Winning concurrent reachability games requires doubly-exponential patience

Kristoffer Arnsfelt Hansen[*] Aarhus University

Michal Koucký Czech Academy of Sciences

Peter Bro Miltersen[†] Aarhus University

Abstract

We exhibit a deterministic concurrent reachability game PURGATORY$_n$ with n non-terminal positions and a binary choice for both players in every position so that any positional strategy for Player 1 achieving the value of the game within given $\epsilon < 1/2$ must use non-zero behavior probabilities that are less than $(\epsilon^2/(1-\epsilon))^{2^{n-2}}$. Also, even to achieve the value within say $1 - 2^{-n/2}$, doubly exponentially small behavior probabilities in the number of positions must be used. This behavior is close to worst case: We show that for any such game and $0 < \epsilon < 1/2$, there is an ϵ-optimal strategy with all non-zero behavior probabilities being at least $\epsilon^{2^{O(n)}}$. As a corollary to our results, we conclude that any (deterministic or nondeterministic) algorithm that given a concurrent reachability game explicitly manipulates ϵ-optimal strategies for Player 1 represented in several standard ways (e.g., with binary representation of probabilities or as the uniform distribution over a multiset) must use at least exponential space in the worst case.

1 Introduction

1.1 Dante in Purgatory - a riddle

There are seven terraces in Purgatory, indexed $1, 2, 3, 4, 5, 6, 7$. Dante enters Purgatory at terrace 1. Each day, if Dante finds himself at some terrace $i \in \{1, 2, \ldots, 7\}$, he must play a game of matching pennies against Lucifer: Lucifer hides a penny, and Dante must try to guess if it is heads up or tails up. If Dante guesses correctly, he proceeds to terrace $i + 1$ the next morning - if $i + 1$ is 8, he enters Paradise and the game ends. If, on the other hand, Dante guesses incorrectly, there are two cases. If he incorrectly guesses "heads", he goes back to terrace 1 the next morning. If he incorrectly guesses "tails" the game ends and Dante forever loses the opportunity of visiting Paradise. How can Dante ensure ending up in Paradise with probability at least 3/4? How long should he expect to stay in Purgatory before the game ends in order to achieve this?

The somewhat striking answer to this riddle is that it *is* possible for Dante to go to Paradise with probability at least 3/4, but any strategy achieving this guarantee has the downside that it allows Lucifer to confine Dante to Purgatory for roughly 10^{25} years (In comparison, the current age of the universe is less than 10^{11} years so even playing one move per nanosecond would not help Dante much.) Other strategies guarantee Dante to go to Paradise with probability at least 99% or 99.9999%, but he would have to be even more patient to play these. Details are given in Section 3.

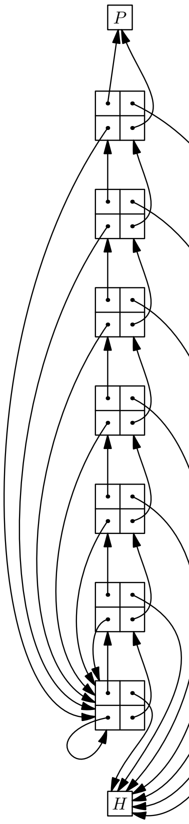

Figure 1. *Dante's Purgatory.*

[*] Supported by a postdoc fellowship from the Carlsberg Foundation.

[†] Work supported by Center for Algorithmic Game Theory, funded by the Carlsberg Foundation.

1.2 Problem statements and results

The central question we address in this paper is: *How patient must one be to play concurrent reachability games games near-optimally?* In the following, we formalize this question, explain how it is motivated by literature on algorithms for solving such games, and we outline our results that partially answers the question.

We consider *finite state, zero-sum, deterministic, concurrent reachability games*. For brevity, we shall henceforth refer to these as just reachability games. The class of reachability games is a subclass of the class of games dubbed *recursive games* by Everett [13] and was introduced to the computer science community in a seminal paper by de Alfaro, Henzinger and Kupferman [10].

A reachability game G is played between two players, Player 1 and Player 2. The game has a finite set of positions, including a special position GOAL. At any point in time during play, some position is the current position. The objective for Player 1 is to eventually make the current position GOAL. The objective for Player 2 is to forever prevent this. GOAL is a terminal position, where the game ends. We also allow additional terminal positions ("traps"). The rest of the positions are non-terminal positions. To each non-terminal position i is associated a set of actions A_i^1, A_i^2 for each of the two players. At each point in time, if the current position is i, Player 1 chooses an action in A_i^1 while Player 2 simultaneously chooses an action in A_i^2. For each position i and each action pair $a \in A_i^1, a' \in A_i^2$ is associated a position $\pi(i, a, a')$. When the current position at time t is i and the players play the action pair (a, a'), the position at time $t+1$ is $\pi(i, a, a')$. The game of Dante in Purgatory (represented graphically in Figure 1) may be viewed as a reachability game with nine positions: One non-terminal position for each terrace, the GOAL position representing Paradise (labeled P in the figure), and one additional non-terminal position H where Dante has lost the game.

A *strategy* for a reachability game is a (possibly randomized) procedure for selecting which action to take, given the history of the play so far. A strategy *profile* is a pair of strategies, one for each player. A *positional strategy* is the special case of a strategy where the choice only depends on the current position. A positional strategy is a family of probability distributions on actions, one distribution for each position. The probability of an action is called a *behavior probability*. A *pure, positional strategy* is the even more special case of this where the choice only depends on the current position and is deterministic. That is, a pure, positional strategy is simply a map from vertices to actions.

An interesting number one may assign to a strategy is its *patience*. We define the patience of a positional strategy to be $1/p$ where p is the *smallest non-zero behavior probability* of any action in any position according to the strategy[1] In particular, a pure positional strategy has patience 1 (the smallest possible patience). The central question we ask in this paper is: What is the patience required for playing reachability games well?

To make this more precise, we should explain what is meant by "well". We let $\mu_i(x, y)$ denote the probability that Player 1 eventually reaches GOAL if the players play using the strategy profile (x, y) and the play starts in position i. The *supinf value* of position i is defined as:

$$\underline{v_i} = \sup_{x \in S^1} \inf_{y \in S^2} \mu_i(x, y)$$

where S^1 (S^2) is the set of strategies for Player 1 (Player 2) Similarly, the *infsup value* of a position i is

$$\overline{v_i} = \inf_{y \in S^2} \sup_{x \in S^1} \mu_i(x, y).$$

The following facts about reachability games are due to Everett (1957), who proved them for the larger class of recursive games:

Fact 1 *For all positions i in a reachability game, the supinf value $\underline{v_i}$ equals the infsup value $\overline{v}_i$ and and this number is therefore called the* value *v_i of that position. The vector v is called the* value vector *of the game.*

Fact 2 *In a reachability game, for any $\epsilon > 0$, there is a positional strategy x^* of Player 1 so that for all positions i*

$$\inf_{y \in S^2} \mu_i(x^*, y) \geq v_i - \epsilon.$$

Similarly, there is a positional strategy y^* of Player 2 so that for all positions i

$$\sup_{x \in S^1} \mu_i(x, y^*) \leq v_i + \epsilon.$$

The strategies x^*, y^* are called *ϵ-optimal*. Note that x^*, y^* do not depend on i. They may however depend on $\epsilon > 0$ and this dependence may be necessary for strategies for Player 1, as shown by examples of Everett. In contrast, in any reachability game (as opposed to the more general recursive games considered by Everett), Player 2 always has an optimal strategy that is guaranteed to achieve the value of the game, without any additive error [11]. Still, such a strategy may involve irrational probabilities, so in computational settings, one may have to settle for ϵ-optimal strategies, even for Player 2. The question we address in this paper is the following: For a worst case reachability game with n positions, what is the minimum patience of ϵ-optimal strategies? In other words, how small behavior probabilities do ϵ-optimal strategies have to use?

[1] While this notion of patience is not completely standard, it in fact goes back to Everett [13, page 77].

Our main results are the following. For any positive integer n, we exhibit a concurrent reachability game PURGATORY$_n$ with n non-terminal positions (Dante's Purgatory described above is PURGATORY$_7$) so that the value of a particular position START is 1 and so that the following holds:

Theorem 3 *Let $n \geq 2$ be an integer and $0 < \delta < 1$ be a real number. Let x be a positional strategy of Player 1 in* PURGATORY$_n$ *such that Player 1 using strategy x is guaranteed to reach GOAL with probability at least δ, against any strategy of Player 2, when play starts in START.*

1. *If $\delta = 2^{-\ell}$ for an integer $1 \leq \ell < n-2$, then the patience of x is at least $2^{2^{n-\ell-2}}$.*
2. *If $\epsilon = 1 - \delta < 1/2$ then the patience of x is at least $\left(\frac{1-\epsilon}{\epsilon^2}\right)^{2^{n-2}}$.*

In short, doubly exponential patience in n is required for playing PURGATORY$_n$ near-optimally. Theorem 3 is shown in Section 2. In Section 3 we show that a similar statement holds even for general (non-positional) strategies and based on this we discuss the striking answer to the riddle about Dante in Purgatory. In Section 4, we show that the patience required for playing PURGATORY$_n$ is, in a sense, close to the worst possible:

Theorem 4 *For any reachability games with a total number of $m \geq 61$ actions in the entire game (collecting actions in all positions belonging to both players), and any $0 < \epsilon < \frac{1}{2}$, Player 1 as well as Player 2 have ϵ-optimal positional strategies with all non-zero behavior probabilities being at least $\epsilon^{2^{42m}}$.*

This latter theorem improves theorems by Chatterjee et al [5, 6, 7] by providing a better dependence on ϵ. It is proved in a similar generic way as these theorems, by appealing to general statements of Basu et al [1] concerning the first order theory of the reals. While the lower bound for the patience required to play PURGATORY$_n$ in some sense is close to the upper bound valid for any reachability game with binary choices and n non-terminal positions (for $\epsilon < 1/2$, the upper bound is $(1/\epsilon)^{2^{O(n)}}$, while the lower is $(1/\epsilon)^{2^{\Omega(n)}}$), they are also tremendously far apart as can be verified by plugging in concrete values of ϵ, n. The upper bound can be improved somewhat without changing the proof at the expense of making it uglier-looking but it will still far from match the lower bound exactly. We leave as an open problem if PURGATORY$_n$ is the reachability game with n non-terminal positions and binary choices that requires the very most patience, or if such games requiring even more patience than PURGATORY$_n$ exist.

Our results are motivated by the recent growing literature by the logic and verification community on *solving* reachability games and *safety games*, the latter games being simply concurrent reachability games with the roles of Player 1 and Player 2 switched[2]). By "solving" a reachability game we may mean a number of different tasks:

1. Determining the set of nodes of value 1. This task has been referred to as *qualitatively* solving the game in the computer science literature.
2. Approximating[3] the value of the game, from below and/or from above. This task has been referred to as *quantitatively* solving the game in the computer science literature.
3. Exhibiting an ϵ-optimal strategy for Player 1 and/or Player 2. This is a stronger notion of solving the game that quantitatively solving it. Indeed, if an ϵ-optimal strategy for Player 1 is given and fixed, we can efficiently approximate the values of all positions of the game from below within ϵ by solving the resulting Markov Decision Process for Player 2. Similarly, if an ϵ-optimal strategy for Player 2 is given, we can efficiently approximate the value of the positions of the game from above. We shall refer to exhibiting an ϵ-optimal strategy as *strongly* solving the game.

Using the terminology above, de Alfaro, Henzinger and Kupferman [10] show that a concurrent reachability game can be qualitatively solved in polynomial time. Etessami and Yannakakis [12] show that it can be quantitatively solved in PSPACE. Chatterjee and co-authors present in a series of papers [8, 4, 5, 6, 7] a series of algorithms that all strongly solve reachability games, by explicitly finding and exhibiting ϵ-optimal strategies (for Player 1 or for Player 2). The strategies obtained use probabilities distributions that are uniform distributions on multisets of actions. As in those works, we use the terminology "k-uniform distribution" to refer to the uniform distribution on a multiset of size k. Note that to satisfy the bound of Theorem 3, a k-uniform distribution that is an ϵ-optimal strategy for PURGATORY$_n$ must have k doubly exponential in n, even for ϵ very close to 1. The default representation of a uniform distribution on a multiset of actions is simply to list the number of times that each possible action appears in the multiset, using binary representation (and no alternative representation has been suggested in the series of papers above). In particular, at least $\log_2 k$ bits is used to represent any k-uniform distribution. This default representation is polynomial time equivalent to representing the behavior probability of each action as a fraction, with the numerator and denominator being integers represented in binary. This latter representation

[2] We shall therefore absorb the discussion on safety games in our discussion on reachability games. Still, the distinction is far from immaterial because of the asymmetry between Players 1 and 2 in a reachability game.

[3] We consider approximations, rather than exact computations, as the value of a reachability game may be an irrational number.

also subsumes fixed point representation, i.e., representing behavior probabilities as finite decimal or binary numbers. Let us refer to all these representations of behavior probabilities as *explicit* representations. With this discussion in mind, an immediate corollary of Theorem 3 is:

Corollary 5 *Any algorithm that takes a reachability game with binary choices as input and manipulates explicitly represented ϵ-optimal strategies for Player 1, even for $\epsilon = 1 - 2^{-n/2}$, where n is the number of positions of the game, uses in the worst case space which is at least exponential in n.*

In particular, any algorithm that strongly solves reachability games and gives an explicitly represented strategy as output uses at least exponential time in the worst case, merely to produce the output. Interestingly, the algorithm of Chatterjee, Majumdar and Jurdziński [8] was originally claimed to be an NP $\cap$ coNP algorithm. The proof of this claim was later found to be incorrect due to subtle issues involving Lipschitz continuity [3], but it was not immediately clear if the correctness and efficiency of the algorithm could be reestablished without significant modifying it. Corollary 5 shows that an entirely different approach would be needed.

While reachability games can be solved qualitatively time-efficiently and quantitatively space-efficiently, we think that solving them in the third strong sense is interesting and important in its own right: For many applications, it is not enough to know that something can be achieved, one also wants to know how to achieve it! Corollary 5 rules out *any* worst case efficient way of doing this, unless we redefine the problem and consider "non-standard" representations of probability distributions. It is interesting to observe that the algorithm of de Alfaro, Henzinger and Kupferman [10] may be used to correctly determine the value of all states in PURGATORY$_n$ as all states except one have value 1 (the last state having trivially value 0). Inspecting its correctness proof, one finds that it is easy to modify the AHK algorithm so that it not only establishes the values, but also "constructs" an ϵ-optimal strategy for Player 1 represented *symbolically*, by assigning to each action either a symbol ϵ_j, $j \in \{1, \ldots, \ell\}$ or as a *formal sum* of the form $1 - \sum_i \epsilon_{j_i}$. The parameterized strategy thus constructed is an ϵ-optimal strategy when the ϵ_j are assigned any sequence of concrete values so that $0 < \epsilon_\ell \cdots \ll \epsilon_3 \ll \epsilon_2 \ll \epsilon_1 \ll \epsilon$, where $\ll$ means "is sufficiently smaller than". This does not contradict Corollary 5, as such sufficiently small values would have exponentially many digits if represented as fractions or decimal numbers. It would be most interesting to prove or disprove that such formal sums containing both actual numbers (for general reachability games, one would need rational numbers other than 1) and symbols representing a sequence of sufficiently small numbers of decreasing magnitude, are sufficient to compactly represent ϵ-optimal strategies for *any* reachability game. Here, by "compactly" we mean in polynomial space in the number of states of the game.

Not only the lower bound on patience of Theorem 3 but also the upper bound of Theorem 4 has relevance for algorithmically solving concurrent reachability games and safety games. In particular, the algorithms of [5, 6, 7] output k-uniform distributions, where k is a parameter supplied by the user. It is not a priori clear for a given game *which* k has to be supplied in order to make it possible to achieve an ϵ-optimal strategy for a desired ϵ. We show in Section 5 that Theorem 4 implies that a certain doubly exponential upper bound on k is sufficient. The upper bound improves similar upper bounds of [5, 6, 7]. Those bounds are also doubly exponential, but our bound has a better dependence on ϵ. Unfortunately (and inevitably, given the Purgatory examples), the bound is still astronomical, even for small games.

1.3 Useful preliminary

The following lemma is well-known:

Lemma 6 *Let x be any fixed strategy of Player 1. let $v_x^i = \inf_{y \in S_2} \mu^i(x, y)$. That is, v_x^i is the value of the game at position i if Player 1's strategy were fixed to x. Then, $v_x^i = \min_{y \in S_2'} \mu^i(x, y)$ where S_2' is the set of positional, pure strategies of Player 2.*

Proof. When Player 1's strategy fixed, the game becomes one of perfect information for Player 2. Then, the statement follows from more general statements, e.g., Liggett and Lippman [15]. □

2 Purgatory

PURGATORY$_n$ is the deterministic reachability game with two terminal positions H and P and n non-terminal positions $\{1, 2, \ldots, n\}$ where the sets of actions associated with position $i \in \{1, \ldots, n\}$ are $A_i^1 = A_i^2 = \{\text{t}, \text{h}\}$ and the transitions from position i associated with actions $a_1 \in A_i^1$ of Player 1 and $a_2 \in A_i^2$ of Player 2 are given in the following table:

a_2 a_1	t	h
t	$i+1$	H
h	1	$i+1$

where we identify the position $n+1$ with position P. The GOAL of Player 1 is to reach position P. For a positional strategy x of Player 1 we will denote by $p_i(x)$ the probability of playing t by Player 1 in the position $i \in \{1, \ldots, n\}$. When x is clear from the context we will omit it. PURGATORY$_7$ (Dante's Purgatory) is depicted in Figure 1. Note that PURGATORY$_n$ essentially consists of a "stack" of n "linked" copies of the game de Alfaro et al [10] call HIDE-AND-RUN, a game that appears also as Example 1 in Everett [13]. Also note that the global structure of PURGATORY$_n$ is reminiscent of examples of random walks on directed graphs with exponential escape time, as applied for instance by Condon [9, Figure 1] to analyze algorithms for simple stochastic games - the latter being *turn-based* rather than concurrent reachability games.

One can verify by induction on n that the value of all positions in PURGATORY$_n$ is 1, except for the value of H, which is 0. In particular, the following strategy for Player 1 ensures that Player 1 wins PURGATORY$_n$ with probability $\geq 1 - \epsilon$. Let the probabilities of playing t by Player 1 be as follows.

$$p_1 = \epsilon^{2^{n-1}}$$
$$p_i = \frac{\epsilon^{2^{n-i}} - \epsilon^{2^{n-i+1}}}{1 - \epsilon^{2^{n-i+1}}} \qquad \text{for } i > 1$$

We will prove by induction that the subgame consisting of positions $1, \ldots, i$, where position $i+1$ is identified with GOAL is won with probability at least $1 - \epsilon^{2^{n-i}}$. This is clearly true for $i = 1$.

For the induction step, assume that the subgame consisting of positions $1, \ldots, i$ is won with probability $1 - \epsilon^{2^{n-i}}$, and consider the subgame consisting of positions $1, \ldots, (i+1)$ and any pure strategy of Player 2. If Player 2 plays h at position $i+1$ Player 1 will win the game with probability at least

$$\begin{aligned} &1 - \epsilon^{2^{n-i}} - (1 - \epsilon^{2^{n-i}})p_{i+1} \\ =&1 - \epsilon^{2^{n-i}} - (\epsilon^{2^{n-(i+1)}} - \epsilon^{2^{n-(i+1)+1}}) = 1 - \epsilon^{2^{n-(i+1)}}. \end{aligned}$$

On the other hand, if Player 2 plays t at position $i+1$, then using Proposition 7 Player 1 will win the game with probability at least

$$\begin{aligned} &1 - \frac{\epsilon^{2^{n-i}}}{\epsilon^{2^{n-i}} + (1 - \epsilon^{2^{n-i}})p_{i+1}} \\ =&1 - \frac{\epsilon^{2^{n-i}}}{\epsilon^{2^{n-(i+1)}}} = 1 - \epsilon^{2^{n-(i+1)}}. \end{aligned}$$

Proposition 7 *Let $0 < p, q < 1$ be real numbers. Consider the Markov chain given in Fig. 2. If we start in state s of the chain then with probability $\frac{q}{q+p-pq}$ we reach the state H and with the remaining probability we reach the state P.*

Proof. Clearly, $\Pr[\text{ reaching } H] = q\sum_{i=0}^{\infty}(1-q)^i(1-p)^i = \frac{q}{1-(1-q)(1-p)} = \frac{q}{q+p-pq}$ □

Alternatively, the following simple strategy of Player 1 achieves winning probability in PURGATORY$_n$ at least $1 - \epsilon$: in position i play t with probability $\epsilon^{2^{n-i+1}}$. We now turn to Theorem 3, which is a consequence of the following statement.

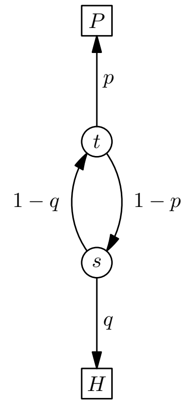

Figure 2. *Two-level Purgatory.*

Theorem 8 *Let $n \geq 2$ be an integer and $0 < \delta < 1$ be a real number. Let x be a positional strategy of Player 1 in* PURGATORY$_n$ *such that the probability that Player 1 reaches P using strategy x starting from position 1 is at least δ, against any counter-strategy of Player 2.*

1. *If an integer $1 \leq \ell(n) < n-2$ satisfies $\delta = 2^{-\ell(n)}$ then for all $i \in \{1, \ldots, n - \ell(n) - 1\}$, $0 < p_i(x) \leq 2^{-2^{n-\ell(n)-1-i}}$.*
2. *If $\epsilon = 1 - \delta < 1/2$ then for all $i \in \{1, \ldots, n-1\}$, $0 < p_i(x) \leq \left(\frac{\epsilon^2}{1-\epsilon}\right)^{2^{n-i-1}}$.*

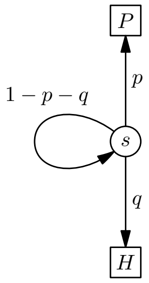

Figure 3. *Loop.*

Proposition 9 *Let $0 < p, q, \delta < 1$ be real numbers. Consider the Markov chain given in Fig. 3. The probability of reaching state P from state s is at least δ iff $p \geq \frac{\delta}{1-\delta}q$.*

Proof. Clearly, starting from S we reach P or H with probability 1. The probability of reaching P is precisely

$\frac{p}{p+q}$. Thus

$$\frac{p}{p+q} \geq \delta$$

iff

$$\begin{aligned} p &\geq \delta(p+q) \\ p &\geq \frac{\delta}{1-\delta} q. \end{aligned}$$

□

Proof of Theorem 8. Part 1. If for some $i \in \{1, \ldots, n\}$, $p_i(x)$ were either 0 or 1 then clearly the value of PURGATORY$_n$ for Player 1 using x would be zero. So all p_i's are non-zero. Assume the claim is false, though. Let $i \in \{1, \ldots, n - \ell(n) - 1\}$ be the largest i such that $p_i(x) > 2^{-2^{n-\ell(n)-1-i}}$. Consider the following strategy of Player 2: in positions $1, \ldots, i$ play h always, in positions $i+1, \ldots, n - \ell(n) - 1$ play t always and in positions $n - \ell, \ldots, n$ play t and h independently with probability $1/2$ each. We claim that the probability of reaching P from position 1 using these strategies is smaller than δ. Let p be the probability of reaching P and q be the probability of reaching H starting from position 1 without returning to position 1. We claim that $p < q\delta/(1-\delta)$ which gives the contradiction by Proposition 9.

To reach P the game has to pass through position $n - \ell(n)$. Clearly, the probability of reaching P starting from position $n - \ell(n)$ without returning to position 1 is precisely $2^{-\ell(n)-1}$. Furthermore, as $p_j \leq 2^{-2^{n-\ell(n)-1-j}}$ for $j = i+1, \ldots, n - \ell(n) - 1$, the probability of reaching P from position $i+1$ is:

$$\begin{aligned} & 2^{-\ell(n)-1} \prod_{j=i+1}^{n-\ell(n)-1} p_j(x) \\ \leq \;& 2^{-\ell(n)-1} \prod_{j=i+1}^{n-\ell(n)-1} 2^{-2^{n-\ell(n)-1-j}} \\ \leq \;& 2^{-\ell(n)-1} \cdot 2^{-\sum_{j=i+1}^{n-\ell(n)-1} 2^{n-\ell(n)-1-j}} \\ = \;& 2^{-\ell(n)} \cdot 2^{-2^{n-\ell(n)-1-i}}. \end{aligned}$$

Hence, $p \leq 2^{-\ell(n)} \cdot 2^{-2^{n-\ell(n)-1-i}}$. However, with probability $q \geq p_i(x) > 2^{-2^{n-\ell(n)-1-i}}$ we reach H as Player 2 plays h in position i. That implies the contradiction.

Part 2. For the second part, the strategy of Player 2 will be similar. Since Player 2 could play always h in position n, $p_n(x) \leq \epsilon$. We pick the largest $i \in \{1, \ldots, n-1\}$ such that $p_i(x) > \left(\frac{\epsilon^2}{1-\epsilon}\right)^{2^{n-i-1}}$. Player 2 will play h in positions $1, \ldots, i$ and t otherwise. The probability p of reaching P from position $i+1$ is given by:

$$\begin{aligned} & \prod_{j=i+1}^{n} p_j(x) \\ \leq \;& \epsilon \cdot \prod_{j=i+1}^{n-1} \left(\frac{\epsilon^2}{1-\epsilon}\right)^{2^{n-j-1}} \\ \leq \;& \epsilon \cdot \left(\frac{\epsilon^2}{1-\epsilon}\right)^{\sum_{j=i+1}^{n-1} 2^{n-j-1}} \\ = \;& \frac{1-\epsilon}{\epsilon} \cdot \left(\frac{\epsilon^2}{1-\epsilon}\right)^{2^{n-i-1}}. \end{aligned}$$

Since Player 2 plays h in position i, the probability of reaching H starting from position 1 without returning to 1 is $q \geq p_i(x) > \left(\frac{\epsilon^2}{1-\epsilon}\right)^{2^{n-i-1}}$. Thus $\frac{1-\epsilon}{\epsilon} q > p$ and we obtain the desired contradiction. □

3 Adaptive strategies and the answer to the riddle

The previous section dealt with non-adaptive strategies for PURGATORY$_n$. This is a natural restriction, as Everett proved that non-adaptive strategies are sufficient for approximately achieving the value of recursive games (Fact 2). It is rather easy to solve the riddle of the introduction based on the material of the last section, *if* one assumes that Dante plays by a non-adaptive strategy. Still, one could speculate that by observing Lucifer's moves, Dante could adapt his strategy to escape sooner while maintaining a particular winning probability. So, to give an answer to the riddle, we should consider adaptive strategies. It is worth pointing out that in more general classes of stochastic games, one might be able to approximate the value of the game by adaptive strategies only. A celebrated example of this phenomenon is the *Big Match* [2], where the near-optimal strategies involve observing the behavior of the opponent and adjusting behavior probabilities accordingly. Thus, adaptive strategies have shown their power in other contexts and are worth investigating. However, the flavor of the results of this section is that adaptive strategies will not allow Dante to escape much faster.

We say that an adaptive strategy for a reachability game *involves* an action with behavior probability at most μ if there is a sequence of actions of both players that will lead the game to a position in which one of the actions given by the strategy has behavior probability at most μ.

Theorem 10 *Let $0 < \delta < 1$ be a real number and $1 \leq \ell(n) < n - 2$ be integers. Let x be an adaptive strategy of Player 1 for* PURGATORY$_n$ *game such that the value of the game for Player 1 using strategy x starting from position 1 is at least δ.*

1. *If $\delta = 2^{-\ell(n)}$ then the strategy x must involve some action with non-zero behavior probability $\leq 2^{-2^{n-\ell(n)-2}}$.*

2. *If $\epsilon = 1 - \delta < 1/2$ then the strategy x must involve some action with non-zero behavior probability $\leq \left(\frac{\epsilon^2}{1-\epsilon}\right)^{2^{n-2}}$.*

Proof. Part 1. Assume that no action under strategy x has behavior probability in the range $(0, 2^{-2^{n-\ell(n)-2}}]$. We will design a strategy y for Player 2 which will achieve against x probability of reaching P smaller than δ. Player 2 will define his strategy y in phases as the game proceeds. A new phase will start whenever the game is at position 1. Player 2 defines his strategy so that in a given phase, the probability q of reaching H and the probability p of reaching P satisfy $\delta q \geq p$. Given that the strategy of Player 2 has this property it is clear that the game ends in P with probability less than δ.

At the beginning of a given phase Player 2 looks at the strategy of Player 1 that will be used within the next up-to n moves of the game. For each $i \in \{1, \ldots, n\}$ and $w \in \{\mathrm{t}, \mathrm{h}\}^{i-1}$, Player 2 calculates the probabilities $p_i(w)$ of Player 1 playing t in position i conditioned on the event that Player 2 plays the first $i-1$ moves in this phase according to the sequence w of actions and the phase did not end yet. (Conditioned on the event that Player 2 plays the first $i-1$ moves in this phase according to w and the phase did not end yet there is a unique sequence of actions of Player 1 that must have been taken during the first $i-1$ moves. Hence $p_i(w)$ is well defined.)

Player 2 decides his actions for positions $i = n, n-1, \ldots, 1$ (i.e. in backward order). For positions $i = n, n-1, \ldots, n-\ell(n)$, Player 2 will play both t and h with probability $1/2$ each. Clearly, if the game reaches position $n-\ell(n)$ then this strategy of Player 2 ensures that the probability of reaching P in this phase is precisely $2^{-\ell(n)-1}$. Next we define the strategy of Player 2 for positions $i = n-\ell(n)-1, n-\ell(n)-2, \ldots, 1$. Assume that Player 2 already decided for each sequence of actions $w \in \{\mathrm{t}, \mathrm{h}\}^i$ what would be his next moves after playing such a sequence of actions in this phase and let $v_{i+1}(w)$ be the probability that the game would reach P if Player 2 plays like that. Player 2 will define his strategy so to maintain the invariant $v_{i+1}(w) \leq \delta 2^{-2^{n-\ell(n)-1-i}}$. Player 2 is going to decide his actions in position i and possibly the overall strategy.

If for some $w \in \{\mathrm{t}, \mathrm{h}\}^{i-1}$, $p_i(w) \geq 2^{-2^{n-\ell(n)-1-i}}$ then Player 2 starting from position 1 will play according to wh and then continue as decided earlier so that the probability of reaching P after passing through position $i+1$ will be $v_{i+1}(w\mathrm{h}) \leq \delta 2^{-2^{n-\ell(n)-1-i}}$. This will guarantee the required relationship between probabilities of reaching H and P in this phase since the only way to reach P in this phase is to pass through position $i+1$ and with probability at least $2^{-2^{n-\ell(n)-1-i}}$ the game will end by entering H from position i.

If for all $w \in \{\mathrm{t}, \mathrm{h}\}^{i-1}$, $p_i(w) < 2^{-2^{n-\ell(n)-1-i}}$ then Player 2 will play t in position i regardless of his previous actions in this phase. Since he plays t, the probability of reaching P from position i in this phase will be $v_i(w) = p_i(w) \cdot v_{i+1}(w\mathrm{t}) < 2^{-2^{n-\ell(n)-1-i}} \cdot \delta 2^{-2^{n-\ell(n)-1-i}} \leq \delta 2^{-2^{n-\ell(n)-1-(i-1)}}$ for all $w \in \{\mathrm{t}, \mathrm{h}\}^{i-1}$. Next Player 2 continues to define his actions for positions $i-1, \ldots, 1$. As the strategy x of Player 1 does not involve actions with behavior probability in the range $(0, 2^{-2^{n-\ell(n)-2}}]$ the process must stop at some point and we obtain a good strategy for Player 2.

Part 2. The proof is similar to the proof of Part 1. Player 2 maintains that a given phase ends in H with probability q and in P with probability p where $q > \frac{\epsilon}{1-\epsilon}p$. To do so he keeps $v_{i+1}(w) \leq \frac{1-\epsilon}{\epsilon}\left(\frac{\epsilon^2}{1-\epsilon}\right)^{2^{n-i-1}}$ for all $i = n-1, n-2, \ldots, 1$ and $w \in \{\mathrm{t}, \mathrm{h}\}^{i-1}$, and he looks for w with $p_i(w) > \left(\frac{\epsilon^2}{1-\epsilon}\right)^{2^{n-i-1}}$. He starts to build his strategy as follows. If $p_n(w) > \epsilon$ for some $w \in \{\mathrm{t}, \mathrm{h}\}^{n-1}$ then his strategy will be given by wh. Otherwise $p_n(w) \leq \epsilon$ for all $w \in \{\mathrm{t}, \mathrm{h}\}^{n-1}$, and his strategy will be to play t in position n for all w. Thus $v_n(w) \leq \epsilon = \frac{1-\epsilon}{\epsilon}\left(\frac{\epsilon^2}{1-\epsilon}\right)^{2^0}$. He continues to build inductively his strategy as in the previous part but keeping the stated invariants. □

The previous theorem is interesting in connection with the following general theorem.

Theorem 11 (No quick exit strategy) *Let $0 < \delta, \mu < 1$ and $0 < \tau$. Let a reachability game G have the property that any strategy of Player 1 with value at least δ involves an action with non-zero behavior probability smaller or equal to μ. Then no strategy of Player 1 guarantees winning the game with probability at least $(1+\tau)\delta$ by plays of length at most $\tau\delta/\mu$.*

Proof. By contradiction. Let G be a game with the required property and x be a strategy of Player 1 which for an arbitrary strategy y of Player 2 reaches GOAL with probability at least $(1+\tau)\delta$ by plays of length at most $\tau\delta/\mu$. For a fixed strategy y of Player 2 consider the winning plays of length at most $\tau\delta/\mu$. Among these plays some of them invoke actions with behavior probability smaller or equal to μ. As the length of these plays is at most $\tau\delta/\mu$ the total contribution towards the winning probability of these plays involving the small probability actions is at most $\mu \cdot \tau\delta/\mu = \tau\delta$. Hence the probability of plays of length at most $\tau\delta/\mu$ that reach GOAL and do not involve any of the small probability actions is at least δ. We claim that this contradicts the assumed properties of the game.

If the strategy of Player 1 were modified as to set to zero behavior probabilities of all actions with behavior probability at most μ and the remaining probabilities were renormalized evenly as to sum to one at each game position then the total probability of plays winning against strategy y of length at most $\tau\delta/\mu$ would still be at least δ, as the probability of the plays not involving the small probability actions may only increase. Since this is true for any strategy y of Player 2 and the modification is always the same we conclude that Player 1 has a strategy that does not involve any action with behavior probability at most μ which guarantees reaching GOAL with probability at least δ. This contradicts the original assumption about G. □

We may now answer the riddle. In order to have $3/4$ probability of success, Dante has to use a strategy involving behavior probabilities smaller than $\left(\frac{0.25^2}{0.75}\right)^{32} \approx 3.4 \cdot 10^{-34}$ (by Part 2 of Theorem 10). In case Lucifer uses the strategy suggested in the above proofs, Dante would spend $3 \cdot 10^{33}$ days, i.e., $8 \cdot 10^{30}$ years, in Purgatory with overwhelming probability, as the number of steps of play would be geometrically distributed. From the preceding theorem, we can derive a somewhat weaker bound on the expected time needed to win the game: for 0.73 winning probability Dante has to use probabilities smaller than $\left(\frac{0.27^2}{0.73}\right)^{32} \approx 9.6 \cdot 10^{-33}$. Setting $\delta = 0.73$ and $\tau = 0.01$ in the previous theorem we get that with probability at least 0.01, the time to win the game is at least $7.6 \cdot 10^{29}$ days which is $2 \cdot 10^{27}$ years. Hence the expected time to win the game is at least $2 \cdot 10^{25}$ years. This would constitute a real Purgatory for Dante! Clearly, we have demonstrated that although one can win the game almost surely, from a practical standpoint it is not winable. A good practical strategy for Dante could be to flip a fair coin and escape the purgatory with probability $1/128$ within a week.

4 Upper bound on patience of both players in all reachability games

In this section we prove Theorem 4, stated in the introduction. We use a theorem of Filar et al. [14] applicable to *stochastic games* [16] with limiting average payoffs [15]. Everett's recursive games (and hence, concurrent reachability games) can be seen as a special case of these. We first state the theorem almost verbatim. Afterwards, we explain the notation and how to apply the theorem to the special case of concurrent reachability games.

Theorem 12 (Filar et al., Theorem 4.1) *Let a two-player, zero-sum, stochastic game be given with set of states S, set of actions A_s^i for player $i \in \{1, 2\}$ in state s, transition function q and reward function r. Let $\hat{f}$ be a behavior strategy profile for the two players. If there exist a feasible setting $\hat{v}, \hat{t}$ of the remaining variables of the following non-linear progam N so that the objective function has value of ϵ or less, then the strategies in $\hat{f}$ are ϵ-optimal. Conversely, if the strategies in $\hat{f}$ are ϵ-optimal, then there exist $\hat{v}$ and $\hat{t}$ such that $(\hat{v}, \hat{f}, \hat{t})$ are feasible in N with objective value $2|S|\epsilon$ or less. The program N has variables $\{v_s^k\}_{k\in\{1,2\},s\in S}$, $\{f_s^k(a)\}_{k\in\{1,2\},s\in S,a\in A_a^k}$, $\{t_a^k\}_{k\in\{1,2\},s\in S}$, objective function $\sum_{s\in S}(v_s^1 + v_s^2)$ (to be minimized) and constraints:*

1. $\forall s \in S, a \in A_s^1$:
   $v_s^1 \geq \sum_{s'\in S} v_{s'}^1 q(s'|s, f^{\langle 1,s,a\rangle})$,
2. $\forall s \in S, a \in A_s^1$:
   $v_s^1 + t_s^1 \geq r_s^1(f^{\langle 1,s,a\rangle}) + \sum_{s'\in S} t_{s'}^1 q(s'|s, f^{\langle 1,s,a\rangle})$,
3. $\forall s \in S, a \in A_s^2$:
   $v_s^2 \geq \sum_{s'\in S} v_{s'}^2 q(s'|s, f^{\langle 2,s,a\rangle})$,
4. $\forall s \in S, a \in A_s^2$:
   $v_s^2 + t_s^2 \geq r_s^2(f^{\langle 2,s,a\rangle}) + \sum_{s'\in S} t_{s'}^2 q(s'|s, f^{\langle 2,s,a\rangle})$,
5. $\forall k \in \{1, 2\}, s \in S : \sum_{a\in A_s^k} f_s^k(a) = 1$,
6. $\forall k \in \{1, 2\}, s \in S, a \in A_s^k : f_s^k(a) \geq 0$.

In the above theorem, $f_s^k(a)$ is the probability that Player k puts on action a in state s according to the profile f while $f^{\langle k,s,a\rangle}$ is the strategy profile obtained from f by altering Player k's behavior in state s so that he puts his entire probability mass on action a.

Expressions of the form $q(s'|s, f')$ is the probability that the state at time $t + 1$ is s' given that the state at time t is s when the players play according to profile f'. Note that for the case of reachability games and all settings of f' occuring in the program N, the expression $q(s'|s, f')$ is simply a sum of products of two variables from $\{f_s^k(a)\}$.

In stochastic games, players receive *rewards* during play. Expression of the form $r_s^k(f')$ is the expected reward Player k receives in state s when Player plays using profile f'. A concurrent reachability game can be modeled as a stochastic game by simply letting $r_s^1(f')$ be 0 when s is not the GOAL state and 1 when s is the GOAL state, no matter what f' is and letting $r_s^2(f') = -r_s^1(f')$.

Applying Theorem 12 to a concurrent reachability game in this way, we see that Fact 2 (the existence of ϵ-optimal strategies) implies that the program N has feasible solutions with arbitrarily small strictly positive value of the objective function. Given a concrete ϵ, we let $M = \lceil 1/\epsilon \rceil$ and add to N the constraint that M times the value of the objective function is at most 1. We know that the resulting program has a feasible solution $(\hat{v}, \hat{f}, \hat{t})$, and that the $\hat{f}$ in any such solution is a pair of ϵ-optimal strategies. Now, *freeze* all variables in $\hat{f}$ that have value 0 to *constants* in N and for all remaining variables $f_s^k(a)$ add a variable $g_s^k(a)$ and a constraint $g_s^k(a) f_s^k(a) = 1$. Let N' be the resulting program. By construction, we already have a feasible solution

$(\hat{v}, \hat{f}, \hat{g}, \hat{t})$ to N', where $\hat{f}$ is the zero-reduced vector. Also, a pair of ϵ-optimal strategies to the game can be obtained from any feasible solution by extending $\hat{f}$ by zeros. The largest $\hat{g}$ value is the patience of the most patient of these strategies. Thus, we want to show that N' has a feasible solution where non-zero values $\hat{g}_s^k(a)$ are not extremely large.

We observe that if m is the total number of actions of both players in the game, then N' has at most $4n + 2m$ variables and $3m + 4n + 1$ constraints, all constraints being inequalities involving polynomials of degree at most three and all coefficients being in $\{-1, 0, 1, M\}$. We now appeal to a version of Proposition 1.3.5 of Basu *et al* [1]. Unfortunately, the published version of this Proposition contains "big-O"s and we want to obtain the "big-O"-less bound of Theorem 4. From personal communication with Basu, we have obtained a non-asymptotic version of the proposition. This non-asymptotic version will appear in a forthcoming publication by Basu and Roy (if the reader prefers to rely on published information only, the original Propostion 1.3.5 of [1] still yields a bound of $\epsilon^{2^{O(n)}}$ on the patience). The non-asymptotic version of the proposition is as follows. Let $\text{bit}(\cdot)$ denote bitsize, i.e., $\text{bit}(j) = \lfloor \log_2(|j|) + 1 \rfloor$.

Proposition 13 (Basu et al., Basu and Roy) *Given a set of s polynomials of degree at most d in k variables with integer coefficients of bitsize at most τ, the ball centered in origin with radius $2^{\tau' 2k(d+4)(d+5)(2d+6)^2(2d+5)^{2(k-1)}}$ where $\tau' = 33 + 2\max(\tau, k\text{bit}(s)) + (k+11)\text{bit}(d+5) + 11\text{bit}(k+1) + 3\text{bit}(5(2d+6)(2d+5)^{k-1})$ intersects the realization of every realizable sign condition on the set of polynomials.*

The statement "N' is feasible" is equivalent to a sign condition on the set of cubic polynomials defining the constraints of N'. To apply Proposition 13, we merely have to plug in the upper bound on the number of variables and polynomials of N' and let $d = 3$. Doing so, we obtain the desired bound of Theorem 4 and are done.

5 Consequences for k-uniform strategies

By k-uniform strategies we understand strategies in which behavior probabilities are given by k-uniform distributions.

Lemma 14 *Let P and P' be matrices of absorbing Markov chains on the same set of n states with the property that all non-zero transition probabilities (of both chains) are bigger than or equal to α. Let $\delta = \epsilon(\ln(4/\epsilon))^{-1} n^{-1} \alpha^n$ and suppose that the entries p_{ij}, p'_{ij} of P, P' satisfy*

$$|p_{ij} - p'_{ij}| \leq \delta.$$

Let b_{ij}, b'_{ij} be the absorption probabilities of the two chains. Then,

$$|b_{ij} - b'_{ij}| \leq \epsilon$$

Proof. For any time t, conditioned on being in any particular state, the conditional probability that each chain will be absorbed within the next n steps is at least α^n. The probability that each chain is not absorbed after $k(1/\alpha)^n$ steps is therefore at most $(1 - \alpha^n)^{k(1/\alpha)^n} \leq e^{-k}$. So we only have to run each chain for $M = \ln(4/\epsilon)(1/\alpha)^n$ steps to approximate an absorption probability from below within $\epsilon/4$. This goes for both chains. Let the absorption probabilities thus approximated be denoted $\bar{b}_{ij}, \bar{b}'_{ij}$. That is, we have for each i, j that $|b_{ij} - \bar{b}_{ij}| \leq \epsilon/4$ and $|b'_{ij} - \bar{b}'_{ij}| \leq \epsilon/4$.

For a fixed state i, the total variation distance between the distributions p_{i*} and p'_{i*} are at most $\delta n/2$. From this it follows that total variation distance between the distributions $\bar{b}_{i*}$ and $\bar{b}'_{i*}$ is at most $M\delta n/2 \leq \epsilon/2$. In particular, for any i, j, $|\bar{b}_{ij} - \bar{b}'_{ij}| \leq \epsilon/2$.

The conclusion of the lemma follows. □

Theorem 15 *For any concurrent reachability game or concurrent safety game with a total number of $m \geq 61$ actions, and for any given $0 < \epsilon \leq \frac{1}{2}$, there is a k-uniform, ϵ-optimal positional strategy for each player, where*

$$k \leq (1/\epsilon)^{2^{43m}}$$

Proof. Let $l = \log(2/\epsilon)2^{42m}$. By Theorem 4, there is an $\epsilon/2$-optimal strategy of patience at most 2^l. Round all behavior probabilities of this strategy to $4lm$ binary digits, by rounding all probabilities except the largest one in each distribution upwards, and rounding the largest probability in each distribution down by the total amount that the rest were rounded up. As the largest probability in each distribution is at least $1/m$, the rounded value is non-negative, and hence the rounded distributions are still probability distributions, and the family of rounded distributions is a positional strategy. This is a k-uniform strategy for the stated magnitude of k. We claim that it is ϵ-optimal. Indeed, when the strategy of Player 1 is fixed, we can fix a bet reply of Player 2 which is pure, by Lemma 6. With both strategies thus fixed, the dynamics of play is a Markov chain. We can assume without loss of generality that this is an absorbing Markov chain, as states from which it is impossible to reach GOAL can be replaced by absorbing states. Now apply Lemma 14, with $\epsilon/2$ substituted for ϵ. □

Acknowledgements

We sincerely thank Kousha Etessami for sharing with us significant insights about concurrent reachability games and the literature about them and for much encouragement.

References

[1] BASU, S., POLLACK, R., AND ROY, M.-F. On the combinatorial and algebraic complexity of quantifier elimination. *J. ACM 43*, 6 (1996), 1002–1045.

[2] BLACKWELL, D., AND FERGUSON, T. The big match. *Annals of Mathematical Statistics* (1968), 159–163.

[3] CHATTERJEE, K. On Nash equilibria in stochastic games, errata. http://www.eecs.berkeley.edu/ ~ c_krish/ publications/ errata-csl04.pdf.

[4] CHATTERJEE, K., DE ALFARO, L., AND HENZINGER, T. A. Strategy improvement for concurrent reachability games. In *Quantitative Evaluation of Systems, 2006. QEST 2006. Third International Conference on* (2006), pp. 291–300.

[5] CHATTERJEE, K., DE ALFARO, L., AND HENZINGER, T. A. Strategy improvement for concurrent safety games. *CoRR abs/0804.4530* (2008).

[6] CHATTERJEE, K., DE ALFARO, L., AND HENZINGER, T. A. Termination criteria for solving concurrent safety and reachability games. *CoRR abs/0809.4017* (2008).

[7] CHATTERJEE, K., DE ALFARO, L., AND HENZINGER, T. A. Termination criteria for solving concurrent safety and reachability games. In *Proceedings of the Twenteeth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'09)* (2009).

[8] CHATTERJEE, K., MAJUMDAR, R., AND JURDZIŃSKI, M. On Nash equilibria in stochastic games. In *CSL 2004* (2004), J. Marcinkowski and A. Tarlecki, Eds., vol. 3210 of *LNCS*, Springer-Verlag, p. 2640.

[9] CONDON, A. On algorithms for simple stochastic games. *Advances in Computational Complexity Theory, DIMACS Series in Discrete Mathematics and Theoretical Computer Science 13* (1993), 51–73.

[10] DE ALFARO, L., HENZINGER, T. A., AND KUPFERMAN, O. Concurrent reachability games. *Theor. Comput. Sci. 386*, 3 (2007), 188–217.

[11] DE ALFARO, L., AND MAJUMDAR, R. Quantitative solution of omega-regular games. *J. Comput. Syst. Sci. 68*, 2 (2004), 374–397.

[12] ETESSAMI, K., AND YANNAKAKIS, M. Recursive concurrent stochastic games. In *ICALP (2)* (2006), M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, Eds., vol. 4052 of *Lecture Notes in Computer Science*, Springer, pp. 324–335.

[13] EVERETT, H. Recursive games. In *Contributions to the Theory of Games Vol. III*, H. W. Kuhn and A. W. Tucker, Eds., vol. 39 of *Annals of Mathematical Studies*. Princeton University Press, 1957.

[14] FILAR, J. A., SCHULTZ, T. A., THUIJSMAN, F., AND VRIEZE, O. J. Nonlinear programming and stationary equilibria in stochastic games. *Math. Programming 50*, 2, (Ser. A) (1991), 227–237.

[15] LIGGETT, T. M., AND LIPPMAN, S. A. Stochastic games with perfect information and time average payoff. *SIAM Review 11*, 4 (1969), 604–607.

[16] SHAPLEY, L. S. Stochastic games. *Proceedings of the National Academy of Sciences, U.S.A.*, 39 (1953), 1095–1100.