. Thus, we generalize the previous proposition via a constructive proof.

**Proposition 2.** *The dynamics of IS-deviations in symmetric ASHG may require exponentially many rounds before converging to an IS partition.*

### Stability under Majority Consent

In this section, we study stability under majority consent. First, the existential results of Theorem 3 and Theorem 4 are contrasted with the non-existence of stable partitions in AEGs under the majority-based relaxations of the respective stability concepts.

**Proposition 3.** *There exists an AEG which contains no MIS (respectively, MOS) partition.*

*Proof.* First, we provide an AEG with no MIS partition. Let  $N = \{c_1, c_2, c_3, c_4\}$ , i.e., there are  $n = 4$  agents, and valuations defined as  $v_{c_1}(c_2) = v_{c_3}(c_4) = -n$  and all other valuations set to 1. The AEG is illustrated in Figure 2 (left).

Assume for contradiction that there exists an MIS partition  $\pi$ . Then,  $c_1 \notin \pi(c_2)$  and  $c_3 \notin \pi(c_4)$ . Also,  $|\pi(c_1)| \leq 1$  (respectively,  $|\pi(c_3)| \leq 1$ ), because otherwise,  $c_2$  (respectively,  $c_4$ ) would join via an MIS-deviation). But then  $\pi(c_1) = \{c_1\}$  and  $\pi(c_3) = \{c_3\}$ , and  $c_1$  could deviate to join  $\pi(c_3)$ , a contradiction.

Second, we provide an AEG without MOS partition. Let  $N = \{d_1, d_2, d_3, d_4\}$ , and define valuations for all  $i, j \in [4]$  with  $i < j$  as  $v_{d_i}(d_j) = 1$  and  $v_{d_j}(d_i) = -4$ . An illustration is provided in Figure 2 (right).

Assume for contradiction that there exists an MOS partition  $\pi$ . Then, every coalition  $C \in \pi$  must fulfill  $|C| \leq 2$ . Otherwise, the agent of  $C$  with the second smallest index would form a singleton via an MOS-deviation. In addition, there cannot be a singleton, because if some agent is in a singleton, there must be a second such agent, and then the one with the smaller index would join the other one. Hence,  $\pi$  consists of two pairs. But then  $d_1$  would deviate to the pair not containing her, a contradiction.  $\square$

We can leverage the AEGs provided in the previous proposition as gadgets in reductions to show hardness of the associated decision problems. This can be interpreted as a more exact boundary (compared to Theorem 1) of the tractabilities encountered in Theorem 3 and Theorem 4 for the special case of AEGs.

**Theorem 6.** *It is NP-complete to decide if there exists an MIS (respectively, MOS) partition in AEGs.*



The utility restrictions in Theorem 6 are not as flexible as in the negative result for Nash stability in Theorem 1 or the positive results for unanimity-based dynamics in Theorem 3 and Theorem 4. In fact, the picture for majority-based notions is more diverse, because we obtain another positive result in the class of AFGs.

**Theorem 7.** *When starting from the grand coalition, the dynamics of MIS-deviations converges after at most  $n$  rounds in AFGs.*

*Proof.* The key insight is that there can only be deviations to form a new singleton coalition yielding no more than  $n$  deviations. Let  $\pi_0 = \{N\}$  be the initial partition, and consider a sequence of  $k$  MIS-deviations

$$\pi_0 \xrightarrow{i_1} \pi_1 \xrightarrow{i_2} \dots \xrightarrow{i_k} \pi_k.$$

We inductively define coalitions evolving from the grand coalition if removing the deviator as  $G_0 = N$ , and  $G_j = G_{j-1} \setminus \{i_j\}$  for  $j > 0$ .

Now, we proceed to simultaneously prove the following claims by induction:

1.  $\forall j \in [k] : \pi_{j-1}(i_j) = G_{j-1}$ .
2.  $\forall j \in [k] : \pi_j(i_j) = \{i_j\}$ .
3.  $\forall j \in [k] : \{i \in \pi_{j-1}(i_j) : v_{i_j}(i) = n\} = \emptyset$ .

The base case  $j = 1$  is immediate. For the induction step, let  $2 \leq j \leq k$  and suppose the claims are true for all  $1 \leq l < j$ . We start with the first claim. By the induction hypothesis,  $\pi_{j-1} = \{G_{j-1}\} \cup \{\{i_l\} : 1 \leq l < j\}$ . This means that if  $\pi_{j-1}(i_j) \neq G_{j-1}$ , we must have  $\pi_{j-1}(i_j) = \{i_l\}$ , indicating  $i_j = i_l$  for some  $l < j$ . Then, the welcoming coalition cannot be  $G_{j-1}$ , as  $i_j$ , by induction hypothesis, abandoned  $G_{l-1}$  due to not having any friends in  $G_{l-1}$ , and thus has, by  $G_{j-1} \subseteq G_{l-1}$ , no friends in  $G_{j-1}$ , either. The alternative is that  $i_j$  joins another singleton coalition  $\{i_m\}$  to form a pair. However, if  $i_m$  abandoned  $G_m$  at some point  $m < l$ , then she dislikes  $i_j$ , and won't allow her to join. If  $i_m$  abandoned  $G_m$  at some point  $m > l$ , then  $i_j$  dislikes  $i_m$ , and has no incentive to join. Hence,  $\pi_{j-1}(i_j) = G_{j-1}$ . For the second claim, note that  $i_j$  cannot join another singleton  $\{i_m\}$ , because  $i_m$  abandoned  $G_{m-1}$  at some point  $m < j$  and thus dislikes  $i_j$ . Hence,  $i_j$  must form a singleton  $\pi_j(i_j) = \{i_j\}$ , which she only wants to do if  $\{i \in \pi_{j-1}(i_j) : v_{i_j}(i) = n\} = \emptyset$ . This accomplishes the third claim, and completes the induction proof.

Finally, as there can be at most  $n$  singletons, the dynamics must terminate after at most  $n$  rounds.  $\square$

The computational boundaries in this section encountered so far only hold for one-sided stability notions where either the welcoming or the abandoned coalition takes a vote. On the other hand, Theorem 5 shows that these are opposed by tractabilities under two-sided majority consent.

For general utilities, existence of SMS (and therefore JMS) partitions is not guaranteed anymore, and we show that the tractabilities break down.

**Theorem 8.** *Deciding whether an ASHG contains an SMS (respectively, JMS) partition is NP-complete.*

## Conclusion and Discussion

We studied stability based on single-agent deviations in additively separable hedonic games with a particular focus on games with restricted utility functions that can be naturally interpreted in terms of friends and enemies. We identified a computational boundary between Nash stability and stability with unanimous consent. The picture is less clear when deviations are governed by majority consent. While stable partitions always exist when considering both the abandoned and the welcoming coalition of the deviating agent, we obtain both positive and negative results if only one of these coalitions is considered. Table 1 summarizes our results and compares them with related results from the literature. Notably, we obtain all of our positive results through the convergence of simple and natural dynamics. This also extends previously known results about IS. Aziz and Brandl (2012) obtain a polynomial algorithm essentially by running a dynamics from the singleton partition, whereas Dimitrov et al. (2006) take a different, graph-theoretical approach considering strongly connected components. The construction of CIS partitions by Aziz, Brandt, and Seidig (2013) is done by iteratively identifying specific coalitions, and it is not known whether CIS-dynamics converge in polynomial time for natural starting partitions such as the singleton partition or grand coalition. An important tool in establishing our results concerning convergence of dynamics is the Deviation Lemma, a general combinatorial insight that allows us to study dynamics from a global perspective.

Table 1: Overview of our computational results. A red cell means existence of games without stable partition and usually comes with computational intractability. A green cell means that a stable partition can be constructed in polynomial time (Function-P), and in the case of our results even by executing a dynamics. A white cell means that it is unknown whether a stable partition always exists.

<sup>a</sup>: Aziz and Brandl (2012), <sup>b</sup>: Aziz, Brandt, and Seidig (2013), <sup>c</sup>: Dimitrov et al. (2006), <sup>d</sup>: Sung and Dimitrov (2010)

	General	FEGs	AEGs	AFGs
NS	NP-c <sup>d</sup>	NP-c (Th. 1)	NP-c (Th. 1)	NP-c (Th. 2)
IS	NP-c <sup>d</sup>	FP (Th. 3)	FP <sup>a</sup> (Th. 3)	FP <sup>c</sup> (Th. 3)
CNS	NP	FP (Th. 4)	FP (Th. 4)	FP (Th. 4)
CIS	FP <sup>b</sup>	FP <sup>b</sup>	FP <sup>b</sup>	FP <sup>b</sup>
MIS	NP-c (Th. 6)	?	NP-c (Th. 6)	FP (Th. 7)
MOS	NP-c (Th. 6)	?	NP-c (Th. 6)	?
JMS	NP-c (Th. 8)	FP (Th. 5)	FP (Th. 5)	FP (Th. 5)
SMS	NP-c (Th. 8)	FP (Th. 5)	FP (Th. 5)	FP (Th. 5)

Our work offers a wide range of interesting follow-up questions. First, Table 1 contains some problems left open in our analysis. Specifically, despite the existence of partitions without CNS partitions, the complexity of the existence problem of CNS partitions remains open for general utilities. Also, the voting-based stability notions deserve further investigation, and might even lead to interesting discoveries in other classes of hedonic games. Lastly, an intriguing further direction is to study further applications of the Deviation Lemma, particularly in domains other than coalition formation.

## Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft under grant BR 2312/12-1.

## References

- Alcalde, J.; and Revilla, P. 2004. Researching with whom? Stability and manipulation. *Journal of Mathematical Economics*, 40(8): 869–887.
- Aziz, H.; and Brandl, F. 2012. Existence of Stability in Hedonic Coalition Formation Games. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 763–770.
- 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.
- 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.
- 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. 2020. Finding and Recognizing Popular Coalition Structures. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 195–203.
- 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.
- 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.
- Cseh, Á. 2017. Popular Matchings. In Endriss, U., ed., *Trends in Computational Social Choice*, chapter 6. AI Access.
- Dimitrov, D.; Borm, P.; Hendrickx, R.; and Sung, S. C. 2006. Simple Priorities and Core Stability in Hedonic Games. *Social Choice and Welfare*, 26(2): 421–433.
- Dimitrov, D.; and Sung, S. C. 2007. On top responsiveness and strict core stability. *Journal of Mathematical Economics*, 43(2): 130–134.
- Drèze, J. H.; and Greenberg, J. 1980. Hedonic Coalitions: Optimality and Stability. *Econometrica*, 48(4): 987–1003.
- Gairing, M.; and Savani, R. 2019. Computing Stable Outcomes in Symmetric Additively Separable Hedonic Games. *Mathematics of Operations Research*, 44(3): 1101–1121.
- Gärdenfors, P. 1975. Match Making: Assignments based on bilateral preferences. *Behavioral Science*, 20(3): 166–173.
- Hoefler, M.; Vaz, D.; and Wagner, L. 2018. Dynamics in matching and coalition formation games with structural constraints. *Artificial Intelligence*, 262: 222–247.
- Kahn, A. B. 1962. Topological sorting of large networks. *Communications of the ACM*, 5(11): 558–562.
- Karp, R. M. 1972. Reducibility among Combinatorial Problems. In Miller, R. E.; and Thatcher, J. W., eds., *Complexity of Computer Computations*, 85–103. Plenum Press.
- Monien, B.; and Tscheuschner, T. 2010. On the power of nodes of degree four in the local max-cut problem. In *Proceedings of the 7th International Conference on Algorithms and Complexity (CIAC)*, number 6078 in Lecture Notes in Computer Science (LNCS), 264–275. Springer-Verlag.
- Newman, M. E. J. 2004. Detecting community structure in networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38(2): 321–330.
- 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.
- Saad, W.; Han, Z.; Basar, T.; Debbah, M.; and Hjørungnes, A. 2011. Hedonic Coalition Formation for Distributed Task Allocation among Wireless Agents. *IEEE Transactions on Mobile Computing*, 10(9): 1327–1344.
- Schäffer, A. A.; and Yannakakis, M. 1991. Simple Local Search Problems that are Hard to Solve. *SIAM Journal on Computing*, 20(1): 56–87.
- Suksompong, W. 2015. Individual and Group Stability in Neutral Restrictions of Hedonic Games. *Mathematical Social Sciences*, 78: 1–5.
- Sung, S. C.; and Dimitrov, D. 2010. Computational Complexity in Additive Hedonic Games. *European Journal of Operational Research*, 203(3): 635–639.