A Gymnasium-Compatible Reinforcement Learning Environment for Monopoly Board Game

Lathitha Nongauza

***Abstract*—Monopoly is a stochastic, multi-agent board game that captures several core aspects of real-world economic decision-making, including negotiation, asset valuation, and long-term strategic planning. These properties make it a challenging and relatively underexplored domain for reinforcement learning research. This paper presents the design and implementation of a Gymnasium-compatible reinforcement learning environment for Monopoly, with an emphasis on economically meaningful interactions such as trading, auctions, and property development. The environment is formulated as a Markov Decision Process with carefully designed observation and action spaces that prioritise financial reasoning over low-level movement decisions. State representations incorporate global board information, detailed player asset portfolios, recent turn dynamics, and dynamic action masking to ensure legal action selection. Complex interactions, including auctions and trades, are abstracted to maintain tractability while preserving strategic depth. Proximal Policy Optimisation (PPO) is used as a validation mechanism to confirm the technical suitability of the environment for reinforcement learning. Rather than serving as a benchmark for algorithmic performance, the proposed environment is intended as a flexible platform for exploring alternative learning paradigms in economically rich, multi-agent settings.**

***Index Terms*—Reinforcement Learning, Gymnasium, Monopoly, Multi-Agent Systems, Economic Decision-Making**

I. Introduction

Monopoly is a widely recognised board game that provides an abstract yet expressive representation of economic interactions such as investment, negotiation, liquidity management, and risk assessment. Its structured ruleset, stochastic dynamics, and long-term strategic dependencies make it a compelling testbed for artificial intelligence research, particularly in the context of reinforcement learning (RL).

Despite its apparent suitability, Monopoly has received comparatively limited attention in reinforcement learning literature, largely due to its multi-agent nature, delayed reward structure, and complex interaction dynamics. This paper presents the design and implementation of a custom reinforcement learning environment for Monopoly using the Gymnasium framework. The primary contribution of this work lies not in algorithmic performance evaluation, but in the formulation of the environment itself, including state representations, action abstractions, and transition dynamics.

The proposed environment is designed to support future experimentation with a wide range of learning paradigms by providing a flexible and economically grounded simulation platform. Particular attention is given to strategic decision-making processes such as trading, auctions, and property development, which are central to the game's economic structure.

II. Environment Design

A. Monopoly Game Implementation

The Monopoly board game is implemented in accordance with the official rules, using the South African edition of the board. The game is inherently stochastic, with player movement determined by dice rolls and card draws. However, the primary objective of this environment is not to optimise movement decisions, which are largely outside the agent's control, but rather to enable learning of strategic economic behaviours.

Specifically, the environment emphasises player-to-player interactions such as property trading and auctions. In these scenarios, agents exercise control over trade structures, pricing decisions, and bidding strategies, requiring valuation, negotiation, and long-term planning. By prioritising these interactions, the environment encourages learning in economically meaningful contexts.

The game implementation includes:

- A complete Monopoly board with properties, railways, utilities, and special squares.
- Enforcement of official rules governing ownership, rent calculation, property development, and bankruptcy.
- Explicit modelling of stochastic elements such as dice rolls and card draws.
- Trading and auction mechanisms that allow agents to propose, evaluate, and execute transactions.

B. Conversion to a Gymnasium Environment

To enable reinforcement learning, the Monopoly game is formulated as a Gymnasium-compatible environment by defining a Markov Decision Process (MDP) with explicit state, action, and transition specifications. The environment inherits from the `gymnasium.Env` interface and implements the required `reset`, `step`, and `render` methods.

Each episode corresponds to a complete game of Monopoly and terminates upon player bankruptcy, victory, or the attainment of a predefined maximum number of turns.

C. Observation Space

The observation space is represented as a fixed-length, continuous vector composed of four semantically distinct components:

$$s_t = \left[s_t^{\text{board}} \;\middle|\; s_t^{\text{player}} \;\middle|\; s_t^{\text{turn}} \;\middle|\; m_t \right], \tag{1}$$

where all components are normalised to ensure numerical stability during learning.

1) Board Representation: The board state encodes global information about all properties, railways, and utilities, including ownership status, mortgage state, and normalised market prices. Binary indicators are used for ownership and mortgage status, enabling the agent to reason about global economic conditions while remaining invariant to absolute currency scale.

2) Player Representation: The player-specific component captures both financial and positional information, including:

- Normalised cash balance
- Normalised board position
- Jail status and possession of a "Get Out of Jail Free" card
- Per-tile indicators for ownership, mortgage status, number of houses, and hotel presence

This representation enables the agent to reason about asset portfolios, development opportunities, and liquidity constraints.

3) Turn Information: To preserve limited temporal context, the observation includes recent turn information such as the previous dice roll, previous position, current position, and current capital. These features allow the agent to condition decisions on recent stochastic outcomes without violating the Markov assumption.

4) Action Masking: An action mask is appended to the observation vector to indicate which actions are legally available in the current state. This dynamic masking prevents the selection of invalid actions and significantly reduces the effective exploration space.

D. Action Space

The action space is discrete and consists of eleven high-level actions, as shown in Table I.

TABLE I
ACTION SPACE DEFINITION

Index	Action Description
0	Buy asset
1	Initiate trade
2	Start auction
3	Build house
4	Build hotel
5	Sell house
6	Sell hotel
7	Mortgage asset
8	Unmortgage asset
9	Pay to leave jail
10	Stay in jail

Low-level actions such as dice rolling are omitted, as they are stochastic and not subject to agent control. The agent instead focuses on economically meaningful decisions involving investment, negotiation, and risk management.

E. Design Rationale

Several considerations guided the environment design:

- **Stochastic isolation**: Random events are handled internally to reduce learning variance.
- **Action abstraction**: Auctions and trades are resolved using heuristics to maintain tractability.
- **Action masking**: Invalid actions are dynamically disabled to stabilise learning.
- **Economic focus**: The environment prioritises financial reasoning over movement optimisation.

A 90-second turn rule, extending the official game rules, is enforced during human gameplay to prevent deadlocks in auctions and trades, but is disabled during training.

III. Discussion and Limitations

Proximal Policy Optimisation (PPO) is employed primarily as a sanity check to verify that the environment is compatible with standard reinforcement learning algorithms. The objective is not to achieve optimal performance, but to confirm that learning signals are well-defined and that training proceeds without instability.

A key limitation of this work is the difficulty of disentangling environment complexity from algorithmic inadequacy. Poor agent performance does not necessarily indicate flaws in the environment; rather, it may reflect the inherent difficulty of the task. This ambiguity is well-documented in challenging benchmark environments such as *Montezuma's Revenge*, where sparse rewards and long-term credit assignment hinder learning across many algorithms. Monopoly exhibits similar characteristics, including delayed rewards, high stochasticity, and complex multi-agent interactions.

IV. Future Work

Several promising research directions emerge from this work. A natural extension is a comparative evaluation of multiple reinforcement learning algorithms across a suite of economically motivated environments. Beyond conventional RL methods, transformer-based architectures offer a compelling alternative. Monopoly is inherently sequential, with each decision influenced by a history of player actions. Encoding full turn sequences using temporal embeddings may allow agents to model opponent strategies, negotiation patterns, and long-term behavioural tendencies more effectively.

V. Conclusion

This paper presented a Gymnasium-compatible reinforcement learning environment for the Monopoly board game, with a focus on economically meaningful decision-making processes such as trading, auctions, and asset management. The environment preserves the strategic depth and stochastic nature of the original game while remaining tractable for reinforcement learning research. PPO-based validation demonstrates technical compatibility, though learning performance remains constrained by the inherent complexity of the task. Rather than serving as a benchmark, the environment is intended as a flexible platform for exploring alternative representations, learning objectives, and architectures in complex, multi-agent economic domains.