Multi-Agent Systems with Quantitative Satisficing Goals

Senthil Rajasekaran[1] , **Suguman Bansal**[2] and **Moshe Y. Vardi**[1]

[1]Rice University
[2]Georgia Institute of Technology
sr79@rice.edu, suguman@gatech.edu, vardi@rice.edu

Abstract

In the study of reactive systems, *qualitative* properties are usually easier to model and analyze than *quantitative* properties. This is especially true in systems where mutually beneficial cooperation between agents is possible, such as multi-agent systems. The large number of possible payoffs available to agents in reactive systems with quantitative properties means that there are many scenarios in which agents deviate from mutually beneficial outcomes in order to gain negligible payoff improvements. This behavior often leads to less desirable outcomes for all agents involved. For this reason we study *satisficing* goals, derived from a decision-making approach aimed at meeting a good-enough outcome instead of pure optimization. By considering satisficing goals, we are able to employ efficient automata-based algorithms to find pure-strategy Nash equilibria. We then show that these algorithms extend to scenarios in which agents have multiple thresholds, providing an approximation of optimization while still retaining the possibility of mutually beneficial cooperation and efficient automata-based algorithms. Finally, we demonstrate a one-way correspondence between the existence of ϵ-equilibria and the existence of equilibria in games where agents have multiple thresholds.

1 Introduction

Work in formal verification has seen recent trends towards generalizations, such as considering *quantitative* properties as opposed to qualitative ones [Bulling and Goranko, 2022; Gupta, 2016; Kwiatkowska, 2007] and reasoning about *multi-agent* systems as opposed to two-agent systems [Grädel *et al.*, 2002; Gupta, 2016; Mogavero *et al.*, 2014; Shoham and Leyton-Brown, 2009; Van der Hoek and Wooldridge, 2008; Wooldridge, 2009]. This motivates combining both generalizations to reason about extremely general systems.

The quantitative properties in the multi-agent formal verification literature are usually derived from modifications of qualitative properties, i.e. temporal logics [Bulling and Goranko, 2022; Bouyer *et al.*, 2019]. Recent work in fields like reinforcement learning and planning [Sutton and Barto,

2018; Lavalle, 2006], however, trends towards a different type of system in which more general quantitative state-based rewards are considered. In these systems, agents choose actions in order to receive rewards, and the quantitative property being considered reasons directly about the sequence of rewards obtained. Since the sum of these rewards is likely to be infinite in an infinite execution of a system, an aggregation function is usually applied to this sequence [Osborne and Rubinstein, 1994; Sutton and Barto, 2018]. One of the most common aggregation functions is the *discounted sum* – as each time step passes all rewards are decreased geometrically by multiplication through a discount factor between 0 and 1 [Osborne and Rubinstein, 1994]. This function is popular since it not only guarantees convergence but also discounts rewards in the future against rewards in the present, incentivizing agents to follow the rational economic behavior of preferring to receive rewards immediately. The use of the discounted sum then allows for us to reason about reward sequences, commonly referred to as *payoffs*, in such systems.

We choose to reason about the *Nash equilibrium* solution concept as our quantitative property of interest as it is one of the most important concepts in the theory of multi-agent systems [Nash, 1950; Shoham and Leyton-Brown, 2009; Wooldridge, 2009]. A Nash equilibrium consists of a strategy for each agent such that no agent in the system can increase his payoff through a unilateral deviation. Since agents are not incentivized to change their strategy, Nash equilibria represent stable points in multi-agent systems. Deterministic behavior is preferred in many settings (i.e. formal verification), so *pure* Nash equilibria, which consist of deterministic strategies, are a preferable concept [Bouyer *et al.*, 2010; Bouyer *et al.*, 2011; Ummels *et al.*, 2015]. We adopt this preference here, referring to pure Nash equilibria whenever we write Nash equilibria.

In these systems it is natural to consider agents that are incentivized by maximizing their payoff – i.e. their *goal* is payoff maximization. A difficulty arises, however, when this type of agent is considered alongside the Nash equilibria solution concept. Agents with a maximization goal are incentivized to deviate for *an arbitrary* improvement in payoff, even if that improvement is minuscule, a situation that naturally arises with the accumulation of a discount factor. We demonstrate that this behavior removes opportunities for mutually beneficial cooperation between agents with an example (Figure 1).

For this reason the notion of an ϵ-equilibria [Daskalakis *et al.*, 2006; Roughgarden, 2010] is often considered, in which an agent is not incentivized to deviate unless it gain at least ϵ more payoff for some fixed ϵ. While this solution concept disincentivizes arbitrary deviations, we are not aware of general solutions to finding ϵ-equilibria in these systems. A system that that disincentivizes minuscule deviations (increasing cooperative opportunities) and allows for efficient algorithms that find Nash equilibria is then missing from the literature.

We address this gap by introducing a new type of agent goal based on the concept of *satisficing* [Simon, 1956]. An agent with a satisficing goal is only interested in searching for some payoff that would meet a fixed threshold, i.e. a "good enough" outcome. The use of this type of goal has been successful in stochastic systems with two agents [Bansal *et al.*, 2021; Bansal *et al.*, 2022b]. Satisficing goals address the problem of agents deviating for negligible gain since agents only have a single threshold to meet. Furthermore, they transform our quantitative system into one with qualitative properties, allowing us to use efficient automata-based techniques [Rajasekaran and Vardi, 2021; Rajasekaran and Vardi, 2022].

Satisficing goals, however, come at the cost of removing much of the expressiveness we wished to originally capture by considering general quantitative systems. To this end, we introduce a novel approach based on *multi-satisficing* goals. An agent with a multi-satisficing goal has multiple thresholds to meet and is incentivized to meet as many of his thresholds as possible. This allows us to recapture much of the expressiveness of our original quantitative system while still not incentivizing minuscule deviations. Furthermore, we show how to extend the constructions used for satisficing goals (with single thresholds) to retain the complexity-theoretic benefit of our automata-theoretic approach, maintaining a PSPACE upper bound for finding Nash equilibria that matches the complexity of the qualitative setting [Rajasekaran and Vardi, 2021; Rajasekaran and Vardi, 2022]. We conjecture that this bound is tight because the problem of finding Nash equilibria requires keeping track of the states of an unbounded number of agents. Since, however, satisficing goals in multi-agent games are a novel concept we found it challenging to construct a complexity-theoretic reduction; the specification of satisficing goals is quite different from the specifications of the canonical PSPACE-hard multi-agent problems [Kozen, 1977; Sipser, 2006; Hopcroft and Ullman, 1979].

Finally, by considering many thresholds that are close together we can establish a relationship between Nash equilibria in systems where agents have multi-satisficing goals to the more traditional notion of ϵ-equilibria. Section 6 demonstrates how to construct thresholds in a system with multisatisficing goals such that the existence of an equilibria in that system corresponds to an ϵ-equilibria in the system where only payoffs are considered. The inverse relationship represents a powerful open question, as a positive result would allow for a highly efficient automata-based algorithm that solves for ϵ-equilibria in a very general form of system.

The main contribution of this this paper is the introduction of a new type of goal (multi-satisficing) for agents in a very general quantitative multi-agent concurrent game setting. Three main reasons are provided in support of these goals. First, they provide a framework that offers more opportunities for collaboration between agents, providing a strategic reason to consider multi-satisficing goals. Second, they admit efficient automata-based algorithms. In Section 4 we discuss how a naive application of the techniques in [Bouyer *et al.*, 2015] would yield a NEXPTIME upper bound as opposed to the PSPACE upper bound provided in our paper, providing a complexity-theoretic reason to consider this approach to multi-satisficing goals. Finally, this paper demonstrates a one-way correspondence between equilbiria in systems with multi-satisficing goals and ϵ-equilibria, tying the concept of multi-satisficing goals back to a widely accepted and used notion in the literature.

2 Preliminaries

2.1 Büchi, Safety, and Co-safety automata

A *Büchi automaton* is a tuple $\mathcal{A} = \langle \Sigma, Q, q_0, \delta, F \rangle$, where Σ is a finite *input alphabet*, Q is a finite set of *states*, $q_0 \in Q$ is the *initial state*, $\delta \subseteq (Q \times \Sigma \times Q)$ is the *transition relation*, and $F \subseteq Q$ is the set of *accepting states*. A Büchi automaton is *deterministic* if for all states q and input symbols $s \in \Sigma$, we have that $|\{q' : (q, s, q') \in \delta\}| \leq 1$. For a word $w = w_0 w_1 \cdots \in \Sigma^\omega$, a *run* ρ of $\mathcal{A}$ over w is a sequence $q_0, q_1 \cdots \in Q^\omega$ such that q_0 is the initial state and $\tau_i = (q_i, w_i, q_{i+1}) \in \delta$ for all $i \in \mathbb{N}$. Let $inf(\rho)$ denote the set of states that occur infinitely often in run ρ. A run ρ is an *accepting run* if $inf(\rho) \cap F \neq \emptyset$. If $\mathcal{A}$ has an accepting run over w then it *accepts* w. The *language* of $\mathcal{A}$, denoted $\mathcal{L}(\mathcal{A})$, is the set of words accepted by $\mathcal{A}$. Languages specified by Büchi automata are called ω-regular and are closed under set-theoretic union, intersection, and complementation [Grädel *et al.*, 2002]. A *safety automaton* is a deterministic Büchi automaton with a single rejecting sink state [Kupferman and Vardi, 2000]; it accepts a word w if no finite prefix of w leads to the rejecting state. A *co-safety automaton* is a deterministic Büchi automaton with a single accepting sink state [Kupferman and Vardi, 2000]; it accepts a word w if some finite prefix of w leads to the accepting state.

Comparator automata. Given an aggregate function $f : \mathbb{Z}^\omega \to \mathbb{R}$, a relation $\mathsf{R} \in \{<, >, \leq, \geq, =, \neq\}$, and a threshold value $v \in \mathbb{Q}$, the *comparator automaton for f with upper bound $\mu \in \mathbb{Z}$, relation R, and threshold $v \in \mathbb{Q}$* is an automaton that accepts an infinite word A over the alphabet $\Sigma = \{-\mu, -\mu + 1, \ldots \mu\}$ iff $f(A) \mathsf{R} v$ holds [Bansal *et al.*, 2018]. The discounted sum of an infinite-length weight-sequence $W = w_0, w_1, \cdots \in \mathbb{Z}^\omega$ with discount factor $\gamma > 1$ is given by $\sum_{i=0}^{\infty} \frac{w_i}{\gamma^i}$. The comparator automaton for the discounted-sum has been shown to be a safety or cosafety automaton when the discount factor $\gamma > 1$ is an integer for all values of R, μ and v [Bansal and Vardi, 2019; Bansal *et al.*, 2021]. Furthermore, no Büchi comparator automata exist for non-integer discount factors $\gamma > 1$ [Bansal *et al.*, 2022a]. Here, we only consider integral discount factors.

2.2 Multi-agent Systems

Multi-agent systems provide a powerful framework for reasoning about interactions between multiple strategic agents. One of the most general multi-agent systems is given by

a *concurrent game*, defined as a 6-tuple $\langle V, \mathbf{v_0}, \Omega, A, \tau, \alpha \rangle$, where **(1)** V is a finite set of states. $\mathbf{v_0} \in V$ is the initial state **(2)** $\Omega = \{0 \ldots k - 1\}$ is a set of agents of size k **(3)** $A = \langle A_0 \ldots A_{k-1} \rangle$ is a tuple of action sets; the set A_i is associated with Agent i, and it represents the actions available to that agent. We denote $D = A_0 \times A_1 \times \ldots \times A_{k-1}$ as the set of *decisions*, which represent global actions by the full set of agents **(4)** $\tau : V \times D \to V$ is the *transition function*. **(5)** $\alpha = \langle \alpha_0 \ldots \alpha_{k-1} \rangle$ is a tuple of *goals*. The goal α_i is associated with Agent i, and it represents some ordering on outcomes of the game that constitutes the agent's preferences. We discuss different types of goals below. Intuitively, at each state $v \in V$ of the game, the agents concurrently choose actions from their set of actions. These actions then determine a decision, which is used to transition the state of the game via τ. Given this framework, we now define the notion of a *play* in a concurrent game. A play $\rho \in (V \times D)^\omega$ is an infinite sequence of states and decisions $(\mathbf{v_0}, d_0), (v_1, d_1) \ldots$ such that $\mathbf{v_0}$ is the initial state and, for all $j \geq 0$, $v_{j+1} = \tau(v_j, d_j)$. Plays represent outcomes of the game, so each α_i represents an ordering on the set of plays as we discuss further below.

Agents are then incentivized to choose actions that create plays with maximal preference. These actions are dictated by a *strategy*. A strategy for Agent i is a function $\pi_i : (V \times D)^* \times V \to A_i$. Intuitively, this is a deterministic function that, given the observed history of the game (represented by a sequence of previously seen state-decision pairs $(V \times D)^*$ and a current state V) returns an action $a_i \in A_i$.

Let Π_i represent the set of strategies for agent i. The set of *strategy profiles* is defined as $\Pi = \times_{i \in \Omega} \Pi_i$. Thus, a strategy profile $\pi \in \Pi$ is a tuple of strategies, one for each agent, of type $(V \times D)^* \times V \to D$. We define π_{-i} to be the strategy profile π with the i-th element removed. By a slight abuse of notation, we combine π_{-i} with a strategy π_i' for Agent i to create a new strategy profile $\langle \pi_{-i}, \pi_i' \rangle$.

Strategies are deterministic, so given a concurrent game $G = \langle V, \mathbf{v_0}, \Omega, A, \tau, \alpha \rangle$ and a strategy profile π, there is a unique play ρ resulting from a strategy profile π, given by **(1)** $d_0 = \pi((\epsilon, \mathbf{v_0}))$ **(2)** $\rho_0 = \langle \mathbf{v_0}, d_0 \rangle$ **(3)** $v_{i+1} = \tau(v_i, d_i)$ **(4)** $d_{i+1} = \pi((\rho_0, \ldots, \rho_i), v_{i+1})$ **(5)** $\rho_{i+1} = \langle v_{i+1}, d_{i+1} \rangle$. We denote this play as ρ_π and call it the *primary trace* of π.

We now define the notion of a *Nash equilibrium*. A Nash equilibrium is a strategy profile $\pi = \langle \pi_0 \ldots \pi_{k-1} \rangle$ such that for every agent $i \in \Omega$ there is no Agent i strategy π_i' such that, under the preference induced by α_i, the play $\rho_{\langle \pi_{-i}, \pi_i' \rangle}$ is strictly preferred to the play ρ_π. Therefore, under a Nash equilibrium strategy profile no agent may deviate from the profile in order to obtain a more preferable result. Since both the notion of strategies for individual agents and the transition function in a concurrent games are deterministic here, when we refer to concepts such as Nash equilibria, we implicitly mean *pure-strategy* Nash equilibria.

Turn-Based games. An important class of games are *turn-based* games, in which the state set V is partitioned among the k agents: $V = V_0 \cup \ldots \cup V_{k-1}$. In a state $v \in V_i$ only the action of Agent i affects the transition. Formally, let d and d' be two decisions such that $d[i] = d'[i]$ (they agree on A_i), then we have that $\tau(v, d) = \tau(v, d')$. It is convenient to view

this as if only Agent i takes an action a in a state $v \in V_i$, and, by slight abuse of notation we write that this action causes a move from the state v to the state $\tau(v, a)$. We denote turn-based games as $\langle V_0, \ldots, V_{k-1}, \mathbf{v_0}, E, \alpha \rangle$, where E is a set of edges over $V = \bigcup_{i=0}^{k-1} V_i$ describing the possible moves in the game, the sets of agents and actions are implicit.

2.3 Reachability and Safety games

Two important types of goals α_i are specified by *Reachability* and *Safety* conditions. A reachability goal is specified by a set of states $R \subseteq V$. An agent with a reachability goal prefers plays that visit a state $r \in R$ over those that do not. A safety goal is the dual of a reachability goal and specified by a set of states $S \subseteq V$. An agent with a safety goal prefers plays that do not visit states in $V \setminus S$ over those that do. In this paper, when we refer to reachability and safety *games* we specifically mean a two-agent turn-based game in which one agent has a reachability or safety goal and the other agent has the negation of this goal [McNaughton, 1993]. In a reachability game $G = \langle V_0, V_1, \mathbf{v_0}, E, R \rangle$ Agent 0 has a preference for plays that visit at least one state in the reachability set specified by $\alpha_0 = R$; Agent 1 has the safety goal $\alpha_1 = V \setminus R$ so the two goals are mutually exclusive.

For both reachability and safety games we use the notation $G = \langle V_0, V_1, E, C \rangle$, where C represents the winning condition (either a reachability or safety set) without specifying an initial state. Instead, we define a set of "winning states" $\mathsf{Win}_0(G)$. We say a state $v \in V$ is in $\mathsf{Win}_0(G)$ if Agent 0 has a *winning strategy* π_0 starting in state v, which means that for all Agent 1 strategies π_1, we have that $\rho_{\langle \pi_0, \pi_1 \rangle}$ is winning for Agent 0 when v is treated as the initial state. If no such Agent 0 strategy exists then Agent 1 has a winning strategy from v and v belongs to the analogously defined $\mathsf{Win}_1(G)$. Both sets can be computed in time $O(|V| + |E|)$ [Martin, 1975; McNaughton, 1993].

3 Satisficing Games

A *discounted-sum game* $\langle V, \mathbf{v_0}, \Omega, A, \tau, \gamma, \mathcal{W}, \alpha \rangle$ is a concurrent game $\langle V, \mathbf{v_0}, \Omega, A, \tau, \alpha \rangle$ with the following additions: **(1)** $\gamma > 1$ is an integral discount factor, and **(2)** $\mathcal{W} : V \times D \to \mathbb{Z}^k$ is a *reward function* that associates each state-decision pair with a k-length integer-valued vector. Intuitively, the reward function associates integer rewards for each agent as a function of the state and the decision. The discount factor devalues rewards in the future with respect to rewards in the present, disincentivizing agents from waiting before taking actions to get rewards. The addition of these quantitative properties allows us to assign numerical values to plays, which we use to define goals α. Given a play $\rho = (\mathbf{v_0}, d_0), (v_1, d_1) \ldots$, the *cumulative reward* earned by Agent i from ρ is defined as $R_i(\rho) = \sum_{j=0}^{\infty} \mathcal{W}(v_j, d_j)[i] \cdot \frac{1}{\gamma^j}$.

An often-used choice for the goal α_i is an *optimization* goal, i.e. a maximization goal which always prefers plays with higher cumulative reward to those with lower cumulative reward. An alternative we consider here is a *satisficing* goal α_i, specified by a pair $\langle \mathsf{R}_i, t_i \rangle$, where $\mathsf{R}_i \in \{<, \leq, >, \geq, =, \neq\}$ is a comparison relation and $t_i \in \mathbb{Q}$ is a threshold. These goals induce a binary preference on plays - an agent
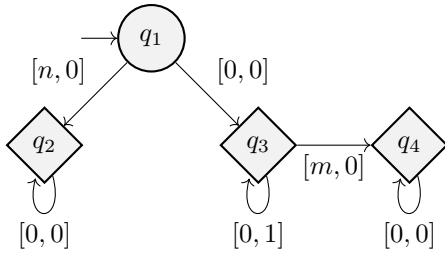
Figure 1: A two-agent turn-based discounted-sum game with states $V = \{q_1, q_2, q_3, q_4\}$. Agent 0 owns the circled state ($V_0 = \{q_1\}$); Agent 1 owns the diamond states ($V_1 = \{q_2, q_3, q_4\}$).

with a satisficing goal prefers plays ρ in which $R_i(\rho) \; \mathsf{R}_i \; t_i$ holds to plays where it does not hold. We demonstrate the utility of considering agents with satisficing goals as opposed to optimization goals under the Nash equilibria solution concept through an example.

Consider the two-player turn-based discounted-sum game in Fig 1. Each edge is labeled with a pair of integers, representing the reward function $\mathcal{W}$ - Agent 0 and Agent 1 receive the first and second reward, respectively. Suppose that $m \gg n$, such that for our given discount factor γ we have $\frac{m}{\gamma^k} > n > \frac{m}{\gamma^{k+1}}$ for a large k. This means there exists a path from q_1 to q_4 that loops in q_3 for several turns such that the (discounted) reward gained by Agent 0 is higher than what it would have achieved on the path from q_1 to q_2.

Assume that both agents are given maximization goals. Note that the single play that goes from q_1 to q_2 and stays there represents the only Nash equilibrium in this game. In this profile, Agent 0 chooses to go to q_2 leading to a reward of n and 0 for both agents, respectively. The alternative Agent 0 strategy of moving to q_3 would lead to a 0 reward for the Agent 0, since Agent 1 is incentivized to loop at q_3 forever – even when the marginal reward gained by each successive loop is minuscule due to the discount factor. This is because Agent 1 always chooses plays with more cumulative reward per her maximization goal α_1. Thus, the game's only stable point emerges from a double threat - Agent 1 threatens to loop at q_3 forever, while Agent 0 does not give her the chance, as Agent 1's desire to loop at q_3 forever for increasingly minuscule rewards means that Agent 0 cannot "negotiate" with her.

For this reason, a notion of ϵ-Nash equilibria is sometimes considered, in which a deviation is not considered preferable unless it gains at least ϵ more reward for some fixed constant $\epsilon > 0$ [Daskalakis *et al.*, 2006; Roughgarden, 2010]. This gives the players a sense of *rationality* in that they will not deviate for arbitrarily minuscule rewards. Here, we choose another form of rationality based on attaining certain thresholds via satisficing goals. Satisficing is a decision-making strategy that entails searching through available alternatives until a *good-enough* threshold is met [Simon, 1956]. It is, therefore, a reasonable approach by a decision-maker in circumstances where an optimal solution cannot be determined or are undesirable. In our example, if Agent 1 had a properly chosen satisficing goal instead of an optimization goal, she could accept a strategy that loops at q_3 a finite number of times before moving to q_4. To this end we study the existence

of Nash equilibria in *satisficing games*, which are discounted-sum games where each agent has a satisficing goal.

We note that the change from optimization to satisficing goals also influences the definition of Nash equilibria since there are only two possible preferences for each agent. Given a fixed play ρ in a satisficing game, for each Agent i it is either the case that $R_i(\rho) \; \mathsf{R}_i \; t_i$ holds or does not. Thus, we introduce the concept of a *winning set* W induced by a play ρ to be the set of agents $W \subseteq \Omega$ such that $i \in W$ iff $R_i(\rho)\alpha_i$ holds. Therefore we can find Nash equilibria by finding *W-Nash Equilibria* (W-NE for short), which are Nash equilibria with winning set W - varying W as needed.

The next section develops the technical tools needed to find W-NE in satisficing games. Satisficing games are able to avoid certain undesirable behaviors from agents, but this comes at the cost of equipping agents with only two possible payoffs. While the tendency for agents with optimization goals to deviate for negligible rewards stemmed from the potentially uncountable number of payoffs available, limiting agents to two possible payoffs in the satisficing case presents its own concerns. In Section 5 (Multi-Satisficing Goals) we show how to extend the analysis of satisficing goals to allow for a finite but unbounded number of payoffs for each agent, which we believe represents an important middle ground between satisficing and optimization goals.

4 Characterization of Nash Equilibria

Agents in multi-agent satisficing games distinguish between two types of plays based on goal satisfaction, allowing us to characterize the existence of Nash equilibria through the existence of W-NE. Our goal is to develop an automata-theoretic characterization of W-NE existence; we begin by describing how to model satisficing goals via automata.

Since the discount factor γ is integral, we construct comparator automata that recognize infinite sequences of rewards that satisfy a given satisficing goal in a multi-agent satisficing game. Let μ_i represent the maximum magnitude of the rewards assigned to Agent i, i.e. the range of $\mathcal{W}(\cdot)[i]$ is in the interval $[-\mu_i, \mu_i]$. By following the construction in [Bansal *et al.*, 2021], we construct a comparator automaton A^i that recognizes reward sequences that satisfy the goal $\alpha_i = \langle \mathsf{R}_i, t_i \rangle$. The size of A^i is $\frac{\mu_i \eta_i}{\gamma - 1}$, where η_i is the length of the base γ representation of $t_i = t_i[0] \dots (t_i[m] \dots t_i[\eta - 1])^\omega$. This is a linear-sized construction under the assumption that μ_i is provided in unary. Furthermore, if $\mathsf{R}_i \in \{\leq, \geq, =\}$ then A^i is a safety automaton; it is a co-safety automaton otherwise [Bansal *et al.*, 2021]. Therefore when given a satisficing goal α_i, we construct a Büchi automaton $A^i = \langle \Sigma^i, Q^i, q_0^i, \delta^i, F^i \rangle$, where $\Sigma^i = \{-\mu_i \dots \mu_i\}$. Note that when A^i is a co-safety automaton, once a state in F^i is reached the automaton accepts and when A^i is a safety automaton, once a state not in F^i is reached the automaton rejects. Since agents may either have a safety or co-safety automaton representation of their goal, we denote Ω_S as the set of agents with a safety goal representations; Ω_C is analogously denoted for co-safety representations. Furthermore, $W_S = \Omega_S \cap W$, $W_C = \Omega_C \cap W$, $\overline{W} = (\Omega \setminus W)$, $\overline{W}_C = \overline{W} \cap \Omega_C$, and $\overline{W}_S = \overline{W} \cap \Omega_S$.

The qualitative nature of satisficing goals means that a strategy profile π is a W-NE iff (**1**) Precisely the goals for agents in W are satisfied on ρ_π and (**2**) For agents $j \notin W$, there is no Agent-j-strategy π'_j such that Agent j's goal is satisfied on $\rho_{\langle \pi_{-j}, \pi'_j \rangle}$. This is because in the qualitative setting, agents in W have achieved their maximal preference and therefore have no incentive to deviate. Therefore, it is enough to only consider deviations from agents outside of W. Following [Rajasekaran and Vardi, 2021], we refer to the first condition as the Primary-Trace condition and the second as the j-Deviant-Trace Condition. We proceed by analyzing each condition separately and then show how to unify them.

4.1 The Primary-Trace Condition

In order to analyze the Primary-Trace Condition for a given set W, we construct a Büchi automaton $A_W = \langle D, Q, q_0^W, \delta_W, F_W \rangle$ that recognizes words $\rho \in D^\omega$ iff they correspond to plays that satisfy the Primary-Trace Condition. Note that since the transition function in a multi-agent satisficing game is deterministic and the initial state $\mathbf{v_0} \in V$ is fixed, plays are uniquely determined by an infinite sequence of decisions, so we take the input alphabet to be D. The other components of A_W are defined as follows: (**1**) $Q = V \times \times_{i \in \Omega} Q^i \times 2^\Omega \times 2^\Omega$ (**2**) $q_0^W = \langle \mathbf{v_0}, q_0^0 \dots q_0^{k-1}, W_C, \overline{W}_S \rangle$ (**3**) For $q = \langle v, q_0 \dots q_{k-1}, S_1, S_2 \rangle$, with $S_1, S_2 \subseteq \Omega$ and d a decision, the transition $\delta^W(q, d)$ is $\langle v', q'_0 \dots q'_{k-1}, S'_1, S'_2 \rangle$, where (**3.i**) $v' = \tau(v, d)$ (**3.ii**) For $i \in \Omega$, we have $q'_i = \delta^i(q_i, \mathcal{W}(v, d)[i])$ (**3.iv**) Finally, we specify how we change S_1 and S_2 to S'_1 and S'_2 by removing states – we always have $S'_i \subseteq S_i$: (**3.iv.a**) If $i \in S_1$ then $i \notin S'_1$ if $q'_i \in F^i$. (**3.iv.b**) If $i \in S_2$, then $i \notin S'_2$ if $q'_i \notin F^i$. (**3.iv.c**) If $i \in W_S$, we specify that the transition is not defined if $q'_i \notin F^i$. Upon attempting such a transition the automaton A_W rejects. (**3.iv.d**) If $i \in \overline{W}_C$, we specify that the transition is not defined if $q'_i \in F^i$. Upon attempting such a transition the automaton A_W rejects. (**4**) F_W is the set of states with $S_1 = S_2 = \emptyset$.

Intuitively, S_1 represents agents in W with co-safety goals that have yet to be satisfied. Once a final state for such a goal has been reached, this goal is satisfied. A dual logic applies to S_2, which represents agents not in W with safety goals. At some point such goals must reach their rejecting state. For co-safety goals for agents not in W, we must make sure they never reach a final state, so we immediately reject if they do. The same holds for safety goals in W, if we reach a rejecting state then A_W must also reject.

Theorem 1 (A_W Correctness). *For a given multi-agent satisficing game* $\mathbb{G}$, *A_W accepts a word $\rho \in D^\omega$ if $R_i(\rho) \, \mathsf{R}_i \, t_i$ holds for precisely the agents $i \in W$.*

All omitted proofs can be found in the technical report [1] [Rajasekaran *et al.*, 2023]

4.2 The j-Deviant-Trace Condition

In order to analyze the j-Deviant-Trace Condition for an agent j we wish to characterize the set of states (v, q^j) and (v, q^j, d) with $v \in V, q^j \in Q^j, d \in D$ from which there

exists some Agent j strategy that leads to a successful deviation. To characterize these states we create a two-agent turn-based game $G_j = \langle V_0, V_1, E, C \rangle$ where (**1**) $V_0 = V \times Q^j$, $V_1 = V \times Q^j \times D$ and (**2**) For $(v, q^j) \in V_0$, we have $\langle (v, q^j), (v, q^j, d) \rangle \in E$ for all $d \in D$. For $d \in D$, let $d[-j]$ represent d with A_j projected out. We have $\langle (v, q^j, d), (v', q^{j'}) \rangle \in E$ iff $\exists d' \in D$ s.t. $d[-j] = d'[-j]$, $\tau(v, d') = v'$, and $\delta^i(q^j, \mathcal{W}(v, d')[j]) = q^{j'}$

The condition C depends on whether Agent j has a cosafety or safety goal. If Agent j has a cosafety goal, C is a safety condition specified by the set $V \times Q^j \setminus F^j \cup V \times Q^j \setminus F^j \times D$. If Agent j has a safety goal, then C is a reachability condition specified in the exact same manner. Intuitively, Agent 0 takes control of the agents who are not j and tries to ensure that no successful deviation from j is possible. Therefore, when Agent j has a cosafety goal Agent 0 tries to prevent Agent 1 (the stand-in for Agent j) from reaching an accepting state. And when Agent j has a safety goal Agent 0 tries to reach the non-accepting state in the corresponding safety automaton A^j.

Theorem 2. *Agent j can only successfully deviate from a profile π from a state in* $\mathsf{Win}_1(G_j)$.

That is to say, for a strategy profile π, if ρ_π has a prefix $(\mathbf{v_0}, d_0) \dots (v_n, d_n)$ and the run of A^j on the corresponding reward sequence of this prefix puts it in a state q^j such that either (v_n, q^j) or (v_n, q^j, d_n) belongs to $\mathsf{Win}_1(G_j)$ then Agent j has a strategy to ensure a successful deviation; otherwise he does not and cannot deviate successfully.

4.3 Characterization of W-NE Existence

With the characterization of deviations given above, we create the automaton $A'_W = \langle D, Q', q_0^W, \delta'_W, F_W \rangle$, defined as A_W with a restricted state space and transition function. Here, Q' consists of states $\langle v, q_0 \dots q_{k-1}, S_1, S_2 \rangle$ such that $\forall j \notin W$ we have $\langle v, q_j \rangle \notin \mathsf{Win}_1(G_j)$. Furthermore, for $q \in Q, d \in D$, $\delta'_W(q, d)$ is now undefined if $\exists j \notin W$ such that $\langle q[0], q[j], d \rangle \in \mathsf{Win}_1(G_j)$ (Note $q[0] \in V$). Our characterization is then given by the following:

Theorem 3. *For a given multi-agent satisficing game $\mathbb{G}$, A'_W is nonempty iff there exists a W-NE strategy profile.*

Therefore, we are able to reduce the problem of W-NE existence to that of non-emptiness in a Büchi automaton. We now analyze the complexity of determining the nonemptiness of A'_W. As mentioned before, if the rewards are specified in unary, then each comparator is linear in the size of the input of a multi-agent satisfying game. A_W, which consists of the cross-products of these comparator automata and 2^Ω is exponential in the size of the input.

In order to determine $\mathsf{Win}_1(G_j)$ for each $j \notin W$, we either construct a reachability or safety game as outlined in Section 4.2. The size of these games is $(|V| \times |Q^j| \times |D|) + (|V| \times |Q^j|)$, which is polynomial in the size of the input. These games can be solved in time linear to the size so this step takes polynomial time. Testing the final A'_W therefore lies in NPSPACE=PSPACE, as Büchi automata can be tested for nonemptiness in NLOGSPACE [Vardi and Wolper, 1994].

Theorem 4. *The problem of deciding whether a W-NE exists in a multi-agent satisficing game is in PSPACE.*

The characterization of Theorem 4 is for a fixed $W \subseteq \Omega$. But we can also find some $W \subseteq \Omega$, or some $W \subseteq \Omega$ of maximal size such that a W-NE exists while remaining in PSPACE. As mentioned in the introduction, one could construct a game consisting of the cross-product of each comparator automaton and the underlying concurrent game in order to apply the algorithms of [Ummels *et al.*, 2015]. This construction is exponential in the size of the input due to the size of the cross-product. Applying the NP algorithm in [Ummels *et al.*, 2015] then yields a NEXPTIME upper bound.

5 Multi-Satisficing Goals

In Section 3 we demonstrated a simple scenario in which the tendency for agents with optimization goals to deviate for negligible payoff improvements disallowed the possibility of mutually beneficial cooperation between agents. Satisficing goals were able to address this problem at the cost of allowing each agent only two possible payoffs. Ideally, we would like a goal type that allows agents more than two types of payoffs but still keeps the possibility of mutual beneficial cooperation that is often lost when using optimization goals. In this section we describe a new type of goal in which each agent is equipped with a single relation but has multiple thresholds and payoffs that increase with the number of thresholds satisfied. This allows us to combine the algorithmic efficiency and possibilities for cooperation achieved by using single-threshold satisficing goals with the expressive payoff structure of optimization-type goals.

To this end, we consider *multi-threshold satisficing goals* (henceforth, multi-satisficing goals), which are once again specified by a pair $\langle R_i, T_i \rangle$. The difference is that instead of a single threshold t_i, Agent i has a monotone sequence T_i of thresholds. In order to accommodate these multiple thresholds, we only consider relations $R_i \in \{<, >, \leq, \geq\}$. On a play ρ, Agent i with a multi-satisficing goal receives a *payoff* (which is distinct from the cumulative reward) equal to the number of thresholds in T_i that are R_i-satisfied on ρ. Since the payoff structure is based on the number of thresholds met and not the actual reward assigned on a play, multi-satisficing agents always have a preference for a higher payoff over a lower one regardless of their relation. For example, for an agent with relation $>$ and threshold sequence $\{5, 10, 15\}$ a payoff of 0 would be given for plays with cumulative reward ≤ 5, representing no thresholds met. If the threshold 5 was met but not 10 this would be a payoff of 1, all the way to her maximal preference of a payoff of 3 for plays with cumulative reward greater than 15. Even an agent with the $<$ relation would still seek to maximize the number of thresholds met; an agent with relation $<$ and thresholds $\{1, 2, 3\}$ would receive a payoff of 3 on a play with cumulative reward of 0 but a payoff of 0 on a play with a cumulative reward of 4.

We once again consider W-NEs but a modification is needed. Earlier, W specified which agents had their goal satisfied. This can be seen as a list of agents that receive a payoff of 1 as opposed to a payoff of 0. In the multi-threshold framework, W corresponds to a payoff assignment to each

agent. So, if each agent had three thresholds, there would be four possible payoffs for each agent and therefore 4^k possible W's. As an example, for an agent with threshold sequence $\{5, 10, 15\}$, if W assigns a payoff of 2, then only strategy profiles that yielded this agent a cumulative reward of more than 10 but less than or equal to 15 are acceptable. While more thresholds lead to more possibilities for W, it is often the case that a system planner has some intention towards the outcome and therefore only needs to consider W that correspond to these desired outcomes. Furthermore, the different W's can still be enumerated in polynomial space.

We now construct an automaton A_W that recognizes plays that assign the payoffs stipulated by W. In Section 4 we showed how to construct a Büchi automaton given a set of co-safety and safety automata and a specification of which automata should accept on infinite words (considering prefix satisfaction) accepted by the Büchi automaton. We can apply this same construction here in multiple ways by considering comparator automata that model the cumulative reward intervals corresponding to the payoffs assigned to each agent by a given W. For example, if we have an agent with relation $>$ and thresholds 1 and 2 that receives a payoff of 1 by a given W, we can construct the Büchi automaton as in Section 4 with a co-safety comparator automaton for > 1 that we specify to accept along with a safety comparator automaton for ≤ 2 that we also specify to accept – here we are modeling payoff specification with two comparators as opposed to one. If an agent has relation $\leq$, thresholds of 0,1, and 2 and receives a payoff of 0 then we can model this with a co-safety comparator automaton for ≤ 2 that we specify not to accept *or* a co-safety comparator for > 2 that we specify to accept. Since there is a choice in how to represent a given cumulative reward interval with comparators, automata may be re-used, an important direction for future work concerning implementations.

Now, we consider the deviation games constructed in Section 4.2 to analyze deviations. In order to do so, we consider the threshold representing the next highest payoff for each agent. To use our running example, if our agent with relation $>$ and thresholds $\{5, 10, 15\}$ was assigned a payoff of 1 by W, then a play with a cumulative reward greater than 10 represents a strictly preferable play to the current play. Therefore, we can construct a safety game (since she has a co-safety relation $>$) w.r.t the comparator that recognizes the set of plays with cumulative reward greater than 10. This lets us reapply the construction from Section 4.2 in order to compute $\mathsf{Win}_1(G_j)$ – which once again represents the set of states from which profitable deviation is possible.

We can now construct the automaton A'_W by restricting A_W with respect to $\mathsf{Win}_1(G_j)$ as described in Section 4.3. Note that, as before, we consider games for *every* agent not achieving their maximal payoff, as only agents achieving their maximal payoff do not consider deviations.

Theorem 5. *For a given multi-agent multi-satisficing game $\mathbb{G}$, we have that A'_W is nonempty iff there exists a W-NE strategy profile.*

We are now able to reduce the problem of W-NE existence to that of non-emptiness of Büchi automata for multi-satisficing games, while retaining the same complexity – A_W

is still exponential w.r.t the input and each deviation game is still polynomial in size, so this algorithm runs in PSPACE.

Theorem 6. *The problem of deciding whether a W-NE exists in a multi-agent multi-satisficing game is in PSPACE.*

This result can also be extended, as before, to problems such as the existence of *some* Nash equilibria in a multi-satisficing game by repeatedly running the algorithm for every possible W. Since the algorithm for deciding a single W is in PSPACE, all W can be checked in PSPACE as well.

6 Relationship to ϵ-Equilibria

The notion of an ϵ-equilibria modifies the standard notion of the Nash equilibria. In a Nash equilibrium strategy profile, no agent could deviate to obtain a strictly preferable payoff - an ϵ-equilibrium adds an extra condition to "preferable". Let $\mathbb{G}$ be a discounted-sum game. A strategy profile $\sigma = \langle \sigma_0 \ldots \sigma_i \ldots \sigma_{m-1} \rangle$ is an ϵ-*equilibrium*, for some fixed constant ϵ, in $\mathbb{G}$ if there is no agent $i \in \Omega$ with a strategy $\sigma_i' \neq \sigma_i$ such that $R_i(\rho_\sigma) + \epsilon \leq R_i(\rho_{\langle \sigma_0 \ldots \sigma_i' \ldots \sigma_{m-1} \rangle})$.

Since deviations must generate at least ϵ more cumulative reward in an ϵ-equilibrium, this solution concept also addresses the problems of minuscule deviations raised in Section 3. This is one of the reasons ϵ-equilibria have become an extremely popular concept in algorithmic game theory [Roughgarden, 2010; Daskalakis *et al.*, 2006]. In this general concurrent game setting solving for ϵ-equilibria entails adding in extra assumptions, see [Gupta, 2016; Ummels, 2011; Chatterjee *et al.*, 2004] for a few examples.

In this section we develop a one-way correspondence between ϵ-equilibria in discounted-sum games and the W-NE in the multi-satisficing games of section 5. Specifically, if we are given a discounted-sum game $\mathbb{G}$ and a fixed constant $\epsilon > 0$, we show how to construct a family of multi-satisficing games of the form $\langle \mathbb{G}, \alpha \rangle$ (representing $\mathbb{G}$ plus appropriately chosen multi-satisficing goals α, so each $\langle \mathbb{G}, \alpha \rangle$ is a multi-satisficing game) such that the existence of some W-NE in some $\langle \mathbb{G}, \alpha \rangle$ corresponds to an ϵ-equilibria in $\mathbb{G}$.

Here, we explore sufficient restrictions on the multi-satisficing goals α such that the correspondence holds. For an agent i in a discounted-sum game, let l_i (g_i) be a lower (upper) bound on the minimal (maximal) possible cumulative reward available to the agent over all possible plays.

Theorem 7 (ϵ-Equilibria Correspondence). *Let $\mathbb{G}$ be a discounted-sum game with m agents and $\epsilon > 0$ a fixed constant. Let $\alpha = [\alpha_0, \alpha_1 \ldots \alpha_{m-1}]$ be a multi-satisficing goal such that each α_i (which has smallest threshold a_i and largest threshold z_i) satisfies the following properties: (1) α_i has the $\geq$ relation (2) $a_i \leq l_i$ (3) $z_i \geq g_i$ (4) For each pair of consecutive thresholds t_j, t_{j+1} in α_i we have that $t_{j+1} - t_j \leq \epsilon$. Let $\sigma = \langle \sigma_0 \ldots \sigma_i \ldots \sigma_{m-1} \rangle$ be a W-NE for some W in $\langle \mathbb{G}, \alpha \rangle$. Then, σ is also an ϵ-equilibrium in $\mathbb{G}$.*

In the theorem statement we use the $\geq$ relation since it most naturally corresponds to the definition of the ϵ-equilibrium in the literature. Although it is not standard, we can define a similar minimizing condition – i.e. agents with the minimizing condition are not incentivized to deviate unless they receive ϵ *less* cumulative reward, corresponding to agents with

the $\leq$ relation in the multi-satisficing setting. This is a natural extension that lets us consider "modified" ϵ-equilibria in which agents seek to minimize total cumulative reward or a mixed game with both maximizing and minimizing agents.

The bounds l_i and g_i are critical. Without the bounds in place, it may be possible for an agent to deviate to gain more than ϵ additional cumulative reward but not satisfy additional thresholds. For example, an agent i may receive a cumulative reward of r and be able to deviate to a play to receive $r + \epsilon$ cumulative reward, but if both of those values are lower than agent i's lowest threshold then the agent has no incentive to deviate in the corresponding multi-satisficing game, breaking the correspondence between the two solution concepts.

For this reason, we demonstrate one simple way to compute sufficient values of l_i and g_i. For each agent i, let b_i be the maximum possible reward available on all edges for i (for biggest) and s_i be the minimum possible reward over all edges (for smallest). Then, $g_i = \sum_{j=0}^{\infty} b_i \cdot \frac{1}{\gamma^j}$ is an upper bound for the cumulative reward for agent i (representing receiving the maximum reward at every time step) and similarly $l_i = \sum_{j=0}^{\infty} s_i \cdot \frac{1}{\gamma^j}$ is a lower bound.

Theorem 7 shows a general but only one-way correspondence between the notions of W-NE and ϵ-equilibria. A reverse correspondence, which we leave for future work, would be a completely characterize ϵ-equilibria purely in terms of W-NE and would therefore be an extremely powerful result as it would allow for our efficient automata-based algorithms to find ϵ-equilibria in an extremely general game setting.

7 Conclusion

In this paper, we have argued for the use of satisficing goals when analyzing Nash equilibria in multi-agent concurrent discounted-sum games. There are three main advantages to consider in favor of satisficing goals.

First, satisficing goals allow for the existence of mutually beneficial equilibria that may not exist when considering optimization goals, as discussed in Section 3. Agents with optimization goals are incentivized to deviate for negligible gain, but satisficing goals address this problem. Second, satisficing goals allow for the use of efficient automata-based techniques. The use of these techniques allowed us to demonstrate a PSPACE upper bound. Furthermore, automata-based techniques allow for the use of heuristics that have been shown to be highly efficient in practice, such as BDD-based encodings [Burch *et al.*, 1992]. Finally, we have shown how multi-satisficing goals allow us to consider a richer payoff structure closer to that of an optimizing agent's while still addressing the negligible deviation problem and allowing for automata-based techniques.

There are several directions available for future work; we mention a few of particular interest here. As mentioned before, it is natural to consider a PSPACE-lower bound to match our upper bound presented, noting that PSPACE-hardness is a reasonable expectation for a multi-agent problem. Furthermore, we wish to continue exploring the relationship between W-NEs and ϵ-equilibria and consider implementations.

Acknowledgements

References

[Bansal and Vardi, 2019] S. Bansal and M. Y. Vardi. Safety and co-safety comparator automata for discounted-sum inclusion. In *Proc. of CAV*, 2019.

[Bansal et al., 2018] S. Bansal, S. Chaudhuri, and M. Y. Vardi. Comparator automata in quantitative verification. In *Proc. of FoSSaCS*, 2018.

[Bansal et al., 2021] Suguman Bansal, Krishnendu Chatterjee, and Moshe Y. Vardi. On satisficing of quantitative games. In *Proc. of TACAS*, 2021.

[Bansal et al., 2022a] Suguman Bansal, Swarat Chaudhuri, and Moshe Y Vardi. Comparator automata in quantitative verification. *Logical Methods in Computer Science*, 18, 2022.

[Bansal et al., 2022b] Suguman Bansal, Lydia Kavraki, Moshe Y Vardi, and Andrew Wells. Synthesis from satisficing and temporal goals. In *Proc. of AAAI*, 2022.

[Bouyer et al., 2010] Patricia Bouyer, Romain Brenguier, and Nicolas Markey. Nash equilibria for reachability objectives in multi-player timed games. In *International Conference on Concurrency Theory*, pages 192–206. Springer, 2010.

[Bouyer et al., 2011] Patricia Bouyer, Romain Brenguier, Nicolas Markey, and Michael Ummels. Nash equilibria in concurrent games with büchi objectives. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2011)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2011.

[Bouyer et al., 2015] Patricia Bouyer, Romain Brenguier, Nicolas Markey, and Michael Ummels. Pure nash equilibria in concurrent deterministic games. *Log. Methods Comput. Sci.*, 11(2), 2015.

[Bouyer et al., 2019] Patricia Bouyer, Orna Kupferman, Nicolas Markey, Bastien Maubert, Aniello Murano, and Giuseppe Perelli. Reasoning about quality and fuzziness of strategic behaviours. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, 2019.

[Bulling and Goranko, 2022] Nils Bulling and Valentin Goranko. Combining quantitative and qualitative reasoning in concurrent multi-player games. *Autonomous Agents and Multi-Agent Systems*, 2022.

[Burch et al., 1992] J.R. Burch, E.M. Clarke, K.L. McMillan, D.L. Dill, and L.J. Hwang. Symbolic model checking: 10^{20} states and beyond. *Information and Computation*, 98(2):142–170, 1992.

[Chatterjee et al., 2004] Krishnendu Chatterjee, Rupak Majumdar, and Marcin Jurdzinski. On nash equilibria in stochastic games. pages 26–40, 09 2004.

[Daskalakis et al., 2006] Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a nash equilibrium. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '06, page 71–78, New York, NY, USA, 2006. Association for Computing Machinery.

[Grädel et al., 2002] E. Grädel, W. Thomas, and T. Wilke. *Automata, Logics, and Infinite Games: A Guide to Current Research*. Lecture Notes in Computer Science 2500. Springer, 2002.

[Gupta, 2016] Anshul Gupta. *Equilibria in finite games*. PhD thesis, University of Liverpool, UK, 2016.

[Hopcroft and Ullman, 1979] John E. Hopcroft and Jeff D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley Publishing Company, 1979.

[Kozen, 1977] Dexter Kozen. Lower bounds for natural proof systems. In *18th Annual Symposium on Foundations of Computer Science, Providence, Rhode Island, USA, 31 October - 1 November 1977*, pages 254–266. IEEE Computer Society, 1977.

[Kupferman and Vardi, 2000] Orna Kupferman and Moshe Y. Vardi. An automata-theoretic approach to modular model checking. *ACM Trans. Program. Lang. Syst.*, 22(1):87–128, jan 2000.

[Kwiatkowska, 2007] Marta Kwiatkowska. Quantitative verification: models techniques and tools. In *Proc. 6th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on Foundations of Software Engineering*, pages 449–458, 2007.

[Lavalle, 2006] Steven M. Lavalle. *Planning Algorithms*. Cambridge University Press, 2006.

[Martin, 1975] Donald A. Martin. Borel determinacy. *Annals of Mathematics*, 102(2):363–371, 1975.

[McNaughton, 1993] Robert McNaughton. Infinite games played on finite graphs. *Ann. Pure Appl. Logic*, 65(2):149–184, 1993.

[Mogavero et al., 2014] Fabio Mogavero, Aniello Murano, Giuseppe Perelli, and Moshe Y Vardi. Reasoning about strategies: On the model-checking problem. *ACM Trans. on Computational Logic*, 15(4):1–47, 2014.

[Nash, 1950] John F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.

[Osborne and Rubinstein, 1994] Martin J. Osborne and Ariel Rubinstein. *A course in game theory*. Cambridge, USA, 1994.

[Rajasekaran and Vardi, 2021] Senthil Rajasekaran and Moshe Y. Vardi. Nash equilibria in finite-horizon multiagent concurrent games. In *Proc. 20th Int'l Conf. on Autonomous Agents and MultiAgent Systems*, AAMAS '21, page 1046–1054, Richland, SC, 2021.

[Rajasekaran and Vardi, 2022] Senthil Rajasekaran and Moshe Y. Vardi. Verification and realizability in finite-horizon multiagent systems. In Gabriele Kern-Isberner,

Gerhard Lakemeyer, and Thomas Meyer, editors, *Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning*, Haifa, Israel, 2022.

[Rajasekaran *et al.*, 2023] Senthil Rajasekaran, Suguman Bansal, and Moshe Y. Vardi. Multi-agent systems with quantitative satisficing goals, 2023.

[Roughgarden, 2010] Tim Roughgarden. Algorithmic game theory. *Communications of the ACM*, 53(7):78–86, 2010.

[Shoham and Leyton-Brown, 2009] Yoav Shoham and Kevin Leyton-Brown. *Multiagent Systems - Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2009.

[Simon, 1956] Herbert A Simon. Rational choice and the structure of the environment. *Psychological review*, 63(2):129, 1956.

[Sipser, 2006] Michael Sipser. *Introduction to the Theory of Computation*. Course Technology, second edition, 2006.

[Sutton and Barto, 2018] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, second edition, 2018.

[Ummels *et al.*, 2015] Michael Ummels, Nicolas Markey, Romain Brenguier, and Patricia Bouyer. Pure nash equilibria in concurrent deterministic games. *Logical Methods in Computer Science*, 11, 2015.

[Ummels, 2011] Michael Ummels. *Stochastic multiplayer games: theory and algorithms*. PhD thesis, RWTH Aachen University, 2011.

[Van der Hoek and Wooldridge, 2008] Wiebe Van der Hoek and Michael Wooldridge. Multi-agent systems. *Foundations of Artificial Intelligence*, 3:887–928, 2008.

[Vardi and Wolper, 1994] M.Y. Vardi and P. Wolper. Reasoning about infinite computations. *Information and Computation*, 115(1):1–37, 1994.

[Wooldridge, 2009] Michael Wooldridge. *An introduction to multiagent systems*. John wiley & sons, 2009.