Mixed Nash Equilibria in Concurrent Games

Patricia Bouyer, Nicolas Markey, Daniel Stan

April 2018

Laboratoire Spécification & Vérification

Mixed Nash Equilibria in Concurrent Games

Patricia Bouyer, Nicolas Markey, and Daniel Stan

LSV – CNRS & ENS Cachan – France

─── **Abstract** ───────────────────────────────────

We study mixed-strategy Nash equilibria in multiplayer deterministic concurrent games played on graphs, with terminal-reward payoffs (that is, absorbing states with a value for each player). We show undecidability of the existence of a constrained Nash equilibrium (the constraint requiring that one player should have maximal payoff), with only three players and 0/1-rewards (*i.e.*, reachability objectives). This has to be compared with the undecidability result by Ummels and Wojtczak for turn-based games which requires 14 players and general rewards. Our proof has various interesting consequences: (*i*) the undecidability of the existence of a Nash equilibrium with a constraint on the social welfare; (*ii*) the undecidability of the existence of an (unconstrained) Nash equilibrium in concurrent games with terminal-reward payoffs.

1 Introduction

Games (especially games played on graphs) have been intensively used in computer science as a powerful way of modelling interactions between several computerised systems [10, 7]. Until recently, more focus had been put on the study of purely antagonistic games (a.k.a. zero-sum games, where the aim of one player is to prevent the other player from achieving her objective), which conveniently model systems evolving in a (hostile) environment.

Over the last ten years, games with non-zero-sum objectives have come into the picture: they allow for conveniently modelling complex infrastructures where each individual system tries to fulfill its own objectives, while still being subject to actions of the surrounding systems. As an example, consider (a simplified version of) the team formation problem [3], an example of which is presented in Figure 1: several agents are trying to achieve tasks; each task requires some resources, which are shared by the players. Achieving a task thus requires the formation of a team that have all required resources for that task: each player selects the task she wants to achieve (and so proposes her resources for achieving that task), and if a task receives enough resources, the associated team receives the corresponding payoff (to be divided among the players in the team). In such a game, there is a need of cooperation (to gather enough resources), and an incentive to selfishness (to maximise the payoff).
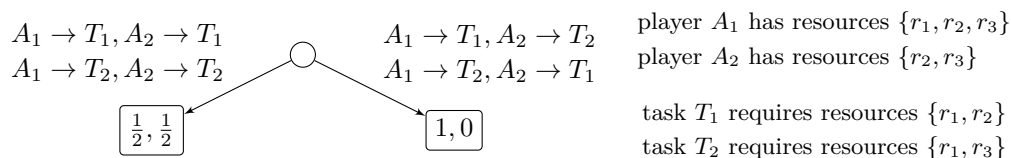

$A_1 \to T_1, A_2 \to T_1$ ◯ $A_1 \to T_1, A_2 \to T_2$ player A_1 has resources $\{r_1, r_2, r_3\}$
$A_1 \to T_2, A_2 \to T_2$ $A_1 \to T_2, A_2 \to T_1$ player A_2 has resources $\{r_2, r_3\}$

$\boxed{\frac{1}{2}, \frac{1}{2}}$ $\boxed{1, 0}$ task T_1 requires resources $\{r_1, r_2\}$
     task T_2 requires resources $\{r_1, r_3\}$

Figure 1 An instance of the team-formation problem. For any deterministic choice of actions, one of the players has an incentive to change her choice: there is no pure Nash equilibrium. However there is one mixed Nash equilibrium, where each player plays T_1 and T_2 uniformly at random.

In that setting, focusing only on optimal strategies for one single agent is not relevant. In game theory, several solution concepts have been defined, which more accurately represents rational behaviours of these multi-player systems; Nash equilibrium [8] is the most prominent such concept: a Nash equilibrium is a strategy profile (that is, one strategy to each player) where no player can improve her own payoff by unilaterally changing her strategy. In other terms, in a Nash equilibrium, each individual player has a satisfactory strategy with regards to the other players' strategies. Notice that Nash equilibria need not exist or be unique, and are not necessarily optimal: Nash equilibria where all players lose may coexist with more interesting Nash equilibria. Therefore, looking for constrained Nash equilibria (*e.g.* equilibria in which some players are required to win, or equilibria with maximal social welfare) is an interesting and important problem to study, which has been suggested both in the game-theory community [4] and in the computer-science community [11].

In this paper, we study (deterministic) concurrent games played on graphs. Such games are indeed a general and relevant model for interactive systems, where the agents take their decision simultaneously (which is the case for instance in distributed systems). Concurrent games subsume turn-based games, where in each state, only one player has the decision for the next move, and which have attracted more focus until now in the computer science community. Notice also that in game theory, models are almost exclusively based on concurrent actions (*e.g.* games in normal form given as matrices indicating the payoff of each player for each concurrent choice of actions, and extensions thereof, such as repeated games).

In this paper we are interested in randomized (a.k.a. mixed) strategies for the players. A mixed strategy consists in choosing, at each step of the game, a probability distribution over the set of available actions; the game then proceeds following the product distribution of the strategies of all players. Strategies may depend on the history of the game, *i.e.*, the sequence of visited states, but do not require to see the actions played by the other agents. In previous works, the first two authors have focused on pure strategies, where at each step, each player proposes exactly one action, and developed algorithms for deciding the existence of constrained Nash equilibria in various settings [1]. In the present paper, we focus on terminal-reward payoffs (where some designated states are absorbing, and each player has a value—or reward—attached to each of these states): the payoff of a player is then her expected reward. We will also consider the subclass of games with terminal-reachability objectives, where the reward in each absorbing state is either 0 or 1 (hence the expected reward for a player is the probability to reach her winning states). The game in Figure 1 has terminal-reward payoffs: they are given by the values labelling the two absorbing states (1 for player A_1 and 0 for player A_2 in the right-most state). This game can be shown to have no pure-strategy Nash equilibria, but it has a mixed-strategy one.

Our results. Our main result is the undecidability of the existence of a 0-optimal Nash equilibrium in concurrent games with terminal-reachability payoff functions, with only three players and strategies insensitive to actions. A 0-optimal Nash equilibrium is a Nash equilibrium in which one designated player is required to have maximal payoff (that is, 1 in the case of terminal-reachability payoffs). A corollary of our result is the undecidability of the existence of unconstrained Nash equilibria in concurrent games with terminal payoffs. We believe that these results are important, as they solve natural questions for basic objectives. Moreover, our constructions give new insight in the understanding of concurrent games and their algorithmics, and contain several intermediary tools that can be interesting on their own in different contexts.

Several results already exist in related settings:

- our result should first be compared with the undecidability of the existence of a 0-optimal Nash equilibrium in turn-based games with terminal-reward payoffs[12], which requires 14 players and general rewards. It should be noticed that this result requires more than 0/1 rewards (contrary to our result), since the existence of a 0-optimal equilibrium can be decided in polynomial time in turn-based games with terminal-reachability payoffs (by combining the reduction to pure 0-optimal Nash equilibria of [11] and the algorithm in [12] for computing such equilibria);
- our result should also be compared with polynomial-time algorithm for deciding the existence of a 0-optimal *pure* Nash equilibrium in concurrent games with terminal-reward payoffs [14];
- our result has several corollaries, that we develop at the end of the paper:
  - the existence of a (unconstrained) Nash equilibrium in terminal-reward games with three players; on the opposite, stationary ϵ-Nash equilibria do always exist in concurrent games for terminal-reachability (and terminal-reward) games [2];
  - the existence of a Nash equilibrium that maximizes the social welfare in games with terminal-reachability payoffs is undecidable with three players. This should be compared with the NP-completeness of the existence of such equilibria for two-player normal-form games [5];
  - the existence of a constrained finite-memory Nash equilibrium in terminal-reachability games is undecidable with three players;
  - the existence of a constrained Nash equilibrium in safety games is undecidable with three players. This can be compared to the result of [9], which states that there always exists a Nash equilibrium (with little memory) in a safety game.

For the sake of readability, all proofs have been put in a separate appendix at the end of the paper.

2 Definitions

▶ **Definition 1.** A concurrent arena $\mathcal{A}$ is a tuple $\mathcal{A} = \langle \mathsf{States}, \mathsf{Agt}, \mathsf{Act}, \mathsf{Tab}, (\mathsf{Allow}_i)_{i \in \mathsf{Agt}} \rangle$ where States is a finite set of states; Agt is a finite set of players; Act is a finite set of actions; For all $i \in \mathsf{Agt}$, $\mathsf{Allow}_i \colon \mathsf{States} \longrightarrow 2^{\mathsf{Act}} \backslash \{\varnothing\}$ is a function describing authorized actions in a given state for Player i; $\mathsf{Tab} \colon \mathsf{States} \times \mathsf{Act}^{\mathsf{Agt}} \to \mathsf{States}$ is the transition function.

A state $s \in \mathsf{States}$ is said *terminal* (or *final*) if $\mathsf{Tab}(s, \cdot) \equiv s$. We write $\mathsf{F}_{\mathcal{A}}$ (or simply F when the underlying arena is clear from the context) for the set of terminal states of $\mathcal{A}$.

A *history* of such an arena $\mathcal{A}$ is a finite, non-empty word $h \in \mathsf{States}^+$. We denote by $\mathsf{first}(h)$ and $\mathsf{last}(h)$ respectively the first and last states of the word h. During a play, players in Agt choose their next moves concurrently and independently from each others, according to the current history h and what they are allowed to do in the current state $\mathsf{last}(h)$.

▶ **Definition 2.** A *strategy* for Player i is a function $\sigma_i \colon \mathsf{States}^+ \to \mathsf{Dist}(\mathsf{Act})$ with the requirement that $\sigma_i(h)^{-1}(\mathbb{R}_{>0}) \subseteq \mathsf{Allow}_i(\mathsf{last}(h))$ for all history h.

Let $\alpha \in \mathsf{Act}$. We write $\sigma_i(\alpha \mid h)$ for the probability mass $\sigma_i(h)(\alpha)$ of action α in the distribution $\sigma_i(h)$. In the sequel, we sometimes write $\sigma_i(h) = \alpha$ when $\sigma_i(\alpha \mid h) = 1$. When $\sigma_i(h) \in \mathsf{Act}$ for all h, the strategy σ_i for player i is said to be *pure*. Otherwise it is said to be *mixed*. We denote by S_i (resp. $\$_i$) the set of pure (resp. mixed) strategies of Player i. A *strategy profile* σ is a mapping assigning one strategy to each player. We write $\$$ for the set of all strategy profiles, and for $\sigma \in \$$, we will write σ_i in place of $\sigma(i)$ for the strategy of Player i.

▶ Remark. While strategies are aware of the sequence of actions played in a turn-based game, we can notice this is generally not the case in the concurrent setting depicted here, since strategies only depend on the sequence of visited states. This is realistic when considering multi-agent systems, where only the global effect of the actions of the players is assumed to be observable. However this partial-information hypothesis makes the detection of strategy deviations (and therefore the computation of Nash equilibria) harder.

Consider a strategy profile $\sigma \in \mathbb{S}$ and an initial state s_0. For any history $h \in \mathsf{States}^+$ and any player $i \in \mathsf{Agt}$, we construct the random variable $\alpha_i(h) \in \mathsf{Act}$ with distribution $\sigma_i(h)$ such that $(\alpha_i(h))_{i \in \mathsf{Agt}, h \in \mathsf{States}^+}$ is a family of independent random variables.

We define the stochastic process $(X_n)_{n \in \mathbb{N}}$ inductively by $X_0 = s_0$ and for every n, $X_{n+1} = X_n \cdot \mathsf{Tab}\left(\mathsf{last}(X_n), (\alpha_i(X_n))_i\right)$. For each n, the random variable X_n takes value in States^{n+1}: $(X_n)_n$ is an increasing sequence of prefixes whose limit is an infinite random run $X_\infty \in \mathsf{States}^\omega$.

We now consider the standard Borel σ-algebra over States^ω from s_0, and define the probability measure $\mathbb{P}^\sigma$ as the probability distribution induced by X_∞, that is, if B is a Borel subset of States^ω, $\mathbb{P}^\sigma(B) = \mathbb{P}(X_\infty \in B)$. It coincides with the standard construction based on cylinders. In the following, to make explicit the initial state, we may write $\mathbb{P}^\sigma(B \mid s_0)$ instead of simply $\mathbb{P}^\sigma(B)$. In the sequel, we sometimes also abusively write h for the cylinder $h \cdot \mathsf{States}^\omega$: then, when we write $\mathbb{P}^\sigma(h \mid s_0)$, we mean $\mathbb{P}^\sigma(X_{|h|} = h)$. If $\mathbb{P}^\sigma(h \mid s_0) > 0$, we say that σ enables h from s_0: in that case we can define the conditional probability $\mathbb{P}^\sigma(B \mid h) = \mathbb{P}^\sigma(B \mid X_{|h|} = h)$.

Finally we say that a node n is activated by a strategy profile whenever it is visited with positive probability under that profile.

▶ **Definition 3.** A *terminal-reward game* $\mathcal{G} = \langle \mathcal{A}, s, (\phi_i)_{i \in \mathsf{Agt}} \rangle$ is given by an arena $\mathcal{A}$, an initial state s, and for every player $i \in \mathsf{Agt}$, a real-valued function ϕ_i ranging over terminal states of $\mathcal{A}$. In the following, we extend ϕ_i to every $r \in \mathsf{States}^\omega$, by $\phi_i(r) = \phi_i(s)$ if r is an infinite path ending in a state $s \in \mathsf{F}$, and $\phi_i(r) = 0$ otherwise.

The game $\mathcal{G}$ will be said a *terminal-reachability game* whenever each function ϕ_i only takes values 0 or 1.

▶ Remark. In the sequel, we represent terminal-reward games as graphs with circle states representing non-terminal states, and rectangle states representing terminal states, decorated with the associated rewards for all players. The self-loop on terminal states will be omitted. The transition table of the underlying arena is encoded by decorating the transitions with the move vectors that trigger it. Move vectors are written as words over Act, by identifying Agt with the subset $[\![0, |\mathsf{Agt}| - 1]\!]$. We will use $\cdot$ as a special symbol representing any action. Also, for a set S of words in $(\mathsf{Act} \cup \{\cdot\})^k$, with $k < |\mathsf{Agt}|$, and for a letter $a \in \mathsf{Act} \cup \{\cdot\}$, we write aS for the words $\{aw \mid w \in S\}$. See Figure 2 (and the subsequent figures) for an example.

Consider a terminal-reward game $\mathcal{G}$, a strategy profile σ, and an enabled history h. One can easily check that ϕ_i is a mesurable function under $\mathbb{P}^\sigma$. The expected payoff of Player i under σ after h is defined as

$$\mathbb{E}^\sigma(\phi_i \mid h) = \sum_{x \in \mathsf{Img}(\phi_i)} x \cdot \mathbb{P}^\sigma\left(\phi_i^{-1}(\{x\}) \mid h\right).$$

In case $\mathcal{G}$ is a terminal-reachability game, the expected payoff of Player i is the probability of reaching terminal states with value 1 under ϕ_i.

Let $\mathcal{G}$ be a terminal-reward game. Let $\sigma \in \$$ be a (mixed) strategy profile in $\mathcal{G}$, and h be a history. A *single-player deviation* (simply called *deviation* hereafter, as we only consider deviations of a single player at a time) of σ for Player i after history h is another strategy profile σ' for which there exists $\sigma_i'' \in \$_i$ satisfying

$$\forall h' \in \mathsf{States}^+. \begin{cases} h \sqsubseteq h' \Rightarrow \sigma_i'(h') = \sigma_i''(h') \\ \forall j \in \mathsf{Agt}. \ (h \not\sqsubseteq h' \vee j \neq i) \Rightarrow \sigma_j'(h') = \sigma_j(h') \end{cases}$$

where $\sqsubseteq$ is the prefix relation. We then write $\sigma' = \sigma[i/\sigma_i'']^h$. The deviation $\sigma[i/\sigma_i'']^h$ is said *deterministic* if σ_i'' is.

▶ **Definition 4.** Let $\mathcal{G}$ be a terminal-reward game. A strategy profile σ forms a *Nash equilibrium* after a history h when the following conditions are met:

- $h \in \mathsf{States}^+$ is enabled by σ from $\mathsf{first}(h)$;
- No player has a profitable deviation; in other terms, for all $i \in \mathsf{Agt}$ and for all $\sigma_i' \in \$_i$, it holds $\mathbb{E}^{\sigma[i/\sigma_i']^h}(\phi_i \mid h) \leq \mathbb{E}^\sigma(\phi_i \mid h)$.

We then write that $\langle \sigma, h \rangle$ is a Nash equilibrium.

A Nash equilibrium $\langle \sigma, h \rangle$ is said 0-*optimal* whenever the expected payoff of Player 0 is optimal, that is, $\mathbb{E}^\sigma(\phi_i \mid h) = \max(\mathsf{Img}(\phi_0))$. In case of a terminal-reachability game, it amounts to saying that the payoff of Player 0 is 1.

The following result will be useful all along the paper:

▶ **Lemma 5.** *Let $\mathcal{G}$ be a terminal-reward game, and $\langle \sigma, h \rangle$ be a Nash equilibrium. If $\langle \sigma, h \rangle$ enables h', then $\langle \sigma, h' \rangle$ is a Nash equilibrium.*

In general, several Nash equilibria may coexist (see *e.g.* Figure 12a). It is therefore very relevant to look for *constrained Nash equilibria*, that is, Nash equilibria that satisfy a constraint on the expected payoff. In this paper, we only consider 0-optimality as the constraint, and we prove that the existence of a 0-optimal Nash equilibrium in a three-player terminal-reachability game is undecidable. To prove this result, we will first show undecidability in the case of terminal-reward games, and then extend the result to terminal-reachability games. Those results will have interesting corollaries, like the undecidability of the existence of a Nash equilibrium (with no constraint) in terminal-reward games, when the rewards are in $\{-1, 0, 1\}$, or the existence of a Nash equilibrium with optimal social welfare.

3 Tools

In this section, we develop several intermediary results that will be useful for our reduction. We first show that we can equivalently define Nash equilibria by considering only deterministic deviations (for non-negative terminal-reward games). We then study a few simple games and constructions which will be used in the encoding.

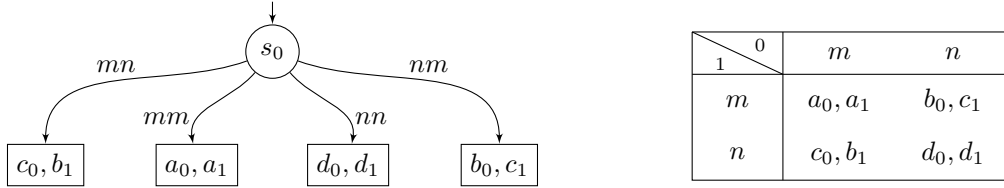### 3.1 Deterministic deviations

We explain in this section that it is enough to consider deterministic deviations in the characterization of a Nash equilibrium.

▶ **Proposition 6.** *Let $\mathcal{G}$ be a terminal-reward game with non-negative rewards. Pick a history $h \in \mathsf{States}^+$, and a strategy profile σ. Then $\langle \sigma, h \rangle$ is a Nash equilibrium if, and only if, for all $i \in \mathsf{Agt}$ and all deterministic deviation $\sigma_i'' \in S_i$, it holds $\mathbb{E}^{\sigma[i/\sigma_i'']^h}(\phi_A \mid h) \leq \mathbb{E}^\sigma(\phi_i \mid h)$.*

▶ Remark. A similar result was proven in [14, Proposition 3.1] for turn-based games with qualitative Borel objectives (the payoff is 1 if the run belongs to the designed objective, and it is 0 otherwise).

3.2   One-stage games

We analyse two-player two-action one-stage games (that is, games that end up in a terminal state in one step), and obtain useful properties of their Nash equilibria. Such games can be represented by a graph as shown in Figure 2a. Alternatively, these games, also known as one-shot games, can be represented as a matrix as in Table 2b (this is the standard representation in the game-theory community).

1＼0	m	n
m	a_0, a_1	b_0, c_1
n	c_0, b_1	d_0, d_1

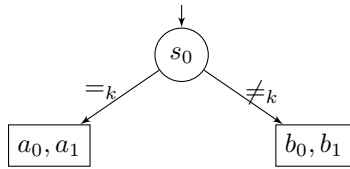**(a)** A generic two-player two-action one-stage game   **(b)** Associated matrix representation

Figure 2 Representations of a one-shot game

▶ **Lemma 7.** *Consider the two-player two-action one-shot concurrent game $\mathcal{G}$ of Figure 2, and pick some strategy profile σ. If $\langle \sigma, s_0 \rangle$ is a Nash equilibrium, then for every player $i \in \{0, 1\}$, it holds*

$$\begin{cases} \sigma_i(m \mid s_0) < 1 & \Rightarrow & [(d_i - c_i) + (a_i - b_i)] \cdot \sigma_{1-i}(m \mid s_0) \leq d_i - c_i \\ \sigma_i(m \mid s_0) > 0 & \Rightarrow & [(d_i - c_i) + (a_i - b_i)] \cdot \sigma_{1-i}(m \mid s_0) \geq d_i - c_i \end{cases}$$
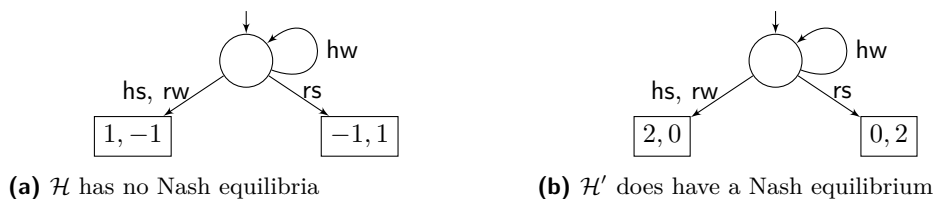
3.3 k-action matching-pennies games

Figure 3 Matching-pennies game

The classical matching-pennies games are a special case of one-stage games, where $a_i = d_i$ and $b_i = c_i$: basically, there are two outcomes, depending on whether the players propose the same action or not. This game can be generalized to k (≥ 2) actions, as depicted on Figure 3. In this figure (and in the sequel), $=_k$ (resp. $\neq_k$) is a shorthand for pairs of identical (resp. different) actions taken from a set of k actions $\Sigma_k = \{c_1, \ldots, c_k\}$. In other terms, $=_k$ represents the set of words $\{c_i c_i \mid 1 \leq i \leq k\}$, and $\neq_k$ is the complement in Σ_k^2.
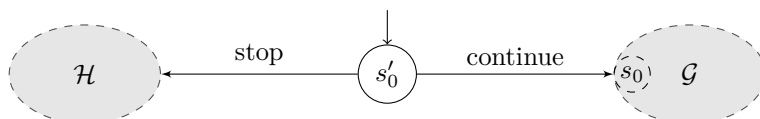
▶ **Lemma 8.** *In the k-action matching-pennies game, playing uniformly at random for both players defines a Nash equilibrium. Moreover, this is the unique Nash equilibrium of the game if, and only if, either $a_0 < b_0$ and $a_1 > b_1$, or $a_0 > b_0$ and $a_1 < b_1$. The payoff of this Nash equilibrium is $\left(\frac{1}{k} \cdot a_0 + \left(1 - \frac{1}{k}\right) \cdot b_0, \frac{1}{k} \cdot a_1 + \left(1 - \frac{1}{k}\right) \cdot b_1\right)$.*

3.4   Games without equilibrium

In this section, we show that there are games that admit no Nash equilibria. We then explain how these games can be used to impose constraints on payoffs.

(a) $\mathcal{H}$ has no Nash equilibria   **(b)** $\mathcal{H}'$ does have a Nash equilibrium

■ **Figure 4** Hide-or-run games

■ **Figure 5** A game that has a Nash equilibrium if, and only if, $\mathcal{G}$ has a 0-optimal Nash equilibrium

Consider the game *hide-or-run*, depicted in Figure 4a. Player 0 can either hide (h) or run home (r), while Player 1 can either shoot him (s), or wait (w). If Player 1 shoots while Player 0 is hiding, she loses her bullet and loses the game. If Player 1 shoots when Player 0 is running, she wins. This game has been shown to have no optimal almost-sure strategy [6], and we adapt the proof to show that it has no Nash equilibria.

▶ **Lemma 9.** *The game $\mathcal{H}$ has no Nash equilibria.*

The payoff function of $\mathcal{H}$ takes negative value. In order to only have nonnegative payoffs, we could shift the values by 1, which yields the game $\mathcal{H}'$ depicted on Figure 4b. But then one easily sees that the strategies $\sigma_0(\mathsf{h} \mid s_0^n) = 1$ and $\sigma_1(\mathsf{s} \mid s_0^n) = 1$ form a Nash equilibrium, contrary to a claim in [2, 12]. The difference is that when shifting the payoffs, we did not modify the payoff of the run that never reaches a terminal state: while this run was a positive deviation for Player 1 in $\mathcal{H}$, this is not the case in $\mathcal{H}'$ anymore.

We now explain how we use the game $\mathcal{H}$ to impose a 0-optimality constraint on the payoff. In the sequel, we restrict[1] to games where $\max_{s \in \mathsf{F}} \phi_0(s) = 1$. Then:

▶ **Lemma 10.** *Let $\mathcal{G}$ be a terminal-reward game. Then we can build a terminal-reward game $\mathcal{G}'$ (see Figure 5) such that $\mathcal{G}$ has a 0-optimal Nash equilibrium if, and only if, $\mathcal{G}'$ has a Nash equilibrium.*
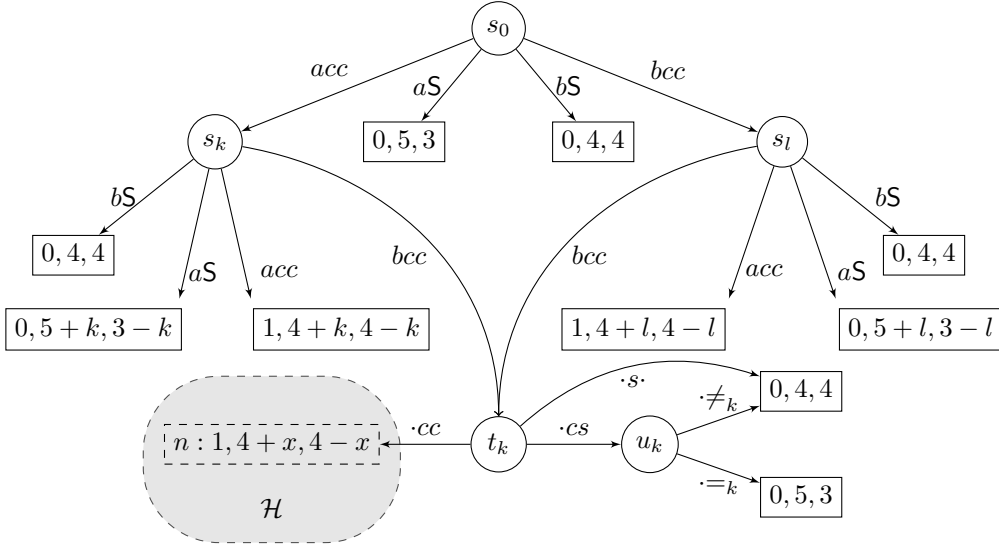
This lemma will be useful for extending the undecidability result from the constrained existence to the existence problem (Corollary 14).

▶ Remark. Note that in the above construction, game $\mathcal{H}$ can be replaced by any game with no Nash equilibria, such that Player 0 can secure a payoff $1 - \varepsilon$ for every $\varepsilon > 0$. For instance, one could use a game with limit-average payoff and nonnegative rewards only [13].

4   Updating values

Our undecidability proof will be based on an encoding of a two-counter machine. In this section and in the next one, we present games that will be building blocks for our proof.

---

[1] This is no loss of generality, since we can linearly rescale payoffs of each player without changing profitable deviations.

Figure 6 The rescale game $\mathcal{G}_k^r$ for $k \in \{1, 2, 3\}$ and $l = k - 1$

Consider the game $\mathcal{G}_k^r$ depicted on Figure 6: in this game, Player 0 has two available actions a and b from s_0, s_k and s_l, while the other two players can either continue (action c), or unilateraly decide to stop the game (action s) and go to a terminal state (where Player 0 will have payoff 0). In node t_k, only players 1 and 2 have a choice: they can either continue to the game $\mathcal{H}$ (when both of them play c), or decide to stop and go to a k-action matching-pennies game (when one of them plays s). In Figure 6, we write S as a shorthand representing any combination of moves of players 1 and 2 where at least one of them decides to stop (action s). Node n is the initial node of a game $\mathcal{H}$ (which is unknown for the moment).

The interesting property of game $\mathcal{G}_k^r$ is that we can relate 0-optimal Nash equilibria from s_0 and those from n: (roughly) there is a Nash equilibrium from n of expected payoff $(1, 4 + x, 4 - x)$ if, and only if, there is a Nash equilibrium from s_0 of expected payoff $(1, 4 + k \cdot x, 4 - k \cdot x)$.
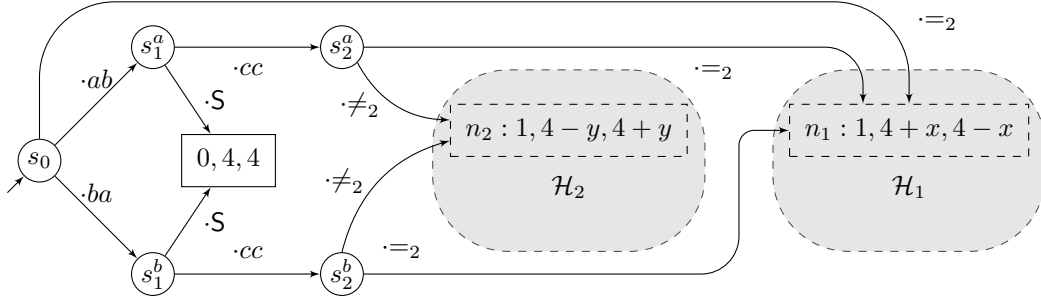
This is because, from s_0 and s_k, there is a threat for Player 0 that one of the players 1 and 2 stops the game immediately, leading to a state with payoff 0 for her. Hence, Player 0 is forced to "collaborate" with players 1 and 2 and help them be satisfied with their payoffs, either by joining one of the interesting terminal states of $\mathcal{G}_k^r$, or in the next game $\mathcal{H}$ after n. Some technical calculations show that Player 0 has to play a with probability $k \cdot x$ at s_0, and with probability $x/(x + 1)$ at s_k and s_l. The gadget to the right of t_k is just for ensuring that $0 \le k \cdot x \le 1$ (this condition is required for having the above-mentioned equivalence between Nash equilibria from s_0 and Nash equilibria from n).

5 Comparing values

5.1 Testing game

We present in this section the construction of a game for comparing the expected payoffs in different nodes. This will be useful in our reduction to encode the zero-tests of our two-counter machine.

Consider the game $\mathcal{G}_t$ depicted on Figure 7. This game has the very interesting property that if we assume there are 0-optimal Nash equilibria from n_1 and n_2 of respective payoffs

Figure 7 The testing game $\mathcal{G}_t$

$(1, 4 + x, 4 - x)$ and $(1, 4 - y, 4 + y)$, then there is a 0-optimal Nash equilibrium from s_0 if, and only if, $x = y$, and the payoff is then $(1, 4 + x/2, 4 - x/2)$. Indeed, unless $x = 0$ or $y = 0$, it should be the case that a 0-optimal Nash equilibrium activates all states s_j^α in the game, and then, as players 1 and 2 have zero-sum objectives, the best way is to play uniformly at random in all states where this makes sense (when actions a and b are available), and to play deterministically action c in all states where c is available.

This gadget allows, by plugging in n_2 a game with known payoffs (the games on the next subsection), to check that the payoff at s_0 has some particular value (which depends on that after n_1).
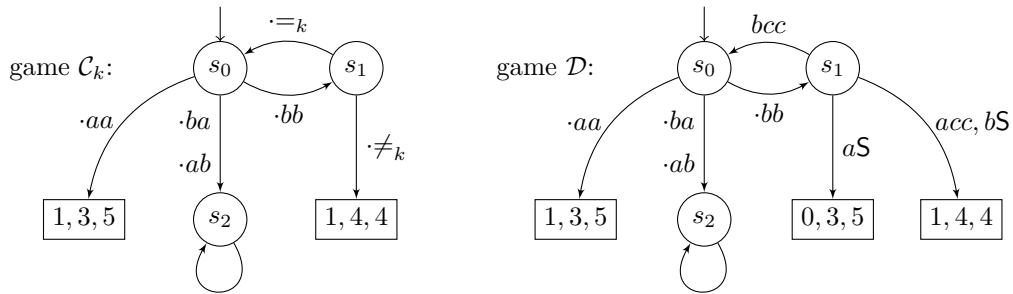
5.2 Counting modules

We now present games that generate a family of Nash equilibria with a particular expected payoffs. These modules will later be plugged at node n_2 of game $\mathcal{G}_t$, and will ensure that the payoff of an Nash equilibrium in $\mathcal{G}_t$ will have a predefined form.

▶ **Lemma 11.** *Consider the games of Figure 8. For $n \in \mathbb{N} \cup \{+\infty\}$, define*

$$r_k(n) = \left(1, 4 - \frac{1}{k^n}, 4 + \frac{1}{k^n}\right) \qquad\qquad s(n) = \left(1, 4 - \frac{1}{n+1}, 4 + \frac{1}{n+1}\right)$$

Then the set of 0-optimal Nash equilibrium values is $\{s(n) \mid n \in \mathbb{N} \cup \{\infty\}\}$ in $\mathcal{D}$, and $\{r_k(n) \mid n \in \mathbb{N} \cup \{\infty\}\}$ in $\mathcal{C}_k$ for all $k \geq 2$.

Figure 8 The modules $\mathcal{C}_k$ (for $k \geq 2$) and $\mathcal{D}$. Notice that state s_2 should be considered terminal, as it only carries a self-loop. We could replace it by a two-state loop. We could also see it as a terminal state with reward $(0, 0, 0)$, but for the proof of Corollary 16, we want the terminal rewards of players 1 and 2 to always sum to 8, which we could not achieve easily in this case.

(a) Input gadget $\mathcal{G}_{\mathcal{M}}^{\mathsf{init}}$

(b) Game $\mathcal{G}_{\mathcal{M}}^{q}$ (where $\{\delta_1, ..., \delta_n\}$ is the set Δ^q of transitions leaving q)

(c) Game $\mathcal{G}_{\mathcal{M}}^{\delta}$ when $\delta = (q, dec(k), q')$

(d) Game $\mathcal{G}_{\mathcal{M}}^{\delta}$ when $\delta = (q, inc(k), q')$

(e) Game $\mathcal{G}_{\mathcal{M}}^{\delta}$ when $\delta = (q, zero(k), q')$

(f) Game $\mathcal{G}_{\mathcal{M}}^{\delta}$ when $\delta = (q, !zero(k), q')$

Figure 9 Description of the subgames $\mathcal{G}_{\mathcal{M}}^{q}$ and $\mathcal{G}_{\mathcal{M}}^{\delta}$

6   Undecidability proof

We now turn to the global undecidability proof of the constrained-existence problem in three-player games. The complete proof is given in Appendix D. The proof is a reduction from the recurring problem of a two-counter machine. We encode the behaviour of a two-counter machine $\mathcal{M}$ as a concurrent game $\mathcal{G}_{\mathcal{M}}$, which connects the various subgames depicted on Figures 9 (one initial gadget, one per state q, one per transition δ). Roughly, this game will encode a configuration (q, c_1, c_2) of $\mathcal{M}$ using a Nash equilibrium $\sigma \in \$$ from q such that $\mathbb{E}^{\sigma}(\phi \mid s) = \left(1, 4 + \frac{1}{2^{c_1} 3^{c_2}}, 4 - \frac{1}{2^{c_1} 3^{c_2}}\right)$ (property $P(q, c_1, c_2)$). Using the various constructions we have made previously, we can show that if $P(q, c_1, c_2)$ is satisfied, then there is a transition $(q, c_1, c_2) \to_{\delta} (q', c_1', c_2')$ in $\mathcal{M}$ such that $P(q', c_1', c_2')$ is satisfied as well, which allows to progress 'along' a Nash equilibrium while building a computation in $\mathcal{M}$.
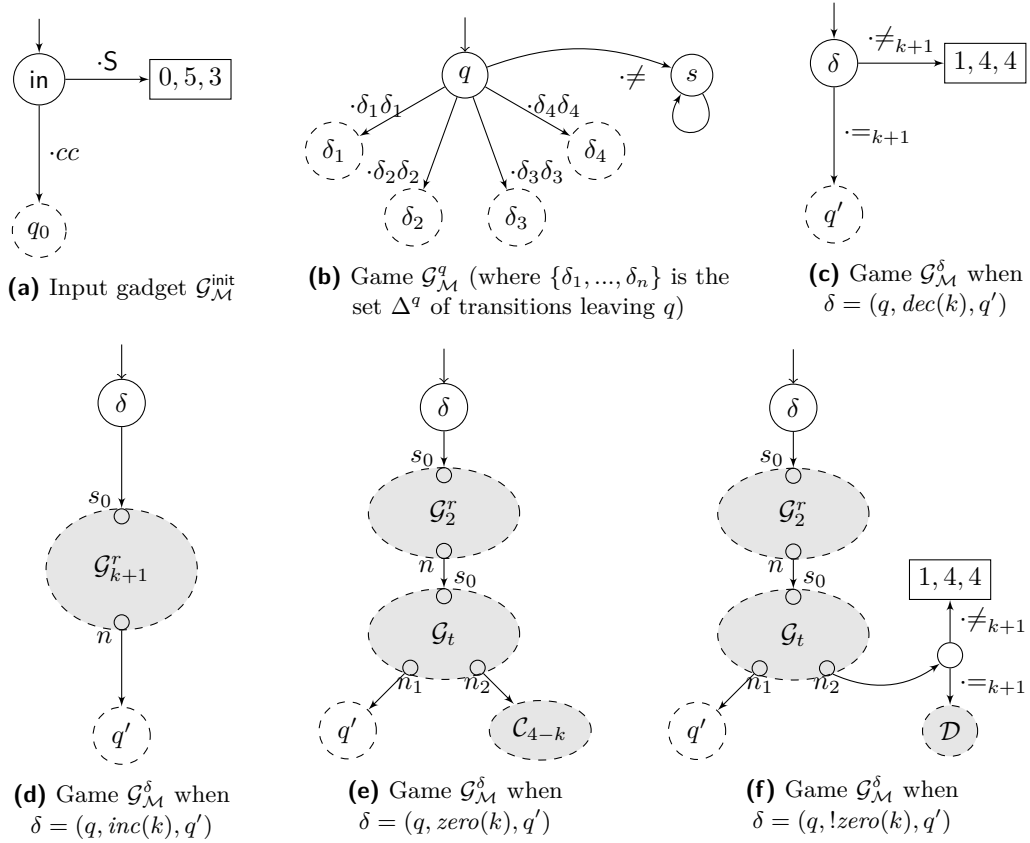
The correspondence between $\mathcal{M}$ and $\mathcal{G}_{\mathcal{M}}$ is made precise as follows:

▶ **Proposition 12.** *The two-counter machine $\mathcal{M}$ has an infinite valid computation if, and only if, there is a 0-optimal Nash equilibrium from state* in *in game $\mathcal{G}_{\mathcal{M}}$.*

This immediately entails:

▶ **Theorem 13.** *We cannot decide whether there exists a 0-optimal Nash equilibrium in three-player games with non-negative terminal-reward payoffs.*

We now consider several extensions of this result. We first state two straightforward
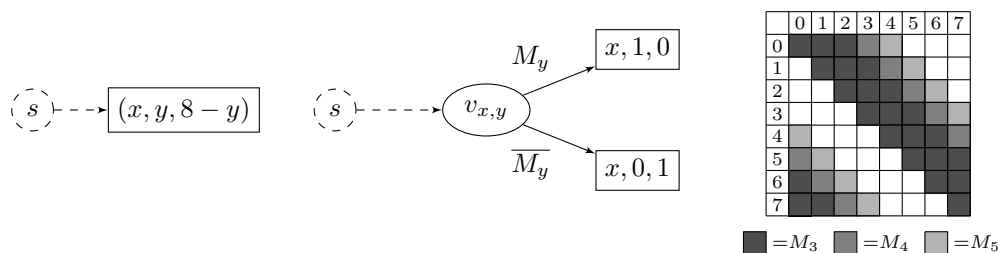
Figure 10 Transformation of a terminal node $(x, y, 8 - y)$ with an intermediate node $v_{x,y}$. The table on the right gives the value of M_y for some values of y (notice that $M_y \subseteq M_{y'}$ when $y \le y'$).

corollaries. First, applying Lemma 10, we can enforce the 0-optimality constraint in the game by inserting an initial module. It follows:

▶ **Corollary 14.** *We cannot decide whether there exists a Nash equilibrium in three-player games with (possibly negative) terminal-reward payoffs.*

Now we realize that in this reduction, there is a 0-optimal Nash equilibrium from in if, and only if, there is a Nash equilibrium with social welfare larger than or equal to 9, where the social welfare is defined as the sum of the expected payoffs of all players. As an immediate corollary, we get:

▶ **Corollary 15.** *We cannot decide whether there exists a Nash equilibrium with some lower bound on the social welfare (or with optimal social welfare) in three-player terminal-reward games with non-negative payoffs.*

We now explain briefly how the main theorem can be extended to terminal-reachability payoffs (the details are given in Appendix E). We indeed realize that the payoffs of players 1 and 2 always sum up to 8 in the reduction (the game between those two players is zero-sum). The idea is then to replace each terminal state with a simple gadget in which the payoffs of players 1 and 2 are $(8, 0)$ or $(0, 8)$, and to use an adequate set of actions which decomposes runs into two sets with proportions mimicking the normal rewards of the terminal state. For instance, for a reward $(x, y, 8 - y)$, the set of actions $M_y = \{\cdot ij \mid \exists 0 \le r < y.\ i - j = r \mod 8\}$ will lead to $(x, 8, 0)$ and its complement to $(x, 0, 8)$, as illustrated on Figure 10. In the game from $v_{x,y}$, there is a unique Nash equilibrium which consists in playing uniformly at random for both players, yielding a payoff of $(x, y, 8 - y)$. It remains to normalize and replace each $(x, 8, 0)$ (resp. $(x, 0, 8)$) by $(x, 1, 0)$ (resp. $(x, 0, 1)$).

▶ **Corollary 16.** *We cannot decide whether there exists a* 0-*optimal Nash equilibrium in three-player games with terminal-reachability payoffs.*

Finally, (roughly) by dualizing reachability and safety conditions, we can prove that the constrained existence of Nash equilibrium in safety games cannot be decided (details in Appendix F). This result is interesting since [9] that there always exists a Nash equilibrium in safety games (called stay-in-a-set games in [9]).

▶ **Corollary 17.** *We cannot decide whether there exists a Nash equilibrium in three-player safety games with payoff* 0 *assigned to Player* 0.

7   Conclusion and future work

In this paper we have shown the undecidability of the existence of a constrained Nash equilibrium in a three-player concurrent game with terminal-reachability objectives. We believe

this result is surprising, since it applies to very simple payoff functions, and with very few players. This result has to be compared with the undecidability result of [12], which on one hand, applies to turn-based games, but requires 14 players and the full power of terminal-reward payoffs. Furthermore, in turn-based games with terminal-reachability payoffs, constrained Nash equilibria can be computed (in polynomial time) through a reduction to pure Nash equilibria [11] and algorithms for computing pure Nash equilibria [12]. We have also mentioned a couple of interesting corollaries that we do not repeat here.

This work lets open the decidability status of the constrained-existence problem in two-player games with terminal-reward and terminal-reachability payoffs. In fact, even the existence of Nash equilibria in such games is an open problem: it was believed until recently that there are two-player games with nonnegative terminal rewards having no Nash equilibrium [2, 12], but the proposed example was actually wrong (as we explained in Section 3.4). If one can find such a game with no Nash equilibrium, then our Corollary 14 extends to nonnegative terminal-reward games, and possibly to terminal-reachability games. Notice that two-player games have been studied quite a lot in the literature, and we know for instance that (uniform) ϵ-Nash equilibria always exist in terminal-reward games [15, 16].

References

1 P. Bouyer, R. Brenguier, N. Markey, and M. Ummels. Pure Nash equilibria in concurrent games. *Logical Methods in Computer Science*, 2014. Submitted, under revision.

2 K. Chatterjee, M. Jurdziński, and R. Majumdar. On Nash equilibria in stochastic games. In *CSL'04*, LNCS 3210, p. 26–40. Springer, 2004.

3 T. Chen, M. Kwiatkowska, D. Parker, and A. Simaitis. Verifying team formation protocols with probabilistic model checking. In *CLIMA XII*, LNAI 6814, p. 190–207. Springer, 2011.

4 V. Conitzer and T. Sandholm. Complexity results about Nash equilibria. In *IJCAI'03*, p. 765–771, 2003.

5 V. Conitzer and T. Sandholm. New complexity results about Nash equilibria. *Games and Economic Behavior*, 63(2):621–641, 2008.

6 L. de Alfaro, T. Henzinger, and O. Kupferman. Concurrent reachability games. *Theoretical Computer Science*, 386(3):188–217, 2007.

7 T. A. Henzinger. Games in system design and verification. In *TARK'05*, p. 1–4, 2005.

8 J. F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences of the United States of America*, 36(1):48–49, 1950.

9 P. Secchi and W. D. Sudderth. Stay-in-a-set games. *International Journal of Game Theory*, 30:479–490, 2001.

10 W. Thomas. Infinite games and verification. In *CAV'02*, LNCS 2404, p. 58–64. Springer, 2002. Invited Tutorial.

11 M. Ummels. The complexity of Nash equilibria in infinite multiplayer games. In *FoSSaCS'08*, LNCS 4962, p. 20–34. Springer, 2008.

12 M. Ummels and D. Wojtczak. The complexity of Nash equilibria in limit-average games. In *CONCUR'11*, LNCS 6901, p. 482–496. Springer, 2011.

13 M. Ummels and D. Wojtczak. The complexity of Nash equilibria in limit-average games. Technical Report abs/1109.6220, CoRR, 2011.

14 M. Ummels and D. Wojtczak. The complexity of Nash equilibria in stochastic multiplayer games. *Logical Methods in Computer Science*, 7(3), 2011.

15 N. Vieille. Equilibrium in two-person stochastic games I: A reduction. *Israel Journal of Mathematics*, 119:55–91, 2000.

16 N. Vieille. Equilibrium in two-person stochastic games II: The case of recursive games. *Israel Journal of Mathematics*, 119:93–126, 2000.

A   Complements for Section 3

A.1   Deterministic deviations

The proof of Proposition 6 requires a first lemma.

▶ **Lemma 18.** *Let $\mathcal{G}$ be a terminal-reward game with non-negative rewards, and $\sigma \in \mathbb{S}$. Let $i \in \mathsf{Agt}$, h be an enabled history, and $\varepsilon > 0$. There exists a deterministic deviation σ_i'' of Player i such that $\sigma' = \sigma[i/\sigma_i'']$ satisfies $\mathbb{E}^{\sigma'}(\phi_i \mid h) \geq \mathbb{E}^{\sigma}(\phi_i \mid h) - \varepsilon$.*

Proof. Let N be the random variable representing the number of steps before reaching a final state along an infinite outcome of σ from h. If $\mathbb{P}^{\sigma}(N < +\infty \mid h) = 0$, then $\mathbb{E}^{\sigma}(\phi_i \mid h) = 0$, and the result holds (any deterministic deviation will work).

Assume now that $\mathbb{P}^{\sigma}(N < +\infty \mid h) > 0$. Since $\mathbb{P}^{\sigma}$ is a probability measure, it holds $1 = \mathbb{P}^{\sigma}(N < +\infty \mid h, N < +\infty) = \sup_n \mathbb{P}^{\sigma}(N < n \mid h, N < +\infty)$. For a fixed $\varepsilon' > 0$, we can choose $n > 0$ such that $\mathbb{P}^{\sigma}(N < n \mid h, N < +\infty) \geq 1 - \varepsilon'$. Hence,

$$\mathbb{P}^{\sigma}(N \geq n \mid h, N < +\infty) \leq \varepsilon'$$

and

$$0 \leq \mathbb{E}^{\sigma}(\phi_i \cdot \mathbb{1}_{N \geq n} \mid h, N < +\infty) \leq (\max \phi_i) \cdot \varepsilon' \leq \frac{\varepsilon}{\mathbb{P}^{\sigma}(h, N < +\infty)}$$

by taking $\varepsilon' \leq \frac{\varepsilon}{(\max \phi_i) \cdot \mathbb{P}^{\sigma}(N < +\infty \mid h)}$. Now we can decompose $\mathbb{E}^{\sigma}(\phi_i \cdot \mathbb{1}_{N \geq n} \mid h, N < +\infty)$ as follows:

$$\mathbb{E}^{\sigma}(\phi_i \cdot \mathbb{1}_{N \geq n} \mid h) = \mathbb{E}^{\sigma}(\phi_i \cdot \mathbb{1}_{N \geq n} \mid h, N < +\infty) \cdot \mathbb{P}^{\sigma}(N < +\infty \mid h) \leq \varepsilon$$

since all runs with $N = +\infty$ have payoff 0.

For this fixed n, we have,

$$\mathbb{E}^{\sigma}(\phi_i \cdot \mathbb{1}_{N < n} \mid h) = \sum_{f \,:\, h\mathsf{States}^{<n} \to \mathsf{Act}} \mathbb{E}^{\sigma}(\phi_i \cdot \mathbb{1}_{N < n} \mid h, \forall h'.\, \alpha_i(h') = f(h')) \cdot \mathbb{P}^{\sigma}(\forall h'.\, \alpha_i(h') = f(h')).$$

which is an average sum over the finite set

$$\left\{ \mathbb{E}^{\sigma}(\phi_i \cdot \mathbb{1}_{N < n} \mid h, \forall h' \, \alpha_i(h') = \sigma_i'(h')) \mid f : h\mathsf{States}^{<n} \to \mathsf{Act} \right\}.$$

Hence, one of the maximal value v is reached for some function f with $\mathbb{P}^{\sigma}(\forall h' \, \alpha_i(h') = f(h')) > 0$ so the following strategy is allowed: $\forall h' \in h\mathsf{States}^{<n} \, \sigma_i'(h') = f(h')$.

Let us complete σ_i' by choosing an arbitrary deterministic action for any $h' \in h\mathsf{States}^{\geq n}$ and then define $\sigma' = \sigma[i/\sigma_i']^h$. So

$$\mathbb{E}^{\sigma'}(\phi_i \cdot \mathbb{1}_{N < n} \mid h) \geq \mathbb{E}^{\sigma}(\phi_i \cdot \mathbb{1}_{N < n} \mid h) \geq \mathbb{E}^{\sigma}(\phi_i \mid h) - \mathbb{E}^{\sigma}(\phi_i \cdot \mathbb{1}_{N \geq n} \mid h) \geq \mathbb{E}^{\sigma}(\phi_i \mid h) - \varepsilon$$

In the case of $\phi_i \geq 0$, $\mathbb{E}^{\sigma'}(\phi_i \mid h) \geq \mathbb{E}^{\sigma'}(\phi_i \cdot \mathbb{1}_{N < n} \mid h)$ which concludes the proof.   ◀

Proof of Proposition 6. If $\langle \sigma, h \rangle$ is not a Nash equilibrium, then there is a deviation $\sigma_i'' \in \mathbb{S}_i$ for some Player i such that $\varepsilon = \mathbb{E}^{\sigma[i/\sigma_i'']^h}(\phi_i \mid h) - \mathbb{E}^{\sigma}(\phi_i \mid h) > 0$. Applying Lemma 18 with $\varepsilon/2$, we get a deterministic deviation τ_i'' for Player i for which

$$\mathbb{E}^{\sigma[i/\tau_i'']^h}(\phi_i \mid h) \geq \mathbb{E}^{\sigma[i/\sigma_i'']^h}(\phi_i \mid h) - \varepsilon/2 = \mathbb{E}^{\sigma}(\phi_i \mid h) + \varepsilon/2.$$ ◀

▶ **Remark.** A similar result was proven in [14, Proposition 3.1] for turn-based games with qualitative Borel objectives (the payoff is 1 if the run belongs to the designed objective and 0 otherwise).

A.2   One-stage games

Proof of Lemma 7. For $x \in \mathbb{R}$, we write $\overline{x} = 1 - x$. Then $\langle \sigma, s_0 \rangle$ is a Nash equilibrium if, and only if, it is resilient to deterministic deviations (Proposition 6). Considering the deterministic deviation of Player 0 returning move m, we get (omitting to mention s_0 in $\sigma_i(m)$):

$$a_0 \sigma_0(m)\sigma_1(m) + b_0 \sigma_0(n)\sigma_1(m) + c_0 \sigma_0(m)\sigma_1(n) + d_0 \sigma_0(n)\sigma_1(n) \geq a_0 \sigma_1(m) + c_0 \sigma_1(n).$$

As $\sigma_0(m) + \sigma_0(n) = 1$, we get $b_0 \sigma_0(n)\sigma_1(m) + d_0 \sigma_0(n)\sigma_1(n) \geq a_0 \sigma_0(n)\sigma_1(m) + c_0 \sigma_0(n)\sigma_1(n)$, which, assuming $\sigma_0(n) > 0$ (or, equivalently, $\sigma_0(m) < 1$), gives

$$[(a_0 - c_0) - (b_0 - d_0)] \cdot \sigma_1(m) \leq d_0 - c_0.$$

The other cases are similar. ◄

A.3 k-action matching-pennies games

Proof of Lemma 8. In the case where $a_0 = b_0$, Player 0 can never improve her payoff, and for any strategy σ_0 of Player 0, there is a strategy σ_1 of Player 1 such that (σ_0, σ_1) is a Nash equilibrium.

Applying the result of Lemma 7, we easily obtain that if $a_i \neq b_i$ for $i = 0, 1$, then there exists at most one non-pure Nash equilibrium, where both players play uniformly at random. Conversely, playing uniformly at random for both players is an equilibrium, since a single deviation would not modify the distribution of the outcomes.

Now if $a_0 > b_0$ and $a_1 > b_1$, then there is a pure Nash equilibrium which consists in deterministically joining the best terminal state.

Now, assume that $a_0 > b_0$ and $a_1 < b_1$ (the symmetric case would be similar; in the other cases, there is a pure Nash equilibrium), and consider a Nash equilibrium where Player 1 does not play uniformly at random: then there is one action, say c_n, receiving the largest probability mass (possibly 1, if the strategy is deterministic). The best response to this strategy for Player 0 is to play the same action c_n purely, since $a_0 > b_0$, so this is what Player 0 would play in a Nash equilibrium. Now, if Player 0 plays c_n purely, then Player 1 would better play an action different from c_n deterministically, which contradicts our assumption. Hence the only Nash equilibrium in such a case is uniform. ◄

A.4   Games without equilibrium

Proof of Lemma 9. Let us first consider the memoryless strategy for Player 0 where $\sigma_0(r \mid s_0^n) = \varepsilon$, where $0 < \varepsilon < 1$. Then Player 1 can either play s, with expected payoff $(1 - \varepsilon, \varepsilon)$, or play w, which either reaches the terminal state with payoff $(1, -1)$ with probability ε, or remains in a_0. This indicates that Player 1 cannot secure a payoff larger than ε in this case, so that the strategy of Player 0 secures a payoff of $1 - \varepsilon$. This entails that any Nash equilibrium can only have payoff $(1, -1)$.

Now, assume there exists a Nash equilibrium $\langle \sigma, s_0 \rangle$. Let n be the least index such that $\sigma_0(r \mid s_0^n) > 0$, assuming it exists. Then Player 1 has a strategy to exit to state $(-1, 1)$ with positive probability, which we proved cannot be the case of a Nash equilibrium. Hence Player 0 has to always play h; then if Player 1 always plays w, the play stays in s_0, with payoff $(0, 0)$, again a contradiction. It follows that $\mathcal{H}$ does not have a Nash equilibrium. ◄

Proof of Lemma 10. The game $\mathcal{G}'$ is depicted on Figure 11, where Player 0 can decide in s_0' whether to go to $\mathcal{H}$ or to $\mathcal{G}$. Assume there is a Nash equilibrium in $\mathcal{G}'$. Since $\mathcal{H}$

has no Nash equilibrium, in any Nash equilibrium of $\mathcal{G}$ from s'_0, Player 0 will play action continue (with probability 1) in s'_0. This entails that $\mathcal{G}$ has a Nash equilibrium (since the payoffs are prefix-independent). Moreover, the payoff of Player 0 in this Nash equilibrium must be 1, as otherwise Player 0 could secure a better payoff by going to $\mathcal{H}$ (see proof of Lemma 9). Conversely, if there is a 0-optimal Nash equilibrium in $\mathcal{G}$, then it gives rise to a Nash equilibrium in $\mathcal{G}'$ by letting Player 0 move to $\mathcal{G}$ in s'_0. This is easily seen to be a Nash equilibrium, in particular because deviating to $\mathcal{H}$ in s'_0 cannot benefit to Player 0. ◄
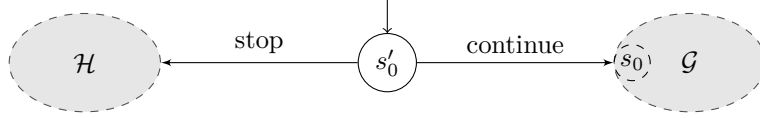
Figure 11 A game that has a Nash equilibrium if, and only if, $\mathcal{G}$ has a 0-optimal Nash equilibrium

A.5 Reduced game

We now present a technical construction that will be useful for our reduction: indeed, our reduction is modular, and consists in plugging subgames at various nodes of other games. When the same subgame can be reached via different histories, the moves returned in the subgame by a strategy may depend on the path that led to the subgame. Our aim here is to avoid this. Since this construction serves technical purposes only, we did not mention it in the main part of the paper. Its role will be made clear at the end of Appendix B.

The basic idea is to transform a terminal-reward game $\mathcal{G}$ into a one-shot game $\mathcal{G}'$ (where all but one state are terminal) embedding all the histories in simple transitions. Intuitively, the states of $\mathcal{G}'$ will be the maximal runs of $\mathcal{G}$, which we will assume to be finitely many. We show that for *action-visible* games, this transformation preserves the payoffs of Nash equilibria:

▶ **Definition 19.** A terminal-reward game $\mathcal{G}$ is *action-visible* if, for every non-final state $s \in \mathsf{States}$, if $\mathsf{Tab}(s, m_1) = \mathsf{Tab}(s, m_2)$, then either $m_1 = m_2$, or $\mathsf{Tab}(s, m_1)$ is terminal.

Let $\mathcal{G}$ be an a terminal-reward game, with $\mathcal{A} = \langle \mathsf{States}, \mathsf{Agt}, \mathsf{Act}, \mathsf{Tab}, (\mathsf{Allow}_i)_{i \in \mathsf{Agt}} \rangle$ as its underlying arena, and $h \in \mathsf{States}^+$ be a history. We denote by $\overline{H}(h)$ the set of the finite possible continuations of h, that is $\overline{H}(h) = \{h' \in \mathsf{States}^+ \mid \exists \sigma \in S. \ \mathbb{P}^\sigma(h' \mid h) > 0\}$, by $H(h) \subseteq \overline{H}(h)$ the subset containing those continuations that do not reach a final state, and by $H^\infty(h)$ the set of infinite continuations of h.

Assume that $H(h)$ is finite. This means that all continuations of h reach a final state after finitely many steps, so that also $H^\infty(h)$ is also finite. We define a new arena $\widetilde{\mathcal{A}}_h = \langle \widetilde{\mathsf{States}}, \mathsf{Agt}, \widetilde{\mathsf{Act}}, \widetilde{\mathsf{Tab}}, \widetilde{\mathsf{Allow}} \rangle$ as follows:

- $\widetilde{\mathsf{States}} = \{h\} \uplus \mathsf{F}$;
- $\widetilde{\mathsf{Act}} = \{\text{-}\} \uplus \{f : H(h) \to \mathsf{Act} \mid f = (\sigma_i)_{|H(h)}, i \in \mathsf{Agt}, \sigma_i \in S_i\}$ contains the (finetely many) deterministic strategies of the players in $\mathcal{G}$ restricted to $H(h)$, together with an extra action -;
- $\widetilde{\mathsf{Allow}}_i(h) = \{f \mid f = (\sigma_i)_{|H(h)}, \sigma_i \in S_i\}$, and $\widetilde{\mathsf{Allow}}_i(s) = \{\text{-}\}$ if $s \in \mathsf{F}$;
- $\widetilde{\mathsf{Tab}}(h, (f_i)_{i \in \mathsf{Agt}})$ is defined as the unique state $s \in \mathsf{F}$ reached from h with the deterministic strategy profile $(\sigma_i)_i = (f_i)_i$ in $\mathcal{G}$. We will also denote by $\mathsf{Tab}^*(h, (f_i)_{i \in \mathsf{Agt}})$ the corresponding unique infinite run X_∞ from h in $\mathcal{G}$. When $s \neq h$, $\widetilde{\mathsf{Tab}}(s, \{\text{-}\}^n) = s$.

The game $\widetilde{\mathcal{G}}_h$ is defined from $\widetilde{\mathcal{A}}_h$ by taking h as initial state, and playing the whole game in one round: final set of states and their associated payoffs $\phi = \tilde{\phi}$ remain unchanged.

▶ **Proposition 20.** *For any action-visible game $\mathcal{G}$ and any history $h \in$ States$^+$ with finite set $H(h)$, the sets of average payoffs of Nash equilibria in $\mathcal{G}$ and in $\widetilde{\mathcal{G}}$ coincide.*

Before proving this result, we begin with a lemma.

▶ **Lemma 21.** *Let $\mathcal{G}$ be a terminal-reward game and $\sigma \in \mathbb{S}$ be a mixed strategy profile in $\mathcal{G}$, and $\widetilde{\sigma} \in \widetilde{\mathbb{S}}$ be a mixed strategy profile in $\widetilde{\mathcal{G}}$, such that for any player i and any strategy f of Player i in $\mathcal{G}$, it holds $\widetilde{\sigma}_i(f \mid h) = \prod_{h' \in H(h)} \sigma_i(f(h') \mid h')$. Then for any $s \in \mathsf{F}$, $\mathbb{P}^\sigma(h\mathsf{States}^* s \mid h) = \mathbb{P}^{\widetilde{\sigma}}(h \cdot s \mid h)$.*

Proof. Denote by $X_s = \left\{ (f_i)_i \in \widetilde{\mathsf{Act}}^n \middle| \widetilde{\mathsf{Tab}}(h, (f_i)_i) = s \right\} = (\mathsf{Tab}^*(h, \cdot))^{-1}(\{h\mathsf{States}^* s^\omega\})$. We have the following sequence of equalities:

$$
\begin{aligned}
\mathbb{P}^\sigma(h\mathsf{States}^* s \mid h) &= \sum_{(f_i)_i \in X_s} \mathbb{P}^\sigma \left(\bigwedge_{i \in \mathsf{Agt}} \bigwedge_{h' \in H(h)} \alpha_i(h') = f_i(h') \right) \\
&= \sum_{(f_i)_i \in X_s} \prod_{i \in \mathsf{Agt}} \prod_{h' \in H(h)} \mathbb{P}^\sigma(\alpha_i(h') = f_i(h')) \\
&= \sum_{(f_i)_i \in X_s} \prod_{i \in \mathsf{Agt}} \prod_{h' \in H(h)} \sigma_i(f_i(h') \mid h') \\
&= \sum_{(f_i)_i \in X_s} \widetilde{\sigma}((f_i)_i \mid h) \\
&= \mathbb{P}^{\widetilde{\sigma}}(h \cdot s \mid h)
\end{aligned}
$$
◀

Proof of Proposition 20. Let $\langle \sigma, h \rangle$ be a Nash equilibrium in $\mathcal{G}$. We define the strategy profile $\widetilde{\sigma}$ in $\widetilde{\mathcal{G}}$ as in Lemma 21. Then $\mathbb{E}^{\widetilde{\sigma}}(\phi \mid h) = \mathbb{E}^\sigma(\phi \mid h)$. For any player $i \in \mathsf{Agt}$ and any deterministic deviation $\widetilde{\sigma}'_i \in \tilde{S}_i$, the associated deterministic deviation in $\mathcal{G}$ is $\widetilde{\sigma}'_i(h)$ (seen as a deterministic strategy). By letting, $\sigma' = \sigma[i/\widetilde{\sigma}'_i(h)]^h$ and $\widetilde{\sigma}' = \widetilde{\sigma}[i/\widetilde{\sigma}'_i]^h$, we check that Lemma 21 still applies to σ' and $\widetilde{\sigma}'$. It follows

$$
\mathbb{E}^{\widetilde{\sigma}'}(\phi \mid h) = \mathbb{E}^{\sigma'}(\phi \mid h) \leq \mathbb{E}^\sigma(\phi \mid h) = \mathbb{E}^{\widetilde{\sigma}}(\phi \mid h).
$$

So $\langle \widetilde{\sigma}, h \rangle$ is a Nash equilibrium with the same payoff as $\langle \sigma, h \rangle$.

Conversely, for any history $h' \in H(h)$, let $\varphi(h')$ indicate whether the chosen action $(\alpha_i(h))_i$ is a strategy profile generating h' in $\mathcal{G}$ (i.e., , $h' \sqsubseteq \mathsf{Tab}^*(h, (\alpha_i(h))_i)$.) Let $\varphi_i(h')$ indicate whether the action $\alpha_i(h)$ of Player i allows h' to occur (as an outcome in $\mathcal{G}$). Since the actions are visible, there exists a single sequence of actions that leads to h', so that we have $\varphi(h') = \wedge_{i \in \mathsf{Agt}} \varphi_i(h')$.

Assume that $\langle \widetilde{\sigma}, h \rangle$ is a Nash equilibrium. We define σ by

$$
\forall h' \in H(h). \ \forall i \in \mathsf{Agt}. \ \forall a \in \mathsf{Act}. \ \sigma_i(a \mid h') = \mathbb{P}^{\widetilde{\sigma}}\left(\alpha_i(h)(h') = a \ \middle| \ \varphi_i(h') \right)
$$

In other terms, $\sigma_i(a \mid h')$ is defined as the probability that the chosen strategy in $\widetilde{\mathcal{G}}$ returns action a in h', given that this strategy enables an outcome that reaches h'. Note that this value does not depend on the strategies $\widetilde{\sigma}_j$ for $j \neq i$ and that $\alpha_i(h)(h')$ and $\varphi_i(h')$ do not depend on the properties $\varphi_j(h')$ for $j \neq i$. Hence, we have for all $h' \in H(h)$,

$$
\mathbb{P}^\sigma\left(\alpha_i(h') = a \right) = \sigma_i(a \mid h') = \mathbb{P}^{\widetilde{\sigma}}\left(\alpha_i(h)(h') = a \mid \varphi_i(h') \right) = \mathbb{P}^{\widetilde{\sigma}}\left(\alpha_i(h)(h') = a \mid \varphi(h') \right),
$$

and

$$\forall A \in \mathsf{Act}^n. \ \mathbb{P}^\sigma\Big((\alpha_i(h'))_i = A\Big) = \mathbb{P}^{\widetilde{\sigma}}\Big((\alpha_i(h)(h'))_i = A \mid \varphi(h')\Big). \tag{1}$$

We show by induction that Equation (1) implies that for all $h' \in \overline{H}(h)$, $\mathbb{P}^\sigma(h' \mid h) = \mathbb{P}^{\widetilde{\sigma}}(\varphi(h'))$:

- If $h' = h$, then the result is trivially true;
- If $h' = h_1 s$ and $\mathsf{last}(h_1) \in \mathsf{F}$, then s is the same terminal state;
- If $h' = h_1 s$ and $\mathsf{last}(h_1) \notin \mathsf{F}$, then $h_1 \in H(h)$. Let us denote the unique joint action $A \in \mathsf{Act}^n$ such that $\mathsf{Tab}(\mathsf{last}(h_1), A) = s$. We have,

$$\begin{aligned}
\mathbb{P}^\sigma(h' \mid h) &= \mathbb{P}^\sigma(h_1 \mid h) \times \mathbb{P}^\sigma((\alpha_i(h_1))_i = A) \\
&= \mathbb{P}^{\widetilde{\sigma}}\Big(\varphi(h_1)\Big) \times \mathbb{P}^{\widetilde{\sigma}}\Big((\alpha_i(h)(h_1))_i = A \mid \varphi(h_1)\Big) \quad \text{(by (1) and by induction)} \\
&= \mathbb{P}^{\widetilde{\sigma}}\Big(\varphi(h_1) \wedge (\alpha_i(h)(h_1))_i = A\Big) \\
&= \mathbb{P}^{\widetilde{\sigma}}(\varphi(h'))
\end{aligned}$$

We conclude finally that for all $s \in \mathsf{F}$, it holds $\mathbb{P}^\sigma(h\mathsf{States}^* s \mid h) = \mathbb{P}^{\widetilde{\sigma}}(h \cdot s \mid h)$ therefore $\mathbb{E}^\sigma(\phi \mid h) = \mathbb{E}^{\widetilde{\sigma}}(\phi \mid h)$.

If we consider a deterministic deviation $\sigma_i' \in S_i$, we construct the corresponding deviation $\widetilde{\sigma}_i' \in \widetilde{S}_i$ in $\widetilde{\mathcal{G}}$ satisfying $\widetilde{\sigma}_i'(h)(h') = \sigma_i'(h')$ for all $h' \in H(h)$. Equation (1) still holds, so $\langle \sigma, h \rangle$ is a Nash equilibrium:

$$\mathbb{E}^{\sigma[i/\sigma_i]^h}(\phi_i \mid h) = \mathbb{E}^{\widetilde{\sigma}[i/\sigma_i]^h}(\phi_i \mid h) \leq \mathbb{E}^{\widetilde{\sigma}}(\phi_i \mid h) = \mathbb{E}^\sigma(\phi_i \mid h) \qquad\qquad \blacktriangleleft$$
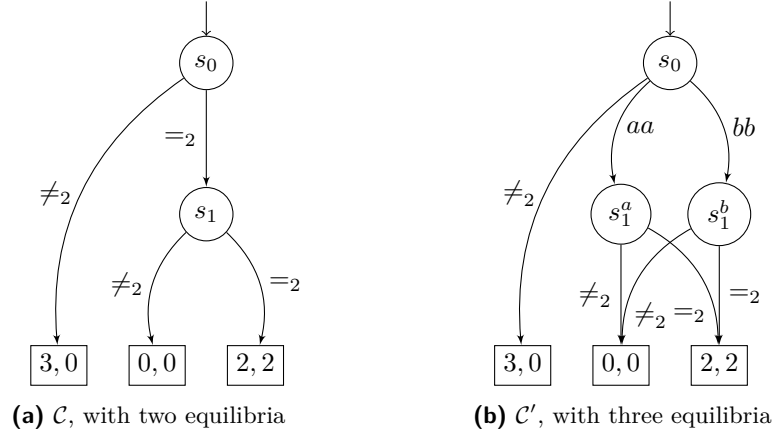
▶ Remark. Notice that the restriction to action-visible games is crucial in Proposition 20. Indeed, consider the games depicted in Figure 12: in game $\mathcal{C}$, there are only two possible equilibria from s_1: either both players agree on playing the same action, yielding payoff $(2, 2)$, or they both play uniformly at random, yielding payoff $(1, 1)$. From state s_0, the second player always has an incentive to go to s_1, whereas the first player is always better off with payoff $(3, 0)$. We conclude that $\mathcal{C}$ has two equilibria:

- $\left(2, \frac{1}{2}\right)$: $\sigma_i(s_0)$ is uniform and $\sigma_0(s_1) = \sigma_1(s_1)$;
- $\left(\frac{5}{2}, 1\right)$: $\sigma_i(s_0)$ and $\sigma_i(s_1)$ are uniform.

Now, $\mathcal{C}'$ is an unfolded version of $\mathcal{C}$ which has been made action-visible by annotating state s_1 with the actions from s_0 that triggered the transition. The previous equilibria are still possible in $\mathcal{C}'$ but a new equilibrium arises: agents can play uniformly from s_1^α and agree on an action in the other state s_1^β (for $\beta \neq \alpha$). Let us consider the case where s_1^a is played uniformly, the other case being symmetric. From state s_0, game can now be seen as a one-shot game as represented in Figure 13a. This game has a unique equilibrium, which can be computed by Lemma 7 as seen in Figure 13b.

However, both reduced games have the same equilibrium payoffs. Indeed, from an equilibrium in one of the reduced game, we can build an equilibrium in the other reduced game, with the same payoff:

- if $\langle \widetilde{\sigma}, s_0 \rangle$ is a Nash equilibrium of $\widetilde{\mathcal{C}}$: for every $f' : \{s_0, s_0 s_1^a, s_0 s_1^b\} \to \{a, b\}$, we define $f : \{s_0, s_0 s_1\} \to \{a, b\}$, by letting $f(s_0) = f'(s_0)$ and $f(s_0 s_1) = f'(s_0 s_1^{f'(s_0)})$, and for all i, $\widetilde{\sigma}_i'(f' \mid s_0) = \widetilde{\sigma}_i(f \mid s_0)$;
- if $\langle \widetilde{\sigma}', s_0 \rangle$ is a Nash equilibrium of $\widetilde{\mathcal{C}'}$: for every $f : \{s_0, s_0 s_1\} \to \{a, b\}$, we define $f' : \{s_0, s_0 s_1^a, s_0 s_1^b\} \to \{a, b\}$, by letting $f'(s_0) = f(s_0)$ and for all α, $f'(s_0 s_1^\alpha) = f(s_0 s_1)$. and for all i, $\widetilde{\sigma}_i(f \mid s_0) = \widetilde{\sigma}_i'(f' \mid s_0)$.

(a) $\mathcal{C}$, with two equilibria

(b) $\mathcal{C}'$, with three equilibria

Figure 12 Two games with different equilibria, but whose reduced games have the same equilibrium payoffs.

(a) Matrix game from $\mathcal{C}'$

$$\begin{cases} \sigma_0(a \mid s_0) = \dfrac{2}{2+1} = \dfrac{2}{3} \\ \sigma_1(a \mid s_0) = \dfrac{-1}{-1-2} = \dfrac{1}{3} \end{cases} \qquad \mathbb{E}^\sigma(\phi_i \mid s_0) = \left(\dfrac{7}{3}, \dfrac{2}{3}\right)$$

(b) Associated equilibrium

Figure 13 The extra equilibrium in $\mathcal{C}'$

One can check that these indeed gives rise to Nash equilibria in the respective games. We conclude from this analysis that Property 20 does not apply to $\mathcal{C}$.

Let $\mathcal{G} = (\mathcal{A}, s, \phi)$ be a terminal-reward game. A subarena $\mathcal{H}$ of $\mathcal{G}$ is defined as a subset of states. The subarena $\mathcal{H}$ is said action-visible whenever for all $s \in \mathcal{H}$ and $m_1, m_2 \in \mathsf{Act}$, $\mathsf{Tab}(s, m_1) = \mathsf{Tab}(s, m_2)$ implies $m_1 = m_2$ or $\mathsf{Tab}(s, m_1) \notin \mathcal{H}$.

We define an equivalence relation $\equiv_\mathcal{H}$ on histories of $\mathcal{G}$ by

$$h \equiv_\mathcal{H} h' \Leftrightarrow \exists n.\ \exists w_0 \ldots w_{n-1} \in (\mathsf{States} \backslash \mathcal{H})^* \mathcal{H}.\ \exists w_n \in \mathsf{States}^+.$$
$$h \text{ and } h' \text{ are in } w_0 \mathcal{H}^* \cdot w_1 \mathcal{H}^* \cdots w_{n-1} \mathcal{H}^* \cdot w_n.$$

That is to say, we enter in $\mathcal{H}$ from an entry point $\mathsf{last}(w_i)$ (for $i < n$) and events that occurred when staying in $\mathcal{H}$ are forgotten.

A strategy profile σ is said *blind to* $\mathcal{H}$ whenever it does not distinguish between histories that are equivalent w.r.t. $\equiv_\mathcal{H}$. The following corollary is an obvious consequence of the study made in this section, and will be heavily used in the rest of the paper:

▶ **Corollary 22.** *Assume $\mathcal{G}$ is a terminal-reward game, and $\mathcal{H}$ is a subarena of $\mathcal{G}$, such that $\mathcal{H}$ is action-visible and for all $s \in \mathcal{H}$, $H(s) \cap \mathcal{H}^*$ is finite (where $H(s)$ is the set of non-terminal histories from s in $\mathcal{G}$). We can construct a game $\mathcal{G}/\mathcal{H}$ such that for every payoff vector $v = (v_i)_{i \in \mathsf{Agt}}$, the following two properties are equivalent:*
- *there is a Nash equilibrium in $\mathcal{G}/\mathcal{H}$ with payoff v;*
- *there is a Nash equilibrium in $\mathcal{G}$ with payoff v, which is blind to $\mathcal{H}$.*

Proof. In order to build $\mathcal{G}/\mathcal{H}$, we keep the same set of agents and states but modify the allowed actions and transition function in $\mathcal{H}$. For Player i and state $s \in \mathcal{H}$, we define:

- $\mathsf{Allow}'_i(s) = \{f = (\sigma_i)_{|H(s)\cap\mathcal{H}} \mid \sigma_i \in S_i\}$ the set of deterministic strategies from s when staying in $\mathcal{H}$ (remember $H(s) \cap \mathcal{H}$ is finite). Act' is updated accordingly.
- $\mathsf{Tab}'(s, (f_i)_i)$ is defined as the first state $s' \in \mathsf{States}\setminus\mathcal{H}$ that left the subgame when applying the deterministic strategy $(f_i)_i$.

We now prove that this construction achieves our equivalence.

- If $\langle\sigma', s_0\rangle$ is a Nash equilibrium in $\mathcal{G}/\mathcal{H}$, for every enabled history hs with $s \in \mathcal{H}$, we can apply Proposition 20 to construct an equilibrium σ for all histories equivalent to hs w.r.t. $\equiv_{\mathcal{H}}$. For every history $h' \equiv_{\mathcal{H}} hs$ with $s \notin \mathcal{H}$, we define $\sigma(h') = \sigma'(h)$. Notice that $\equiv_{\mathcal{H}}$ is transitive so this definition is sound, and induces a strategy profile that is blind to $\mathcal{H}$.
- Conversely, assume $\langle\sigma, s_0\rangle$ is a Nash equilibrium in $\mathcal{G}$, which is blind to $\mathcal{H}$. Let hs be a history of $\mathcal{G}/\mathcal{H}$. If $s \notin \mathcal{H}$, we define $\sigma'_i(h) = \sigma_i(h)$. Otherwise, $s \in \mathcal{H}$ and we define $\sigma'_i(h)$ as the distribution over strategies induced by $(\sigma_i(h'))_{h'\in H(s)}$ (see Lemma 21 for details.) ◀

▶ **Remark.** Notice that in the games that we will define in the sequel, the subgames $\mathcal{H}$ have a state e that is the only predecessor of the others. The construction of $\mathcal{G}/\mathcal{H}$ will then totally avoid the ability of reaching states in $\mathcal{H}\setminus\{e\}$. This subset will thus be omitted in the figures, and we will refer to $\mathcal{H}$ as a black-box with a unique entry state e and some possible exit points in $\mathsf{States}\setminus\mathcal{H}$.

B   Details for Section 4

Consider the game $\mathcal{G}^r_k$ depicted on Figure 14: in this game, Player 0 has two available actions a and b from s_0, s_k and s_l, while the other two players can either continue (action c), or unilateraly decide to stop the game (action s) and go to a terminal state (where Player 0 will have payoff 0). In node t_k, only players 1 and 2 have a choice: they can either continue to the game $\mathcal{H}$ (when both of them play c), or decide to stop and go to a k-action matching-pennies game (when one of them plays s). In Figure 14, we write S as a shorthand to represent any combination of moves of players 1 and 2 where at least one of them decides to stop (action s). Node n is the initial node of a game $\mathcal{H}$ (which is unknown for the moment).

We relate 0-optimal Nash equilibria from s_0 and those from n:

▶ **Proposition 23.** *Consider the (action-visible) game $\mathcal{G}^r_k$ of Figure 14, with $k \in \{1, 2, 3\}$.*

- *Assume that $0 \le k \cdot x \le 1$ and that there is a Nash equilibrium in $\mathcal{H}$ from n with expected payoff $(1, 4 + x, 4 - x)$. Then there is a Nash equilibrium from s_0 with expected payoff $(1, 4 + k \cdot x, 4 - k \cdot x)$: it consists for Player 0 in playing a with probability $k \cdot x$ at s_0, and a with probability $x/(x + 1)$ at s_k or s_l; the other two players play c almost-surely everywhere except in u_k, where they both play uniformly at random; in $\mathcal{H}$, they follow the given 0-optimal Nash equilibrium.*
- *Assume that, in $\mathcal{H}$, all terminal states with reward 1 for Player 0 are such that the rewards of players 1 and 2 sum up to 8. Assume that $\langle\sigma, s_0\rangle$ is a 0-optimal Nash equilibrium such that $\sigma(s_0 s_k t_k \cdot)$ and $\sigma(s_0 s_{k-1} t_k \cdot)$ coincide (that is, the strategy after t_k is indifferent of whether the history went through s_k or s_{k-1}). Then:*
  - *states t_k and n are activated by σ;*
  - *$\langle\sigma, s_0 s_k t_k\rangle$ and $\langle\sigma, s_0 s_k t_k n\rangle$ (and equivalently $\langle\sigma, s_0 s_{k-1} t_k\rangle$ and $\langle\sigma, s_0 s_{k-1} t_k n\rangle$) are 0-optimal Nash equilibria of expected payoff $(1, 4 + x, 4 - x)$ for some $0 \le x \le 1/k$;*

- $\mathbb{E}^{\sigma}(\phi \mid s_0) = (1, 4 + k \cdot x, 4 - k \cdot x)$.

The proof of this proposition requires a careful analysis of the game. We decompose the proof into several intermediary lemmas, which all refer to the game $\mathcal{G}_k^r$ of Figure 14.

▶ **Lemma 24.** *Assume that there is a Nash equilibrium from t_k with payoff $(1, 4+x, 4-x)$. If $\langle \sigma, s_k \rangle$ is a 0-optimal Nash equilibrium from s_k, then the state t_k is activated by σ, and the expected payoff of σ from s_k is $(1, 4 + (k+1) \cdot \frac{x}{x+1}, 4 - (k+1) \cdot \frac{x}{x+1})$. Furthermore such a 0-optimal Nash equilibrium from s_k exists and consists for player 0 in playing a with probability $x/(x+1)$ and b with probability $1/(x+1)$, and for the other two players, in playing c almost-surely in s_k, and then follow the given equilibrium from t_k.*

An analogous result applies from s_l: if $\langle \sigma, s_l \rangle$ is a 0-optimal Nash equilibrium from s_l, then the state t_k is activated by σ, and the expected payoff of σ from s_l is $(1, 4+l \cdot \frac{x}{x+1}, 4-l \cdot \frac{x}{x+1})$. Furthermore such a 0-optimal Nash equilibrium from s_l exists and consists for player 0 in playing a with probability $x/(x+1)$ and b with probability $1/(x+1)$, and for the other two players, in playing c almost-surely in s_l, and then follow the given equilibrium from t_k.

Proof. Let $\langle \sigma, s_k \rangle$ be a 0-optimal Nash equilibrium. Because the equilibrium is 0-optimal, players 1 and 2 do not play action s: we have $\sigma_1(c \mid s_k) = \sigma_2(c \mid s_k) = 1$. Considering this fixed action for Player 2, we can look at the 2-player game between players 0 and 1, which is represented in matrix form in Table 1 (left).

Applying Lemma 7, using the fact that $\sigma_1(s \mid s_k) < 1$, we get that $(x+1) \cdot \sigma_0(a \mid s_k) \le x$. The same argument applied to the projection to players 0 and 2 (see Table 1 (right)) gives $(x+1) \cdot \sigma_0(a \mid s_k) \ge x$. Hence $(x+1) \cdot \sigma_0(a \mid s_k) = x$. This entails that $x \ne -1$ (actually, x will be forced to be nonnegative in the sequel), so that the action vector bcc has probability $1/(x+1)$ in the Nash equilibrium $\langle \sigma, s_k \rangle$. We conclude that t_k is reached with positive probability, and that

$$\mathbb{E}^{\sigma}(\phi \mid s_k) = \left(1, \frac{(4+k)x + (4+x)}{x+1}, \frac{(4-k)x + (4-x)}{x+1}\right).$$
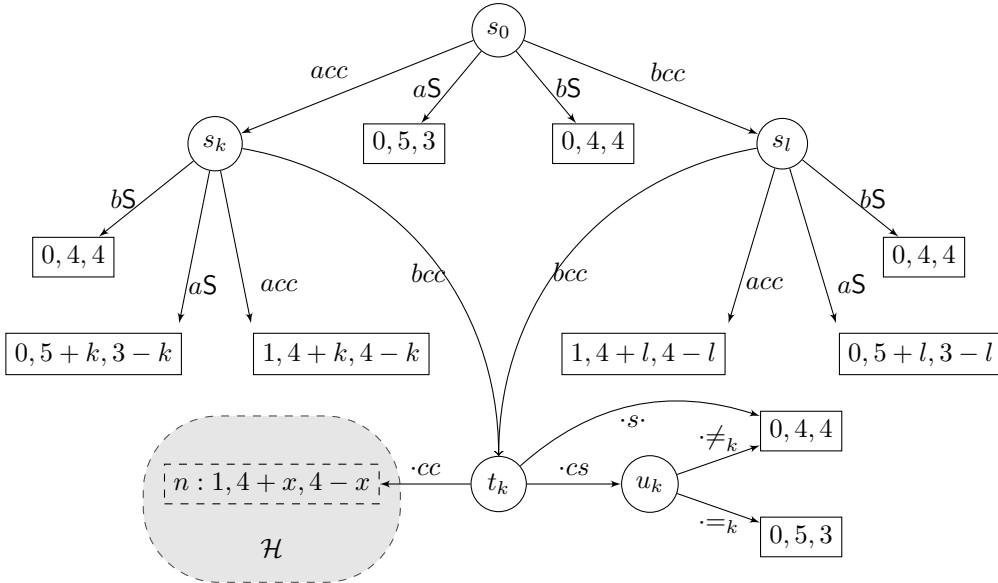
Figure 14 The rescale game $\mathcal{G}_k^r$ for $k \in \{1, 2, 3\}$ with $l = k - 1$

1 \ 0	a	b
s	$0, 5+k$	$0, 4$
c	$1, 4+k$	$1, 4+x$

2 \ 0	a	b
s	$0, 3-k$	$0, 4$
c	$1, 4-k$	$1, 4-x$

■ **Table 1** Two-player projections of $\mathcal{G}_k^r$ in s_k assuming Player 2 (left), resp. Player 1 (right), plays c almost-surely

1 \ 0	a	b
s	$0, 5$	$0, 4$
c	$1, 4+y_k$	$1, 4+y_l$

2 \ 0	a	b
s	$0, 3$	$0, 4$
c	$1, 4-y_k$	$1, 4-y_l$

■ **Table 2** Two-player projections of $\mathcal{G}_k^r$ in s_0 assuming Player 2 (left), resp. Player 1 (right), plays c almost-surely

Finally it is not hard to realize that the strategy profile from s_k where player 0 plays a with probability $x/(x+1)$ and b with probability $1/(x+1)$ in s_k, and where the other two players play c almost-surely, is a 0-optimal Nash equilibrium.

The same reasoning can be done from state s_l. ◄

We now consider the game $\mathcal{G}_k^r$ from its initial state s_0.

▶ **Lemma 25.** *Assume that $0 \le k \cdot x \le 1$ and that there is a Nash equilibrium from t_k with expected payoff $(1, 4+x, 4-x)$. If $\langle \sigma, s_0 \rangle$ is a 0-optimal Nash equilibrium from s_0, then the state t_k is activated by σ, and the expected payoff of σ from s_0 is $(1, 4+k \cdot x, 4-k \cdot x)$. Furthermore such a 0-optimal Nash equilibrium from s_0 exists and consists for player 0 in playing a at s_0 with probability $k \cdot x$, and for the other two players, to play c almost-surely at s_0, and then in following the profile described in Lemma 24.*

Proof. Since the equilibrium is 0-optimal, it holds $\sigma_1(c \mid s_0) = \sigma_2(c \mid s_0) = 1$. We first consider the cases when only one of the states s_k and s_l is enabled:

- if only s_k is enabled, *i.e.*, $\sigma_0(a \mid s_0) = 1$, then from the previous lemma, we have $\mathbb{E}^\sigma(\phi_1) = 4 + (k+1) \cdot \frac{x}{x+1}$. This quantity should be greater than 5 (otherwise Player 1 would better deviate), so that $k \cdot x \ge 1$, and using our hypothesis, $k \cdot x = 1$. It follows that the payoff of σ from s_0 is $(1, 5, 3)$ in this case.
- if only s_l is enabled, *i.e.*, $\sigma_0(a \mid s_0) = 1$, the value for Player 2 is $4 - (l+1) \cdot \frac{x}{x+1}$. This must be greater than or equal to 4 (otherwise Player 2 has a profitable deviation). We get $x = 0$, and the expected payoff of σ is $(1, 4, 4)$.

We now consider the case where both states s_k and s_l are enabled, *i.e.*, $0 < \sigma_0(a \mid s_0) < 1$. We again separately consider the strategies of players 1 and 2, as shown in Table 2. Let us fix y_i (for $i \in \{k, l\}$) and y such that $\mathbb{E}^\sigma(\phi \mid s_0 s_i) = (1, 4+y_i, 4-y_i)$, and $\mathbb{E}^\sigma(\phi \mid s_0) = (1, 4+y, 4-y)$. Applying Lemma 7 twice, we get $(y_l + 1 - y_k) \cdot \sigma_0(a \mid s_0) = y_l$. Then the expected payoff for Player 1 is

$$\mathbb{E}^\sigma(\phi_1 \mid s_0) = 4 + y = \sigma_0(a \mid s_0) \cdot (4+y_k) + (1 - \sigma_0(a \mid s_0)) \cdot (4+y_l).$$

This simplifies as $y = y_l/(y_l - y_k + 1)$. Replacing y_l and y_k with their values (Lemma 24), we end up with $y = k \cdot x$. Now to reconstruct a 0-optimal Nash equilibrium from s_0, it is sufficient for Player 0 to play a at s_0 with probability $k \cdot x$, and for the other two players, to play c almost-surely at s_0, and then to follow the 0-optimal Nash equilibrium from s_k and s_l. This yields a 0-optimal Nash equilibrium from s_0.   ◀

We now explain how the k-action matching-pennies game from u_k allows to lift the constraint that $0 \leq k \cdot x \leq 1$.

▶ **Lemma 26.** *Assume that there is a Nash equilibrium from n with expected payoff $(1, 4 + x, 4 - x)$ (with no constraints on x). There is a 0-optimal Nash equilibrium from t_k if, and only if, $0 \leq k \cdot x \leq 1$. Furthermore a strategy profile achieving that condition consists in playing almost-surely c for both players in t_k, and in playing uniformly at random in u_k. Then, the expected payoff of this equilibrium from t_k is also $(1, 4 + x, 4 - x)$.*

Proof. Assume that there is a 0-optimal Nash equilibrium from t_k: since it is 0-optimal, it goes to state n almost-surely (hence players 1 and 2 play action c almost-surely in t_k). Moreover, the expected payoff of Player 1 from n must then be larger than or equal to 4, as otherwise in t_k she could deviate and secure a payoff of 4. Hence $4 + x \geq 4$, so that $x \geq 0$. Finally, in u_k, Player 2 can secure payoff $4 - 1/k$ by playing uniformly at random. Since this should not give rise to a profitable deviation for Player 2 from t_k, it must be $4 - x \geq 4 - 1/k$, so that $k \cdot x \leq 1$.

The converse is straightforward: players 1 and 2 should play c in t_k and u_k, and play uniformly at random in v_k.   ◀

Proof of Proposition 23. The first property follows from Lemmas 25 and 26.

We now turn to the second property. Fix a 0-optimal Nash equilibrium $\langle \sigma, s_0 \rangle$ such that $\sigma(s_0 s_k t_k \cdot) = \sigma(s_0, s_{k-1} t_k \cdot)$. Then in $\mathcal{G}_k^r$, only terminal states assigning reward 1 to player 0 are activated. Assume s_k is visited with positive probability: from that state players 1 and 2 play c almost-surely (otherwise Player 0 would not have payoff 0). Then from s_k, action b has to be played with positive probability, otherwise Player 1 would better stop the game and get payoff $5 + k$. The same reasoning applies to s_{k-1}, and we deduce that t_k is activated. Now since the payoff of Player 0 after u_k is 0, it means that u_k is not activated by σ, and therefore that n is activated. This implies in particular that $\langle \sigma, s_0 s_k t_k \rangle$ and $\langle \sigma, s_0 s_k t_k n \rangle$ are Nash equilibrium, and they are 0-optimal.

Now, since the rewards of players 1 and 2 in terminal states of $\mathcal{H}$ sum up to 8 when Player 0 has reward 0, it holds that $\mathbb{E}^\sigma(\phi_1 \mid s_0 s_k t_k n) + \mathbb{E}^\sigma(\phi_2 \mid s_0 s_k t_k n) = 8$, which we can rewrite as: $\mathbb{E}^\sigma(\phi_1 \mid s_0 s_k t_k n) = 4 + x$ and $\mathbb{E}^\sigma(\phi_2 \mid s_0 s_k t_k n) = 4 - x$ for some x. Applying Lemma 26, we get that $0 \leq x \leq 1/k$, and then, applying Lemma 25, that the expected payoff of $\langle \sigma, s_0 \rangle$ is $(1, 4 + k \cdot x, 4 - k \cdot x)$, which concludes the proof.   ◀

▶ Remark. The condition in Proposition 23 that the strategy profile should not distinguish between histories visiting states s_k and s_{k-1} can be lifted by considering the reduced game associated to $\mathcal{G}_k^r$, after noticing that $\mathcal{G}_k^r$ is indeed action-visible, and applying Corollary 22. Notice that in the main reduction (see Appendix D and in particular Figure 17), we indeed use the reduced version $\widetilde{\mathcal{G}_k^r}$ of $\mathcal{G}_k^r$.

C   Details for Section 5

C.1   Testing game

We present in this section the construction of a game for comparing the expected payoffs in different nodes. This will be useful in our reduction to encode the zero-tests of our two-counter machine.
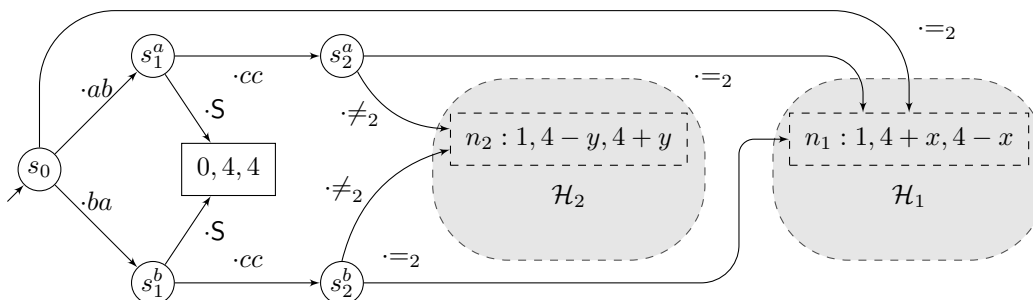
Figure 15 The testing game $\mathcal{G}_t$

▶ **Proposition 27.** *Consider the (action-visible) game $\mathcal{G}_t$ of Figure 15.*

▬ *Assume that there is a Nash equilibrium in $\mathcal{H}_1$ from n_1 with expected payoff $(1, 4+x, 4-x)$, and a Nash equilibrium in $\mathcal{H}_2$ from n_2 with expected payoff $(1, 4-y, 4+y)$, and that $x, y \geq 0$. Then there exists a 0-optimal Nash equilibrium from s_0 if, and only if, $x = y$. Moreover, when this condition is satisfied, there is a Nash equilibrium from s_0 with expected payoff $(1, 4+x/2, 4-x/2)$: it consists for both players 1 and 2 in playing action c almost-surely in s_1^α, and in playing uniformly at random in s_0 and s_2^α (for $\alpha \in \{a, b\}$); and then, they should follow the equilibria from n_1 and n_2.*

▬ *Assume that in $\mathcal{H}_1$, all terminal states with reward 1 for Player 0 are such that the rewards of player 1 and 2 sum up to 8. Assume that $\langle \sigma, s_0 \rangle$ is a 0-optimal Nash equilibrium with expected payoff $(1, 4+z, 4-z)$ with $z > 0$, such that σ is blind to subgame $\mathcal{G}_t$. Assume furthermore that in $\mathcal{H}_2$, there are 0-optimal Nash equilibria, and that their expected payoffs are all of the form $(1, 4-y, 4+y)$ with $y \geq 0$. Then:*
  - *states n_1 and n_2 are activated by σ;*
  - *$\langle \sigma, s_0 \; \mathcal{G}_t \; n_1 \rangle$ [2] and $\langle \sigma, s_0 \; \mathcal{G}_t \; n_2 \rangle$ are 0-optimal Nash equilibria of expected payoff $(1, 4+x, 4-x)$, respectively $(1, 4-y, 4+y)$ with $x, y > 0$, and $x = y$;*
 - *$z = x/2$, that is: $\mathbb{E}^\sigma(\phi \mid s_0) = (1, 4+x/2, 4-x/2)$.*

Proof. We prove both items separately.

▬ We first prove the first item. We first assume $x = y$. We consider the strategy profile σ where both players 1 and 2 play action c in s_1^α, and where they play uniformly at random in s_0 and s_2^α (for $\alpha \in \{a, b\}$). Then,

$$\mathbb{E}^\sigma(\phi \mid s_1^\alpha) = \mathbb{E}^\sigma(\phi \mid s_2^\alpha) = \left(1, 4+\frac{x-y}{2}, 4-\frac{x-y}{2}\right) = (1, 4, 4)$$

so that none of the players have an incentive to deviate. Finally, we have in s_0 the exact situation of a matching pennies, so that the uniform strategy forms an equilibrium, with payoff $\mathbb{E}^\sigma(\phi \mid s_0) = \left(1, 4+\frac{x}{2}, 4-\frac{x}{2}\right)$.

---

[2] As σ is blind to subgame $\mathcal{G}_t$, this is strategy $\langle \sigma, h \rangle$ after any path h in $\mathcal{G}_t$ from s_0 to n_1.
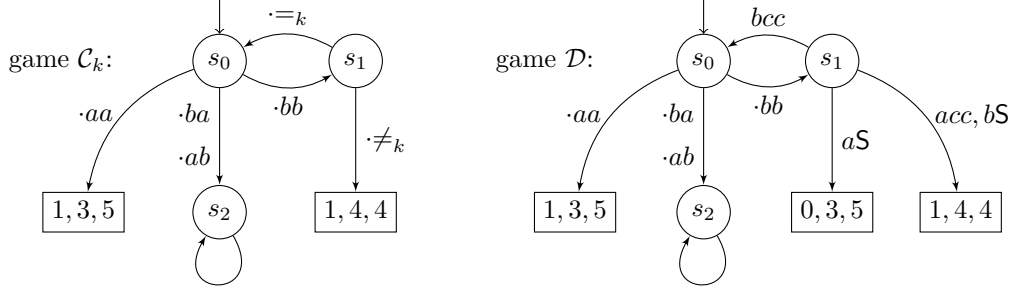
Figure 16 The modules $\mathcal{C}_k$ (for $k \geq 2$) and $\mathcal{D}$. Notice that state s_2 should be considered terminal, as it only carries a self-loop. We could replace it by a two-state loop. We could also see it as a terminal state with reward $(0, 0, 0)$, but for the proof of Corollary 16, we want the terminal rewards of players 1 and 2 to always sum to 8, which we could not achieve easily in this case.

Conversely, assume there is a 0-optimal Nash equilibrium from s_0. The payoff of Player 1 (resp 2) is of the form $4 + z$ (resp $4 - z$) for some value $z \in \mathbb{R}$.

- if none of the transitions $s_0 \to s_1^\alpha$ ($\alpha \in \{a, b\}$) is enabled, then $z = x$. Moreover, we must have $x = 0$, since otherwise Player 2 could deviate to s_1^α and ensure payoff 4. Similarly, $y = 0$, since otherwise Player 2 could improve her payoff by deviating to s_2^α and then n_2.
- A similar reasoning holds if only one of the $s_0 \to s_1^\alpha$ transition is enabled.
- Finally, if all three transitions are enabled from s_0, then action cc should be played from s_1^a and s_1^b, because the equilibrium is 0-optimal. From s_2^α, actions should be chosen uniformly (unless $x = y = 0$) to ensure equilibrium, and the expected payoff will be

$$\mathbb{E}^\sigma(\phi \mid s_1^\alpha) = \mathbb{E}^\sigma(\phi \mid s_2^\alpha) = \left(1, 4 + \frac{x - y}{2}, 4 - \frac{x - y}{2}\right).$$

  By stability, and since $(0, 4, 4)$ is not enabled, we must have $4 \leq 4 + \frac{x-y}{2}$ and $4 \leq 4 - \frac{x-y}{2}$, that is, $x = y$ and $\mathbb{E}^\sigma(\phi \mid s_1^\alpha) = (1, 4, 4)$. It follows that in s_0, both players have to play uniformly to ensure equilibrium, with resulting payoff $\mathbb{E}^\sigma(\phi \mid s_0) = \left(1, 4 + \frac{x}{2}, 4 - \frac{x}{2}\right)$.

- Assume now that there is a 0-optimal Nash equilibrium σ from s_0 with expected payoff $(1, 4 + z, 4 - z)$ with $z > 0$. Towards a contradiction, assume that n_1 is not activated by σ: then only one of the states s_1^α is reached almost-surely, and in this case Player 2 has a profitable deviation to terminal state labelled by $(0, 4, 4)$. Hence n_1 is activated by σ, and there should be a Nash equilibrium from n_1 of expected payoff of the form $(1, 4 + x, 4 - x)$ with $x \in \mathbb{R}$. Now since $z > 0$, it should be the case that $x > 0$ (since all Nash equilibria from n_2 have a payoff (strictly) smaller than $4 + z$ for Player 2). We next realize that σ cannot go almost-surely to n_1, otherwise Player 2 would have a profitable deviation to terminal state labelled by $(0, 4, 4)$ as well. Hence at least one of the states s_1^α is activated by σ, and for Player 2 not to be willing to deviate to terminal state $(0, 4, 4)$, it should be the case that n_2 is activated: there is a 0-optimal Nash equilibrium from n_2. We are now with the hypotheses of the first item, which allows to conclude the proof. ◄

C.2   Counting modules

Proof of Lemma 11. Consider the games of Figure 16.

Fix $k \geq 2$. We begin with proving that these values are indeed the payoffs of Nash equilibria. For this, we define the witnessing strategy profiles. For all history h, we let

$$\gamma^{\infty}(hs_0) = abb \qquad \gamma^{\infty}(hs_1) = ac_1c_2 \qquad \delta^{\infty}(hs_0) = abb \qquad \delta^{\infty}(hs_1) = bss$$

One easily observes that $\langle \sigma^{\infty}, s_0 \rangle$ and (δ^{∞}, s_0) are Nash equilibria with payoff $(1, 4, 4)$ in $\mathcal{C}_k$ and in $\mathcal{D}$, respectively.

For $n \in \mathbb{N}$, we define γ^n and δ^n inductively. First, we let $\gamma^0(s_0) = \delta^0(s_0) = aaa$, which gives rise to Nash equilibria with payoff $(1, 3, 5)$ from s_0 both in $\mathcal{C}_k$ and in $\mathcal{D}$.

Then, for $n \in \mathbb{N}$, we define the strategies inductively on the length of the history. The base cases are $\gamma^{n+1}(s_0) = \delta^{n+1}(s_0) = abb$ and

$$\gamma_0^{n+1}(s_0s_1) = a \qquad \forall 1 \leq j \leq k. \; \gamma_1^{n+1}(c_j \mid s_0s_1) = \gamma_2^{n+1}(c_j \mid s_0s_1) = 1/k$$

$$\delta_0^{n+1}(a \mid s_0s_1) = \frac{1}{n+2} \qquad \qquad \delta_1^{n+1}(s_0s_1) = \delta_2^{n+1}(s_0s_1) = c.$$

The inductive case is

$$\gamma_i^{n+1}(s_0s_1h) = \gamma_i^n(h) \qquad \qquad \delta_i^{n+1}(s_0s_1h) = \delta_i^n(h)$$

It is not difficult to check that these strategy profiles form Nash equilibria with the expected payoffs.

Conversely, let us fix a 0-optimal Nash equilibrium $\langle \sigma, s_0 \rangle$, and let us show that the expected payoff of that Nash equilibrium is one of the above values. By 0-optimality, players 1 and 2 have to play deterministically the same actions after all histories ending up in s_0 (state s_2 should not be enabled under $\langle \sigma, s_0 \rangle$). We can reason on the number of histories enabled from s_0. For that, we define $N_{s_0}(\sigma) = \sup\{|h| \mid h \text{ enabled by } \sigma \text{ and } \mathsf{last}(h) = s_0\}$ (this is somehow the maximal number of visits of s_0 enabled by σ).

- If $N_{s_0}(\sigma, s_0) = +\infty$, then the transition $\cdot aa$ from s_0 is never taken (since it would then be played deterministically and it would then stop the game immediately). Since (σ) is 0-optimal, it then means that the terminal state $(1, 4, 4)$ is reached almost-surely (only possibility for Player 0 to get payoff 1).

- Otherwise, $N_{s_0}(\sigma)$ is finite and we reason by induction on this number:
  - first if $N_{s_0}(\sigma, s_0) = 1$, the game ends up immediately in $(1, 3, 5)$.
  - if $N_{s_0}(\sigma, s_0) > 1$, then a transition to s_1 occurs with probability 1. If the path $s_0 \cdot s_1 \cdot (1, 4, 4)$ has probability 1 under σ, then the results holds; otherwise $s_0s_1s_0$ is enabled from s_0, and $\langle \sigma', s_0 \rangle$, with $\sigma' \colon h \mapsto \sigma(s_0s_1h)$, is another 0-optimal Nash equilibrium such that $N_{s_0}(\sigma') < N_{s_0}(\sigma)$. By induction, it has an expected payoff of the form $(1, 4 - x, 4 + x)$, with either $x = \frac{1}{k^n}$ (for $\mathcal{C}_k$) or $x = \frac{1}{n+1}$ (for $\mathcal{D}$) for some $n \in \mathbb{N} \cup \{\infty\}$. If $x = 0$, the results holds immediately, as the expected payoff of $\langle \sigma, s_0 \rangle$ is also $(1, 4, 4)$. Now assume $x > 0$, and consider the game $\mathcal{C}_k$, and the distributions proposed by the strategies σ_1 and σ_2 after s_0s_1. For this to be a Nash equilibrium, both distributions must be uniform; this leads to payoff $(1, 4 - \frac{x}{k}, 4 + \frac{x}{k})$, and proves the result. For $\mathcal{D}$, if it were $\sigma_0(s_0s_1) = b$, then Player 1 would have a profitable deviation. Hence $\sigma_0(a \mid s_0s_1) > 0$, and the best response for players 1 and 2 is to play c. We can then analyze the projections on agents $0, 1$ and $0, 2$ (as done in Section 4) and apply Lemma 7, which yields $(x + 1)\sigma_0(a \mid s_0s_1) \geq x$ and $(x + 1)\sigma_0(a \mid s_0s_1) \leq x$; it follows $\sigma_0(a \mid s_0s_1) = \frac{1}{n+2}$ and $\mathbb{E}^{\sigma}(\phi \mid s_0s_1) = (1, 4 - \frac{1}{n+2}, 4 + \frac{1}{n+2})$. ◄

D Complete undecidability reduction of Section 6

In this annex, we give the complete reduction to show the undecidability of the exsitence of a 0-optimal Nash equilibrium in a three-player game.

We first recall the definition of a two-counter machine as a tuple $\mathcal{M} = \langle Q, q_0, \Delta \rangle$ where:

- Q is a finite set of states,
- $q_0 \in Q$ is an initial state,
- $\Delta \subseteq Q \times \Gamma \times Q$ is the transition table with $\Gamma = \{inc(j), dec(j), zero(j), !zero(j) \mid j \in \{1, 2\}\}$ is the set of operations on counters.

W.l.o.g. we assume that $\mathcal{M}$ never decreases a counter with value 0 (this can be enforced by using non-zero tests $!zero(\cdot)$ before any $dec(\cdot)$ operation).

The semantics of $\mathcal{M} = \langle Q, q_0, \Delta \rangle$ is given as a transition system where configurations are tuples $C = \langle q, c_1, c_2 \rangle \in Q \times \mathbb{N} \times \mathbb{N}$ and for any two configurations $C = \langle q, c_1, c_2 \rangle$ and $C' = \langle q', c'_1, c'_2 \rangle$, for every $\delta = (q, \gamma, q') \in \Delta$, there is a transition $C \rightarrow_\delta C'$ if, and only if:

- $c'_k = c_k + 1$ and $c'_{3-k} = c_{3-k}$, if $\gamma = inc(k)$;
- $c'_k = c_k - 1$ and $c'_{3-k} = c_{3-k}$, if $\gamma = dec(k)$;
- $c_k = 0$ and $(c'_1, c'_2) = (c_1, c_2)$, if $\gamma = zero(k)$;
- $c_k > 0$ and $(c'_1, c'_2) = (c_1, c_2)$, if $\gamma = !zero(k)$.

We fix for the rest of this section a two-counter machine $\mathcal{M} = \langle Q, q_0, \Delta \rangle$, and we build a terminal-reward game $\mathcal{G}_{\mathcal{M}}$ as follows. For every state $q \in Q$ (resp. every $\delta \in \Delta$) we have a subgame $\mathcal{G}_{\mathcal{M}}^q$ (resp. $\mathcal{G}_{\mathcal{M}}^\delta$), as depicted on Figure 17b (resp. Figures 17c to 17f), and an initial subgame as depicted in Figure 17a. The subgames are connected in the obvious way. We write ϕ for the terminal-reward payoff function that is given by $\mathcal{G}_{\mathcal{M}}$.

The main nodes of game $\mathcal{G}_{\mathcal{M}}$ are $Q \cup \Delta$, and all other nodes (belonging to gadgets in grey) are internal nodes. We will evaluate the existence of a 0-optimal Nash equilibrium from all the main nodes of the game. The relation between $\mathcal{M}$ and $\mathcal{G}_{\mathcal{M}}$ is made explicit thanks to the following predicate. Let $s \in Q \cup \Delta$, and $c_1, c_2 \in \mathbb{N}$. We denote by $P(s, c_1, c_2)$ the predicate:

$$\exists \sigma \in \mathbb{S}. \left[\sigma \text{ is a Nash equilibrium from } s \text{ and } \mathbb{E}^\sigma(\phi \mid s) = \left(1, 4 + \frac{1}{2^{c_1} 3^{c_2}}, 4 - \frac{1}{2^{c_1} 3^{c_2}} \right) \right]$$

▶ **Lemma 28.** *Assume $C = \langle q, c_1, c_2 \rangle$ is a configuration of $\mathcal{M}$ such that $P(q, c_1, c_2)$ holds. Then there are a transition δ and a configuration $C' = \langle q', c'_1, c'_2 \rangle$ such that (i) $C \rightarrow_\delta C'$ and (ii) $P(\delta, c_1, c_2)$ and $P(q', c'_1, c'_2)$ hold.*
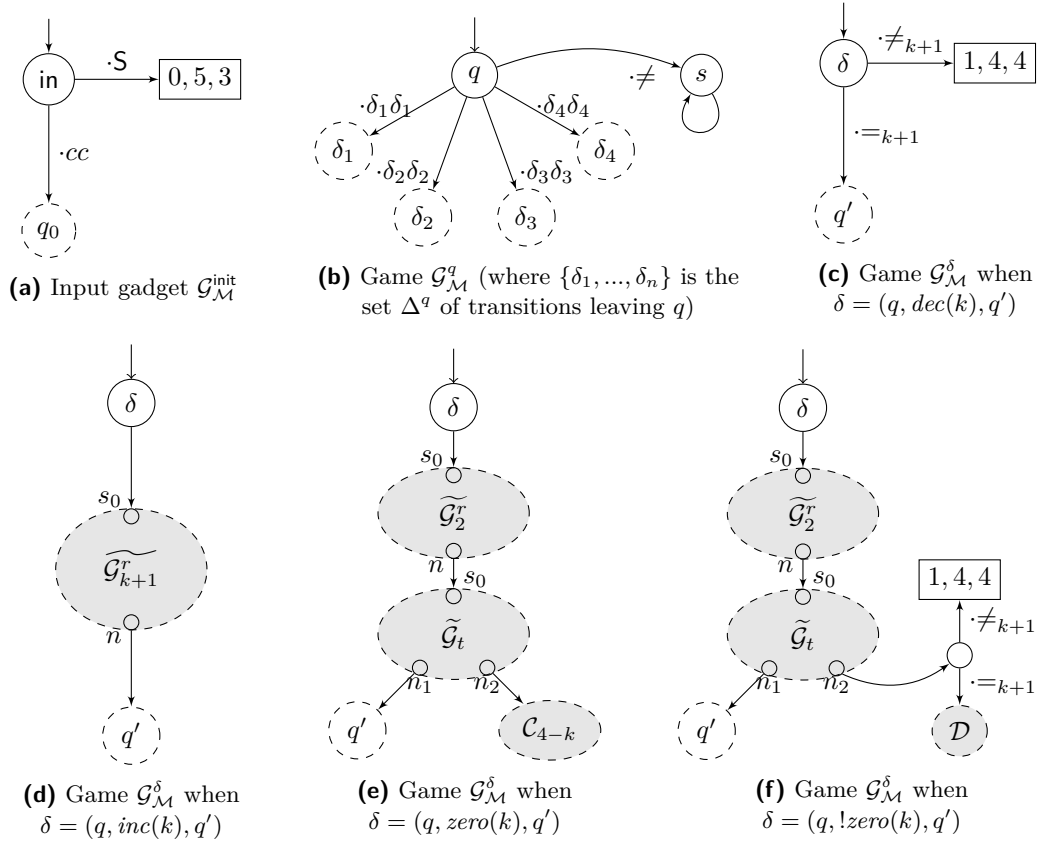
Proof. Write σ for a 0-optimal Nash equilibrium witnessing the truth of predicate $P(q, c_1, c_2)$. In particular,

$$\mathbb{E}^\sigma(\phi \mid q) = \left(1, 4 + \frac{1}{2^{c_1} 3^{c_2}}, 4 - \frac{1}{2^{c_1} 3^{c_2}} \right)$$

In $\mathcal{G}_{\mathcal{M}}$ (more precisely, in subgame $\mathcal{G}_{\mathcal{M}}^q$), only one state $\delta \in \Delta^q$ is activated, otherwise σ would not be 0-optimal, as with positive probability the play would end up in state s of Figure 17b. Hence $P(\delta, c_1, c_2)$ holds as well.

We write $x = \frac{1}{2^{c_1} 3^{c_2}}$, and we distinguish the different cases for δ.

- First assume $\delta = (q, dec(k), q')$. Since $x > 0$, the next state q' has to be activated by σ, and $\langle \sigma, q\delta q' \rangle$ needs to be a 0-optimal Nash equilibrium as well. Applying the analysis of k-action matching-pennies games of Section 3.3, it must be the case that the payoff of σ after $q\delta q'$ is $(1, 4 + y, 4 - y)$ with $y = (k + 1) \cdot x$. If $k = 1$, it is the case that $y = \frac{1}{2^{c_1 - 1} 3^{c_2}}$,

(a) Input gadget $\mathcal{G}_\mathcal{M}^{\mathsf{init}}$

(b) Game $\mathcal{G}_\mathcal{M}^q$ (where $\{\delta_1, ..., \delta_n\}$ is the set Δ^q of transitions leaving q)

(c) Game $\mathcal{G}_\mathcal{M}^\delta$ when $\delta = (q, dec(k), q')$

(d) Game $\mathcal{G}_\mathcal{M}^\delta$ when $\delta = (q, inc(k), q')$

(e) Game $\mathcal{G}_\mathcal{M}^\delta$ when $\delta = (q, zero(k), q')$

(f) Game $\mathcal{G}_\mathcal{M}^\delta$ when $\delta = (q, !zero(k), q')$

Figure 17 Description of the subgames $\mathcal{G}_\mathcal{M}^q$ and $\mathcal{G}_\mathcal{M}^\delta$

and if $k = 2$, it is the case that $y = \frac{1}{2^{c_1} 3^{c_2} - 1}$. Writing $c_k' = c_k - 1$ and $c_{3-k}' = c_{3-k}$, we get that $(q, c_1, c_2) \to_\delta (q', c_1', c_2')$ and that $P(q', c_1', c_2')$.

- Then assume $\delta = (q, inc(k), q')$. Applying Proposition 23, we get that q' is activated by σ, and that there is a 0-optimal Nash equilibrium from q' whose expected payoff is $(1, 4 + y, 4 - y)$ with $y = x/(k + 1)$. As in the previous case, writing $c_k' = c_k + 1$ and $c_{3-k}' = c_{3-k}$, we get that $(q, c_1, c_2) \to_\delta (q', c_1', c_2')$ and that $P(q', c_1', c_2')$.

- Assume $\delta = (q, zero(k), q')$. Applying Proposition 23, we get that the first node s_0 of $\mathcal{G}_t$ in $\mathcal{G}_\mathcal{M}^\delta$ is activated by σ, and that there is a 0-optimal Nash equilibrium from that node whose payoff is $(1, 4 + x/2, 4 - x/2)$. Then, applying Proposition 27, we get that q' and the initial node of $\mathcal{C}_{k+1}$ are activated, and that σ after q' and σ after entering $\mathcal{C}_{k+1}$ are 0-optimal Nash equilibrium. Furthermore, writing $(1, 4 + z, 4 - z)$ and $(1, 4 - y, 4 + y)$ for the payoffs of those equilibria respectively, we should have $z = y$ and $x/2 = z/2$. In particular, $P(q', c_1, c_2)$ holds. Now thanks to Lemma 11, we know that there exists m such that $y = \frac{1}{(4-k)^m}$. This implies that $c_k = 0$: $(q, c_1, c_2) \to_\delta (q', c_1, c_2)$.

- Finally, assume that $\delta = (q, !zero(k), q')$. Again applying Proposition 23, we get that the first node of $\mathcal{G}_t$ in $\mathcal{G}_\mathcal{M}^\delta$ is activated by σ, and that there is a 0-optimal Nash equilibrium from that node whose payoff is $(1, 4 + x/2, 4 - x/2)$. Then, applying Proposition 27, we get that q' and node m are activated, and that σ after q' and σ after entering m are 0-optimal Nash equilibrium. Furthermore, writing $(1, 4 + z, 4 - z)$ and $(1, 4 - y, 4 + y)$ for the payoffs of those equilibria respectively, we should have $z = y$ and $x/2 = z/2$. In particular, $P(q', c_1, c_2)$ holds.

Note that there is a 0-optimal Nash equilibrium from m of payoff $(1, 4 - y, 4 + y)$ with $y > 0$ if, and only if, there is a 0-optimal Nash equilibrium in $\mathcal{D}$ of payoff $(1, 4 - y', 4 + y')$ with $y' = (k + 1) \cdot y$. Now, thanks to Lemma 11, we know that there exists h such that $y' = \frac{1}{h+1}$, hence $y = \frac{1}{(k+1)\cdot(h+1)}$. This implies that $c_k > 0$: $(q, c_1, c_2) \rightarrow_\delta (q', c_1, c_2)$. ◀

We can now show the following correspondence between $\mathcal{M}$ and $\mathcal{G}_\mathcal{M}$:

▶ **Proposition 29.** *The two-counter machine $\mathcal{M}$ has an infinite valid computation if, and only if, there is a 0-optimal Nash equilibrium from state* in *in game $\mathcal{G}_\mathcal{M}$.*

Proof. We use the reduction from the non-halting problem of a two-counter machine we have described. Given a two-counter machine $\mathcal{M}$, we construct game $\mathcal{G}_\mathcal{M}$ as on Figure 9. For technical reasons, we require that each incrementation is followed by a non-zero test: since decrementations are preceded with a zero-test, this enforces infinitely many visits to module $\widetilde{\mathcal{G}}_t$ in $\mathcal{G}_\mathcal{M}$ along any infinite run, so that infinite runs will have probability zero in strategy profiles we will build (we will see that later).

Assume a 0-optimal Nash equilibrium exists from the initial state in. Then it must be the case that its payoff is $(1, 5, 3)$ (in terminal states where Player 0 has reward 1, the sum of the rewards of Player 1 and Player 2 is 8), otherwise there would be a profitable deviation for one of the players. Hence the predicate $P(q_0, 0, 0)$ is true. We can then inductively apply Lemma 28 to build the corresponding valid infinite run of the counter machine.

Conversely assume that $\mathcal{M}$ has an infinite (valid) run $C_0 \rightarrow_{\delta_0} C_1 \rightarrow_{\delta_1} \ldots$ with $C_i = \langle q_i, c_1^i, c_2^i \rangle$, and $c_1^0 = c_2^0 = 0$. We build a strategy profile σ inductively as follows:

- in state in, players 1 and 2 should play c almost-surely;
- we assume we have built σ for the prefix $C_0 \rightarrow_{\delta_0} C_1 \rightarrow_{\delta_1} \ldots C_i$, and that the "main stream" of σ traverses successively the gadgets $\mathcal{G}_\mathcal{M}^{\text{init}}$ $\mathcal{G}_\mathcal{M}^{q_0}$ $\mathcal{G}_\mathcal{M}^{\delta_0}$ $\ldots$ $\mathcal{G}_\mathcal{M}^{\delta_{i-1}}$ and arrives in state q_i, from which we now need to define the strategy profile σ. In state q_i, the players should select transition δ_i almost-surely, and then enter gadget $\mathcal{G}_\mathcal{M}^{\delta_i}$. We now distinguish the possible cases for δ_i:
  - if $\delta_i = (q_{i-1}, dec(k), q_i)$, then players 1 and 2 should play uniformly at random among the $k + 1$ actions;
  - if $\delta_i = (q_{i-1}, inc(k), q_i)$, then, in $\mathcal{G}_{k+1}^r$, the players should follow the strategy described in Proposition 23 with $x = \frac{1}{(k+1)\cdot 2^{c_1^{i-1}}\cdot 3^{c_2^{i-1}}}$;
  - if $\delta_i = (q_{i-1}, zero(k), q_i)$, then, in $\mathcal{G}_2^r$, the players should follow the strategy described in Proposition 23 with $x = \frac{1}{2\cdot 2^{c_1^{i-1}}\cdot 3^{c_2^{i-1}}}$; in $\mathcal{G}_t$, the players should follow the strategy described in Proposition 27, and in $\mathcal{C}_{4-k}$, they should follow the strategies described in the proof Lemma 11, for the correct values of the counters;
  - if $\delta_i = (q_{i-1}, !zero(k), q_i)$, then we apply a strategy as described in the previous item (except that we replace the strategy in $\mathcal{C}_{4-k}$ by that in $\mathcal{D}$).

First, due to the hypothesis on (non-)zero-tests in every syntactic loop of the machine, under the above strategy profile, the game ends up almost-surely in a terminal state, where Player 0 has reward 1. This is because in any subgame $\mathcal{G}_\mathcal{M}^\delta$ where δ is a test-to-zero, or a test-to-nonzero, the game ends up in gadget $\mathcal{C}_{4-k}$ or $\mathcal{D}$ with probability 1/2. For every $\varepsilon > 0$, there is a length N_ε that we can easily compute such that

$$\mathbb{P}^\sigma(\text{reach terminal in no more than } N_\varepsilon \text{ steps} \mid \text{in}) \geq 1 - \varepsilon.$$

We write $\mathcal{G}_\mathcal{M}(N_\varepsilon)$ the game $\mathcal{G}_\mathcal{M}(N_\varepsilon)$ truncated after N_ε computation steps of $\mathcal{M}$, in which we replace any outgoing transitions by a terminal node with reward $(1, 0, 0)$. We note σ_ε the truncated strategy profile.

We have that

$$\mathbb{E}^{\sigma_\varepsilon}(\phi_j \mid \mathsf{in}) \leq \mathbb{E}^{\sigma}(\phi_j \mid \mathsf{in}) \leq \mathbb{E}^{\sigma_\varepsilon}(\phi_j \mid \mathsf{in}) + 8\varepsilon$$

We can now show by induction on $i \leq N_\varepsilon$ that $\mathbb{E}^{\sigma_\varepsilon}(\phi \mid \mathsf{in})$ is $(1, u_1^\varepsilon, u_2^\varepsilon)$ with $|5 - u_1^\varepsilon| \leq 8\varepsilon$ and $|3 - u_2^\varepsilon| \leq 8\varepsilon$, which entails that $\mathbb{E}^{\sigma}(\phi \mid \mathsf{in}) = (1, 5, 3)$. We can then show inductively that for every i,

$$\mathbb{E}^{\sigma}(\phi \mid \mathsf{in} \ \mathcal{G}_{\mathcal{M}}^{\mathsf{init}} \ q_0 \ \mathcal{G}_{\mathcal{M}}^{q_0} \ \delta_0 \mathcal{G}_{\mathcal{M}}^{\delta_0} \ \ldots \ \delta_{i-1} \ \mathcal{G}_{\mathcal{M}}^{\delta_{i-1}} \ q_i) = \left(1, 4 + \frac{1}{2^{c_1^i} 3^{c_2^i}}, 4 - \frac{1}{2^{c_1^i} 3^{c_2^i}}\right)$$

In particular, in any subgame $\mathcal{G}_{\mathcal{M}}^{\delta_i}$, these payoffs yield a local Nash equilibrium.

Assume that σ is not a Nash equilibrium, and pick σ'_j a deviation of Player j (with $j \in \{1, 2\}$) which improves her payoff. By Lemma 18 we can assume that σ'_j is deterministic. Under $\sigma' = \sigma[\sigma'_j]$, we can first notice that the same gadgets are visited than under σ, since Player j cannot improve her payoff by switching the choice of the transitions (gadgets $\mathcal{G}_{\mathcal{M}}^q$ with $q \in Q$). We also realize that in all states of the game, under σ, the choice of Player j is either deterministic (play c) or she plays matching-pennies games uniformly at random against Player $3 - j$. Switching the choice of Player j in a matching-penny game does not change the probabilities of the two output-edges. So only a switch from action c to action s can possibly improve the payoff of Player j.

This is not the case, since by construction, we have a local Nash equilibrium in every gadget. Hence, no deterministic deviation of Player j can improve her payoff. ◄

E   Reachability games: Proof of Corollary 16

We now explain how to extend our main theorem to games with terminal-reachability objectives (in other terms, with terminal payoffs in $\{0, 1\}$). The crucial point to achieve this is that in all our terminal states, the sum of the rewards of players 1 and 2 is 8. Our construction amounts to replacing these terminal rewards with a simple module in which the payoffs of players 1 and 2 are $(8, 0)$ and $(0, 8)$.

▶ **Proposition 30.** *Let $\mathcal{G} = \langle \mathcal{A}, s_0, \phi \rangle$ be a 3-player terminal-reward game such that in any final state s, the terminal payoff $\phi(s) = (x, y, z)$ satisfies the following conditions:*

$$x \in \{0, 1\} \qquad\qquad y, z \in \mathbb{N} \qquad\qquad y + z = 8$$

Then we can construct an arena $\mathcal{A}'$, and sets of (final) states R_0, R_1, R_2 such that $\langle \sigma, s_0 \rangle$ is a 0-optimal Nash equilibrium in $\mathcal{G}$ if, and only if, it is a 0-optimal Nash equilibrium in $\mathcal{G}' = \langle \mathcal{A}', s_0, \phi' \rangle$ where $\phi' = (\mathbb{1}_{R_0}, \mathbb{1}_{R_1}, \mathbb{1}_{R_2})$.

Proof. We replace every final node $(x, y, 8 - y)$ with a zero-sum game as depicted in Figure 18. In this figure, for all $i \in \{1, 2\}$, the set $\mathsf{Allow}_i(v_{x,y})$ of allowed actions is the set $[\![0, 7]\!]$, and $M_y = \{\cdot ij \mid \exists 0 \leq r < y. \ i - j = r \mod 8\}$ (see Figure 18).

By playing uniformly at random, Player 1 can ensure winning (*i.e.*, reaching state $(x, 8, 0)$) with probability $y/8$, whatever Player 2 does. She then gets payoff y. Similarly, Player 2 can ensure winning with probability $1 - y/8$ by playing uniformly. We conclude that $(x, y, 8 - y)$ is the only equilibrium payoff.

We built this way a new game $\mathcal{G}'$ where all final payoffs are of the form $(x, 8, 0)$ or $(x, 0, 8)$. Every Nash equilibrium in $\mathcal{G}$ can be converted into a Nash equilibrium with the same payoff (by playing uniformly in every new node $v_{x,y}$) in $\mathcal{G}'$.

Conversely, if $\langle \sigma, s_0 \rangle$ is a Nash equilibrium in $\mathcal{G}'$, then for every $hv_{x,y} \in \mathsf{States}^+$,
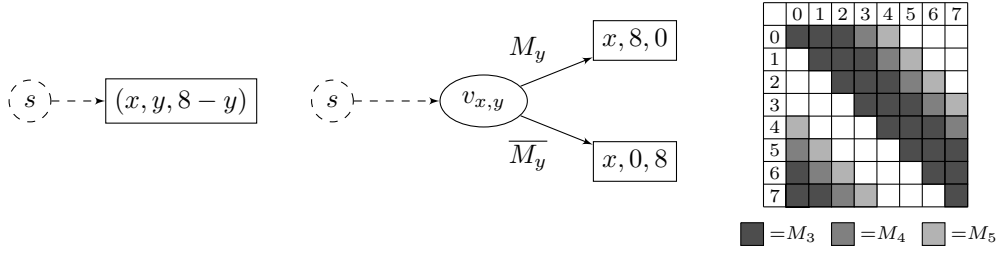
Figure 18 Transformation of a terminal node $(x, y, 8 - y)$ with an intermediate node $v_{x,y}$. The table on the right gives the value of M_y for some values of y (notice that $M_y \subseteq M_{y'}$ when $y \leq y'$, so that for instance $|M_4| = 32$).

- If $hv_{x,y}$ is enabled by σ, we have $\mathbb{E}^\sigma(\phi \mid hv_{x,y}) = (x, y, 8 - y)$;
- Otherwise, $hv_{x,y}$ is not enabled and we can assume $\sigma_i(hv_{x,y})$ is the uniform distribution for both $i \in \{1, 2\}$. This assumption does not change the final reward of the game (as $hv_{x,y}$ is not enabled) and preserves the equilibrium because a deviation of Player 1 in this branch can already ensure a least payoff y (respectively at least $8 - y$ for Player 2).

Finally, every branch ending up in $v_{x,y}$ has payoff $(x, y, 8 - y)$ so $\langle \sigma, s_0 \rangle$ is in fact an equilibrium in $\mathcal{G}$ with the same value.

To conclude, we can divide every terminal reward for players 1 and 2 by 8, so that every final state satisfies $\phi(s) \in \{0, 1\}^{\mathsf{Agt}}$. By linearity, every 0-optimal Nash equilibrium in the original game is a Nash equilibrium in $\mathcal{G}'$ with average payoffs for players 1 and 2 divided by 8. ◄

F Safety games: Proof of Corollary 17

A safety game is described by $\langle \mathcal{A}, s_0, (G_i)_{i\mathsf{Agt}} \rangle$ where $\mathcal{A}$ is an arena, s_0 an initial state and $(G_i)_i \in \left(2^{\mathsf{States}}\right)^{\mathsf{Agt}}$ is a family of goals for every player. An agent i wins the game if she stays in the subset of states G_i, so we define the payoff function $\phi_i^s = \mathbb{1}_{G_i^\omega}$.

Proof of Corollary 17. Proposition 30 along with the reduction presented in Proposition 29 allows us to compute for every two-counter machine $\mathcal{M}$ a concurrent reachability game $\mathcal{G}$, with objectives $R_0, R_1, R_2 \subseteq \mathsf{F}$ such that $\mathcal{M}$ does not halt if, and only if, $\mathcal{G}$ has a 0-optimal Nash equilibrium, with the reward function $\phi^r = (\mathbb{1}_{R_1}, \mathbb{1}_{R_2}, \mathbb{1}_{R_2})$.

We now define safety conditions for this arena by:

$$G_0 = \mathsf{States} \backslash R_0$$
$$\forall i\{1, 2\}. \; G_i = \mathsf{States} \backslash \mathsf{F} \uplus R_i$$

First remark that we defined all internal states as winning for all players so an infinite run is a possible Nash equilibrium for the safety game. Let us now consider the constraint $\mathbb{E}^\sigma(\phi_0^s \mid s_0) = 0$. In the following, we will say that a $\langle \sigma, s_0 \rangle$ is an 0-unsafe Nash equilibrium if it is a Nash equilibrium of the safety game which satisfies the above constraint.

Let us notice that $\phi_0^s \equiv 1 - \phi_0^r$ and for $i \in \{1, 2\}$, $\mathbb{1}_{(\mathsf{States} \backslash \mathsf{F} \uplus R_i)^\omega} \equiv \mathbb{1}_{\mathsf{States}^+ R_i^\omega} + \mathbb{1}_{(\mathsf{States} \backslash \mathsf{F})^\omega}$ The following analysis is mostly concerned with the term $\mathbb{1}_{(\mathsf{States} \backslash \mathsf{F})^\omega}$ that is the difference between reachability objectives and safety objectives. Based on the reduction of Proposition 29, we will show that this term can be neglected.

- If σ is a 0-unsafe Nash equilibrium, we have $\mathbb{E}^{\sigma}(\phi_0^r \mid s_0) = 1$ so σ is 0-optimal (for the reachability objective). Moreover, $R_0 \subset \mathsf{F}$ is reached with probability 1 so $\mathbb{E}^{\sigma}(\mathbb{1}_{(\mathsf{States} \setminus \mathsf{F})^{\omega}} \mid s_0) = 0$. So $\forall i \in \{1, 2\}$, $\mathbb{E}^{\sigma}(\phi_i^s \mid s_0) = \mathbb{E}^{\sigma}(\phi_i^r \mid s_0)$
  For $\sigma_i' \in S_i$, let $\sigma' = \sigma[i/\sigma_i']$, then $\mathbb{E}^{\sigma'}(\phi_i^s \mid s_0) = \mathbb{E}^{\sigma'}(\mathbb{1}_{\mathsf{States}^+ R_i^{\omega}} \mid s_0) + \mathbb{E}^{\sigma'}(\mathbb{1}_{(\mathsf{States} \setminus \mathsf{F})^{\omega}} \mid s_0) \geq \mathbb{E}^{\sigma'}(\phi_i^r \mid s_0)$. So $\mathbb{E}^{\sigma'}(\phi_i^r \mid s_0) \leq \mathbb{E}^{\sigma'}(\phi_i^s \mid s_0) \leq \mathbb{E}^{\sigma}(\phi_i^s \mid s_0) = \mathbb{E}^{\sigma}(\phi_i^r \mid s_0)$.
  Player 0 cannot ensure staying in safe states: any history before leaving to a terminal state is safe, and the only possible deviations to an unsafe state for Player 0 occur in games $\mathcal{G}_k^r$ and $\mathcal{D}$. However players 1 and 2 are then forced to play cc for staying in their own safe states.

- Conversely, if there exists a 0-optimal Nash equilibrium, there exists one constructed from an infinite run of $\mathcal{M}$. Without loss of generality, we can assume such run has infinitely many counter tests, so that the underlying Nash equilibrium enables the testing module $\widetilde{\mathcal{G}_t}$ infinitely often. This strategy profile σ makes both players 1 and 2 play uniformly at random so even if one decides to deviate, there is still a fixed positive probability $\frac{1}{4}$ to branch to submodules n_2 and eventually reach a final state. We conclude from this analysis, that for every deviation $\sigma_i \in \$$, $\mathbb{E}^{\sigma[i/\sigma_i]}(\mathbb{1}_{(\mathsf{States} \setminus \mathsf{F})^{\omega}} \mid s_0) = 0$. Hence σ is resilient to deviations of 1 and 2 for safe objectives. Similarly to the other implication, Player 0 cannot deviate unilateraly to a terminal safe state.
  We conclude that there exists a strategy profile σ that is a 0-unsafe Nash equilibrium. ◄