

Deviation Dynamics in Cardinal Hedonic Games*

Valentin Zech¹ and Martin Bullinger²

¹ Department of Computer Science, University of Oxford, UK

² School of Engineering Mathematics and Technology, University of Bristol, UK
zech@vzech.de, martin.bullinger@bristol.ac.uk

Abstract

Computing stable partitions in hedonic games is a challenging task because there exist games in which stable outcomes do not exist. Even more, these No-instances can often be leveraged to prove computational hardness results. We make this impression rigorous in a dynamic model of cardinal hedonic games by providing meta theorems. These imply hardness of deciding about the possible or necessary convergence of deviation dynamics based on the mere existence of No-instances. Our results hold for additively separable, fractional, and modified fractional hedonic games (ASHGs, FHGs, and MFHG). Moreover, they encompass essentially all reasonable stability notions based on single-agent deviations. In addition, we propose dynamics as a method to find individually rational and contractually individual stable (CIS) partitions in ASHGs. In particular, we find that CIS dynamics from the singleton partition possibly converge after a linear number of deviations but may require an exponential number of deviations in the worst case.

1 Introduction

The field of Computational Social Choice (COMSOC) is concerned with aggregating potentially conflicting individual preferences of different agents into a compromise solution (Brandt et al., 2016). With various applications to, among others, politics, multi-agent systems, and economic processes, coalition formation is among the primary areas of interest within COMSOC (Ray and Vohra, 2015). Here, a group of agents must be divided into distinct coalitions, with each agent having preferences for these divisions.

A common restriction on agents' preferences is that their utility depends only on which agents are present in their own coalition. This restriction describes the model of so-called *hedonic games* (Drèze and Greenberg, 1980). Since their introduction, they have been a constant area of interest in the literature on artificial intelligence and multi-agent systems (Aziz and Savani, 2016; Bullinger et al., 2024). Hedonic games have been successfully utilized to model many interesting real-world settings, such as research team formation (Alcalde and Revilla, 2004), allocation of indivisible goods (Peters, 2016), task allocation for wireless agents (Saad et al., 2011), and community detection in social networks (Aziz et al., 2019). Further, they have proven to be a powerful theoretical model in the context of clustering (Feldman et al., 2015; Ahmadi et al., 2022; Cohen-Addad et al., 2022), one of the central research topics in the realm of machine learning.

*We dedicate this work to Dr. Jochen Zech. Rest in peace.

A prominent measure for the desirability of outcomes in hedonic games is stability, defined as the absence of beneficial deviations by agents to join other coalitions (Bogomolnaia and Jackson, 2002). In certain scenarios, it is sensible to additionally require partial or unanimous consent of the otherwise affected agents, which give rise to a wide landscape of notions of stability (Aziz and Savani, 2016). We will focus on those defined by deviations of single agents.

Some stability notions guarantee a stable outcome in any hedonic game, e.g., contractual individual stability where a deviation requires unanimous consent of all involved agents. For most stability notions, however, stable partitions are not guaranteed to exist, even in fairly restricted game classes. This gives rise to the problem of deciding whether a given hedonic game admits a stable partition. A common observation is that No-instances, i.e., games without a stable partition, can be used as gadgets to prove computational boundaries of the existence problem (see, e.g., Sung and Dimitrov, 2010; Aziz et al., 2013; Peters and Elkind, 2015; Brandt et al., 2024).¹

The work discussed so far is only concerned with whether an outcome is stable or not, while it matters less how this outcome is obtained. One natural way to model the process of obtaining stable outcomes are deviation dynamics, where the agents start in some initial state and then iteratively perform deviations as long as they have an incentive to do so, see, e.g., (Bilò et al., 2018; Gairing and Savani, 2019; Brandt et al., 2023, 2024). Such dynamics have previously been utilized successfully, e.g., to show that partitions satisfying a specific stability notion always exist in a particular game class (Bogomolnaia and Jackson, 2002; Boehmer and Elkind, 2020; Brandt et al., 2024; Çaskurlu and Kizilkaya, 2024), to study the complexity of computing stable outcomes (Gairing and Savani, 2019) or to place an upper bound on the price of stability in terms of achieving high social welfare (Bilò et al., 2018; Monaco et al., 2020). While dynamics are, therefore, a powerful general tool, scenarios in which dynamics are guaranteed to converge offer a decentralized approach to reaching desirable partitions. Thus, they give rise to interesting questions in their own right. Specifically, Brandt et al. (2023) ask whether, given a hedonic game and a starting partition, dynamics possibly or necessarily converge, i.e., reach a stable partition.

1.1 Contribution

We will make the intuition that No-instances lead to computational intractabilities explicit. In contrast to previous work that explicitly constructs No-instances and uses them to prove individual hardness results (see, e.g., Sung and Dimitrov, 2010; Brandt et al., 2024), we present meta-theorems that treat No-instances as a black box. This approach enables future hardness results to be derived by identifying a single suitable instance. The meta theorems concern the intractability of possible and necessary convergence of dynamics, and apply to three prominent classes of hedonic games: additively separable (Bogomolnaia and Jackson, 2002), fractional (Aziz et al., 2019), and modified fractional (Olsen, 2012) hedonic games. They hold for most reasonable stability notions based on deviations between Nash deviations (which simply need to make the deviator better off) and contractual individual deviations (which additionally require the consent of all other agents). We demonstrate the generality of our meta theorems by applying them for a general class of voting-based stability notions that encompass a wide range of known and new stability notions.

Finally, we zoom in on a special case of dynamics that necessarily converge, namely those based on contractual individual deviations for additively separable hedonic games. When starting from

¹Two notable exceptions are aversion-to-enemies games and the locally egalitarian variant of hedonic games. In both cases, the core is nonempty but an outcome in the core is NP-hard to compute (Dimitrov et al., 2006; Bullinger and Kober, 2021).

the singleton partition, the resulting partition additionally is individually rational, i.e., at least as good for each agent as being on her own. We show that fast convergence is always possible. It is, however, unclear how to efficiently identify the associated deviations. Simply running any sequence of deviations may take an exponential number of steps. Nonetheless, we identify the structural reason behind this result, leading to a fixed-parameter tractability result based on the number of certain valuation pairs.

1.2 Related Work

Hedonic games were first introduced by Drèze and Greenberg (1980), and later popularized by Bogomolnaia and Jackson (2002), Banerjee et al. (2001), and Cechlárová and Romero-Medina (2001). An overview is provided in the book chapters by Aziz and Savani (2016) and Bullinger et al. (2024).

The axiomatic and computational properties of stability have been studied extensively in cardinal hedonic games (see, e.g., Dimitrov et al., 2006; Sung and Dimitrov, 2010; Aziz et al., 2013; Woeginger, 2013; Bilò et al., 2018; Aziz et al., 2019; Boehmer and Elkind, 2020; Brandt et al., 2024). Sung and Dimitrov (2010), specifically, provide a detailed overview of stability based on single-agent deviations in additively separable hedonic games. Related to our efforts to study the computational complexity of finding a partition that is individually rational and contractually individually stable, Aziz et al. (2013) provide an algorithm for computing a (not necessarily individually rational) partition that is contractually individually stable in additively separable hedonic games. Further, Peters and Elkind (2015) utilize a meta approach to show hardness for several game classes and stability notions simultaneously, similar to our unified theory. In contrast to our investigation of deviation dynamics, their paper concerns the general existence of stable outcomes.

In this light, a recent trend has been to study the dynamic aspects of coalition formation based on beneficial deviations, which offer a decentralized approach to finding stable outcomes and can thus model specific real-world scenarios more realistically. Most related is the work by Brandt et al. (2023) that studies the computational complexity of possible and necessary convergence of dynamics in a variety of game classes. The only overlap with our work is the consideration of fractional hedonic games. While Brandt et al. (2023) only study individual stability, our meta theorems work for a much larger set of stability notions and additionally concerns other classes of cardinal hedonic games. Subsequently, Bullinger and Suksompong (2024) study possible and necessary convergence for the equivalent of Nash stability in a generalization of additively separable hedonic games.

Further, Bilò et al. (2018) study Nash stability in fractional hedonic games and, for instance, utilize dynamics to design an algorithm that approximates the maximum social welfare of a Nash stable outcome in polynomial time. Gairing and Savani (2019) settle the complexity of deciding whether a stable partition exists in symmetric additively separable hedonic games by treating this question as local search problems. Brandt et al. (2024) also study computational questions related to the existence of stable partitions, where all their positive results are obtained by proving convergence of dynamics. Boehmer et al. (2023) propose a version of hedonic games specifically adapted to a dynamic setting, where utilities change after a deviation takes place. Their work also has implications for a fixed-utility setting: In particular, they consider the computational complexity of convergence in a given time limit and prove hardness results for additively separable hedonic games. In addition, Hoefer et al. (2018), Bullinger and Kober (2021), and Fanelli et al. (2021) study dynamics in hedonic games based on group deviations. Finally, we note that similar

dynamic approaches to finding stable solutions have been studied in the context of stable matchings (Abeledo and Rothblum, 1995; Hoefer et al., 2018; Brandt and Wilczynski, 2024).

Further, the study of dynamic processes has recently received increased interest in the research community of computational social choice and collective decision-making at large (see, e.g., Zech et al., 2024; Elkind et al., 2024; Igarashi et al., 2024; Caragiannis and Narang, 2024).

2 Preliminaries

In this section, we introduce preliminaries. We use the convention that $\mathbb{N}$ is the set of nonnegative integers, including 0. For $i \in \mathbb{N}$, $i \geq 1$, we denote $[i] := \{1, \dots, i\}$.

2.1 Hedonic Games

We consider a finite set N of $n := |N|$ agents. A nonempty subset of agents is called a *coalition*. We aim to partition the agents in N into disjoint coalitions. A *coalition structure* (or *partition*) of N is a subset $\pi \subseteq 2^N$ with $\bigcup_{C \in \pi} C = N$, where, for all $C, D \in \pi$, it holds that $C = D$, or $C \cap D = \emptyset$. Given an agent $a \in N$, we denote by $\pi(a)$ the coalition in π that contains a . Let $\mathcal{N}_a = \{C \subseteq N \mid a \in C\}$ denote the set of all coalitions that a can belong to. We refer to the partition $\pi = \{\{a\} \mid a \in N\}$ as the *singleton partition*, and to $\pi = \{N\}$ as the *grand coalition*. Further, for each agent $a \in N$, we call $\{a\}$ the *singleton coalition* of a .

A *hedonic game* $G = (N, \succsim)$ consists of a set N of agents, and a *preference profile* $\succsim = (\succsim_a)_{a \in N}$ where $\succsim_a \subseteq \mathcal{N}_a \times \mathcal{N}_a$ is a complete, reflexive, and transitive binary relation called agent a 's *preference relation* (Dr ze and Greenberg, 1980). Given two coalitions $C, D \in \mathcal{N}_a$, we write $C \succ_a D$ if $C \succsim_a D$ but not $D \succsim_a C$ (i.e., a *strictly prefers* C over D). We say that a partition π is *individually rational* if $\pi(a) \succsim_a \{a\}$ for each agent $a \in N$, i.e., no agent would strictly prefer to be in her respective singleton coalition.

Agents have preferences over partitions based on preferences over coalitions. Given two partitions π, π' of N , we say that $\pi \succsim_a \pi'$ if and only if $\pi(a) \succsim_a \pi'(a)$. Further, we denote by $G - a$ the game with agent set $N \setminus \{a\}$ that is induced by G by removing agent a . We write $\pi - a$ to mean the partition of $N \setminus \{a\}$ that resulted from π by removing a from her coalition, formally, $\pi - a := \{C \setminus \{a\} \mid C \in \pi, C \neq \{a\}\}$.

We consider classes of hedonic games in which preference relations evolve from cardinal utility functions, i.e., agents have numeric value for each coalition and preferences are based on comparing these values. Formally, a *cardinal hedonic game* is given by the pair (N, u) where N is the agent set and $u = (u_a : \mathcal{N}_a \rightarrow \mathbb{Q})_{a \in N}$ a profile of *utility functions*. Then, (N, u) induces the hedonic game $(N, \succsim)$ where, for every agent $a \in N$ and coalitions $C, D \in \mathcal{N}_a$, we define $C \succsim_a D$ if and only if $u_a(C) \geq u_a(D)$. We say that $u_a(C)$ is a 's utility for coalition C and extend this to utilities for partitions by setting $u_a(\pi) := u_a(\pi(a))$.

Cardinal hedonic games generally require to specify a utility for an exponentially large set of coalitions. To avoid listing these all explicitly, several classes of cardinal hedonic games have been proposed where utility functions are represented succinctly by merely specifying valuations for single agents. Let $G = (N, u)$ be a cardinal hedonic game and let $(v_a : N \rightarrow \mathbb{Q})_{a \in N}$ be a collection of valuation functions.

Following Bogomolnaia and Jackson (2002), G is called an *additively separable hedonic game* (ASHG) if for all $a \in N, C \in \mathcal{N}_a$ it holds that $u_a(C) = \sum_{b \in C \setminus \{a\}} v_a(b)$. Following Aziz et al.

(2019), G is called a *fractional hedonic game* (FHG) if for all $a \in N, C \in \mathcal{N}_a$ it holds that $u_a(C) = \sum_{b \in C \setminus \{a\}} \frac{v_a(b)}{|C|}$. Following Olsen (2012), G is called a *modified fractional hedonic game* (MFHG) if for all $a \in N$, it holds that $u_a(\{a\}) = 0$ and for all $C \in \mathcal{N}_a, C \neq \{a\}$ it holds that $u_a(C) = \sum_{b \in C \setminus \{a\}} \frac{v_a(b)}{|C|-1}$.

In other words, the utility in an ASHG is the sum of valuations for agents in the considered coalition, and the utility in an FHG and MFHG is the average valuation, where FHGs include the consideration of the agent herself. All three game classes are fully specified by the valuation functions and we therefore also represent an ASHG, FHG, or MFHG G by the pair (N, v) , where $v = (v_a : N \rightarrow \mathbb{Q})_{a \in N}$ is a profile of *valuation functions*.

Note that the valuation functions of ASHGs, FHGs, and MFHGs can be represented as a weighted directed graph, where the vertices are agents, and, given two agents $a, b \in N$, there is an edge from a to b with weight $v_a(b)$.

2.2 Single-Agent Stability

We now formalize how to capture stability based on beneficial deviations by single agents. Given a hedonic game $G = (N, \succ)$, a *single-agent deviation* of an agent $a \in N$ transforms a partition π of N into a partition π' of N , where $\pi(a) \neq \pi'(a)$, and, for all agents $b \in N \setminus \{a\}$, it holds that $\pi(b) \setminus \{a\} = \pi'(b) \setminus \{a\}$. We denote such a deviation by $\pi \xrightarrow{a} \pi'$. Intuitively, agent a deviates away from coalition $\pi(a)$, to join coalition $\pi'(a)$ (importantly, $\pi'(a)$ can be a 's singleton coalition), while all other coalitions remain unchanged.

A minimum requirement for the desirability of a deviation is whether the deviator is better off by performing this deviation. A *Nash deviation* is a single-agent deviation $\pi \xrightarrow{a} \pi'$ of an agent $a \in N$ such that $\pi'(a) \succ_a \pi(a)$. A partition π which does not admit a Nash deviation is said to be *Nash stable* (NS), and π is called an NS partition.

While Nash stability offers a very strong and desirable solution concept, NS deviations completely disregard the opinion of members in the abandoned and welcoming coalition. In this light, several stability notions enforce additional requirements to be satisfied for a deviation to be valid. We introduce a general class of such stability notions based on voting among the involved agents.

Let $C \subseteq N$ be a coalition and $a \in N$ an agent. Following Brandt et al. (2024), we define the *favour-in set* $F_{\text{in}}(C, a)$ and *favour-out set* $F_{\text{out}}(C, a)$ of C with respect to a as

$$F_{\text{in}}(C, a) := \{b \in C \setminus \{a\} \mid C \cup \{a\} \succ_b C \setminus \{a\}\} \text{ and} \\ F_{\text{out}}(C, a) := \{b \in C \setminus \{a\} \mid C \setminus \{a\} \succ_b C \cup \{a\}\}.$$

These capture the agents in C that prefer a inside or outside the coalition C . Note that the definition is valid regardless of whether a is part of C .

Let $q_{\text{out}}, q_{\text{in}} \in [0, 1]$ be two real numbers interpreted as quotas. A Nash deviation $\pi \xrightarrow{a} \pi'$ of an agent $a \in N$ is called a $(q_{\text{out}}, q_{\text{in}})$ -vote deviation if

1. $|F_{\text{out}}(\pi(a), a)| \geq q_{\text{out}}(|F_{\text{in}}(\pi(a), a)| + |F_{\text{out}}(\pi(a), a)|)$ and
2. $|F_{\text{in}}(\pi'(a), a)| \geq q_{\text{in}}(|F_{\text{in}}(\pi'(a), a)| + |F_{\text{out}}(\pi'(a), a)|)$.

Hence, such a deviation requires that at least a q_{out} -fraction of the nonindifferent members of the abandoned coalition and a q_{in} -fraction of the nonindifferent members of the welcoming coalition are

strictly in favor of the deviation. Now, a partition is said to be $(q_{\text{out}}, q_{\text{in}})$ -*voting-stable* $((q_{\text{out}}, q_{\text{in}})$ -VS) if it does not admit a $(q_{\text{out}}, q_{\text{in}})$ -vote deviation.

Our stability framework captures most single-deviation stability notions commonly considered in the literature. If $q_{\text{out}}, q_{\text{in}} \in \{0, 1\}$, we obtain stability notions based on unanimous consent whenever consent is required. Specifically, $(0, 0)$ -VS is NS, $(0, 1)$ -VS is called *individual stability* (IS), $(1, 0)$ -VS is called *contractual Nash stability* (CNS), and $(1, 1)$ -VS is called *contractual individual stability* (CIS) (Bogomolnaia and Jackson, 2002; Sung and Dimitrov, 2007). In addition, $(q_{\text{out}}, q_{\text{in}})$ -VS generalizes previously studied voting-based stability concepts: Gairing and Savani (2019) consider $(0, q_{\text{in}})$ -VS and $(q_{\text{out}}, 0)$ -VS under the names of *vote-in* and *vote-out stability* (VIS and VOS), and Brandt et al. (2024) consider $(0, \frac{1}{2})$ -VS, $(\frac{1}{2}, 0)$ -VS, and $(\frac{1}{2}, \frac{1}{2})$ -VS. Brandt et al. (2024) call the latter *separate-majorities stability* (SMS). Among all of these, only CIS guarantees the existence of stable partitions.

Given a stability notion χ , we refer to the corresponding deviations and stable partitions as χ deviations and χ partitions, respectively.

Given two stability notions χ and χ' , we write $\chi \subseteq \chi'$ if every χ deviation is also a χ' deviation. For instance, for every $q_{\text{out}}, q_{\text{in}} \in [0, 1]$, it holds that $(q_{\text{out}}, q_{\text{in}})$ -VS $\subseteq$ NS and CIS $\subseteq (q_{\text{out}}, q_{\text{in}})$ -VS.

2.3 Standard Stability Notions

In the last section, we introduced a class of specific stability notions based on voting. To state our meta theorems, we propose a novel condition to capture an even more general class of stability notions between NS and CIS. These are defined for cardinal hedonic games and should satisfy two properties:

1. the feasibility of deviations only depends on the utility changes of the involved agents, not their identities, i.e., deviations are anonymously hedonic,
2. deviations that are stronger than feasible deviations are also feasible, i.e., deviations are monotonic.

We formalize this in the following.

Let $G = (N, u)$ be a cardinal hedonic game, let $a \in N$ be an agent, and let π, π' be two partitions of N . We refer to $uc_G(a, \pi, \pi') := (u_a(\pi), u_a(\pi'))$ as the *utility-change tuple* of a with respect to G , π and π' . Further, we refer to the multisets $UC_G^{\text{out}}(a, \pi, \pi') := \{uc_G(b, \pi, \pi') \mid b \in \pi(a) \setminus \{a\}\}$ and $UC_G^{\text{in}}(a, \pi, \pi') := \{uc_G(b, \pi, \pi') \mid b \in \pi'(a) \setminus \{a\}\}$ as the *utility-change-out multiset* and *utility-change-in multiset* of a with respect to G , π and π' , respectively. We denote by $\mathcal{UC}$ the set of all utility-change multisets (that is, both utility-change-out and utility-change-in multisets). For all functions uc_G , UC_G^{out} , and UC_G^{in} , we will omit the game G whenever it is clear from the context.

We say that a stability notion χ is *anonymously hedonic* if there exists a polynomial-time computable function $f_\chi : \mathcal{UC} \times \mathcal{UC} \times (\mathbb{Q} \times \mathbb{Q}) \rightarrow \{0, 1\}$, such that for all single-agent deviations $\pi \xrightarrow{a} \pi'$, it holds that

$$f_\chi(UC^{\text{out}}(a, \pi, \pi'), UC^{\text{in}}(a, \pi, \pi'), uc(a, \pi, \pi')) = \begin{cases} 1, & \text{if } \pi \xrightarrow{a} \pi' \text{ is a } \chi \text{ deviation,} \\ 0, & \text{otherwise.} \end{cases}$$

Simply put, the validity of a deviation with respect to an anonymously hedonic stability notion solely depends on the changes in the utility of abandoned and welcoming coalitions and that of the

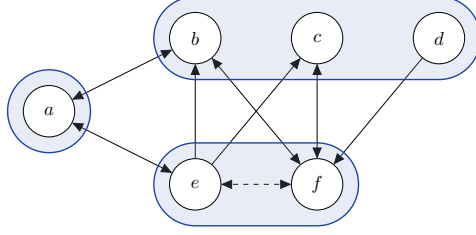


Figure 1: Illustration of an ASHG. Blue boxes indicate the initial partition. A straight arrow from an agent x to an agent y indicates $v_x(y) = 1$ while a dashed arrow indicates $v_x(y) = -1$. Missing arrows indicate a valuation of 0.

deviator. In particular, this captures all stability notions that are implied by NS, and further only depend on the sizes of the favor-in and favor-out sets of the abandoned and welcoming coalitions. However, the class of anonymously hedonic stability notions allows for more nuanced requirements. For example, a deviation may be allowed if it is an NS deviation, and increases the utilitarian welfare, defined as $\sum_{a \in N} u_a(\pi)$ for partition π .

Next, given two multisets $X, Y \in \mathcal{UC}$, we say that X *dominates* Y , written $Y \trianglelefteq X$, if it holds that $|Y| \leq |X|$ and:

$$\forall (y, y') \in Y, (x, x') \in X : y' - y \leq x' - x.$$

Now, an anonymously hedonic stability notion χ is *monotone* if, for all $X, X', Y, Y' \in \mathcal{UC}$, and $z, z' \in \mathbb{Q} \times \mathbb{Q}$, where $X \trianglelefteq X'$, $Y \trianglelefteq Y'$, and $\{z\} \trianglelefteq \{z'\}$, it holds that:

$$f_\chi(X, Y, z) \leq f_\chi(X', Y', z'),$$

i.e., whenever a deviation is allowed with parameters X, Y, z , then it must also be allowed with parameters X', Y', z' . We will refer to anonymously hedonic monotone stability notions as *standard stability notions*.

We illustrate standard stability with an example.

Example 1. Consider the ASHG depicted in Figure 1 with agents acting according to a standard stability notion. First, assume that agent b has a deviation to join $\{a\}$. Then one can verify that both b and c must also be allowed to deviate to join $\{e, f\}$. However, we cannot infer any further deviations. In particular, despite the fact that the abandoned agent of a deviation of an agent in $\{e, f\}$ strictly increases her utility, we cannot infer whether, e.g., e can deviate to join $\{a\}$, since $|\{e, f\}| < |\{b, c, d\}|$.

Next, assume that f cannot deviate to join $\{b, c, d\}$. Then, it holds that e can also not deviate to join $\{b, c, d\}$ or $\{a\}$. However, perhaps somewhat surprisingly, we cannot say whether, e.g., agent d can deviate to join $\{e, f\}$, again since $|\{e, f\}| < |\{b, c, d\}|$.

We note that our voting-based stability notions are standard stability notions, as the relevant favor-in and favor-out sets can be reconstructed with the information captured in the utility-change multisets. We defer the formal proof to Appendix A.

Proposition 2. Let $q_{\text{out}}, q_{\text{in}} \in [0, 1]$. Then, $(q_{\text{out}}, q_{\text{in}})$ -VS is a standard stability notion.

We remark that our notion of monotonicity does not capture all stability notions between NS and CIS, e.g., it fails to capture some notions that rely on the egalitarian welfare.

2.4 Deviation Dynamics

We are ready to introduce the central concept of this paper, which we will utilize to formulate our decision problems.

Stability notions naturally induce *dynamics*, where, given a hedonic game and a starting partition of the agents, we iteratively obtain successor partitions by letting agents perform deviations from the current partition in alignment with the stability notion.

Formally, let χ be a stability notion, let $G = (N, \succsim)$ be a hedonic game with a set N of agents, and let π_0 be a partition of N . Then, an *execution of the χ dynamics* of (G, π_0) is a finite or infinite sequence $(\pi_i)_{0 \leq i \leq t}$ of partitions, i.e., $t \in \mathbb{N} \cup \{+\infty\}$, together with a corresponding sequence $(a_i)_{1 \leq i \leq t}$ of deviating agents, such that for every $1 \leq i \leq t$, it holds that $\pi_{i-1} \xrightarrow{a_i}_\chi \pi_i$, i.e., π_i evolves from π_{i-1} by a χ deviation of a_i . We say that an execution of the χ dynamics of (G, π_0) *converges* if π_t is a χ partition.

We say that the χ dynamics of (G, π_0) *possibly converges* if some execution of (G, π_0) converges. Moreover, we say that the χ dynamics of (G, π_0) *necessarily converges* if every execution of the χ dynamics of (G, π_0) is finite. This means that we necessarily reach a χ partition if we continue applying χ deviations. By contrast, if the χ dynamics of (G, π_0) does not converge necessarily, there have to be executions where the same partition is reached infinitely often. In this case, we say that the dynamics *cycles*.

As computational decision problems, possible and necessary convergence can be captured as follows.

POSSIBLE CONVERGENCE OF DYNAMICS (χ -PCD)

Input: A hedonic game G and a starting partition π_0 .

Question: Is there a sequence of χ deviations on G that results in a χ partition when starting from π_0 ?

NECESSARY CONVERGENCE OF DYNAMICS (χ -NCD)

Input: A hedonic game G and a starting partition π_0 .

Question: Is every sequence of χ deviations on G finite when starting from π_0 ?

Typically, we consider χ -PCD and χ -NCD for a specific class of hedonic games, such as ASHG.

We conclude with the simple observation that CIS dynamics necessarily converge. This follows immediately because we operate on a finite game and every CIS deviation increases the utilitarian welfare $\sum_{a \in N} u_a(\pi)$ (Aziz et al., 2013).

Observation 3. *Every execution of the CIS dynamics converges necessarily.*

3 Presentation of Meta Theorems

We now present our meta theorems. A proof sketch can be found in Section 4 and the full proof is provided in Appendix B. Our first theorem states that the existence of a cycling dynamics implies hardness of deciding about possible convergence of dynamics.

Theorem 4. *Let χ be a standard stability notion with $\chi \subsetneq NS$ and $CIS \subsetneq \chi$. Assume that there exists an ASHG, FHG, or MFHG G_χ and partition π_χ such that the χ dynamics of (G_χ, π_χ) must cycle. Then χ -PCD is NP-hard for the game class of G_χ (e.g., for ASHGs if G_χ is an ASHG).*

Moreover, if there exists an instance in which the dynamics can cycle but necessarily converge after the removal of a *singleton* coalition, we obtain hardness of deciding about necessary convergence of dynamics.

Theorem 5. *Let χ be a standard stability notion with $\chi \subseteq NS$ and $CIS \subseteq \chi$. Assume that there exists an ASHG, FHG, or MFHG G_χ and partition π_χ that contains a singleton coalition $\{a\} \in \pi_\chi$, such that the χ dynamics can cycle on (G_χ, π_χ) , but necessarily converge on $(G_\chi - a, \pi_\chi - a)$. Then χ -NCD is coNP-hard for the game class of G_χ .*

The precondition for the required game in Theorem 5 may seem intricate, but it is quite weak. For instance, it is satisfied whenever there exists a game in which the dynamics starting from the singleton partition can cycle. Indeed, in this case, one can obtain the desired game by iteratively removing agents until the dynamics from the singleton coalition necessarily converges. Then, the penultimate game in this procedure satisfies the prerequisites of Theorem 5. Moreover, both theorems hold whenever there exists an instance without a stable partition. In this case, the dynamics from any starting partition (e.g., the singleton partition) must cycle. We state the latter observation in the following corollary.

Corollary 6. *Let χ be a standard stability notion with $\chi \subseteq NS$ and $CIS \subseteq \chi$. Assume that there exists an ASHG, FHG, or MFHG G_χ without a χ partition. Then, χ -PCD is NP-hard and χ -NCD is coNP-hard for the game class of G_χ .*

We can directly apply our corollary for established stability notions of which it is known that instances without stable partitions exist. For instance, there exist ASHGs without an IS or CNS (and, therefore, no NS) partition (Bogomolnaia and Jackson, 2002, Example 5; Sung and Dimitrov, 2007, Example 2). Our meta theorems (Theorems 4 and 5) apply uniformly to all standard stability notions, including NS, IS, CNS, VIS, VOS, and SMS in ASHGs, FHGs, and MFHGs. All of these also follow from Theorem 7 below.

In fact, we now demonstrate the applicability of our meta theorems for any deviation concept between NS deviations and voting-based notions weaker than CIS deviations. More precisely, consider (q_{out}, q_{in}) -VS for any $q_{out}, q_{in} \in [0, 1]$. In case that $q_{out} = q_{in} = 1$, this is CIS, for which dynamics necessarily converge (Observation 3). In all other cases, we show that Theorems 4 and 5 can be applied for all three game classes. We thus obtain a dichotomy that separates CIS from other voting-based stability notions.

Theorem 7. *Let χ be a standard stability notion such that $\chi \subseteq NS$ and (q_{out}, q_{in}) -VS $\subseteq \chi$ for some $q_{out}, q_{in} \in [0, 1]$. Then, χ -PCD is NP-hard and χ -NCD is coNP-hard for ASHGs, FHGs, and MFHGs if $q_{out} < 1$ or $q_{in} < 1$.*

The full proof of Theorem 7 is presented in Appendix C. It relies on constructing two games for which we apply Theorems 4 and 5 once each. We further distinguish whether for the relevant stability notion χ it holds that $(q_{out}, 1)$ -VS $\subseteq \chi$ or $(1, q_{in})$ -VS $\subseteq \chi$. All constructed games consist of a large set of deviating agents and a small set of gadget agents that never perform deviations (and, in fact, their valuation function is the 0-function, under which all coalitions yield an identical utility). Starting from a predetermined partition, there always exists precisely one deviating agent that can perform a permissible χ deviation, while no other deviation is possible that is even an NS deviation. Performing this deviation yields a partition that is identical up to a permutation of agents. Hence, we establish inevitable cycling, and, therefore, games suitable to apply Theorem 4. The starting partitions can then be turned into partitions satisfying the preconditions of Theorem 5 by removing the first deviator from her coalition and placing her in a singleton coalition.

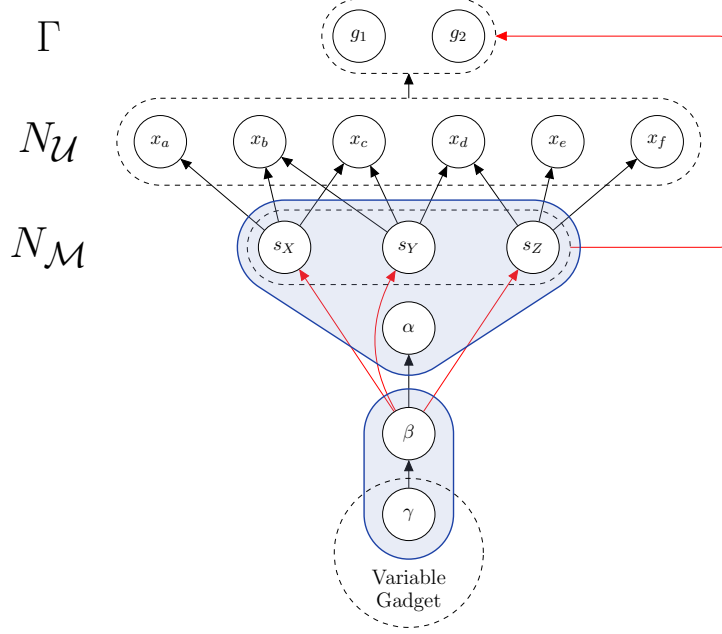


Figure 2: Illustration of the reduction. A covering instance $(\mathcal{U}, \mathcal{M})$ is represented by agents $N_{\mathcal{U}}$ and $N_{\mathcal{M}}$. Here, we have $\mathcal{U} = \{a, \dots, f\}$ and $\mathcal{M} = \{X, Y, Z\}$ with $X = \{a, b, c\}$, $Y = \{b, c, d\}$, and $Z = \{d, e, f\}$. Black and red arrows indicate potential utility increases and decreases, respectively. Important coalitions of the starting partition are indicated in blue. In Yes-instances, dynamics can lead to agent γ ending up in a singleton coalition.

4 Proof Sketch of Meta Theorems

In this section, we outline the proofs of Theorems 4 and 5. Both rely on a reduction from RESTRICTED EXACT COVER BY 3-SETS (RX3C). An instance of RX3C consists of a finite set of elements $\mathcal{U} = \{e_1, \dots, e_{3h}\}$ and a family $\mathcal{M} = \{M_1, \dots, M_{3h}\}$ subsets of $\mathcal{U}$ of size 3 such that every element of $\mathcal{U}$ belongs to exactly three sets in $\mathcal{M}$. An instance is a Yes-instance if and only if there is a selection of exactly h sets from $\mathcal{M}$ whose union is $\mathcal{U}$. RX3C is known to be NP-complete (Karp, 1972; Gonzalez, 1985).

Both proofs are performed in two steps: first, we encode the combinatorial structure of an RX3C instance as deviation dynamics, then we use the games assumed by the respective theorem as a gadget. The first step is the same for both theorems and is outlined in Figure 2. Given an instance $(\mathcal{U}, \mathcal{M})$ of RX3C, we introduce sets $N_{\mathcal{U}}$ and $N_{\mathcal{M}}$ of element agents and set agents representing $\mathcal{U}$ and $\mathcal{M}$, respectively. Set agents receive a positive utility from the element agents corresponding to their contained elements. At the top, there is a set Γ of grouping agents, identical in size to the number of sets in an exact cover, e.g., 2 agents if $|\mathcal{U}| = 6$. Further down, there are special agents α and β . The latter has a very high valuation for α but dislikes set agents. At the bottom, there is a variable gadget containing a dedicated agent γ who is the only agent that can interact with the other gadget agents through deviations.

In Figure 2, black arrows indicate deviation incentives, while red arrows represent deviation obstacles. The two important coalitions of the starting partition are indicated in blue. Generally, agents perform deviations “upwards.”

Element agents can freely join the coalitions of grouping agents, which can in principle lead to coalitions containing any grouping agent and any subset of element agents. However, set agents can only join the coalition of a grouping agent if it contains exactly the agents corresponding to its contained elements.² Once this happens, a coalition is created from and towards which no more deviations happen.

Over time, the coalition of α contains less and less set agents. This allows β to join this coalition if and only if the deviated set agents correspond to an exact cover of $\mathcal{U}$. This in turn allows the abandoned γ to engage in deviations within the gadget. In the deviation sequence up to this step, almost all performed deviations are CIS deviations and, therefore, χ deviations. The only deviation that is possibly not a CIS deviation is when β joins α . When performing this deviation, it is the only time in the proof that we use that we need a standard stability notion.

By specifying the variable gadget, we can leverage this general reduction to prove Theorems 4 and 5. For possible convergence, we use the game in which cycling dynamics must happen. For each of the coalitions of the starting partition causing necessary cycling, we append a copy of the construction in Figure 2. If the source instance was a No-instance, then agents of type γ (in the multiple copies) never end up in singleton coalitions. Hence, the gadget agents have to cycle inevitably. If, however, the source instance was a Yes-instance, then agents of type γ can join the coalitions from the gadget with CIS deviations, leading to a stable partition. Hence, dynamics possibly converge if and only if the source instance was a Yes-instance.

We now turn to necessary convergence. Note that **coNP**-hardness for necessary convergence is identical to **NP**-hardness of the question whether dynamics possibly cycle. We now use the possibly cycling game with its dedicated agent a as a variable gadget and identify γ with a . Hence, if the source instance was a No-instance, dynamics can never change the coalition of a , and, therefore, dynamics have to converge in the variable gadget. Otherwise, if the source instance was a Yes-instance, agent a can initiate cycling once she is in a singleton coalition.

5 Contractual Individual Stability

As CIS dynamics necessarily converge in any hedonic game (cf. Observation 3), CIS-PCD and CIS-NCD are trivially polynomial-time solvable. Moreover, Aziz et al. (2013) provide an algorithm to compute *some* CIS partition in polynomial time for ASHG_s.³ Unfortunately, their algorithm fails to produce partitions that satisfy individual rationality, i.e., some agents might have a large negative utility. Notably, as CIS deviations preserve individual rationality, CIS dynamics from the singleton coalition guarantee the existence of individually rational CIS partitions.

Observation 8. *Let G be a hedonic game together with an individually rational partition π_0 . Then, any execution of the CIS dynamics of (G, π_0) converges to an individually rational CIS partition.*

By contrast, it is **NP**-hard to decide whether CIS dynamics lead to individually rational outcomes, when starting from a general partition. This result holds even for fairly restricted valuations, e.g., to $\{-1, 1\}$. We defer all missing proofs in this section to Appendix D.

Theorem 9. *Let $f^+ : \mathbb{N} \rightarrow \mathbb{Q}^+$ and $f^- : \mathbb{N} \rightarrow \mathbb{Q}^-$ be two functions with $f^+(n) \geq |f^-(n)|$ for all $n \in \mathbb{N}$. It is **NP**-hard to decide whether the CIS dynamics in an ASHG, FHG, or MFHG can*

²Initially, coalitions of element agent contain an additional restricting agent that prevents set agents from joining. These are omitted from the figure for simplicity.

³Bullinger et al. (2025) correct an inaccuracy in this algorithm.

converge to an individually rational partition from a given starting partition π , even when valuations are restricted to $\{f^-(n), f^+(n)\}$ for games with n agents.

Hence, for each hedonic game, one can compute an individually rational CIS partition by running CIS dynamics from the singleton partition. However, it is not clear whether one can efficiently find a short converging sequence of CIS deviations, i.e., a sequence that consists of polynomially many steps. We, therefore, dedicate the remainder of this section to this question, and focus our attention on ASHG's.

First, we show that short converging sequences taking a linear number of CIS deviations always exist.

Theorem 10. *Let G be an ASHG and let π be a CIS partition that was reached through an execution of the CIS dynamics on G when starting from the singleton partition. Then π can be reached from the singleton partition after exactly $|N| - |\pi|$ CIS deviations.*

Proof. Consider an execution of the CIS dynamics on G when starting from the singleton partition. Our proof relies on the following claim which is proved in the appendix.

Claim 11. *Every coalition C in π contains exactly one agent that never deviated in the execution of the CIS dynamics.*

We denote the agents that never deviate to reach π as per Claim 11 as the *owners* of their respective coalitions in π . Moreover, given an arbitrary agent $a \in N$, we denote by o_a the owner of the coalition $\pi(a)$. Now, given the original (possibly exponential length) sequence of CIS deviations that resulted in π , consider the last deviation of each agent. We construct a new, shortened sequence of $|N| - |\pi|$ deviations, where each agent a that is not the owner of a coalition performs exactly one deviation from her singleton coalition to join o_a . We order this new deviation sequence by when the agents performed their last deviation in the original sequence. It is clear that this new deviation sequence results in the same partition π after exactly $|N| - |\pi|$ steps.

It remains to show that the new sequence consists only of CIS deviations. As each agent deviates from her singleton coalition, no agent will ever be blocked from leaving. Now, given a nonowner agent a , let C_{new} be the coalition that she joins in the new sequence and let C_{ori} be the coalition that she joins in the original sequence. Observe that $C_{\text{new}} \subseteq C_{\text{ori}}$ must hold. Then, a not being blocked from joining C_{new} directly follows from the fact that the original sequence consists only of CIS deviations. Further, in case there exists an agent $b \in C_{\text{ori}} \setminus C_{\text{new}}$ with $v_a(b) > 0$, then b must have deviated from a coalition that contains a in the original sequence, which cannot have been a CIS deviation. Hence, $v_a(b) \leq 0$, and thus $u_a(C_{\text{new}}) \geq u_a(C_{\text{ori}}) > 0$ must hold, where the strict inequality follows because C_{ori} was reached in a CIS dynamics starting from the singleton partition by a deviation of a . Therefore, the deviation of a is a CIS deviation. Since a was chosen arbitrarily, this concludes the proof. $\square$

An additional observation from the last theorem is that in the constructed dynamics every agent deviates at most once. However, finding this sequence needed knowledge of a possibly much longer sequence. This raises the question whether all CIS dynamics starting from the singleton partition are short. We answer this question negatively by constructing a family of instance where CIS dynamics can have exponential length with respect to the game size.

Theorem 12. *Let χ be a stability notion with $\text{CIS} \subsetneq \chi$. Then the χ dynamics starting from the singleton partition may take an exponential number of steps with respect to the game's input size.*

It remains an interesting open problem to determine the complexity of computing an individually rational CIS partition (even without using dynamics). We make first progress towards this question by identifying the structural reason behind Theorem 12. The games constructed in its proof heavily rely on valuations that are positive in one direction but 0 in the other. If we bound the number of agents with such valuations, we can efficiently compute individually rational CIS partitions. To this end, for an ASHG $G = (N, u)$, define $s(G) := |\{a \in N \mid \exists b \in N : v_a(b) > 0 \wedge v_b(a) = 0\}|$.

The proof idea is as follows. We construct the desired CIS dynamics in three phases. Define $X := \{a \in N \mid \exists b \in N : v_a(b) > 0 \wedge v_b(a) = 0\}$, i.e., $|X| = s(G)$. In the first phase, the agents not in X deviate. After at most one deviation each, a partition is reached in which these agents cannot deviate again. In the second phase, agents in X deviate at most once, joining best coalitions containing agents not in X . The first two phases comprise at most n deviations. In the third phase, arbitrary CIS deviations are performed. It can be shown that, after the second phase, deviations can only be performed by agents in X , joining other agents in X . Hence, this can lead to at most $s(G)^{s(G)}$ unique partitions.

Theorem 13. *An execution of the CIS dynamics starting from the singleton partition taking at most $s(G)^{s(G)} + n$ deviations can be computed in polynomial time with respect to the game's input size.*

6 Conclusion

We presented a meta approach to determine the computational complexity of deciding whether the deviation dynamics possibly or necessarily converge in a hedonic game based on the mere existence of simple No-instances. Our results encompass all standard stability notions based on deviations between NS and CIS deviations. Moreover, they hold for the prominent game classes of additively separable, fractional, and modified fractional hedonic games. We also investigated the computational complexity of finding an individually rational CIS partition in an ASHG. Here, dynamics may converge in a linear number of steps, but we can only efficiently extract the deviations for fast convergence when restricting the number of certain valuation pairs.

Natural directions for future work include reevaluating our hardness results for restricted domains of valuations, such as, utilities based on *friend-and-enemy* evaluations (Dimitrov et al., 2006), different classes of hedonic games, including ordinal models, or stability notions that rely on group deviations. Further, while Boehmer et al. (2023) discuss the structure of outcomes and running time of simulations for NS dynamics, an interesting direction would be a comprehensive experimental evaluation for a broader set of stability notions. Finally, an intriguing open question is the computational complexity of computing an individually rational CIS partition, and the applicability of our established results to game classes other than ASHGs.

Acknowledgments

Most of this work was done when Martin Bullinger was at the University of Oxford. Martin Bullinger was supported by the AI Programme of The Alan Turing Institute.

References

- Hernan Abeledo and Uriel G. Rothblum. Paths to marriage stability. *Discrete Applied Mathematics*, 63(1):1–12, 1995.
- Saba Ahmadi, Pranjal Awasthi, Samir Khuller, Matthäus Kleindessner, Jamie Morgenstern, Pat-tara Sukprasert, and Ali Vakilian. Individual preference stability for clustering. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, pages 197–246, 2022.
- José Alcalde and Pablo Revilla. Researching with whom? Stability and manipulation. *Journal of Mathematical Economics*, 40(8):869–887, 2004.
- Haris Aziz and Rahul Savani. Hedonic games. In Felix Brandt, Vincent Conitzer, Ulle Endriss, J. Lang, and Ariel D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 15. Cambridge University Press, 2016.
- Haris Aziz, Felix Brandt, and Hans Georg Seedig. Computing desirable partitions in additively separable hedonic games. *Artificial Intelligence*, 195:316–334, 2013.
- Haris Aziz, Florian Brandl, Felix Brandt, Paul Harrenstein, Martin Olsen, and Dominik Peters. Fractional hedonic games. *ACM Transactions on Economics and Computation*, 7(2):1–29, 2019.
- Suryapratim Banerjee, Hideo Konishi, and Tayfun Sönmez. Core in a simple coalition formation game. *Social Choice and Welfare*, 18:135–153, 2001.
- Vittorio Bilò, Angelo Fanelli, Michele Flammini, Gianpiero Monaco, and Luca Moscardelli. Nash stable outcomes in fractional hedonic games: Existence, efficiency and computation. *Journal of Artificial Intelligence Research*, 62:315–371, 2018.
- Niclas Boehmer and Edith Elkind. Individual-based stability in hedonic diversity games. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, pages 1822–1829, 2020.
- Niclas Boehmer, Martin Bullinger, and Anna M. Kerkmann. Causes of stability in dynamic coalition formation. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI)*, pages 5499–5506, 2023.
- Anna Bogomolnaia and Matthew O. Jackson. The stability of hedonic coalition structures. *Games and Economic Behavior*, 38(2):201–230, 2002.
- Felix Brandt and Anaëlle Wilczynski. On the convergence of swap dynamics to Pareto-optimal matchings. *Journal of Artificial Intelligence Research*, 80:1063–1098, 2024.
- Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.
- Felix Brandt, Martin Bullinger, and Anaëlle Wilczynski. Reaching individually stable coalition structures. *ACM Transactions on Economics and Computation*, 11(1–2):4:1–65, 2023.
- Felix Brandt, Martin Bullinger, and Leo Tappe. Stability based on single-agent deviations in additively separable hedonic games. *Artificial Intelligence*, 334:104160, 2024.

- Martin Bullinger and Stefan Kober. Loyalty in cardinal hedonic games. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 66–72, 2021.
- Martin Bullinger and Warut Suksompong. Topological distance games. *Theoretical Computer Science*, 981:114238, 2024.
- Martin Bullinger, Edith Elkind, and Jörg Rothe. Cooperative game theory. In Jörg Rothe, editor, *Economics and Computation: An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*, chapter 3, pages 139–229. Springer, 2024.
- Martin Bullinger, Adam Dunajski, Edith Elkind, and Matan Gilboa. Single-deviation stability in additively separable hedonic games with constrained coalition sizes. Technical report, <https://arxiv.org/abs/2510.12641>, 2025.
- Ioannis Caragiannis and Shivika Narang. Repeatedly matching items to agents fairly and efficiently. *Theoretical Computer Science*, 981:114246, 2024.
- Bugra Çaskurlu and Fatih Erdem Kizilkaya. On hedonic games with common ranking property. *Annals of Mathematics and Artificial Intelligence*, 92(3):581–599, 2024.
- Katarína Cechlárová and Antonio Romero-Medina. Stability in coalition formation games. *International Journal of Game Theory*, 29:487–494, 2001.
- Vincent Cohen-Addad, Silvio Lattanzi, Andreas Maggiori, and Nikos Parotsidis. Online and consistent correlation clustering. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, pages 4157–4179, 2022.
- Dinko Dimitrov, Peter Borm, Ruud Hendrickx, and Shao C. Sung. Simple priorities and core stability in hedonic games. *Social Choice and Welfare*, 26(2):421–433, 2006.
- Jacques H. Drèze and Joseph Greenberg. Hedonic coalitions: Optimality and stability. *Econometrica*, 48(4):987–1003, 1980.
- Edith Elkind, Svetlana Obraztsova, and Nicholas Teh. Temporal fairness in multiwinner voting. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence (AAAI)*, pages 22633–22640, 2024.
- Angelo Fanelli, Gianpiero Monaco, and Luca Moscardelli. Relaxed core stability in fractional hedonic games. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 182–188, 2021.
- Moran Feldman, Liane Lewin-Eytan, and Joseph Naor. Hedonic clustering games. *ACM Transactions on Parallel Computing (TOPC)*, 2(1):1–48, 2015.
- Martin Gairing and Rahul Savani. Computing stable outcomes in symmetric additively separable hedonic games. *Mathematics of Operations Research*, 44(3):1101–1121, 2019.
- Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985.

- Martin Hoefer, Daniel Vaz, and Lisa Wagner. Dynamics in matching and coalition formation games with structural constraints. *Artificial Intelligence*, 262:222–247, 2018.
- Ayumi Igarashi, Martin Lackner, Oliviero Nardi, and Arianna Novaro. Repeated fair allocation of indivisible items. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence (AAAI)*, pages 9781–9789, 2024.
- Richard M. Karp. Reducibility among combinatorial problems. In Raymond E. Miller and James W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972.
- Gianpiero Monaco, Luca Moscardelli, and Yllka Velaj. Stable outcomes in modified fractional hedonic games. *Auton. Agents Multi Agent Syst.*, 34(1):4, 2020.
- Martin Olsen. On defining and computing communities. In *Proceedings of the 18th Computing: The Australasian Theory Symposium (CATS)*, volume 128 of *Conferences in Research and Practice in Information Technology (CRPIT)*, pages 97–102, 2012.
- Dominik Peters. Graphical hedonic games of bounded treewidth. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, 2016.
- Dominik Peters and Edith Elkind. Simple causes of complexity in hedonic games. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 617–623, 2015.
- Debraj Ray and Rajiv Vohra. Coalition formation. In H. Peyton Young and Shmuel Zamir, editors, *Handbook of Game Theory with Economic Applications*, volume 4, chapter 5, pages 239–326. Elsevier, 2015.
- Walid Saad, Zhu Han, Tamer Basar, Mérouane Debbah, and Are Hjørungnes. Hedonic coalition formation for distributed task allocation among wireless agents. *IEEE Transactions on Mobile Computing*, 10(9):1327–1344, 2011.
- Shao C. Sung and Dinko Dimitrov. On myopic stability concepts for hedonic games. *Theory and Decision*, 62(1):31–45, 2007.
- Shao C. Sung and Dinko Dimitrov. Computational complexity in additive hedonic games. *European Journal of Operational Research*, 203(3):635–639, 2010.
- Gerhard J. Woeginger. A hardness result for core stability in additive hedonic games. *Mathematical Social Sciences*, 65(2):101–104, 2013.
- Valentin Zech, Niclas Boehmer, Edith Elkind, and Nicholas Teh. Multiwinner temporal voting with aversion to change. In *Proceedings of the 27th European Conference on Artificial Intelligence (ECAI)*, 2024.

Appendix

In the appendix, we provide additional material, such as missing proofs.

A Proof of Proposition 2

In this appendix, we provide the proof that voting-based stability notions are standard stability notions.

Proposition 2. *Let $q_{\text{out}}, q_{\text{in}} \in [0, 1]$. Then, $(q_{\text{out}}, q_{\text{in}})$ -VS is a standard stability notion.*

Proof. Let $q_{\text{out}}, q_{\text{in}} \in [0, 1]$. Consider an agent $a \in N$ and a single-agent deviation $\pi \xrightarrow{a} \pi'$. For $b \in N$, define $\Delta u_G(b, \pi, \pi') := u_b(\pi') - u_b(\pi)$, which only depends on $u_G(b, \pi, \pi')$.

Define $f : \mathcal{UC} \times \mathcal{UC} \times (\mathbb{Q} \times \mathbb{Q}) \rightarrow \{0, 1\}$ such that for all $X, Y \in \mathcal{UC}$, and $z \in \mathbb{Q} \times \mathbb{Q}$, we have that $f(X, Y, z) = 1$ if and only if

- $\Delta z > 0$,
- $|\{x \in X : \Delta x > 0\}| \geq q_{\text{out}}(|\{x \in X : \Delta x < 0\}| + |\{x \in X : \Delta x > 0\}|)$, and
- $|\{y \in Y : \Delta y > 0\}| \geq q_{\text{in}}(|\{y \in Y : \Delta y < 0\}| + |\{y \in Y : \Delta y > 0\}|)$.

We first show that f precisely encapsulates $(q_{\text{out}}, q_{\text{in}})$ -VS. Note that it holds that

- $F_{\text{out}}(\pi(a), a) = |\{x \in UC_G^{\text{out}}(a, \pi, \pi') : \Delta x > 0\}|$,
- $F_{\text{in}}(\pi(a), a) = |\{x \in UC_G^{\text{out}}(a, \pi, \pi') : \Delta x < 0\}|$,
- $F_{\text{out}}(\pi'(a), a) = |\{x \in UC_G^{\text{in}}(a, \pi, \pi') : \Delta x < 0\}|$, and
- $F_{\text{in}}(\pi'(a), a) = |\{x \in UC_G^{\text{in}}(a, \pi, \pi') : \Delta x > 0\}|$.

Hence, $\pi \xrightarrow{a} \pi'$ is a $(q_{\text{out}}, q_{\text{in}})$ -VS deviation if and only if $f(F_{\text{out}}(\pi(a), a), F_{\text{in}}(\pi(a), a), u_G(a, \pi, \pi')) = 1$. It follows that $(q_{\text{out}}, q_{\text{in}})$ -VS is an anonymously hedonic stability notion.

Moreover, consider $X, X', Y, Y' \in \mathcal{UC}$, and $z, z' \in \mathbb{Q} \times \mathbb{Q}$ such that $X \trianglelefteq X'$, $Y \trianglelefteq Y'$, and $\{z\} \trianglelefteq \{z'\}$. If $f(X, Y, z) = 0$, then $f(X, Y, z) \leq f(X', Y', z')$ is immediate. Assume, therefore, that $f(X, Y, z) = 1$. Since $\{z\} \trianglelefteq \{z'\}$, it holds that $\Delta z' \geq \Delta z > 0$. The second condition in the definition of f holds for X' if $q_{\text{out}} = 0$. If $q_{\text{out}} > 0$, then $\{x \in X : \Delta x < 0\} = \emptyset$ or there exists $\hat{x} \in \{x \in X : \Delta x > 0\}$. Hence, $\max\{\Delta x : x \in X\} \geq 0$. Since, $X \trianglelefteq X'$, it follows that $\min\{\Delta x' : x' \in X'\} \geq 0$, and therefore the second condition in the definition of f is satisfied for X' .

Finally, the third condition in the definition of f is satisfied for Y' by an analogous argument. We conclude that $f(X', Y', z') = 1$. Hence, f is monotone. $\square$

B Proof of Theorems 4 and 5

In this section, we will provide the full proof of Theorems 4 and 5. Both proofs use the same overall construction, which we will introduce first, and analyze in subsequent lemmas. The reduction is from EXACT COVER BY THREE SETS (X3C), which is defined as follows.

EXACT COVER BY 3-SETS (X3C)

Input: A finite set of elements $\mathcal{U}$ and a family $\mathcal{M}$ of subsets of $\mathcal{U}$ of size 3.

Question: Is there a selection of exactly $|\mathcal{U}|/3$ sets from $\mathcal{M}$ whose union is $\mathcal{U}$, i.e., is there an exact cover of $\mathcal{U}$ with sets from $\mathcal{M}$?

It is known that X3C is NP-complete (Karp, 1972). We use the following variant that assumes further restrictions on the structure of the set $\mathcal{M}$. This variation is known to remain NP-complete (Gonzalez, 1985).

RESTRICTED EXACT COVER BY 3-SETS (RX3C)

Input: A finite set of elements $\mathcal{U} = \{e_1, \dots, e_{3h}\}$ and a family $\mathcal{M} = \{M_1, \dots, M_{3h}\}$ of subsets of $\mathcal{U}$ of size 3 such that every element of $\mathcal{U}$ belongs to exactly three sets in $\mathcal{M}$.

Question: Is there a selection of exactly h sets from $\mathcal{M}$ whose union is $\mathcal{U}$, i.e., is there an exact cover of $\mathcal{U}$ with sets from $\mathcal{M}$?

Throughout the remaining section, we assume that χ is a standard stability notion such that $\chi \subseteq \text{NS}$ and $\text{CIS} \subseteq \chi$.

B.1 Reduction from RX3C

Consider an RX3C instance $\mathcal{I} = (\mathcal{U}, \mathcal{M})$, where $|\mathcal{U}| = 3h$. The reduction is illustrated in Figure 3. We construct an ASHG, FHG, or MFHG $G = (N, v)$ as follows. We define the set N of agents as $N_{\mathcal{M}} \cup N_{\mathcal{U}} \cup R \cup \Gamma \cup D \cup A \cup \{\alpha, \beta\}$, where

- $N_{\mathcal{M}} = \{s_M\}_{M \in \mathcal{M}}$ is a set of *set agents*,
- $N_{\mathcal{U}} = \{x_e\}_{e \in \mathcal{U}}$ is a set of *element agents*,
- $R = \{r_e\}_{e \in \mathcal{U}}$ is a set of *restricting agents*,
- $\Gamma = \{g_i\}_{i \in [h]}$ is a set of *grouping agents*,
- A is a set of *gadget agents* which contains a dedicated agent γ , and
- D is a (possibly empty) set of $\max(0, |A| - 2h)$ *dummy agents*.

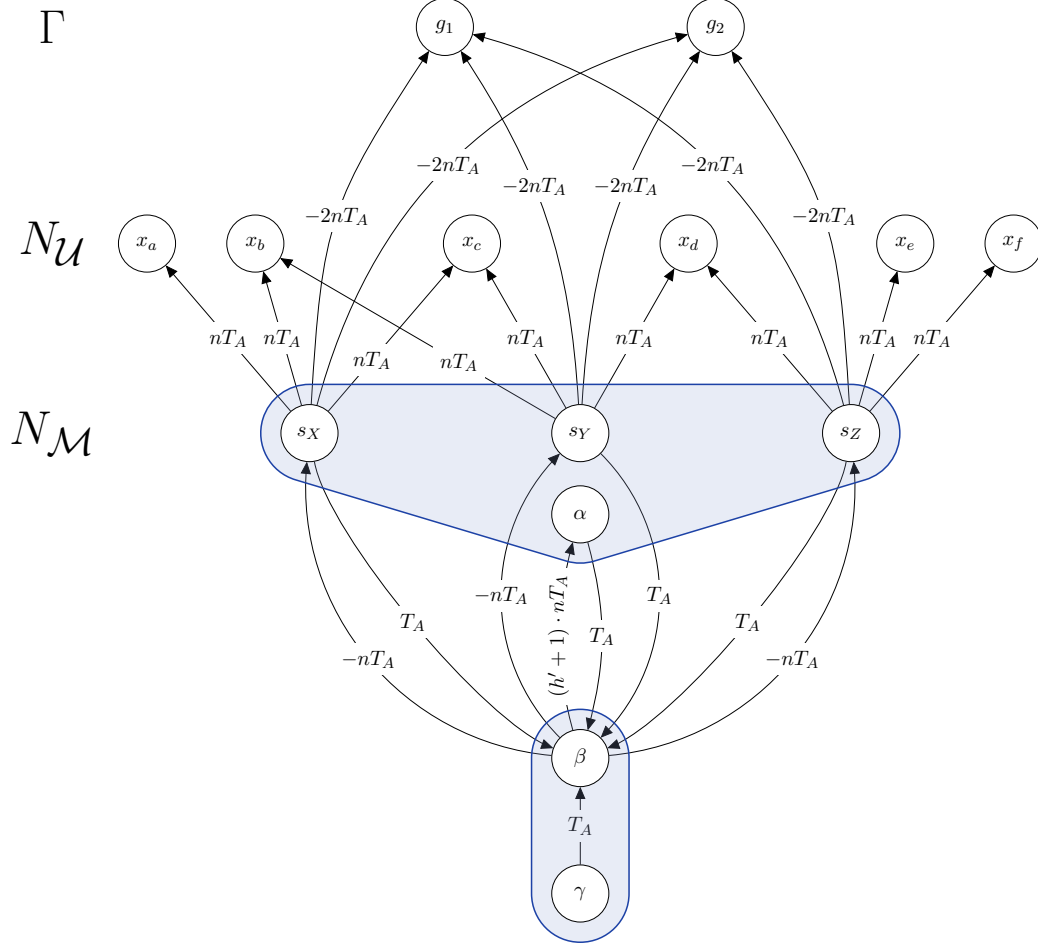
We assume that valuations among agents in A are already defined. Based on this, define

$$t_A := 1 + \sum_{p, d \in A} |v_p(d)| \quad \text{and} \quad T_A := nt_A.$$

There, n is, as usual, the number of agents.

We set $v_\gamma(\beta) = T_A$ and $v_\beta(\gamma) = 0$. Moreover, we define valuations among the agents in $N \setminus A$ as follows:

1. Let $v_\beta(\alpha) = (2h + 1) \cdot nT_A$, and $v_\beta(s_M) = -nT_A$ for every $M \in \mathcal{M}$.
2. Let $v_\alpha(\beta) = T_A$.
3. For every $d \in D$, let $v_d(\beta) = T_A$.
4. For every $M \in \mathcal{M}$, $e \in M$, $e' \in \mathcal{U} \setminus M$, and $g \in \Gamma$, let
 - (a) $v_{s_M}(x_e) = nT_A$,
 - (b) $v_{s_M}(r_e) = v_{s_M}(r_{e'}) = v_{s_M}(x_{e'}) = -nT_A$,



(a) Schematic of the overall construction, where we omit the agents in $R \cup D \cup (A \setminus \{\gamma\})$ and the corresponding valuations. Further, we omit all outgoing valuations of agents in $N_{\mathcal{U}}$, and the negative valuations of set agents for element agents that do not belong to the corresponding sets. Technically, $v_{\beta}(\alpha)$ is only well-defined for RX3C source instances, which we indicate by replacing $2h$ by h' . One can assume $h' = |\mathcal{M}| - |\mathcal{U}|/3$. We display the nonadapted instance, where the valuation between agents β and γ is not flipped.

Figure 3: Illustration of the reduction. For the sake of simplicity, the depicted reduction is from X3C instead of RX3C. However, the schematic is analogous apart from a small change to the valuation $v_{\beta}(\alpha)$. The reduced instance for the source instance $(\mathcal{U}, \mathcal{M})$ is displayed, where $\mathcal{U} = \{a, \dots, f\}$, and $\mathcal{M} = \{X, Y, Z\}$ with $X = \{a, b, c\}$, $Y = \{b, c, d\}$, and $Z = \{d, e, f\}$. A directed edge from agent p to agent d represents the valuation $v_p(d)$. Whenever two or more displayed agents belong to the same coalition in the starting partition π_0 , we indicate this by blue boxes.

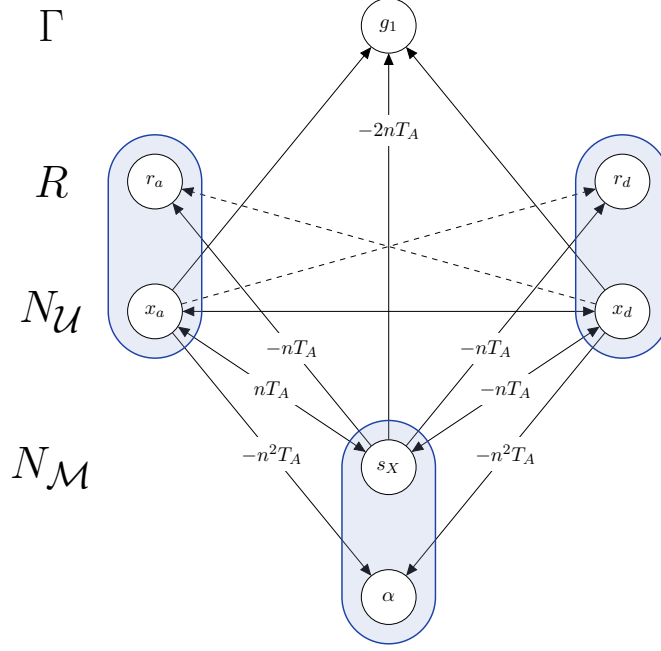
(c) $v_{s_M}(g) = -2nT_A$, and

(d) $v_{s_M}(\beta) = T_A$.

5. For every $e, e' \in \mathcal{U}$ with $e \neq e'$, $M, M' \in \mathcal{M}$ with $e \in M$ and $e \notin M'$, and $g \in \Gamma$, let

(a) $v_{x_e}(x_{e'}) = v_{x_e}(g) = 1$,

(b) $v_{x_e}(r_{e'}) = -1$,



(b) Schematic of the construction with focus only on the agents in $\{g_1, r_a, r_d, x_a, x_b, s_X, \alpha\}$, where $a \in X$ but $d \notin M$. Unlabeled edges represent a valuation of 1, and dashed edges represent a valuation of -1 .

Figure 3: Illustration of the reduction. For the sake of simplicity, the depicted reduction is from X3C instead of RX3C. However, the schematic is analogous apart from a small change to the valuation $v_\beta(\alpha)$. The reduced instance for the source instance $(\mathcal{U}, \mathcal{M})$ is displayed, where $\mathcal{U} = \{a, \dots, f\}$, and $\mathcal{M} = \{X, Y, Z\}$ with $X = \{a, b, c\}$, $Y = \{b, c, d\}$, and $Z = \{d, e, f\}$. A directed edge from agent p to agent d represents the valuation $v_p(d)$. Whenever two or more displayed agents belong to the same coalition in the starting partition π_0 , we indicate this by blue boxes.

(c) $v_{x_e}(\alpha) = -n^2T_A$,

(d) $v_{x_e}(s_M) = nT_A$, and

(e) $v_{x_e}(s_{M'}) = -nT_A$.

Based on the construction so far, we define

$$t_{N \setminus A} := 1 + \sum_{p, d \in N \setminus A} |v_p(d)| \quad \text{and} \quad T_{N \setminus A} := nt_{N \setminus A}.$$

We define valuations such that, with the exception of coalition $\{\beta, \gamma\}$, no two agents $(p, d) \in A \times (N \setminus A)$ can ever be part of a joint individually rational coalition, and we refer to these valuations by *sub-game restricting valuations*. Specifically, for all $(p, d) \in A \times (N \setminus A)$ with $(p, d) \neq (\beta, \gamma)$, let $v_p(d) = -T_A$, and $v_d(p) = -T_{N \setminus A}$.

Finally, we set all valuations that have not yet been specified to 0.

For the reader's convenience, valuations are often chosen much larger than necessary for our construction to work as intended, simplifying many of our following arguments.

We set the starting partition π_0 to

$$\begin{aligned} & \{\{g\} \mid g \in \Gamma\} \cup \{\{x_e, r_e\} \mid e \in \mathcal{U}\} \cup \\ & \{\{D \cup N_{\mathcal{M}} \cup \{\alpha\}\}, \{\beta, \gamma\}\} \cup \pi_A, \end{aligned} \quad (1)$$

where π_A is an arbitrary but fixed partition of the agents in $A \setminus \{\gamma\}$, whose exact composition is left open.

Let $\Lambda \subseteq N_{\mathcal{M}}$ be an arbitrary subset of exactly $2h$ set agents, and consider the coalitions $C = \Lambda \cup D \cup \{\alpha\}$ and $\{\beta, \gamma\}$. It will later become crucial whether β can perform a χ deviation from $\{\beta, \gamma\}$ to join a coalition of type C . First, note that the utility-change tuples that are relevant for whether β can deviate towards C do not depend on the composition of C (they are identical regardless of the agent set Λ). Hence, since we consider a standard stability notion, we can check whether such a deviation is permitted under our stability notion, by arbitrarily fixing Λ .

Now, note that a deviation of β leaving $\{\beta, \gamma\}$ and joining C is an NS deviation because β would increase her utility from 0 to nT_A . However, the deviation might not be a χ deviation, e.g., because other involved agents might block the deviation. We can check this by applying the polynomial-time computable function f_χ associated to χ for the utility-change tuples of this deviation. In case that β cannot perform a χ deviation to join C , we amend our construction by changing the valuations between β and γ and setting $v_\gamma(\beta) = 0$ and $v_\beta(\gamma) = T_A$ (we “reverse” the direction of the weighted edge between agents β and γ). This will ensure that the deviation of β would be a CIS deviation and hence a χ deviation.⁴

We remark that the reduced instance (G, π_0) can be constructed in polynomial time with respect to the source instance of RX3C, once the subgame induced by the agents in A has been fixed.

B.2 Investigation of Dynamics in the Reduced Instance

We will now in detail consider dynamics in the reduced instance. Throughout this section, we refer to G as the reduced game and π_0 the chosen starting partition. We first capture the behavior induced by the sub-game restricting valuations. Since these are sufficiently negative, no agent in A performing an NS deviation can ever join an agent in $N \setminus A$ and vice versa. We can use this to prove the following lemma.

Lemma 14. *In every execution of the χ dynamics of (G, π_0) , the only coalition ever containing an agent from A and an agent from $N \setminus A$ is $\{\beta, \gamma\}$.*

Proof. Assume for contradiction that agents $p \in A$ and $d \in N \setminus A$ where $(d, p) \neq (\beta, \gamma)$ end in a joint coalition at some point. Recall that $v_p(d) = -T_A$, and $v_d(p) = -T_{N \setminus A}$. Hence, when this happens for the first time, the agent of d and p that performs the deviation does not increase her utility. Thus the performed deviation is not an NS deviation. This is a contradiction as every χ deviation is an NS deviation. $\square$

The previous lemma implies that no agent can ever join the coalition $\{\beta, \gamma\}$. Hence, this coalition can only change if either β or γ performs a deviation herself. The next two lemmas reason about their deviations. This is where we make use of the potential adaptation of the constructed

⁴The exact reason for this case distinction will become apparent from the proof later on, see Lemmas 15 and 16. In particular, if we change the valuations, we will make use of the fact that β was not allowed to perform the deviation to join C before the alteration of valuations.

game. It is important to recall that, whenever agent β can perform a χ deviation in the original construction to leave $\{\beta, \gamma\}$ and join some coalition $C = \Lambda \cup D \cup \{\alpha\}$ with $\Lambda \subseteq N_{\mathcal{M}}$ that contains exactly $2h$ set agents, then we use this original construction. Otherwise, if this is not the case, then we adapt the original construction by “reversing” the direction of the weighted edge between β and γ . First, we consider deviations by β .

Lemma 15. *In every execution of the χ dynamics of (G, π_0) , agent β can deviate to leave a coalition that contains γ , to join a coalition $C = \Lambda \cup D \cup \{\alpha\}$, where $\Lambda \subseteq N_{\mathcal{M}}$ with $|\Lambda| = 2h$.*

Proof. By Lemma 14, we can assume that the abandoned coalition of the deviation of β is $\{\beta, \gamma\}$. If we did not adapt G , then β can perform the deviation by assumption.

Otherwise, we show that such a deviation of β is a CIS deviation and hence a χ deviation. To see this, observe that:

1. The deviation is an NS deviation because it holds that $u_\beta(\{\beta, \gamma\}) \leq T_A$ (achieved in case of an ASHG or MFHG) while $u_\gamma(C \cup \{\beta\}) \geq \frac{(2h+1) \cdot nT_A - 2hnT_A}{n-1} > T_A$ (observe that $|C \cup \{\beta\}| < n$).
2. Agent γ has a valuation of 0 for β , and is, therefore, indifferent between coalitions $\{\gamma\}$ and $\{\beta, \gamma\}$.
3. For each agent $p \in C$, it holds that $u_p(C) = 0$ while $u_p(C \cup \{\beta\}) \geq \frac{T_A}{n} > 0$. $\square$

Next, we consider deviations of γ abandoning β .

Lemma 16. *In every execution of the χ dynamics of (G, π_0) , agent γ cannot deviate to leave a coalition that contains β .*

Proof. Again, because of Lemma 14, for a deviation of γ away from β , we can assume that the abandoned coalition of the deviation of γ is $\{\beta, \gamma\}$, and the welcoming coalition is some $C_A \subseteq A$.

In case we made no adaptation to G , it holds that $u_\gamma(\{\beta, \gamma\}) \geq \frac{T_A}{2} > t_A$, and, by definition of t_A , agent γ can never improve her utility by deviating to coalition C_A .

Otherwise, we show the statement by virtue of χ being a standard stability notion and β not being able to deviate from $\{\beta, \gamma\}$ in the nonadapted construction to join a coalition $C = \{\Lambda \cup D \cup \{\alpha\}\}$ with $\Lambda \subseteq N_{\mathcal{M}}$ and $|\Lambda| = 2h$.

Let $\bar{G}$ and $\bar{G}'$ refer to the nonadapted and adapted game, respectively. Further, let $\bar{\pi}$ be a partition of N with $\{\{\beta, \gamma\}, C, C_A\} \subseteq \bar{\pi}$, and let $\bar{\pi}_{\rightarrow}^\beta$ and $\bar{\pi}_{\rightarrow}^\gamma$ be the partitions that resulted from $\bar{\pi}$ after β and γ performed their deviations in $\bar{G}$ and $\bar{G}'$ to join C and C_A , respectively. Let us compare the relevant utility-change multisets:

1. The utility-change multisets for the abandoned coalition $\{\beta, \gamma\}$ of β 's deviation in $\bar{G}$, namely, $X = UC_{\bar{G}}^{\text{out}}(\beta, \bar{\pi}, \bar{\pi}_{\rightarrow}^\beta)$, and γ 's deviation in $\bar{G}'$, namely $X' = UC_{\bar{G}'}^{\text{out}}(\gamma, \bar{\pi}, \bar{\pi}_{\rightarrow}^\gamma)$, are identical. Specifically, they are both $\{(T_A/\delta, 0)\}$ with $\delta \in \{1, 2\}$ (dependent on the game type). Hence X dominates X' .
2. Consider the utility-change multisets with respect to the two welcoming coalitions, namely, $Y = UC_{\bar{G}}^{\text{in}}(\beta, \bar{\pi}, \bar{\pi}_{\rightarrow}^\beta)$, and $Y' = UC_{\bar{G}'}^{\text{in}}(\gamma, \bar{\pi}, \bar{\pi}_{\rightarrow}^\gamma)$. Now, since $D \cup \Lambda \subseteq \bar{\pi}_{\rightarrow}^\gamma(y)$ with $|D \cup \Lambda| \geq |A| - 2h + 2h \geq |A|$, and $C_A \subseteq A$, it must hold that $|Y| \geq |Y'|$. Further, for every agent $p \in C$, it holds that $u_p(C) = 0$ in all game classes, while $u_p(C \cup \{\beta\}) \geq T_A/n \geq t_A$. On the other hand, any agent in C_A can have a valuation of at most t_A for agent γ , and hence only experience an increase of t_A by γ joining C_A . Thus, it must hold that Y dominates Y' .

3. Let $z = uc_{\bar{G}}(\beta, \bar{\pi}, \bar{\pi}_{\rightarrow}^{\beta}) = (0, u_{\beta}(C \cup \{\beta\}))$ with $u_{\beta}(C \cup \{\beta\}) \geq nT_A/n = T_A$, and let $z' = uc_{\bar{G}'}(\gamma, \bar{\pi}, \bar{\pi}_{\rightarrow}^{\gamma}) = (0, u_{\gamma}(C_A \cup \{\gamma\}))$ with $u_{\gamma}(C_A \cup \{\gamma\}) \leq t_A$. Then $\{z\}$ dominates $\{z'\}$.

But then, by the definition of a standard stability notion, $f_{\chi}(X, Y, z) \geq f_{\chi}(X', Y', z')$ must hold for the above-defined sets. Now, since β cannot deviate from $\{\beta, \gamma\}$ to join C in $\bar{G}$, i.e., $f_{\chi}(X, Y, Z) = 0$, agent γ cannot deviate from $\{\beta, \gamma\}$ in $\bar{G}'$, from which we can immediately follow the statement. $\square$

In the following lemma, we examine the χ dynamics of (G, π_0) in greater detail, where we fully characterize all coalitions that can result from the dynamics.

Lemma 17. *Let π' be a partition of N that resulted from (G, π_0) through an execution of the χ dynamics. Then each coalition in π' is of one of the following types:*

- I. $\{r_e\}$,
- II. $\{x_e, r_e\}$,
- III. $\{g\} \cup N'_{\mathcal{U}}$ for some $g \in \Gamma$, and $N'_{\mathcal{U}} \subseteq N_{\mathcal{U}}$,⁵
- IV. $\{g, x_e, x_f, x_g, s_M\}$ for some $g \in \Gamma$, $s_M \in N_{\mathcal{M}}$ with $M = \{e, f, g\}$,
- V. $\{\alpha\} \cup \Lambda \cup D$, where $\Lambda \subseteq N_{\mathcal{M}}$ with $|\Lambda| \geq 2h$,
- VI. $\{\alpha, \beta\} \cup \Lambda \cup D$, where $\Lambda \subseteq N_{\mathcal{M}}$ with $|\Lambda| = 2h$,
- VII. $\{\beta, \gamma\}$, and
- VIII. $C_A \subseteq A$.

Proof. We show the statement by induction over the number of deviations. First, the statement is true for the initial partition π_0 : Agents in Γ are in coalitions of Type III and the other agents are in coalitions of Type II, V, VII, or VIII.

Next, let π' be a partition of N that resulted from (G, π_0) through an execution of the χ dynamics, and assume that each coalition in π' is of one of the above types. Therefore, let π'' be a partition and let $p \in N$ be a deviator such that $\pi' \xrightarrow{p}_{\chi} \pi''$, i.e., we increase the execution of the dynamics by another deviation. We will show that both the abandoned coalition $\pi'(p) \setminus \{p\}$ and the welcoming coalition $\pi''(p)$ are of one of the above types.

Note that $p \notin \Gamma \cup R$, as these agents are indifferent over all coalitions they can possibly be in, and, thus, none of their deviations would be an NS and, therefore, χ deviation. Also note that there can only ever be one coalition of Types V and VI and one of Types VI and VII.

Now, assume that $p = x_e$ for some $e \in \mathcal{U}$. Then p cannot be part of a coalition of Type IV in π' , as the presence of the relevant set agent ensures that p cannot increase her utility by deviating. Hence, p is part of a coalition of Type II or III in π' , and has a utility of at least 0 in all game classes. Thus, the deviation must lead to a strictly positive utility for p . In particular, this implies that the welcoming coalition cannot be of Type I, II, VII, or VIII, as p has no positive valuation for any agent in these coalitions and, therefore, no overall positive utility. It can also not be of Type V or VI, as the presence of α ensures that p has strictly negative utility for such a coalition.

⁵Here, $N'_{\mathcal{U}} = \emptyset$ is explicitly allowed.

Next, it cannot be of Type IV, as such a coalition contains s_M for some $M \in \mathcal{M}$ together with all corresponding element agents. Hence, for the $e \in \mathcal{U}$ with $p = x_e$, it holds that $e \notin M$, and p would obtain a negative utility from joining the coalition of s_M . Finally, if the welcoming coalition is of Type III, then it is so after the deviation. This concludes the consideration of the case $p = x_e$.

In case $p = s_M$ for some $M \in \mathcal{M}$, then we claim that $\pi''(p)$ must be of Type IV. First, assume that $\pi'(p)$ is of Type IV. But then, p has a utility of at least $nT_A/n \geq T_A$ in partition π' , and it is clear that p cannot gain by deviating, as the only agent outside of $\pi'(p)$ that p has a positive valuation for is β , with $v_p(\beta) = T_A$. Hence, $\pi'(p)$ must be of Type V or VI, and, since there is only one coalition of Type V or VI, she cannot join a coalition of Type V or VI. Now, p cannot deviate to a coalition of Type I or II, as the presence of a restricting agent would lead to a negative utility. Also, she cannot deviate to a coalition of Type IV, as one of the present element agents would not correspond to the set M , and p can have a utility of at most 0 for any such coalition. Moreover, she cannot deviate to a coalition of Type VII or VIII due to Lemma 14. As we have excluded all other cases, the welcoming partition $\pi''(p) \setminus \{p\}$ must be of Type III. Further, it has to exactly contain those element agents corresponding to the set M with $p = s_M$. Otherwise, the negative valuation of s_M for agents in Γ would lead to p not performing an NS deviation. However, if $\pi'(p)$ contains at most $2h$ set agents, then each of the remaining set agents not contained in $\pi'(p)$ form coalitions of Type IV in π' . Hence, π' would not contain a coalition of Type III. Hence, $\pi'(p)$ must contain at least $2h + 1$ set agents and, therefore, be of Type V. But then, the abandoned coalition is still of Type V after the deviation, and the welcoming partition $\pi''(p)$ is of Type IV. This concludes the consideration of the case $p = s_M$.

Next, note that α can only ever have an incentive to deviate to join β . However, if β is in a coalition of Type VII, then α cannot deviate because of Lemma 14, and otherwise, β is already in the same coalition as α . It follows that α never deviates and, therefore, $p \neq \alpha$.

In case that $p = \beta$, the abandoned coalition cannot be of Type VI, as then, p would already have a utility of $nT_A/n \geq T_A$ in π' , while γ is the only outside agent that p *might* have a positive valuation for, with $v_p(\gamma) \leq T_A$. Thus, the abandoned coalition must be of Type VII and, after the deviation, $\pi''(p) \setminus \{p\} = \{\beta, \gamma\} \setminus \{\beta\} = \{\gamma\} \subseteq A$, which is of Type VIII. Moreover, deviating from $\{\beta, \gamma\}$, $p = \beta$ can only have incentive to join α , i.e., the welcoming coalition must be of Type V containing $\Lambda \subseteq N_{\mathcal{M}}$. However, if $|\Lambda| > 2h$ then $\sum_{d \in \pi''(p)} v_p(d) \leq (2h + 1) \cdot nT_A - (2h + 1) \cdot nT_A \leq 0$, and p does not have an incentive to deviate. In addition, all set agents outside of this coalition must be in a coalition of Type IV, and there are at most h such coalitions. Thus, we have $|\Lambda| \geq 2h$ and, therefore, $|\Lambda| = 2h$ must hold. We conclude that the welcoming coalition is of Type VI after the deviation.

Finally, let us consider the case where $p \in A$. If $p = \gamma$, then, by Lemma 16, $\pi'(p) \neq \{\beta, \gamma\}$. Hence, whether $p = \gamma$ or not, p must abandon a coalition of Type VIII, which remains such a coalition (if nonempty after the deviation). Also, β then is in a coalition of Type VI. By Lemma 14, p cannot join a coalition containing an agent outside of A . Therefore, the welcoming coalition must be of Type VIII as well. $\square$

Next, we show the defining behavior of the constructed instance, namely that whether agent β can deviate from $\{\beta, \gamma\}$ is directly corresponding to whether the source instance of RX3C is a Yes-instance.

Lemma 18. *There exists an execution of the χ dynamics of (G, π_0) where agent β can deviate from $\{\beta, \gamma\}$ if and only if the source instance $\mathcal{I} = (\mathcal{U}, \mathcal{M})$ of RX3C is a Yes-instance. Moreover,*

if $\mathcal{I}$ is a No-instance, then $\{\beta, \gamma\}$ is part of every occurring partition in every execution of the χ dynamics of (G, π_0) .

Proof. Note that, since $\{\beta, \gamma\} \in \pi_0$ and because of Lemma 14, no other agent can ever deviate to join $\{\beta, \gamma\}$. Additionally, because of Lemma 16, agent γ can never abandon $\{\beta, \gamma\}$. Thus, it suffices to show that β can deviate from $\{\beta, \gamma\}$ in some execution of the χ dynamics if and only if the source instance $\mathcal{I} = (\mathcal{U}, \mathcal{M})$ of RX3C is a Yes-instance.

($\Rightarrow$) Assume that β can deviate from $\{\beta, \gamma\}$. By Lemma 17, this deviation must result in a coalition of Type VI, i.e., a coalition $C = \{\alpha, \beta\} \cup \Lambda \cup D$ with $\Lambda \subseteq N_{\mathcal{M}}$ and $|\Lambda| = 2h$. But then, there must be exactly h set agents who left their initial coalition with α to join a coalition of Type IV. Since such a coalition consists of some grouping agent and all element agents that correspond to the relevant sets, it is easy to see that these set agents correspond to an exact cover of $\mathcal{U}$ with h sets from $\mathcal{M}$.

($\Leftarrow$) Assume that $\mathcal{I}$ is a Yes-instance, i.e., there is a set $\mathcal{M}' = \{M_1, \dots, M_h\}$ such that $\bigcup_{M \in \mathcal{M}'} M = \mathcal{U}$. Note that, since $\text{CIS} \subseteq \chi$, every CIS deviation also is a χ deviation. We thus provide a sequence of CIS deviations, followed by a single χ deviation of β to join the coalition of α . For all CIS deviations, we will argue that (i) it is an NS deviation, (ii) the favor-in set of the abandoned coalition is empty, and (iii) the favor-out set of the joined coalition is empty. The sequence of deviations can now be given as follows (unless otherwise specified, within each step, deviations are performed in an arbitrary order):

1. For each $i \in [h]$ and $e \in M_i$, agent x_e deviates to join the coalition of grouping agent g_i . Since $\mathcal{M}'$ is an exact cover of $\mathcal{U}$ of size h , no agent in $N_{\mathcal{U}}$ is asked to join two different agents in Γ . These are CIS deviations since:

- (i) agent x_e increases her utility from 0 to at least $1/2$,
- (ii) agent x_e leaves the coalition $\{x_e, r_e\}$, and $v_{r_e}(x_e) = 0$, and
- (iii) the coalition of g_i is a subset of $\{g_i\} \cup \{x_{e'}\}_{e' \in M_i}$, where g_i is indifferent of x_e 's deviation, and the agents $\{x_{e'}\}_{e' \in M_i}$, in case of an ASHG or FHG, are strictly in favor of the deviation, or indifferent in case of an MFHG.

2. For each $i \in [h]$, agent s_{M_i} joins the coalition of g_i . These are CIS deviations since:

- (i) agent s_{M_i} increases her utility from 0 to at least $(-2nT_A + 3nT_A)/5 \geq T_A$,
- (ii) agent s_{M_i} leaves a coalition that is a subset of $\{\alpha\} \cup N_{\mathcal{M}} \cup D$, where all these agents have a valuation of 0 for s_M and utility of 0 for the whole coalition and are hence indifferent to the deviation in all game classes, and
- (iii) the coalition of g_i is $\{g_i, x_e, x_f, x_g\}$ with $M_i = \{e, f, g\}$, where g_i is indifferent to the deviation, and the set agents strictly increase their utility from at most 3 to at least $nT_A/5 \geq T_A$.

3. Agent β performs a χ deviation to join the coalition of α , which, at this point in the dynamics, is of the form $\{\alpha\} \cup \Lambda \cup D$, where $\Lambda \subseteq N_{\mathcal{M}}$ with $|\Lambda| = 2h$. This is a χ deviation, as shown in Lemma 15. $\square$

Finally, we show that the χ dynamics of the subgame restricted to the agents in $N \setminus A$ must converge when starting from the partition π_0 .

Lemma 19. *In every execution of the χ dynamics of (G, π_0) , every agent in $N \setminus A$ can only perform a finite number of deviations.*

Proof. With the arguments given in Lemma 17, one can verify that no agent in $\Gamma \cup R \cup D \cup \{\alpha\}$ ever has an incentive to deviate. Similarly, agent β has no incentive to abandon a coalition of Type VI and can, therefore, deviate at most once. Moreover, a set agent in $N_{\mathcal{M}}$ does not have an incentive to deviate from a coalition of Type IV, and hence deviates at most once.

It thus suffices to consider deviations of element agents. Given a partition $\bar{\pi}$ of N , consider the following potential function:

$$\Phi(\bar{\pi}) := \sum_{g \in \Gamma} |\bar{\pi}(g)|^2.$$

Intuitively, Φ describes the sum of squared coalition sizes of all grouping agents. Due to Lemma 17, it is easy to see that for any $\bar{\pi}$ that resulted from π through the χ dynamics, the value $\Phi(\bar{\pi})$ is at most $|\Gamma| \cdot (5 - 1)^2 = 16h$.

We claim that each deviation of an element agent strictly increases the value of potential function Φ . As every other agent in $N \setminus A$ can perform at most one deviation, this immediately implies the statement.

Let π' be a partition that resulted from π through an execution of the χ dynamics, and let $x \in N_{\mathcal{U}}$ be an element agent that performs a deviation $\pi' \xrightarrow{\mathcal{F}_\chi} \pi''$ for some π'' .

If x deviates from a coalition of Type II, an increase of Φ is immediate. Next, a deviation of x from a coalition of Type IV would result in a forbidden coalition type, so it cannot happen. It remains to consider deviations of x from coalitions of Type III.

First, x has no incentive to deviate from a coalition of Type III to a coalition of Type I, as she receives strictly positive utility for any of the former and a utility of 0 for the latter. Further, as a coalition of Type IV cannot be formed through a deviation of x , this leaves a deviation of x from a coalition C of Type III to a different coalition of Type III. Note that x only increases her utility via such a deviation in case $|\pi'(x)| < |\pi''(x)|$. Let $i := |\pi'(x)|$, and let $j := |\pi''(x)|$. Then it holds that

$$\begin{aligned} & \Phi(\pi'') - \Phi(\pi') \\ &= (|\pi'(x) \setminus \{x\}|^2 + |\pi''(x)|^2) - (|\pi'(x)|^2 + |\pi''(x) \setminus \{x\}|^2) \\ &= ((i - 1)^2 + j^2) - (i^2 + (j - 1)^2) \\ &= 2 \cdot (j - i) > 0. \end{aligned}$$

This completes the proof. □

B.3 Hardness Results

We are now ready to leverage our reduction to prove our hardness results stated in Theorems 4 and 5. First, we prove the theorem for possible convergence by utilizing the reduced games from Appendix B.1 as gadgets in a larger construction.

Theorem 4. *Let χ be a standard stability notion with $\chi \subseteq NS$ and $CIS \subseteq \chi$. Assume that there exists an ASHG, FHG, or MFHG G_χ and partition π_χ such that the χ dynamics of (G_χ, π_χ) must cycle. Then χ -PCD is NP-hard for the game class of G_χ (e.g., for ASHGs if G_χ is an ASHG).*

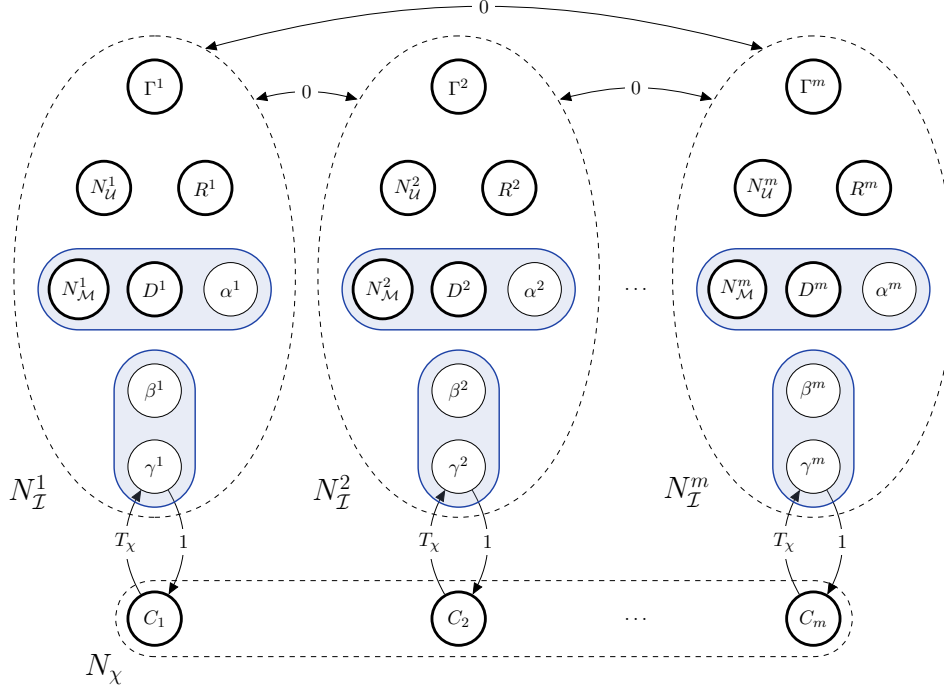


Figure 4: Illustration of the reduction for the proof of Theorem 4.

Proof. Consider an ASHG, FHG, or MFHG $G_\chi = (N_\chi, v_\chi)$ and a starting partition π_χ such that every execution of the χ dynamics of (G_χ, π_χ) cycles. Let $m := |\pi_\chi|$ be the number of coalitions in π_χ and assume that $\pi_\chi = \{C_1, \dots, C_m\}$. For our hardness reduction, we will utilize several copies of the game constructed in Appendix B.1. An illustration is provided in Figure 4. Specifically, we consider one copy for each coalition in π_χ . All coalitions introduce a disjoint set of agents except for the gadget agents that share N_χ and only have an individual dedicated agent.

Formally, consider an RX3C instance $\mathcal{I} = (\mathcal{U}, \mathcal{M})$. We construct a reduced game $G = (N, v)$ where $N = N_\chi \cup \bigcup_{i \in [m]} N_{\mathcal{I}}^i$. For $i \in [m]$, $N_{\mathcal{I}}^i = N_{\mathcal{M}}^i \cup N_{\mathcal{U}}^i \cup R^i \cup \Gamma^i \cup D^i \cup \{\gamma^i\}$ where all sets are chosen according to the reduced instance corresponding to $\mathcal{I}$ constructed in Appendix B.1.

Valuations among agents in N_χ are according to v_χ . Next, define $T_\chi := 1 + |N_\chi| \sum_{p, d \in N_\chi} |v_p(d)|$. For $i \in [m]$ and $p \in N_\chi$, we have

$$v_{\gamma^i}(p) = \begin{cases} 1 & p \in C_i \\ 0 & p \in N_\chi \setminus C_i \end{cases}$$

and

$$v_p(\gamma^i) = \begin{cases} T_\chi & p \in C_i, \\ 0 & p \in N_\chi \setminus C_i. \end{cases}$$

Moreover, among the agent set $N_{\mathcal{I}}^i \cup N_\chi$, we define valuations according to the construction in Appendix B.1, where we identify the variable subset of agents with $A^i = \{\gamma^i\} \cup N_\chi$. Note that these sets are not disjoint: we have $\bigcup_{i \in [m]} A^i = N_\chi$. However, the valuations are still well defined because we assume the given valuations among agents A^i as just defined. Finally, for $1 \leq i < j \leq m$, $p \in N_{\mathcal{I}}^i$, and $d \in N_{\mathcal{I}}^j$, we define $v_p(d) = 0$.

We set the starting partition π_0 to the union of the starting partition defined in Appendix B.1 for the agents in N^i and π_χ for the agents in N_χ , i.e.,

$$\begin{aligned} \pi_0 = & \pi_\chi \cup \bigcup_{i \in [m]} \{ \{g^i\} \mid g^i \in \Gamma^i \} \cup \{ \{x_e^i, r_e^i\} \mid e \in \mathcal{U} \} \\ & \cup \{ \{D^i \cup N_{\mathcal{M}}^i \cup \{\alpha^i\}\}, \{\beta^i, \gamma^i\} \}. \end{aligned}$$

For $i \in [m]$, the subgame of G induced by the agent set $N_{\mathcal{I}}^i \cup N_\chi$ is identical to the reduced instance constructed in Appendix B.1 and the starting partition restricted to this agent set is identical to the respective starting partition. Note that the partition π_0 is individually rational for all agents in $N_{\mathcal{I}}^i$ and, by Lemma 17, remains individually rational for these agents while deviations only happen among coalitions among $N_{\mathcal{I}}^i \cup N_\chi$. However, while coalitions for these agents are individually rational, they have no incentive to join the coalition of an agent in $N_{\mathcal{I}}^j$ for $j \in [m] \setminus \{i\}$. Hence, throughout any execution of the χ dynamics of (G, π_0) , agents in $N_{\mathcal{I}}^i$ and $N_{\mathcal{I}}^j$ will not form joint coalitions. Hence, we can apply all of our results from Appendix B.2, especially Lemmas 17 to 19, to the subgame induced by $N_{\mathcal{I}}^i \cup N_\chi$.

We now claim that the χ dynamics of (G, π_0) can converge if and only if the source instance $\mathcal{I}$ of RX3C is a Yes-instance.

($\Rightarrow$) Assume that $\mathcal{I}$ is a No-instance. Let $i \in [m]$. By Lemma 18, in any execution of the χ dynamics of (G, π_0) , $\{\gamma^i, \beta^i\}$ will remain a coalition throughout. But then, the agents in N_χ must remain in coalitions among themselves. As a consequence, the χ dynamics in the subgame (G_χ, π_χ) must cycle, and thus, must also cycle in (G, π_0) .

($\Leftarrow$) Assume that $\mathcal{I}$ is a Yes-instance. Then, as shown in Lemma 18, there exists a sequence of χ deviations that results in all agents in $\{\gamma^i \mid i \in [m]\}$ being in their respective singleton coalitions.

Next, for $i \in [m]$, we let each agent γ_i deviate to her corresponding coalition $C_i \in \pi_\chi$. These are CIS deviations (and thus χ deviations), since γ_i leaves her singleton coalition, has positive valuations for all agents in C_i (and hence strictly increases her utility), and since the utility of all agents in C_i strictly increases in all game classes based on the definition of T_χ . It is easy to see that, after these deviations, no agent in N_χ , nor an agent in $\{\gamma^i \mid i \in [m]\}$ can increase their utility by performing any further deviation. Further, by Lemma 19, each agent in $N_{\mathcal{I}}^i$ can only perform a finite number of deviations. Hence, the χ dynamics of (G, π_0) must converge subsequently. $\square$

Next, we prove our theorem for necessary convergence.

Theorem 5. *Let χ be a standard stability notion with $\chi \subseteq NS$ and $CIS \subseteq \chi$. Assume that there exists an ASHG, FHG, or MFHG G_χ and partition π_χ that contains a singleton coalition $\{a\} \in \pi_\chi$, such that the χ dynamics can cycle on (G_χ, π_χ) , but necessarily converge on $(G_\chi - a, \pi_\chi - a)$. Then χ -NCD is coNP-hard for the game class of G_χ .*

Proof. Assume that there exists a game $G_\chi = (N_\chi, v_\chi)$ and partition π_χ with the properties of the statement of the theorem. We show NP-hardness of the complement problem, i.e., whether it is possible for the χ dynamics to cycle.

Given an RX3C instance $\mathcal{I} = (\mathcal{U}, \mathcal{M})$, consider the constructed instance (G, π_0) from Appendix B.1. We now specify the subgame by setting $A = N_\chi$ and all valuations among the agents in N_χ are defined as in v_χ . Further, let the dedicated agent γ in A be a , where $\{a\} \in \pi_\chi$, and we set the subpartition π_A that was referenced in Equation (1) in Appendix B.1 to $\pi_\chi \setminus \{a\}$.

We claim that the χ dynamics of the constructed instance (G, π) can cycle if and only if the source instance $\mathcal{I}$ of RX3C is a Yes-instance.

($\Rightarrow$) Assume that $\mathcal{I}$ is a No-instance. Then, by Lemmas 14 and 18, agent $a = \gamma$ must remain in the coalition $\{\beta, a\}$. Now, again because of Lemma 14, all agents in $N_\chi \setminus \{a\}$ must remain in coalitions among themselves, and, by definition, the χ dynamics of the sub-game G_χ when starting on π must converge. Together with Lemma 19, this directly implies that the χ dynamics of (G, π) must converge.

($\Leftarrow$) Assume that $\mathcal{I}$ is a Yes-instance. Then, because of Lemma 18, in the χ dynamics of G , agent β can deviate to join α . But then, agent a is left in her singleton coalition, and the agents in N_χ are partitioned as in π_χ . Now, by definition of G_χ , the χ dynamics can cycle. $\square$

C Proof of Theorem 7

In this section, we present the full proof of Theorem 7 restated as follows.

Theorem 7. *Let χ be a standard stability notion such that $\chi \subseteq NS$ and $(q_{\text{out}}, q_{\text{in}})\text{-VS} \subseteq \chi$ for some $q_{\text{out}}, q_{\text{in}} \in [0, 1]$. Then, χ -PCD is NP-hard and χ -NCD is coNP-hard for ASHG, FHG, and MFHG if $q_{\text{out}} < 1$ or $q_{\text{in}} < 1$.*

We want to apply Theorems 4 and 5, and, therefore, have to construct games with the desired properties.

We further distinguish whether for the relevant stability notion χ it holds that $(q_{\text{out}}, 1)\text{-VS} \subseteq \chi$ or $(1, q_{\text{in}})\text{-VS} \subseteq \chi$. We start with the former and construct two games for the respective applications of Theorems 4 and 5.

Lemma 20. *Let χ be a stability notion such that $\chi \subseteq NS$ and $(q_{\text{out}}, 1)\text{-VS} \subseteq \chi$ for some $q_{\text{out}} \in [0, 1]$. Then, there exists an ASHG, FHG, and MFHG $G = (N, v)$, and partition π of N such that the χ dynamics on G must cycle when starting on π .*

Proof. Since $q_{\text{out}} < 1$, there exists $t' \in \mathbb{N}$ such that $t' \geq q_{\text{out}}(1 + t')$. Let $t_{\text{out}} := \min\{t' \in \mathbb{N} : t' \geq q_{\text{out}}(1 + t')\}$. In other words, t_{out} is the smallest number of agents required to favor some inside agent to leave so that this agent can leave if exactly one agent is against this deviation. Now, define

$$t := \max\{3, t_{\text{out}} + 1\} \quad \text{and} \quad T := t^2 - t.$$

We construct a game $G = (N, v)$ with agent set $N = A \cup \Gamma$, where $A = \{a_0, \dots, a_{3t+1}\}$ is a set of $m := 3t + 2$ *deviating agents*, and $\Gamma = \{g_1, g_2, g_3\}$ is a set of *grouping agents*. In this proof, we read indices modulo m mapping to the representative in $\{0, \dots, m-1\}$. For every integer $0 \leq i \leq 3t+1$, let

- $N_1^i := \{a_{i+j} \mid j \in \{2, \dots, t\}\},$
- $N_2^i := \{a_{i+j} \mid j \in \{t+1, \dots, 2t+1\}\},$ and
- $N_3^i := \{a_{i+j} \mid j \in \{2t+2, \dots, 3t+1\}\}.$

We will use these sets to define partitions. They essentially subdivide the deviating agents into three intervals of agents that encompass all deviating agents except for a_i and a_{i+1} . It holds that $|N_1^i| = t - 1$, $|N_2^i| = t + 1$, and $|N_3^i| = t$.

For $0 \leq i \leq 3t+1$, agent a_i has the following valuations:

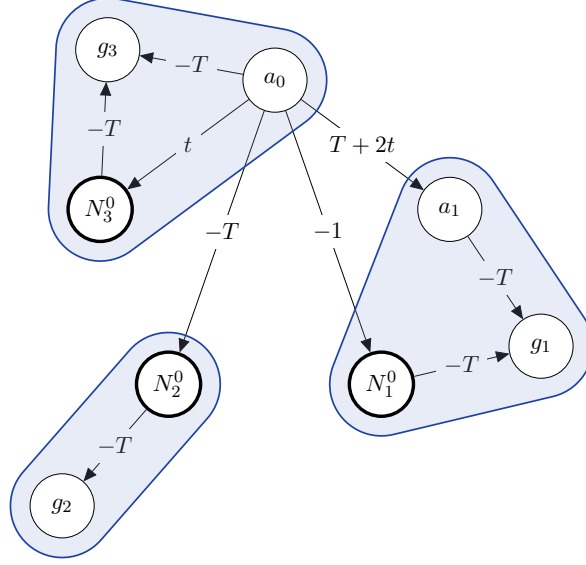


Figure 5: Illustration of our construction in the proof of Lemma 20. A bold node represents a group of agents. A weighted directed edge from an agent a to an agent b represents the valuation $v_a(b)$ and, in the case of a group of agents, applies to all group members individually. The initial partition π is indicated in blue. We omit all valuations of agents in $A \setminus \{a_0\}$ to other agents in A , as well as the valuations from agents in A to agents in Γ that are not part of the same coalition in π .

1. Let $v_{a_i}(a_{i+1}) = T + 2t$.
2. For each $a \in N_1^i$, let $v_{a_i}(a) = -1$.
3. For each $a \in N_2^i$, let $v_{a_i}(a) = -T$.
4. For each $a \in N_3^i$, let $v_{a_i}(a) = t$.
5. For each $g \in \Gamma$, let $v_{a_i}(g) = -T$.

All other valuations (i.e., the outgoing valuations of agents in Γ) are 0. We illustrate our construction in fig. 5.

For $0 \leq i \leq 3t+1$, define partition $\pi_i = \{C_1^i, C_2^i, C_3^i\}$ where $C_1^i = \{a_{i+1}, g_1\} \cup N_1^i$, $C_2^i = \{g_2\} \cup N_2^i$, and $C_3^i = \{a_i, g_3\} \cup N_3^i$. We claim that in π_i , agent a_i has a $(q_{\text{out}}, 1)$ -VS deviation to join C_1^i , whereas there exists no other NS deviation (including by a_i). Therefore, there exists a unique χ deviation in π_i as $\chi \subseteq \text{NS}$ and $(q_{\text{out}}, 1)\text{-VS} \subseteq \chi$.

Note that the deviation by a_i to join C_1^i results in the partition π_{i+2t+1} , which is identical in terms of valuations up to shifting indices by $2t+1$. Hence, proving our claim establishes that the dynamics starting from π_0 must cycle. Without loss of generality, we prove our claim for the case $i = 0$, and set $\pi := \pi_0$, and $C_j := C_j^i$ for $j \in [3]$.

First, we will show that a_0 can perform a $(q_{\text{out}}, 1)$ -VS deviation to join coalition C_1 . Observe that $\sum_{a \in C_3 \setminus \{a_0\}} v_{a_0}(a) = t \cdot t - T = t$, while $\sum_{a \in C_1} v_{a_0}(a) = T + 2t - T + (t-1) \cdot (-1) = t+1$. Hence, since $|C_3| = |C_1 \cup \{a_0\}|$, in all game classes, it holds that a_0 strictly prefers $C_1 \cup \{a_0\}$ over C_3 .

Now, let $\bar{N}_3^0 = N_3^0 \setminus \{a_{3t+1}\}$ and consider $a \in \bar{N}_3^0$. Then, $\sum_{b \in C_3 \setminus \{a\}} v_a(b) \geq T + t - T + (t - 1) \cdot (-1) = 1$, but $v_a(a_0) = -1$. Hence, in all game classes, it holds that a prefers $C_1 \setminus \{a_0\}$ over C_1 . Moreover, g_3 is indifferent over all coalitions, and only a_{3t+1} votes against the deviation of a_0 . Hence, by the choice of t , a_0 is allowed to abandon C_3 .

Next, let us consider the welcoming coalition C_1 . Observe that, again, g_1 is indifferent between all coalitions. Further, in the case of an ASHG and since all agents in $\{a_1\} \cup N_1^0$ have valuation t for a_0 , they will never object to a_0 joining. Otherwise, in case of an FHG or MFHG, given some $1 \leq i \leq t-1$, observe that $u_{a_i}(C_1) = \frac{T+2t-T+(i-1) \cdot t - (t-2-(i-1))}{\delta} = \frac{i \cdot (t+1) + 1}{\delta} \leq \frac{t \cdot (t+1)}{\delta}$, for some $\delta \in \{t+1, t+2\}$, and thus, $u_{a_i}(C_1) \leq t$. Then, since $v_{a_i}(a_0) = t$, no agent in $\{a_1\} \cup N_1^0 \setminus \{a_t\}$ objects to a_0 joining. Finally, it holds that $\sum_{b \in C_1 \setminus \{a_t\}} v_{a_t}(b) = -T + (t-1) \cdot t = 0$, and, as $v_{a_t}(a_0) = t$, agent a_t will also not object. Together, the deviation of a_0 is a $(q_{\text{out}}, 1)$ -VS deviation.

It remains to show that no other NS deviation is available. First, it is clear that a_0 cannot perform any different deviation, as she is individually rational in C_1 and has strictly negative utility for C_2 . Further, no grouping agent $g \in \Gamma$ can ever perform an NS deviation, as they have a utility of 0 for all coalitions.

Next, for an agent $a \in N_3^0$, it holds that $\sum_{b \in C_3 \setminus \{a\}} v_a(b) \geq T + 2t - T - t = t$. On the other hand, observe that $\sum_{b \in C_1} v_a(b) \leq 2 \cdot (-T) + (t-1) \cdot (-1) < 0$, and $\sum_{b \in C_2} v_a(b) \leq 2 \cdot (-T) + t \cdot t = -2t^2 + 2t + t^2 = 2t - t^2$, and, since $t \geq 3$, it follows that $\sum_{b \in C_2} v_a(b) < 0$. Thus, a has no incentive to deviate in any of the considered game classes.

For an agent $a \in N_1^0$, it holds that $\sum_{b \in C_1} v_a(b) \geq (t-1) \cdot t - T = 0$, and it holds that $\sum_{b \in C_1} v_{a_1}(b) = T + 2t - T + (t-1) \cdot (-1) \geq t$. By contrast, for any agent $a \in N_1^0 \cup \{a_1\}$, we have $\sum_{b \in C_3} v_a(b) \leq 2 \cdot (-T) + t \cdot t < 0$, using $t \geq 3$ once again, and $\sum_{b \in C_2} v_a(b) \leq 2 \cdot (-T) + T + 2t + t \cdot (-1) = t - T < 0$. Thus, no agent in C_1 has an incentive to deviate.

Finally, for the agents in N_2^0 , it holds that $\sum_{b \in C_2 \setminus \{a_{t+1}\}} v_{a_{t+1}}(b) = T + 2t - T + (t-1) \cdot (-1) \geq t$, and $\sum_{b \in C_2 \setminus \{a_{2t+1}\}} v_{a_{2t+1}}(b) = -T + t \cdot t = t$, with this value being at least as high for any other agent in N_2^0 . However, for $a \in N_2^0$, we have $\sum_{b \in C_3} v_a(b) \leq T + 2t + t \cdot (-1) + 2 \cdot (-T) = 2t - T = 3t - t^2 < 0$, using $t \geq 3$, and agent a has no NS deviation to join C_3 . Finally, it holds that $\sum_{b \in C_1} v_a(b) \leq t \cdot t - T = t$, and a has no incentive to deviate in all game classes, since $|C_1 \cup \{a\}| = |C_2|$.

This shows that no other NS deviation exists, which completes the proof. $\square$

The game constructed in the previous proof is constructed in a way that the first deviating agent could also start in a singleton coalition. Then, the first deviation is still a permissible χ deviation, and the game can still cycle. Moreover, if this agent is removed from the game, the χ dynamics can be shown to quickly converge. In this way, we can also construct the game required to apply Theorem 5.

Lemma 21. *Let χ be a stability notion such that $\chi \subseteq \text{NS}$ and $(q_{\text{out}}, 1)\text{-VS} \subseteq \chi$ for some $q_{\text{out}} \in [0, 1]$. Then, there exists an ASHG, FHG, and MFHG $G = (N, v)$, and partition π of N that contains a singleton coalition $\{a\} \in \pi_\chi$, such that the χ dynamics can cycle on (G_χ, π_χ) , but must converge on $(G_\chi - a, \pi_\chi - a)$.*

Proof. Consider the game $G = (N, v)$ constructed in the proof of Lemma 20. We alter π^0 to obtain a starting partition π' by placing agent a_0 in her singleton coalition while leaving all other agents untouched, i.e., we define $\pi' = \{\{a_0\}\} \cup \{C_1, C_2, C_3\}$ with $C_1 = \{a_1, g_1\} \cup N_1^0$, $C_2 = \{g_2\} \cup N_2^0$, and $C_3 = \{g_3\} \cup N_3^0$. We claim that the altered partition satisfies the assertion.

First, to see that the χ dynamics of (G, π') can cycle, observe that, in the first deviation of the χ dynamics of (G, π^0) , agent a_0 had a strictly positive utility for the welcoming coalition. Thus, a_0

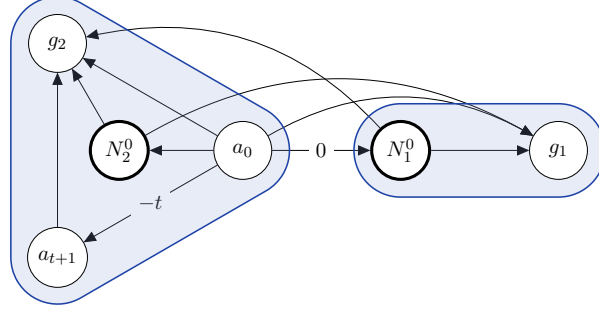


Figure 6: Illustration of our construction in the proof of Lemma 22. A bold node represents a group of agents. A weighted directed edge from an agent a to an agent b represents the valuation $v_a(b)$ and, in the case of a group of agents, applies to all group members individually. Unweighted edges represent a valuation of 1. The initial partition π is indicated in blue. We omit all valuations of agents in $A \setminus \{a_0\}$ to other agents in A .

can still increase her utility by deviating from her singleton coalition to join C_1 , and the deviation is furthermore still a $(q_{\text{out}}, 1)$ -VS deviation, as the support in coalition C_1 remains unchanged. But then, the resulting partition is identical to a partition in the cycling dynamics in the proof of Lemma 20 after the first deviation, and, going forward, the χ dynamics must cycle.

Next, we claim that the χ dynamics of $(G - a_0, \pi' - \{a_0\})$ must converge. First, one can verify that, from the initial state, only agent a_{2t+1} can perform an NS deviation to join C_3 . Afterwards, only agent a_t can perform an NS deviation to join C_2 . However, in the resulting partition, agent a_{3t+1} cannot deviate to join C_1 (as she could in the original construction), because the positive valuation for agent a_0 is missing as an incentive. In fact, based on the arguments given in the proof of Lemma 20, it is easy to see that no other NS deviation is possible at this stage. Hence, the NS dynamics, and, therefore, the χ dynamics converges. $\square$

We continue with constructing games for deviation concepts implying $(1, q_{\text{in}})$ -VS deviations. The constructed example is similar, and even simpler, as we now have deviating agents that only alternate between two coalitions.

Lemma 22. *Let χ be a stability notion such that $\chi \subseteq \text{NS}$ and $(1, q_{\text{in}})\text{-VS} \subseteq \chi$ for some $q_{\text{in}} \in [0, 1[$. Then, there exists an ASHG, FHG, and MFHG $G = (N, v)$, and partition π of N such that the χ dynamics on G must cycle when starting on π .*

Proof. We start by defining a threshold similar to the proof of Lemma 20. Since $q_{\text{in}} < 1$, there exists $t' \in \mathbb{N}$ such that $t' \geq q_{\text{in}}(1 + t')$. Let $t_{\text{in}} := \min\{t' \in \mathbb{N} : t' \geq q_{\text{in}}(1 + t')\}$. In other words, t_{in} is the smallest number of agents required to favor some outside agent to join a coalition so that this agent can join if exactly one agent of the coalition is against this deviation. We define

$$t := \max\{3, t_{\text{in}} + 1\}.$$

We construct a game $G = (N, v)$ with agent set $N = A \cup \Gamma$, where $A = \{a_0, \dots, a_{2t}\}$ is a set of $m := 2t + 1$ deviating agents, and $\Gamma = \{g_1, g_2\}$ is a set of grouping agents. As in the proof of Lemma 20, we read indices modulo m mapping to the representative in $\{0, \dots, m - 1\}$.

For every integer $0 \leq i \leq 2t$, define

- $N_1^i := \{a_{i+[n]j} \mid j \in \{1, \dots, t\}\}$ and
- $N_2^i := \{a_{i+[n]j} \mid j \in \{t+2, \dots, 2t\}\}.$

We will use these sets to define partitions. They essentially subdivide the deviating agents into two intervals of agents that encompass all deviating agents except for a_i and a_{i+t+1} . It holds that $|N_1^i| = t$ and $|N_2^i| = t - 1$.

For $0 \leq i \leq 2t$, agent a_i has the following valuations:

1. Let $v_{a_i}(a_{i+t+1}) = -t$.
2. For each $a \in N_2^i$, let $v_{a_i}(a) = 1$.
3. For each $g \in \Gamma$, let $v_{a_i}(g) = 1$.

All other valuations are 0. We remark that the two grouping agents are indifferent over all possible coalitions. The construction is illustrated in Figure 5.

For $0 \leq i \leq 2t$, define partition $\pi_i = \{C_1^i, C_2^i\}$ where $C_1^i = \{g_1\} \cup N_1^i$, and $C_2^i = \{a_i, a_{i+t+1}, g_2\} \cup N_2^i$. We claim that in π_i , agent a_i has a $(1, q_{\text{in}})$ -VS deviation to join C_1^i , whereas there exists no other NS deviation (including by a_i). Therefore, there exists a unique χ deviation in π_i as $\chi \subseteq \text{NS}$ and $(1, q_{\text{in}})$ -VS $\subseteq \chi$.

Note that the deviation by a_i to join C_1^i results in the partition π_{i+t} , which is identical in terms of valuations up to shifting indices by t . Hence, proving our claim establishes that the dynamics starting from π_0 must cycle. Without loss of generality, we prove our claim for the case $i = 0$, and set $\pi := \pi_0$, and $C_j := C_j^i$ for $j \in [2]$.

First, we show that a_0 has a $(1, q_{\text{in}})$ -VS deviation to join coalition C_1 . Indeed, observe that $\sum_{a \in C_2 \setminus \{a_0\}} v_{a_0}(a) = (t-1) + 1 - t = 0$, while $\sum_{a \in C_1} v_{a_0}(a) = t \cdot 0 + 1 = 1$, and agent a_0 strictly prefers C_1 over C_2 in all game classes.

Next, observe that all agents in $N_2^0 \cup \{a_{t+1}\}$ have a utility strictly greater than 0 for coalition C_2 , but a valuation of 0 for agent a_0 , and hence do not object to agent a_0 leaving.

Further, for $1 \leq i \leq t-1$, it holds that $v_{a_i}(a_0) = 1$. Hence, a_i is in favor of a_0 joining in the case of an ASHG. Otherwise, in case of an FHG or MFHG, we have $u_{a_i}(C_1) = \frac{i-1}{\delta} < \frac{i}{\delta+1} = u_{a_i}(C_1 \cup \{a_0\})$ with $\delta \in \{t, t+1\}$. It follows that a_i is also in favor of a_0 joining. Moreover, only a_t is against the deviation. Hence, by the choice of t , the deviation of a_0 is a $(1, q_{\text{in}})$ -VS deviation.

It remains to show that no other NS deviation is possible. First, as a_0 is in an individually rational coalition, she cannot deviate to her singleton coalition.

Moreover, for every $i \in [2t]$, agents a_i and a_{i+t+1} are in different coalitions. Thus, a_i is in an individually rational coalition, and cannot receive a positive utility by deviating to the other nonempty coalition. Therefore, a_i cannot perform an NS deviation. $\square$

Once again, we can modify the example but extracting the first deviating agent from the initial partition and place her in a singleton coalition to obtain a game suitable for applying Theorem 5.

Lemma 23. *Let χ be a stability notion such that $\chi \subseteq \text{NS}$ and $(1, q_{\text{in}})$ -VS $\subseteq \chi$ for some $q_{\text{in}} \in [0, 1[$. Then, there exists an ASHG, FHG, and MFHG $G = (N, v)$, and partition π of N that contains a singleton coalition $\{a\} \in \pi_\chi$, such that the χ dynamics can cycle on (G_χ, π_χ) , but must converge on $(G_\chi - a, \pi_\chi - a)$.*

Proof. Consider the game $G = (N, v)$ constructed in the proof of Lemma 22. We alter π^0 to obtain a starting partition π' by placing agent a_0 in her singleton coalition while leaving all other agents untouched, i.e., we define $\pi' = \{\{a_0\}\} \cup \{C_1, C_2\}$ with $C_1 = \{g_1\} \cup N_1^0$ and $C_2 = \{a_t + 1, g_2\} \cup N_2^0$. We claim that the altered partition satisfies the assertion.

First, to see that the χ dynamics on (G, π') can cycle, observe that $u_{a_0}(C_1 \cup \{a_0\}) > 0$. Now, by virtue of the first deviation of a_0 to C_1 being a $(1, q_{\text{in}})$ -VS deviation when a_0 was abandoning C_2 , this deviation is still a $(1, q_{\text{in}})$ -VS deviation when a_0 abandons her singleton coalition. Afterwards, due to the arguments given in the proof of Lemma 22, the χ -dynamics must cycle.

Next, on $(G - a_0, \pi' - a_0)$, observe that for every $i \in [2t] \setminus \{t\}$, agents a_i and a_{i+t+1} are still in different coalitions and, therefore, a_i has no NS deviation, as argued in the proof of Lemma 22. Further, for agent a_t , it holds that $\sum_{a \in C_1 \setminus \{a_t\}} v_{a_t}(a) = t \geq 3$, while $\sum_{a \in C_2} v_{a_t}(a) = 1$, while $|C_1| < |C_2 \cup \{a_t\}|$. Thus, a_t strictly prefers to stay in C_1 and has no NS deviation. Hence, $\pi' - a_0$ is an NS partition and, therefore, a χ partition on $G - a_0$, and the χ dynamics converge without a single deviation. $\square$

D Missing Proofs in Section 5

In this section, we present the missing proof concerning contractual individual stability.

D.1 Proof of Theorem 9

In this section, we present the full proof of Theorem 9. We provide a reduction from INDEPENDENT SET (INDSET), defined as follows.

INDEPENDENT SET (INDSET)

Input: An undirected graph $G = (V, E)$ and a positive integer k .

Question: Does G admit an independent set $I \subseteq V$ of size at least k , i.e., $|I| \geq k$ and $\{v, w\} \notin E$ for all $v, w \in I$?

It is well-known that INDSET is NP-complete (Karp, 1972).

Theorem 9. Let $f^+ : \mathbb{N} \rightarrow \mathbb{Q}^+$ and $f^- : \mathbb{N} \rightarrow \mathbb{Q}^-$ be two functions with $f^+(n) \geq |f^-(n)|$ for all $n \in \mathbb{N}$. It is NP-hard to decide whether the CIS dynamics in an ASHG, FHG, or MFHG can converge to an individually rational partition from a given starting partition π , even when valuations are restricted to $\{f^-(n), f^+(n)\}$ for games with n agents.

Proof. We reduce from INDSET. Let $(G = (V, E), k)$ be an instance of INDSET, and without loss of generality, assume that $k \geq 3$. We construct a game $G = (N, v)$ with a set $N = N_V \cup R_V \cup D \cup \{g^+, g^-, x, y, p\}$ of agents. There, $N_V = \{a_v\}_{v \in V}$ is a set of *vertex agents*, $R_V = \{r_v^+, r_v^-\}_{v \in V}$ is a set of *restricting vertex agents*, and $D = \{d_i\}_{i \in [k-1]}$ is a set of *dummy agents*. We define valuations as follows:

1. Let $v_y(a) = f^+(n)$ for all $a \in D \cup N_V \cup \{x, g^+\}$.
2. Let $v_d(d') = v_d(p) = v_p(d) = f^+(n)$ for all $d, d' \in D$ with $d \neq d'$.
3. Let $v_{g^+}(a) = v_{g^-}(a) = f^+(n)$ for all $a \in N_V \cup \{y\}$, and let $v_{g^+}(g^-) = v_{g^-}(g^+) = f^+(n)$.

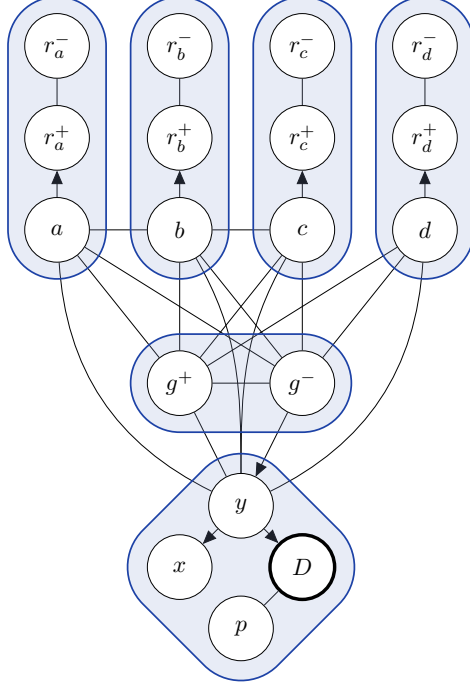


Figure 7: Illustration of reduction in Theorem 9. The reduced instance for the source instance $(\mathcal{G} = (V, E), k)$ is displayed, where $V = \{a, b, c, d\}$, $E = \{\{a, c\}, \{a, d\}, \{b, d\}, \{c, d\}\}$, and $k = 2$. An edge from agent a to agent b represents the valuation $v_a(b) = f^+(n)$. We note that the bold node D represents the whole set D of agents, and we omit the positive valuations among all agents in D . All other omitted edges represent the valuation $f^-(n)$. The initial coalitions according to π are displayed by blue boxes.

4. Let $v_{a_v}(a_w) = f^+(n)$ for all $v, w \in V$ with $v \neq w$ and $\{v, w\} \notin E$.
5. Let $v_{a_v}(r_v^+) = v_{r_v^+}(r_v^-) = v_{r_v^-}(r_v^+) = f^+(n)$.
6. Let all other valuations be $f^-(n)$.

We set the initial partition $\pi = \{\{x, y, p\} \cup D, \{\{a_v, r_v^+, r_v^-\} \mid v \in V\}, \{g^+, g^-\}\}$, i.e., we place agents g^+ and g^- together, agents x, y , and p with D , and all vertex agents with their corresponding restricting vertex agents. We will sometimes refer to the coalition that contains agents g^+ and g^- as the *grouping coalition* Γ . We depict the schematic of our construction in Figure 7.⁶

Note that π is neither a CIS partition, as any vertex candidate can deviate to join g^+ , nor individually rational, as, e.g., x would rather be in her singleton coalition.

Further, since G only contains one negative and one positive valuation value, we can make the following observation.

Observation 24. *Given an agent $a \in N$, a coalition C , and an agent $c \in C \setminus \{a\}$, for the game classes ASHG and FHG, it holds that $c \in F_{\text{in}}(C, a)$ if $v_c(a) = f^+(n)$, and $c \in F_{\text{out}}(C, a)$ otherwise.*

⁶Note that our reduction assumes that $k \geq 3$, but we chose an example with $k = 2$ to avoid overloading the illustration.

For the game class MFHG, it holds that $c \in F_{\text{in}}(C, a)$ if $v_c(a) = f^+(n)$ and $\exists c' \in C : v_c(c') = f^-(n)$ or $C \setminus \{a\} = \{c\}$, and $c \in F_{\text{out}}(C, a)$ if $v_c(a) = f^-(n)$ and $\exists c' \in C : v_c(c') = f^+(n)$ or $C \setminus \{a\} = \{c\}$.

We now begin our in-depth analysis of the CIS dynamics of the constructed instance (G, π) by proving several claims.

Claim 25. *No agent in $D \cup \{x, p\}$ can perform a CIS deviation abandoning coalition $D \cup \{x, y, p\}$.*

Proof. Observe that agent y has valuation $v_y(a) = f^+(n)$ for all agents in $a \in D \cup \{x\}$, and $v_y(p) = f^-(n)$. Hence, by Observation 24, it holds that $y \in F_{\text{in}}(D \cup \{x, y, p\}, a)$, and no such agent a can perform a CIS deviation from $D \cup \{x, y, p\}$. Further, since $k \geq 3$, there exists an agent $d \in D$. Using Observation 24 again, and since $v_d(p) = f^+(n)$ and $v_d(y) = f^-(n)$, one can see that p cannot deviate. $\triangleleft$

Claim 26. *Let $v \in V$. Then no restricting vertex agent $r_v \in \{r_v^+, r_v^-\}$ can deviate from her initial coalition $\{a_v, r_v^+, r_v^-\}$, or from the coalition $\{r_v^+, r_v^-\}$.*

Proof. Observe that both r_v^+ and r_v^- get nonnegative utility from both coalitions $\{a_v, r_v^+, r_v^-\}$ and $\{r_v^+, r_v^-\}$, since $v_{r_v^+}(r_v^-) = v_{r_v^-}(r_v^+) = f^+(n)$, and $f^+(n) \geq |f^-(n)|$. Hence, neither agent can deviate to her singleton coalition. Further, since both agents r_v^+ and r_v^- have valuation $f^-(n)$ for all agents in $N \setminus \{r_v^+, r_v^-\}$, any deviation to another nonempty coalition would strictly decrease their utility. $\triangleleft$

Claim 27. *A vertex agent $a_v \in N_V$ can CIS-deviate from her initial coalition $\{a_v, r_v^+, r_v^-\}$ to join a nonempty coalition $C \subseteq \{g^+, g^-\} \cup N_V \cup \{y\}$ if and only if there is no $a_w \in N_V \cap C$ with $\{v, w\} \in E$. Further, a_v cannot deviate to her singleton coalition, a coalition $\{r_w^-, r_w^+\}$ for some $w \in V \setminus \{v\}$, or any coalition $C' \in \pi \setminus \{\{a_v, r_v^+, r_v^-\}\}$.*

Proof. To see that a_v can deviate to such a coalition C that contains no such agent a_w , observe that, by Observation 24, and since $v_c(a_v) = f^+(n)$ for all $c \in C$, it holds that $F_{\text{out}}(C, a_v) = \emptyset$. Further, since a_v has positive valuation for all (at least one) agents in C , but dislikes r_v^- in her original coalition, she strictly improves her utility in all game classes.

Second, assume that C contains such an agent a_w . Then, since $v_{a_w}(a_v) = f^-(n)$, again by Observation 24, agent a_v cannot join C . Finally, to see that a_v cannot deviate to her singleton coalition, observe that, since $f^+(n) \geq |f^-(n)|$, agent a_v has positive utility for her initial coalition $\{a_v, r_v^+, r_v^-\}$. Further, a_v cannot join coalition $D \cup \{x, y, p\}$, since, e.g., $v_p(a_v) = f^-(n)$, and a_v cannot join $\{r_w^+, r_w^-\}$ or $\{a_w, r_w^+, r_w^-\}$ with $w \in V \setminus \{v\}$ by Observation 24 since, e.g., $v_{r_w^+}(a_v) = f^-(n)$ but $v_{r_w^+}(r_w^-) = f^+(n)$. $\triangleleft$

Claim 28. *Let $N'_V \subseteq N_V$ be a subset of $m \in \mathbb{N}$ vertex agents, and let $\tilde{\pi}$ be the partition that resulted from π after all agents in N'_V joined the grouping coalition Γ . Then, y can deviate to join Γ if and only if $m \geq k$, and otherwise, y cannot deviate to join any coalition in $\emptyset \cup \tilde{\pi} \setminus \{\tilde{\pi}(y)\}$.*

Proof. Due to Observation 24, and because y is “disliked” by all agents in $\tilde{\pi}(y) = \pi(y)$, and “liked” by all agents in $\tilde{\pi}(g^+)$, no agent will ever object to y performing such a deviation. Hence, we will focus our argument on whether y can improve her utility.

Observe that $\sum_{c \in \tilde{\pi}(y)} v_y(c) = k \cdot f^+(n) - f^-(n)$, and $\sum_{c \in \tilde{\pi}(g^+)} v_y(c) = (m+1) \cdot f^+(n) - f^-(n)$. Then, for the game class ASHG, the statement is immediate. Otherwise, for the game classes FHG

and MFHG, it holds that $u_y(\tilde{\pi}(y)) = \frac{k \cdot f^+(n) - f^-(n)}{k+s}$ and $u_y(\tilde{\pi}(g^+) \cup \{y\}) = \frac{(m+1) \cdot f^+(n) - f^-(n)}{m+s+1}$ with $s = 2$ and $s = 1$, respectively. Then, it holds that:

$$\begin{aligned}
& u_y(\tilde{\pi}(g^+) \cup \{y\}) - u_y(\tilde{\pi}(y)) \\
&= \frac{(m+1) \cdot f^+(n) - f^-(n)}{m+s+1} - \frac{k \cdot f^+(n) - f^-(n)}{k+s} \\
&= \frac{(m+1) \cdot (s \cdot f^+(n) + f^-(n)) - k \cdot (s \cdot f^+(n) + f^-(n))}{(m+s+1) \cdot (k+s)} \\
&= (m+1-k) \cdot \frac{s \cdot f^+(n) + f^-(n)}{(m+s+1) \cdot (k+s)}.
\end{aligned}$$

Now, since $s, k > 0$, $m \geq 0$, and $f^+(n) \geq |f^-(n)|$, we have $\frac{s \cdot f^+(n) + f^-(n)}{(m+s+1) \cdot (k+s)} > 0$, and hence, $u_y(\tilde{\pi}(g^+) \cup \{y\}) > u_y(\tilde{\pi}(y))$ if and only if $m \geq k$.

Now, to see that y cannot deviate to her singleton coalition, observe that, since $\sum_{c \in \tilde{\pi}(y)} v_y(c) = k \cdot f^+(n) - f^-(n) > 0$, it holds that $u_y(\pi(y)) > 0$ for all game classes. Finally, by Observation 24, y cannot deviate to join a coalition $\{r_v^+, r_v^-\}$ or $\{a_v, r_v^+, r_v^-\}$ for some $v \in V$, since, e.g., $v_{r_v^+}(y) = f^-(n)$ but $v_{r_v^+}(r_v^-) = f^+(n)$. $\triangleleft$

We claim that there is a sequence of CIS deviations on (G, π) that leads to an individually rational CIS partition if and only if there is a size- k independent set in $\mathcal{G}$.

($\Rightarrow$) Assume that there is a sequence of CIS deviations on (G, π) that leads to an individually rational CIS partition. Observe that no agent in $\{g^+, g^-\}$ has ever has incentive to deviate, as they both like each other in their original coalition, but, due to the behaviour described in the above claims, the agents in $N_V \cup \{y\}$ (i.e., the only other agents that are liked by g^+ and g^-) can only deviate to join $\{g^+, g^-\}$. Now, because x is not individually rational in the original coalition, and due to Claims 25 and 26, it is easy to see that y needs to deviate away to join the grouping coalition. However, as shown in Claim 28, this can only happen if there is a size- k independent set in $\mathcal{G}$.

($\Leftarrow$) Assume that there is a size- k independent set $I \subseteq V$. Then, by Claim 27, all agents in $\{a_v \in N_V\}_{v \in I}$ can deviate to join G . Afterwards, due to Claim 28 and since $|I| = k$, agent y can deviate to join Γ . Finally, let x deviate from $D \cup \{x, p\}$ to her singleton coalition, which is a CIS deviation, since $v_x(c) = v_c(x) = f^-(n)$ for all $c \in D \cup \{p\}$.

Now, to see all agents have nonnegative utility in the resulting coalition, observe that all agents in $D \cup \{p\}$ like each other, all restricting agents in a coalition $\{r_v^+, r_v^-\}$ (i.e., $v \in I$) like each other, and all agents in a coalition $\{a_v, r_v^+, r_v^-\}$ (i.e. $v \notin I$) have positive utility for their coalitions, since $f^+(n) \geq f^-(n)$. Further, the statement trivially holds for agents g^+, g^- , and for all agents that have deviated. Now, due to Observations 3 and 8, it is clear that the CIS dynamics must converge to an individually rational CIS partition. $\square$

D.2 Proof of Claim 11

Claim 11. *Every coalition C in π contains exactly one agent that never deviated in the execution of the CIS dynamics.*

Proof. We prove the statement by induction over the number of deviations. Clearly, the statement holds for the base case of the singleton partition. Now, assume that the statement holds for all

executions of the CIS dynamics of length k . Then, in case the next deviation is performed by an agent that has previously deviated, it cannot be to her singleton coalition, as each CIS deviation must strictly increase the deviator's utility, and weakly increase that of everyone else. Hence, the assertion follows after the $(k + 1)$ st deviation.

Otherwise, in case the subsequent deviation is performed by an agent that has never deviated, observe that such an agent can only deviate from her singleton coalition. Indeed, it is easy to see that as long as such an agent is in a coalition that contains an agent that CIS-deviated to join her, she is prohibited from leaving via a CIS deviation. But then, she must join a different, nonempty coalition, which, by the induction hypothesis, contains exactly one agent that never deviated. Then, the abandoned coalition is empty, and the welcoming coalition is joined by an additional agent who now performed a deviation. Hence, the statement holds for all CIS deviation sequences of length $k + 1$. $\square$

D.3 Proof of Theorem 12

In this section, we present the full proof of Theorem 12 restated as follows.

Theorem 12. *Let χ be a stability notion with $\text{CIS} \subsetneq \chi$. Then the χ dynamics starting from the singleton partition may take an exponential number of steps with respect to the game's input size.*

Proof. Let $k \in \mathbb{N}$ be a natural number with $k \geq 5$. We will construct an ASHG G such that the size of G is polynomially bounded in k , whereas, in the χ dynamics on G when starting from the singleton partition, there exists a sequence of at least 2^{k-1} deviations. Let $G = (N, v)$ be a game with a set $N = B \cup T$ of agents, where $B = \bigcup_{i \in [k]} B_i$ with $B_i = \{b_i^j\}_{j \in [i]}$ for each $i \in [k]$ is a set of *counting agents*, and $T = \{t_i\}_{i \in [k+1]}$ is a set of *track agents*.

Let $f : \mathbb{N} \rightarrow \mathbb{N}$ be a function with $f(x) = x^{kx}$. We define valuations as follows:

1. For each $b \in B$ and $t_i \in T$, let $v_b(t_i) = i$.
2. For each $b_i \in B_i$ and $b_j \in B_j$ with $i < j$, let $v_{b_i}(b_j) = f(j)$.
3. Let all other valuations be 0.

Note that G consists of $|T| + \sum_{i \in [k]} |B_i| \leq k + 1 + \sum_{i \in [k]} k = k + 1 + k^2$ agents. Further, the maximum utility in G is $f(k) = k^{k^2}$, and can thus be represented with $\log_2(k^{k^2}) = k^2 \cdot \log_2(k)$ bits. Hence, the space required to represent G is polynomial in k ; specifically, it holds that $|G| \in \mathcal{O}(k^5)$.

The high-level idea is that we count up in binary, where a counting agent from group B_i contributes not 2^i to the total score, but $f(i) = i^{ki}$. Further, after each step, we will “collapse” all counting agents from particular groups onto a single track, which ensures that the polynomial number of tracks suffices to represent an exponential number of different values.

As the name suggests, we will use the track agents to form coalitions in the form of “tracks”, where at each point during the relevant part of the dynamics, there exist exactly $k + 1$ coalitions. We often denote to the coalition that some agent $t_i \in T$ belongs to by *track- t_i coalition* C_{t_i} or simply by *track- i coalition* C_i .

Given a coalition $C \subseteq N$, and an agent $a \in N$, we denote to the value $u_a(C) - (u_a(C) \bmod (k + 1))$ by the *truncated utility* of a for C . Since no two track agents can ever be part of the same coalition, and as $f(2) = 2^{2k} > k + 1$, the truncated utility value directly corresponds to a counting agents utility for a coalition when not taking the track agents into account. Given

an $i \in [k]$ and a boolean compactor relation $\triangleleft \subseteq \mathbb{N} \times \mathbb{N}$, we define $B_{\triangleleft i} = \{B_j \in B \mid j \triangleleft i\}$ (e.g., we write $B_{\leq i}$ to denote the set $\{B_j \in B \mid j \leq i\}$).

In the following, we will show that, when starting from the singleton partition, in the χ -dynamics on G , there is a sequence of at least 2^{k-1} deviations. Since $\text{CIS} \subsetneq \chi$, it suffices to show that each performed deviation is a CIS deviation.

Observe that, as there are no negative valuations, no agent will ever block an agent from joining a coalition, and only agents in B_i can be blocked from leaving a coalition if it contains some agent in B_j with $j < i$. Thus, when arguing that a deviation is a CIS deviation, we will focus solely on showing that the latter part is not satisfied and that it is an NS deviation.

Starting from the singleton partition, we first let all agents in $B_{<k}$ deviate to join the track- t_1 coalition, and each agent in B_k to join a unique track- j coalition with $j > 1$. Then, we call the following recursive procedure with parameter $i = k-1$ with respect to the set $\{t_j\}_{j>1}$. We illustrate the procedure in Figure 8.

Let $i \in [k]$ and assume that the following conditions hold:

- I. all agents in $B_{\leq i}$ belong to the track- t coalition C_t for some arbitrary $t \in T$, and
- II. there exists a set $\tilde{T} \subseteq T \setminus \{t\}$ of $i+1$ track agents, such that for each $b \in B_{\leq i}$ and $\tilde{t}, \tilde{t}' \in \tilde{T}$, it holds that
 - (a) b has the same truncated utility for the two track coalitions $C_{\tilde{t}}, C_{\tilde{t}'}$, and
 - (b) b has a strictly higher utility for $C_{\tilde{t}}$ than for her current coalition C_t .

Then, our procedure will produce a partition such that all agents in $B_{>i}$, as well as the track agents in T remain the coalitions they were in before the procedure was called on i , and all agents in $B_{\leq i}$ move to the track coalition $C_{\tilde{t}_{\max}}$ where $\tilde{t}_{\max} \in \tilde{T}$ and such that no agent has a higher valuation for any other agent in $\tilde{T}$ (i.e., the track agent with the highest index among those in $\tilde{T}$).

First, let $\tilde{T}' = \tilde{T} \setminus \{t_{\max}\}$ be a collection of i track agents, and observe that all above conditions are also satisfied for $i-1$ with respect to $\tilde{T}'$. Hence, in case $i > 1$, we call the procedure recursively on $i-1$. Afterwards, as per the above assumption, all agents in $B_{<i}$ belong to some coalition $C_{\tilde{t}'}$ for a $\tilde{t}' \in \tilde{T}'$, and let $\tilde{T}'' = \tilde{T} \setminus \{\tilde{t}'\}$.

Second, we let each agent $b_i \in B_i$ deviate to a unique track- $\tilde{t}''$ coalition for some $\tilde{t}'' \in \tilde{T}''$. To verify that these are indeed CIS deviations, observe that any agent with nonzero valuation for any agent in B_i is part of the track- $\tilde{t}'$ coalition and b strictly increases her utility as per condition II.b.

Next, in case $i > 1$, we recursively call the procedure again on $i-1$ with respect to the set $\tilde{T}''$, where conditions I., and II.a trivially follow from the fact that they initially held for i with respect to $\tilde{T}$. To see that II.b also holds, observe that, as II.a initially held for i with respect to $\tilde{T}$, the truncated utility of any agent $b_j \in B_{<i}$ for a track- $\tilde{t}''$ coalition with $\tilde{t}'' \in \tilde{T}''$ with respect to only the agents in $B_{>i}$ must be exactly $f(i)$ higher than that for her current track- $\tilde{t}'$ coalition. In addition, agent b_j gets at most $k+1$ further utility points from the track agents, and $\sum_{j \in [i-1]} j \cdot f(j) \leq (i-1)^2 \cdot f(i-1) \leq k^2 \cdot f(i-1)$ further utility points from the counting agents

in $B_{<i}$. Therefore, with $i \geq 2$ and $k \geq 5$, we can follow that

$$\begin{aligned}
& u_{b_j}(C_{\tilde{t}''}) - u_{b_j}(C_{\tilde{t}'}) \\
& \geq f(i) - (k + 1 + k^2 \cdot f(i - 1)) \\
& = k^{ki} - (k + 1 + k^2 \cdot k^{k(i-1)}) \\
& \geq k^k \cdot k^{k(i-1)} - (k^2 + k^2 \cdot k^{k(i-1)}) \\
& \geq k^5 \cdot k^{k(i-1)} - (k^4 \cdot k^{k(i-1)}) \\
& \geq k^{k(i-1)} \\
& \geq 1.
\end{aligned}$$

Hence, condition II.b is also satisfied for $i - 1$ with respect to $\tilde{T}''$.

Finally, as per the above assumption, after the recursive call has terminated, all agents in $B_{<i}$ belong to coalition $C_{\tilde{t}_{\max}}$. Hence, as per condition II.b, and based on how $\tilde{t}_{\max}$ was chosen, all agents in B_i (apart from the one that already belongs to $C_{\tilde{t}_{\max}}$) can perform CIS deviations to join $C_{\tilde{t}_{\max}}$. Thus, our claim about the final partition after the recursive procedure on i terminated was justified.

One can verify that the specified conditions hold for the initial call on $i = k - 1$ with respect to $\{t_j\}_{j>1}$. Further, each recursive call requires at least one CIS deviation, and, in case $i \in [2, k - 1]$, induces two recursive calls on $i - 1$. Thus, the sequence requires at least 2^{k-1} deviations, which concludes the proof. $\square$

D.4 Proof of Theorem 13

Theorem 13. *An execution of the CIS dynamics starting from the singleton partition taking at most $s(G)^{s(G)} + n$ deviations can be computed in polynomial time with respect to the game's input size.*

Proof. Define $X := \{a \in N \mid \exists b \in N : v_a(b) > 0 \wedge v_b(a) = 0\}$ as the set of agents defining $s(G)$. Moreover, let $Y := N \setminus X$ be the set of the remaining agents.

We claim that the statement holds for any execution of the CIS dynamics that is constructed according to the following three phases:

1. Iterate through the agents in Y in an arbitrary order. For each agent $y \in Y$, let y perform a CIS deviation if possible.
2. Iterate through the agents in X in an arbitrary order. For each agent $x \in X$, let x perform a CIS deviation to join an agent in Y if possible, where we choose the deviation that maximizes x 's utility in case there are multiple such deviations.
3. Perform arbitrary CIS deviations until a CIS partition is reached.

We begin by showing the following two auxiliary claims.

Claim 29. *After the first phase, no agent in Y can perform a further CIS deviation.*

Proof. First, observe that no agent that is placed in a nonsingleton coalition after the first phase can perform any further CIS deviation. This is because, for an agent $a \in N$ such that some $y \in Y$ deviated to join a 's singleton coalition, it must have hold that $v_y(a) > 0$. In addition, since $y \notin X$, it holds that $v_a(y) \neq 0$. Hence, since a approved of y joining, it follows that $v_a(y) > 0$. Hence, y and a block each other from leaving. Similarly, for any $y' \in Y$ that joined some nonsingleton coalition C , there must have been an $a' \in C$ with $v_{y'}(a') > 0$. Hence, $v_{a'}(y') > 0$ as above and a' blocks y' from leaving.

Next, assume for contradiction that there is an agent $y \in Y$ that is in her singleton coalition after the first phase but can perform a CIS deviation to join some coalition C at some later point in the dynamics. Let $C' \subseteq C$ be the agents that joined C before y 's turn in the first phase. If $u_y(C') > 0$, then y could have joined C' at her turn in the first phase. Hence, $u_y(C') \leq 0$ must hold and there must be some agent $a \in C \setminus C'$ such that $v_y(a) > 0$. Note that if a had ended in a nonsingleton coalition in the first phase, she would not have been able to perform another deviation, and the coalition C would never have formed. Hence, a must have been in her singleton coalition throughout the first phase. Thus, when y had her turn in the first phase, she could have joined a , a contradiction to y ending up in a singleton coalition after the first phase. $\triangleleft$

Claim 30. *Each agent in Y that is in her singleton coalition after the second phase will stay in her singleton coalition for the remaining dynamics.*

Proof. Assume for contradiction that some agent $y \in Y$ is in a singleton coalition at the end of the second phase but is joined by an agent in the third phase. By Claim 29, the joining agent must be an agent $x \in X$. In addition, since y is in a singleton coalition at the end of the second phase, she was in a singleton coalition throughout the second phase. Now, recall that in CIS dynamics utilities weakly increase after each deviation. Hence, x obtains a higher utility than at the end of the second phase. But then, when it was the turn of agent x in the second phase, she did not perform her best deviation, because joining y was an option. $\triangleleft$

Now, because of Claim 29, all deviations in the third phase must be performed by agents in X . Further, because of Claim 30, each of these deviations must be to join one of the coalitions that contained at least one agent in X after the second phase. Note that there are at most $s(G)$ such deviations. Hence, after the first and second phases are complete, there are at most $s(G)^{s(G)}$ unique partitions reachable in the CIS dynamics. Since the CIS dynamics are acyclic, there can be at most n deviations in the first two phases and $s(G)^{s(G)}$ in the third phase, until the dynamics converge. $\square$

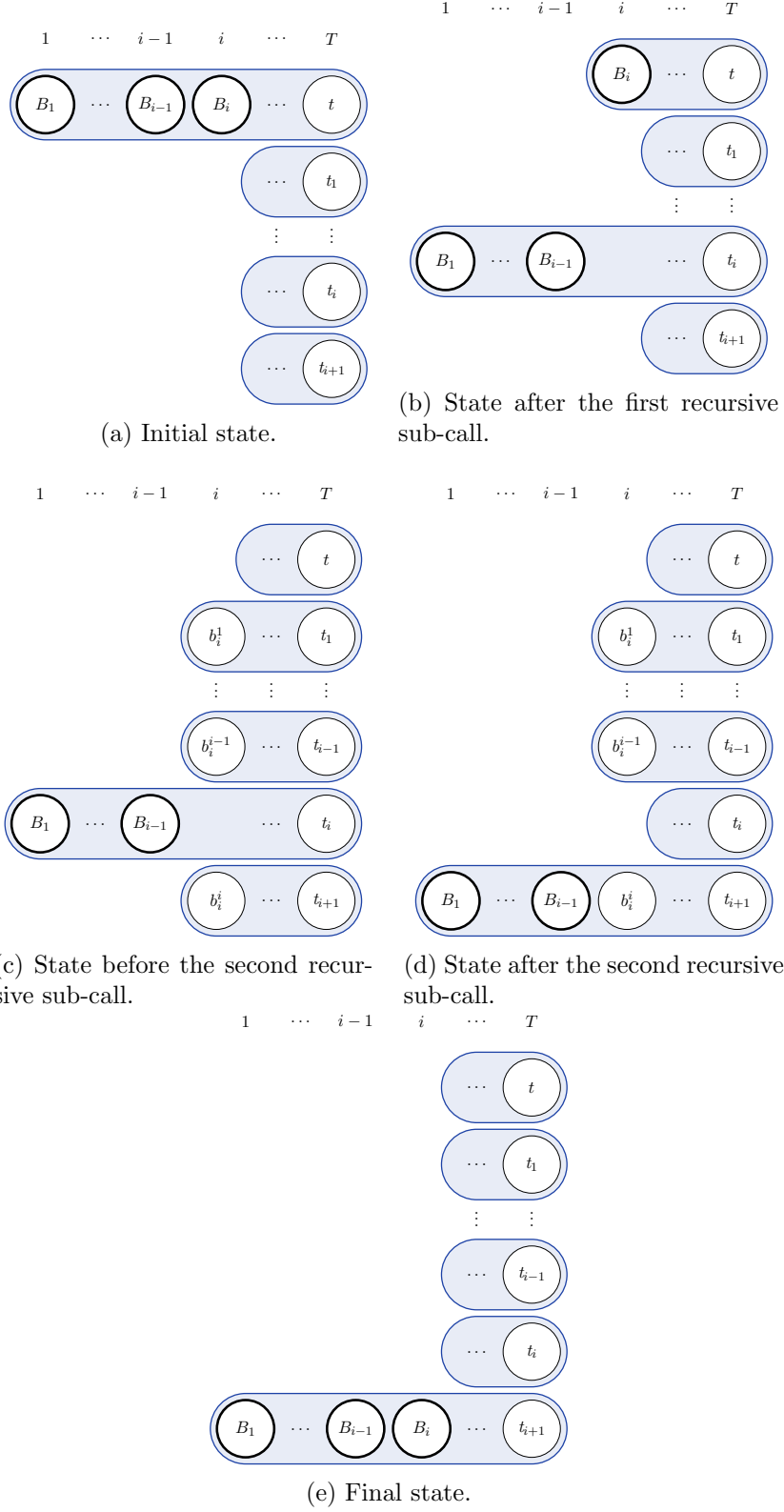


Figure 8: Illustration of the recursive procedure of Theorem 12. The individual sub-figures display the partition of the constructed game after certain steps of the recursive procedure. Bold nodes refer to groups of agents.