Sequential Decision Making in Stochastic Games with Incomplete Preferences over Temporal Objectives

Abhishek Ninad Kulkarni[1], **Jie Fu**[2], **Ufuk Topcu**[1]

[1]University of Texas at Austin, Austin, TX, USA.
[2]University of Florida, Gainesville, FL, USA.
abhishek.kulkarni@austin.utexas.edu, fujie@ufl.edu, utopcu@utexas.edu

Abstract

Ensuring that AI systems make strategic decisions aligned with the specified preferences in adversarial sequential interactions is a critical challenge for developing trustworthy AI systems, especially when the environment is stochastic and players' incomplete preferences leave some outcomes unranked. We study the problem of synthesizing preference-satisfying strategies in two-player stochastic games on graphs where players have opposite (possibly incomplete) preferences over a set of temporal goals. We represent these goals using linear temporal logic over finite traces (LTLf), which enables modeling the nuances of human preferences where temporal goals need not be mutually exclusive and comparison between some goals may be unspecified. We introduce a solution concept of non-dominated almost-sure winning, which guarantees to achieve a most preferred outcome aligned with specified preferences while maintaining robustness against the adversarial behaviors of the opponent. Our results show that strategy profiles based on this concept are Nash equilibria in the game where players are risk-averse, thus providing a practical framework for evaluating and ensuring stable, preference-aligned outcomes in the game. Using a drone delivery example, we demonstrate that our contributions offer valuable insights not only for synthesizing rational behavior under incomplete preferences but also for designing games that motivate the desired behavior from the players in adversarial conditions.

Introduction

Sequential decision-making in adversarial environments is an important problem when designing trustworthy AI systems, especially when the environment is stochastic and preferences are incomplete (Wing 2021; Dalrymple et al. 2024). This approach enables AI systems to make strategic decisions over time and adapt to changing circumstances to achieve the best possible outcome while remaining robust to the opponent's behavior.

We study sequential decision-making in two-player stochastic games on graphs, with the goal of synthesizing strategies for each player that align with their individual preferences, where these preferences are adversarial and potentially incomplete. A game on graph (Grädel, Thomas, and Wilke 2003) is a widely studied model in computer science for verification (Pnueli and Zuck 1993; Baier and Katoen 2008), synthesis (Chatterjee and Henzinger 2012) and testing (Blass et al. 2006) of reactive systems.

We specify player preferences in the game as a preorder on a set of temporal goals defined using linear temporal logic over finite traces (LTL_f) (De Giacomo and Vardi 2013). This representation enables us to express preferences formally using an English-like language (Finucane, Jing, and Kress-Gazit 2010) and possibly be extracted from human language (Liu et al. 2022; Brunello, Montanari, and Reynolds 2019).

While much of the existing literature on games with preferences has studied normal-form games (Bade 2005; Bosi and Herden 2012; Özgür Evren and Ok 2011; Kokkala et al. 2019; Sasaki 2019), there is a growing interest in studying sequential decision-making within these games. Recently, games on graphs have been studied under lexicographic preferences (Chatterjee et al. 2023). To the best of our knowledge, the problem has not been studied for the class of incomplete preferences, which subsumes both complete and lexicographic preferences.

The synthesis problem presents two key challenges. First, human preferences over temporal goals are often combinative (Hansson and Grüne-Yanoff 2022), meaning the alternatives over which the preferences are specified are not mutually exclusive. For example, a cleaning robot may have a preference for "cleaning the living room" over "cleaning the bedroom," but if the battery allows, the robot could clean both rooms, satisfying both goals. Decision-making with combinative preferences is challenging because it requires simultaneously evaluating the possibility of satisfying various subsets of alternatives. However, existing literature on games on graphs with preferences (Chatterjee et al. 2023) frequently assumes that alternatives are exclusive, leaving the strategic planning of combinative preferences over temporal goals in stochastic games largely unexplored.

Second, human preferences over temporal goals are often incomplete because it is difficult to specify comparisons between all possible subsets of goals (Barbera and Pattanaik 1984; Dalrymple et al. 2024). Synthesizing strategies in the presence of such incomparability is challenging because the utility-based approaches for rational decision-making are inapplicable (Sen 1997). Although recent research has studied sequential decision-making in Markov decision process (MDP) with incomplete preferences (Li et al. 2020; Kulkarni and Fu 2022; Rahmani, Kulkarni, and Fu 2023), this problem remains underexplored in the context of stochastic games on graphs involving two or more non-cooperative players.

Contributions. This paper makes fundamental contributions to game theory and formal methods by introducing a novel automata-theoretic approach for strategy synthesis in two-player stochastic games with adversarial preferences. Our method provides formal guarantees in line with the goals of trustworthy AI (Tegmark and Omohundro 2023) and formally verified AI (Seshia, Sadigh, and Sastry 2022), while shifting the emphasis from verification to the synthesis of robust and correct-by-construction strategies. Our key contributions are as follows.

1. **Solution Concept.** We introduce a solution concept called *non-dominated almost-sure winning (ND-ASWin)*, designed to achieve the most-preferred outcome for a player while remaining robust to various opponent behaviors. This concept builds on the established notion of almost-sure winning in stochastic games with a single temporal goal (De Alfaro, Henzinger, and Kupferman 2007), extending it to games with incomplete preferences.
2. **Scalar Metric.** Since traditional models of incomplete preferences do not admit a utility representation (Sen 1997), multi-utility (vector-based) representations (Ok 2002) are employed to study these preferences. However, in our context, using this representation requires solving multi-objective stochastic games, which is computationally hard (Chen et al. 2013b). To this end, we propose a scalar metric called *rank* based on the undominance principle (Sen 1997) such that any outcome with a lower rank is no worse than that with a higher rank. Although rank does not capture the full complexity of incomplete preferences, we show that it is sufficient to synthesize ND-ASWin strategies.
3. **Nash equilibrium.** We show that any pair of ND-ASWin strategies of the players constitutes a Nash equilibrium. This result shows how a weaker solution concept can characterize Nash equilibrium in games with incomplete preferences.

Using a drone delivery scenario, we demonstrate that our results are particularly useful not only in computing strategies aligned with preference specification but also in designing games that motivate the desired behavior from the players given incomplete preferences.

Preliminaries

Notations. The set of all finite (resp., infinite) words over a finite alphabet Σ is denoted Σ^* (resp., Σ^ω). The empty string is denoted as ϵ and the set of non-empty strings is denoted by Σ^+. We denote the set of all probability distributions over a finite set X by $\mathcal{D}(X)$. Given a distribution $\mathbf{d} \in \mathcal{D}(X)$, the probability of an outcome $x \in X$ is denoted $\mathbf{d}(x)$.

Given a countable set U, a preference relation $\succeq$ on U is *preorder* on U. An element $u \in U$ is called *maximal* if there is no $v \in U$ such that $u \succeq v$, and it is called *minimal* if there is no $v \in U$ such that $v \succeq u$. The sets of all maximal and minimal elements in U are denoted by $\mathsf{Max}(U, \succeq)$ and $\mathsf{Min}(U, \succeq)$, respectively.

Game Model

Definition 1. A stochastic two-player concurrent game on graph is a tuple, $G = \langle S, A, T, s_0, AP, L\rangle$, where S is a set of states. $A = A_1 \times A_2$ is a set of actions, where A_1, A_2 represents the set of actions of P1 and P2, respectively. $T : S \times A \to \mathcal{D}(S)$ is a probabilistic transition function. Given any two states $s, s' \in S$ and any action $a \in A$, $T(s, a, s')$ denotes the probability that the game transitions from state s to s' when action a is chosen at s. $s_0 \in S$ is an initial state. AP is a set of atomic propositions. $L : S \to 2^{AP}$ is a labeling function that maps every state $s \in S$ to the set of atomic propositions $L(s) \subseteq AP$ that hold true at s.

A *path* in a game G is a sequence of states $\rho = s_0 s_1 s_2 \ldots$ such that, for every $i \geq 0$, there exists an action $a_i \in A$ such that $T(s_i, a_i, s_{i+1}) > 0$. The path is said to be *finite* if it terminates after a finite number of steps, otherwise it is *infinite*. The last state of a finite path ρ is denoted by $\mathsf{Last}(\rho)$. The set of all finite paths in G is denoted by $\mathsf{Paths}(G)$ and that of infinite paths is denoted by $\mathsf{Paths}_\infty(G)$. Every finite path ρ induces a finite word $L(\rho) = L(s_0)L(s_1)\ldots L(s_k) \in (2^{AP})^*$ called the *trace* of ρ.

A strategy in G is a function $\pi : S^+ \to \mathcal{D}(A)$ that maps every finite path in $\mathsf{Paths}(G)$ to a probability distribution over the action set A. A strategy π is called *memoryless* if for any two paths $\rho s, \rho' s \in \mathsf{Paths}(G)$, we have $\pi(\rho s) = \pi(\rho' s)$. Otherwise, π is called a *forgetful* strategy (Grädel, Thomas, and Wilke 2003). A strategy π is called *deterministic* if, for any path $\rho \in \mathsf{Paths}(G)$, $\pi(\rho)$ is a Dirac delta distribution. Otherwise, π is said to be a *randomized* strategy.

The temporal goals considered in this paper are interpreted over finite traces. Therefore, throughout this paper, we restrict the set of strategies to contain *proper strategies*, which only produce finite paths.

Definition 2 (Proper Strategy)**.** A strategy $\pi : S^+ \to A$ is said to be *proper* if, for every infinite path $s_0 s_1 \ldots \in \mathsf{Paths}_\infty(M)$, there exists an integer $n \geq 0$ such that $\pi(s_0 s_1 \ldots s_n)$ is undefined.

A *strategy profile* (π_1, π_2) is a tuple of strategies for each player in G. A path $\rho = s_0 s_1 s_2 \ldots s_k \in \mathsf{Paths}(G)$ is said to be *consistent* with a strategy profile (π_1, π_2) if, for every non-negative integer $i < k$, there exist $a_1 \in \mathsf{Supp}(\pi_1(s_i))$ and $a_2 \in \mathsf{Supp}(\pi_2(s_i))$ such that $T(s_i, (a_1, a_2), s_{i+1}) > 0$. Given a path $\rho \in \mathsf{Paths}(G)$, the *cone* of G defined by the strategy profile (π_1, π_2) is the set, $\mathsf{Cone}(\rho, \pi_1, \pi_2) = \{\rho\rho' \in \mathsf{Paths}(G) \mid \rho\rho' \text{ is consistent with } (\pi_1, \pi_2)\}$.

Specifying Temporal Goals

The temporal goals of players in the game G are specified formally using temporal logic formulas interpreted over finite traces (De Giacomo and Vardi 2013).

Definition 3. Given a set of atomic propositions AP, a LTL_f is produced by the following grammar:

$$\varphi := p \mid \neg\varphi \mid \varphi \land \varphi \mid \bigcirc \varphi \mid \varphi \, \mathsf{U} \, \varphi,$$

made of atomic propositions $p \in AP$, the standard Boolean operators $\neg$ (negation) and $\land$ (conjunction), as well as temporal operators $\bigcirc$ ("Next") and U ("Until").

The formula $\bigcirc \varphi$ denotes that φ holds at the next time instant, while $\varphi_1 \mathsf{U} \varphi_2$ means φ_2 holds at some future instant, and φ_1 holds at all preceding instants. From these operators, the temporal operators $\Diamond$ ("Eventually") and $\Box$ ("Always") are derived: $\Diamond \varphi := \mathsf{true} \mathsf{U} \varphi$ signifies that φ holds at some future instant, and $\Box \varphi := \neg \Diamond \neg \varphi$ indicates φ holds at the current and all future instants. For formal semantics of LTL$_f$, see (De Giacomo and Vardi 2013).

Automata-theoretic Planning with LTL$_f$

A P1 strategy π_1 is said to be *almost-sure winning* to satisfy an LTL$_f$ formula φ if, for any P2 strategy π_2, any path $\rho \in \mathsf{Cone}(s_0, \pi_1, \pi_2)$ satisfies φ with probability one. The automata-theoretic approach to synthesizing an almost-sure winning strategy in G leverages the fact that every LTL$_f$ formula over AP defines a regular language over the alphabet $\Sigma = 2^{AP}$. Such a regular language can be represented using a finite automaton (De Giacomo and Vardi 2013).

Definition 4. A deterministic finite automaton (DFA) is a tuple $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$ where Q is a finite state space. Σ is a finite alphabet. $\delta : Q \times \Sigma \to Q$ is a deterministic transition function. $q_0 \in Q$ is an initial state, and $F \subseteq Q$ is a set of accepting (final) states.

Given a stochastic game and a DFA, the almost-sure winning strategy is computed by applying (De Alfaro, Henzinger, and Kupferman 2007, Algorithm 3) on their product (Baier and Katoen 2008): $\mathcal{G} = \langle V, A, \Delta, v_0, \mathcal{F} \rangle$, where $V = S \times Q$ is the set of states. $\Delta : V \times A \to \mathcal{D}(V)$ is the probabilistic transition function such that, for any states $(s, q), (s', q') \in V$ and $a \in A$, we have $\Delta((s, q), a, (s', q') = T(s, a, s')$ whenever $\delta(q, L(s')) = q'$ and $\Delta((s, q), a, (s', q') = 0$ otherwise. $\mathcal{F} = S \times F$ is a set of final states. $v_0 = (s_0, L(s_0))$ is the initial state.

Preference Modeling

We represent preferences over temporal goals as a binary relation $\trianglerighteq$ on a set of LTL$_f$ formulas Φ (Rahmani, Kulkarni, and Fu 2024). Given two LTL$_f$ formulas, φ_1, φ_2, we write $\varphi_1 \trianglerighteq \varphi_2$ to state that satisfying φ_1 is weakly preferred to satisfying φ_2. We write $\varphi_1 \triangleright \varphi_2$ to state that satisfying φ_1 is strictly preferred to satisfying φ_2.

Every preference relation $\trianglerighteq$ on Φ induces a preorder on Σ^*, which can be represented using a preference automaton.

Definition 5. A preference automaton is defined as a tuple, $\mathcal{P} = \langle Q, \Sigma, \delta, q_0, E \rangle$, where Q is a finite set of states., Σ is the alphabet, $\delta : Q \times \Sigma \to Q$ is a deterministic transition function, $q_0 \in Q$ is the initial state, and $E \subseteq Q \times Q$ is a preorder on Q.

Definition 5 augments the semi-automaton $\langle Q, \Sigma, \delta, q_0 \rangle$ with the preference relation E, instead of a set of accepting states as is typical with a DFA. We write $q \succeq_E q'$ to denote that state q is weakly preferred to q' under preorder E.

The preference automaton encodes a preference relation $\succeq$ on Σ^*. Given $w, w' \in \Sigma^*$, let $q, q' \in Q$ be the two states such that $q = \delta(q_0, w)$ and $q' = \delta(q_0, w')$. Then, the following statements hold (Rahmani, Kulkarni, and Fu 2024, Theorem 1). (a) If $(q, q') \in E$ and $(q', q) \notin E$, then $w \succ w'$; (b) If $(q, q') \notin E$ and $(q', q) \in E$, then $w' \succ w$; (c) If $(q, q') \in E$ and $(q', q) \in E$, then $w \sim w'$; (d) If $(q, q') \notin E$ and $(q', q) \notin E$, then $w \parallel w'$. The procedure to construct a preference automaton is enlisted in (Rahmani, Kulkarni, and Fu 2023) and an online tool is available at https://akulkarni.me/prefltlf2pdfa.html.

Problem Formulation

In a stochastic game where players have adversarial and possibly incomplete preferences, an important decision problem is determining the most desirable outcome that a player can achieve, regardless of the strategy followed by the opponent. We introduce a new solution concept to study these games called non-dominated almost-sure winning (ND-ASWin). This concept extends the solution concept of almost-sure winning—traditionally used for qualitative analysis of stochastic games with a single temporal objective (De Alfaro, Henzinger, and Kupferman 2007)—to study the games with preferences over a set of temporal objectives.

First, we define the notion of a dominating strategy in the context of stochastic games. Let $\succeq_1$ be a preorder on Σ^* induced by P1's preference $\trianglerighteq_1$ on a set of LTL$_f$ objectives, Φ.

Definition 6 (Dominating Strategy Profile)**.** Let (π_1, π_2) and (π_1', π_2') be two strategy profiles in G. We say (π_1, π_2) *strictly dominates* (π_1', π_2') if and only if the following conditions hold:

1. For any $\rho \in \mathsf{Min}(\mathsf{Cone}(s_0, \pi_1, \pi_2), \succeq_1)$ and any $\rho' \in \mathsf{Min}(\mathsf{Cone}(s_0, \pi_1', \pi_2'), \succeq_1)$, we have $L(\rho') \not\succ_1 L(\rho)$.
2. There exists $\rho \in \mathsf{Min}(\mathsf{Cone}(s_0, \pi_1, \pi_2), \succeq_1)$ and $\rho' \in \mathsf{Min}(\mathsf{Cone}(s_0, \pi_1', \pi_2'), \succeq_1)$ such that $L(\rho) \succ_1 L(\rho')$.

Intuitively, (π_1, π_2) *strictly dominates* (π_1', π_2') if (1) none of the least preferred paths produced under the strategy profile (π_1', π_2') is strictly preferred for P1 to any least preferred path produced under (π_1, π_2), and (2) there exists a least preferred path produced under the strategy profile (π_1, π_2) that P1 strictly prefers to some least preferred path produced under (π_1', π_2').

Definition 7 (ND-ASWin)**.** A P1 strategy π_1 is said to be *ND-ASWin* for P1 if, for any P2 strategy π_2, there is no P1 strategy π_1' such that (π_1', π_2) strictly dominates (π_1, π_2).

A ND-ASWin strategy for P1 guarantees the best worst-case outcome for P1, regardless of the strategy followed by P2. A ND-ASWin strategy for P2 is defined analogously.

In this paper, we assume that P1 and P2 have adversarial preferences: If P1 prefers satisfying an LTL$_f$ formula φ over φ', then P2 prefers satisfying φ' over φ.

Problem 1. Given a stochastic game G, P1's preference relation $\trianglerighteq_1$ and P2's adversarial preference relation $\trianglerighteq_2$, both defined on a shared set of LTL$_f$ formulas Φ, synthesize a ND-ASWin strategy for P1 and P2.

Main Results

Product Game

We follow an automata-theoretic approach to synthesize the non-dominated almost-sure winning (ND-ASWin) strategies for P1 and P2. This approach enables us to transform a preference relation over LTL$_f$ objectives to a preference relation over the states of the product game defined below.

Definition 8. Given a game G, a set of LTL$_f$ objectives Φ, and player preferences $\trianglerighteq_i$ for $i = 1, 2$, let $\mathcal{P}_i = (Q, \Sigma, \delta, q_0, E_i)$ be the preference automata corresponding with the relation $\trianglerighteq_i$. The product game is a tuple,

$$H = \langle V, A, \Delta, v_0, \mathcal{E}_1, \mathcal{E}_2 \rangle,$$

where $V = S \times Q$ is the set of states. A is the set of actions. $\Delta : V \times A \to \mathcal{D}(V)$ is a probabilistic transition function. Given two states $v = (s, q)$ and $v' = (s', q')$ in V and an action $a \in A$, we have $\Delta(v, a, v') = T(s, a, s')$ if $q' = \delta(q, L(s'))$ and $\Delta(v, a, v') = 0$, otherwise. $v_0 = (s_0, \delta(q_0, L(s_0)))$ is the initial state. For $i = 1, 2$, $\mathcal{E}_i$ is a preorder on V such that (s, q) is weakly preferred to (s', q') under $\mathcal{E}_i$, denoted $(s, q) \succeq_{\mathcal{E}_i} (s', q')$, if and only if $q \succeq_{E_i} q'$.

The preorder $\mathcal{E}_i$ is called a lifting of the relation E_i (Maly 2020; Barbera and Pattanaik 1984). That is, player-i prefers a state (s, q) to (s', q') in the product game if q is preferred to q' under player-i's preference automaton $\mathcal{P}_i$.

Every path $\rho = s_0 s_1 \dots s_n$ in G induces a unique path $\varrho = v_0 v_1 \dots v_n$ in H where, for all $j = 0, \dots, n$, $v_j = (s_i, q_j)$ and $q_j = \delta(q_0, L(s_0 s_1 \dots s_j))$. We call this path ϱ as the *trace* of ρ in H.

Proposition 1. *Given any finite paths ρ, ρ' in G, let ϱ, ϱ' be their traces in H. Then, $L(\rho) \succeq_1 L(\rho')$ holds if and only if $\mathsf{Last}(\varrho) \succeq_{\mathcal{E}_1} \mathsf{Last}(\varrho')$.*

Proof. Suppose that $L(\rho) \succeq_1 L(\rho')$ holds. By (Rahmani, Kulkarni, and Fu 2024, Thm. 1), we have $q \succeq_{E_1} q'$, where $q = \delta(q_0, L(\rho))$ and $q' = \delta(q_0, L(\rho'))$. Then, for $(s, q) = \mathsf{Last}(\varrho)$ and $(s', q') = \mathsf{Last}(\varrho')$, it must be the case that $(s, q) \succeq_{\mathcal{E}_1} (s', q')$ because $\mathcal{E}_1$ is a lifting of the relation E_1. The proposition is established by observing that the converse of each aforementioned statement is true by definition. □

Recall that an outcome in G is the path generated in the game when P1 and P2 follow their chosen strategies. Following Proposition 1, the outcome in H can be understood as the last state of the trace of this path in H. Hence, to compare two paths in G under preference relation $\succeq_i$ amounts to comparing the last states of the traces of two paths under $\mathcal{E}_i$.

Ranks: A Measure of Quality of Outcome

In this subsection, we introduce a scalar metric called *rank*, a key contribution of this paper, to measure the quality of an outcome in H. Unlike standard approaches to representing incomplete preferences—which rely on multi-utility representations (Rahmani, Kulkarni, and Fu 2024) and are computationally hard (Chen et al. 2013a)—rank offers a simplified and computationally efficient representation. While it is not intended to replace the multi-utility approach, we show that it is sufficient to synthesize ND-ASWin strategies.

Rank is an integer-valued metric that compare two states in H under a preference relation $\mathcal{E}$ based on undominance principle (Sen 1997); a state with smaller rank is *no worse than* a state with higher rank.

Definition 9 (Rank). Given a preorder $\mathcal{E}$ on V, let $Z_0 = \mathsf{Max}(V, \mathcal{E})$ and, for all $k \geq 0$, $Z_{k+1} = \mathsf{Max}(V \setminus \bigcup_{j=0}^{k} Z_j, \mathcal{E})$. The rank of any state $v \in V$, denoted by $\mathsf{rank}_{\mathcal{E}}(v) = k$, is the smallest integer $k \geq 0$ such that $v \in Z_k$.

Definition 9 assigns a unique, finite rank to every state in V given a preorder $\mathcal{E}$. Since the set $\mathsf{Max}(U, \mathcal{E})$ is non-empty for any non-empty subset $U \subseteq V$, the inductive assignment of ranks terminates only when the subset $V \setminus \bigcup_{j=0}^{k} Z_j$ is empty, i.e., when a rank has been assigned to all states in V. Additionally, the sets $Z_0, Z_1 \dots$ are mutually exclusive and exhaustive subsets of V. Therefore, every state in V has a unique rank under a given preorder.

Proposition 2. *The following statements hold for any two states $v, v' \in V$,*

1. *If $\mathsf{rank}_{\mathcal{E}}(v) = \mathsf{rank}_{\mathcal{E}}(v')$ then either $v \sim_{\mathcal{E}} v'$ or $v \parallel_{\mathcal{E}} v'$.*
2. *If $\mathsf{rank}_{\mathcal{E}}(v) > \mathsf{rank}_{\mathcal{E}}(v')$ then $v \not\succeq_{\mathcal{E}} v'$.*
3. *If $v \succ_{\mathcal{E}} v'$ then $\mathsf{rank}_{\mathcal{E}}(v) < \mathsf{rank}_{\mathcal{E}}(v')$.*

Proof. (1) We first show that when $\mathsf{rank}_{\mathcal{E}}(v) = \mathsf{rank}_{\mathcal{E}}(v') = k$, neither $v \succ_{\mathcal{E}} v'$ nor $v' \succ_{\mathcal{E}} v$ can be true. Suppose that $v \succ_{\mathcal{E}} v'$ is true. Then, by Definition 9, both states v and v' must be elements of Z_k, which means that v and v' must be elements of the set $\mathsf{Max}(Y, \mathcal{E})$ where $Y = V \setminus (Z_0 \cup Z_1 \cup \dots \cup Z_{k-1})$. But v and v' cannot both be maximal elements of Y because $v \succ_{\mathcal{E}} v'$, which contradicts our supposition. A similar argument can be used to establish that $v' \succ_{\mathcal{E}} v$. If neither $v \succ_{\mathcal{E}} v'$ nor $v' \succ_{\mathcal{E}} v$ is true, then it must be the case that either $v \sim_{\mathcal{E}} v'$ or $v \parallel_{\mathcal{E}} v'$.

(2) Let $\mathsf{rank}_{\mathcal{E}}(v') = k$. If $\mathsf{rank}_{\mathcal{E}}(v) > \mathsf{rank}_{\mathcal{E}}(v')$ then, by Definition 9, v and v' are both included in the set $Y = V \setminus (Z_0 \cup Z_1 \cup \dots \cup Z_{k-1})$. If $v \succeq_{\mathcal{E}} v'$, then by definition it must be included in $Z_k = \mathsf{Max}(Y, \mathcal{E})$. Since this is not the case, the statement $v \not\succeq_{\mathcal{E}} v'$ must be true.

(3) The proof follows a similar argument as (2). □

Proposition 2 formalizes the connection between preferences between two states and their ranks. It asserts that two states with equal ranks are either indifferent or incomparable under the given preorder. It specifies that a state with a higher rank is not preferred to one with a lower rank. Finally, it establishes that if one state is strictly preferred to another, the former has a strictly smaller rank than the latter.

The converse of statements in Proposition 2 do not necessarily hold, primarily due to the possibility of incomparability arising from incomplete preferences. Specifically, the following statements are not valid: (1′) If $v \parallel_{\mathcal{E}} v'$, then $\mathsf{rank}_{\mathcal{E}}(v) = \mathsf{rank}_{\mathcal{E}}(v')$. (2′) If $v \not\succeq_{\mathcal{E}} v'$, then $\mathsf{rank}_{\mathcal{E}}(v) > \mathsf{rank}_{\mathcal{E}}(v')$. (3′) If $\mathsf{rank}_{\mathcal{E}}(v) < \mathsf{rank}_{\mathcal{E}}(v')$, then $v \succ_{\mathcal{E}} v'$. We provide a counterexample to demonstrate these points.

Example 1. Consider a game with five states $\{v_1, \dots, v_5\}$ where the state v_1 is strictly preferred to v_2, v_3 is strictly preferred to v_4, and v_5 is strictly preferred to v_4. Following Definition 9, the states v_1, v_3, v_5 have rank 0 and the states v_2, v_4 have rank 1. Statement ($1'$) is invalid because v_2 and v_3 are incomparable but they have different ranks. Statement ($3'$) is invalid because the rank of v_3 is smaller than that of v_3 but $v_2 \succ_{\mathcal{E}} v_3$ does not hold since are v_2 and v_3 incomparable. To see the invalidity of statement ($2'$), first note that $v \not\succeq_{\mathcal{E}} v'$ holds when either $v' \succ_{\mathcal{E}} v$ or $v \parallel_{\mathcal{E}} v'$. Now, observe that states v_3 and v_5 are incomparable but they have the same ranks.

Synthesis of Non-dominated Almost-sure Winning Strategy

In this subsection, we characterize P1's ND-ASWin strategy in terms of the rank metric and propose a synthesis procedure by leveraging the algorithm to synthesize an almost-sure winning strategy.

We first establish two properties connecting the outcomes under a strategy profile in H with the rank metric.

Lemma 1. *Given a strategy profile* (π_1, π_2)*, let* $\Omega(\pi_1, \pi_2) = \{v \mid \exists \rho \in \mathsf{Cone}_H(v_0, \pi_1, \pi_2) : v = \mathsf{Last}(\rho)\}$ *be the set of possible outcomes in* H *under* (π_1, π_2)*. For any* $v \in \Omega(\pi_1, \pi_2)$*, if* $\mathsf{rank}_1(v) = \max\{\mathsf{rank}_1(v) \mid v \in \Omega(\pi_1, \pi_2)\}$ *then* $v \in \mathsf{Min}(\Omega(\pi_1, \pi_2), \mathcal{E}_1)$*.*

Proof. By contradiction. Suppose that $\mathsf{rank}_1(v) = \max\{\mathsf{rank}_1(v) \mid v \in \Omega(\pi_1, \pi_2)\}$ but $v \notin \mathsf{Min}(\Omega(\pi_1, \pi_2), \mathcal{E}_1)$. Then, there must exist a state $v' \in \mathsf{Min}(\Omega(\pi_1, \pi_2), \mathcal{E}_1)$ such that $\mathsf{rank}_1(v') = \mathsf{rank}_1(v)$ and $v \succ_1 v'$—which contradicts Proposition 2(1). The proposition is thus established. □

The first property formalized by Lemma 1 states that, for any P2 strategy π_2, every maximum ranked outcome in the set $\Omega(\pi_1, \pi_2)$ is a minimal outcome in $\Omega(\pi_1, \pi_2)$ under $\mathcal{E}_1$. For convenience of notation, we write

$$\mathsf{MaxRank}_1(\pi_1, \pi_2) = \max\{\mathsf{rank}_1(v) \mid v \in \Omega(\pi_1, \pi_2)\}$$

Lemma 2. *Let* π_1 *be a P1 strategy such that, for any P2 strategy* π_2*,* $\mathsf{MaxRank}_1(\pi_1, \pi_2) = \min\{\mathsf{MaxRank}_1(\pi_1', \pi_2) \mid \pi_1' \in \Pi_1\}$*. Then, there is no P1 strategy* π_1' *such that* (π_1', π_2) *strictly dominates* (π_1, π_2)*.*

Proof. By contradiction. Suppose there is π_1' such that (π_1', π_2) strictly dominates (π_1, π_2) but $\mathsf{MaxRank}_1(\pi_1, \pi_2) = \min\{\mathsf{MaxRank}_1(\pi_1', \pi_2) \mid \pi_1' \in \Pi_1\}$. Then, by Definition 6 and Lemma 1, there must exists a state $v' \in \mathsf{Min}(\Omega(\pi_1', \pi_2'))$ such that $v' \succ_1 v$ for some $v \in \mathsf{Min}(\Omega(\pi_1, \pi_2'))$. In this case, Proposition 2(3) implies that $\mathsf{rank}_1(v') < \mathsf{rank}_1(v)$. But we know that π_1 ensures the smallest possible $\mathsf{MaxRank}_1$, which results in a contradiction. Therefore, it must be the case that there is no strategy profile (π_1', π_2) that strictly dominates (π_1, π_2). □

The second property formalized by Lemma 2 states that any strategy that minimizes the rank of the minimal outcomes cannot be dominated by any other P1 strategy.

Together with Definition 7, Lemma 2 enables us to establish that every ND-ASWin strategy π_1 of P1 minimizes the maximum rank achieved by P1 under π_1, regardless of the strategy chosen by P2.

Theorem 1. *Given any P2 strategy* π_2*, every P1 strategy* π_1 *that satisfies* $\mathsf{MaxRank}_1(\pi_1, \pi_2) = \min\{\mathsf{MaxRank}_1(\pi_1', \pi_2) \mid \pi_1' \in \Pi_1\}$ *is a P1's ND-ASWin strategy.*

Algorithm 1: ND-ASWin Strategy.

Input: H: Product game.
Output: π: ND-ASWin strategy.

for $k = 0 \dots k_1^{\max}$ **do**
  $Y_k = \{v \in V \mid \mathsf{rank}_1(v) \leq k\}$.
  $V_k = \mathsf{ASWin}_1(Y_k)$.
  if $v_0 \in V_k$ **then**
    Let π_1 be P1's almost-sure winning strategy to visit V_k from v_0.
    return π_1.

Note that, in every game, P1 has a ND-ASWin strategy. Suppose the rank of initial state v_0 is k and that P2 has a strategy to prevent P1 from achieving any outcome with rank smaller than k. Then, by setting $\pi_1(v_0)$ to undefined, P1 can ensure that it achieves no worse outcome than v_0. In other words, if P1 does not have a strategy to almost-surely achieve a better than v_0, then it has a ND-ASWin strategy to remain at v_0.

Algorithm 1 presents a procedure to compute a ND-ASWin strategy for P1. The algorithm iteratively identifies the smallest rank k for which P1 has a strategy to visit some state with rank k or smaller with probability one. For this purpose, the procedure iteratively computes the almost-sure winning regions $Y_0, Y_1, \dots, Y_j$ until v_0 is included in $\mathsf{ASWin}_1(Y_j)$, for $j = 0, \dots, k_1^{\max}$. Assuming k is the smallest integer for which $v_0 \in Y_k$, the following result establishes that the strategy returned by Algorithm 1 is indeed a ND-ASWin strategy for P1.

Theorem 2. *Every P1 strategy* π_1 *returned by Algorithm 1 is a ND-ASWin strategy for P1.*

Proof. By contradiction. Suppose π_1 is not a ND-ASWin strategy for P1. Then, there exist strategies $\pi_1' \in \Pi_1$ and $\pi_2 \in \Pi_2$ such that (π_1', π_2) strictly dominates (π_1, π_2), which implies the existence of states $v' \in \max\{\mathsf{rank}_1(v) \mid v \in \Omega(\pi_1', \pi_2)\}$ and $v \in \max\{\mathsf{rank}_1(v) \mid v \in \Omega(\pi_1, \pi_2)\}$ such that $\mathsf{rank}_1(v') < \mathsf{rank}_1(v)$. This means that every state in $\Omega(\pi_1', \pi_2)$ has a rank strictly smaller than k. This observation contradicts our hypothesis because Algorithm 1 returns a strategy that almost-surely visits Y_k for the smallest value of k. This concludes our proof. □

Complexity. The time complexity of the procedure to compute the ND-ASWin strategy scales quadratically in the size of the game and linearly with the maximum rank assigned to any state in V. This is because the complexity of solving for an almost-sure winning strategy is quadratic (De Alfaro, Henzinger, and Kupferman 2007) and the above procedure invokes almost-sure winning computation at most

k_1^{max}-times, where $k_1^{\max} = \max\{\mathsf{rank}_1(v) \mid v \in V\}$ is the maximum assigned rank to any state in V.

Qualitative Nash Equilibrium

In this subsection, we show that a pair of ND-ASWin strategies of P1 and P2 is a Nash equilibrium in the stochastic game given qualitative outcomes. We first define the Nash equilibrium in a stochastic game and then characterize P2's ND-ASWin strategy in terms of P1's ND-ASWin strategy.

Definition 10 (Nash Equilibrium)**.** A strategy profile (π_1^*, π_2^*) is a Nash equilibrium in H if and only if the following conditions hold:

$$\forall \pi_1 : \mathsf{Min}(\Omega(\pi_1, \pi_2^*), \succeq_1) \not\succ_1 \mathsf{Min}(\Omega(\pi_1^*, \pi_2^*), \succeq_1)$$
$$\forall \pi_2 : \mathsf{Min}(\Omega(\pi_1^*, \pi_2), \succeq_2) \not\succ_2 \mathsf{Min}(\Omega(\pi_1^*, \pi_2^*), \succeq_2)$$

Definition 10 represents a risk-averse interpretation of Nash equilibrium since it is defined based on the worst-case outcome. Intuitively, a Nash strategy profile guarantees that neither player has a strategy following which the player can achieve a better least preferred path than any least preferred path possible under the Nash strategy profile. Given the combinative nature of preferences over temporal goals, there are multiple ways to define Nash equilibrium (c.f. the eight semantics of preference logic (van Benthem, van Otterloo, and johan 2005).) Given the stochastic nature of our problem, where each strategy profile defines a set of possible paths, we adopt the interpretation that each player evaluates the quality of a strategy profile based on the worst-case outcome possible under that profile.

We now characterize P2's ND-ASWin strategy. Recall that $\mathcal{E}_2$ represents P2's preference relation on V. Since $\mathcal{E}_2$ is adversarial to $\mathcal{E}_1$, for any two states $u, v \in V$, we have $u \succeq_{\mathcal{E}_2} v$ if and only if $v \succeq_{\mathcal{E}_1} u$.

Let $\mathsf{rank}_2(v)$ denote the rank of state v under $\mathcal{E}_2$, analogous to P1's rank function rank_1. Define $k_2^{\max} \triangleq \max\{\mathsf{rank}_2(v) \mid v \in V\}$ as the highest rank assigned to any state in V under $\mathcal{E}_2$.

The key insight behind the characterization of P2's ND-ASWin strategy is that the sum of P1 and P2 ranks of any state in V is constant.

Lemma 3. *Given any state $v \in V$, $\mathsf{rank}_1(v) + \mathsf{rank}_2(v) = k_1^{\max} = k_2^{\max}$.*

Proof. We will show that for any set $U \subseteq V$, a maximal state in U under $\mathcal{E}_1$ is the minimal element in U under $\mathcal{E}_2$ (recall that a state $v \in V$ is minimal under $\mathcal{E}_1$ if there is no state $u \in V$ such that $v \succeq_{\mathcal{E}_1} u$).

First, we note that the minimal states in V under $\mathcal{E}_1$ are all included in the set $Z_{k_1^{\max}}$. This is because $k_1^{\max}$ is the maximum rank assigned to any state in V and if there were a state $u \in V$ which was minimal but not included in $Z_{k_1^{\max}}$, then it must have a rank greater than $k_1^{\max}$, by Proposition 2.

Now, consider a state $v \in Z_{k_1^{\max}}$. We will show that v is a maximal state under $\mathcal{E}_2$, i.e., $\mathsf{rank}_2(v) = 0$. For this, we observe that every state $u \in Z_j$ for any $j < k_1^{\max}$ satisfies $u \succeq_{\mathcal{E}_1} v$ or $u \parallel_{\mathcal{E}_1} v$. Thus, under the opposite preference relation $\mathcal{E}_2$ it must satisfy $v \succeq_{\mathcal{E}_2} u$ or $v \parallel_{\mathcal{E}_2} u$. Since v was a minimal element in V under $\mathcal{E}_2$, there is no u such that $u \succeq_{\mathcal{E}_2} v$. In other words, v is a maximal element in V under $\mathcal{E}_2$. By definition, $\mathsf{rank}_2(v) = 0$.

It follows that every v such that $\mathsf{rank}_1(v) = k_1^{\max}$ has a rank 0 under $\mathcal{E}_2$. For $j = 0, 1, \ldots$, let Y_j, , denote the set of states with rank j under $\mathcal{E}_2$. Using a similar argument, the minimal elements of $\bigcup_{j=0}^{k} Z_j$ are the maximal elements of the set $V \setminus \bigcup_{j=0}^{k} Y_j$. Therefore, every state in Y_j is a state with rank $k_1^{\max} - j$ under $\mathcal{E}_2$.

From this observation, it follows that $k_1^{\max} = k_2^{\max}$ and $\mathsf{rank}_1(v) + \mathsf{rank}_2(v) = k_1^{\max}$. □

Lemma 3 highlights a key property of P2's ND-ASWin strategy; if P1's ND-ASWin strategy ensures a rank of at least k, then every ND-ASWin strategy of P2 must visit a rank $k_2^{\max} - k$ state with a positive probability.

Lemma 4. *Let π_1^* be a ND-ASWin strategy. If $\min_{\pi_2 \in \Pi_2} \mathsf{MaxRank}_1(\pi_1^*, \pi_2) = k$, then every ND-ASWin strategy π_2^* of P2 satisfies $\mathsf{MaxRank}_2(\pi_1^*, \pi_2^*) = k_2^{\max} - k$.*

Proof. When the smallest possible $\mathsf{MaxRank}_1(\pi_1^*, \pi_2)$ when P1 follows its ND-ASWin strategy π_1^* is equal to k, Theorem. 1 implies the existence of a P2 strategy $\widehat{\pi}_2$ such that, for some $v \in \Omega(\pi_1^*, \widehat{\pi}_2)$, $\mathsf{rank}_1(v) = k$. Clearly, by definition of $\mathsf{MaxRank}_1$, we have $\mathsf{MaxRank}_1(\pi_1^*, \pi_2) \leq k$ for any P2 strategy π_2. Together with Lemma 3, this observation means that $k_2^{\max} - k$ is the smallest $\mathsf{MaxRank}_2$ that P2 can achieve when P1 follows π_1^*. Therefore, $\widehat{\pi}_2$ must be a ND-ASWin strategy for P2 and that, for any such P2 strategy $\widehat{\pi}_2$, we have $\mathsf{MaxRank}_2(\pi_1^*, \pi_2^*) = k_2^{\max} - k$. □

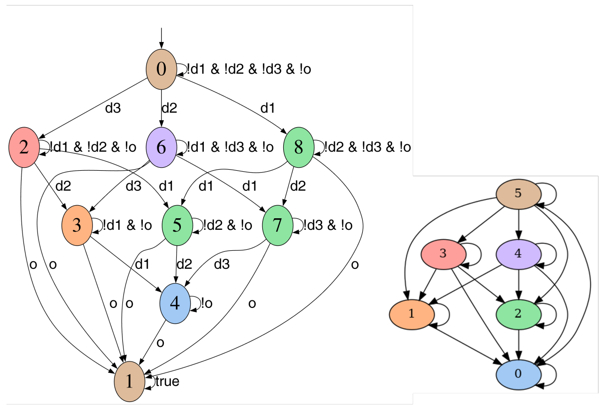
Lemma 4 establishes that when P1 and P2 both follow their ND-ASWin strategies, neither player has a strategy to achieve a better worst-case outcome than that ensured by its ND-ASWin strategy. Together with Definition 10, it follows immediately that a strategy profile consisting of ND-ASWin strategies is a Nash equilibrium.

Theorem 3. *Every strategy profile consisting of ND-ASWin strategies constitutes a Nash equilibrium in H.*

Experiment

We illustrate our theoretical results using a drone delivery scenario in a hostile environment, modeled as a 5×5 stochastic gridworld with two drones, A and B, as shown in Figure 2. This example demonstrates how trustworthy AI systems can navigate hostile settings, aligning with specified preferences while ensuring safety constraints are met.

In this scenario, drone A must deliver three packages from locations p_1, p_2, and p_3 to destinations d_1, d_2, and d_3, while adhering to its delivery preferences. Drone B's objective is to obstruct A from achieving a highly preferred outcome. Both drones can move in four directions (`N, E, S, W`) with an 0.8 probability of reaching the intended cell and a 0.2 probability of landing in an adjacent valid cell. Drones pick up packages by entering the corresponding cells and can attack each other if they occupy the same cell. Black



(a) Semi-automaton (b) Preference Graph

Figure 1: Preference automaton for the relation defined by $\varphi_1 \triangleright \varphi_2$, $\varphi_1 \triangleright \varphi_3$, $\varphi_4 \triangleright \varphi_2$, and $\varphi_4 \triangleright \varphi_3$.

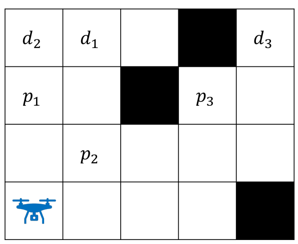


Figure 2: Drone delivery in hostile environment.

cells indicate no-go zones. All deliveries must be completed within 10 rounds.

The central design question we ask is: Given that drone B starts at $(3, 0)$ and both drones act rationally, which starting cell should drone A choose to best satisfy its preferences?

Suppose drone A prefers delivering package 1 over any single package and delivering both packages 2 and 3 over either one individually. This specification is formalized using four LTL$_f$ formulas: $\varphi_i = \Diamond d_i \wedge \Box \neg o$ for $i = 1, 2, 3$ and $\varphi_4 = \Diamond d_2 \wedge \Diamond d_3 \wedge \Box \neg o$, where the preorder $\trianglerighteq$ contains represents four atomic preference relations, $\varphi_1 \triangleright \varphi_2$, $\varphi_1 \triangleright \varphi_3$, $\varphi_4 \triangleright \varphi_2$ and $\varphi_4 \triangleright \varphi_3$. We assume that drone A prefers delivering at least one package over delivering none. The preferences of drone B are completely opposite to those of A. Observe that satisfying φ_2 is incomparable to satisfying φ_3, and satisfying φ_1 is incomparable to satisfying $\varphi_2 \wedge \varphi_3$. Note that treating incomparability as indifference can lead to undesirable outcomes, such as excluding meaningful Nash equilibria (Bade 2005).

Figure 1 depicts P1's preference automaton. The semi-automaton tracks the progress made towards satisfaction of $\varphi_1 \dots \varphi_4$ and the preference graph encodes the comparison between the states of semi-automaton. For example, when drone A delivers packages 1 and 2, the semi-automaton state transitions from 0 to 8 and then to 7. Similarly, when it delivers only package 3, the semi-automaton state is 2. Since there exists an edge from node 3 (red) to 2 (green) in preference graph, we determine that delivering packages 1 and 2

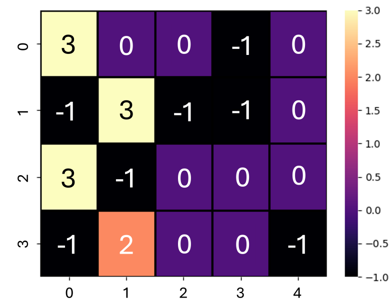


Figure 3: The smallest rank achievable by drone A by following a ND-ASWin strategy.

is strictly preferred over delivering only package 3.

Figure 3 shows in each cell the smallest rank that drone A can guarantee to achieve regardless of the strategy employed by drone B, when it starts from that cell. The invalid starting locations for A are assigned rank -1. For instance, the value 3 at cells $(0, 1)$, $(0, 3)$, or $(1, 2)$ denotes that drone A cannot deliver any packages if it starts from any of these cells. Specifically, in this case, drone B has a strategy to prevent A from either picking or dropping packages. For instance, if drone A starts at the cell $(0, 1)$, then drone B has a strategy to prevent drone A from picking up any package. This is because B can reach the cells labeled p_1, p_2 and p_3 before drone A can reach them, and A cannot enter the cell with B since B can use the `attack` action to disable it. However, when B cannot prevent A from picking up any of the three packages, A can enforce a rank 0 outcome against every possible strategy of B. Therefore, we conclude that drone A must start at any cell with rank 0 to achieve the best possible outcome for itself.

Conclusion

In this paper, we proposed a novel automata-theoretic approach to synthesizing preference satisfying strategies in a two-player stochastic game with adversarial, incomplete preferences. We introduced the concept of non-dominated almost-sure winning strategies in two-player stochastic games, which provably guarantee robustness against adversarial actions while still remaining aligned with the specified, potentially incomplete, human preferences. By utilizing LTL$_f$, we effectively modeled the complexities of human preferences, where some outcomes may remain unranked or incomparable. Our results demonstrated that these strategies lead to Nash equilibria, ensuring stable and preference-aligned outcomes in the game.

While this work contributes to the state-of-the-art in game theory and formal methods, it also supports the broader effort of developing trustworthy AI systems capable of making strategic decisions that align with human values in complex, real-world settings. We believe these insights can pave the way for further advancements in preference-aligned AI, offering new avenues for research in robust AI decision-making under uncertainty.

Acknowledgements

This work is partially supported by the Air Force Office of Scientific Research under award number FA9550-21-1-0085, Army Research Office grant under award number W911NF-23-1-0317, and Office of Naval Research grant under award number N00014-24-1-2797.

References

Bade, S. 2005. Nash equilibrium in games with incomplete preferences. *Economic Theory*, 26: 309–332.

Baier, C.; and Katoen, J.-P. 2008. *Principles of model checking*. MIT press.

Barbera, S.; and Pattanaik, P. K. 1984. Extending an order on a set to the power set: some remarks on Kannai and Peleg's approach. *Journal of Economic Theory*, 32(1): 185–191.

Blass, A.; Gurevich, Y.; Nachmanson, L.; and Veanes, M. 2006. Play to test. In *Formal Approaches to Software Testing: 5th International Workshop, FATES 2005, Edinburgh, UK, July 11, 2005, Revised Selected Papers 5*, 32–46. Springer.

Bosi, G.; and Herden, G. 2012. Continuous multi-utility representations of preorders. *Journal of Mathematical Economics*, 48: 212–218.

Brunello, A.; Montanari, A.; and Reynolds, M. 2019. Synthesis of LTL formulas from natural language texts: State of the art and research directions. In *26th International symposium on temporal representation and reasoning (TIME 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Chatterjee, K.; and Henzinger, T. A. 2012. A survey of stochastic ω-regular games. *Journal of Computer and System Sciences*, 78(2): 394–413.

Chatterjee, K.; Katoen, J. P.; Mohr, S.; Weininger, M.; and Winkler, T. 2023. Stochastic games with lexicographic objectives. *Formal Methods in System Design*, 1–41.

Chen, T.; Forejt, V.; Kwiatkowska, M.; Simaitis, A.; and Wiltsche, C. 2013a. On stochastic games with multiple objectives. In *Mathematical Foundations of Computer Science 2013: 38th International Symposium, MFCS 2013, Klosterneuburg, Austria, August 26-30, 2013. Proceedings 38*, 266–277. Springer.

Chen, T.; Kwiatkowska, M.; Simaitis, A.; and Wiltsche, C. 2013b. Synthesis for multi-objective stochastic games: An application to autonomous urban driving. In *Quantitative Evaluation of Systems: 10th International Conference, QEST 2013, Buenos Aires, Argentina, August 27-30, 2013. Proceedings 10*, 322–337. Springer.

Dalrymple, D.; Skalse, J.; Bengio, Y.; Russell, S.; Tegmark, M.; Seshia, S.; Omohundro, S.; Szegedy, C.; Goldhaber, B.; Ammann, N.; et al. 2024. Towards Guaranteed Safe AI: A Framework for Ensuring Robust and Reliable AI Systems. *arXiv preprint arXiv:2405.06624*.

De Alfaro, L.; Henzinger, T. A.; and Kupferman, O. 2007. Concurrent reachability games. *Theoretical computer science*, 386(3): 188–217.

De Giacomo, G.; and Vardi, M. Y. 2013. Linear temporal logic and linear dynamic logic on finite traces. In *IJCAI'13 Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, 854–860. Association for Computing Machinery.

Finucane, C.; Jing, G.; and Kress-Gazit, H. 2010. LTLMoP: Experimenting with language, temporal logic and robot control. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1988–1993. IEEE.

Grädel, E.; Thomas, W.; and Wilke, T. 2003. *Automata, logics, and infinite games: a guide to current research*, volume 2500. Springer.

Hansson, S. O.; and Grüne-Yanoff, T. 2022. Preferences. In Zalta, E. N., ed., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2022 edition.

Kokkala, J.; Berg, K.; Virtanen, K.; and Poropudas, J. 2019. Rationalizable strategies in games with incomplete preferences. *Theory and Decision*, 86: 185–204.

Kulkarni, A. N.; and Fu, J. 2022. Opportunistic Qualitative Planning in Stochastic Systems with Preferences over Temporal Logic Objectives. *arXiv preprint arXiv:2203.13803*.

Li, M.; Turrini, A.; Hahn, E. M.; She, Z.; and Zhang, L. 2020. Probabilistic preference planning problem for markov decision processes. *IEEE transactions on software engineering*, 48(5): 1545–1559.

Liu, J. X.; Yang, Z.; Schornstein, B.; Liang, S.; Idrees, I.; Tellex, S.; and Shah, A. 2022. Lang2ltl: Translating natural language commands to temporal specification with large language models. In *Workshop on Language and Robotics at CoRL 2022*.

Maly, J. 2020. Lifting preferences over alternatives to preferences over sets of alternatives: The complexity of recognizing desirable families of sets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 2152–2159.

Ok, E. A. 2002. Utility Representation of an Incomplete Preference Relation. *Journal of Economic Theory*, 104: 429–449.

Pnueli, A.; and Zuck, L. D. 1993. Probabilistic verification. *Information and computation*, 103(1): 1–29.

Rahmani, H.; Kulkarni, A. N.; and Fu, J. 2023. Probabilistic planning with partially ordered preferences over temporal goals. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 5702–5708. IEEE.

Rahmani, H.; Kulkarni, A. N.; and Fu, J. 2024. Preference-Based Planning in Stochastic Environments: From Partially-Ordered Temporal Goals to Most Preferred Policies. *arXiv preprint arXiv:2403.18212*.

Sasaki, Y. 2019. Rationalizability in multicriteria games. *International Journal of Game Theory*, 48(2): 673–685.

Sen, A. 1997. Maximization and the Act of Choice. *Econometrica: Journal of the Econometric Society*, 745–779.

Seshia, S. A.; Sadigh, D.; and Sastry, S. S. 2022. Toward verified artificial intelligence. *Communications of the ACM*, 65(7): 46–55.

Tegmark, M.; and Omohundro, S. 2023. Provably safe systems: the only path to controllable AGI. *arXiv preprint arXiv:2309.01933*.

van Benthem, J.; van Otterloo, S.; and johan, O. R. 2005. Preference logic, conditionals and solution concepts in games. *Journal of Pharmacology and Experimental Therapeutics*.

Wing, J. M. 2021. Trustworthy AI. *Commun. ACM*, 64(10): 64–71.

Özgür Evren; and Ok, E. A. 2011. On the multi-utility representation of preference relations. *Journal of Mathematical Economics*, 47: 554–563.