Sequential Commitment Games

Itai Arieli[*,§] Yakov Babichenko[†,§] Moshe Tennenholtz[‡,§]

July 29, 2015

Abstract

We study an implementation problem of Pareto efficiency as a subgame perfect equilibrium outcome in extensive form games, faced by a mediator who is ignorant about the payoff structure of the game. We introduce a novel class of sequential commitment games, where players make voluntary unconditional commitments in a prescribed order. Our main result is surprisingly positive: We show that a particular type of order can implement a Pareto efficient outcome in every such two-player game structure *regardless of the actual payoffs.* We also show an impossibility result for the case of four players or more.

1 Introduction

A main social puzzle is how a set of participants finds its way to a Pareto-efficient outcome while each participant maximizes his own utility. In his seminal book Schelling [13] introduced several concepts that may serve this purpose. While Schelling emphasizes the existence of multiple equilibria and the concept of focal point, the main ideas and concepts carry out to other contexts as well. As noted in Myerson [9] two of the main issues discussed in Schelling's book that led to central developments in game theory and that one may further exploit, are *credible commitments* and *legitimate authority.* Commitments can be used in order to promise

[*]iarieli@tx.technion.ac.il

[†]yakovbab@tx.technion.ac.il

[‡]moshet@ie.technion.ac.il

[§]Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology.

cooperation or threatening against deviation from cooperation by other participants, but must subscribe to sequential rationality. A legitimate authority can be used in order to facilitate moving the society from one equilibrium to another.

Considering a setting of a game, and the potential use of credible commitments obeying sequential rationality, and of legitimate authority in imposing such commitments, one may ask what are the capabilities of the legitimate authority: Can it impose behavior? Does it know the players' utilities? Can it enrich the set of actions, by e.g. communicating with the participants or making monetary transfers? Of central interest is the question of whether by only offering services of imposing voluntary commitments, the legitimate authority can lead to a Pareto-efficient outcome even if it does not know the participants' utilities.

Consider an interaction that is modeled as an extensive form game with complete information. Assume that the game tree structure is known also to a mediator (i.e. legitimate authority) but the utility functions (modeled as payoffs at the tree leaves) are unknown to him. The mediator can however provide a voluntary commitment protocol before the actual play starts. In such a protocol the participants are approached with requests for voluntary commitments in different nodes of the tree, and in a well-defined order. Namely, the nodes of the game tree are ordered, and at each point after hearing the previous commitments, the corresponding player can choose to commit to one of the actions in that node, or make no commitment. The role of the mediator is to enforce the voluntary commitments made when the actual game is played. Notice that the voluntary commitments stage, followed by the play of the original game subject to the commitments made, defines a new extensive form game[1], which will be called the *commitment game*, where we can require sequential rationality: as from the players' perspectives this is yet again an extensive form game with complete information wherein we will be interested in subgame perfect equilibrium. A major question expanding on Schellling's concepts, manifested in extensive form games, can be now fully formalized: Can the mediator lead the way to a Pareto-optimal outcome?

The voluntary commitments can be viewed as modifications of the rules of the game, as advocated by political economists. In the words of Buchanan[1]:

"Normatively, the task for the constitutional political economist is to assist individuals, as

[1]Using the Hurwicz [5] terminology the second stage "true game" includes legal and non-legal moves according to the commitments enforced by the "guardian."

citizens who ultimately control their own social order, in their continuing search for those rules of the political game that will best serve their purposes, whatever these might be."

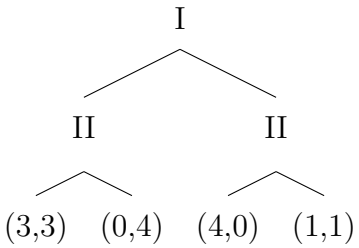
Notice that the above setup is vastly different from approaches such as the ones discussed in the theory of mechanism design and implementation theory, where in order to implement desired outcomes, new games where new actions (such as communication of messages of various forms, monetary transfers, etc.) are added with the aim to overcome the information asymmetry while yielding economic efficiency (as in auction theory [8]) or to overcome the existence of multiple equilibria (as in classical implementation theory for games with complete information [6]). It is also vastly different from the approaches discussed in the literature on conditional commitments implementing threats and promises by mediators in complete information games [7, 14]. The setting discussed above is perhaps the most minimal bridge one may consider in connecting the concepts originated by Schelling of credible commitments and legitimate authority who wishes to lead to desired sequentially rational behavior. Commitments are voluntary on the one hand, and are not conditioned on any event; at the same time, the mediator is trustful but does not have the private information held by the players about their utilities.

In the same spirit as this work, Hamilton and Slutsky [3, 4], van Damme and Hurkens [15], and Renou [11] consider unconditional commitment mechanisms. In these works players are engaged in a preplay stage that enable them to commit simultaneously to a strategy (or in the case of Renou, a subset of strategies). Similarly to the conditional commitment case, this commitment device may enrich the set of equilibria of the underlying game or in some cases generates an effective tool to select among multiple equilibria. In any of these works, however, Pareto efficiency is not guaranteed as a unique outcome of a subgame perfect equilibrium.

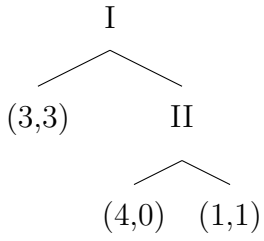
To gain more intuitive insight consider the following classical problem that arises in sequential voting. Two coalition parties are needed to pass law 1 and law 2, which are voted in that order. Party 1 (resp. 2) strongly desires law 2 (resp. 1) to pass and slightly prefers law 1 (resp. 2) to fail. In order to pass a law a unanimous support is required by all members of the coalition. Any equilibrium of the game induced by the above strategic situation has a unique outcome in which the two laws fail. In contrast consider a mediated game where the coalition head may offer a commitment protocol. Can he implement the efficient outcome where all members of the coalition support the two laws?

One can easily see that the scenario described above is a variant of the classical prisoner's

dilemma. For simplicity assume that all members of every party vote unanimously. Assume further that the benefit of passing law 2 (resp. 1) to party 1 (resp. 2) is 3, and the benefit to party 1 (resp. 2) from a failure of law 1 (resp. 2) is 1. The game is then equivalent to the following extensive form prisoner's dilemma.



Party I decides whether to support law 1 (play $left$),[2] or to fail law 1 (play $right$). Party II observes the decision of party I and as a function of I's choice can support law 2 (play $left$) or fail it (play $right$). As mentioned, this game comprises a unique equilibrium, which is also *subgame perfect equilibrium* where every party fails in each of the decision nodes and the resulting outcome is the Pareto dominated outcome $(1, 1)$. Consider the preplay stage where the head of the coalition allows the parties to commit to a decision prior to the voting stage. If party 2 commits to play $left$ in the left-hand decision node and supports law 2 conditional on party 1 supporting law 1, then the resulting game tree would be



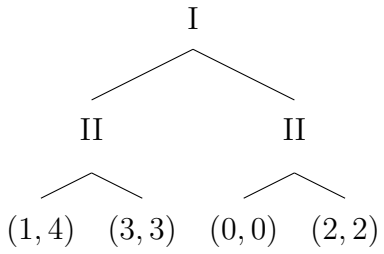
with $(3, 3)$ as a unique equilibrium. Hence, it is rational for party II to commit to support law 2. Therefore, in that case the opportunity of commitment before the game is played may lead to efficiency.

In the above game, in fact, every order over the decision nodes induces a commitment game for which $(3, 3)$ is the unique subgame perfect equilibrium outcome.[3] This highlights that the

[2]Naturally party 1 supports law 2, and party 2 supports law 1.

[3]The outcome $(3, 3)$ is implementable also using a simultaneous commitment, but not as the unique subgame perfect equilibrium. The inefficient outcome $(1, 1)$ is also a subgame perfect equilibrium in this case. For further discussion on simultaneous commitments, and an example where *all* equilibria are inefficient see Section 5 Comment 2.

opportunity to commit may induce an efficient outcome. However, the above reasoning strongly depends on the payoff structure. Consider for example the following game.



Consider a commitment protocol under which player I has the opportunity to commit first to an action and then player II may commit to an action, first in his right-hand side decision node and thereafter in his left-hand side decision node. If player I chooses not to commit to an action in the first step of the commitment stage then player II may force the outcome $(1, 4)$. To see this note that player II can commit to play $left$ in his two decision nodes. Doing this he allows player I to choose between two actions in the play stage: $left$ that yields $(1, 4)$ or $right$ that yields $(0, 0)$. Since player I prefers the $right$ option he must choose $right$ in the play stage, which yields $(1, 4)$. Knowing the above, player I may alternatively commit to $right$ in his first decision node and as a result guarantee 2. Hence, the subgame perfect equilibrium outcome of this commitment game is inefficient. This highlights that achieving efficiency as an outcome of the commitment game without knowing the payoff structure is not obvious. Namely, even in simple examples not every order over the decision nodes leads to a Pareto efficient outcome.

Say that Pareto efficiency is *implementable* for a particular game tree if there exists an order over the decision nodes for which the subgame perfect equilibrium outcome of the commitment game is Pareto efficient *no matter what the payoffs are.* The following question naturally arises: *What are the conditions on the game structure such that Pareto efficiency is implementable?*

Our first result provides a striking positive answer to this question for two-player games. In Theorem 1 we show that *every* two-player extensive form game is implementable. Our proof is constructive: For every two-player game tree we consider the family of *depth first search orders*[4] (henceforth DFS) and show that for every payoff assignment the commitment game induces by any DFS order has a pareto efficient subgame perfect equilibrium outcome.

For general extensive form games we show that our proof does not carry forward. We then provide a sharp negative result for games with more than three players. Particularly,

[4]This family is well known in computer science and algorithms, see Section 2.3.

we construct an extensive form game tree with four players for which Pareto efficiency is not implementable.[5] That is, for every order over the decision nodes we show that there exists a payoff assignment for which the subgame perfect equilibrium outcome of the resulting commitment is pareto dominated.

Following our negative result one may ask whether Pareto efficiency is implementable in some classes of (beyond two-player) extensive form games. We consider the class of *quitting games* where the decision nodes can be ordered in such a way that at every node but the last the corresponding player has two alternatives: either stop the game and force a particular outcome, or stay and pass the decision to the next node. The two alternatives that are available to the last node stop the game and each alternative yields a certain outcome. As an example the well known *centipede game* by Rosenthal [12] is a quitting game. Lastly we show that Pareto efficiency is implementable for the class of multi-player quitting games, for any number of players.

2 Preliminaries

2.1 Extensive form games

An extensive-form perfect information game $G = (N, T, d, u)$ is a tuple which comprises:

- A set of players N.
- A rooted directed tree $T = (V, E)$. For every node v let $A_v = \{w \in V : (v, w) \in E\}$ denotes the *sons* of v. The *terminal nodes* are denote by $C = \{v \in V : A_v = \emptyset\}$. The non-terminal nodes $D := V \setminus C$ are called *decision nodes*. For every node v we denote by T_v the rooted directed subtree (of T) with the root v. The nodes in T_v excluding v itself are called *descendants* of v.
- A labelling of the decision nodes $d : D \to N$. We say that $d(v)$ is *the acting player at node* v. For every player $i \in N$ let $D_i := \{v \in V : d(v) = i\}$ be the decision nodes of i.
- Payoff function[6] $u : C \to \mathbb{R}^{|N|}$.

[5] The case of three-player games remains an open question.

[6] For simplicity on notations only, we consider *cardinal* preferences of the players over the outcomes. All the results in the paper hold for the case where players have *ordinal* preferences over the outcomes.

The first three components (without the payoffs) are called the *game structure*.

A strategy of player i in the game is a choice of a son at each one of his decision nodes (i.e., $s_i = (a_v)_{v \in D_i}$ where $a_v \in A_v$). Every profile of strategies $s = (s_i)_{i \in N}$ induces a unique terminal node $c(s)$. The payoffs of the players for the profile s are given by $u(c(s))$.

A game is called *generic* if for every two terminal nodes $x, y \in C$ such that $u_i(x) = u_i(y)$ for some player i, it holds that $u(x) = u(y)$, i.e., the set of outcomes is generic, but we allow the same outcome to appear several times. For simplicity, throughout the paper we focus on *generic* games only. Genericity guarantees uniqueness of a subgame perfect equilibrium outcome[7] which will be denoted by $\mathsf{Val}(G)$, and will be called *the value of the game* G.

An outcome $u(x)$ *Pareto dominates* the outcome $u(y)$ if $u_i(x) > u_i(y)$ for every $i \in N$. An outcome $u(y)$ is *Pareto efficient* if there is no outcome $u(x)$ that Pareto dominates $u(y)$.

2.2 Commitment Protocols

In this section we formalise the idea of commitment protocols. Given an extensive form game G and an order over all the decision nodes $\overline{v} = (v_1, v_2, ..., v_n)$ such that $\{v_1, v_2, ..., v_n\} = D$, the *commitment game* $Com(G, \overline{v})$ is the perfect information game where first players have the option to commit to an action at every decision node v_i, according to the order $\overline{v}$, and thereafter the game G is played subject to the commitments that have been made at the first stage. Formally, the game $Com(G, \overline{v})$ has two stages which are described below.

The Commitment Stage

In this stage, each player is asked to commit to an action at a node according to the order $\overline{v}$. Commitment is optional but irreversible. That is, first player $d(v_1)$ may commit to an action at v_1. Then player $d(v_2)$ observes the commitment of player $d(v_1)$ at v_1 and decides whether to commit to an action at v_2, and so forth. Formally, at every time $i \geq 1$ the player $d(v_i)$ may commit to an action $a_{v_i} \in A_{v_i}$ at node v_i, as a function of the commitment history up to time i. Thus the decision available to $d(v_i)$ at time i may be seen as a mapping $\mathbf{a}_i : \prod_{k<i}(A_{v_k} \cup \{\phi_k\}) \to A_{v_i} \cup \{\phi_i\}$ that assigns an action $a_{v_i} \in A_{v_i}$ or the null action ϕ_i (the option not to commit), as a function of the commitment history up to time i.

The Play Stage

Let $l = (l_{v_1}, \ldots, l_{v_n})$ be the vector of commitments from the commitment stage, where for every

[7] Note that the subgame perfect equilibrium path may not be unique.

$1 \leq i \leq n$, $l_{v_i} \in A_{v_i} \cup \{\phi_i\}$. In the play stage the original game tree T is being played with the distinction that if a node $v \in D$ is reached such that $l_v \in A_v$ then l_v is played.

Note that any vector of commitments h defines a new game tree T' which is a subtree of the original game tree T. Essentially in the first stage, players, using commitments, determine the game that they are going to play in the play stage.

For a generic game G, the commitment game $Com(G, \overline{v})$ is also a generic extensive form game, and therefore has a value $\mathsf{Val}(Com(G, \overline{v}))$. In cases where it is obvious what is the order $\overline{v}$, we will simply write $\mathsf{Val}(Com(G))$. We now turn to our notion of implementation.

Definition 1. We say that Pareto efficiency is *implementable* for the game structure (N, T, d) if there exists an order $\overline{v}$ such that *for every* generic game $G = (N, T, d, u)$ the subgame perfect equilibrium outcome $\mathsf{Val}(Com(G, \overline{v}))$ is Pareto efficient.

Our notion of implementation is compatible with the fact that the mediator is ignorant of the actual payoff u. We therefore require that Pareto efficiency should be achieved regardless of the actual payoff function.

We shall next turn to some central notions that will be used in the sequel.

Definition 2. Given an order $\overline{v} = (v_1, ..., v_n)$, a node v_i is called *pre-terminal with respect to* $\overline{v}$ if $i \geq j$ for every $v_j \in T_{v_i}$. Namely, v_i appears in the order after all its descendants. An order $\overline{v}$ is *pre-terminal* if all nodes are pre-terminal with respect to $\overline{v}$.

The following observation will be useful to exclude the option of not committing in some cases:

Lemma 1. *Let G be a game, and let v_i be a pre-terminal node with respect to $\overline{v} = (v_1, \ldots, v_n)$. There exists a subgame prefect equilibrium of the game $Com(G, \overline{v})$ where at node v_i player $d(v_i)$ commits to some action $a_{v_i} \in A_v$ (i.e., does not choose ϕ_i).*

Proof. The idea is that choosing the action ϕ_i at the commitment stage is equivalent to a commitment to the subgame perfect equilibrium action of the subtree with the root v_i. More formally, let T'_{v_i} be the game subtree with the root v_i at the moment player $d(v_i)$ is asked to commit at the node v_i (after taking into account all commitments made up to this point). Since v_i is pre-terminal no additional commitments will be made at T'_{v_i}. Let a_{v_i} be a subgame perfect equilibrium action of the game T'_{v_i} (without commitments). We argue that the action ϕ_i (i.e.,

not committing to any action) leads to the same subgame perfect equilibrium outcome as the one that is obtained by committing to a_{v_i}. This follows since all nodes in T_{v_i} have already decided if and to what action they are committed. Therefore, if in the play stage the node v_i will be reached the action a_{v_i} is the action that will be eventually played by $d(v_i)$ at v_i under the subgame perfection assumption. Hence, if in a subgame perfect equilibrium ϕ_i is chosen, then there exists another subgame perfect equilibrium in which a_{v_i} is chosen instead of ϕ_i. □

From Lemma 1 we deduce the following useful remark.

Remark 2.1. In what follows, we will study $\mathsf{Val}(Com(G, \overline{v}))$ as a function of the order $\overline{v}$. Since $\mathsf{Val}(Com(G, \overline{v}))$ is independent of the subgame perfect equilibrium strategy, we can apply Lemma 1 and assume that players at pre-terminal nodes always commit to an action.

2.3 DFS orders

Definition 3. An order $\overline{v} = (v_1, \ldots, v_n)$ is called a *depth first search* (DFS) order if every node v_i appears in $\overline{v}$ precisely after all of its descendants. Namely, if $w_1, ..., w_m$ are the descendants of v_i, then $\{v_{i-1}, v_{i-2}, ..., v_{i-m}\} = \{w_1, ..., w_m\}$.

DFS order is a fundamental notion in computer science and algorithms (see e.g., [2]). There are three types of DFS orders: pre-order, in-order, and post-order. Our notion of DFS is identical to the post-order.

The following properties of a DFS order are easily verified:

- Every DFS order is a pre-terminal order.
- Every DFS order induced a DFS order over any subtree T_v of T for every node $v \in D$.
- Once we set an order of the sons $(w_1^i, w_2^i, ..., w_{m(i)}^i)$ of every vertex v_i, there exists a unique DFS order that satisfies the property: w_j^i is visited before w_l^i for every i and every $j < l$.

3 Two-player games

Our main positive result shows that Pareto efficiency is implementable for every game structure of two players. More particularly we will show that in every two-player extensive form game every DFS order induces a Pareto efficient subgame perfect equilibrium outcome.

Theorem 1. *For every generic two-player extensive form game G and every DFS order $\overline{v}$, the unique subgame perfect equilibrium outcome of the commitment game $Com(G, \overline{v})$ is Pareto efficient.*

In particular, Theorem 1 shows that every two-player game structure is implementable.

Note that, since *any* DFS order implements Pareto efficiency, the designer of the protocol may be ignorant, in addition to the payoffs, as to the identity of the players that act at any given node.

Outline of the proof of Theorem 1: Since the DFS order is pre-terminal, Remark 2.1 allows us to consider an equivalent simpler case where players always commit to some action at every decision node. In such a case the play-stage of the commitment game is redundant, since are already committed to an action at all nodes. This in particular implies that for every node $v \in D$, when v's turn to commit has reached every action of v, along with the previous commitments, defines a unique outcome. Hence at every time of the commitment stage we can identify any action of every node v with a unique outcome. Therefore, we can view the commitment game as follows: At every node v, player $d(v)$ decides which among the available outcomes would replace the subtree T_v in the remaining game tree.

The proof is by induction on the number of decision nodes. Let v be the root node. Let $w_1, w_2, ..., w_m$ be the non-terminal sons of v. The DFS order goes over all nodes in T_{w_1}, then in T_{w_2},..., then in T_{w_m}, and finally at v. The idea is to identify the commitment game with the following two-step procedure.

Step 1: Players in T_{w_1} commit to an action. By the above considerations the commitment of all nodes in T_{w_1} defines a unique outcome b. This outcome may be viewed as the outcome that "replaces" the subtree T_{w_1} in the remaining subgame.

Step 2: Players commit on the remaining parts of the tree (i.e., $T_{w_2} \cup ... \cup T_{w_m} \cup \{v\}$).

We claim that the above two steps define two commitment games of smaller size. Since the decisions of the agents at T_{w_1} determine a unique outcome b it is clear that Step 2 is equivalent to a commitment game of smaller size in which we replace the subtree T_{w_1} with a terminal node yielding the outcome b. Clearly, the value of this commitment game is the same as the value of the original commitment game $\mathsf{Val}(Com(G, \overline{v}))$. Interestingly we may also identify step 1 with a commitment game, denoted by $Com(\widehat{G}, \overline{v}_*)$, that is defined over T_{w_1}. To do it we replace

every outcome x in T_{w_1} with $r(x)$ which is defined by the value of the commitment game of step 2 had we replaced the subtree T_{w_1} with the outcome x (see Figure 3). We show in Lemma 2 that the value of this commitment game, $\mathsf{Val}(Com(\widehat{G}, \overline{v}_*))$, is equal to the value of the original commitment game, $\mathsf{Val}(Com(G, \overline{v}))$.

We now apply the induction hypothesis twice for each of the two smaller commitment games. Assume by way of contradiction that the value of the game is Pareto dominated by some outcome $c \in T$. The outcome c cannot appear in $T_{w_2} \cup ... \cup T_{w_m}$ since by the induction hypothesis the value of the commitment game of step 2 is Pareto efficient. We argue that c cannot appear in the subtree T_{w_1} either. To prove it, we use Lemma 3 to infer that if c dominates the value of the commitment game $\mathsf{Val}(Com(G, \overline{v}))$ then $r(c)$ also dominates $\mathsf{Val}(Com(G, \overline{v}))$. But by Lemma 2 the value of the commitment of the first stage $\mathsf{Val}(Com(\widehat{G}, \overline{v}_*))$ is $\mathsf{Val}(Com(G, \overline{v}))$. This stands in contradiction to the induction hypothesis for $Com(\widehat{G}, \overline{v}_*)$.

As we will show in Example 1, Lemma 3, which is the heart of the proof, is no longer valid for games with four players or more.

3.1 Proof of Theorem 1

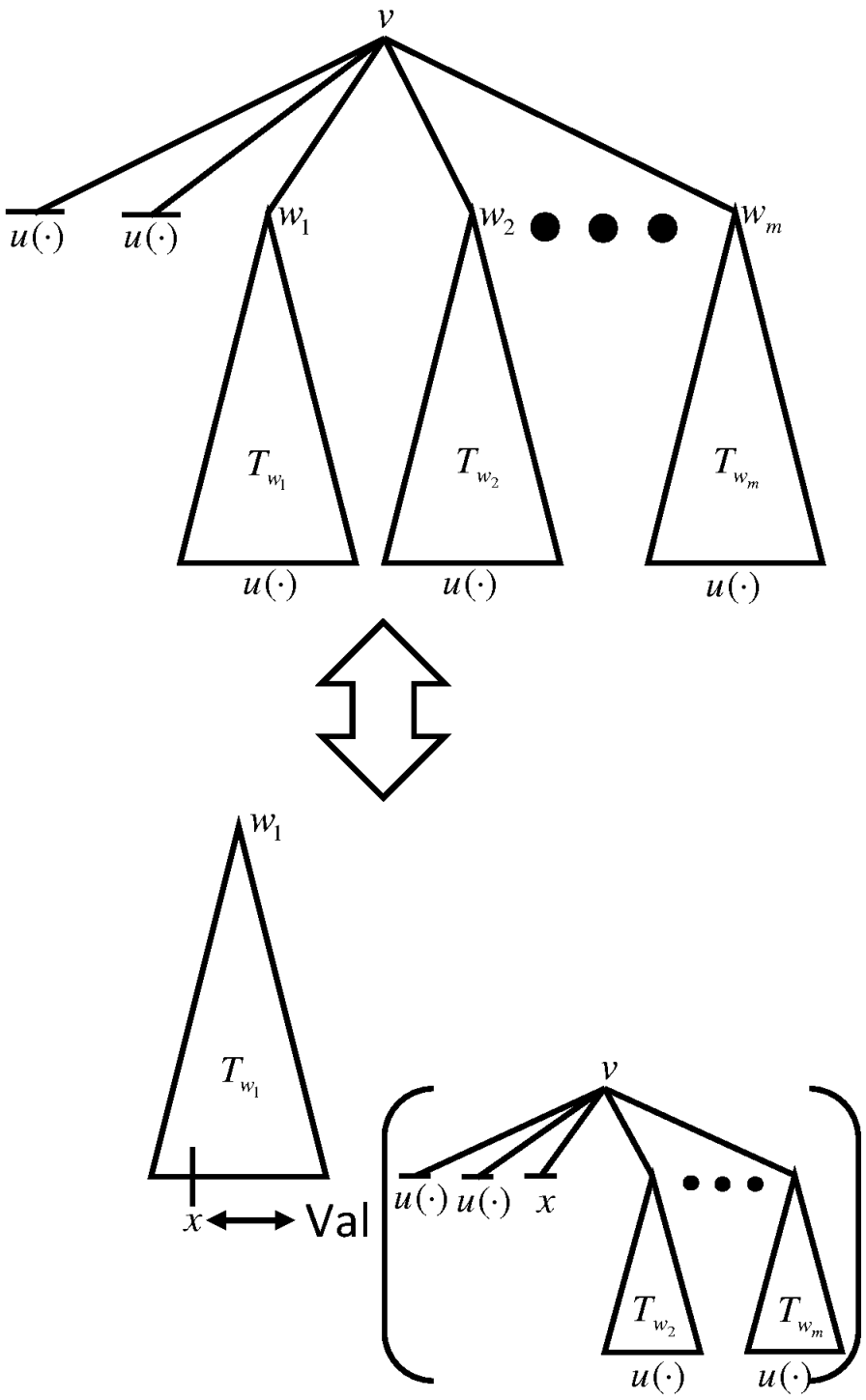
By Remark 2.1, we note that for every DFS order $\overline{v}$ the commitment game $Com(G, \overline{v})$ is equivalent (in terms of values) to the game $Act(G, \overline{v})$ where at each step of the commitment procedure players *must* commit to an action. In $Act(G, \overline{v})$ the commitment stage defines a unique outcome.

Remark 3.1. Remark 2.1 implies that for every pre-terminal order $\overline{v}$ it holds that $\mathsf{Val}(Com(G, \overline{v})) = \mathsf{Val}(Act(G, \overline{v}))$.

We recall the notations from the outline of the proof, and present several others to formalize the ideas presented there. Node v is the root node, while $\overline{v}$ denotes the DFS order. The non-terminal sons of v are $w_1, w_2, ..., w_m$, where w_1 (and the subtree T_{w_1}) is visited before all other sons by the DFS order $\overline{v}$. Note that the order $\overline{v}$ induces a DFS order on the nodes of T_{w_1}. In addition, if we let $T \setminus T_{w_1}$ be the tree that is obtained when we replace T_{w_1} with a terminal node, then $\overline{v}$ also induces a DFS order on $T \setminus T_{w_1}$. We denote these two induced orders on T_{w_1} and on $T \setminus T_{w_1}$ by $\overline{v}_*$ and $\overline{v}^*$ respectively.

For every outcome x denote by $\tilde{G}(x)$ the game over the tree $T \setminus T_{w_1}$ where the outcome

Figure 1: Division of the commitment game into two steps.



at the terminal node replacing T_{w_1} is x, and all other outcomes are identical to the outcomes in G. The value of the commitment game on $\tilde{G}(x)$ with the order $\overline{v}^*$ is denoted by $r(x) := \mathsf{Val}(Com(\tilde{G}(x), \overline{v}^*))$.

We denote by $\widehat{G}$ the game on the tree T_{w_1} where we replace every outcome x of the game G with $r(x)$ and denote by $\mathsf{Val}(Com(\tilde{G}(x), \overline{v}^*))$ the value of the commitment game on $\tilde{G}$ with the order $\overline{v}_*$. We claim first that

Lemma 2. $\mathsf{Val}(Com(G, \overline{v})) = \mathsf{Val}(Com(\widehat{G}, \overline{v}_*))$

Proof. By Remark 3.1 it is sufficient to prove $\mathsf{Val}(Act(G, \overline{v})) = \mathsf{Val}(Act(\widehat{G}, \overline{v}_*))$.

The extensive form game $Act(\widehat{G}, \overline{v}_*)$ is of depth $|T_{w_1}|$. The extensive form game $Act(G, \overline{v})$ is of depth $|T|$, where the nodes at depths $1, 2, ..., |T_{w_1}|$ correspond to commitments in the tree T_{w_1}, and the nodes at depths $|T_{w_1}| + 1, |T_{w_1}| + 2, ..., |T|$ correspond to commitments in the tree $T \setminus T_{w_1}$. We apply backward induction on the game $Act(G, \overline{v})$ up to depth $|T_{w_1}|$, and we denote this game by H. We argue that $H = Act(\widehat{G}, \overline{v}_*)$ (payoff by payoff equality, not only in terms of values).

A terminal node t in the game H that corresponds to a history that selects the outcome x in the subtree T_{w_1} has the payoff vector $r(x)$ (because the subgame starting at t in the large game $Act(G, \overline{v})$ is exactly the game $\mathsf{Val}(Act(\tilde{G}(x), \overline{v}^*)))$. Same is true for the game $Act(\widehat{G}, \overline{v}_*)$: A terminal node t in the game $Act(\widehat{G}, \overline{v}_*)$ that corresponds to a history that selects the outcome x in the subtree T_{w_1} has the payoff vector $r(x)$ (simply by definition).

Summarising, the game $Act(\widehat{G}, \overline{v}_*)$ is obtained from the game $Act(G, \overline{v})$ by applying partial backward induction. Therefore both games have the same value. □

The following lemma is the core of the proof of Theorem 1.

Lemma 3. *Let G be a game such that the root v has at least one terminal son $t \in C$. Let $G(c)$ be the game G where we replace the outcome $u(t)$ by the outcome c. Let $\overline{v} = (v_1, ..., v_n)$ be a pre-terminal order. If $c_i > \mathsf{Val}_i(Com(G, \overline{v}))$ for both players $i = \mathrm{I}, \mathrm{II}$, then $\mathsf{Val}_i(Com(G(c), \overline{v})) \geq c_i$ for both players $i = \mathrm{I}, \mathrm{II}$.*

The lemma states the following. If we replace the outcome of a terminal son of the root v with an outcome c that dominates the value of the commitment game $\mathsf{Val}(Com(G, \overline{v}))$, then the value of the new commitment game is better than c_i for each player $i = 1, 2$.

The proof is based on a general lemma regarding general two-player extensive form games (without commitments).

Lemma 4. *Let G be a generic two-player game where at every pre-terminal node the acting player is* I. *Let $c = (c_{\mathrm{I}}, c_{\mathrm{II}})$ be an outcome such that $c_{\mathrm{I}} > \mathsf{Val}_{\mathrm{I}}(G)$. Let G' be any game that is obtained from G when we replace by c exactly one of the terminal outcomes of every pre-terminal node. Then $\mathsf{Val}_i(G') \geq c_i$ for $i = \mathrm{I}, \mathrm{II}$.*

Proof. We prove the result using induction on the depth of the tree. For a game G of depth one, by assumption, only player I gets to choose and it clearly holds that $\mathsf{Val}(G') = c$. Assume the result holds for all games with a depth smaller than k. Let G be of depth k. We shall consider two cases.

Case 1: Player I is the acting player at the root. Let $w_1, \ldots, w_m$ be the sons of the root, and denote by $G_{w_1}, ..., G_{w_m}$ the corresponding subgames. Since $\mathsf{Val}_{\mathrm{I}}(G) < c_{\mathrm{I}}$, and since $\mathsf{Val}_{\mathrm{I}}(G) = \max_{1 \leq j \leq m} \mathsf{Val}_{\mathrm{I}}(G_{w_j})$ we must have that $\mathsf{Val}_{\mathrm{I}}(G_{w_j}) < c_{\mathrm{I}}$, for every $1 \leq j \leq m$. Hence we can apply the induction hypothesis to deduce that for every $1 \leq j \leq m$ it holds that $\mathsf{Val}_i(G'_{w_j}) \geq c_i$ for $i = \mathrm{I}, \mathrm{II}$, and therefore $\mathsf{Val}_i(G') \geq c_i$ for $i = \mathrm{I}, \mathrm{II}$.

Case 2: Player II is the acting player at the root. As before let $w_1, \ldots, w_m$ be the sons of the root, and $G_{w_1}, ..., G_{w_m}$ the corresponding subgames. Since $\mathsf{Val}_{\mathrm{I}}(G) < c_{\mathrm{I}}$ we must have l for which $\mathsf{Val}_{\mathrm{I}}(G_{w_l}) < c_{\mathrm{I}}$. Hence by the induction hypothesis $\mathsf{Val}_i(G'_{w_l}) \geq c_i$ for $i = \mathrm{I}, \mathrm{II}$. Since player II controls the root, we must have that $\mathsf{Val}_{\mathrm{II}}(G') \geq c_{\mathrm{II}}$. Furthermore, since player I controls all pre-terminal nodes we must have that $\mathsf{Val}_{\mathrm{I}}(G'_{w_j}) \geq c_{\mathrm{I}}$ for every $1 \leq j \leq m$. Hence we must also have that $\mathsf{Val}_{\mathrm{I}}(G') \geq c_{\mathrm{I}}$. This concludes the proof of Lemma 4. □

Now, Lemma 3 follows immediately from Lemma 4 and Remark 3.1.

Proof of Lemma 3. By Remark 3.1 the lemma is equivalent to the following statement: If $c_i > \mathsf{Val}_i(Act(G, \overline{v}))$ for both players i =I,II, then $\mathsf{Val}_i(Act(G(c), \overline{v})) \geq c_i$ for both players i =I,II.

We note that $Act(G, \overline{v})$ is also an extensive form game. Consider its game tree. Since player $d(v)$ that controls the root v is the last to make a commitment in $Act(G, \overline{v})$ ($v = v_n$) it follows that in the game tree of $Act(G, \overline{v})$, player $d(v)$ is acting at all pre-terminal nodes. Replacing the outcome $u(t)$ by the outcome c is translated to replacing by c exactly one of the terminal

outcomes of every pre-terminal node. Lemma 4 states that indeed the value of the new game (after the replacement) is weakly above c_i for both players i =I,II. □

We turn next to the proof of Theorem 1.

Proof of Theorem 1. As before, by Remark 3.1 it is sufficient to prove that $\mathsf{Val}(Act(G,\overline{v}))$ is Pareto efficient.

The proof is by induction on the number of decision nodes in T. Clearly, for $n = 1$ the result holds, because the acting player maximizes his own payoff. Assume that $\mathsf{Val}(Act(G,\overline{v})$ is Pareto efficient for every game with less than n nodes; we now prove it for games with n nodes.

Consider a subgame perfect equilibrium of the game $Act(G,\overline{v})$, and denote by b the outcome that is chosen in the subtree T_{w_1}. We have

$$\mathsf{Val}(Act(\widehat{G},\overline{v}_*)) = \mathsf{Val}(Act(G,\overline{v})) = \mathsf{Val}(Act(\tilde{G}(b),\overline{v}^*)) = r(b).$$

The first equality follows from Lemma 2. The second equality follows by noting that the remaining subgame of $Act(G,\overline{v})$ after all nodes at T_{w_1} have committed is equivalent to the game $Act(\tilde{G}(b),\overline{v}^*)$. The third equality follows from the definition of $r(\cdot)$.

It is easy to see that both games $\widehat{G}$ and $\tilde{G}(b)$ are games of size strictly smaller than n.

Assume by way of contradiction that $\mathsf{Val}(Act(G,\overline{v})) = r(b)$ is Pareto dominated by an outcome c.

Case 1: The outcome c lies in the tree T_{w_1}. Consider the game $Act(\widehat{G},\overline{v}_*)$. By Lemma 3 the outcome $r(c)$ is weakly better than c for both players. Hence $r(c)$ pareto dominates $\mathsf{Val}(Act(\widehat{G},\overline{v}_*)) = r(b)$. This stands in contradiction to the induction hypothesis applied to the game $Act(\widehat{G},\overline{v}_*)$, because $r(b)$ is the chosen outcome and there exists a better outcome for both players, $r(c)$.

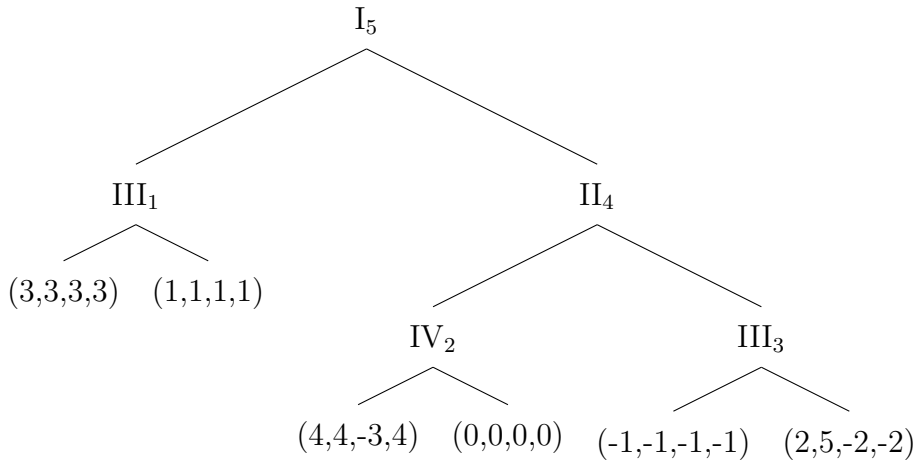
Case 2: The outcome c is not in T_{w_1}, which means that it lies in $T \setminus T_{w_1}$. This stands in contradiction to the induction hypothesis applied to the game $Act(\tilde{G}(b),\overline{v}^*)$, since its value $\mathsf{Val}(Act(\tilde{G}(b),\overline{v}^*))$ equals $r(b)$ which is pareto dominated by c. □

4 Multi-Player Games

Theorem 1 shows that any DFS commitment protocol implements a Pareto efficient outcome in every two-player game. The following example demonstrates that the DFS protocol fails in

four-player games.

Example 1. The DFS order over the decision vertices appears at the bottom-right side of the players.



(1)

Proposition 1. *The outcome of the commitment protocol in Example 1 is* $(1, 1, 1, 1)$.

Before providing the formal proof we outline the intuition for it. The idea is that player III wants to avoid the outcome $(4, 4, -3, 4)$. His only way to avoid it is by committing to the outcome $(2, 5, -2, -2)$ at the third decision node (because this way III would convince II not to choose $(4, 4, -3, 4)$).

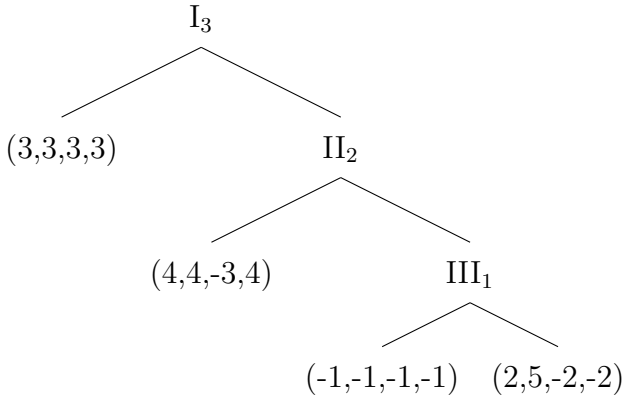
In case III commits to $(1, 1, 1, 1)$ at the first decision node, III has threatening power on IV: "If you (IV) will commit to $(4, 4, -3, 4)$ then the best option for me (III) would be to commit to $(2, 5, -2, -2)$, which also will be chosen as the final outcome, and this is bad for both of us." Therefore IV has to commit to $(0, 0, 0, 0)$, and this eventually leads to the outcome $(1, 1, 1, 1)$.

In case III commits to $(3, 3, 3, 3)$ at the first decision node, he loses his threatening power: IV no longer afraid of $(2, 5, -2, -2)$ because it will not be chosen by I. Therefore, IV commits to $(4, 4, -3, 4)$ and it is chosen as a final outcome.

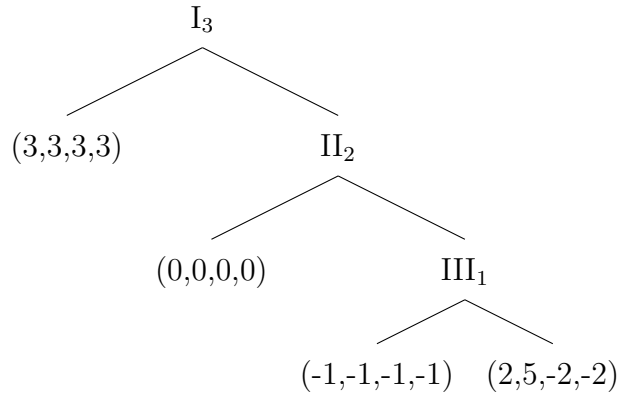
Hence, III prefers to commit to $(1, 1, 1, 1)$ at the first decision node; moreover, it is eventually selected as the final outcome.

Formal proof of Proposition 1. Consider the following four games:

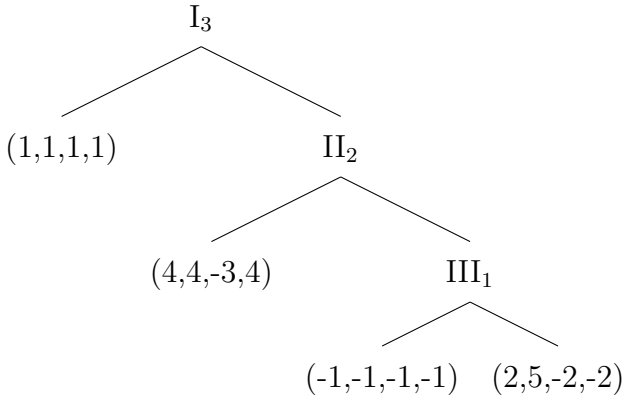
G_1 :



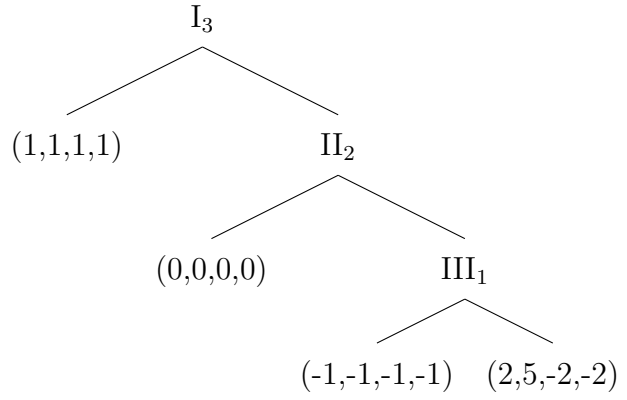
G_2 :



G_3 :



G_4 :



It is easy to see that $\mathsf{Val}(Com(G_1)) = (4, 4, -3, 4)$, $\mathsf{Val}(Com(G_2)) = (3, 3, 3, 3)$, $\mathsf{Val}(Com(G_3)) = (2, 5, -2, -2)$, and $\mathsf{Val}(Com(G_4)) = (1, 1, 1, 1)$. Therefore, a choice of $(3, 3, 3, 3)$ by III, will

be proceeded by a choice of $(4,4,-3,4)$ by IV (because $\mathsf{Val}_{\mathrm{IV}}(Com(G_1)) > \mathsf{Val}_{\mathrm{IV}}(Com(G_2))$). A choice of $(1,1,1,1)$ by III, will be proceeded by a choice of $(0,0,0,0)$ by IV (because $\mathsf{Val}_{\mathrm{IV}}(Com(G_4)) > \mathsf{Val}_{\mathrm{IV}}(Com(G_3))$). When III takes it into account, he prefers to choose $(1,1,1,1)$ and we have $\mathsf{Val}(Com(G)) = \mathsf{Val}(Com(G_4)) = (1,1,1,1)$. □

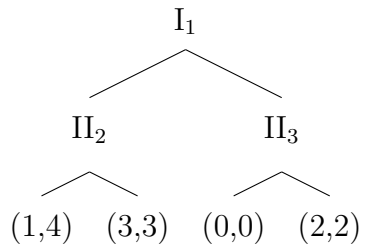
Proposition 1 demonstrates that Pareto efficiency is not implementable by any DFS order, as in the two-player case, where the number of players is greater than 3. However it doesn't rule out implementation, as there still may exist some order that implements Pareto efficiency in these games. The following Theorem excludes such a possibility and shows that Pareto efficiency cannot be implementable in all extensive form games with 4 players or more.

We recall that *extensive form game structure* (N,T,d) comprises all components of a game but for the payoffs (i.e., the set of players N, a game tree T, and a labelling of the decision nodes d) and an order $\overline{v}$ implements Pareto efficiency in a game structure (N,T,d) if for every payoff function u the resulting outcome of the commitment game is Pareto efficient.

Theorem 2. *There exists a four-player extensive form game structure (N,T,d), such that for every order $\overline{v}$ there exists a game $G = (N,T,d,u)$ for which $\mathsf{Val}(Com(G,\overline{v}))$, the subgame perfect equilibrium outcome of the commitment game, is Pareto dominated.*

Idea of the proof of Theorem 2. The proof is based on two examples. The first is Example 1. For the proof of the theorem we have to extend the set of orders for which Example 1 leads to the inefficient outcome $(1,1,1,1)$. It turns out (see Lemma 6) that $(1,1,1,1)$ is the outcome of the commitment game not only for the DFS order presented in Example 1 but also for a much richer set of orders. The second example we shall use is the following example, which was discussed also in the introduction.

Example 2.



Lemma 5. *The outcome of the game in Example 2 is* $(2,2)$.

Proof. If I does not commit at the first step, then II will commit to play $left$ in both nodes (i.e., $(1,4)$ and $(0,0)$), and the outcome will be $(1,4)$. The same outcome is the result of a

commitment to $left$ of player I at the first step. On the other hand, if I commits to right at the first step the outcome will be $(2, 2)$. Therefore $(2, 2)$ is the Pareto dominated outcome of this commitment game. □

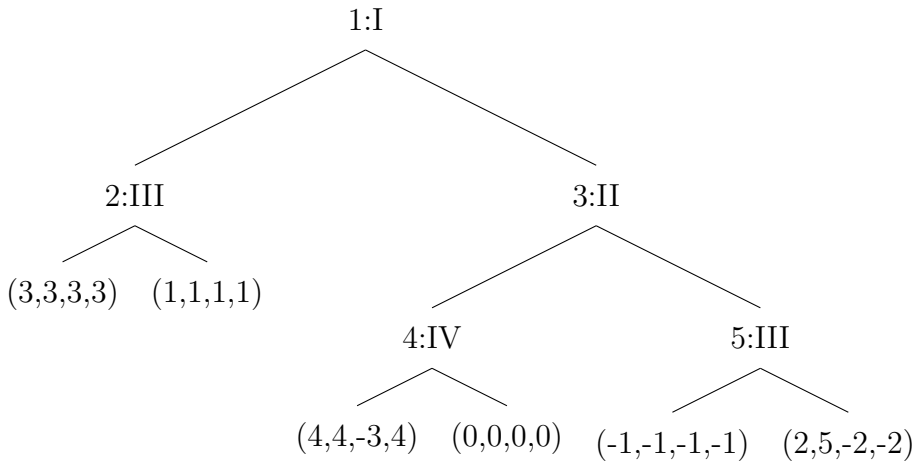
We say that an order over vertices *admits uncertainty in two subtrees* if there exists a vertex that appears before two of his descendants in two different subtrees, or formally, if there exists a triple $i < j < k$ such that $v_j \in T_{w_1}, v_k \in T_{w_2}, w_1 \neq w_2$ and w_1, w_2 are sons of v_i.

The idea is to use Example 2 to argue that every order that admits uncertainty in two subtrees fails for some payoffs. This is because we can "plant" the payoffs of Example 2 in the relevant leaves of v_j and v_k, and make all other payoffs irrelevant. This allows us to focus only on orders that do not admit uncertainty in two subtrees.

Recall that in Example 1 the order in which players commit is (III, IV, III, II, I). Define the game structure to be a complete binary tree of depth five where players I, II, III, IV, III are making the decision in *all* nodes of depth 1,2,3,4,5 respectively. This structure, together with the fact that the order does not admit uncertainty in two subtrees allows us to focus only on orders that satisfy the following conditions:

- III is asked to commit first (at depth 5).
- In both subtrees that starts at depth 2, IV is asked to commit at depth 4 before II, III are asked to commit at depths 2,3.

It turns out that these two properties are sufficient for "planting" the payoffs of Example 1 into the large tree (i.e., the depth five tree), such that the induced order on the relevant nodes of the large tree fits the set of orders in Lemma 6 for which Example 1 fails efficiency.



(2)

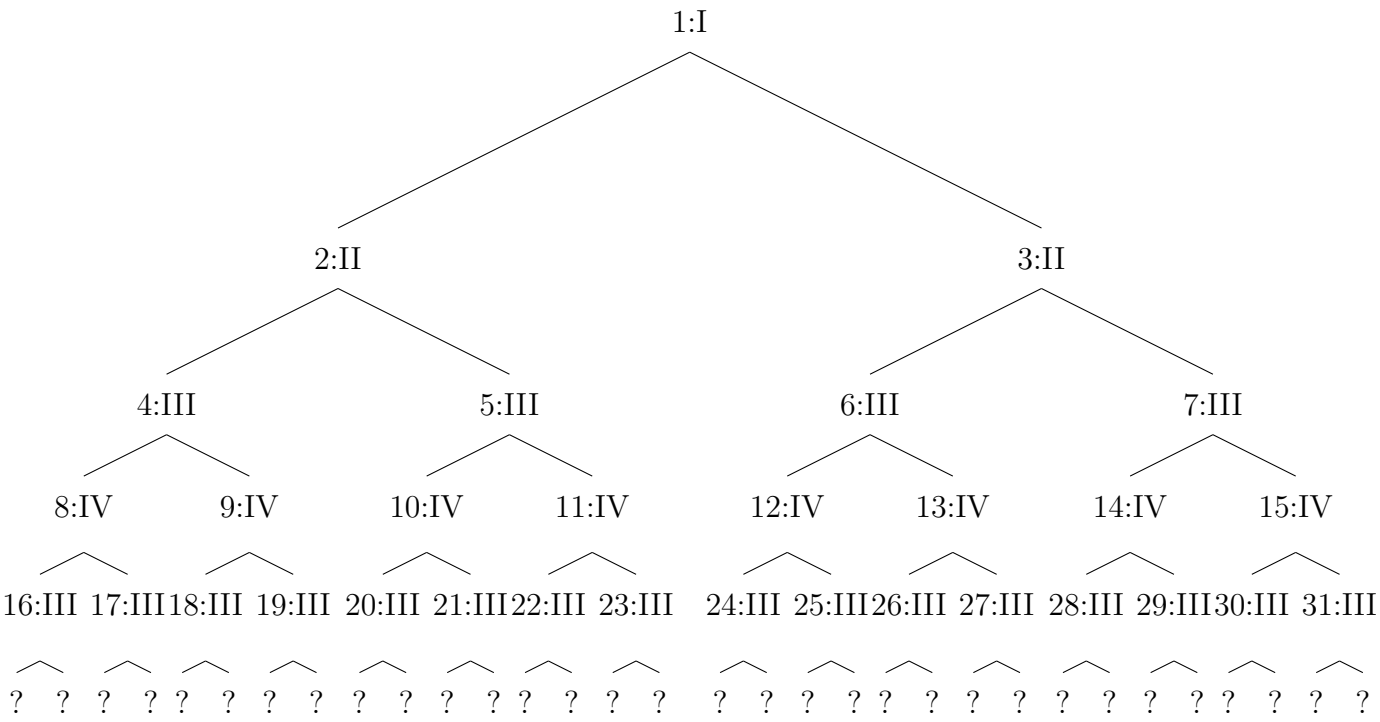
Lemma 6. *Let G be the extensive form game of Example 1 (see (2)). For every order that satisfies:*

- *Node 2 appears first.*
- *Node 4 appears before 3 and before 5.*

the outcome of the commitment game is $(1, 1, 1, 1)$.

The proof of the Lemma is relegated to Appendix A.

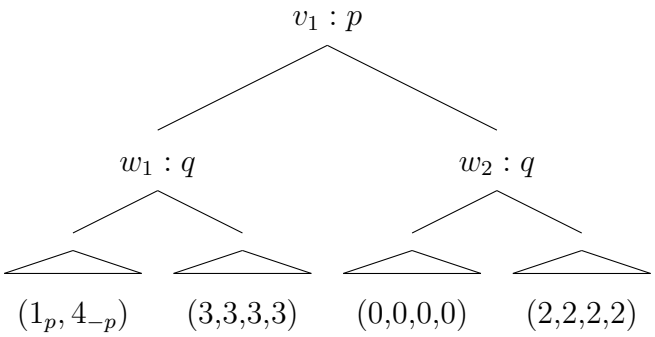
Proof of Theorem 2. Let the structure of the game be a complete binary tree of depth five where players I, II, III, IV, III act at all nodes of depth 1,2,3,4,5 respectively. We enumerate the decision nodes as follows:



Let $v_1, v_2, ..., v_{31}$ be an order of the decision nodes.

If $v_1 \in \{1, 2, ..., 15\}$, we denote $p = d(v_1)$ the acting player at v_1, and q the acting player at one depth below v_1. We plant the payoffs of Example 2 as follows:

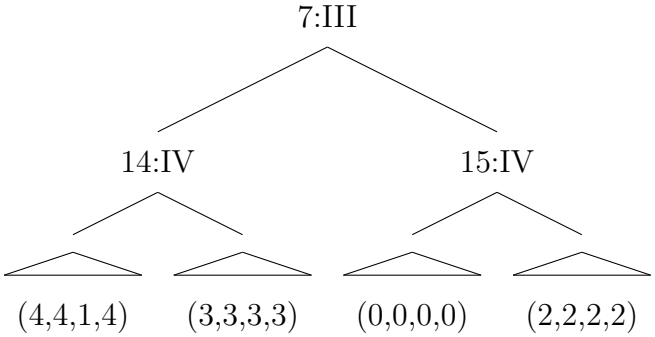
On the subtree with the root v_1 we set the payoffs to be



where $\underset{(u)}{\triangle}$ denotes a subtree in which all payoffs are u, and $(1_p, 4_{-p})$ denotes the payoff profile where player p gets 1 and all other players get 4. Out of the subtree with the root v_1 we set the payoffs to be $(-1,-1,-1,-1)$. The outcome of this game is $(2,2,2,2)$, because in all decision nodes other than v_1, w_1, w_2 players will commit on (or equivalently will not commit but play this action at the final stage of playing the game) actions that lead to v_1 (this is their only option to avoid the payoff -1). By the arguments in the proof of Lemma 5 at node v_1 player p commits on w_2, which essentially leads to the outcome (2,2,2,2).

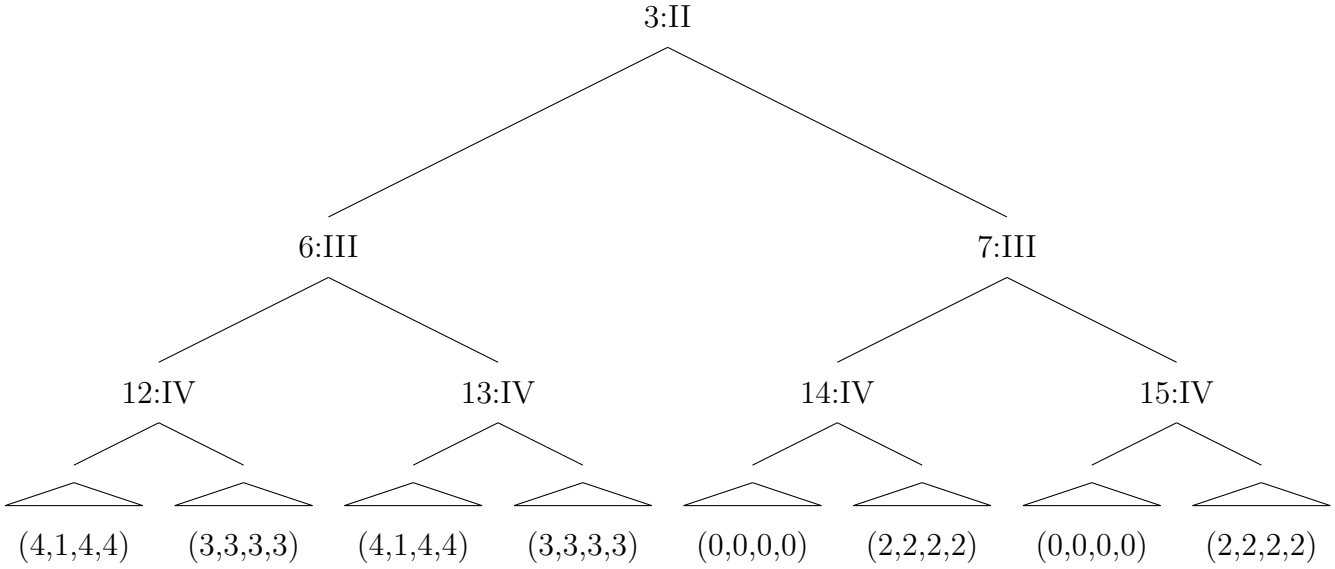
Therefore, we can focus only on the case where $v_1 \in \{16, 17, ..., 31\}$. With no loss of generality, assume that $v_1 = 16$ and that 12 appear in the order before 13,14, and 15.

If 7 appears in the order before 12, then we set the payoffs to be



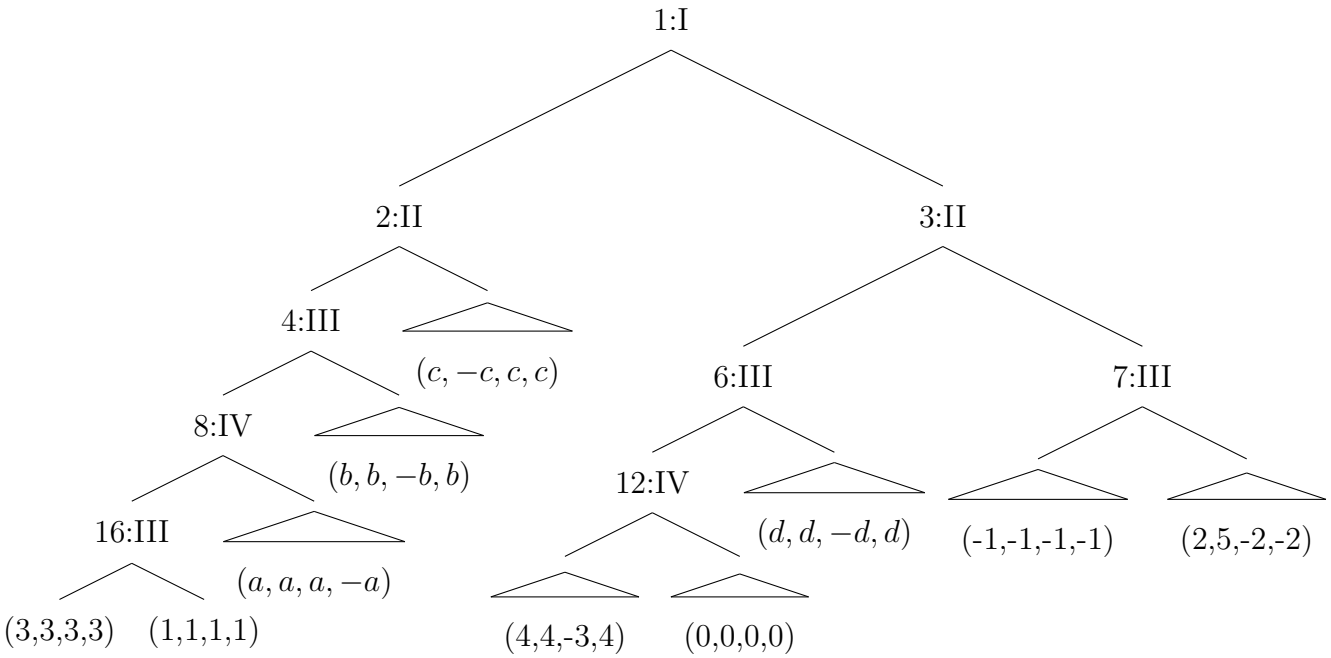
and $(-1,-1,-1,-1)$ everywhere else. By similar arguments, since 14 and 15 appear after 7 in the order, the outcome of the game is $(2,2,2,2)$.

If 3 appears in the order before 12, then we set the payoffs to be



and $(-1,-1,-1,-1)$ everywhere else. Again by similar arguments, since 12,13,14, and 15 appears after 3, only commitment on right of player II leads to a payoff above 1 for him (because player IV can commit on left in all vertices 12,13,14,15), and the outcome is $(2,2,2,2)$.

The remaining case is the one where the first node in the order is 16, and 12 appears before 3 and 7 in the order. In such a case we set the payoffs as follows:



where $a,b,c,d>5$. We order the values of a,b,c,d to be in opposite order according to which the vertices $8,4,2,6$ are visited: for example, if along the order we first visit 8, after that 6, after that 4, and after that 2, then we set the values to satisfy $a>d>b>c>5$. We argue that in nodes $2,4,6,8$ players II,III,III,IV players do not commit to the outcomes $(a,a,a,-a),...,(d,d,-d,d)$ respectively and do not play it in the play stage of the commitment

game. Assume, with no loss of generality, that 8 is the first node among 2,4,6,8 that the order visits.. Note that $-a$ is the worst payoff player IV can get in the game; a is the best payoff players I, II, III can get in the game. If IV commits to $(a, a, a, -a)$ then it also will be chosen as a final outcome. Therefore IV would not commit to $(a, a, a, -a)$ and obviously won't play it at the final stage of playing the game. Assume, with no loss of generality, that 6 is the second node among 2,4,6,8 that the order visits. Note that $-d$ is the worst payoff for III; d is the best payoff for players I, II, IV because we already ruled out the possibility of IV receiving the payoff a. If III commits to $(d, d, -d, d)$ then it also will be chosen as a final outcome. Therefore III does not commit to $(d, d, -d, d)$ and obviously does not play it in the play stage of the commitment game. We proceed similarly for the third and the fourth nodes in the order. Therefore we can ignore the subtrees where the outcomes are $(a, a, a, -a), ..., (d, d, -d, d)$ and we get a tree that has exactly the same outcomes as in Example 1. Finally, the order over nodes 1,3,7,12,16 satisfies:

1. 16 if the first.

2. 12 is before 3 and 7.

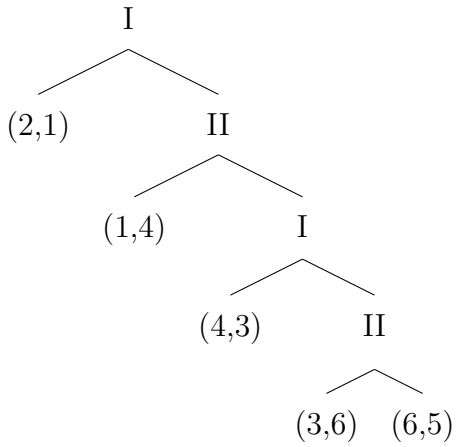
By Lemma 6 the outcome of the game for every such order is the Pareto dominated outcome $(1, 1, 1, 1)$. □

4.1 Quitting games

In this section we study a special class of extensive form games that is called *quitting games*. Our main result shows that Pareto efficiency is implementable in this class of games.

Definition 4. An extensive-form perfect information game G is called a *quitting game* if $D = \{v_1, \ldots, v_n\}$ such that every node v_j has two actions $\{exit, stay\}$, a choice of *stay* by $d(v_j)$ for $j < n$ leads to v_{j+1} and a choice of *exit* leads to a terminal node. Both choices at v_n lead to a terminal node.

A well-known example of a quitting game is the *centipede game* below.



A slightly more intricate example of a quitting game is the three-player centipede game studied by Rapoport et al. [10]. For the above centipede game, we know from Theorem 1 that the unique pre-terminal order $(v_n, v_{n-1}, ..., v_1)$, which is also a DFS order, leads to Pareto efficiency. The following result shows that this is in fact true for every multi-player quitting game.

Theorem 3. *For every quitting game G the unique pre-terminal order $\overline{v} = (v_n, \ldots, v_1)$ leads to Pareto efficiency, i.e., $\mathsf{Val}(Com(G, \overline{v}))$ is Pareto efficient.*

Proof. The proof is obtained using simple induction on n. Clearly, for $n = 1$ the theorem holds. We assume the result for $n - 1$ and prove it for n.

We denote the acting player at v_1 by $d(v_1) = j$. We denote by y the outcome after *exit* at node v_1. We use the same notations as in Theorem 1. We recall that $\widehat{G}$ is the game defined on T_{v_2} when we replace every outcome x with $r(x)$, where $r(x)$ is the value of the game $\tilde{G}(x)$ that is obtained when we replace the subtree T_{v_2} with the outcome x. In the quitting games case, $\tilde{G}(x)$ translates to a simple decision of player j between two options x or y. Therefore, $r(x)$ gets a particular simple form:

$$r(x) = \begin{cases} x & \text{if } x_j \geq y_j \\ y & \text{if } x_j < y_j. \end{cases} \tag{3}$$

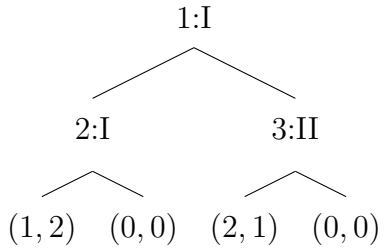
By Lemma 2, we know that $\mathsf{Val}(Com(G, \overline{v})) = \mathsf{Val}(Com(\widehat{G}, \overline{v}_*))$ where in this case $\overline{v}_* = (v_n, \ldots, v_2)$. So applying the induction hypothesis on $Com(\widehat{G}, \overline{v}_*)$ we can deduce that $\mathsf{Val}(Com(G, \overline{v}))$ is Pareto efficient with respect to all outcomes $\{r(x) : x \in T_{v_2}\}$. If, by contradiction, $\mathsf{Val}(Com(G, \overline{v}))$ is Pareto dominated by some outcome x' (in the original game), then

$$x'_j > \mathsf{Val}_j(Com(G, \overline{v})) \geq y_j$$

where the second inequality follows from the fact that player j can guarantee y_j. Therefore, by equation (3) we have $r(x') = x'$. This contradicts the fact that the value is Pareto efficient with respect to $\{r(x) : x \in T_{v_2}\}$. □

5 Comments

1. The dependence of the outcome on the DFS order. Theorem 1 shows that every DFS order of commitments leads to a Pareto efficient outcome. It is natural to ask whether *every* DFS order leads to the same outcome in every two-player extensive form game. The answer is negative, as demonstrated by the following example.



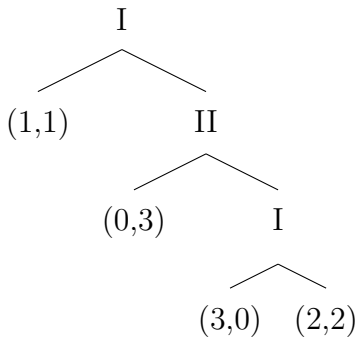
For the DFS order (3,2,1) player II commits to the right action (i.e., the outcome $(0, 0)$), which enforces player I to choose left in both of his nodes, and the outcome is $(1, 2)$.

For the DFS order (2,3,1) player I commits to $(0, 0)$, which leads player II to choose $(2, 1)$, and it is also the final outcome.

2. On the necessity of sequentiality The commitment protocols that we consider in this paper are sequential in two aspects. First, players commit sequentially (and not simultaneously). Second, each player is asked to commit on a single decision node at each step. We argue that both aspects of sequentiality are necessary for efficiency to be implemented, even in two-player games.

Regarding the first aspect, consider a commitment protocol where players simultaneously commit to actions, and then turn to the play stage. This is the model that is discussed in the previous works on commitments [3, 4, 15, 11]. The sequential prisoner's dilemma (see page 4) demonstrates that in such a case inefficient outcomes may be obtained as a subgame perfect equilibrium outcome. The following example demonstrates the existence of extensive form games where *all* equilibria outcomes are inefficient.

Example 3.



In the commitment stage player II has three actions: Committing to left (L), committing to right (R), or not committing (ϕ). Similarly, player I has nine actions which represent a choice of an action (among the three) in each one of his decision nodes.

After the simultaneous commitment, both players will follow the subgame perfect equilibrium of the remaining game. Therefore, the simultaneous commitment game is equivalent to the one-shot game presented in Figure 2.

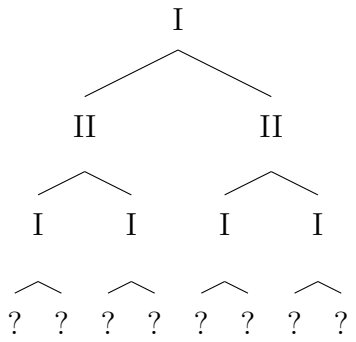
Figure 2: Simultaneous commitment game for Example 3

	ϕ	L	R
ϕ, ϕ	(1,1)	(1,1)	(3,0)
ϕ, L	(1,1)	(1,1)	(3,0)
ϕ, R	(1,1)	(1,1)	(2,2)
L, ϕ	(1,1)	(1,1)	(1,1)
L, L	(1,1)	(1,1)	(1,1)
L, R	(1,1)	(1,1)	(1,1)
R, ϕ	(0,3)	(0,3)	(3,0)
R, L	(0,3)	(0,3)	(3,0)
R, R	(0,3)	(0,3)	(2,2)

It is easy to check, using weak dominance arguments that the unique Nash equilibrium outcome (pure or mixed) of this game is $(1, 1)$, which is inefficient.

Regarding the second aspect of sequentiality, consider the class of commitment protocols where the order is not on decision nodes but on players. Player i, at his turn, is allowed to commit on actions in all his decision nodes. This is the model that is discussed in the existing

literature [3, 4, 15, 11]. We do not assume that a player has an option to commit only once; i.e., the order may contain player i several times. A game structure that excludes the possibility of Pareto efficient implementation is the following:



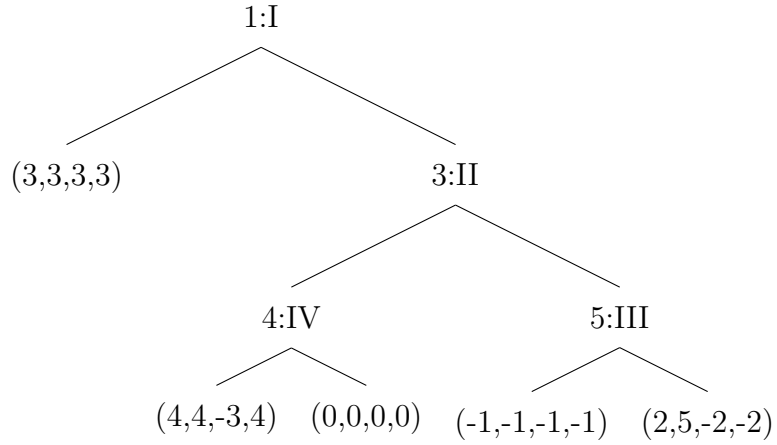
Depending on the player that commits first (player I or player II) we can "plant" into this structure the game of Example 2 to get inefficiency.

Here, we demonstrate inefficiency in the extreme case, where each player is allowed to commit in *all* of his nodes. Theorem 2 demonstrated inefficiency in the other extreme case, where players are restricted to commit to a single node at each step. We believe that the counter examples developed in this paper may also show inefficiency in four-player games for a much richer class of commitment protocols. In particular, intermediate protocols between the two extremes, where players are allowed to commit on subsets of nodes, and the option to commit at a certain node may appear several times along the protocol.

A Appendix: Proof of Lemma 6

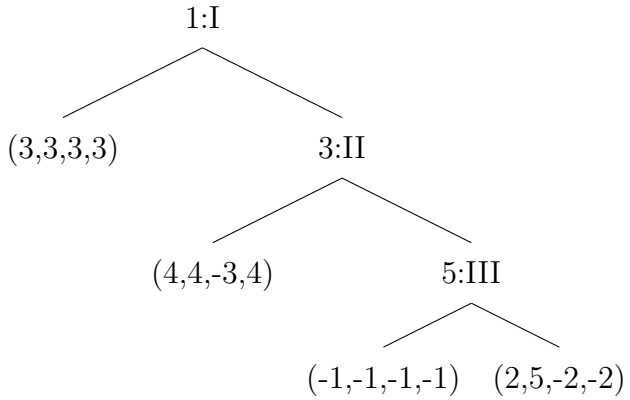
Player III commits first at node 2. By Remark 2.1 we can assume that he commits to an action. We consider first the case where he commits to (3,3,3,3). We argue that for every order over the vertices 1,3,4,5 the outcome of the commitment game:

G_1:



is (4,4,-3,4). The outcome cannot be below 3 for player I, because I can guarantee 3 by choosing left (irrespective of commitments). Therefore the outcome is either (3,3,3,3) or (4,4,-3,4). By Remark 2.1 we can assume that player IV commits to an action at node 4. A commitment to (0,0,0,0) by player IV (at any stage of the commitment procedure) guarantees that the outcome will be (3,3,3,3) (because it excludes the possibility of (4,4,-3,4)), therefore it is weakly better for player IV to choose (4,4,-3,4). Therefore it is sufficient to focus on the game:

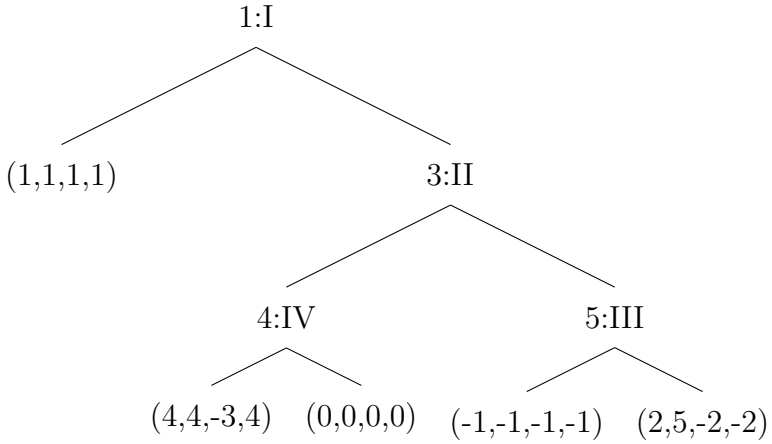
G_1':



It can be verified that every order over the vertices 1,3,5 leads to the outcome (4,4,-3,4), where in equilibrium player I always chooses not to commit, and player II commits to the outcome (4,4,-3,4).

Since -3 is the worst outcome for player III, he never will choose to commit to (3,3,3,3) at node 2. It must be the case that at the first step player III commits to (1,1,1,1). So we remain with the game

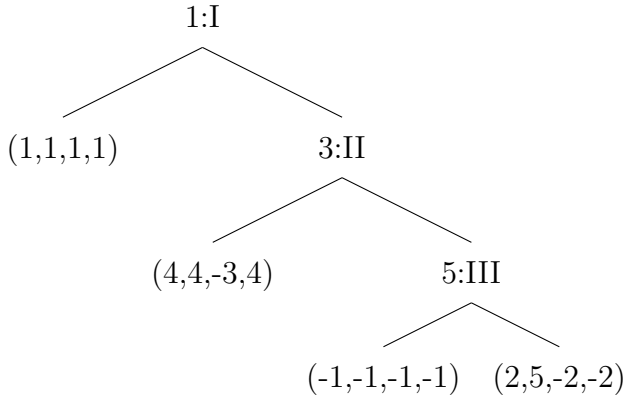
G_2:



We argue that for every order where IV commits before II and before III the outcome is (1,1,1,1).

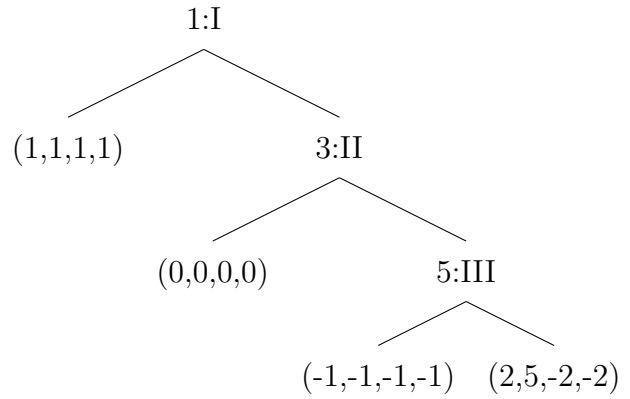
Case 1: Player IV commits first. Again, by Remark 2.1 we can assume that IV commits to an action. If IV commits to (4,4,-3,4), then we remain with the game:

G_2':



It can be verified that every order over the vertices 1,3,5 leads to the outcome (2,5,-2,-2), where in equilibrium player III always commits on (2,5,-2,-2). Since -2 is the worst outcome for player IV, he does not commit to (4,4,-3,4) at node 4, and therefore commits to (0,0,0,0). So we remain with the game

G_2':

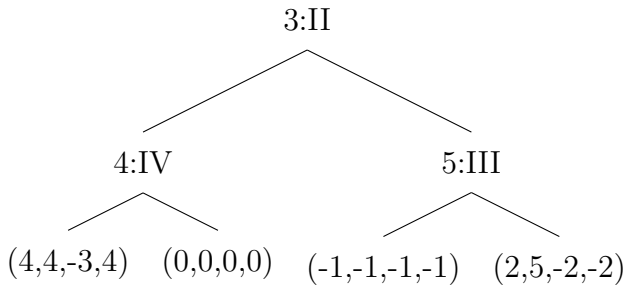


where (2,5,-2,-2) is the worst outcome for player III. Moreover if he chooses it, then it will

be selected (because it is the best outcome for both players I and II). Therefore, III always commits to (-1,-1,-1,-1) and then (1,1,1,1) is the best outcome for player I, and it is selected.

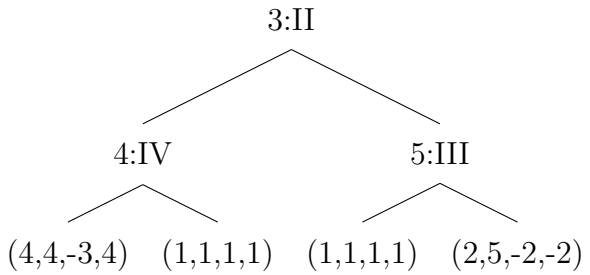
Case 2: Player I acts first. If I commits to (1,1,1,1) we are done. If I commits on right, then we remain with the game:

G_3:



player IV is the first player to commit. If he commits on (4,4,-3,4), then the outcome will be (2,5,-2,-2), because in equilibrium II always can choose not to commit while III always commits to (2,5,-2,-2). Therefore IV commits to (0,0,0,0). In such a case, in equilibrium III always commits on (-1,-1,-1,-1), which leads to the outcome (0,0,0,0). Therefore I will never commit to right. The remaining option of I is to choose not to commit. In such a case, by the arguments in the proof of Theorem 3, we get that the remaining game is equivalent to the game:

G_4:



Here, similar arguments to those of the analysis of game G_3 prove that the outcome of the game is $(1, 1, 1, 1)$.

References

[1] James M. Buchanan. *The Constitution of Economic Policy.* Lecture to the memory of Alfred Nobel, 1986.

[2] Shimon Even. *Graph Algorithms.* Cambridge University Press, 2011.

[3] Jonathan H. Hamilton and Steven M. Slutsky. Endogenous timing in duopoly games: Stackelberg or cournot equilibria. *Games and Economic Behavior*, 2(1):29–46, 1990.

[4] Jonathan H. Hamilton and Steven M. Slutsky. Endogenizing the order of moves in matrix games. *Theory and Decision*, 34(1):47–62, 1993.

[5] Leonid Hurwicz. But who will guard the guardians? *American Economic Review*, 98(3):577–585, 2008.

[6] Matthew O. Jackson. A crash course in implementation theory. *Social Choice and Welfare*, 18(4):655–708, 2001.

[7] Adam T. Kalai, Ehud Kalai, Ehud Lehrer, and Dov Samet. A commitment folk theorem. *Games and Economic Behavior*, 69:127–137, 2010.

[8] V. Krishna. *Auction Theory*. Academic Press, 2002.

[9] Roger B. Myerson. Learning from schelling's 'strategy of conflict'. *Journal of Economic Literature*, 47(4):1109–1125, 2009.

[10] Amnon Rapoport, William E Stein, James E Parco, and Thomas E Nicholas. Equilibrium play and adaptive learning in a three-person centipede game. *Games and Economic Behavior*, 43(2):239–265, 2003.

[11] Ludovic Renou. Commitment games. *Games and Economic Behavior*, 66(1):488–505, 2009.

[12] Robert Rosenthal. Games of perfect information, predatory pricing, and the chain store. *Journal of Economic Theory*, 25(1):92–100, 1981.

[13] Thomas C. Schelling. *The Strategy of Conflict*. Harvard U Press, 1960.

[14] M. Tennenholtz. Program equilibrium. *Games and Economic Behavior*, 49:363–373, 2004.

[15] Eric van Damme and Sjaak Hurkens. Commitment robust equilibria and endogenous timing. *Games and Economic Behavior*, 15(2):290–311, 1993.