Finding Equilibria: Simpler for Pessimists, Simplest for Optimists

Léonard Brice
Université Libre de Bruxelles, Belgium

Thomas A. Henzinger
Institute of Science and Technology Austria, Klosterneuburg, Austria

K. S. Thejaswini
Institute of Science and Technology Austria, Klosterneuburg, Austria

Abstract

We consider equilibria in multiplayer stochastic graph games with terminal-node rewards. In such games, Nash equilibria are defined assuming that each player seeks to maximise their expected payoff, ignoring their aversion or tolerance to risk. We therefore study risk-sensitive equilibria (RSEs), where the expected payoff is replaced by a *risk measure*. A classical risk measure in the literature is the *entropic risk measure*, where each player has a real valued parameter capturing their risk-averseness. We introduce the *extreme risk measure*, which corresponds to extreme cases of entropic risk measure, where players are either extreme optimists or extreme pessimists. Under extreme risk measure, every player is an extremist: an extreme optimist perceives their reward as the maximum payoff that can be achieved with positive probability, while an extreme pessimist expects the minimum payoff achievable with positive probability. We argue that the extreme risk measure, especially in multi-player graph based settings, is particularly relevant as they can model several real life instances such as interactions between secure systems and potential security threats, or distributed controls for safety critical systems. We prove that RSEs defined with the extreme risk measure are guaranteed to exist when all rewards are non-negative. Furthermore, we prove that the problem of deciding whether a given game contains an RSE that generates risk measures within specified intervals is decidable and NP-complete for our extreme risk measure, and even PTIME-complete when all players are extreme optimists, while that same problem is undecidable using the entropic risk measure or even the classical expected payoff. This establishes, to our knowledge, the first decidable fragment for equilibria in simple stochastic games without restrictions on strategy types or number of players.

2012 ACM Subject Classification Theory of computation → Models of computation

Keywords and phrases Nash equilibria, stochastic games, graph games, risk-sensitive equilibria

Digital Object Identifier 10.4230/LIPIcs.MFCS.2025.30

Related Version *Full Version*: https://arxiv.org/abs/2502.05316

Funding This work is a part of project VAMOS that has received funding from the European Research Council (ERC), grant agreement No 101020093.

Acknowledgements We thank anonymous reviewers for pointing us to the Hurwicz criterion and to the work of Gallego-Hernández and Mansutti [13]. We thank Marie van den Bogaard for her valuable feedback on the first author's PhD dissertation, which helped improve the quality of this work.

1 Introduction

Stochastic systems have been used extensively in several areas including verification [11], learning theory [1], epidemic processes [20] to name a few. Several real-world systems however do not work with a centralised control. Therefore, modelling using stochastic systems with multiple agents makes for more faithful abstractions of such systems without a centralised control. Some examples of fields in which multi-agent stochastic modelling include cyber

50th International Symposium on Mathematical Foundations of Computer Science (MFCS 2025).
Editors: Paweł Gawrychowski, Filip Mazowiecki, and Michał Skrzypczak; Article No. 30; pp. 30:1–30:18
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

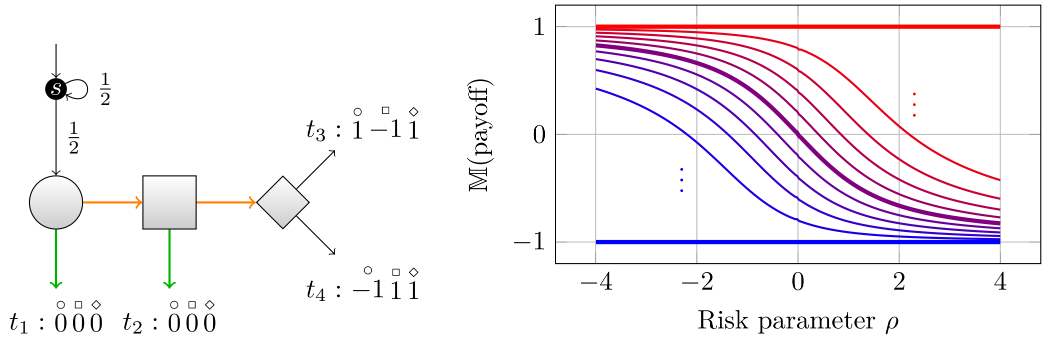
physical systems [29], distributed and probabilistic computer programs [8], probabilistic planning [30]. In such cases, the problem of reasoning about multiple agents with several, often times orthogonal objectives, becomes important. For multi-agent systems modelled with stochasticity on the underlying arena, a fundamental question to ask is the existence or finding of an equilibrium – typically, an important decision problem in such games is the *constrained existence problem* that asks whether a given game contains an equilibrium where each player's (perceived) payoff lies within a given interval. The most popular equilibria in literature are Nash equilibria (NEs) [23], where each player plays optimally to maximise their expected payoff, with regard to the other players' strategies. However, Nash equilibria come with their own downsides. In terms of computational complexity, the cost of the constrained existence problem of NEs in stochastic games is prohibitively expensive, or even undecidable in the general case [31]. But more importantly, NEs do not faithfully model agents in real world settings since they do not consider their tolerance or averseness to risk.

An example. We illustrate this claim with the help of a simple game, depicted in Figure 1a. Each vertex other than the start vertex is owned by one player among the players $\{○, □, ◇\}$ and a play in this game refers to a sequence of moves of a token along the edges of the graph. The initial vertex s denotes a stochastic vertex, which moves the token to the next vertex with probability $\frac{1}{2}$ at each step. When the token is on a vertex owned by a player i, that player decides where the token goes next, and when a terminal vertex (vertex t_1, t_2, t_3, or t_4) is reached, the numbers at the terminal vertex indicate the payoff obtained by each player.

If the token moves out of vertex s, player ○ then chooses between a green (↓) or an orange action (→). If the action → was selected, the player □ must make a similar choice (between actions ↓ and →). If both players selected the action →, then the player ◇ is given the opportunity to reach a terminal that offers negative reward to either player ○ or □, while the other player gets an additional profit (payoff 1 instead of 0).

A strategy profile is a *Nash equilibrium* (NE) if no player can increase their expected payoff by changing their strategy as long as the other players stick to theirs. If player ◇ does not use randomisation, there is no NE in this game where she gets an expected payoff 1, since along any play that reaches deterministically either the vertex t_3 or t_4, one of the two players ○ or □ gets the expected payoff 0 and has an incentive to deviate to the action ↓ at their vertex. However, using randomisation, there is such an NE if both player ○ and □ choose action →, and player ◇ chooses at random to go either to vertex t_3 or t_4, each with probability $\frac{1}{2}$. Then, both player ○ and player □ have expected payoff 1, the same as if they reach the terminals vertices t_1 or t_2. However, if both the players ○ and □ were modelling, say, safety-critical systems, such a Nash equilibrium is dubious because neither player can afford, even a low probability of, the reward -1 as they do not have this tolerance to *risk*. Such a strategy profile would not be stable even if only one of the two players ○ or □, let alone both, were *risk-averse*. However, the strategy where player ○ choses the action →, player □ the action ↓, and player ◇ randomises between her choices with equal probability, is a potential equilibrium in the case where player □ is risk averse. On the other hand, if all players ○, □, and ◇ were *risk loving*, say they model entities like hackers of a system that are happy with *some chance* at a small positive reward, the same strategy would *not* be an equilibrium, as player □ would rather gamble for chance at the reward $+1$, despite the risk of a negative reward, thus deviating from the action ↓ instead to the "risky" alternative →.

This example thus generalises a line of reasoning that is often made in finance or decision theory: consider a participant of a certain lottery where they lose \$100 with probability $\frac{99}{100}$, and wins \$10000 with probability $\frac{1}{100}$. While the expected payoff is positive (\$1), expectation alone is insufficient as a decision criterion to participate in the lottery.



(a) A 3-player stochastic game with a stochastic vertex.

(b) Entropic risk measures for player ○ for varying risk-parameters.

Figure 1 Entropic risk measure.

Risk measures. A *risk measure* captures the perception that a player has of what their payoff will be. Thus, it generalises the notion of expected payoff. Various risk measures exist in the literature, and have been used extensively in the field of economics and finance. They include expected shortfall (ES), value at risk (VaR) [2], variance [4], entropic risk measure (ER) [10]. In terms of graph modelling, these risk measures have been studied mainly over Markov decision processes (MDPs) using variance (along with mean) as a risk measure [9, 27, 21], ES [28, 19, 22] (also referred to as conditional value at risk (CVaR), average value at risk (AVaR), expected tail loss (ETL), and superquantile in literature) and ER [15, 7, 3]. Studying the entropic risk measure in MDPs appears more practical compared to ES or using variance-penalised risk measures, due to the intractable exponential memory [14] and time required to compute optimal strategies [27] for the other measures, even in MDPs. On the other hand, when the risk measure used is ER, players have optimal positional strategies in MDPs [16].

Entropic risk measure. The entropic risk measure is computed by assigning a *risk parameter* to an agent, i.e., a value $\rho \in \mathbb{R}$. The entropic risk measure of a random variable X is then defined as $\mathbb{M}_\rho[X] = -\frac{1}{\rho}\log_e\left(\mathbb{E}\left[e^{-\rho X}\right]\right)$ [12] for $\rho \neq 0$. Assume X is a player's payoff. If the risk parameter ρ is positive, the corresponding players are risk-averse and therefore more weight is given to worse payoffs. Conversely, players with a negative ρ are risk-loving. When ρ tends to 0, the entropic risk measure converges to the classical expectation $\mathbb{E}[X]$.

Consider again the game in Figure 1a, and the strategy profile in which player ○ and player □ choose action →, and player ◇ goes to the vertex t_3 with probability $p \in [0, 1]$, and to vertex t_4 with probability $1 - p$. Then, player ○'s perception of such a strategy profile, depending on her risk parameter ρ, is defined by the entropic risk measure $\mathbb{M}[\mu_\circ] = -\frac{1}{\rho}\log_e\left(pe^{-\rho} + (1-p)e^{\rho}\right)$. Figure 1b illustrates this formula, where each curve captures the perceived payoff for a fixed probability p, the thick blue line corresponds to the case $p = 0$, the thick red line to $p = 1$, and mixtures of blue and red curves to intermediary cases – the thick purple curve corresponds to $p = \frac{1}{2}$. Note that this purple curve reaches exactly the value 0 when $\rho = 0$. This is because 0 is the expected payoff of player ○ when $p = \frac{1}{2}$, and the entropic risk measure when $\rho = 0$ – defined as the limit when ρ approaches zero – corresponds exactly to the expected payoff. For higher values of ρ, the entropic risk measure is lower, since it corresponds to cases where player ○ is more risk-averse and focuses on the event of reaching the vertex t_4; for symmetric reasons, for lower values of ρ, the values of entropic risk measure are higher.

Extreme risk measure. Consider the perception of risk our agents may have in the previous example. If indeed such entities were modelling safety-critical agents where rewards below a certain threshold are unacceptable, such agents usually do not accept *any* strategy profile where they have any positive probability of such rewards, disregarding what the probabilities actually are, behaving as *extremely pessimistic* players – they assume that the worst of the cases that may happen. On the other hand, we may model certain external environment factors, or hackers of a system as an *extremely optimistic* player: they are happy with a small probability of success – they can repeat such attempts with a large number of trials – and seek to maximise the best payoff they may receive with positive probability.

We define the *pessimistic risk measure* of a random variable X as the supremum of values x such that X almost surely takes values above x, or its essential infimum ($\operatorname{ess\,inf}$); the *optimistic risk measure* is defined symmetrically (using $\operatorname{ess\,sup}$). In Figure 1b, the reader may have observed that all curves (except the thick blue and the thick red ones) converge to the value -1 when ρ tends to $+\infty$, and to the value 1 when ρ tends to $-\infty$. These extreme cases correspond to the pessimistic and the optimistic risk measure, respectively, and we group them under the umbrella term *extreme risk measure* (XR).

Risk-sensitive equilibria. *Risk-sensitive equilibria* are equilibria in multi-player games defined using a specified risk measure for each player, where no player can improve the risk measure of their payoff by deviating unilaterally from their strategy. Risk-sensitive equilibria takes into account the risk-sensitivities of a player to ensure that a player does not have an incentive to deviate. When the risk measure used is the entropic risk measure, we call such equilibria *entropic risk-sensitive equilibria (ERSEs).* When using our novel risk measure – the extreme risk measure – we refer to them as *extreme risk-sensitive equilibria (XRSEs).*

▶ **Example 1.** In the game depicted by Figure 1a, if players ○ and □ are extreme pessimists, and player ◇ is an extreme optimist, then in order for player ◇ to get the risk measure 1, it suffices that either the terminal vertex t_3 or t_4 is reached with positive probability. But such a strategy profile cannot be an XRSE since there is a play such that either player ○ or □ gets the payoff -1 with positive probability and, as a pessimist, they get then the risk measure 0 if they deviate to using the ↓. They both have a profitable deviation by avoiding player ◇'s vertex. The only XRSEs in this game are therefore the strategy profiles that almost-surely reach terminals t_1 or t_2, leaving then player ◇ with the risk measure 0.

Risk-sensitive equilibria is particularly relevant in security contexts, where, for example, defenders may act conservatively while attackers behave opportunistically. Similarly, in autonomous systems, different components – such as safety controllers, efficiency optimizers, or compliance monitors – may operate under independent objectives and distinct risk profiles. We study ERSEs and XRSEs in a general enough framework where it might also be suitable to model a variety of situations including epidemic processes and probabilistic planning, where equilibria on graph games are considered. Incorporating risk measures while considering equilibria enables more realistic and nuanced models of such multi-agent interactions.

Our results. We consider *simple quantitative multiplayer stochastic games played on graphs*, that is, games in which the payoffs that the players receive depend on the *terminal vertex* that is reached – and in which an infinite play is associated to the the payoff zero for all players. In such games, we consider two questions in particular: the existence and the complexity of the constrained existence problem of both ERSEs and XRSEs.

In Section 3, we show that the constrained existence problem of ERSEs is undecidable, extending from the same problem for Nash equilibria in the work of Ummels and Wojtczak [31]. However, building on results on the two-player zero-sum case by Baier et al. [3], we find restrictions on strategies to recover decidability. Using known results about NEs, we also show that, when the rewards are all nonnegative, such an equilibrium always exists.

We define *extreme risk measure (XR)* as a novel risk measure to consider in multi-agent stochastic systems. In Section 4, we show that our new definition is robust, since it exactly captures the well-studied entropic risk measure when the risk parameters tend to $\pm\infty$. Our main technical contributions are about the extreme risk measure, and *extreme risk-sensitive equilibria (XRSEs).* We show that the constrained existence problem of XRSEs is decidable and NP-complete. The technical crux of proving NP membership lies in proving that if an XRSE satisfying the constraints exists, then there exists one that has finite memory, and a polynomial representation. Such a succinct representation then can therefore can be guessed and verified in polynomial time. We show that the problem remains NP-complete when strategies are required to be stationary, pure, or positional. Finally, we show that if all players are extreme optimists, the problem is PTIME-complete.

As we do for ERSEs, we also prove the existence of XRSEs in games with nonnegative rewards – and we show that there exists such a *stationary* equilibrium (where players use no memory and only randomness) that can be algorithmically constructed in polynomial time.

Related work. Hurwicz criterion is used in decision theory in situations where probabilities are unknown, and it assigns for a given random variable X and a parameter $\alpha \in [0,1]$, the objective of maximising the quantity $\alpha \max(X) + (1-\alpha)\min(X)$ [17, 5], which generalises Wald's maximin criterion [32] that constitutes to one extreme ($\alpha = 0$) of Hurwicz criterion. Our definition of pessimistic risk measure corresponds to Wald's maximin criterion *when only outcomes of positive probability are considered*, while our pessimistic risk measure corresponds to the other extreme of Hurwicz's criterion ($\alpha = 1$). Even classical notions such as considering the *worst-case scenario* obtained by abstracting any stochastic environment and treating it instead as an adversary can be seen as Wald's maximin criterion applied on *all outcomes, including probability* 0 *events.*

To the best of our knowledge, equilibria defined using any risk measure, and in particular, the entropic risk measure, have not been studied in multiplayer graph games. Two works have considered risk in the specific case of two-player zero-sum games on stochastic arenas. The first of these works is by Bruyère, Filiot, Randour, and Raskin [6] who have studied *"beyond worst-case synthesis problem"* which considers some measures that address risk averseness in a player in the context of synthesis. They consider *"strongly risk-averse"* strategies, those which avoid outcomes below a certain threshold and maximise the expected payoff, resulting in an entirely different risk measure. Baier, Chatterjee, Meggendorfer, and Piribauer have considered two-player zero-sum stochastic games with total-reward objectives (payoff is the sum of the rewards seen along the way), when one player is risk-sensitive and wants to optimise their entropic risk measure [3]. They show that computing optimal strategies can be done in PSPACE (when the base e is replaced by an algebraic number). If the base of the exponent is e, computing optimal strategies is in 3EXPTIME due to a recent result of Gallego-Hernández and Mansutti [13].

The two-player zero-sum case is a specific case of the constrained existence problem in two-agent systems, where the payoff functions of the two agents as well as their risk parameters are exactly the negation of each other. On the other hand, another subcase of the constrained existence problem of equilibria defined using the entropic risk measure is of course the constrained existence problem of NEs, when the risk parameters of all agents are set to 0, which is known to be undecidable from the work of Ummels and Wojtczak [31].

2 Preliminaries

We assume that the reader is familiar with the basics of probability and graph theory. However, we define some concepts for establishing notation.

Probabilities. Given a (finite or infinite) set of outcomes Ω and a probability measure $\mathbb{P}$ over Ω, let X be a random variable over Ω, that is, a mapping $X : \Omega \to \mathbb{R}$. We then write $\mathbb{E}^{\mathbb{P}}[\mathbb{X}]$, or simply $\mathbb{E}[X]$, for the expectation of X, when it is defined. Given a finite set S, a *probability distribution* over S is a mapping $d : S \to [0, 1]$ that satisfies the equality $\sum_{x \in S} d(x) = 1$. We write $\mathsf{Supp}(d)$ for the *support* of the distribution d, that is, the set of elements $x \in S$ such that $d(x) > 0$.

Risk measures. Given a set Ω of outcomes, a *risk measure* over Ω is a mapping M which maps a probability measure $\mathbb{P}$ over Ω and a random variable X to a real value $M^{\mathbb{P}}[X]$. Sometimes, in the literature, risk measures are expected to have the following three properties: (1) they are *normalised*, i.e., we have $M^{\mathbb{P}}[0] = 0$; (2) they are *monotone*, i.e., the pointwise inequality $X \leqslant Y$ implies $M^{\mathbb{P}}[X] \leqslant M^{\mathbb{P}}[Y]$; and (3) they are *translative*, i.e., $M^{\mathbb{P}}[X + c] = M^{\mathbb{P}}[X] + c$ for every constant c. In particular, the expectation of a random variable $\mathbb{E}$ satisfies those properties. We will not need those properties, and only state them here to remark that the risk measures we consider satisfy them. Note that translativity sometimes refers to the opposite of the definition given here, i.e., the property $M^{\mathbb{P}}[X + c] = M^{\mathbb{P}}[X] - c$ for every c. This is a matter of whether we use a risk measure or its negation.

Graph, paths, games. A directed graph (V, E) consists of a set of *vertices* V and a set of ordered pair of vertices, called *edges*, E. In a directed graph (V, E), for each vertex u, we write $E(u)$ to denote the set $E \cap (\{u\} \times V)$. For simplicity, we often write uv for an edge $(u, v) \in E$. A *path* in the directed graph (V, E) is a (finite or infinite) word $\pi = \pi_0\pi_1 \ldots$ over the alphabet V such that $\pi_n\pi_{n+1} \in E$ for every n such that π_n and π_{n+1} exist. We write $\mathsf{Occ}(\pi)$ for the set of vertices occurring along π, and $\mathsf{Inf}(\pi)$ for those that occur infinitely often, if there are any. The prefix $\pi_0 \ldots \pi_n$ is written as $\pi_{\leqslant n}$ or $\pi_{<n+1}$, and the suffix $\pi_n\pi_{n+1} \ldots$ is written as $\pi_{\geqslant n}$ or $\pi_{>n-1}$. A finite path $\pi = \pi_0 \ldots \pi_n$ is *simple* if every vertex occurs at most once along π. It is a *cycle* if its last vertex π_n is such that $\pi_n\pi_0 \in E$.

▶ **Definition 2** (Game). *A* game *is a tuple* $\mathcal{G} = (V, E, \Pi, (V_i)_{i\in\Pi}, \mathsf{p}, \mu)$*, where we have:*

- *a directed graph* (V, E)*, called the* underlying *graph of* $\mathcal{G}$*;*
- *a finite set* Π *of* players*;*
- *a partition* $(V_i)_{i\in\Pi\cup\{?\}}$ *of the set* V*, where* V_i *denotes the set of vertices* controlled *by player* i*, and the vertices in* $V_?$ *are called* stochastic vertices*;*
- *a* probability function $\mathsf{p} : E(V_?) \to [0, 1]$*, such that for each stochastic vertex* s*, the restriction of* p *to* $E(s)$ *is a probability distribution;*
- *a mapping* $\mu : T \to \mathbb{R}^{\Pi}$ *called* payoff function*, where* T *is the set of* terminal vertices*, i.e. vertices of the graph* (V, E) *that have no outgoing edges. We also write* μ_i*, for each player* i*, for the function that maps a terminal vertex* t *to the* i^{th} *coordinate of the tuple* $\mu(t)$*.*

In a more general framework, payoffs can be assigned to all infinite paths. Here, we only focus on what is usually called *simple quantitative games*, i.e. games in which the underlying graph contains terminal vertices and the payoffs depend only on which terminal vertex is eventually reached. We thus extend the mapping μ to the set $(V \backslash T)^{\omega} \cup (V \backslash T)^*T$ by defining $\mu(v_1 \ldots v_k t) = \mu(t)$, and $\mu(v_1v_2 \ldots) = (0)_{i\in\Pi}$ (if no terminal vertex is reached, everyone gets the payoff 0). A game is *Boolean* if all payoffs belong to the set $\{0, 1\}$.

An *initialised game* is a tuple $(\mathcal{G}, v_0)$, usually written $\mathcal{G}_{\upharpoonright v_0}$, where $v_0 \in V$ is an *initial vertex*. In what follows, when the context is clear, we use the word *game* also for an initialised game. We often assume that we are given a game $\mathcal{G}_{\upharpoonright v_0}$ and implicitly use the same notations as in the definition above.

▶ **Example 3.** The example of initialised game given in Figure 1a is redrawn in Figure 2a. There, the vertex a is controlled by player ○, the vertex b by player □, and the vertex c by player ◇. The vertex s is stochastic, with probability $\frac{1}{2}$ of going to a and probability $\frac{1}{2}$ back to the stochastic vertex s.

▶ **Definition 4** (Markov decision process, Markov chain)**.** *A (initialised or not)* Markov decision process *is a game with one player. A* Markov chain *is a game with zero players.*

Histories and plays. We call *play* a path in the underlying graph that is infinite, or whose last vertex is a terminal vertex. Other paths are called *histories*. We will then use the notations $\mathsf{Hist}(\mathcal{G})$ to denote finite paths in the graph of the game, and $\mathsf{Plays}(\mathcal{G})$ to denote both finite and infinite paths. For a history $h = h_0 \dots h_n$, we write $\mathsf{last}(h) = h_n$. We will also write $\mathsf{Hist}_i(\mathcal{G})$ for the set of histories whose last vertex is controlled by player i. A history or play in an initialised game $\mathcal{G}_{\upharpoonright v_0}$ is a history or play in $\mathcal{G}$ whose first vertex is v_0.

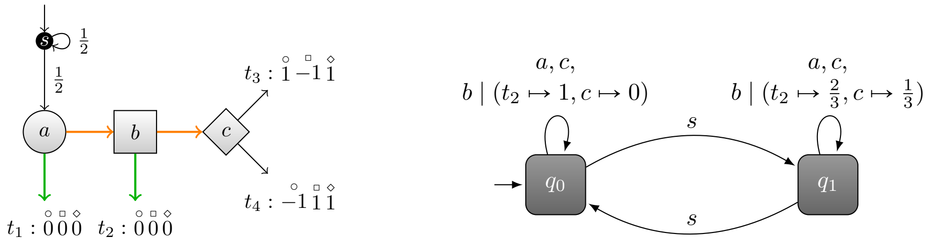
Strategies, and strategy profiles. In a game $\mathcal{G}_{\upharpoonright v_0}$, a *strategy* for player i is a mapping σ_i that maps each history $hu \in \mathsf{Hist}_i(\mathcal{G}_{\upharpoonright v_0})$ to a probability distribution over $E(u)$. The set of possible strategies for player i in $\mathcal{G}_{\upharpoonright v_0}$ is written as $\mathsf{Strat}_i(\mathcal{G}_{\upharpoonright v_0})$. A path $\pi_0\pi_1 \dots$ (be it a history or a play) is *compatible* with the strategy σ_i if for each k such that $\pi_k \in V_i$, the probability that the strategy σ_i proposes h_{k+1} after history $h_{\leqslant k}$ is positive, that is, we have $\sigma_i(h_{\leqslant k})(h_{k+1}) > 0$. A *strategy profile* for a subset $P \subseteq \Pi$ is a tuple $(\sigma_i)_{i \in P}$. A strategy profile for the set P of players is written $\bar{\sigma}_P$, or simply $\bar{\sigma}$ when $P = \Pi$. We also write $\bar{\sigma}_{-i}$ for $\bar{\sigma}_P$ where $P = \Pi \backslash \{i\}$. Similarly, we use (σ_{-i}, σ'_i) to denote the strategy profile $\bar{\tau}$ defined by $\tau_i = \sigma'_i$ and $\tau_j = \sigma_j$ for $j \neq i$. We sometimes write $\bar{\sigma}(hv)$ to mean $\sigma_i(hv)$ where i is the player controlling v, or $\mathsf{p}(v)$ when $v \in V_?$.

For some history h, and a strategy σ_i, we define the strategy truncated to a history h, written $\sigma_{i \upharpoonright hv}$, as the strategy $\sigma'_i : h' \mapsto \sigma_i(hh')$ in the game $\mathcal{G}_{\upharpoonright v}$.

A strategy profile $\bar{\sigma}_{-i}$ in the game $\mathcal{G}_{\upharpoonright v_0}$ defines an initialised Markov decision process $\mathcal{G}_{\upharpoonright v_0}(\bar{\sigma}_{-i})$, where the vertices of the (infinite) underlying graph are the histories of $\mathcal{G}_{\upharpoonright v_0}$ and the edges are added from hu to each history huv iff $uv \in E$. Similarly, a strategy profile $\bar{\sigma}$ for Π defines an initialised Markov chain $\mathcal{G}_{\upharpoonright v_0}(\bar{\sigma})$. Thus, it also defines a probability measure $\mathbb{P}_{\bar{\sigma}}$ over plays – which turns the payoff functions μ_i into random variables.

Pure, stationary, and positional strategies. We say that a strategy σ_i is *pure* when for each history hu, there is a vertex v such that $\sigma_i(hu)(v) = 1$; then we often just write $\sigma_i(hu) = v$. We say that σ_i is *stationary* when for every two histories $hu, h'u \in \mathsf{Hist}_i(\mathcal{G}_{\upharpoonright v_0})$, we have $\sigma_i(hu) = \sigma_i(h'u)$. In that case, we sometimes assume that strategy σ_i is defined in every game $\mathcal{G}_{\upharpoonright u}$ and simply write $\sigma_i(u)$ for $\sigma_i(hu)$. Finally, the strategy σ_i is *positional* when it is pure and stationary. Those concepts are naturally generalised to strategy profiles.

Memory structures. A *memory structure* for player i is a tuple $(S_i, s_0, \delta_i, \nu_i)$, where S_i is a finite set of *memory states*, where $s_0 \in S_i$ is an *initial state*, where δ_i is a *memory-update mapping* that maps each pair $(s, v) \in S_i \times V$ to a memory state s', and where ν_i is an *output mapping* that maps each pair $(s, v) \in S_i \times V_i$ to a distribution d over $E(v)$. The


s
$\frac{1}{2}$
$\frac{1}{2}$
a
b
c
$t_3 : \overset{\circ}{1}\,\overset{\square}{-1}\,\overset{\diamond}{1}$
$t_4 : \overset{\circ}{-1}\,\overset{\square}{1}\,\overset{\diamond}{1}$
$t_1 : \overset{\circ}{0}\,\overset{\square}{0}\,\overset{\diamond}{0}$
$t_2 : \overset{\circ}{0}\,\overset{\square}{0}\,\overset{\diamond}{0}$
$a, c,$
$b \mid (t_2 \mapsto 1, c \mapsto 0)$
$a, c,$
$b \mid (t_2 \mapsto \frac{2}{3}, c \mapsto \frac{1}{3})$
s
s
q_0
q_1

(a) Figure 1a redrawn for convenience.

(b) A memory structure.

Figure 2 A game and a memory structure.

memory-update mapping can be extended to a mapping $\delta_i^* : \mathsf{Hist}(\mathcal{G}_{\restriction v_0}) \to S_i$ with $\delta_i^*(\varepsilon) = s_0$ and $\delta_i^*(hu) = \delta_i(\delta_i^*(h), u)$ for each history hu. The memory structure then induces a strategy σ_i defined by $\sigma_i(hu) = \nu_i(\delta_i^*(h), u)$ for each history $hu \in \mathsf{Hist}_i\mathcal{G}_{\restriction v_0}$. A strategy induced by a memory structure is called *finite-memory strategy.* Note that stationary strategies are exactly the strategies that are induced by a memory structure with $|S_i| = 1$.

We analogously define memory structures for a subset of players $P \subseteq \Pi$, which also defines *finite-memory strategy profiles.* Note that if $\bar{\sigma}_P$ is a finite-memory strategy profile, then each σ_i with $i \in P$ is finite-memory – the memory structure inducing that strategy is obtained by replacing ν by its restriction to $S \times V_i$. Conversely, if each σ_i with $i \in P$ is finite-memory, then the strategy profile $\bar{\sigma}_P$ is finite-memory – a collective memory structure can be obtained by constructing the product of individual memory structures.

▶ **Example 5.** Figure 2b depicts an example of memory structure, for player □, on the game of Figure 2a. The strategy induced can be presented as follows: when the vertex b is reached, player □ goes deterministically to the vertex t_1 if s has been visited an even number of times. Otherwise, he goes to c with probability $\frac{1}{3}$. Note that the output only depends on the history that was seen: player □ does not see whether player ○ deterministically chose to go to the vertex b, or did it as the outcome of a randomised choice.

Risk-sensitive equilibria, constrained existence problem. In multiplayer stochastic games, we wish to study generalisations of the classical Nash equilibria where the expectation is replaced by other risk measures. Such generalisations are called *risk-sensitive equilibria* [24]. When M is a risk measure and $\bar{\sigma}$ is a strategy profile, we write $M(\bar{\sigma})$ for $M^{\mathbb{P}_{\bar{\sigma}}}$.

▶ **Definition 6** (Risk-sensitive equilibrium). *Let $\mathcal{G}_{\restriction v_0}$ be a game, and let $\bar{M} = (M_i)_{i\in\Pi}$ be a tuple of risk measures. Let $\bar{\sigma}$ be a strategy profile in $\mathcal{G}_{\restriction v_0}$, let i be a player, and let σ'_i be a strategy for player i, called* deviation *of player i from $\bar{\sigma}$. The deviation σ'_i is* profitable *with regards to the risk measure M_i if we have $M_i(\bar{\sigma}_{-i}, \sigma'_i)[\mu_i] > M_i(\bar{\sigma})[\mu_i]$. The strategy profile $\bar{\sigma}$ is a $\bar{M}$*-risk-sensitive equilibrium*, or $\bar{M}$-RSE, if no player i has a profitable deviation from $\bar{\sigma}$ with regards to M_i.*

Nash equilibria in a game $\mathcal{G}_{\restriction v_0}$ can then be defined as $\bar{M}$-risk-sensitive equilibria, where $M_i = \mathbb{E}$ for each player i. The following problem is the main focus throughout our paper.

▶ **Question** (Constrained existence of risk-sensitive equilibria). *Given a game $\mathcal{G}_{\restriction v_0}$, a tuple of risk measures $\bar{M}$, and two payoff vectors $\bar{x}, \bar{y} \in \mathbb{Q}^\Pi$, does there exist a $\bar{M}$-RSE $\bar{\sigma}$ in $\mathcal{G}_{\restriction v_0}$ such that for each $i \in \Pi$, we have $x_i \leqslant M_i(\bar{\sigma})[\mu_i] \leqslant y_i$?*

We mainly consider the *entropic risk measure* and the *extreme risk measure.*

3 Entropic risk measure: intractable in multiplayer games

The entropic risk measure is defined using a *risk parameter*, i.e. a real value $\rho \in \mathbb{R}\backslash\{0\}$: large positive values indicate risk-averseness, large negative values risk-inclination.

▶ **Definition 7** (Entropic risk measure)**.** *Given a risk parameter ρ, the* entropic risk measure *is defined for every probability measure $\mathbb{P}$ and random variable X as:*

$$\mathbb{M}^{\mathbb{P}}_{\rho}[X] = -\frac{1}{\rho}\log_e\left(\mathbb{E}^{\mathbb{P}}\left[e^{-\rho X}\right]\right).$$

This definition is generalised by replacing Euler's constant with some base $\beta > 1$. The entropic risk measure with base β *is then defined by $\mathbb{M}^{\mathbb{P}}_{\beta\rho}[X] = -\frac{1}{\rho}\log_\beta\left(\mathbb{E}^{\mathbb{P}}\left[\beta^{-\rho X}\right]\right)$. The probability measure $\mathbb{P}$ is omitted when clear from the context.*

To see a visual representation of the entropic risk measure, see Figure 1 in the introduction.

▶ Remark 8.

- The generalisation to any base $\beta > 1$ follows only computational goals, since algebraic bases will be easier to handle. Baring such concerns, the definitions are equivalent, since for every β we have $\mathbb{M}_{\beta\rho} = \mathbb{M}_{\rho'}$, where $\rho' = \rho\log_e(\beta)$.
- Definition 7 implies that for $\rho = 0$, the function is not defined. However, it is known that for all $\mathbb{P}$, β and X, the quantity $\mathbb{M}_\rho[X]$ converges to $\mathbb{E}[X]$ when ρ tends to 0 (see e.g. [26]). Therefore, we henceforth assume that $\mathbb{M}_{\beta 0}[X] = \mathbb{E}[X]$ to make risk entropy defined for all finite risk parameters ρ.

When we are given a profile $\bar{\rho} = (\rho_i)_{i\in\Pi}$ of risk parameters, we write $\mathbb{M}_{\beta\bar{\rho}}[\mu]$ for the tuple $(\mathbb{M}_{\beta\rho_i}[\mu_i])_{i\in\Pi}$. Risk entropy defines a family of RSEs, namely the $(\mathbb{M}_{\beta\rho_i})_i$-RSEs, that we also call $(\beta,\bar{\rho})$*-entropic risk-sensitive equilibria*, or $(\beta,\bar{\rho})$-ERSEs. We now study the constrained existence problem of $(\beta,\bar{\rho})$-ERSEs. Unfortunately, it is undecidable in the general case.

▶ **Theorem 9.** *The constrained existence problem of $(\beta,\bar{\rho})$-ERSEs with $\bar{\rho} \in \mathbb{Q}^\Pi$ is undecidable, even for any fixed value of β, for $\bar{\rho} = (0)_i$, and with only nonnegative payoffs.*

Proof. The constrained existence problem of Nash equilibria is undecidable [31, Theorem 4.9]. Since Nash equilibria are ERSEs with $\rho_i = 0$ for each player i, our result follows. ◀

We therefore briefly study cases where the class of strategies considered is restricted.

▶ **Theorem 10.** *The constrained existence problem of $(\beta,\bar{\rho})$-ERSEs, in quantitative simple stochastic games, with rational risk parameters:*

1. *remains undecidable when players are restricted to pure strategies;*
2. *is decidable when players are restricted to stationary strategies, or positional strategies:*
   a. *in* 3EXPTIME *if $\beta = e$;*
   b. *in* PSPACE *if the base β is algebraic:*
      - *it is* NP*-hard when restricted to positional strategies, and*
      - *$\exists\mathbb{R}$-complete when restricted to stationary strategies.*

We end this section with this result: ERSEs are guaranteed to exist, if all rewards are nonnegative.

▶ **Theorem 11.** *Let $\mathcal{G}_{\restriction v_0}$ be a simple stochastic game with only nonnegative payoffs. Then, there exists a (pure) (β,ρ)-ERSE over $\mathcal{G}_{\restriction v_0}$.*

We conjecture that this result remains true when negative rewards are involved.

4 Extreme risk measure: limit of entropic risk measure

This section introduces a new risk measure that provides a tractable alternative to existing risk measures available in the literature. Let X be a random variable ranging over $\mathbb{R}$. The *pessimistic risk measure* of X is the highest value x such that X almost-surely takes a value above x. When X takes finitely many values, that corresponds to the least value that it takes with positive probability. In probability theory, that measure is sometimes referred to as *essential infimum*, written $\operatorname{ess\,inf}$. The definition of *optimistic risk measure* is symmetric.

▶ **Definition 12** (Optimistic, pessimistic risk measure). *The pessimistic risk measure of a random variable X is defined by $\mathbb{PM}[X] = \operatorname{ess\,inf}(X) = \sup\{x \in X \mid \mathbb{P}(X \geqslant x) = 1\}$. Analogously, the* optimistic risk measure *of X is $\mathbb{OM}[X] = \operatorname{ess\,sup}(X) = \inf\{x \in X \mid \mathbb{P}(X \leqslant x) = 1\}$.*

Note that the values of probabilities do not influence the optimistic or the pessimistic risk measures, which depend only on which events have zero or nonzero probability. Thus, those risk measures are well-suited to model players that do not care about probabilities because they need certainties, or simply because they do not know them – which is often the case in real-world stochastic processes.

Given a game $\mathcal{G}_{\restriction v_0}$, we can assign a risk measure for each player by defining a partition (P, O) of Π, where the set P represents the set of players that are *pessimists*, whose perceived payoffs are defined by the pessimistic risk measure, while O represents the *optimists*, who intend to maximise their optimistic risk measure. For convenience, we group both measures under the umbrella term *extreme risk measure (XR)*, and often assume that (P, O) is given; then, we write $\mathbb{X}_i$ for $\mathbb{PM}$ when $i \in P$, and for $\mathbb{OM}$ when $i \in O$. Since each player i is interested only in the risk measure of their own payoff, we also write $\mathbb{X}_i(\bar{\sigma})$ for the quantity $\mathbb{X}_i(\bar{\sigma})[\mu_i]$. We define *extreme risk-sensitive equilibria*, or XRSEs for short, as $(\mathbb{X}_i)_i$-RSEs.

Our definition, which considers only outcomes of positive probability, is slightly different from a situation where each player treats every other player or stochastic vertex as an adversary. Indeed, in Figure 1, such an adversarial treatment would mean that player ○ assumes that the probability 0 event of the play staying in the vertex s could also be realised, whereas in for extreme risk measure, such a probability 0 event is disregarded. The same would hold if the stochastic vertex was instead assigned to any player whose strategy randomises between staying and leaving that vertex.

We show that our definition of extreme risk measure corresponds to the limit cases of entropic risk measure. Observe that in Figure 1, when player ◇ follows a randomized strategy, player ○'s risk measure converges to -1 when ρ tends to $+\infty$, and to $+1$ when ρ tends to $-\infty$. We formally prove that our definition of extreme risk measure coincides with the limit case of the entropic risk measure.

▶ **Theorem 13.** *Let X be a random variable that ranges over $\mathbb{R}$, and let $\beta > 1$.*

- *The limit of the risk entropy of X when ρ tends to $+\infty$ exists and is equal to the pessimistic risk measure, that is, we have $\lim_{\rho \to +\infty} \mathbb{M}_{\beta\rho}[X] = \mathbb{PM}[X]$.*
- *Similarly, the limit of the risk entropy of X when ρ tends to $-\infty$ exists and is equal to the pessimistic risk measure, that is, we have $\lim_{\rho \to -\infty} \mathbb{M}_{\beta\rho}[X] = \mathbb{OM}[X]$.*

5 Constrained existence of extreme risk sensitive equilibria

We now study the computational complexity of the constrained existence problem of XRSEs. The main result of this section will show that, contrary to the same problem with ERSEs, it is a decidable fragment of the constrained existence of RSEs, as it is NP-complete. We will also study some subcases, showing that NP-completeness remains true when players are restricted to positional, stationary, or pure strategies. Finally, we will show that when all players are optimists, the problem becomes PTIME-complete.

5.1 Membership in NP

NP-membership is a consequence of this fact: when an XRSE exists, there also exists one with the same extreme risk measures that uses finite memory, polynomial in the size of the game. Let us first illustrate, with examples, how and why memory is required in such XRSEs.

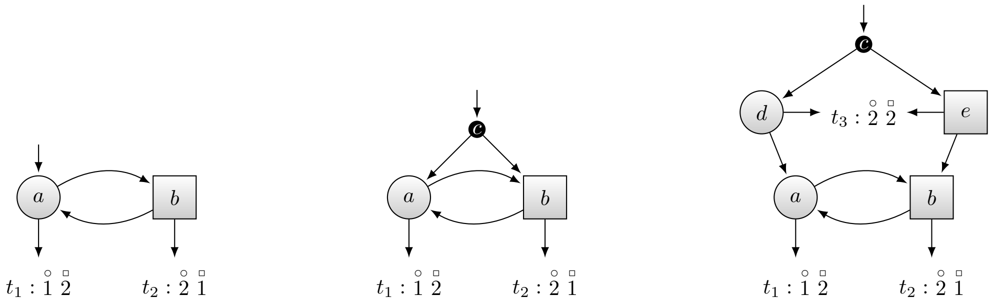
▶ **Example 14.** We consider the following constrained existence question and analyse the same question on three example graphs.

($*$) Is there an XRSE $\bar{\sigma}$ such that we have $\mathbb{X}_{\Box}(\bar{\sigma}) = \mathbb{X}_{\circ}(\bar{\sigma}) = 1$?

Game in Figure 3a. The answer to Question ($*$) in this game is *no.* Indeed, if there is such an XRSE $\bar{\sigma}$ in this game, then necessarily, following $\bar{\sigma}$, both terminal vertices t_1 and t_2 are reached with positive probability, and the probability of following the cycle ab forever is 0 (see Appendix E of full version). Intuitively, the problem here is that such a strategy profile would require a randomisation: for both players to get the payoff 1 with positive probability, there must be a random event that decides which of them will actually have that payoff. In our world, such a situation would be solved by tossing a coin. In the game, it means that one of the players proceeds to a randomised action. But that player, say □ again, can deviate from such a strategy and refuse to leave the cycle. Such refusals form *undetectable* deviations from the strategy, since player ○ only sees that player □ went from b to a, which can be interpreted as a possible outcome of the expected randomisation. In other words, the player that tosses a coin is the only one that sees the result, and can therefore lie about it. This phenomenon can be avoided only if such randomised choices are made by other "impartial" players or by stochastic vertices.

Game in Figure 3b. Consider now a slight modification, as shown in Figure 3b. There, the first player that plays is determined at random by the edge that is taken from an initial stochastic vertex. The answer to Question ($*$) for this game is *yes.* Indeed, the pure strategy profile defined by a strategy of player ○ that maps the history ca to the terminal t_1 and the strategy of player □ that maps the history cb to terminal t_2, and both strategies maps any other history to vertex b, is an XRSE. Indeed, if any player deviates and refuses to leave the cycle, the other one will immediately refuse too, making it impossible to gain any benefit from such a deviation. Compared to the previous example, the existence of the stochastic vertex c provides the players with a trustable common coin, that they can use to decide which of them will get payoff 1 and which will get payoff 2.

Game in Figure 3c. Finally, consider the game depicted in Figure 3c. Here, along every play, one player is given the opportunity of getting payoff 2, by going to the terminal vertex t_3. But the answer to Question ($*$) remains *yes.* Indeed, the pure strategy profile $\sigma_{\circ}$ where from vertex d, player ○ goes from vertex d to a and then to b, from which player □ leaves to t_2, and symmetrically, from vertex e, player □ goes to a through b from which player ○ goes to t_1 is an XRSE. For instance, if player ○ deviates and goes from d to t_3, then there is still the play $cebat_1$ that is generated with positive probability and from which player ○ cannot deviate without triggering a punishing strategy that would make such a deviation non-profitable.


a
b
$t_1 : \overset{\circ}{1}\ \overset{\square}{2}$
$t_2 : \overset{\circ}{2}\ \overset{\square}{1}$
a
b
$t_1 : \overset{\circ}{1}\ \overset{\square}{2}$
$t_2 : \overset{\circ}{2}\ \overset{\square}{1}$
d
e
$t_3 : \overset{\circ}{2}\ \overset{\square}{2}$
a
b
$t_1 : \overset{\circ}{1}\ \overset{\square}{2}$
$t_2 : \overset{\circ}{2}\ \overset{\square}{1}$

(a) Two pessimists.

(b) Two pessimists with a common coin.

(c) Two pessimists with a common coin and some temptation.

Figure 3 Some games involving two pessimistic players.

We see in this last example that to build an XRSE that generates a given tuple of risk measure values, we need one play for each player that *anchors* that player – i.e., a play in which they get their risk measure as a payoff (the lowest/highest value they obtain with positive probability), and from which they cannot deviate in a profitable way. Then, randomisation (or stochastic vertices) is required to *split* plays into two or more potential future plays where each anchors different subset of players, and memory is required to remember the subset of players that are being anchored – or whether a player has deviated from the strategy and must be punished.

In line with this intuition, the following theorem bounds the amount of memory required by an XRSE. We further show that there is a succinct encoding of this bounded memory that has size that is polynomial in the number of players and vertices in the game.

▶ **Theorem 15.** *Let $\mathcal{G}_{\restriction v_0}$ be a game with n vertices and p players, let (P, O) be a partition of the set Π, and let $\bar{\sigma}$ be an XRSE in $\mathcal{G}_{\restriction v_0}$. Then, there exists a finite-memory XRSE $\bar{\sigma}^\star$ with at most $3np - 2n + p + 1$ many memory states, such that $\mathbb{X}(\bar{\sigma}^\star) = \mathbb{X}(\bar{\sigma})$. Furthermore, if $\bar{\sigma}$ is pure, then there is such a strategy profile $\bar{\sigma}^\star$ that is pure.*

Proof sketch. We prove this theorem by formalising the idea of *anchoring plays.* To do so, we define a labelling function Λ, that maps each history h compatible with $\bar{\sigma}$ to the set of players that is, after the history h, currently being *anchored.* In Lemma 16, we show the existence of such a labelling, with some properties. In the sequel, we write $\bar{z}$ to denote the risk measure of each player in the strategy profile $\bar{\sigma}$, that is, the tuple $\bar{z} = (z_i)_i = \mathbb{X}(\bar{\sigma})$.

▶ **Lemma 16** (The labelling Λ)**.** *There exists a labelling Λ that maps each history $h \in \mathsf{Hist}\mathcal{G}_{\restriction v_0}$ compatible with $\bar{\sigma}$ to a subset of players, that is, $\Lambda(h) \subseteq \Pi$, such that for each such h, if we write $\{v_1, \dots, v_k\} = \mathsf{Supp}(\bar{\sigma}(h))$, the labelling Λ satisfies the following properties.*

1. *If the vertex $\mathsf{last}(h)$ is stochastic, or belongs to some player $i \notin \Lambda(h)$, then the sets $\Lambda(hv_1), \dots, \Lambda(hv_k)$ form a partition of $\Lambda(h)$.*
2. *If the vertex $\mathsf{last}(h)$ belongs to some player $i \in \Lambda(h)$, then the sets $\Lambda(hv_1)\backslash\{i\}$, …, and $\Lambda(hv_k)\backslash\{i\}$ form a partition of $\Lambda(h)\backslash\{i\}$, and i belongs to all sets $\Lambda(hv_1), \dots, \Lambda(hv_k)$.*
3. *For each optimistic player $i \in \Lambda(h)$, we have $\mathbb{X}_i(\bar{\sigma}_{\restriction h}) = z_i$.*
4. *For each pessimistic $i \in \Lambda(h)$, for all strategies τ_i of player i, we have $\mathbb{X}_i(\bar{\sigma}_{-i\restriction h}, \tau_i) \leqslant z_i$.*
5. *If there is a successor v_ℓ such that $\Lambda(hv_\ell) = \Lambda(h)$, then all other successors $v_{\ell'}$ are such that $\mathbb{X}_i(\bar{\sigma}_{\restriction hv_{\ell'}}) < z_i$ for each optimist $i \in \Lambda(h)$, and there exists τ_i with $\mathbb{X}_i(\bar{\sigma}_{-i\restriction hv_{\ell'}}, \tau_i) > z_i$ for each pessimist $i \in \Lambda(h)$.*

Since Λ labels vertices with sets of players, potentially there are exponentially many subsets of players A that are in the range of Λ. However, for a Λ that satisfies Items 1–5 of Lemma 16, we show that there are at most $3p-2$ subsets A such that $\lambda(h) = A$ for some history h by an inductive counting argument. We use those subsets to create $(3p-2)n$ memory states, to remember which players are being anchored, and what was the last vertex visited (to detect deviations). We add one punishing state for each player, used when a deviation by that player is detected, adding up to the desired number. We construct an XRSE $\bar{\sigma}^\star$ from Λ that uses only those memory states. ◀

Using Theorem 15, we can show the following lemma.

▶ **Lemma 17.** *The constrained existence problem of XRSEs is in* NP*. The same problem when players are restricted to pure strategies is still in* NP*.*

Proof. Let $\mathcal{G}_{\restriction v_0}$ be a game. Let (P, O) be a partition of Π, and let $\bar{x}$ and $\bar{y}$ be threshold vectors. By Theorem 15, if there exists a (pure) XRSE with $\bar{x} \leqslant \mathbb{X}(\bar{\sigma}) \leqslant \bar{y}$, then there exists one with at most $3np-2n+p+1$ memory states, where p is the number of players and n is the number of vertices. Such a strategy profile can be guessed in polynomial time. We now show that, once such a finite-memory strategy profile $\bar{\sigma}$ is guessed, one can check in polynomial time whether it is an XRSE, and satisfies the constraint $\bar{x} \leqslant \mathbb{X}(\bar{\sigma}) \leqslant \bar{y}$.

- First, given $\bar{\sigma}$, for each player i, the quantity $\mathbb{X}_i(\bar{\sigma})$ can be computed in polynomial time, since it reduces to computing player i's risk measure in the Markov chain induced by $\bar{\sigma}$ (which has polynomial size).
- Second, checking that $\bar{x} \leqslant \mathbb{X}(\bar{\sigma}) \leqslant \bar{y}$ can be done in polynomial time.
- Third, for each player i, we check that player i has no profitable deviation. This can also be done in polynomial time by computing the best risk measure player i can get in the MDP induced by $\bar{\sigma}_{-i}$ (which has polynomial size). ◀

5.2 Membership in NP with restrictions on strategies

We have shown NP-membership of the constrained existence problem of XRSEs in the general case. We now consider variants where the space of a strategies is restricted to stationary, positional or pure strategies. We show that the problem is in NP for each of these cases.

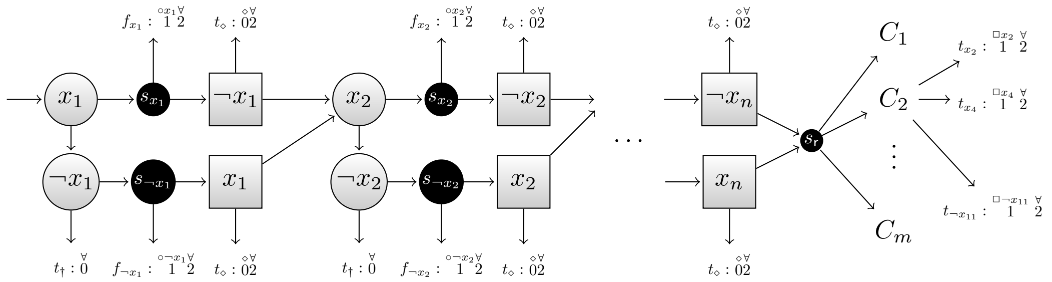
▶ **Lemma 18.** *The constrained existence problem, when all the players are restricted to positional, stationary, or pure strategies, is in* NP*.*

Proof. We show that we can still guess a strategy profile, and verify in polynomial time whether it is indeed an XRSE. For the positional and stationary cases, guessing a strategy profile is straightforward, since such a strategy profile can always be represented using polynomially many bits. We can then verify that a given strategy profile $\bar{\sigma}$ satisfies the constraints and also is an XRSE in polynomial time. For pure strategies, memory may be required. But we showed, alongside Theorem 15, that if there is a pure XRSE satisfying the constraints, there is one with polynomial memory: our result follows. ◀

5.3 NP-hardness

We now prove hardness, for the general setting as well as when strategies are restricted.

▶ **Lemma 19.** *The constrained existence problem of XRSEs is* NP*-hard, even when all players are pessimists and all rewards are nonnegative. It remains* NP*-hard when the strategies are restricted to stationary, pure, or positional ones.*



$f_{x_1} : \overset{\circ x_1}{1}\overset{\forall}{2}$
$t_\diamond : \overset{\diamond}{0}\overset{\forall}{2}$
$f_{x_2} : \overset{\circ x_2}{1}\overset{\forall}{2}$
$t_\diamond : \overset{\diamond}{0}\overset{\forall}{2}$
$t_\diamond : \overset{\diamond}{0}\overset{\forall}{2}$
C_1
$t_{x_2} : \overset{\square x_2}{1}\ \overset{\forall}{2}$
x_1
s_{x_1}
$\neg x_1$
x_2
s_{x_2}
$\neg x_2$
$\neg x_n$
s_r
C_2
$t_{x_4} : \overset{\square x_4}{1}\ \overset{\forall}{2}$
$\cdots$
$\neg x_1$
$s_{\neg x_1}$
x_1
$\neg x_2$
$s_{\neg x_2}$
x_2
x_n
$\vdots$
$t_{\neg x_{11}} : \overset{\square \neg x_{11}}{1}\ \overset{\forall}{2}$
C_m
$t_\dagger : \overset{\forall}{0}$
$f_{\neg x_1} : \overset{\circ \neg x_1}{1}\overset{\forall}{2}$
$t_\diamond : \overset{\diamond}{0}\overset{\forall}{2}$
$t_\dagger : \overset{\forall}{0}$
$f_{\neg x_2} : \overset{\circ \neg x_2}{1}\overset{\forall}{2}$
$t_\diamond : \overset{\diamond}{0}\overset{\forall}{2}$
$t_\diamond : \overset{\diamond}{0}\overset{\forall}{2}$

Figure 4 Construction of a game $\mathcal{G}_\Phi$ from a 3SAT formula Φ.

Proof sketch. We reduce the 3SAT problem to our problem. From a given formula Φ, we construct the game depicted by Figure 4, where all players are pessimistic, and the symbol $\forall$ is used to mean *"every other player"*. For each literal $\ell \in \{x_i, \neg x_i \mid i \in \{1, \dots, n\}\}$, we define two players, namely player $\circ\ell$ and player $\square\ell$. We also define a player C_j for each clause C_j. Each of those players controls one vertex, labelled by their name. We show in the complete proof that this game contains a (positional) XRSE where a witness player, player $\diamond$, gets risk measure 2 if and only if Φ is satisfiable.

Intuitively, the game consists of a sequence of gadgets where a value is given to each variable x_i, depending on whether player $\circ x_i$ takes the edge to the vertex s_{x_i} (variable set to true) or to the vertex $\neg x_i$ (variable set to false). That player cannot randomise between those two edges, because she does not get the same risk measure on both sides, and therefore would have a profitable undetectable deviation. When it is decided that the literal ℓ is true, player $\square\overline{\ell}$ is given the possibility to deviate and get the payoff 2. In the final gadget, each clause player C_j must choose a literal ℓ of the clause C_j that she claims to be true, under the valuation that has been defined with the previous gadgets. Then, player $\square\ell$ gets risk measure 1: that player will thus have a profitable deviation if and only if he was given the possibility to deviate, i.e., if the literal ℓ was actually set to false. ◀

This lemma, along with Lemma 17 and Lemma 18, proves the following theorem. Observe that our construction did not use any negative rewards, and hence the hardness results hold already for nonnegative rewards, while our membership result works with any combination of positive and negative rewards.

▶ **Theorem 20.** *The constrained existence problems for XRSEs, for pure XRSEs, for stationary XRSEs, and for positional XRSEs, are all* NP*-complete. The lower bound holds even when all players are pessimistic and all rewards are non-negative.*

We will now turn to a last subcase, where our problem turns out to be decidable in deterministic polynomial time.

5.4 Things get easier when everyone is optimistic

Our NP-hardness results required only pessimistic players. We now show that optimists are easier to deal with. The constrained existence problem of XRSEs becomes PTIME-complete when the perceived reward of each player is defined by the risk measure $\mathbb{OM}$.

We first show an upper bound by giving a polynomial-time algorithm. The intuition is that an optimistic player has a profitable deviation as soon as a vertex is reached, with positive probability, from which that player can get a payoff higher than their risk measure. This is in contrast to pessimists, who have a profitable deviation only when they can avoid getting their risk measure as a payoff in all plays compatible with the strategy profile. Our algorithm iteratively identifies vertices that must never be reached, and then removing from the game the edges that must consequently never be taken.

▶ **Lemma 21.** *If all players are optimists, then the constrained existence problem of XRSEs is in* PTIME*, and there is an algorithm that decides it in time $\mathcal{O}(pm^2)$, where m is the number of edges in $\mathcal{G}$ and p the number of players. Moreover, the algorithm can be modified to output an XRSE that satisfies the constraints, if one exists, in time $\mathcal{O}(pm^2 + m^3)$. If all the upper thresholds y_i are nonnegative, it takes time $\mathcal{O}(pm^2)$.*

Proof Sketch. We want to decide whether there exists an XRSE $\bar{\sigma}$ satisfying the constraint $\bar{x} \leqslant \mathbb{X}(\bar{\sigma}) \leqslant \bar{y}$. The algorithm deals with two cases, that we call *cycle-friendly* and *cycle-averse* cases, separately. In the cycle-friendly case, we have $y_i \geqslant 0$ for all players i. Then, an XRSE could have positive probability of reaching no terminal vertex. However that is impossible in the cycle-averse case, where there is a player i such that $y_i < 0$. In this proof sketch, we describe only the algorithm in the cycle-friendly case.

The algorithm constructs a decreasing sequence of sets of edges $E_0, E_1, \dots$ until it reaches a fixed point. For each set E_k, it considers the strategy profile $\bar{\sigma}^{E_k}$, defined as follows: from each non-stochastic vertex v, when v is seen for the first time, it randomises uniformly between all edges $vw \in E_k$. Later, if v is visited again, it always repeats the same choice. If some player i deviates and takes an edge that they are not supposed to take, then all the players switch to a positional strategy profile designed to minimise their risk measure. Such a strategy profile is finite-memory, but requires $2^{|V|}|V| + p$ memory states to be represented as a memory structure: we therefore use the set E_k as a succinct representation.

At each iteration k, the algorithm identifies new sets of vertices $V^k_{\frownie}$ that must be avoided. This includes the terminals that give some player i a payoff that is larger than y_i, or vertices from which player i can deviate and obtain a higher value than the value offered by the strategy profile $\mathbb{X}_i(\bar{\sigma}^{E_k})$. If it is not possible to avoid reaching the set $V^k_{\frownie}$, the answer No is returned. Otherwise, the set E_{k+1} is defined from E_k by removing edges that ensure that $V^k_{\frownie}$ is not reached with positive probability. The algorithm stops when there are no more edges to remove and answers Yes and if we have $\mathbb{X}_i(\bar{\sigma}^{E_k}) \geqslant x_i$ for each i, and No otherwise.

Each iteration requires time $\mathcal{O}(mp)$ to identify and remove edges. Since there are $\mathcal{O}(m)$ many edges, the algorithm terminates in time $\mathcal{O}(pm^2)$. ◀

Finally, we show that the problem is PTIME-hard, even when there are only two players.

▶ **Lemma 22.** *The constrained existence problem of XRSEs with optimistic players is* PTIME*-hard even with only two players.*

Proof sketch. We give a log-space reduction from the problem of deciding two-player zero-sum reachability games, which is known to be PTIME-complete [18, Proposition 6]. ◀

We can now conclude this section with the following theorem.

▶ **Theorem 23.** *The constrained existence problem of XRSE is* PTIME*-complete when all players are optimists, that is, when $P = \varnothing$.*

6 The existence of extreme risk sensitive equilibria

This section finally answers the classical question about equilibria: are they guaranteed to exist? We show that (stationary) XRSEs exist when all rewards are nonnegative. Although the result is reminiscent of the same result we proved for ERSEs (Theorem 11), it requires a different, and constructive proof that we discuss below.

We motivate the constructive algorithm with an example. Consider the game depicted in Figure 3a that involves two pessimists: player ○ and player □. Both players want to leave the cycle, but each of them would prefer that the other player leaves. If we first consider the strategy profile that always randomises between all the available edges, then both terminal vertices are reached with positive probability, and it is almost sure that one of them is reached: both players get therefore risk measure 1. Then, player □ (and symmetrically player ○) has a profitable deviation by refusing to leave the cycle, and always going back to the vertex a. Note that player ○ cannot detect such a deviation, since she does not have access to the internal coins tossed by player □. Then, we remove the edge bt_2 (or at_1). This results in a set of edges where player □ gets the payoff 2, and player ○ cannot get more than 1, ensuring that the new strategy profile that we obtain is a (stationary) XRSE.

▶ **Theorem 24.** *Let $\mathcal{G}_{\restriction v_0}$ be a game with only non-negative rewards, and let (P, O) be a partition of Π. Then, there exists a stationary XRSE in $\mathcal{G}_{\restriction v_0}$. Moreover, there exists an algorithm that, given such a game, outputs the representation of such an XRSE in time $\mathcal{O}(m^2 p)$, where m is the number of edges, and p the number of* pessimistic *players.*

Proof sketch. Our algorithm constructs a decreasing sequence $E = E_0, E_1, \ldots$ of sets of edges, and considers, for each k, the stationary strategy profile that randomises between all the outgoing edges in E_k from all vertices. If this strategy profile is not an XRSE, there is a player i who has a profitable deviation. We carefully identify edges used by player i, and remove them. This process always terminates, and the set obtained defines an XRSE. ◀

Like in the case of ERSEs, we conjecture that existence, and even existence of a stationary XRSE, remain true in the general case. We remark that even for Nash equilibria, existence of an NE in such simple stochastic game is known only if all the rewards are nonnegative.

7 Discussion

Our definition of extreme risk measure opens up several promising directions for future research. One immediate extension of our work would be to study games with more sophisticated objectives, such as mean payoff or discounted sum. Another extension is to study the concurrent version of such games, where players choose actions concurrently rather than in a turn-based setting. Finally, our definition of risk-sensitive equilibria is modelled after Nash equilibria and suffers from several of their limitations. Like Nash equilibria, RSEs allow irrational behaviours when one player deviates and must be punished, as in our game of Figure 3c. Exploring alternative definitions of RSE, modelled after other equilibria concepts more suited for games on graphs [25, Section 7.1], such as subgame-perfect equilibria, could provide a relevant framework for player decision-making.

References

1 Agarwal Alekh, Nan Jiang, Sham M. Kakade, and Sun Wen. *Reinforcement Learning: Theory and Algorithms*. https://rltheorybook.github.io/, 2021.

2 M. Auer. *Hands-On Value-at-Risk and Expected Shortfall: A Practical Primer*. Management for Professionals. Springer International Publishing, 2018. URL: https://books.google.at/books?id=4EFKDwAAQBAJ.

3 Christel Baier, Krishnendu Chatterjee, Tobias Meggendorfer, and Jakob Piribauer. Entropic risk for turn-based stochastic games. *Information and Computation*, 301:105214, 2024. doi:10.1016/j.ic.2024.105214.

4 Hans Wolfgang Brachinger. From variance to value at risk: A unified perspective on standardized risk measures. In Wolfgang Gaul and Hermann Locarek-Junge, editors, *Classification in the Information Age*, pages 91–99, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg.

5 Richard Bradley. *Decision Theory: A Formal Philosophical Introduction*, pages 611–655. Springer International Publishing, 2018. doi:10.1007/978-3-319-77434-3_34.

6 Véronique Bruyère, Emmanuel Filiot, Mickael Randour, and Jean-François Raskin. Meet your expectations with guarantees: Beyond worst-case synthesis in quantitative games. *Information and Computation*, 254:259–295, 2017. SR 2014. doi:10.1016/j.ic.2016.10.011.

7 Nicole Bäuerle and Ulrich Rieder. More risk-sensitive markov decision processes. *Mathematics of Operations Research*, 39(1):105–120, 2014. doi:10.1287/MOOR.2013.0601.

8 Luca de Alfaro, Thomas A. Henzinger, and Ranjit Jhala. Compositional methods for probabilistic systems. In *CONCUR 2001 — Concurrency Theory*, pages 351–365, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg. doi:10.1007/3-540-44685-0_24.

9 Jerzy Filar and Lodewijk Kallenberg. Variance-penalized markov decision process. *Mathematics of Operations Research*, 14, February 1989. doi:10.1287/moor.14.1.147.

10 Hans Föllmer and Alexander Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–447, October 2002. doi:10.1007/s007800200072.

11 Vojtěch Forejt, Marta Kwiatkowska, Gethin Norman, and David Parker. *Automated Verification Techniques for Probabilistic Systems*, pages 53–113. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. doi:10.1007/978-3-642-21455-4_3.

12 Hans Föllmer and Alexander Schied. *Stochastic Finance: An Introduction in Discrete Time*. Walter de Gruyter, Berlin, Boston, 4 edition, 2016.

13 Jorge Gallego-Hernández and Alessio Mansutti. On the existential theory of the reals enriched with integer powers of a computable number. In Olaf Beyersdorff, Michal Pilipczuk, Elaine Pimentel, and Kim Thang Nguyen, editors, *42nd International Symposium on Theoretical Aspects of Computer Science, STACS 2025, March 4-7, 2025, Jena, Germany*, volume 327 of *LIPIcs*, pages 37:1–37:18. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi:10.4230/LIPICS.STACS.2025.37.

14 Christoph Haase and Stefan Kiefer. The odds of staying on budget. In *Automata, Languages, and Programming*, pages 234–246, Berlin, Heidelberg, 2015. Springer Berlin Heidelberg. doi:10.1007/978-3-662-47666-6_19.

15 Ronald A. Howard and James E. Matheson. Risk-sensitive markov decision processes. *Management Science*, 18(7):356–369, 1972.

16 Ronald A. Howard and James E. Matheson. Risk-sensitive markov decision processes. *Management Science*, 18(7):356–369, 1972. URL: http://www.jstor.org/stable/2629352.

17 Leonid Hurwicz. Optimality criteria for decision-making under ignorance. Technical report, Cowles Commission for Research in Economics, 1951. URL: https://cowles.yale.edu/sites/default/files/2024-03/s-0355.pdf.

18 Neil Immerman. Number of quantifiers is better than number of tape cells. *J. Comput. Syst. Sci.*, 22:384–406, 1981. doi:10.1016/0022-0000(81)90039-8.

19 Jan Křetínský and Tobias Meggendorfer. Conditional value-at-risk for reachability and mean payoff in markov decision processes. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '18, pages 609–618. Association for Computing Machinery, 2018. doi:10.1145/3209108.3209176.

20 Claude Lefèvre. Optimal control of a birth and death epidemic process. *Operations Research*, 29(5):971–982, 1981. doi:10.1287/OPRE.29.5.971.

21 Shie Mannor and John N. Tsitsiklis. Mean-variance optimization in markov decision processes. In *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, pages 177–184. Omnipress, 2011. URL: https://icml.cc/2011/papers/156_icmlpaper.pdf.

22 Tobias Meggendorfer. Risk-aware stochastic shortest path. *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 9858–9867, 2022. doi:10.1609/AAAI.V36I9.21222.

23 John F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950. doi:10.1073/pnas.36.1.48.

24 Andrzej S. Nowak. *Notes on Risk-Sensitive Nash Equilibria*, pages 95–109. Birkhäuser Boston, Boston, MA, 2005. doi:10.1007/0-8176-4429-6_5.

25 Martin J Osborne. An introduction to game theory. *Oxford University Press google schola*, 2:672–713, 2004.

26 Bernardo K. Pagnoncelli, Oscar Dowson, and David P. Morton. Multistage stochastic programs with the entropic risk measure, 2020. URL: https://optimization-online.org/2020/08/7984/.

27 Jakob Piribauer, Ocan Sankur, and Christel Baier. The variance-penalized stochastic shortest path problem. In *49th International Colloquium on Automata, Languages, and Programming, ICALP*, volume 229 of *LIPIcs*, pages 129:1–129:19. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPICS.ICALP.2022.129.

28 Mickael Randour, Jean-François Raskin, and Ocan Sankur. Variations on the stochastic shortest path problem. In *Verification, Model Checking, and Abstract Interpretation*, pages 1–18. Springer, 2015. doi:10.1007/978-3-662-46081-8_1.

29 Dawei Shi, Robert J Elliott, and Tongwen Chen. On finite-state stochastic modeling and secure estimation of cyber-physical systems. *IEEE Transactions on Automatic Control*, 62(1):65–80, 2016. doi:10.1109/TAC.2016.2541919.

30 Florent Teichteil-Königsbuch, Ugur Kuter, and Guillaume Infantes. Incremental plan aggregation for generating policies in MDPs. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1 - Volume 1*, AAMAS '10, pages 1231–1238. International Foundation for Autonomous Agents and Multiagent Systems, 2010. URL: https://dl.acm.org/citation.cfm?id=1838366.

31 Michael Ummels and Dominik Wojtczak. The Complexity of Nash Equilibria in Stochastic Multiplayer Games. *Logical Methods in Computer Science*, Volume 7, Issue 3, September 2011. doi:10.2168/LMCS-7(3:20)2011.

32 Abraham Wald. *Statistical Decision Functions.* John Wiley & Sons, New York, 1950. Classic formulation of the minimax (Wald) criterion.