Simulation to Strategy: Investigating P-LEO Satellite Internet Market Dynamics via Gaming and Reinforcement Learning

by

Rehman S. Qureshi

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 2, 2026

Keywords: P-LEO satellite constellations, reinforcement learning, gaming

Approved by

Davide Guzzetti, Chair, Associate Professor of Aerospace Engineering
Ehsan Taheri, Associate Professor of Aerospace Engineering
Akhil Rao, Visiting Research Scholar at Middlebury College
Samuel Mulder, Associate Professor of Computer Science and Software Engineering
Daniel Tauritz, COLSA Corporation Cyber Security and Information Assurance Endowed Professor of Computer Science and Software Engineering

Abstract

An influx of private actors and state-owned agencies are entering the space industry to capitalize on space-based infrastructure and resources. Specifically, low Earth orbit (LEO) is seen as an untapped satellite communications (SATCOM) resource. With the deployment of SpaceX's Starlink, the concept of a proliferated low Earth orbit (P-LEO) satellite constellation for internet services has become a reality. These networks of thousands of satellites coordinate to deliver internet access directly to millions of user terminals. Despite mixed results among P-LEO operators, the joint market and orbital dynamics governing operational and orbital sustainability remain largely unexplored. This dissertation investigates these dynamics by modeling the P-LEO SATCOM market as a dynamical system to be explored.

First, using *Satellite Tycoon*, a tabletop board game developed as a multi-player simulation environment, a randomized controlled trial (RCT) is conducted to study how human participants develop constellation management strategies in response to economic and policy instruments. Results reveal that while players' revenue efficiency improves across repeated play, the tested policy treatments had a limited effect on satellite overproduction and debris generation, though derelict satellite debris showed a measurable reduction under the treatment condition. Next, a parameterized, single-agent reinforcement learning (RL) environment was constructed to model the specific dynamics and effects experienced by a single P-LEO constellation operator. An orbit plane catalog was created to give the agent a wide variety of constellation design choices, and RL agents were trained across a range of environmental configurations to explore which environment parameters most influence P-LEO constellation feasibility. Finally, the single-agent framework was extended into a multi-agent environment in which multiple RL agents act as constellation operators interacting and competing with one another for a limited market share. The utility function assigning customers to the single agent was replaced with a more sophisticated, reverse-bidding multi-attribute decision-making (MADM) model to more accurately model the zero-sum nature of customer acquisitions in SATCOM.

Artificial Intelligence (AI) Use Disclosure Statement

In the preparation of this dissertation, the following Artificial Intelligence (AI) tools were used: ChatGPT and Claude Sonnet 4.6. These tools were used primarily to provide clarity and suggestions on sentence structure improvements. The author acknowledges full responsibility for the intellectual content of this work and has ensured that all AI-assisted sections have been reviewed and revised for accuracy and appropriate academic style. All AI-generated content was reviewed and validated for relevance, appropriateness, and accuracy before incorporation into the final document to maintain scholarly integrity of this research.

Digital Accessibility Use Disclosure Statement

In the preparation of this dissertation, the following digital accessibility tools were used to ensure this document complies with federal requirements: Adobe Acrobat. The author acknowledges full responsibility for the intellectual content of this work and has made a good faith effort to comply with digital accessibility requirements in publishing, wherein the nature of the content does not significantly change in order to do so. Furthermore, all content has been reviewed and revised to meet these requirements prior to final publication.

Acknowledgments

This dissertation would not exist without the generosity, encouragement, support, and camaraderie of so many people, and I am deeply grateful to each of them. The common saying is, ”it takes a village”, and Auburn has truly been the Loveliest Village on the Plains.

First and foremost, I owe an immeasurable debt of gratitude to my advisor, Dr. Davide Guzzetti. I entered the 3i Space Dynamics Lab at the peak of COVID, during a very uncertain time for us all. You took a chance on me, accepted me into the lab, gave me numerous opportunities for growth and success, and grew me into the researcher that I am today. The level of impact an advisor has on a graduate experience was completely lost on me when I entered graduate school, so I consider myself extremely lucky to have had an advisor as caring, patient, and thoughtful as yourself. From our very first conversation, you set a standard for what it means to pursue research with rigor, creativity, and purpose. Our conversations since then have been both amusing, philosophical, and indicative of your passion. This work is a reflection of your mentorship, and I am profoundly thankful to carry that forward into the next chapter of my research career, a career I owe entirely to you.

I am equally grateful to the members of my doctoral committee, whose guidance shaped the direction and quality of this research in ways I could not have anticipated. Your thoughtful questions, candid feedback, patience with my close submissions, and generosity with your time made this dissertation stronger at every turn. Dr. Ehsan Taheri, I have never met anyone as passionate about optimal control and your class was one of my favorite at Auburn, in both quality and rigor. Dr. Akhil Rao, your mentorship, advice, and encouragement have always come at times when I most needed to hear them. Dr. Samuel Mulder, I will most certainly miss the difficult questions you ask me; you’ve constantly pushed me to be a better researcher, ask the deeper questions, and figure out how to break a well tuned board game. Dr. Daniel R. Tauritz, whenever anyone asks me what a good scientist, engineer, or researcher looks like, I always refer them to you and your quizzical natural. Additionally, I would like to thank my university reader,

Dr. Daniel Silva. Dr. Silva, your class on dynamic programming led me towards Q-learning all the way back in 2017, when DQN was just starting to show promise. That sparked my life-long interest in machine learning and I owe that to you.

To my colleagues on the Satellite Tycoon development team (Cody, Jay, Robert, Emily, Lucy, and everyone else who was part of that effort): thank you. The hours we spent building, debugging, arguing, and laughing together were some of the most energizing of my entire graduate experience. You made the hard days feel worthwhile and the good days feel like genuine milestones. I will and already do miss you guys. To all the individuals who participated in the board game study — thank you for giving your time and engaging so thoughtfully with the experience. This research rests on your willingness to participate, and I do not take that lightly.

To my labmates in Davis 153, Joe, Nick, Tithe, Ana, Keziban, Jack, Praveen, and the rest: you are my Auburn Family and I cannot overstate how much your presence meant to me throughout this journey. Thank you for letting me come into the lab, disrupting everyone's work, and having for some truly interesting and comical conversations. I hope my path continues to cross everyone of you as we all go forth into our new research careers. Specifically to Joe: these last three years could not have been possible without you. When I found out that you were joining our lab, I remembering excitedly telling Manuel over cannolis that the lab got another all-star like him. From letting me randomly crash on your couch to offering heartfelt advice right before my qualifying exam, I could not have done this PhD without your support.

To the broader Aerospace Engineering department at Auburn — faculty, staff, and fellow graduate students — thank you for creating an environment where intellectual curiosity is taken seriously and supported with genuine care. You are some of my deepest friends and I has been a privilege to be among you. It is a community that I am proud to be a part of and hope to give back to throughout my life.

To Dr. Scott Erwin and the team at AFRL/RVSW, as well as Loren Anderson and everyone involved in AstroCraft: thank you for setting me on the path to work at the intersection of astrodynamics and reinforcement learning. Your enthusiasm and your willingness to engage with me on this project has put me on a path towards a life-long career in service of our national

space needs. Your curiosity, work ethic, and service at AFRL sets a standard for which I aspire to meet one day.

To Asher Sinclair, Dan Carpenter, Dr. Scott Sievert, Dr. Matthew Klawonn, my colleagues at AFRL/RISA, and the entire SMART program: you have given me a golden opportunity to develop a meaningful research career, and serve my nation: I hope that I can live up to the high standard of excellence you have of me and that i make the most of this opportunity. Your acceptance and support of me throughout my PhD is heartwarming, and I hope that my research contributions far outweigh the investment that you have made in my career.

To my parents, nana phopho, and chotu: you are, and have always been, the foundation of everything I do. Your love is not something I earned or could ever repay — it is simply the ground beneath my feet. I hope this work, in some small way, reflects the values you instilled in me. Now I will actually admit (for once): you were right and graduate school was totally worth it! Inshallah, I continue to make you proud in all that I accomplish. I love you all and hope we can celebrate with our first trip to Manhattan together!

To Nazoo: you have been my rock, my emotional support throughout this PhD, and my biggest cheerleader. Your calming presence during my most stressful graduate milestones got me through so much of my journey. This is only the beginning of the story I want us to have together, and I hope you are proud of me.

Finally, to my fra (bro), my homie, and the best friend I made in graduate school: Manuel, thought I forgot about you? You have been a constant source of grounding and humor across every time zone and every stretch of my PhD. Going for walks to Panera, having Life & Cannolis discussions, playing late-night delirious paper basketball games in the lab, and swapping quips about pop culture kept me sane these 6 years. Your friendship is something I cherish just as much as this dissertation and my research career. From attending our first conference together to running a half-marathon, you have been with me through it all and set the bar for excellence at a point I can only hope to work towards. Thank you for always being there, no matter the distance or time.

As an Alabama native and previous graduate of Auburn, I have always felt very appreciative and protective of the university and our traditions. After all, It's my home. It's been home for

almost 10 years (in aggregate) and help me become the adult I am today. Leaving the first time was difficult, but I hope this time is less emotionally challenging and that I can go out into the world to live the Creed. War Eagle!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

EOL end-of-life

FCC Federal Communications Commission

GNC Guidance, Navigation, and Control

ISP Internet Service Provider

LEO Low Earth Orbit

MADM Multi-Attribute Decision-Making

MARL Multi-Agent Reinforcement Learning

MDP Markov decision process

P-LEO Proliferated Low Earth Orbit

POSG Partially Observable Stochastic Game

RL Reinforcement Learning

RSO Resident Space Object(s)

SATCOM Satellite Communications

SSA Space situational awareness

TOPSIS Technique for Order of Preference by Similarity to Ideal Solution

Chapter 1

Introduction

Low Earth orbit (LEO) has seen a significant surge in activity related to Earth observation, satellite communications, and national security missions. Primarily, this increase has been led by private space companies, such as SpaceX, leveraging reusable launch vehicles and secondary payload ride-share programs. Specifically, the rapid expansion of proliferated low Earth orbit (P-LEO) satellite constellations is revolutionizing the satellite communications (SATCOM) industry. Such P-LEO satellite constellations are designed to deliver global connectivity and high-bandwidth internet services by operating thousands of (relatively) cheap satellites in low orbit altitudes. While operating these "mega-constellations" in such low altitude regimes decreases satellite-to-ground latency and can potentially provide broadband internet services to remote regions, lower satellite access areas, and higher likelihoods of conjunctions produce new challenges for operational efficiency and long-term space sustainability.

1.1 Motivation & Research Context

While SpaceX has aggressively deployed thousands of satellites as part of its Starlink and Starshield constellations [1] (becoming a dominant "first-mover" in the industry [2]), networks such as Amazon's Kuiper [3] and EutelSat's OneWeb [4] are also in development. Additionally, multiple Chinese P-LEO constellations such as Xingwang, Qianfan [5], and Guangwang [6] are all in various stages of the constellation development cycle. While these P-LEO constellations promise to capture untapped markets and provide access to underserved communities, they also expand the SATCOM market complexity and introduce emergent dynamics not previously

seen in the space sector. Space sustainability, market competition, impacts to national security infrastructure, and long-term business strategies are all factors that P-LEO SATCOM operators must weigh when evaluating mission design and constellation profitability.

Space sustainability has often referred to the proper, responsible, and fair usage of orbital capacity and space; however, in recent years, this term has also come to represent the financial sustainability of companies operating space-based assets. The primary driver of space sustainability in LEO is the mitigation and remediation of space debris. Operators can mitigate space debris generation by de-orbiting their derelict satellite to safely burn up in the atmosphere [7]. However, debris generation by increased launch activity poses a significant risk to ongoing space missions. When developing initial constellation designs, and making subsequent strategic business decisions, SATCOM operators must consider both the immediate impact of their constellations and the long-term implications for the overall LEO regime. Since LEO has been a popular orbital regime for satellites and space stations, the amount of active and derelict resident space objects (RSOs) has grown (and continues to grow) at an alarming pace, primarily due to P-LEO satellite constellation development. Failing to account for sustainable behavior in LEO could lead to a Kessler Syndrome [8] in which cascading collisions render LEO unusable by anyone. This scenario would be an example of a *Tragedy of the Commons*, a scenario in which a resource (LEO space in our case) becomes restricted to all competitors due to self-serving growth strategies [9].

While the SATCOM industry has historically been stagnant, P-LEO entrants are rapidly evolving and advancing market competition. Modern P-LEO SATCOMs offer broadband internet services to a wider array of customers by operating across many more competitive frequency bands, having shorter latencies, and providing greater swathes of total coverage [10], as compared to traditional SATCOM [11]. This allows P-LEO operators to offer broadband internet services to a wider customer base in the hopes of reaching profitability with economies of scale; however, this strategy also draws competition with traditional internet service providers (ISPs) and telecommunications incumbents, such as AT&T and Spectrum. As most of the terrestrial telecommunications industry is a regional oligopoly and the SATCOM market has been implied to be a monopoly [12], it remains uncertain how many satellite internet providers

are capable of sustainable business in the SATCOM marketplace [13]. Alternate technologies adopted by terrestrial ISPs such as fiber optic internet make the tradeoff between P-LEO satellite internet and terrestrial solutions a much more complex and difficult decision for consumers, thus leading to an increasingly noisy marketplace and a potential speculative bubble in the SATCOM market.

While much of LEO activity has seen commercial growth, national security interests have also continued to drive growth by adding new capabilities for command and control in space. This has led to space becoming a contested domain by multiple nations. Such expansions of national security infrastructure can shorten communication times between warfighters, serve as robust and redundant communications methods, and provide warfighters and rescue operations communications in remote regions that would otherwise not have them. However, the growing activity by militaries entering LEO is causing many to grow concerned by the escalating brinkmanship that is leading many private companies to reevaluate their space strategy. While this work does not *directly* address space warfare, the business strategies in response to congestion, strategic constellation design, and launch cadence may inform space policy for similar threats to nations and militaries. Similarly, space strategy from a national security perspective could also inform the mission design and profitability of operating a P-LEO constellation in a contested space environment.

1.2 The Satellite Communications Industry in the P-LEO Era

Historically, the commercial SATCOM sector has been divided into two primary service categories: unidirectional broadcasting and bidirectional broadband internet. Previous generation systems from Iridium [14], Globalstar [15], Teledesic [16], and Orbcomm [17] struggled to gain traction in the broadband internet marketplace due to substantial expenditures, relatively poor network performances, and limited practical use cases. The financial burden associated with satellite production, launch vehicles, and maintenance offered an unappealing option to consumers outside of very specific use cases. However, an increased data demand coupled with advancements in antenna technologies [18] and reusable launch vehicles [19, 20] has facilitated a surge in investment and subsequent development activities of constellations operating in low

Earth orbit (LEO) [21]. Such Proliferated Low Earth Orbit (P-LEO) constellations call for thousands of *relatively* cheap satellites to be deployed in LEO and provide low-latency internet access directly to consumers via user terminals. This constellation development strategy differs from conventional satellite systems in both volume and price per satellite, utilizing economies of scale to drive down production costs [22]. The goal of such development strategies is to entice military, maritime, aviation, and residential [23] consumers to consider new broadband capabilities made possible through low-latency P-LEO SATCOMs.

While P-LEO SATCOMs have the potential to connect millions of new consumers from underserved/remote regions and offer flexible broadband service alternatives to existing consumers in established markets, the widespread development of such systems introduces several challenges. The telecommunications landscape has become highly competitive, with numerous entrants vying for market share in a sector historically dominated by a few key players [24]. Given the substantial capital expenditures required to launch a P-LEO SATCOM network coupled with uncertainties surrounding service adoption rates and constellation development, the operational sustainability of new entrants in the satellite internet domain remains inconclusive. Beyond financial considerations, the increasing physical congestion of LEO introduces new risks to orbit sustainability, including the heightened potential for orbital conjunctions [25], electromagnetic interference [26], and spectrum allocation conflicts [27].

1.3 Challenges to P-LEO Constellations

The P-LEO SATCOM market remains a complex dynamical system with various challenges specific to its domain. Among these are environmental, operational, and technological challenges that constellation operators must consider when developing a constellation architecture and actually deploying their SATCOM networks.

A difficult environmental factor to study or model in this domain is the temporal dissonance between short-term orbit periods (on the time scale of minutes and hours) and long-term strategic planning (on the scale of years or decades). This timescale mismatch becomes a critical driver of strategic constellation development in multiple corporate strategies and development timelines. While some works have analyzed constellation architectures in a comparative manner [21], no

works have explicitly accounted for both the short-term and long-term dynamics simultaneously in their modeling approaches.

Additionally operating a satellite constellation, at scale, in the LEO environment presents challenges regarding collisions and limited lifespans. Given the the accumulation of RSOs and debris fragments in LEO, operators must now consider the risks and rewards behind launching to particular orbit shells. Additionally, compounding this risk is the finite operational lifespan of LEO satellites, approximately five years, which necessitates continuous and costly replenishment launches to sustain network coverage and service quality. Taken together, operators must balance broader financial competition with sustainable usage of LEO.

1.4 Research Objectives

This dissertation attempts to address the above challenges by proposing and developing comprehensive modeling frameworks to advance the study and understanding of dynamics within the P-LEO SATCOM domain. In pursuit of this overarching goal, three central research objectives are established, each accompanied by a set of supporting sub-objectives:

Objective-1: Quantify the effects of satellite insurance and deorbit penalties on operators' constellation development strategies via an empirical study.

1. Identify the tradeoffs SATCOM operators must consider when trying to generate profits (specifically, the development of space debris created from LEO satellite collisions).

2. Assess how the inclusion of insurance policies and deorbit penalties affects: orbit sustainability (evolution of the amount of debris in an orbit slot), business sustainability (evolution of average profits per player per round), and SATCOM operator strategy (broad differences in player behavior).

3. Measure the progression of participant strategy improvement across multiple simulated scenarios.

Objective-2: Develop a parameterized environment of a single operator's performance in a dynamic P-LEO SATCOM market.

1. Identify necessary environment parameters and space operations when modeling the performance of an operator's constellation development strategy.

2. Develop an environment to model competitive, temporal market dynamics and performance of an operator's constellation development strategy (actions at each epoch) within the P-LEO SATCOM environment.

3. Characterize which combinations of environmental parameters lead to feasible operational strategies and characterize the domain (environment conditions required for operationally sustainable P-LEO networks).

Objective-3: Develop a multi-agent framework for multiple P-LEO constellation operators to interact within the same competitive environment.

1. Expand the single-agent framework into a multi-agent environment that models multiple P-LEO satellite operators (AI agents) taking actions, interacting with each other, and competing within the shared competitive environment simultaneously.

2. Conduct both decentralized training and execution as well as centralized training with decentralized execution.

1.5 Structure of this Dissertation

Given our three objectives, we are able to break this complex dynamical systems problem into three parts that each elucidate a unique understanding of the P-LEO constellation market and addresses each of our research objectives. Hence, the structure of this work is organized as follows:

- Chapter 2 describes the field of research as it stand today. This covers how P-LEO satellite constellations are modeled, the usage of reinforcement learning in aerospace and space systems, multi-agent reinforcement learning, and gamification of space operations. Gaps are identified in the existing literature with this work filling them.

- Chapter 3 is dedicated to the first objective which investigates the impact of economic and policy instruments on the generation of space debris. Development of a tabletop board game as a surrogate simulation tool to model the underlying dynamics is presented. The use of randomized controlled trials as our experimental design, the data collection process, and subsequent results with analyses are also discussed.

- Chapter 4 describes our second research objective: the creation of a single-agent reinforcement learning environment designed to mimic the competitive P-LEO marketplace. The environmental design along with the associated dynamics are carefully described along with the implementation details required to create the environment. The environment is then instantiated and a variation of parameters is performed by allowing an agent to optimize its strategy in each individual case. The results of this investigation are then presented and discussed.

- Chapter 5 expands on the previous chapter by creating a new, multi-agent reinforcement learning (MARL) version of the single agent environment. We discuss the MARL algorithms used and report results from the MARL learning.

- Chapter 6 concludes the manuscript with discussion and presents future potential avenues of research from each objective.

1.5.1 Dissertation Paper Contributions

Each of the following publications contributed to a specific chapter of this dissertation. Publications which are in preparation for publication are stated as such.

- **Satellite Tycoon: Modeling Economic Competition in the Business of P-LEO Constellations (Chapter 3 [28])**
  D. Guzzetti, D. R. Tauritz, **R. Qureshi**, C. Roberts, M. Indaco, L. Bone, E. Kimbrell
  In *11th International Workshop on Satellite and Constellations Formation Flying*. Milan, Italy, June 2022

- **Modeling and Gamification Framework of Business Competition Between P-LEO Constellations (Chapter 4 [29])**

  R. Qureshi, C. Roberts, M. Indaco, L. Bone, E. Kimbrell, S. Mulder, D. R. Tauritz and D. Guzzetti, In *2022 AIAA/AAS Astrodynamics Specialist Conference*. Charlotte, NC, August 2022

- **A Table-Top Game to Simulate Competition Between P-LEO Satellite Internet Constellations (Chapter 3 [30])**

  R. Qureshi, C. Roberts, E. Kimbrell, S. Mulder, A. Rao, D. R. Tauritz and D. Guzzetti, In *2023 AIAA/AAS Astrodynamics Specialist Conference*. Big Sky, MT, August 2023

- **A Tabletop Game to Study Business Wargaming in the P-LEO SATCOM Marketplace (Chapter 3 [31])**

  R. Qureshi, R. Gleason, A. Rao, S. Mulder, D. R. Tauritz and D. Guzzetti, In *IEEE Conference on Games (CoG)*, Milan, Italy, 2024, pp. 1-8, doi: 10.1109/CoG60054.2024.10645581

- **A Reinforcement Learning Framework for Strategy Exploration in the Dynamical P-LEO SATCOM Marketplace (Chapter 4)**

  R. Qureshi, A. Rao, S. Mulder, D. R. Tauritz, and D. Guzzetti, In prep for publication in *Acta Astronautica*.

- **Modeling Economic Instruments in a Tabletop Strategy Game for Space Sustainability in LEO Satellite Constellations (Chapter 3)**

  R. Qureshi, R. Gleason, S. Mulder, D. R. Tauritz, and D. Guzzetti, In prep for publication in the *AIAA Journal of Spacecraft and Rockets*.

Chapter 2

Literature Review

The rise of P-LEO satellite constellations has introduced a unique dynamical systems problem that sits at the intersection of astrodynamics, economics, and strategic decision-making. Addressing this problem requires drawing from a diverse body of literature spanning satellite systems, machine learning, game theory, and market modeling. This chapter surveys the state-of-the-art across each domain, identifies gaps in the existing literature, and motivates the methodology proposed in subsequent chapters.

2.1 P-LEO Satellite Constellation Modeling and Analysis

The modern SATCOM architecture leverages P-LEO constellations, comprising thousands of satellites in low Earth orbit that provide low-latency internet access directly to end users. This constellation development strategy differs from conventional satellite missions in both volume and price per satellite, utilizing economies of scale to drive down production costs [22]. The goal of such development strategies is to entice military [32], maritime [33], aviation [34], and residential [23] consumers to consider new broadband capabilities made possible through low-latency P-LEO SATCOMs.

While P-LEO SATCOMs have the potential to offer flexible broadband service alternatives to established and emerging markets, the widespread development of these space systems introduces several challenges. The telecommunications landscape has become highly competitive, with numerous entrants vying for market share in a sector historically dominated by a few key players [24]. Given the substantial capital expenditures required to launch a P-LEO

SATCOM network coupled with uncertainties surrounding service adoption rates and constellation deployment risk, the operational sustainability of new entrants into the satellite internet domain remains inconclusive. Beyond financial considerations, increasing physical congestion of LEO introduces new risks to orbit sustainability, including the heightened potential for orbital conjunctions [25], electromagnetic interference [26], and spectrum allocation conflicts [27].

Prior studies of the P-LEO regime have examined space safety [35, 36, 37], compared constellation architectures [38, 21], tested anomaly detection methods [39], and monitored space debris accumulation [40, 41, 42]. However, none of these works address the economic or logistical impacts of LEO congestion.

Research surrounding the financial and strategic decision-making of P-LEO SATCOMs has largely focused on the financial feasibility of select constellations or specific areas of interest such as resource allocation [43], customer demand estimation [44], and the effects of orbital use fees [45, 13]. Techno-economic frameworks have been developed to assess the economic feasibility of specific constellation designs [46, 47]; however, these works lack a key feature of dynamical systems: the ability to support sequential decision-making. Such decision-making is essential for understanding and adapting strategies over time in response to changing market and environmental conditions. While certain dynamical systems have been introduced to study staged constellation deployment [48] and satellite inventory control [49], these works do not model dynamic, competitive markets. Furthermore, traditional combinatorial approaches such as grid search are either too coarse to be of value or yield an infeasible number of strategies to propagate temporally, and they cannot account for the interdependent actions of all competing constellation operators. Considering these factors, it becomes clear that alternative approaches are needed to adequately capture both the temporal dynamics and the granularity of the decision space inherent to the competitive P-LEO environment.

Modeling of the space environment more broadly depends on the use case and application of the space system. SpaceNet Cloud [50] is a web-based mission logistics planning tool, while GPS satellite constellation replenishment has been formulated as a Markov decision process (MDP) and solved via value iteration to produce replenishment policies [49, 51]. In each of these modeling scenarios, competition is entirely ignored and proposed constellations are assessed

independently. Many factors critical to the competitive P-LEO environment remain absent from these frameworks, including dynamic strategies and development timelines, competitive markets with limited customers, potential revenues, orbital congestion and debris, and satellite replenishment costs.

2.1.1 Space Sustainability

The Secure World Foundation has defined space sustainability to be the process of "ensuring that all humanity can continue to use outer space for peaceful purposes and socioeconomic benefit now and in the long term" [52]. The practical considerations of space sustainability in LEO becomes responsible space debris mitigation as well as conjunction avoidance. Space debris itself is defined as "all man-made objects, including their fragments and parts, [...] that are non-functional with no reasonable expectation of their being able to assume or resume their intended functions" [53]. As the number of satellites, space debris, and all other resident space objects (RSOs) in LEO continue to grow, constellation operators face a congested environment with a high probability of collisions. If a collision were to happen, more debris would be generated and the operator would also need to potentially replace their lost satellite; thus launching even more satellites into an even more congested environment. This feedback loop of launching more satellites into an already congested environment would be a natural Kessler Syndrome [8] and eventually render LEO inaccessible. Additionally, it has been shown that unsustainable SATCOM launches and debris generation will cause LEO to become economically infeasible prior to becoming physically inaccessible [54]. This "Economic Kessler Syndrome" implies a first-movers advantage if sustainable launch and deorbit behavior are not a standard practice. In subsequent works [55], it was shown that 70% of expected economic losses due to collisions would occur in LEO. When viewing LEO access through the lens of game and economic theory as a shared resource, both the physical and economic Kessler Syndromes can be described as a complex "Tragedy of the Commons" [9] that require economic incentives and policy instruments to avoid.

While regulations promote space sustainability, the effects of proposed economic and policy instruments have not been tested or validated. Regulators such as the United States

Federal Communications Commission (FCC) and policy makers such as the United Nations Office of Outer Space Affairs (UNOOSA) have each prescribed a 5-year deorbit rule [7] and a 25-year deorbit limit [40], respectively. While stricter time limits on post-mission disposal and deorbit are shown to decrease rates of potential conjunctions [41, 56], the efficacy of specific enforcement tools is still unclear. Monetary fines, license revocation, and denial of market access are all emerging policy instruments for improper satellite end-of-life (EOL) disposal. Additionally, mandatory insurance models are also gaining interest as an indirect enforcement tool. However, these economic and policy instruments have not been empirically tested in the real world, and the window for proactive intervention may close before their efficacy can be established. Furthermore, satellite constellation operators cannot be subjected to controlled experimental conditions in which differing enforcement and regulation are applied inconsistently across operators. This creates a fundamental barrier to empirical evaluation of policy instruments and necessitates the use of a suitable modeling approach.

2.2 Reinforcement Learning in Aerospace and Space Systems

Reinforcement learning (RL) is a branch of machine learning in which an agent learns to make sequential decisions by interacting with an environment, receiving scalar reward signals, and updating its policy to maximize cumulative return [57]. Formally, RL problems are cast as Markov decision processes (MDPs), where a state space, action space, transition dynamics, and reward function define the learning problem. The emergence of deep neural network function approximators gave rise to deep reinforcement learning (DRL), unlocking RL agents to operate effectively over high-dimensional, continuous state and action spaces that were previously intractable [58]. Groundbreaking results such as AlphaGo's superhuman performance at Go [59] and AlphaStar's Grandmaster-level play in StarCraft II [60] demonstrated that RL agents trained through self-play and large-scale simulation can discover highly complex, non-intuitive strategies that emerge from the reward signal alone—without being explicitly programmed. These results showed that general-purpose learning methods can be applied to other complex domains in which agents must compete or coordinate with one another in environments with limited information and variable time-scale decision-making. Emergent strategy discovery is one of the most useful

properties of RL for domains characterized by large combinatorial decision spaces, competitive dynamics, and evolving environmental conditions such those found in the P-LEO SATCOM market.

The application of RL to aerospace systems has grown considerably in recent years, with spacecraft guidance, navigation, and control (GNC) representing the most mature application. A broad survey of RL-based approaches to spacecraft control identifies guidance and navigation for landing on celestial bodies, interplanetary trajectory design, attitude control, rendezvous and docking proximity maneuvers, and constellation orbital control as active application areas, demonstrating that RL can address emerging needs for highly autonomous on-board capabilities that are robust to system uncertainties and adaptive to changing environments.

In the domain of low-thrust trajectory design, RL has been applied to the problem of asteroid approaching under stochastic dynamics. By formulating the trajectory guidance problem as an MDP and incorporating navigation performance into the reward function via the Fisher information matrix, RL agents trained in this setting learn policies that jointly optimize propellant consumption and optical observation quality [61]. Similarly, RL has been applied to robust interplanetary trajectory design under severe uncertainties, where trained agents produce guidance policies that outperform classical optimal control methods in the presence of model errors and disturbances [62].

Spacecraft proximity operations and autonomous docking represent perhaps the most active RL application in spacecraft GNC. By enabling guidance strategies to be learned rather than analytically designed, RL-based controllers can be trained entirely in simulation [63]. The learning burden is subsequently lowered by combining RL with classical control theory for velocity tracking [64]. The problem of multi-phase autonomous docking in cluttered orbital environments has also been addressed with hierarchical RL architectures [65].

Beyond GNC, RL has been successfully applied to higher-level decision-making and scheduling. The feasibility of adapting DRL-driven policy generation to spacecraft decision-making has been demonstrated by framing operations scheduling as an MDP [66]. For Earth-observing satellite constellations, the scheduling of imaging and downlinking tasks across large Walker-delta constellations has been formulated as a single-agent RL problem [67]. Additionally,

a policy trained in a single-satellite environment can be deployed across each satellite in a constellation, with satellites communicating to avoid redundant observations, and constellation design and communication assumptions are shown to substantially impact system performance. Space situational awareness (SSA) represents another important application domain. DRL has been applied to the SSA sensor tasking problem and outperformed myopic policies on both RSO state uncertainty reduction and the number of unique objects observed [68].

A strength of RL in the case of complex systems is its capacity to produce emergent behaviors that are neither explicitly designed nor anticipated. In competitive game environments, RL agents trained via self-play have developed strategies that surprised human experts [69]. This emergent quality is particularly relevant when modeling strategic behavior in competitive markets, where rational actors adapt their strategies in response to evolving competition. Furthermore, the Decision Transformer [70] further extends the RL paradigm by recasting the sequential decision-making problem as a sequence modeling problem, allowing transformer architectures to generate actions conditioned on desired returns. This offline policy learning enables learning from previously collected trajectories without requiring online interaction. Reinforcement learning from human feedback (RLHF) [71] provides yet another mechanism by which agent behavior can be shaped by human judgment rather than programmatic reward signals.

When collectively reviewed, the body of RL literature in aerospace and space systems demonstrates that RL agents can solve problems of substantial complexity with the possibility of emergent behaviors. However, existing applications have largely focused on single-agent physical control problems or cooperative scheduling tasks. The competitive, multi-agent dynamics of the P-LEO SATCOM market have not been addressed within an RL framework.

2.3 Multi-Agent Reinforcement Learning

Multi-Agent Reinforcement Learning (MARL) extends single-agent RL to environments where multiple agents interact with the environment and one another, simultaneously. In contrast to single-agent systems, MARL introduces additional complexity due to agent interactions, cooperation, and competition. A comprehensive overview of MARL, including its foundations

and modern approaches, is provided by Albrecht et. al. [72] and much of this section is taken from this book.

One common approach in MARL research is to reduce multi-agent problems into a single-agent version. The environment is then treated as if it is fully controlled by a single agent, allowing standard single-agent RL algorithms to be applied [73]. While this reduction simplifies learning and enables the use of well-established methods such as DQN or PPO, it often fails to capture the non-stationary dynamics inherent in multi-agent systems, where other agents are concurrently learning and adapting [74].

Centralized learning approaches aggregate the observations, actions, and rewards of all agents into a single learning problem [75]. Using complete state information allows the centralized learners to coordinate agent behaviors and optimize global objectives. Such a method is preferred for MARL problems with few agents working cooperatively. However, due to the exponential growth in joint state-action space, centralized methods typically suffer from scalability issues as the number of agents increases [76].

Independent learning treats each agent as a separate learner, optimizing its own policy without explicit coordination with other agents [77]. This approach is simpler and scales easier; however, it often involves challenges associated with non-stationarity: from the perspective of any single agent, the environment changes as other agents update their policies, leading to potentially difficult or even unstable learning [78].

To avoid this, MARL algorithms often employ operational modes to facilitate learning:

- **Self-Play:** Agents learn by competing or cooperating with copies of themselves, which can stabilize learning and lead to potentially emergent behavior [79]. Self-play is widely used in zero-sum games and environments with symmetric roles, and is analogous to scenarios in P-LEO satellite constellations, where agents compete for shared pools of customers [80].

- **Mixed-Play:** Agents interact with each other using different algorithms or policies [81]. This mode allows for heterogeneous agent populations and can increase robustness by exposing agents to diverse behaviors, but may also cause non-stationarity [82].

2.4 Game-Based Approaches in Space Strategy and Operations

A series of works [83, 84, 85, 86] relating to the development of a game modeling the dynamics of federated space systems (FSS) are precursors to the tabletop game presented in this work. Laying the groundwork for a fully developed simulation game, a space-based resource economy in the context of wargaming and federated systems is described [83]. Their work presents an example from lunar in-situ resource production applications. In [84], the FSS approach is described as a system of systems (SoS), and a simulation architecture is described using high-level architecture standards. The *Orbital Federates* simulation game describes FSS logistics and is presented in [85]. Using the *Orbital Federates* game, federated systems are modeled as a Stag Hunt game and risk-dominant (non-cooperative) and payoff-dominant (cooperative) strategies are investigated [86]. While these works are able to capture the dynamics of space systems working in tandem, external competitors and markets are not considered. Additionally, the area of interest for these works primarily revolves around lunar missions rather than the LEO regime. To adequately study P-LEO SATCOM markets, competitive simulations or strategic planning games are needed.

To properly model and study the P-LEO SATCOM marketplace, we employ business wargaming as it can quickly model and describe dynamics of P-LEO business situations as a tabletop role-playing simulation [87]. Such games [88] typically assign players to specific roles such as competitors, regulators, or customers and last several rounds of gameplay (simulating years of real-world decision-making). Business wargames are defined by several key characteristics and challenge preconceived mental models, help participants overcome cognitive barriers, detect weak signals of change in a broader system (or marketplace), and assist with the development of foresight [89].

Although "foresight" has an ambiguous definition in many works, it has often been described as decision-making, given input informed by a view of future, long-term events. Rather than simply predicting system developments, foresight involves recognizing emergent patterns and interpreting relevant features which are likely to impact players downstream [90]. As an individual's foresight improves, their ability to recognize the faintest of changes within the

business environment helps to show alternative scenarios of how the market might evolve. This is often referred to as a player's "memory of the future" [91].

Business wargames yield several unique benefits not available with other modeling techniques such as: multi-player strategy-testing, development of foresight [89], anticipation of future competitive dynamics [92], and training/education [93]. While scenario planning tends to be time-consuming and limited in its ability to incorporate future dynamics into the scope of study, business wargames rigorously examine a situation from several perspectives by actively involving players and providing them with opportunities to learn and improve from previous games. Additionally, business wargames sidestep common cognitive barriers such as: groupthink, overconfidence, mental filters, reduction of ambiguity, and acceptance of confirming evidence rather than skepticism. While humans visualize the future through their own past experiences [94], business wargaming allows us broaden this scope to capture increasingly diverse scenarios and test strategic possibilities.

To study P-LEO SATCOM market dynamics, prior work introduced a digital real-time strategy game dubbed *Satellite Tycoon* that was designed to model competition between players who took the role of P-LEO constellation operators in the game [29, 28, 95]. A screenshot of this game is shown in Figure 2.1. However, lags between modeling and user interface design updates coupled with extensive play-testing to improve playability proved challenging. Using lessons learned and the need to accurately capture dynamics of the P-LEO marketplace, a tabletop version of *Satellite Tycoon* was developed using simplified play-testing and rapid prototyping of individual game elements representing the P-LEO marketplace [31].

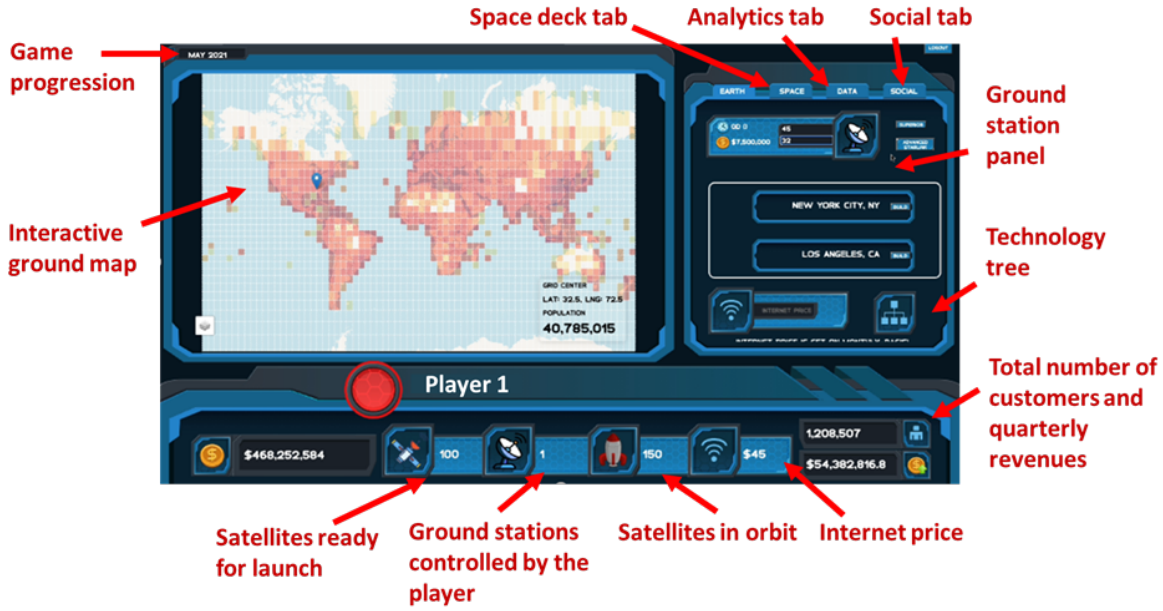

Figure 2.1: The digital version of the *Satellite Tycoon* multi-player game.

Chapter 3

Modeling Space Sustainability via Tabletop Board Game

This chapter is dedicated to the development of a research board game called *Satellite Tycoon* and its use as an integrated assessment framework to investigate how economic instruments affect: orbit sustainability (evolution of the amount of debris in an orbit slot), business sustainability (evolution of average profits per player per round), and operators' constellation development strategies. A pilot randomized controlled trial (RCT) with human participants was performed with recruited participants randomly assigned to one of two groups: an unregulated control group or a regulated treatment group. Regulation of the treatment group included mandatory deorbit penalties when a player did not deorbit their satellites sustainably as well as an optional satellite insurance to protect against possible collision events. All participants played three games in groups of four player cohorts that were persistent across the RCT. The following hypotheses and research questions were formulated at the start of the study:

- **Hypothesis 1:** As players better understand the board game (surrogate simulation environment), more complex and profitable strategies will emerge from their gameplay. Specifically, players' overall constellation value and profits should increase across games.
- **Hypothesis 2:** The introduction of insurance fees will deter players from overproduction and still allow for profitable competition within the marketplace.
- **Research Question 1:** How does the introduction of insurance policies and fees impact the generation of space debris?
- **Research Question 2:** Does the introduction of an insurance framework lead to less space debris, and subsequently, less conjunctions?

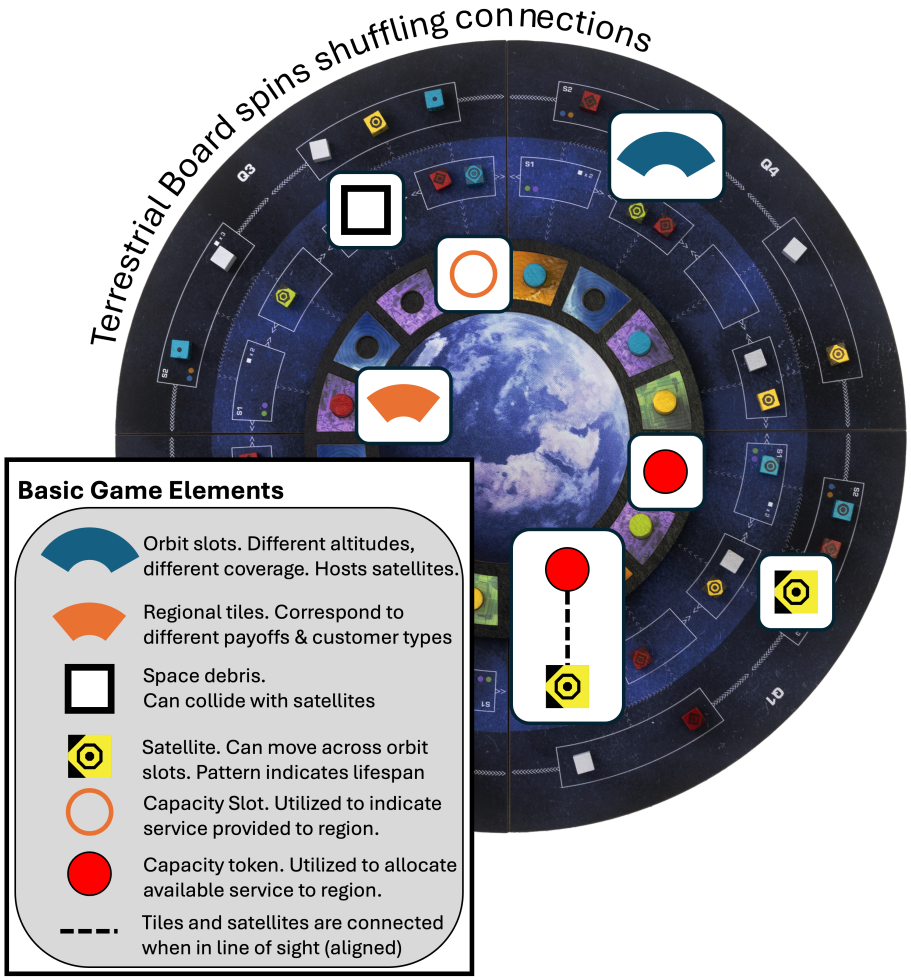

Figure 3.1: The basic elements of the *Satellite Tycoon* tabletop board game used to model P-LEO market dynamics and competition.

3.1 *Satellite Tycoon* as a Simulation Environment

Inspired by several different game mechanics (resource management, social interaction, feedback-and-reward, and turn-based), the *Satellite Tycoon* tabletop game is formulated as a multi-player game in which players act as P-LEO constellation operators to develop their networks, gain service capacity, and acquire customer profits using that capacity. Described below is the current version of the game with mappings to the underlying P-LEO SATCOM environment. A summary of the board game with descriptions of all game elements is shown in Figure 3.1.

3.1.1 Game Elements & Player Objectives

Drawing from other tabletop games in which players control locations on the board to gain a strategic advantage, the *Satellite Tycoon* board is divided into two distinct regions: earth and space. The fixed, outer two rings of the game board represent the space domain. Specifically, they abstractly represent two different orbit shells in LEO: S_1 and S_2. Each orbit shell is further sub-divided into orbit slots which abstractly represent regions in LEO. Note that multiple players and/or debris can occupy an orbit slot, thus operating in a congested region that is vulnerable to a conjunction. As shown, space is also sub-divided into quadrants, but this is only used during collision determinations.

In the game, satellites are consumable resources with limited lifespans as indicated by numbered, colored cubes that players purchase and place in orbit slots. Each satellite's lifespan (or fuel) is indicated by the upward facing graphical numbering system designed on each satellite cube. All satellites are launched with 6 rounds of fuel; however, players may choose to expend a unit of fuel to directly launch to S_2 orbit slots for more coverage over terrestrial regions.

The internal section of the board represents the different markets in which satellite internet services may be offered. There are four different types of markets (regional tiles): desert, rural, city, and sea. The distinction between our two orbit shells comes into play with coverage and latency: S_1 orbit slots cover a single region per round with lower latency, while S_2 orbit slots have higher latency and cover twice as many regions. As a model of the different use cases for P-LEO constellations, lower latency is preferred by both the rural and city regions whereas desert and sea regions prefer greater coverage. Within each regional tile, a capacity slot exists so that players may allocate their network capacity to specific regions in the form of a capacity token.

One of the novel modeling techniques introduced by this tabletop game is the design and representation of short-term performance characteristics of P-LEO satellites while also giving players the capability to design long-term strategies. The solution to this modeling challenge is a rotation of the terrestrial portion of the game board, within each round, throughout the game. This method abstracts away complex computations required in P-LEO satellite internet

constellation modeling by instead allowing players to take advantage of *average* coverage information (a common practice for first-order modeling of satellite constellations) [96, Chapter 9]. This allows players to consider both their short-term actions (the current alignment of both the orbit slot and terrestrial regions) and future strategies (any future possible alignment combinations of orbit slot and terrestrial region).

The objective of all players is to collect as much revenue from the development of their satellite networks as possible. At the end of the game, the player with the most combined funds and residual satellite value wins. A residual value is assigned to each satellite at the end of the 8-round game and is based on the remaining lifespan of each satellite (see Table 3.1). The residual value linearly decreases from the purchase price of a new satellite to zero. In the event of all players going bankrupt prior to the game ending, all players lose.

Table 3.1: Residual Satellite Value based on Lifespan.

Value($)	**Lifespan**
0	1
4	2
8	3
12	4
16	5
20	6

3.1.2 Game Setup & Initialization (Starting States)

To initially setup the game board, regional tiles are turned over, shuffled, and placed randomly onto the rotating board. This simulates a new set of initial conditions and customer locations, relative to one another. Once all regions have been placed, 2 pieces of space debris (white cubes with no numbering system) are placed randomly in each quadrant. This simulates an initial environment with pre-existing debris from prior launch activity.

At the beginning of the game, players are each allocated $150 of in-game currency as their starting funds and roll a D6 die to determine the first-mover during the initial round. In subsequent rounds, first-movers are assigned based on the player with the largest constellation, with players each rolling their D6 dice if there happens to be a tie. Upon each player's initial

turn, they place two free satellites in any orbit slots and receive the requisite capacity token. The player can then place that capacity token immediately in any open regional tile.

3.1.3 Gameplay (Environment Dynamics)

Every round of the game has four distinct phases, each of which involves strategic actions and/or a stochastic process. The different phases of gameplay are detailed below with the order of phases structured as a central game loop and shown in Figure 3.2).

Collision Check

To model stochastic conjunctions and the possibility of collision, each round of gameplay begins with a collision check. In this phase of the game, a D6 die is rolled to determine in which quadrant collisions might occur. If a 5 is rolled, a collision check is performed on all applicable S_1 orbit slots. If a 6 is rolled, a collision check is performed on all applicable S_2 orbit slots. If multiple assets owned by different players (or debris) are in the same orbit slot, that orbit slot is considered congested. Within the quadrant, a D6 die is rolled for each congested orbit slot. Even numbers indicate a collision; odd numbers indicate safety. If a collision occurs, all active satellites in the orbit slot are replaced with corresponding white debris cubes.

Board Rotation

To model changes in average coverage over each regional tile, the terrestrial portion of the board is rotated clockwise based on the roll of a D6 die. This phase of the round is stochastic to maintain an analog with global coverage computations.

Fuel Reduction & Deorbit Decisions

A simplified modeling approach to simulate the natural satellite lifespan is to decrease each active satellite's upward facing life counter in each round of gameplay by 1. The decrease is shown on each player's satellites using a graphical numbering system as shown in Figure 3.3 and allows us to investigate end-of-life practices and incentives. If a satellite has a single unit of fuel left, its owner has the option to use that remaining fuel to sustainably deorbit their satellite

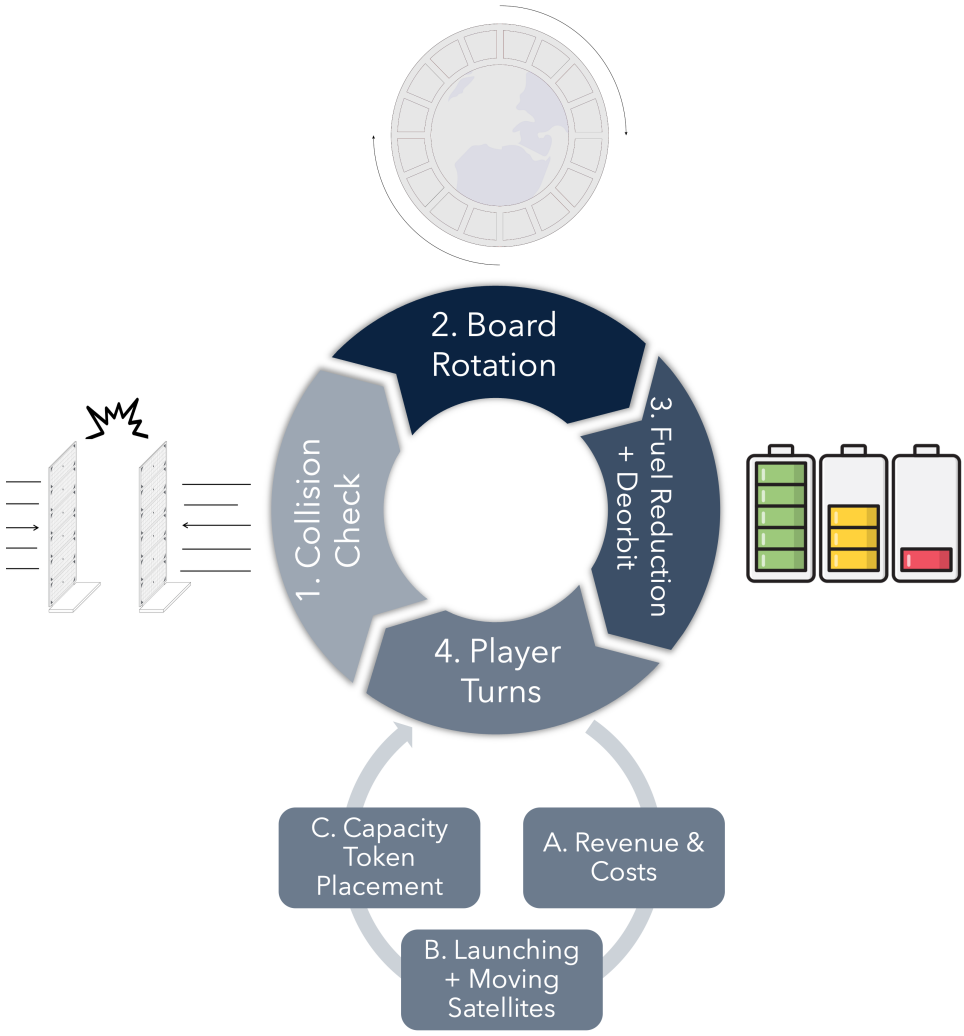

Figure 3.2: The central game loop describes the four phases of each round of gameplay: (1) a stochastic collision check is conducted to determine if any conjunctions occur between satellites or debris, (2) the terrestrial board rotates a stochastic number of slots, (3) the fuel is reduced on all satellites in orbit and satellites are possibly deorbited near the end of their lives, and (4) players take sequential turns collecting profits from their capacity tokens and reinvesting into more satellites and capacity for more profits.

Figure 3.3: The progressive decrease of each player's satellite lifespan from 6 (leftmost) to 1 (rightmost) using a graphical numbering system.

or use that remaining fuel to provide continued operations for one more round of gameplay. All satellites that have a single unit of fuel are converted into debris during the next round.

Player Turns

Modeling positional timing advantages, turn order is assigned with the player owning the most satellites in orbit as first and continuing clockwise until all players have had their turns. Each player's turn consists of three steps:

1. *Collecting Revenues:* Each capacity token placed in a regional tile generates \$5 in base revenue. A line-of-sight connection is formed when a player's satellite is directly above a regional tile with their capacity token, as shown in Figure 3.4. Each such connection yields a bonus determined by the satellite's orbital altitude and the regional tile type, as shown in Table 3.2. Note that both desert and sea regional tiles have stochastic bonus structures while rural and city each have a deterministic bonus. Bonuses of desert regional tiles are based on a card draw with an expected return of \$8. An example desert contract card is shown in Figure 3.5.

   Table 3.2: Line-of-sight bonuses for each regional tile and corresponding altitude

Region	Corresponding Altitude	Line-of-Sight Bonus
Desert	S_2	Draw a desert contract card.
Rural	S_1	Collect an additional \$5.
City	S_1	Collect an extra \$1 per active satellite in your network.
Sea	S_2	Roll a D6 Die and multiply the number by 5 for the bonus.

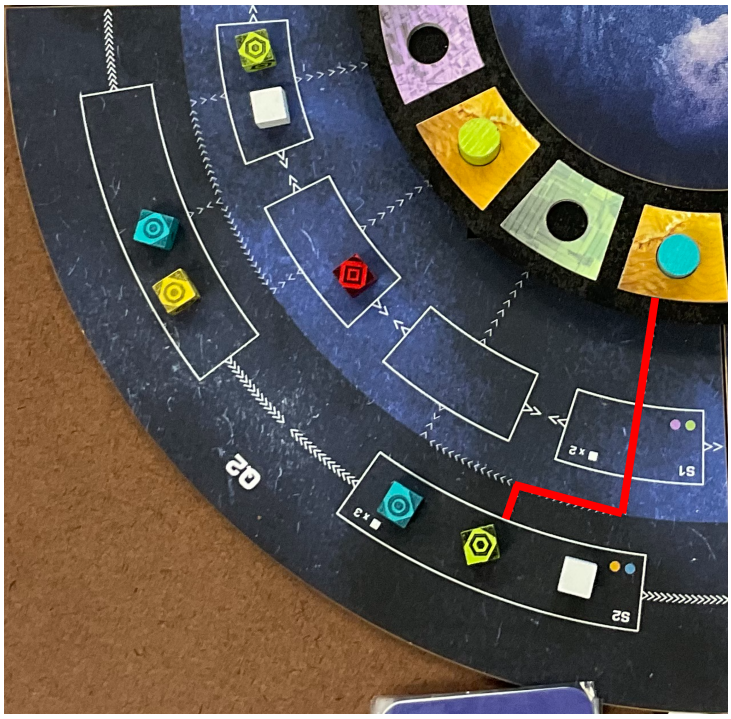

Figure 3.4: An example line-of-sight between the Blue Player's occupied regional tile and their satellite in S_2.

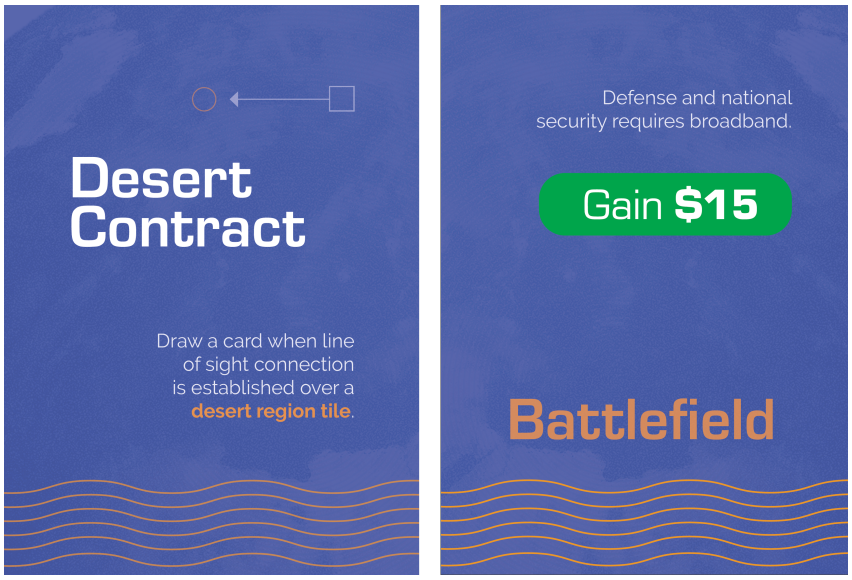

Figure 3.5: An example desert contract card awarding the player a one-time bonus of $15 during their turn.

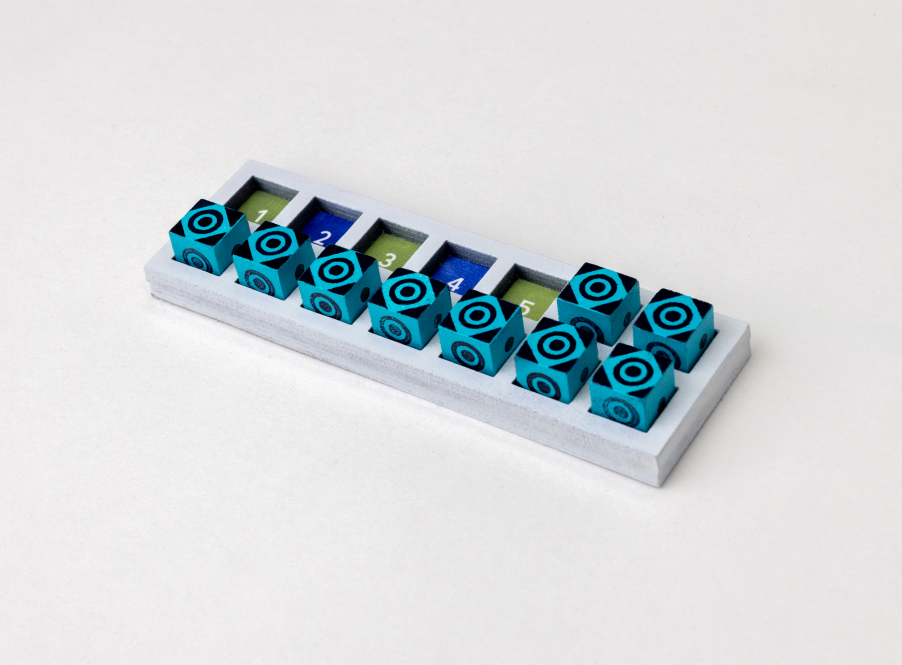

Figure 3.6: The satellite tray of the blue player indicating that they have 5 active satellites in orbit and have unlocked 3 capacity tokens (green squares).

2. *Launching & Moving Satellites:* Once a player has collected their profits, they may purchase new satellites or move existing satellites. New satellites begin with 6 units of fuel and may be launched to any available S_1 slot for $20 per satellite. To directly launch to S_2, the player's new satellite will begin with 5 units of fuel. Each subsequent movement to a different orbit slot requires a decrease of the fuel counter on each satellite.

3. *Placing Capacity Token:* Players acquire capacity tokens based on the number of satellites they have in orbit (displayed as the total number of green squares in their satellite trays, as shown in Figure 3.6). A player may only undercut or replace a competitor's capacity token in a given regional tile if all regions of the same type are claimed and the replacement has more satellites in line-of-sight, at the time the token is being replaced. If a player has their token removed, they may only replace it on their next turn.

Prescribed Debris Decay

To simulate the gradual decay of space debris in LEO, all space debris in S_1 is removed from the board at the end of round 4 (the midpoint of the game) and all debris in S_2 is shifted down (when available) to a S_1 orbit slot.

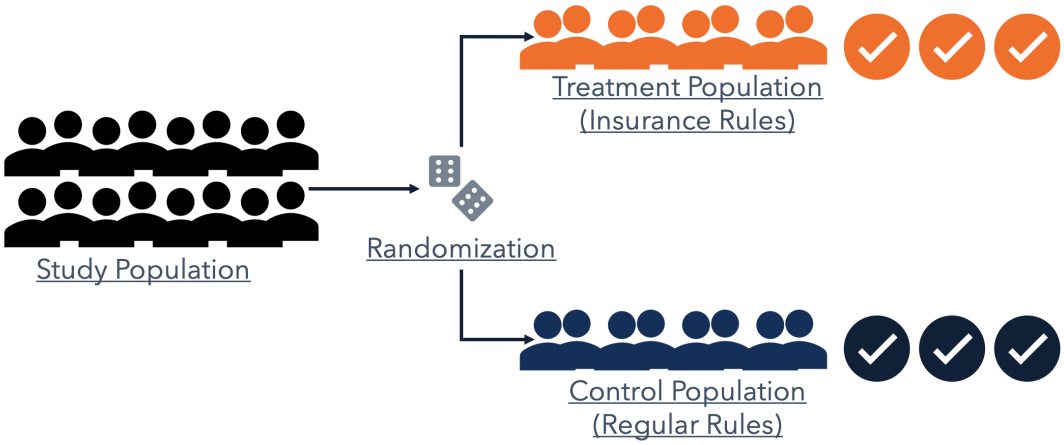

Figure 3.7: Participants were randomly assigned to either a REGULATED policy treatment group or an UNREGULATED control group to play 3 games.

3.2 Experimental Design (Randomized Controlled Trial (RCT))

A randomized controlled trial (RCT), as illustrated in Figure 3.7, is an experimental research design used to evaluate the causal effects of a policy treatment by randomly assigning participants to different conditions. Random assignment ensures that differences in outcomes between groups arise from the experimental policy variation rather than participant characteristics such as prior experience, risk tolerance, or strategic preference. Each participant group operated under identical game mechanics, resource constraints, and competitive structures, with the only difference being the presence or absence of the policy intervention.

In an idealized real-world setting, this study would examine two otherwise identical satellite markets operating under different regulatory conditions. However, conducting such controlled experiments with real satellite operators is infeasible due to cost, operational risk, and regulatory impracticality. Therefore, this study employs an analog experimental framework using human participants engaged in the *Satellite Tycoon* tabletop board game as a simulation environment.

3.2.1 Policy Treatment

Players in the policy treatment operate under the same rules that are described in the previous section; however, additional regulation introduces an optional first-party satellite insurance and a mandatory debris penalty. Players in the REGULATED policy treatment may opt into satellite

insurance at the beginning of each game; for simplicity and uniformity, they may not add or drop insurance within a game. During the revenue collection phase of their turns, players who opted for first-party satellite insurance are required to pay $2 (10% of the value of a new satellite) for each of their active satellites in orbit. This mechanism reduces the financial shock of unexpected losses and encourages more deliberate risk management when launching satellites into congested orbits. The REGULATED policy treatment also introduces a debris penalty for satellites that are not responsibly deorbited at the end of their operational life. If a player allows their satellite to become derelict instead of safely deorbiting it, they must pay one of the following liability fees per derelict satellite due to the debris risk created by that object:

- A $3 fee over 3 rounds
- A one-time $9 fee

This debris penalty represents third-party liability costs and incentivizes responsible end-of-life disposal decisions.

3.2.2 Data Collection & Metrics

In accordance with, and approval of, Auburn University's Institutional Review Board (IRB), a pilot study (STUDY00000287) was conducted in 2025 with 16 total participants. From this pool of participants, 8 players were assigned to play the control version and 8 players were assigned to play the policy treatment version. Both groups of 8 players were further subdivided into cohorts of 4 players. Each cohort played their variant of the game 3 times in-person, with audio and video recording. Note that in 3 out of 12 total games, video recording equipment did not properly capture video of the gameplay so photographs of the board were taken very frequently and matched with the audio recordings to transcribe the gameplay. This did not affect the quality of data collected and an example image is shown in Figure 3.8. The same variables were tracked across each round of every game and are shown in Table 3.3. We note that 16 participants is a nominal sample size; however, each player committed to three games and trends were still found across variables of interest.

Figure 3.8: An example image captured during an active game session.

Table 3.3: Variables recorded in each round of every game, where i is for each Player $\{1, 2, 3, 4\}$.

Game Variables Recorded Per Round
Number of Congested Orbit Slots
Amount of Debris from Collision Check
Number of slots Terrestrial Board is Rotated
Player i's Number of Satellites with 6 Fuel
Player i's Number of Satellites with 5 Fuel
Player i's Number of Satellites with 4 Fuel
Player i's Number of Satellites with 3 Fuel
Player i's Number of Satellites with 2 Fuel
Player i's Number of Satellites with 1 Fuel
Player i's Total Satellites
Player i's Derelict Satellites
Player i's Profit & Loss
Player i's Number of New Satellites
Player i's Funds

3.3 Results & Analysis

Following collection, the data was post-processed with final player funds and final player value being computed as values at the end of each game. We note that economic metrics such as net present value and compound annual growth rate were not computed; rather, cumulative funds were computed due to no grounded numbers being available to map a game round to real-world time. Although debris decay is modeled as a natural midpoint within each game and may approximate the mapping of a single round to real-world time, too many additional factors influence the orbital decay process and timeline. Additionally, performance is not compared across different lengths of time. Since the number of rounds and the initial funds were consistent throughout the study, growth rate computations would simply be a monotonic transformation of the cumulative final funds. Final player value is also computed as the sum of a player's final funds and the value of each our their remaining satellites at the end of the game based on their residual value. Finally, both cumulative final funds and final player value are normalized with relation to the initial funds the players started with.

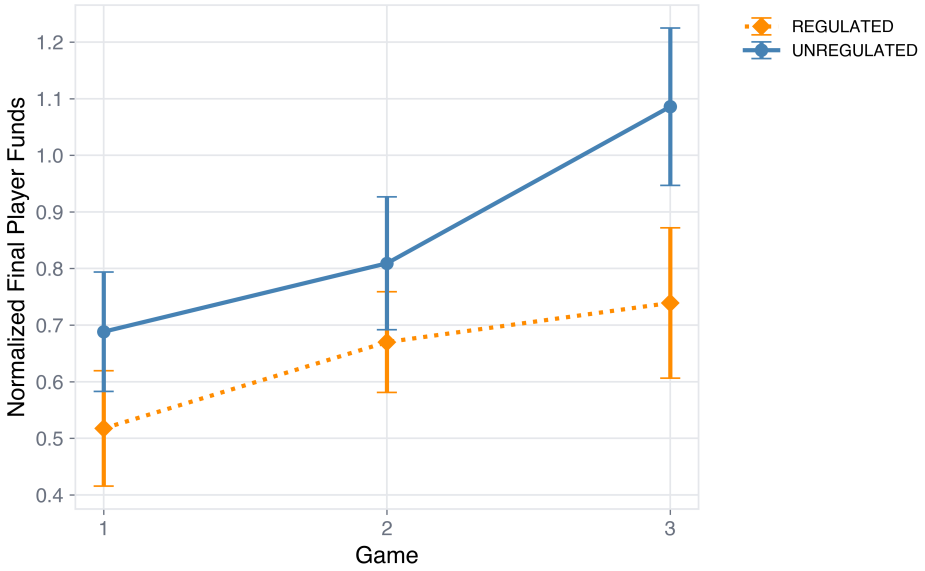

Figure 3.9: The average, normalized final player funds across all 16 players.

3.3.1 Learning Effects Across Games

Testing our first hypothesis, both final player value and final player funds are tracked to understand how players' understanding evolves and influences their strategies. Although this study only conducted 3 games per cohort, trends regarding learned behavior and strategy still emerged. Figure 3.9 shows the averaged, normalized final player funds continuously increasing for both the UNREGULATED control as well as the REGULATED treatment groups across all 3 games. Note that the standard error of the mean (SEM), is also shown as positive and negative bars and given as:

$$\text{SEM} = \frac{\sigma}{\sqrt{n-1}} \tag{3.1}$$

where σ is the standard deviation of the data and n is the number of players. It is noted that only the final average of the UNREGULATED control is above the initial value; all other games generated averaged final player funds below their initial funds. This could be due to several factors but is most likely due to large amounts of collisions in the initial games. The normalized final player funds can be visualized on an individual basis in Figure 3.10. By not averaging them, it is observed that only 1 out of 16 players made a cash profit at the end of Game 1. For Game 2, 4 players ended with more than their original funds. Finally, 8 out of 16 players finished with

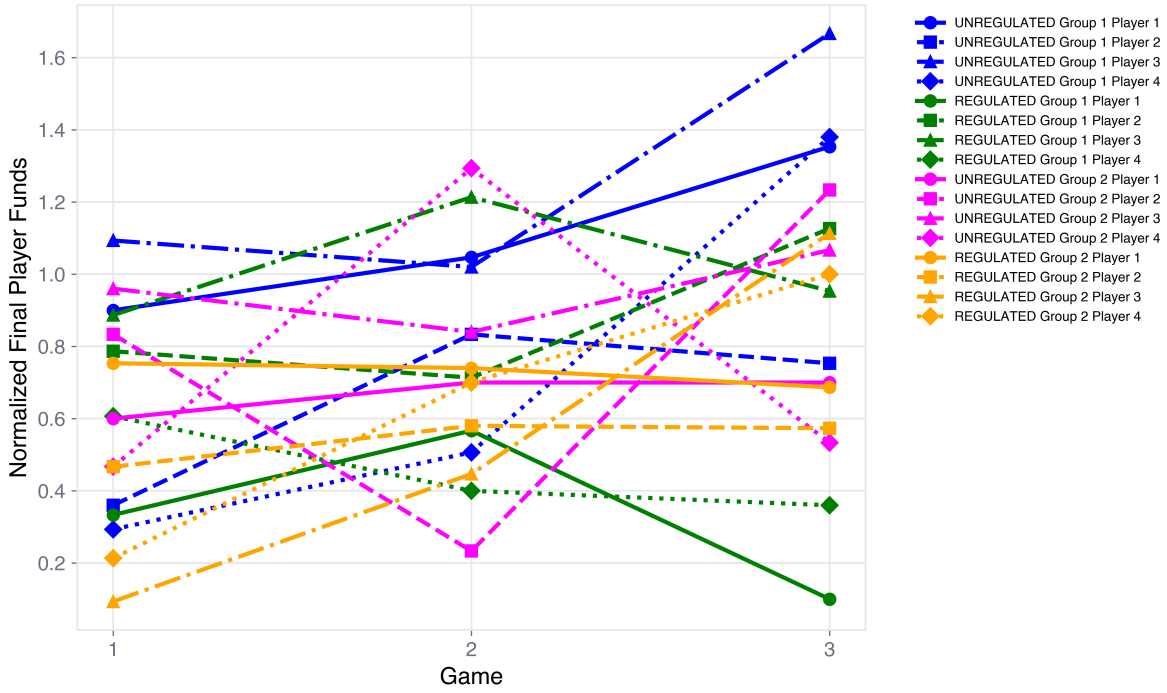

Figure 3.10: The normalized final funds of all 16 players.

a net gain in funds. This illustrates that as players spent more time playing the game, more of them were likely to turn a profit.

Learning effects shift slightly as final player values (the combination of final funds and residual satellites values) are considered. Figure 3.11 shows the normalized, average player value at the end of each game. This demonstrates an increasing trend for the UNREGULATED control group but not for the REGULATED treatment group. Even when accounting for residual satellite value, the average treatment group player never broke even or made a final profit. Similarly, investigating the normalized final player values on an individual basis illustrates that 3 out of 16 players made a true profit at the end of Game 1. In Game 2, 7 out of 16 players were able to make an overall profit. Finally, 12 out of 16 players were able to turn a true profit by Game 3. This further shows that as players became comfortable with the game, their strategy and performance also increased. Additionally, convergence of player values is observed in Game 3 and may be due to the existence of higher-order market equilibria or converging strategy performances.

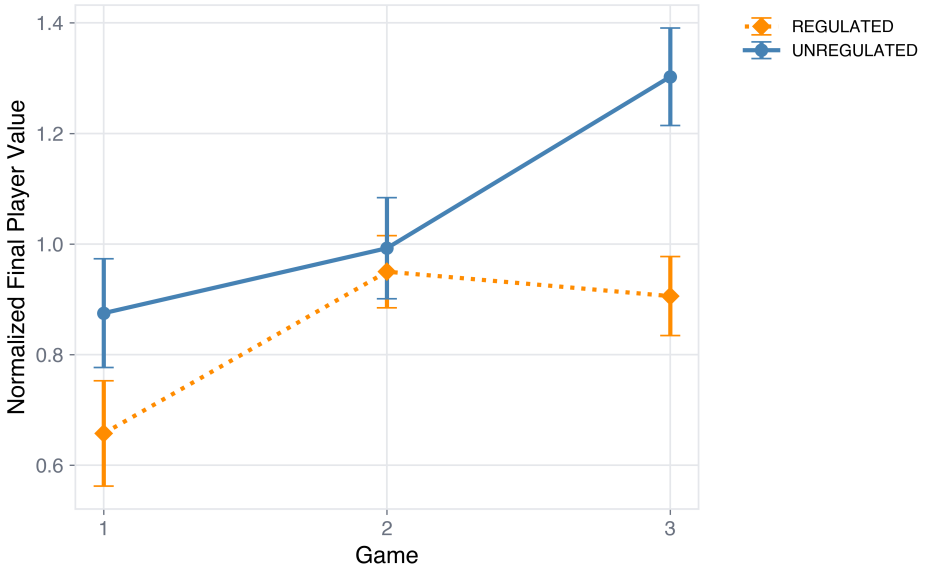

Figure 3.11: The average, normalized final player value across all 16 players.

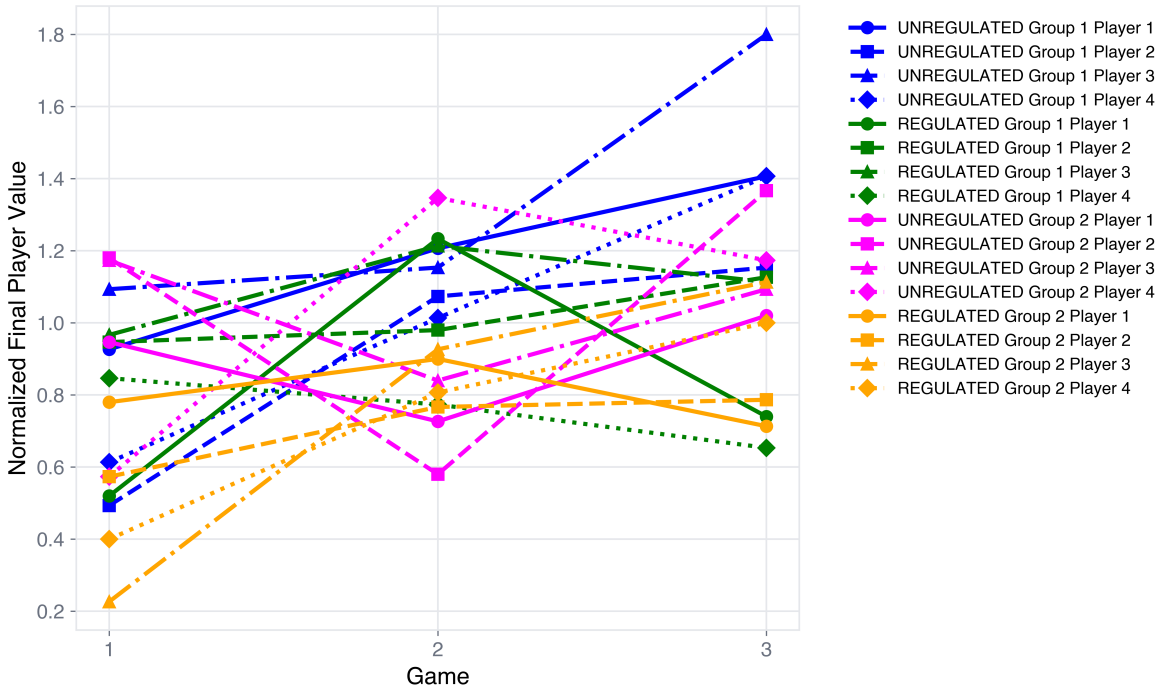

Figure 3.12: The normalized final value of all 16 players.

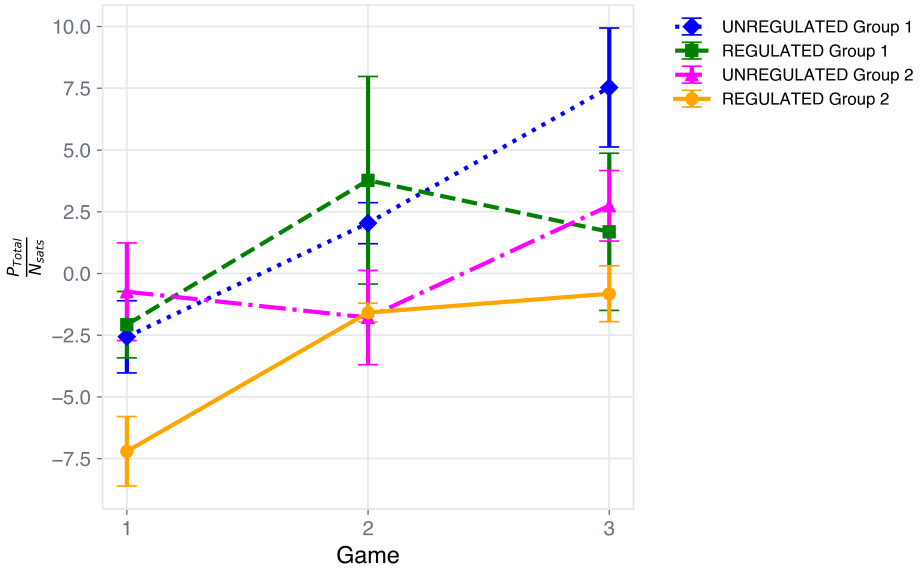

Figure 3.13: Revenue efficiency of each cohort within the study. While not all groups consistently increased their efficiency, the average revenue efficiency after all 3 games was higher than each group's respective average in Game 1.

Another indicator of player learning is revenue efficiency: total profits, P_{Total}, over the total number of satellites owned across a game, N_{Sats},:

$$\frac{P_{Total}}{N_{Sats}} \tag{3.2}$$

This value is not only a key indicator of constellation management but also the strategic positioning of a player's satellites. Additionally, this helped to understand the revenue efficiency of particular strategies. As shown in Figure 3.13, all group averages of revenue efficiency trend upwards indicating that players were not just getting luckier; rather, they were evolving their mental models of the environment and learning how to play the game better. This also indicates that players, on average, deployed satellites in a more revenue-efficient manner by generating more profits from less satellites as they progressed through the study. The breakdown of each player's revenue efficiency, shown in Figure 3.14, illustrates a general trend of improvement, except one outlier: UNREGULATED Group 1 Player 3. This player's strategy was unique and warrants further mention: the player opted into insurance, placed their initial two free satellites in an orbit slot with pre-existing debris, and received an insurance payment due to a collision at the beginning of round 1. They then proceeded to disengage from the market with their insurance

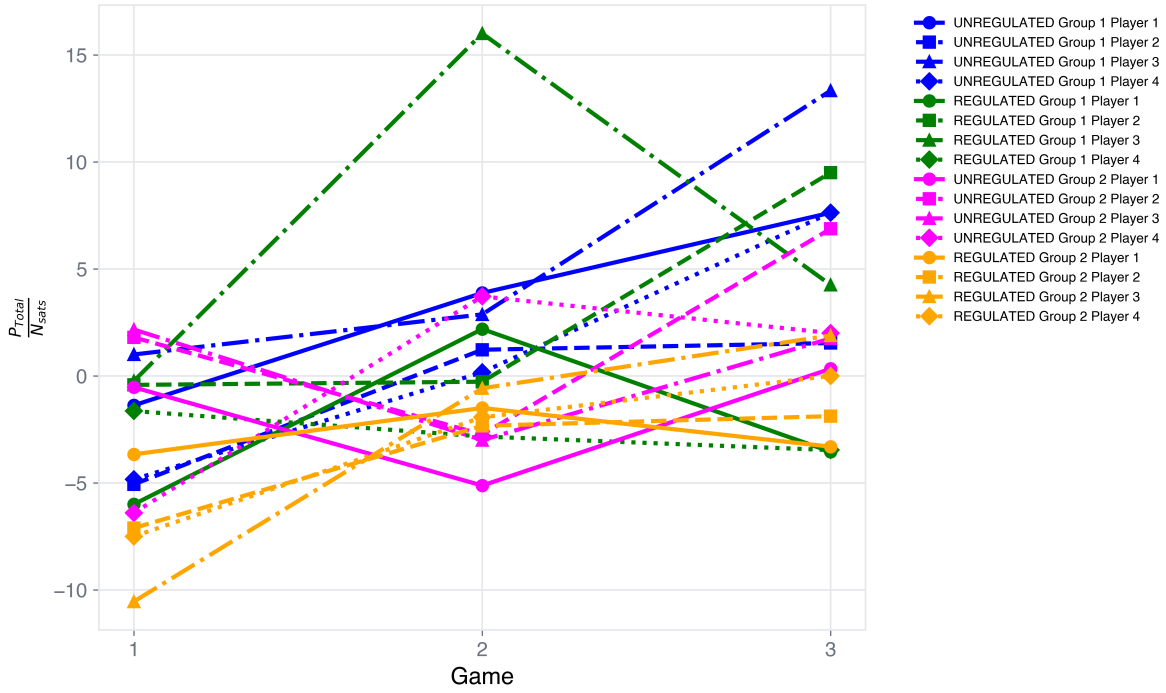

Figure 3.14: Revenue efficiency of all players in the study with an outlier due to a novel strategy.

payment being the majority of their profits that game. Even with this player self-removing themselves from the market, the profits and cumulative values of the other 3 players did not have a marked change from their previous game or from that of other groups. Given the same amount of resources (orbit slots) divided up among a smaller group of competitors, one would assume this would induce greater profits due to less competition in the market; however, it seems that the remaining 3 players acted very conservatively with their launch cadences.

3.3.2 Effects of Policy Instruments on Overproduction & Profitability

To test our second hypothesis, the number of total debris produced per game and the total satellites launched per player were both measured. Figure 3.15 shows the averaged total satellites launched per player across both the REGULATED and UNREGULATED groups. Figure 3.16 shows the average total debris generated by each group. Both of these figures indicate that treatment does not seem to have any measurable impact on total satellites launched. However, Figure 3.16 indicates an initial effect from the policy treatment on the amount of total debris generated. When reviewing profitability, the control group was observed to end the study with a much higher revenue efficiency (as shown in Figure 3.17). This indicates that while the

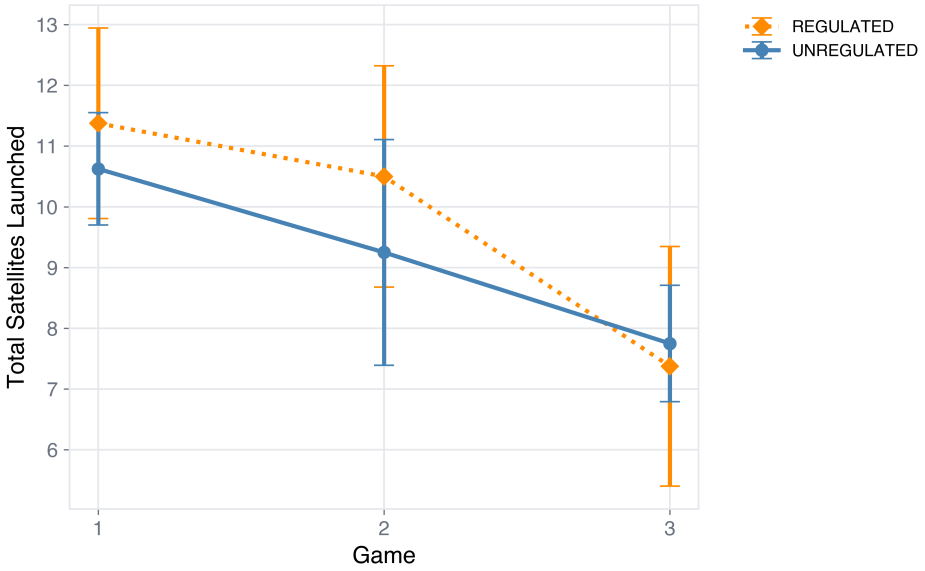

Figure 3.15: The averaged total satellites launched per player does not seem to be tied to either the unregulated control nor the regulated policy treatment.

UNREGULATED control group was more profitable, the REGULATED policy treatment group generated less total debris. A caveat to both profitability and overproduction is mentioned due to the heavy collusion employed by one of the cohorts in the UNREGULATED control group (see Section 3.3.4 for more details).

3.3.3 Effects of Policy Instruments on Debris Generation

To answer the posed research questions and determine the effects of the policy treatment, three variables were tracked across each game: (1) total congested orbit slots, (2) total debris generated from conjunctions, and (3) total debris generated from derelict satellites.

First, if an orbit slot is not congested, then additional debris due to conjunctions cannot be created. Although Figure 3.18 does not show a trend between the UNREGULATED control and the REGULATED policy treatment groups, the decreasing trend of all cohorts' total congested orbit slots indicates that players generated less congested orbit slots as they progressed in the the study. Note that the outlier in Figure 3.18 is due to heavy collusion by UNREGULATED Group 1 during their second game. Additionally, market risk can be approximated and shown to to be decreasing in these cases due to the decreasing probability of conjunctions.

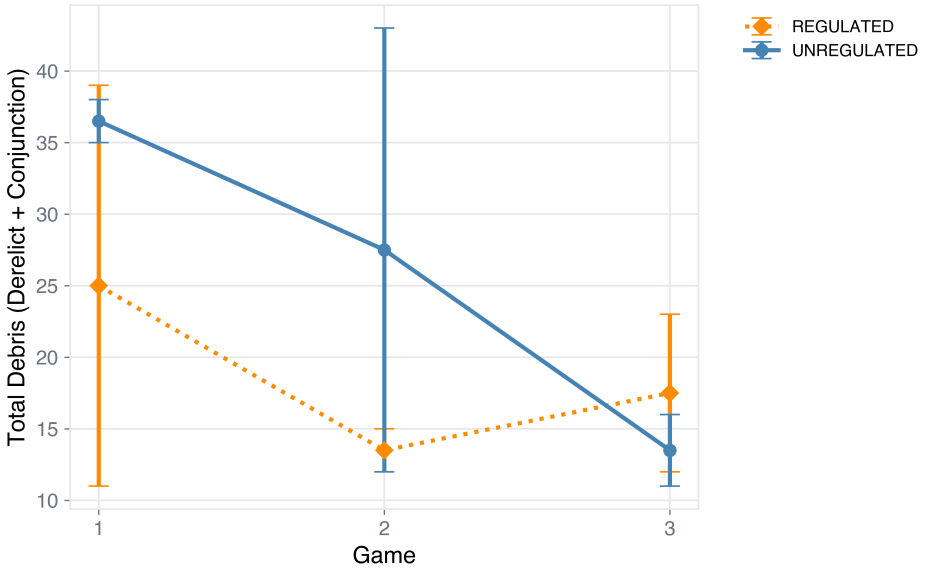

Figure 3.16: The average total debris generated by the REGULATED policy treatment seems to be less than the UNREGULATED control group.

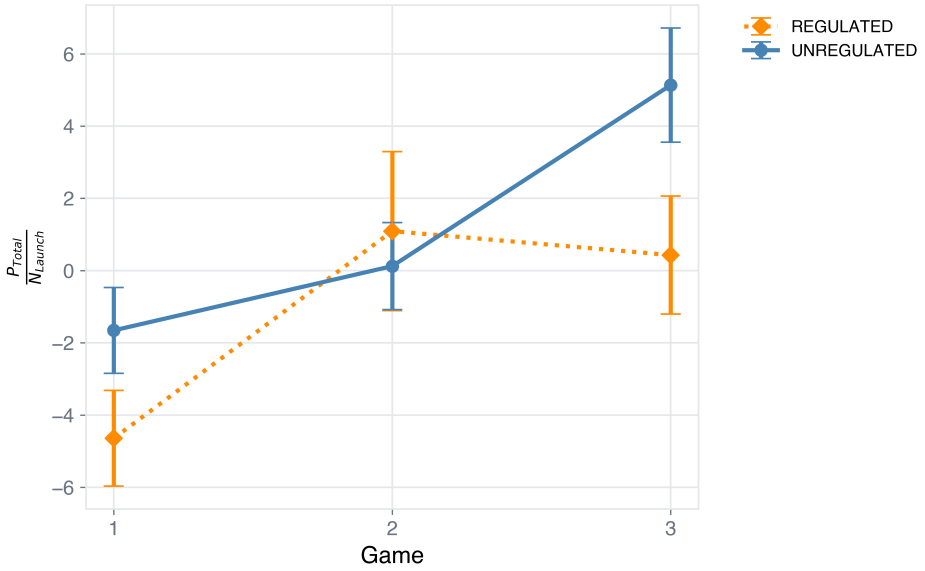

Figure 3.17: Revenue efficiency does not indicate any measurable effect due to the policy treatment.

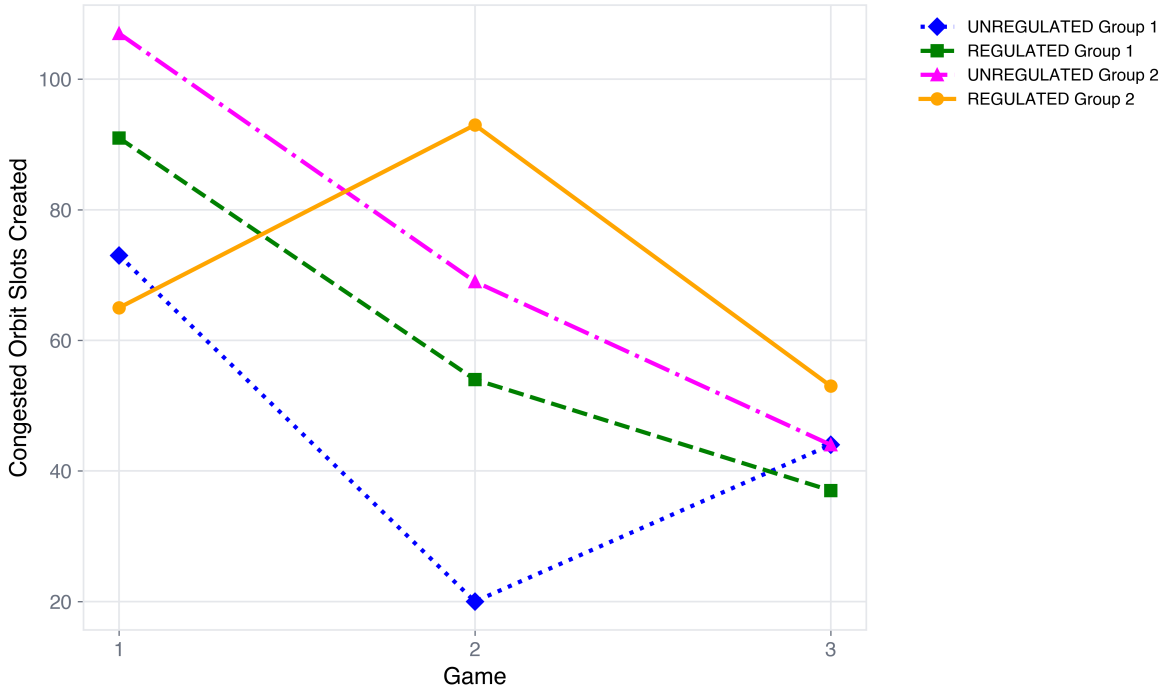

Figure 3.18: Total number of congested orbit slots per game. Policy treatment does not seem to affect congestion; however, both collusion and general experience seem to reduce the total congested orbit slots as the study progressed.

Reviewing debris generated from conjunctions in Figure 3.19, the policy treatment itself does not seem to have a measurable impact on the amount of space debris generated from collisions. This indicates little effect from the optional insurance mechanism. However, a very clear trend is shown in debris generated from derelict satellites. First, Figure 3.20 indicates that the policy treatment strongly dis-incentivized players from creating derelict satellites. This is not observed in the control group which start off with relatively high amounts of derelict satellites to eventually converging with the treatment group. As each cohort played subsequent games, the control cohorts became more cautious while the policy treatment cohorts demonstrated strategic risk-reward behavior. This indicates that the derelict debris penalty was an effective tool in decreasing initial debris generation. However, both groups seemed to learn the importance of deorbiting satellites as a part of the underlying environment dynamics and a constellation development strategy. However, when combining both debris from conjunctions and debris from derelict satellites (as shown in Figure 3.16), it is shown that: (1) debris from conjunctions is the dominating term, and (2) there is weak correlation between the policy treatment and the

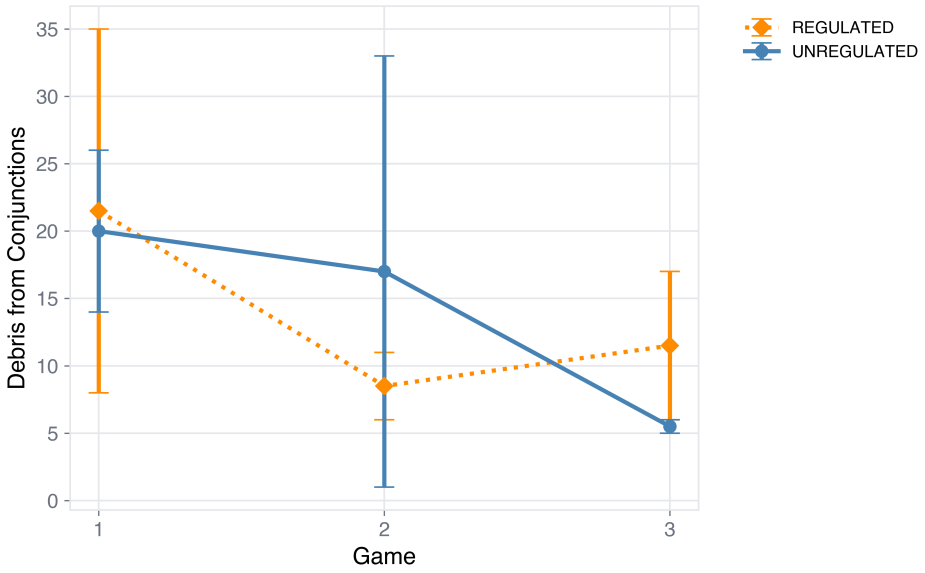

Figure 3.19: Debris from conjunctions shows no measurable impact from the policy treatment.

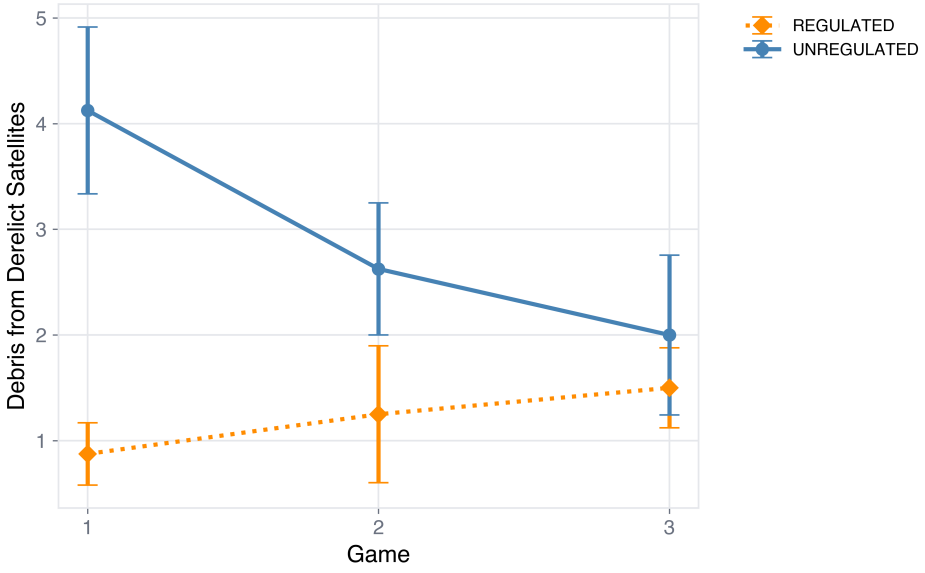

Figure 3.20: The implementation of our policy treatment shows a clear trend in the generation of debris from derelict satellites.

generation of total debris. This may be due to several factors, including imbalanced penalties, frequencies of conjunctions, or overall diversity of player strategy.

3.3.4 Qualitative Results & Strategies of Note

Throughout the pilot RCT, several unique and emergent strategies and dynamics were observed. First, and most importantly, a single cohort heavily colluded after their first game and this led to them generating both the least amount of debris and the most overall profits. Defection was strongly and swiftly condemned for the first player who suggested it, but not so for the second player. This eventually broke down into a tit-for-tat dynamic between players when capacity slots became contested. Additionally, other groups were seen negotiating, subsidizing competitors' satellites, and also creating a derivatives market among themselves as a means to self insure without having to pay for insurance costs each round. The third unique market behavior was the purposeful creation of debris to both clear a desirable orbit slot while inconveniencing an opponent. While highly unethical in the real world, this strategy worked quite well. Finally, across the course of the study multiple players began the game, made a small profit, and disengaged with the market. While not entirely accurate to the reality of the SATCOM market, this strategy proved to be profitable but not game-winning.

3.4 Discussion

As a modeling tool, the *Satellite Tycoon* board game was able to capture and map previously unexplored dynamics within the P-LEO SATCOM marketplace. The novel design of the physical game coupled with the board rotation allowed for comparative modeling of different P-LEO architectures, performances, and deployment strategies. Elements of the game such as fuel reduction and deorbiting rules also captured the end-of-life decisions and their subsequent effects on the rest of LEO. The introduction of stochastic collisions also introduced the possibility of conjunction events and their impact on a constellation's economic performance. By its very nature, the game was also able to model competition in the P-LEO marketplace, which is a largely unexplored area of research.

Data from the pilot RCT seemed to confirm our first hypothesis: As players' economic performance (final player funds and final player value) generally increased across the study, the average total number of satellites launched decreased. This indicates that players were indeed able to refine their strategies across the three games played while also becoming more revenue-efficient by launching fewer satellites and earning more revenues.

Total debris generated had weak correlation with the policy treatment and the number of total satellites launched had no correlation with the policy treatment, thus refuting our second hypothesis (that insurance fees will deter overproduction of satellites). While overproduction and profitability were not directly correlated to the policy treatment, the total satellites launched and total debris produced both decreased across the study. This can be interpreted as all players learning sustainable behavior by simply having more time to interact with the underlying dynamics of the marketplace and rediscovering the *Tragedy of the Commons*.

The first research question asked how insurance policies and fees impacted the generation of space debris. This question was separated into two variables: debris generated from conjunctions and debris generated from derelict satellites. Since deorbit penalties only affected debris generated from derelict satellites, the amount of derelict debris could be decoupled from the broader amount of debris generated. In doing so, it was found that deorbit penalties successfully incentivized more sustainable usage of LEO. Conversely, debris generated from conjunctions did not seem to be affected by the optional satellite insurance available to players. Additionally, the insurance framework itself did not explicitly lead to less space debris or less conjunctions, thus addressing our second research question.

3.5 Limitations & Validations

Both the RCT conducted in this study as well as the underlying modeling of the board game have limitations that must be stated and addressed. Within the RCT, the population size of our study was quite limited to just 16 players and any inferential analysis would have weak statistical significance. In subsequent studies, this can be addressed by further incentivizing participants and recruiting much more heavily. Furthermore, the RCT only had participants play three game

sessions. While this was over a 6 hour time commitment, additional data could be gathered by having future participants play more games to establish a more statistically significant result.

When considering the underling board game mechanics, multiple limitations exist in the modeling approach. First, the modeling of coverage and capacity is not as high-fidelity as in other works. This was a tradeoff made during the development of the game with balancing that occurred to allow novices access to the game; however, the lack of fidelity may lead to more nuanced service quality dynamics being omitted from the game mechanics. Similarly, the omission of user terminals and their production might cause similar nuanced dynamics to be abstracted due to simplicity. Finally, aspects of P-LEO satellites constellation development are not consistent between the game and reality. For instance, logistical delays and scheduled launches are currently not allowed in the game. Additionally, the total number of satellite cubes is not representative of the thousands of satellites that operators must contend with. A simple method to rectify this would be to model each satellite cube as a representation of an entire orbit plane of satellites rather than simply one satellite.

Although limitations exist in the modeling approach, two similar studies validate our use of a tabletop board game as a modeling tool within a broader RCT. First, the *Fishbanks* tabletop game shares many commonalities with *Satellite Tycoon* and has parallel structures [97]. Specifically, the idea of a shared commons with finite utility is present in both games: fish stock in *Fishbanks* and orbit slots in *Satellite Tycoon*. Additionally, both games have a possible Tragedy of the Commons scenario: overfishing in *Fishbanks* and over-populating congested orbit slots in *Satellite Tycoon*. Finally, and most importantly, both games executed policy interventions to curb the Tragedy of the Commons with learning effects and results being documented in both cases. Next, the *Beer Distribution Game*, or simply the *Beer Game* established a precedent of using a tabletop game as a controlled environment in which human subjects are randomly assigned to a treatment [98]. Similar to *Satellite Tycoon*, this work also analyzed quantitative and qualitative emergent behaviors from players. Additionally, this study also investigated learning effects across repeated play sessions.

Chapter 4

Single-Agent RL Environment for P-LEO SATCOM

An efficient method of modeling temporal dynamics is the Markov Decision Process (MDP) [99]: A mathematical framework for modeling sequential decision-making under the control of an agent. It consists of states describing the agent's current performance, actions the agent can take to explore the environment, transition dynamics that determine how the system evolves based on the agent's strategy (sequence of actions), and a reward function that gives the agent rewards or penalties for each action it takes. While others have formulated MDPs for describing smaller constellation replenishment dynamics and solved them using dynamic programming [49, 51], these strategies are very limited to smaller quantities of satellites in which a reward function is not based off of competition or economic gains. Additionally, while the use of dynamic programming offers a globally optimal solution, it suffers from the curse of dimensionality, ultimately reducing the number of satellites that can be modeled [100].

We now formulate the P-LEO SATCOM environment as a MDP, where an agent (representing a SATCOM operator) strategically develops its constellation by launching individual orbit planes (groups of satellites) and setting broadband prices to maximize net present value. Launching new orbit planes enhances coverage, while price adjustments influence customer adoption. This formulation captures the logistical and economic dynamics of constellation development, enabling systematic exploration of strategic decisions. To address the high dimensionality of the state and action spaces, we employ reinforcement learning (RL) to derive optimal policies through agent–environment interactions. The proposed framework, as shown in Figure 4.1, models complex market dynamics and supports training across varied parameterized environments to assess performance across diverse conditions. While this framework itself is

built in existing methods, the resulting environment covers dynamics and parameters that have not adequately investigated in existing literature.

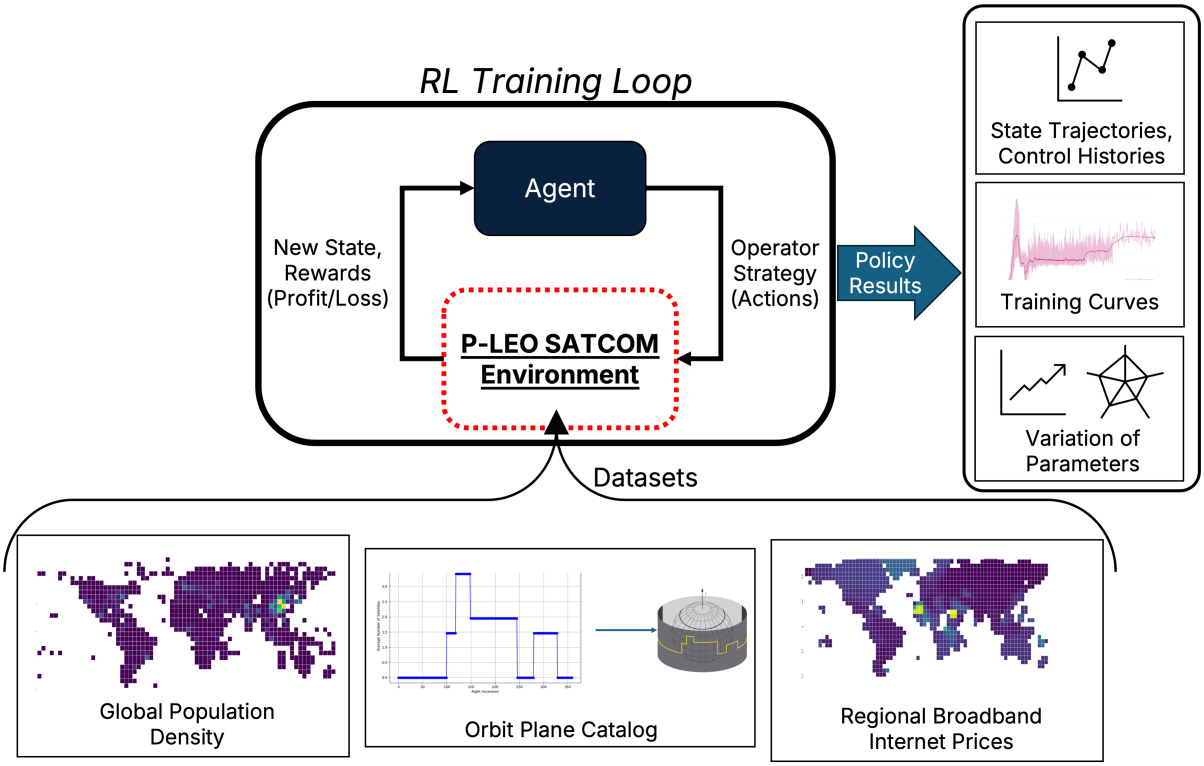

Figure 4.1: The agent-environment framework of our dynamic SATCOM system

4.1 Preliminaries

We outline the key assumptions in constellation design, customer behavior, and business modeling used in the framework. These assumptions are specifically made to capture both the temporal evolution of P-LEO constellations and their associated investment and pricing decisions. Additionally, we specify the simulation assumptions required to model the RL environment in a computationally feasible manner.

4.1.1 P-LEO Constellation Design Assumptions

We begin by abstractly defining a P-LEO constellation, denoted by $\mathcal{C}$, as a collection of orbit planes:

$$\mathcal{C} = \{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_N\}, \tag{4.1}$$

where N is the total number of orbit planes comprising the constellation. Each orbit plane, $\mathcal{P}_j$, is characterized by the tuple:

$$\mathcal{P}_j = \{i_j, \Omega_j, h_j, n_j\}, \tag{4.2}$$

where i_j denotes the inclination in degrees (any real number in the range of $[-90, 90]$), Ω_j the right ascension of the ascending node in degrees (any real number in the range of $[0, 360]$), h_j the altitude in kilometers (any real, positive number), and n_j the number of satellites within plane j (any real, positive integer). Each orbit plane is assumed to be circular, with its n satellites uniformly distributed in angular position along the orbit. An example orbit plane is shown in Figure 4.2a. Orbit planes in similar altitudes (but different inclinations or right ascensions of the ascending node) constitute an orbit shell, as shown in Figure 4.2b. The most common types of P-LEO constellations: Walker-Delta, flower, and polar orbit, can be represented by one or more orbit shells. Examples of P-LEO constellations are shown in Figure 4.2c.

A significant advantage of modeling the P-LEO SATCOM environment as a dynamical system is the ability to incorporate the orbital decay of satellites with realistic, limited lifespans. Operators are required to invest in an initial constellation development strategy but also periodically replenish satellites that decay out of orbit. While the orbit decay period of different satellite buses is variable across different systems and dependent upon infrastructure requirements, we assume a linear decay process on all satellites in the system. Given our above definition of a P-LEO constellation, we further assume that all satellites within the same orbit plane, $\mathcal{P}_j$, are launched concurrently and de-orbit at roughly the same time. This is formalized in the dynamics of our MDP formulation, shown in the next section. As an additional simplifying assumption, individual technologies such as satellite interlinks, user terminal connections, and launch vehicle optimizations are not explicitly modeled; rather, we assume comparable performance technologies throughout the simulation and focus individual constellation comparisons on geometric coverage.

As a P-LEO constellation is updated over time (by adding or removing orbit planes), the geometric figures of merit (FOMs) such as average coverage gap, maximum coverage gap, or average number of satellites in line-of-sight, at any given time, will also change. Unfortunately,

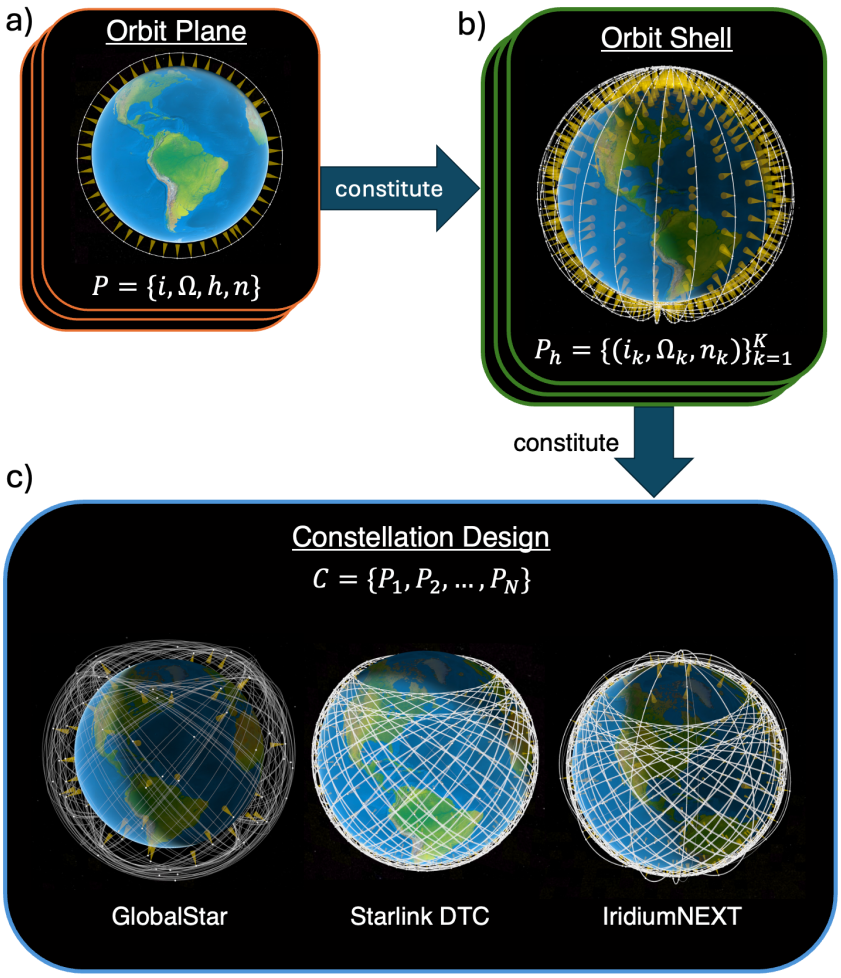

Figure 4.2: P-LEO Constellations: (a) Orbit planes are defined by the tuple, $\{i_j, \Omega_j, h_j, n_j\}$, with satellites uniformly distributed angularly along the orbit; (b) multiple orbit planes (tuples) with the same altitude constitute an orbit shell; (c) orbit shells make up different types of common constellation designs (Walker-Delta, Polar, etc.).

these FOMs (which would be much more computationally efficient to store) cannot be used to represent the state of the constellation because all active orbit planes will still be required to produce new collective FOMs each time an update to the network is made. Since FOMs cannot be used to represent the state of a constellation, orbital parameters themselves must constitute part of the agent's state. The nature of an MDP also requires that the agent's state be Markovian (i.e., the state must capture all information of the associated system without having to rely upon external information or previous states and actions). In this case, we assume an agent operating in the P-LEO SATCOM environment must save all parameters of each active orbit plane in their constellation. This is further explained in our MDP formulation.

In recent years, reusable launch vehicles have dramatically reduced the previously prohibitive costs associated with deploying large numbers of P-LEO constellations [101]. Consequently, the total cost and available launch mass of these vehicles become key considerations when developing cost models and evaluating strategies for P-LEO deployment. To simplify the strategic and logistical aspects of launch vehicle services, ride-share missions, and launch scheduling, we assume a fixed launch cost per orbit plane.

The steady access of reusable launch vehicles coupled with the accumulating space debris increases risk of collision between a P-LEO constellation and other resident space objects (including other satellite constellations). Detailed modeling of such conjunctions inside the environment would increase the computational time required to simulate such congestion and, perhaps more importantly, add a layer of stochasticity to the framework. Although MDPs are known to handle such systems well, we choose to model the environment as a deterministic system to better understand the underlying couplings between environment parameters and agent performance.

4.1.2 Assumptions on Competition Modeling

As previous generation SATCOMs have typically not been profitable investments [102], increased scrutiny is being placed on the business model of P-LEO SATCOM ventures to understand if they are robust to competition (or oligopolies), fluctuations in customer demand, and shifting use cases. Due to computational limitations of constellation management and race conditions associated with multi-agent dynamics, modeling multiple, truly independent agents asynchronously competing with one another for virtual customers quickly becomes infeasible without extensive groups of computing resources. Therefore, we describe three potential alternatives: pre-computed random action competitors, rule-based competitors, and static competitors. Pre-computed random action competitors may give interesting dynamics and deployment strategies, but would not represent any meaningful dynamics seen outside of simulation. While rule-based competitors are an appealing option, even slight variations in the environment parameters may distort strategies produced by both the agent and the rule-based competitor if not properly managed. Although static competitors are not commonly used in

practice, these types of competitors can instead be interpreted as a *target SATCOM performance rather than a true competitor*. In this regard, the agent attempting to beat this baseline competition is equivalent to finding a policy that will achieve a certain target performance across several unique metrics. Modeling such a competitor gives a qualitative indication of which strategies may prove useful to new entrants of the SATCOM industry. Additionally, static thresholds allow us to systematically explore the relative relationship between the amount of investment needed and payback period required before a new SATCOM can become profitable. To further this equivalency, we model indirect competition in the form of user acquisitions of virtual customers. Limited resources, inventory, and supply chain logistics, are abstracted and we assume the static competitor has a constellation replenishment strategy that is hidden from the agent.

4.1.3 Simulation Assumptions

Additional simplifying assumptions are made regarding individual satellite performance, launch availability, and virtual customer modeling. The agent's objective is to maximize their SATCOM profits by two interwoven, environmental mechanisms: construction of a satellite constellation network and collection of revenues from virtual broadband internet customers. However, this must be done through careful design and simplification of both the constellation design and customer modeling. Quality metrics for individual satellites such as data rate and power are assumed to be fixed throughout the simulation, thus simplifying computations for FOMs. We further assume that satellite launches are available at each epoch, given that the agent has enough funds for the launch and subsequent satellites. However, this yields an instantaneous buy-and-launch mechanism that is not seen in the real world due to safety, availability, and supply chain. We make this assumption as a means to simplify these and other detailed logistical concerns such an any potential launch scheduling or inventory management concerns that may occur.

When designing the environment, simplifying assumptions regarding customer preferences are also needed to ensure that the framework is tractable. Virtual customers seeking SATCOM broadband internet are assumed to be a specific percentage of the global population and are geographically segmented into 5° grid cells. Customer preferences are then assessed collectively

within each grid cell, i.e., all customers within a single grid cell either give their business to the agent or do not. A heatmap of these population grid cells is shown in Figure 4.3. This modeling decision avoids the computational complexity of managing the preferences of potentially millions of virtual customers and, instead, models a discrete number of regions that have non-zero populations.

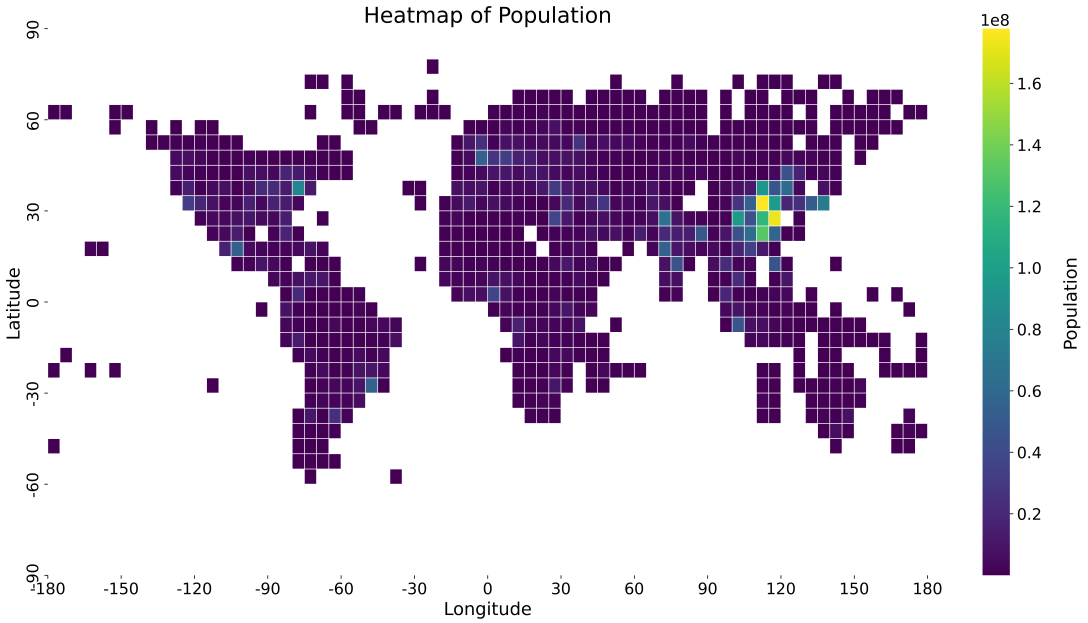

Figure 4.3: Heatmap showing each $5° \times 5°$ grid cell with non-zero populations of customers. Customers in each grid cell are assumed to be a single aggregate population.

4.2 Environment Design and Dynamics

With consideration of the above assumptions, we model our RL environment as a single-agent system in which the agents plays the role of a constellation operator attempting to maximize the net present value. We formulate the Markov Decision Process (MDP) to describe the environment as a dynamical system with sequential decision-making. The notation of our formulation (with minor modification) is taken from Chapter 2.1 of Puterman [99]. An MDP is characterized as the quintuple $\{T, S, A, p, r\}$, which represent the time horizon, state space, action space, state transition probabilities function, and reward function, respectively.

4.2.1 Time Horizon

We formulate our MDP with a finite-horizon, where T is the time horizon (or episode length) and t denotes an indexing on T:

$$t \in \{0, 1, ..., T\} \tag{4.3}$$

For each episode, we define t as months of simulation time; therefore, the agent is allowed to select business decisions on a monthly basis. While SpaceX launches of Starlink satellites have occurred with greater frequency, we note that the majority of competitors in this sector do not launch at such high frequencies yet. Additionally, while an infinite-horizon would be the best possible way to study long-term strategy evolution, this is impractical in simulation; therefore, we set the time horizon sufficiently high such that an episode encapsulates multiple generations/life-cycles of orbit planes. This necessitates the agent to replenish or reorganize its constellation as satellites decay out of orbit.

4.2.2 States and State Space

The state of the decision-maker's system at a specific epoch, t, is represented as the following vector:

$$\vec{s_t} = \begin{bmatrix} \vec{C} \\ f \\ p \\ T - t \\ M \end{bmatrix} \tag{4.4}$$

where $\vec{C}$ is itself a vector representing the remaining lifespan (x) of each orbit plane:

$$\vec{C} = [x_1, x_2, \ldots, x_N]^\top \tag{4.5}$$

Note that $\vec{C}$ is of size N, where N is the total number of possible orbit planes from a pre-computed catalog of P-LEO orbit planes. This catalog is detailed in Section 4.3.1. The agent's current funds, service price, remaining timesteps in the episode, and total active satellites in orbit

are represented by f, p, $T - t$, and M, respectively. We note that the service price is charged by the agent to its virtual customers. Additionally, we explicitly include the remaining timesteps of the episode as a state variable as it has been empirically shown to improve performance and stability of existing RL algorithms [103]. The state space S is therefore any admissible vector that is of length $N + 4$.

4.2.3 Actions and Action Space

An action available during epoch t, at state $\vec{s_t}$, $\vec{a}(s_t)$, is represented by the following vector of decision variables:

$$\vec{a}(\vec{s_t}) = \begin{bmatrix} b \\ o \\ p_{t+1} \end{bmatrix} \tag{4.6}$$

where b is a boolean to launch a new orbit plane or not, o is a specific orbit plane from the pre-computed catalog of orbit planes, and p_{t+1} is the new service price offered to virtual customers. The launch boolean, b, is a binary value the agent sets at each epoch to launch or not launch the selected orbit plane, o. If b is False, the selected orbit plane is ignored. Although the service price can be any floating point value, we set allowable discrete values to simplify analysis and action space interpretation. Although the tradespace of possible orbit plane designs is potentially limitless, a finite set of orbit planes is used to bound the possible size of the action space. Since all three of our decision variables are discrete, we may formulate the complete action space of our environment as the cartesian product of the three discrete action spaces:

$$A = \mathrm{B} \times \mathrm{O} \times \mathrm{P} \tag{4.7}$$

where B represents the decision variable to launch a new orbit plane:

$$b \in \mathrm{B} = \{0, 1\}, \tag{4.8}$$

O represents the catalog of N unique orbit planes to choose from:

$$o \in \mathrm{O} = \{O_1, O_2, \ldots, O_N\}, \tag{4.9}$$

and P represents the set of service prices the agent can feasibly charge virtual customers (ranging from 0 to 200 in increments of 10):

$$p_{t+1} \in \mathrm{P} = \{0, 10, 20, \ldots, 200\} \tag{4.10}$$

Since A is the cartesian product of discrete spaces, it can be modeled as a single, large discrete action space that encodes all feasible combinations of the action vector in one discrete value. At each epoch, the agent selects an action $a \in A$, which is then decoded into the three-component action vector $\vec{a}$. This action decoding is simply done for ease of implementation, and the components of this vector are subsequently used in the environment dynamics.

4.2.4 Dynamics and State Transition Probabilities

Given this is a deterministic system, the probability of reaching state $\vec{j}$, from state $\vec{s}$, by taking action $\vec{a}$, otherwise known as the state transition probabilities function, can be written as:

$$p_t(\vec{j} \mid \vec{s}, \vec{a}) = \begin{cases} 1 & \iff \vec{j} = L(\vec{s}, \vec{a}) \\ 0 & \text{otherwise} \end{cases} \tag{4.11}$$

where function L explicitly defines the deterministic law of motion (otherwise known as the environment dynamics). These environment dynamics describe agent actions at the level of a single timestep. Each episode constitutes a sequence of at most T timesteps, during which the agent observes the current state, takes an action, transitions to a new state, and receives a reward based off of the previous state-action pair. The mechanics executed within the environment, at every timestep, are summarized in Figure 4.4 and illustrate how an action is decoded, how constellation coverage and customer utility are evaluated, how rewards are computed, and how the next state is assembled before the process repeats for the agent.

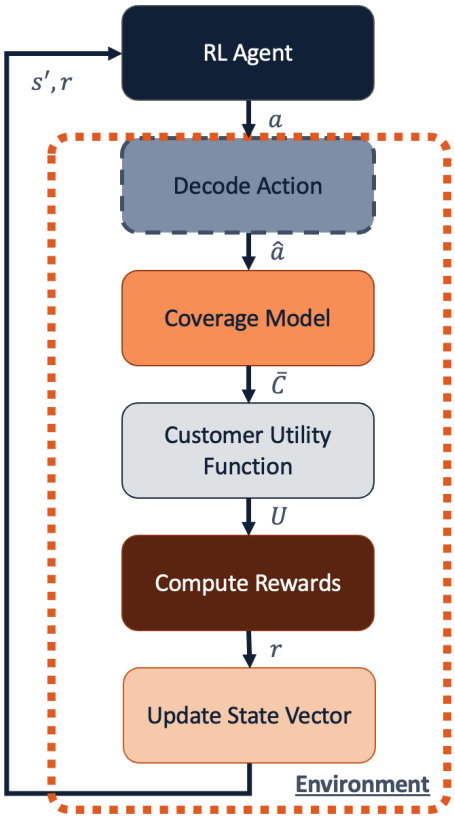

Figure 4.4: System dynamics that constitute individual timesteps of an episode are sub-divided into five components: (1) decoding the discrete action into the action vector, (2) updating the agent's constellation coverage metrics, (3) evaluating the agent's performance using a customer utility function, (4) computing rewards (profits and losses), and (5) assembling an updated state vector to pass back to the agent.

Given a discrete action, a, the size of our orbit catalog, O, and the possible levels of service price the agent is allowed to set, P, we may decode the action into an action vector. We find the launch boolean, b, using the following equation:

$$b = \left\lfloor \frac{a}{|O| \times |P|} \right\rfloor \tag{4.12}$$

Given the size of the discrete action space, A, the launch boolean will always be either 0 or 1. Similarly, we can compute the orbit plane selection, o, using the following conversion:

$$o = \left\lfloor \frac{a}{|O|} \right\rfloor \mod |P| \tag{4.13}$$

Finally, the new service price p_{t+1} is given by:

$$p_{t+1} = 10 \times (a \mod |P|) \tag{4.14}$$

A core challenge of modeling P-LEO constellation development strategies is reconciling (relatively short) orbit periods of individual satellites in LEO with the much slower time scales over which strategic decisions produce economic effects. While orbit periods in LEO are approximately 90 minutes, strategic actions made by operators (such as launching multiple orbit planes) might require several months to influence market share or revenues. We address this time scale issue by computing right ascension coverage masks that are meant to estimate an average day. As with customer grid cells, latitude bands are created at 5° increments, thus generating 36 evenly spaced bands across the 180° latitude. For every orbit plane, o, in the catalog O, 36 coverage masks (one per latitude band) are pre-computed and stored in memory. An individual coverage mask over a single latitude band is shown in Figure 4.5, and the complete algorithm to generate right ascension coverage masks from orbit plane parameters is given in Appendix A.

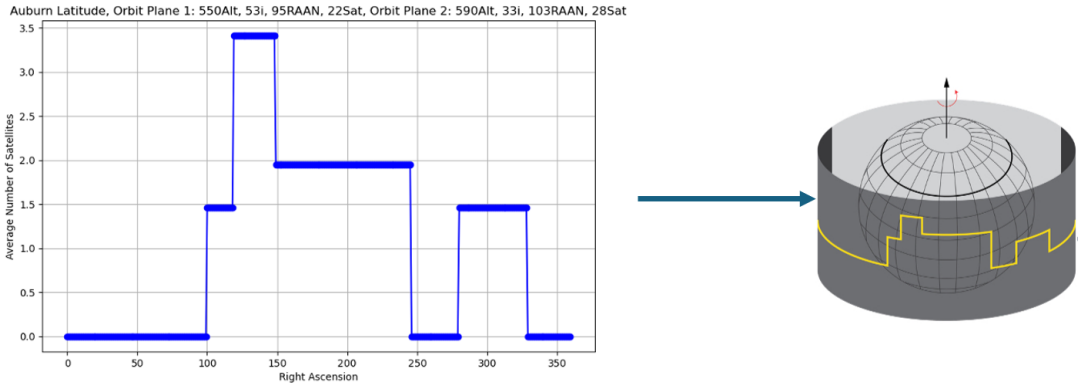

Figure 4.5: A visual representation of the total right ascension coverage mask of two orbit planes. The latitude of Auburn, Alabama is used as an example over which the coverage mask is evaluated. It is presented as the average number of satellites visible across the right ascension of the ascending node ($[0, 360]$).

Using the constellation lifespans, $\vec{C}$, from the state vector, active orbit planes are identified, coverage masks for each orbit plane are pulled from catalog O, and total coverage masks are generated across latitude bands. If the agent chooses to launch a new orbit plane of satellites, the corresponding coverage masks at all altitudes are also pulled from the catalog and added to

the total coverage mask. A proof for the additivity of right ascension coverage masks is given in Appendix B.

The total coverage masks across each latitude band are then used to compute coverage FOMs for that specific latitude band. Total coverage gap (TCG), maximum coverage gap (MCG), and average coverage gap (ACG) are all computed, but only TCG is used in the final utility function to ensure the utility is not overly constraining. The average constellation altitude, h, is also computed based off a weighted average of satellites at each altitude and used as a comparative proxy for latency estimation. The collection of coverage FOMs for all latitude bands is denoted as $\bar{C}$.

Once the agent's constellation FOMs are computed, they are passed to the virtual customer utility function to compute the number of total customers awarded to the agent at the current timestep. The utility function for this work is a threshold-based model that encourages the agent to hit target performance metrics and surpass them through RL. We define the total number of customers served by the agent's satellite internet service as follows:

$$U = \sum_{k=1}^{D} \text{Subscribers}_k \cdot 1(TCG_k, h, p) \tag{4.15}$$

where D is the total number of grid cells, Subscribers_k is the number of active subscribers to the agent's internet services in region k, and $1(TCG_k, h, p)$ is an indicator function that checks the thresholds for each performance metric used:

$$1(TCG_k, h, p) = \begin{cases} 1, & (TCG_k < \alpha) \wedge (h < \beta) \wedge (p < \gamma_k) \\ 0, & \text{otherwise} \end{cases} \tag{4.16}$$

where TCG_k is the total coverage gap produced by the agent's constellation over region k, α is the maximum total coverage gap threshold, h is the average altitude of the agent's constellation, β is the maximum altitude threshold, p is the service price from the agent's state vector, and γ_k is the price threshold specific to region k. The service price, p, is taken as the agent's current service price, p_t; the updated value from the agent's action vector, p_{t+1}, is set after revenue

computations as part of the state vector update. Note that α and β are global thresholds, but the underlying constellation FOMs (namely, TCG_k) are variable based on latitude, i.e., the total coverage mask is not guaranteed to be the same across multiple latitude bands.

While the reward function of our MDP is a portion of the dynamics, it is more formally defined in the next sub-section. For purposes of state updates, we may simply say that it is the cash flow per epoch:

$$r_t(\vec{s}_t, \vec{a}_t) = CF_t \tag{4.17}$$

The updated state variables in timestep $t+1$ following the agent's action $\vec{a}(\vec{s})$ and subsequent environment dynamics in timestep t are given as:

$$C(j)_{t+1} = \begin{cases} x_{max} & \text{if} \quad (b = 1 \wedge o = j) \\ \max(0, C(j)_t - 1) & \text{otherwise} \end{cases}, \forall j \in \{1, \ldots, N\} \tag{4.18}$$

$$f_{t+1} = f_t + r_t(\vec{s}_t, \vec{a}_t) \tag{4.19}$$

$$p_{t+1} = p_{t+1} \quad \text{(from the action vector)} \tag{4.20}$$

$$(T - t)_{t+1} = (T - t) - 1 \tag{4.21}$$

$$M_{t+1} = \sum_{j=1}^{N} n_j \cdot \begin{cases} 1 & \text{if} \quad C(j)_{t+1} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.22}$$

where x_{max} represents the maximum lifespan an orbit plane may have in LEO, N is the total number of orbit planes in the catalog, and n_j is the number of satellites in orbit plane j.

Each episode is also bound by termination and truncation conditions. An episode terminates upon reaching the final epoch, T, while it is truncated if the agent goes bankrupt, i.e., when $f_{t+1} \leq 0$.

4.2.5 Reward Function

We now detail how the reward function from Equation (4.17) is formally defined and computed. First, cash flow can be broken into revenues and costs for epoch t:

$$r_t(\vec{s}_t, \vec{a}_t) = CF_t = \text{revenues}_t - \text{costs}_t \tag{4.23}$$

where revenues are simply computed using the agent's total acquired customers, U_t, and the agent's current service price, p_t:

$$\text{revenues}_t = U_t \cdot p_t \tag{4.24}$$

Costs at each epoch are computed as:

$$\text{costs}_t = \begin{cases} C_L + C_R = C_L + (M_t \cdot K) & \text{if} \quad b = 1 \\ C_R = M_t \times K & \text{otherwise} \end{cases} \tag{4.25}$$

where C_L is the fixed cost to launch an orbit plane, C_R is the recurring cost to maintain the overall constellation infrastructure and company, K is the recurring cost per active satellite, and M is total number of active satellites taken from the agent's state vector. Note that C_L corresponds to the entire cost of an orbit plane's deployment (including the satellite cost itself). It is also assumed that recurring costs scale linearly with the total number of satellites the agent has in orbit.

4.2.6 Bellman Optimality & Net Present Value

A key advantage of formulating the P-LEO SATCOM environment as an MDP is that it enables RL to approximate a near-optimal policy π^* for the dynamics of the system using Bellman Optimality [104]. A policy is composed of decision rules:

$$\pi^* = \{d_0, d_1, \ldots, d_{T-1}\} \tag{4.26}$$

where each decision rule is a function that takes the state at the current epoch as input, and outputs an action:

$$\vec{a}_t = d_t\left(\vec{s}_t\right) \tag{4.27}$$

The Bellman Principle of Optimality states that an optimal policy from time t onward must select actions that maximize the sum of immediate rewards and expected future returns. Formally, the total discounted return over an entire trajectory is:

$$G_0 = \sum_{t=0}^{T-1} \gamma^t r_t(\vec{s}_t, \vec{a}_t), \tag{4.28}$$

where the discount factor, γ, can be related to the financial discount rate, R, via $\gamma = 1/(1+R)$. If rewards correspond to cash flow, $r_t = CF_t$, (as in our reward function) and the discount factor applied during training is the same as applied during evaluation, then the total discounted return over the full trajectory can be rewritten as:

$$G_0 = \sum_{t=0}^{T-1} \frac{CF_t}{(1+R)^t} = \sum_{t=0}^{T-1} PV_t = NPV, \tag{4.29}$$

where PV_t is the present value at epoch t and NPV is the net present value. Equation (4.29) shows that following an optimal policy maximizes the total discounted return, which is also the NPV of the trajectory.

4.3 Experimental Setup

Given the complete MDP formulation, this section now details the different datasets utilized throughout training, explains the training setup with convergence criteria, and describes the economic performance metrics used to evaluate a learned policy. Environment variables are also given with rationale behind each value.

4.3.1 Datasets

As mentioned in the MDP formulation, a catalog of orbit planes was constructed using parameters drawn from existing and proposed LEO constellation architectures. For each orbit plane in

the catalog, right ascension coverage masks were pre-computed over the 36 uniformly-spaced latitude bands. This catalog serves two purposes: (1) it provides a tractable and standardized representation of constellation configurations, and (2) it enables substantial computational savings by avoiding repeated coverage computations during training. A unique challenge in modeling P-LEO constellation designs is that orbit planes evolve over time: older satellites decay out of service, and new satellites are introduced in discrete, staged, deployments. If the state vector were to grow or shrink dynamically as orbit planes are added or removed, the resulting state representation would violate the structure of an MDP's state space, rendering it ill-defined. While common implementation techniques such as zero-padding or masking could enforce a uniform state dimension, these approaches still require coverage computation of each orbit plane and subsequent FOM evaluations at every timestep, thus incurring significant computational overhead when training. Instead, by using a catalog of orbit planes with pre-computed coverage masks, we maintain a fixed and well-defined state space while also eliminating the need for online coverage computation. The training cost associated with on-demand coverage computation is:

$$O(GpET^2) \tag{4.30}$$

where G is the number of latitude bands, p is the probability that the agent will launch an additional orbit plane at each epoch, E is the total number of training episodes, and T is the maximum number of timesteps per episode. Alternatively, the training cost associated with pre-computed coverage computation is:

$$O(NG + ET) \tag{4.31}$$

where N is an additional variable that represents the size of the orbit plane catalog. As the number of training episodes grow very large (as required by most RL training), the cost to pre-compute becomes negligible while the on-demand cost remains quadratic due to the ongoing cost of computing coverage at each timestep.

The full orbit catalog is drawn from two sources and compiled in Table 4.1. The first source is a technical survey which provides performance and orbital configuration summaries for select commercial LEO broadband constellations [21]. The second source is a comprehensive commercial satellite database maintained by Jonathan McDowell [105]. McDowell tracks the most up-to-date launches of all major commercial satellite constellations.

Data from both sources are aggregated into orbit shells, represented as tuples of the form: $\{i, Y, h, n\}$, where i denotes the common inclination of all orbit planes within the shell, Y is the number of orbit planes in the shell, h is the altitude of all orbit planes within the shell, and n is the number of satellites per orbit plane. Within each shell, the orbit planes are evenly distributed in right ascension of the ascending node (RAAN). Specifically, the RAAN of plane j is given by:

$$\Omega_j = \frac{2\pi}{Y}\, j, \quad j = \{1, \ldots, Y\}, \tag{4.32}$$

resulting in a uniform angular separation of:

$$\Delta\Omega = \frac{2\pi}{Y} \tag{4.33}$$

between adjacent orbital planes. Across all shells, a total of 1139 orbit planes were generated, collectively forming the orbit plane catalog.

Table 4.1: P-LEO orbit shells used to generate the orbit plane catalog (Updated Nov. 13, 2025).

Orbit Shell	h **(km)**	i **(deg)**	**Orbit Planes**	**Satellites/Plane**
Starlink Group 1	550	53.0	72	22
Starlink Group 2	570	70.0	36	20
Starlink Group 3	560	97.6	6	58
Starlink Group 3.5	560	97.6	4	43
Starlink Group 4	540	53.2	72	22
Starlink Group 5	530	43.0	4	43
Starlink 2A V2M-1	525	53.0	28	120

Orbit Shell	h **(km)**	i **(deg)**	**Orbit Planes**	**Satellites/Plane**
Starlink 2A V2M-2	523	43.0	28	120
Starlink 2A DTC	535	53.0	28	89
Starlink 2-1	340	53.0	48	110
Starlink 2-2	345	46.0	48	110
Starlink 2-3	350	38.0	48	110
Starlink 2-4	360	96.9	30	120
Starlink 2-5	530	43.0	28	30
Starlink 2-6	525	53.0	28	30
Starlink 2-7	535	33.0	28	30
Starlink 2-8	604	148.0	12	12
Starlink 2-9	614	115.7	18	18
OneWeb Gen-1	1200	87.9	36	49
OneWeb MEO-1	1200	55.0	32	72
OneWeb MEO-2	1200	40.0	32	72
Amazon Kuiper 1	590	33.0	28	28
Amazon Kuiper 2	610	42.0	36	36
Amazon Kuiper 3	630	51.9	34	34
Telesat LEO 1	1015	98.98	27	13
Telesat LEO 2	1325	50.88	40	33
Guangwang 1	590	85.0	16	30
Guangwang 2	600	50.0	40	50
Guangwang 3	508	60.0	60	60
Guangwang 4	1145	30.0	48	36
Guangwang 5	1145	40.0	48	36

Orbit Shell	h **(km)**	i **(deg)**	**Orbit Planes**	**Satellites/Plane**
Guangwang 6	1145	50.0	48	36
Guangwang 7	1145	60.0	48	36

To generate virtual customers in the environment, we incorporate the *Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11* dataset [106] which provides global population density estimates at multiple resolutions for the years between 2000 and 2020. GPWv4 uses a combination of distributed national and international administrations that collect and report census data. Population density rasters are computed by dividing the gridded population counts by estimates for each individual cell. While GPWv4 has data in many resolutions (up to 1° cells), the native data is not available in larger formats. To maintain computational tractability during training, we aggregate the GPWv4 rasters into 5° grid cells while preserving the total population within each original region. We then introduce a population adoption rate, κ, to estimate the number of active subscribers, from the overall population, within each grid cell:

$$\text{Subscribers}_k = \kappa \cdot \mathcal{N}_k \quad \forall k \in D \tag{4.34}$$

where $\mathcal{N}_k$ is the population in grid cell k and D is the total number of grid cells. The resulting active potential customer dataset balances fidelity and efficiency while modeling realistic global market demand. A heatmap of this subscriber population grid cells was shown in Figure 4.3.

The final dataset we utilize is global broadband pricing data [107] to derive threshold price per month of broadband internet services, γ_k, for each grid cell that is used in the utility function, Equation (4.16). This dataset reports the average broadband service cost to consumers by country, rather than by geographic grid cells used in our framework. To translate these national prices into our grid cell threshold prices, we identify the country that is at the geometric center of each grid cell (by latitude and longitude) and assign that country's average broadband price as the grid cell's price threshold. These thresholds represent a maximum price the agent must remain below to provide a competitive option in the region. A heatmap of these price threshold grid cells is shown in Figure 4.6.

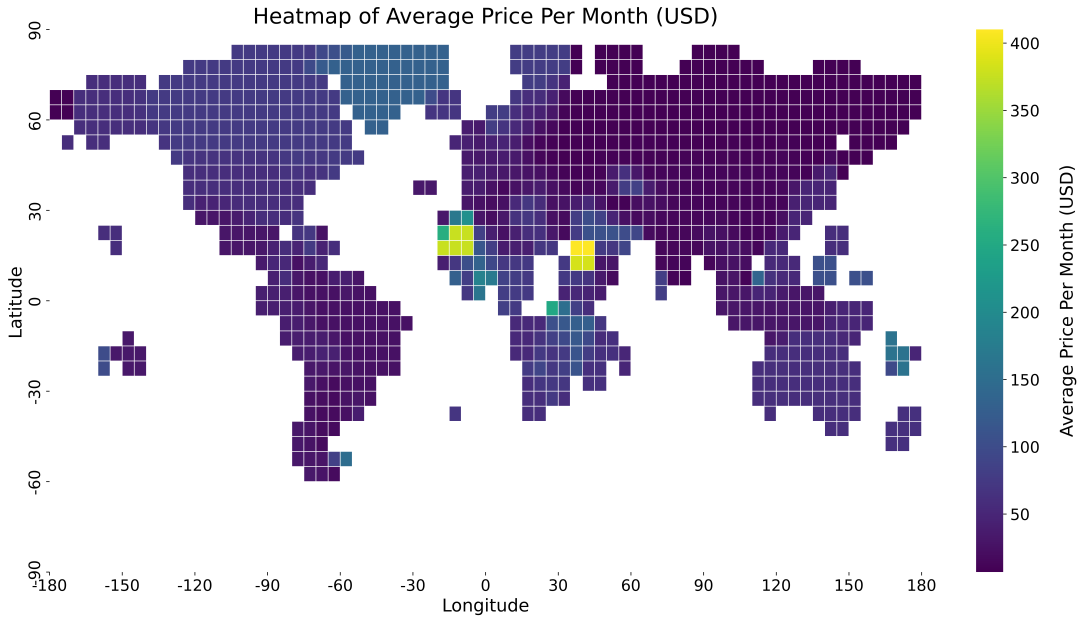

Figure 4.6: Heatmap illustrating each $5° \times 5°$ grid cell with price thresholds. These thresholds simulate competitive benchmarks that the agent must beat in order to attract customers.

4.3.2 Training Setup

To develop and train RL policies, the MDP formulation was implemented in the Farama Foundation's Gymnasium [108] framework, with the Stable-Baselines3 [109] library used for algorithm implementations. Gymnasium was chosen due to its broad adoption within the RL research community and its standardized interface, which promotes reproducibility and compatibility with several RL agent implementations. Stable-Baselines3 served as the algorithm library, providing an industry-standard set of implementations that are well-validated within the RL community.

Using the Stable-Baselines3 library, several different algorithms were tested in the environment, each with their own hyper-parameters. Value-based methods such as Deep Q-Network (DQN) learn a greedy, deterministic policy, but explore stochastically during training (using ϵ-greedy approaches). Policy-gradient methods such as Proximal Policy Optimization (PPO) or Advantage Actor-Critic (A2C) learn a stochastic policy during training by sampling actions from the evolving probability distribution $\pi(a|s)$. However, during evaluation, we execute the learned policy deterministically by selecting the action with the highest probability or Q-value (action-value). Consequently, during the policy-evaluation phase, the return, G_0, can be simplified to the realized sum of discounted rewards (cash flows) along the deterministic trajectory, without the need to take an expectation over stochastic actions.

For each algorithm, default Stable-Baselines3 hyper-parameters were largely retained, with the exception of learning rate and convergence criteria. The learning rate for all experiments was fixed at 10^{-6}; higher learning rates were also tested but led to unstable training and overshooting. To ensure that the reported results were indicative of fully converged policies, training was conducted over at least 25 million episodes. After this minimum episode count, training proceeded until no further improvement in evaluation performance was observed over a subsequent window of 10 million episodes, at which point training was automatically terminated and the policy was evaluated a final time. To account for stochasticity in training and environment dynamics, each experiment was repeated 10 times using distinct, random seed values. The resulting training curves and learning curves were then aggregated and the best learned policy was used for evaluation.

All experiments were executed on the Auburn University Easley High Performance Computing (HPC) cluster, with training parallelized across homogeneous compute nodes. Each node was equipped with dual Intel Xeon Gold 6248R processors (24 cores per CPU, 48 cores total) operating at a base frequency of 3.0 GHz, with 192 GB of system memory. All nodes ran a 64-bit (x86_64) Linux operating system. While we conducted a thorough variation of parameters, specific environment and training variables for the baseline configuration, as well as the initial conditions, are given in Appendix C.

4.3.3 Economic Performance Metrics

Once a policy was trained, a final environment evaluation was completed. State trajectory and control history plots were generated from this final evaluation episode. Using the state trajectory, traditional economic FOMs were computed. Specifically, using the funds state variable, f, and the net present value (from Equation (4.29)), G_0, we are able to approximate the total profits over the time-horizon (P_{total}), the time-to-profit (TTP), the compound monthly growth rate ($CMGR$), and the compound annual growth rate ($CAGR$). Total profits over the time horizon are computed as the difference between the initial funds and the net present value of investment:

$$P_{total} = G_0 - f_0 \tag{4.35}$$

Time-to-profit is computed simply as the epoch in which the agent's state of funds crossed above the initial funds:

$$TTP = \min_t \{t \mid f_t > f_0\} \tag{4.36}$$

If TTP returns a null value, then a profit was not successfully returned. Although economic investments are typically considered on a yearly basis, our simulation epochs are monthly. Therefore, the compound monthly growth rate, given as a percentage, is computed as:

$$CMGR = \left(\left(\frac{f_T}{f_0} \right)^{\frac{1}{T}} - 1 \right) \times 100 \tag{4.37}$$

where f_T are the agent's funds at the end of the episode and f_0 are the agent's initial funds. To supplement the $CMGR$, we also compute the compound *annual* growth rate and give it as a percentage:

$$CAGR = \left(\left(\frac{f_Y}{f_0} \right)^{\frac{1}{Y}} - 1 \right) \times 100 \tag{4.38}$$

where

$$Y = \frac{T}{12} \tag{4.39}$$

Both $CMGR$ and $CAGR$ are effectively average growth rates of an agent's constellation deployment strategy across a single episode. Therefore, both metrics are indicators of long-term performance trends but cannot give insight at a more granular level. Since both values are also computed as percentages, they can also be used to quantitatively compare different business strategies and investment decisions while accounting for compounding.

4.4 Results & Analysis

Having fully defined the environment and training parameters, we investigate the utility function thresholds, compare different RL algorithm performances, and plot an example state trajectory and control history. The environment is systematically explored by performing a sweep across several parameters and evaluating the retrained agent's economic performance on each of these new sets of conditions. This allows us to better understand the sensitivity of each parameter. These experiments are aggregated to show how the economic FOMs are affected by the variations to the environment.

4.4.1 Baseline Determination & Evaluation

Because the utility function is based on static thresholds, careful selection of these environment parameters is critical, particularly in the absence of real-world data to validate against. Specifically, the total coverage gap threshold, α, expressed as a percentage of the total coverage mask, must be chosen sufficiently low to

ensure a realistic level of service quality for which an end user would be willing to pay. Conversely, setting α too low can lead to reduced coverage and the need for the RL agent to design an excessively costly satellite network to compensate for this drop in coverage. The maximum altitude threshold, β,(which is used as a proxy for a latency constraint) must be set sufficiently low to prevent large communication lags and degraded quality. However, lower altitudes reduce the instantaneous coverage area of each satellite, thereby increasing the overall number of satellites required to maintain adequate coverage. To properly account for the competing objectives in both thresholds, a sweep was conducted. For each pair of thresholds, a new environment was created and a newly initialized RL agent was trained on the environment. The net present value (NPV), G_0, was then computed following a final evaluation episode. Figure 4.7 shows a surface plot of G_0 against both thresholds. Using this information, we identified the corresponding thresholds that were most likely to yield a high G_0 while still maintaining some level of operational difficulty for the RL agent to overcome: 40% for α and 1000 for β.

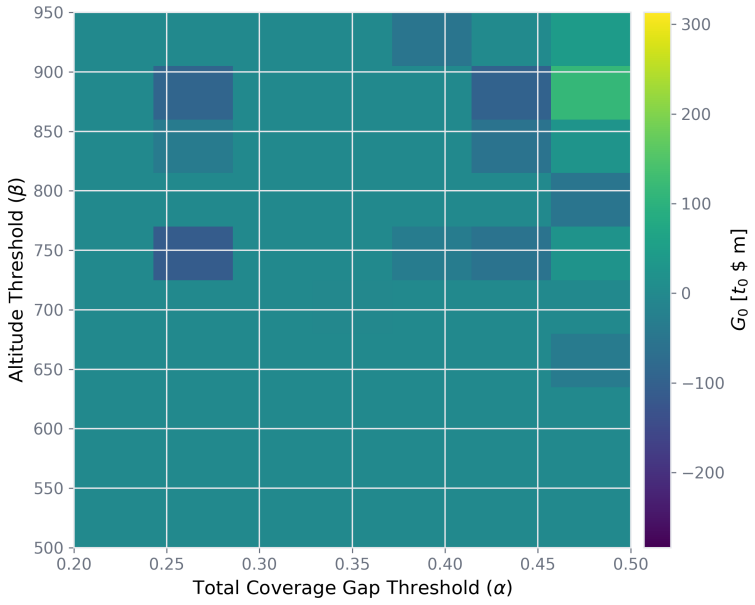

Figure 4.7: Both thresholds, α and β, were explored using grid search techniques to identify a combination that was both suitably difficult for the RL agent and also led to a positive net present value (G_0). In total, 70 RL agents were trained across each threshold configuration.

Following threshold exploration and determination, all parameters were fixed to establish a baseline environment. These values are given in Appendix C. A comparison of unmodified StableBaselines3 algorithms was conducted to determine which was best suited for all further experiments and training. The mean episodic reward, averaged across a window of 100 previous episodes and shown in Figure 4.8, highlights that the DQN agent was able to learn a policy that produced positive rewards. The PPO

and A2C agents were not able to efficiently learn any policies that maximize reward. Hyper-parameter tuning and custom network architectures could result in further learning improvements for all algorithms; however, the default DQN algorithm was used in this study. Figure 4.8 further illustrates the DQN agent's gradual shift from exploration to exploitation and confirms convergence on a positive episodic reward. It should be noted that due to the nature of stochastic optimization, not all training seeds converged in the same amount of time.

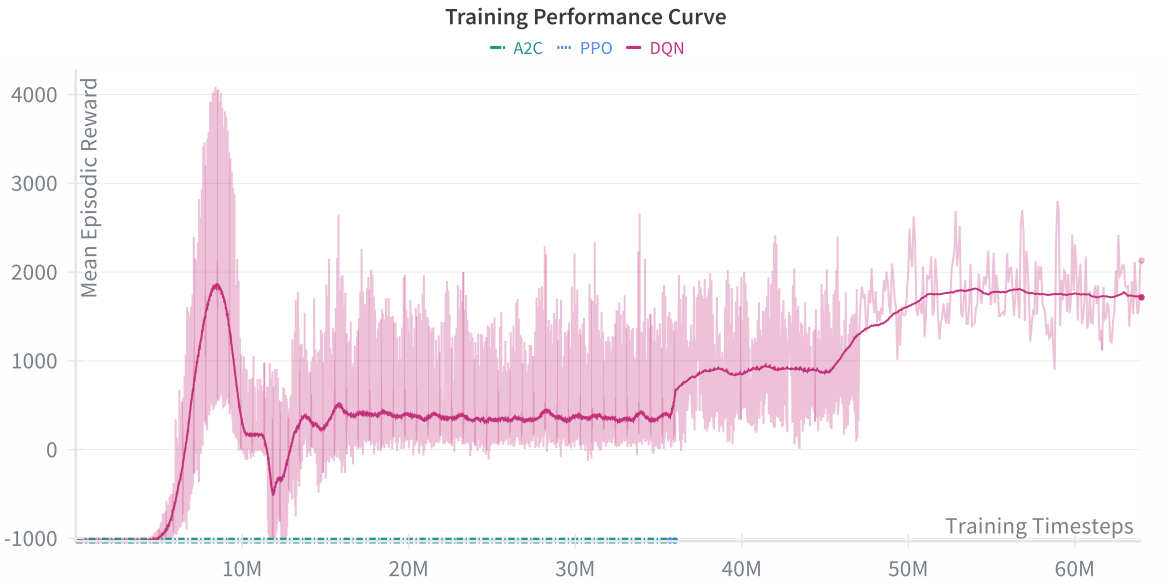

Figure 4.8: Training curves of RL algorithms on the baseline environment configuration. Smoothed and averaged values are shown with minimum and maximum ranges. Note that this plot summarizes 30 different training instances (10 randomly seeded instances for each algorithm).

Convergence is also dictated by the agent's continued improvement in the evaluation environment and the decrease in total loss as training progresses. Training continues until agent performance in the evaluation environment ceases to improve across 10 million timesteps. Additionally, Figure 4.9 shows the total loss of the DQN agent as a function of training timesteps. In both Figure 4.9 and Figure 4.8, large fluctuations at the beginning of training indicate an exploration phase followed by an asymptotic decrease in loss value and a corresponding plateauing of mean episodic reward. Taken together, these trends provide compelling evidence that the DQN algorithm converges to a learned policy.

The final evaluation episode of the DQN experiment with the highest G_0 is used to generate state trajectory and control history plots, as shown in Figure 4.10. The constellation lifespan vector, $\vec{C}$, is not visualized due to its high dimensionality; however, the total active satellites, M, captures a snapshot

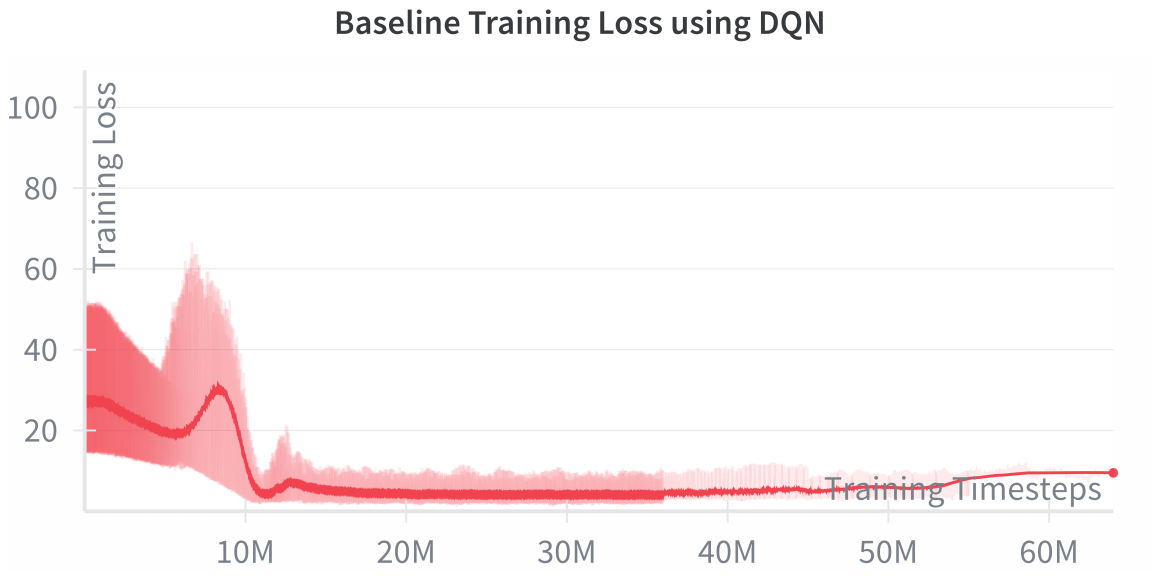

Figure 4.9: Loss over the total training timesteps for the DQN agent on the baseline environment configuration. Smoothed and averaged values from 10 DQN experiments are shown with minimum and maximum ranges.

of how many overall satellites are contributing to the agent's constellation at any given timestep. As shown in the state trajectory, the agent initially expends funds to deploy multiple orbit planes. This initial investment is then rapidly recouped in subsequent profit-generating timesteps. However, as the episode continues and satellites decay out of orbit, replenishment dynamics emerge and the agent attempts to deploy more satellites to improve its diminished constellation performance. Note that although the agent's policy has converged, sub-optimal actions such as oscillating service prices and inadequate satellite replenishment are still taken by the agent.

Using the state trajectories of all 10 DQN experiments, economic FOMs are extracted and averaged to give a full, quantitative snapshot of the trained policies on the environment. These FOMs serve as benchmarks against which alternate agents and environment configurations can be evaluated. Table 4.2 presents the consolidated summary of these metrics. Note that units of [(t_0) $ m] represent the economic value in month t_0 dollars, in millions. This is typically done in economics as a means of discounting later costs and rewards for investment strategy purposes. Additionally, we note that the total profits, when discounted to t_0 dollars, are observed to be a negative value. This is due to the relative value computed using the net present value. We note that a policy in which no investment action is taken would result in a net present value of zero, and therefore, a total profit of -1000 due to the opportunity cost imposed by the time value of money. The time to profit (TTP) is simply the number of timesteps needed to exceed the

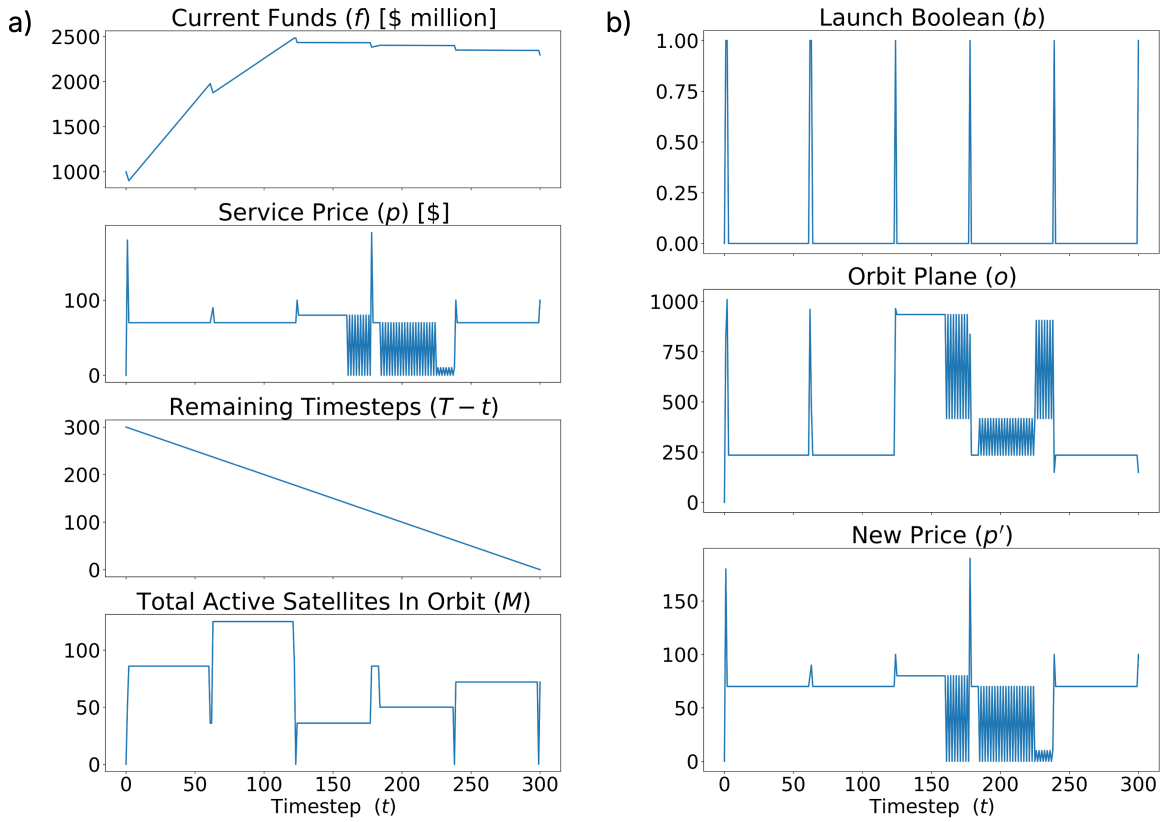

Figure 4.10: State trajectory (a) and control history (b) of the trained DQN agent achieving the highest G_0 in the final evaluation episode of the baseline environment.

initial funds, f_0. This can be readily visualized from the state trajectory as the first timestep in which f goes above the initial value. While net present value is a measure of absolute value generated with considerations for the time value of money, both the compound monthly growth rate and compound annual growth rate can be used to measure an agent's growth efficiency across a specific period. A policy which produces lower total profit with a relatively, high $CMGR$ may be preferable when long-term growth and scalability are prioritized. Both compounding growth rates are further used to compare the performance of the DQN agent across different environment and training configurations.

Table 4.2: The economic FOMs from DQN performance of 10 randomly seeded experiments.

Performance Merit	$\mu \pm \sigma$
G_0 [(t_0) \$ m]	14.515 ± 68.965
P_{total} [(t_0) \$ m]	-985.485 ± 68.965
TTP [months]	28.250 ± 27.515
$CMGR$ [%]	0.118 ± 0.148
$CAGR$ [%]	1.438 ± 1.810

4.4.2 Variation of Environment Parameters

Once a baseline was established, the parameterized environment was systematically adjusted within ranges predetermined from feasibility, historical values, and current technologies. Table 4.3 describes the environment parameters that are varied. Each parameter was varied independently in a one-at-a-time manner: all parameters were held constant while a single parameter was modified, a new environment was generated, and a RL agent was trained on this altered environment. Parameters whose variation significantly impacted the economic FOMs, specifically the $CAGR$, were identified as key drivers of the environment dynamics. Episode length, T, was varied to explore the effects of shorter or longer time horizons relative to the 25-year baseline, investigating whether extreme horizon lengths produce emergent behaviors. Maximum satellite lifespan, x_{max}, was chosen to examine how the frequency of replenishment affects the agent's ability to learn replenishment dynamics. Launch cost, C_L, and the recurring cost multiplier, K, were included because they are both primary drivers for commercial space mission feasibility, allowing exploration of the wide range of starting conditions that commercial operators can begin from. Finally, the service population multiplier, κ, was varied to assess how changes in market size influence the ease or difficulty of maintaining operational sustainability. All results are reported as averaged over 20 independent, random seeds per environment configuration, with uncertainty shown as standard error of the mean (SEM), comprising 700 experiments across both training and evaluation.

Table 4.3: Environment Parameters to vary with ranges

Parameter	**Range of Values**
Episode Length (T)	100, 200, 300, 400, 500, 600
Max Satellite Lifespan (x_{max})	24, 36, 48, 60, 72, 84
Launch Cost (C_L)	30, 50, 70, 80, 90, 100
Service Population Multiplier (κ)	[2, 4, 6, 8, 10]$\times 10^{-4}$
Recurring Cost Multiplier (K)	[0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]$\times 10^{-3}$

Varying individual environment parameters revealed several distinct trends in net present value, G_0. As shown in Figure 4.11a, the service population multiplier, κ, exhibited strong sensitivity with values below 8×10^{-4} (corresponding to approximately 0.08% of the global population), yielding either zero or negative returns. The exponential growth of G_0 indicates the existence of a potential market threshold in which entry by a new operator only becomes economically feasible after a certain percentage of the global population are willing to adopt broadband satellite internet. Although counterintuitive, episode

length, T, displayed an inverse relationship with G_0, as shown in Figure 4.11b. Longer time-horizons yielded lower net present value, a behavior consistent with discounted MDP theory: rewards that are obtained at later timesteps contribute negligibly to present value despite potentially strong longer-term performance. Figure 4.11c shows that varying the maximum satellite lifespan, x_{max}, also seemed to have a potential market threshold. Returns seem to become positive as the lifespan (and utility) of satellites increases. However, this trend is not linear indicating that satellite lifespan (and therefore the need for replenishment frequency) alone may not be a dominant factor in learning replenishment strategies. As expected, Figure 4.11d shows the decrease of returns as the cost to deploy orbit planes increases. This reflects a higher capital burden associated with deploying orbit planes if the market size is not increasing and launch costs are outside of the operator's control. Finally, variations of the recurring cost multiplier, K, are shown in Figure 4.11e. Unlike launch costs, increases in K show no distinct (or weak) correlations in the return. This may indicate the existence of a highly non-linear relation with the extrema. This result warrants further investigation into how recurring costs influence risk preferences in the learned policies.

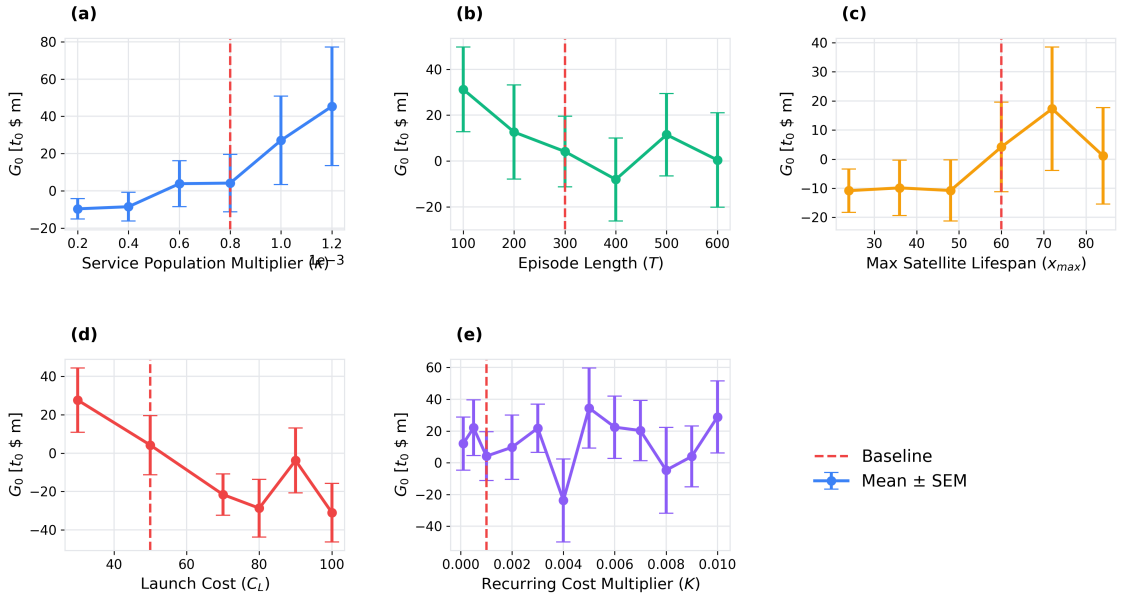

Figure 4.11: Net present value (G_0) as a function of environment parameters, averaged over 20 random seeds per configuration, with the baseline indicated by the red dashed line: (a) service population multiplier, (b) episode length, (c) maximum satellite lifespan, (d) launch cost, and (e) recurring cost multiplier.

Across all parameter variations, the compound annual growth rate ($CAGR$) remained relatively stable, with deviations of less than 5% across all strategies. This indicates that learned policies all

exhibit comparable growth across all parameter variations. As shown in Figure 4.12a, the service population multiplier demonstrated a monotonic positive relationship with $CAGR$. This can be explained as larger market sizes generating higher growth rates. It is notable that this trend was linear until some threshold leading to an accelerating growth rate. Conversely, Figure 4.12b shows an inverse relationship between episode length and $CAGR$. This mirrors the trend observed for G_0 and remains consistent with discounted MDP theory. Figure 4.12c illustrates a departure from the G_0 results shown in Figure 4.11c: as maximum satellite lifespan increases, a clear and intuitive increase in $CAGR$ is observed. While longer satellite lifespans (and reduced replenishment frequency) may lead to higher growth rates, satellite lifespan may still be a secondary driver when learning replenishment strategies. Reduced replenishment frequency may also benefit growth rate disproportionately more than net present value. Launch costs demonstrated in Figure 4.12d show an approximately linear negative relationship with $CAGR$. This is consistent with the trend seen in net present value and further shows that higher expenditures constrain an agent's ability to grow quickly. Figure 4.12e demonstrates that increasing the recurring cost to an agent leads to an approximately logarithmic increase in $CAGR$. Combined with the corresponding net present value results, this further suggests emergent strategies that are sensitive to recurring cost and merit further investigation.

4.5 Discussion of Limitations & Alternatives

This study is subject to several limitations spanning computational resolution, modeling structure, and the gap between abstraction and operational reality. From a numerical standpoint, a larger number of randomly seeded environments could reveal higher-performing policies and reduce variance in performance metrics across the tested parameter space. Likewise, denser parameter sampling may expose clearer trends and smaller-scale features that are unresolved at the current discretization level. Naturally, the results are conditional on the assumptions detailed in the model formulation; deviation from these assumptions reduces validity of the insights. We have not yet formally identified the dominant parameters or the hierarchy of governing dynamics that mostly shape successful policies and drive the competition dynamics.

While we have not formally identified dominant parameters or the hierarchy of governing dynamics that shape a successful policy, assumptions that may suffer more acutely from the gap between simulation and reality are identified. While grouping behavior is used to retain a computationally tractable customer

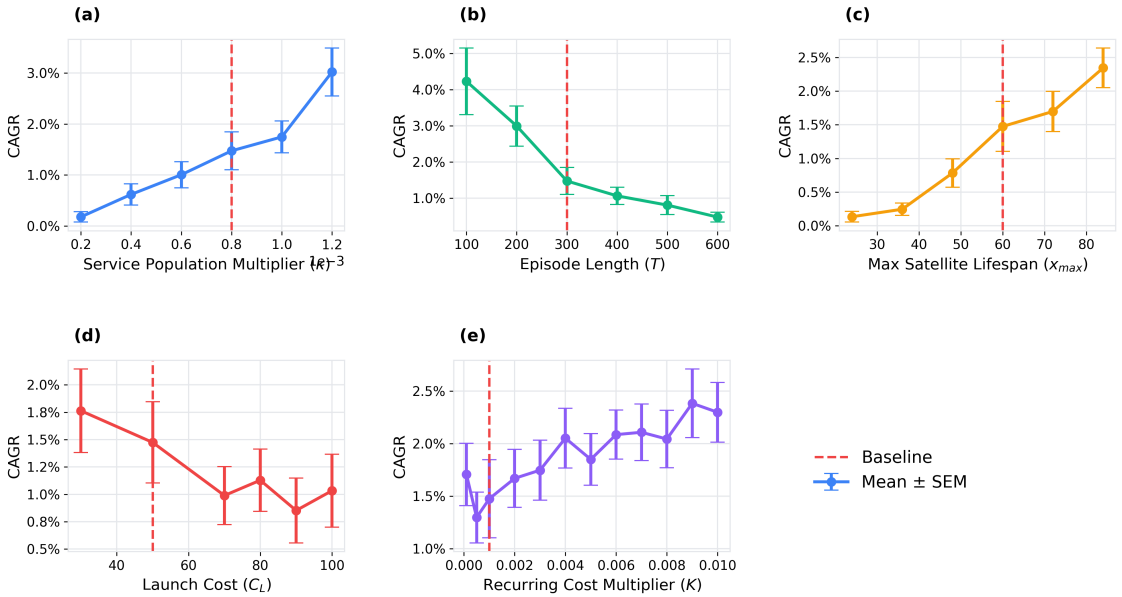

Figure 4.12: Compound Annual Growth Rate ($CAGR$) as a function of environment parameters, averaged over 20 random seeds per configuration, with the baseline indicated by the red dashed line: (a) service population multiplier, (b) episode length, (c) maximum satellite lifespan, (d) launch cost, and (e) recurring cost multiplier.

population in each region, modeling customers with different preferences within each region would allow more realistic and complex customer acquisition tradeoffs. While the service population multiplier was kept constant throughout this study, large upticks of this parameter can model rapid adoption. Finally, the largest assumption made is that of a static competitor. While a static competitor can be re-framed as a threshold performance target for RL agents, pure competition between multiple dynamic constellation operators would reveal broader competition dynamics.

Chapter 5

Multi-Agent Reinforcement Learning (MARL) Formulation of P-LEO

Building upon the single-agent formulation and parameter exploration from the previous chapter, we extend the MDP formulation into a Partially Observable Stochastic Game (POSG) [110]. This is a common generalization of the MDP that supports both multi-agent dynamics as well as partial observability. We reformulate the single-agent problem, keeping or extending many of the existing elements of the MDP, and replace the static utility function with a multi-attribute decision-making (MADM) method known as the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) . Finally, we describe two MARL algorithms (IDQN and IPPO) and show training results for each.

5.1 POSG Formulation

While most POSG formulations explicitly state the set of observations, states and observations are designed to be equivalent in this case, i.e., full information. Additionally, an initial state distribution is not defined as all agents begin with the same starting state,s_0.

5.1.1 States and State Space

In the POSG formulation, the environment state at epoch t is defined as a *global state* shared across all agents:

$$s_t \in \mathcal{S} \tag{5.1}$$

The global state vector at time t is represented as:

$$s_t = \begin{bmatrix} \vec{C}^{\,1} \\ f^1 \\ p^1 \\ M^1 \\ \vdots \\ \vec{C}^{\,n} \\ f^n \\ p^n \\ M^n \\ T - t \end{bmatrix} \tag{5.2}$$

where the superscript $i \in \{1, \ldots, n\}$ indexes each agent in the set of agents. For each agent i, the vector

$$\vec{C}^{\,i} = [x_1^i, x_2^i, \ldots, x_N^i]^\top \tag{5.3}$$

represents the remaining lifespan (x) of each orbit plane belonging to agent i. Note that each vector $\vec{C}^{\,i}$ is of size N, where N is the total number of possible orbit planes from the pre-computed catalog of P-LEO orbit planes. This catalog and the global customer population dataset are both reused from the single-agent RL environment. The quantities f^i, p^i, and M^i denote the current funds, service price, and total active satellites in orbit for agent i, respectively. The remaining timesteps in the episode, $T - t$, are included as a component of the global state. The state space $\mathcal{S}$ therefore consists of all admissible global state vectors formed by concatenating the individual system states of all agents along with the remaining timesteps in the episode.

5.1.2 Actions and Action Spaces

In the POSG formulation, each agent $i \in \{1, \ldots, n\}$ selects an action at epoch t based on its local information. The action available to agent i at state s_t is represented by the following vector of decision

variables:

$$\vec{a}^{\,i}(s_t) = \begin{bmatrix} b^i \\ o^i \\ p^i_{t+1} \end{bmatrix} \tag{5.4}$$

where b^i is a boolean indicating whether agent i launches a new orbit plane, o^i is the selected orbit plane from the pre-computed catalog of orbit planes, and p^i_{t+1} is the new service price offered to virtual customers by agent i.

The launch boolean, b^i, is a binary value set at each epoch to launch or not launch the selected orbit plane, o^i. If $b^i = 0$, the selected orbit plane is ignored. Although the service price can be any floating point value, allowable discrete values are used to simplify analysis and interpretation of the action space. While the tradespace of possible orbit plane designs is potentially limitless, a finite set of orbit planes is used to further bound the size of the action space.

Since all three decision variables are discrete, the individual action space for each agent i is defined as the Cartesian product of three discrete action sets:

$$\mathcal{A}_i = \mathrm{B} \times \mathrm{O} \times \mathrm{P} \tag{5.5}$$

where B represents the decision variable to launch a new orbit plane:

$$b^i \in \mathrm{B} = \{0, 1\}, \tag{5.6}$$

O represents the catalog of N unique orbit planes to choose from:

$$o^i \in \mathrm{O} = \{O_1, O_2, \ldots, O_N\}, \tag{5.7}$$

and P represents the set of service prices the agent can feasibly charge virtual customers (ranging from 0 to 200 in increments of 10):

$$p^i_{t+1} \in \mathrm{P} = \{0, 10, 20, \ldots, 200\}. \tag{5.8}$$

The joint action at epoch t is defined as the tuple of all individual agent actions:

$$\mathbf{a}_t = \left(a^1_t, a^2_t, \ldots, a^n_t\right) \in \mathcal{A} = \prod_{i=1}^{n} \mathcal{A}_i \tag{5.9}$$

Since each $\mathcal{A}_i$ is the Cartesian product of discrete spaces, it may be encoded as a single discrete action index representing all feasible combinations of the action vector components. This global discrete action is then decoded into individual discrete actions prior to further actions vector creation per agent. At each epoch, agent i selects an action $a_t^i \in \mathcal{A}_i$, which is then decoded into the corresponding three-component action vector $\vec{a}^{\,i}$. This decoding is performed for implementation convenience, after which the components of the decoded vector are used in the environment dynamics.

5.1.3 Dynamics and State Transition Probabilities

In the POSG formulation, the environment dynamics are defined over the global state and joint actions of all agents. Given that the system dynamics are deterministic, the probability of transitioning to global state s_{t+1} from state s_t under joint action $\mathbf{a}_t$ is given by:

$$p_t(s_{t+1} \mid s_t, \mathbf{a}_t) = \begin{cases} 1 & \iff s_{t+1} = L(s_t, \mathbf{a}_t) \\ 0 & \text{otherwise} \end{cases} \tag{5.10}$$

where $\mathbf{a}_t = (a_t^1, a_t^2, \ldots, a_t^n)$ denotes the joint action taken by all agents at epoch t, and function L defines the deterministic global law of motion governing the environment dynamics. Each episode consists of a sequence of at most T timesteps. At each timestep t, all agents simultaneously select individual actions $a_t^i \in \mathcal{A}_i$, forming the joint action $\mathbf{a}_t$. The environment then transitions deterministically to the next global state s_{t+1} according to the law of motion $L(s_t, \mathbf{a}_t)$. The mechanics executed within the environment, at every timestep are summarized in Figure 5.1 and illustrate how the single agent dynamics are extended to the multi-agent case. Notably, aside from the virtual customer model, most of the dynamics are unchanged. This is especially useful as the same orbit plane catalog can be used from the single-agent RL environment.

Given the discrete action representation used for each agent i, the encoded action a_t^i is decoded into its component action vector:

$$\vec{a}_t^{\,i} = \begin{bmatrix} b_t^i \\ o_t^i \\ p_{t+1}^i \end{bmatrix} \tag{5.11}$$

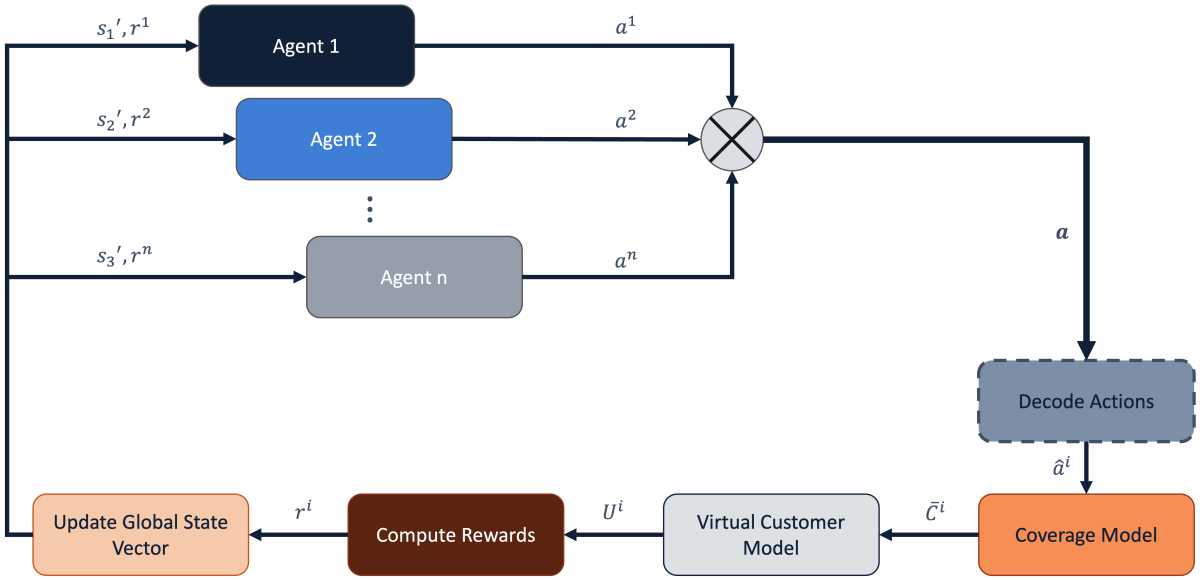

Figure 5.1: Multi-agent system dynamics that constitute individual timesteps of an episode are sub-divided into five components: (1) decoding the discrete actions of each agent into their action vectors, (2) updating each agents' constellation coverage metrics, (3) allocate virtual customers to each agent based on their service price and constellation metrics, (4) computing rewards (profits and losses) for each agent, and (5) assembling an updated global state vector to pass back to each agent.

The launch decision b_t^i for agent i is obtained using:

$$b_t^i = \left\lfloor \frac{a_t^i}{|\mathrm{O}| \times |\mathrm{P}|} \right\rfloor \tag{5.12}$$

The orbit plane selection o_t^i is computed as:

$$o_t^i = \left\lfloor \frac{a_t^i}{|\mathrm{O}|} \right\rfloor \mod |\mathrm{P}| \tag{5.13}$$

The updated service price is given by:

$$p_{t+1}^i = 10 \times (a_t^i \mod |\mathrm{P}|) \tag{5.14}$$

The global state update at timestep $t+1$ is determined by applying the decoded actions of all agents to the system dynamics. For each agent $i \in \{1, \ldots, n\}$ and each orbit plane index $j \in \{1, \ldots, N\}$, the

constellation lifespan variables evolve according to:

$$C^i(j)_{t+1} = \begin{cases} x_{max} & \text{if} \quad (b_t^i = 1 \wedge o_t^i = j) \\ \max\left(0, C^i(j)_t - 1\right) & \text{otherwise} \end{cases} \tag{5.15}$$

The funds for each agent evolve according to its individual reward:

$$f_{t+1}^i = f_t^i + r_t^i(s_t, \mathbf{a}_t) \tag{5.16}$$

The updated service price component is given by:

$$p_{t+1}^i = p_{t+1}^i \quad \text{(from the decoded action vector)} \tag{5.17}$$

The remaining timestep counter evolves as:

$$(T-t)_{t+1} = (T-t) - 1 \tag{5.18}$$

The number of active satellites for each agent i is computed as:

$$M_{t+1}^i = \sum_{j=1}^{N} n_j \cdot \begin{cases} 1 & \text{if} \quad C^i(j)_{t+1} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.19}$$

Termination conditions remain defined over the global timestep. An episode terminates upon reaching the final epoch T. An individual agent i is considered inactive if bankruptcy occurs, defined as:

$$f_{t+1}^i \leq 0 \tag{5.20}$$

In this case, the agent's actions are ignored by the environment. The deterministic transition function L therefore maps the current global state and joint action into the subsequent global state:

$$s_{t+1} = L(s_t, \mathbf{a}_t) \tag{5.21}$$

Each agent i's total acquired customers, U_t^i, are computed using the TOPSIS method described in detail in Section 5.2.

5.1.4 Reward Function

In the POSG formulation, each agent $i \in \{1, \ldots, n\}$ receives an individual reward at epoch t based on the global state s_t and joint action $\mathbf{a}_t$. The reward for agent i is defined as its cash flow:

$$r_t^i(s_t, \mathbf{a}_t) = CF_t^i = \text{revenues}_t^i - \text{costs}_t^i \tag{5.22}$$

where revenues are computed using agent i's total acquired customers U_t^i and current service price p_t^i:

$$\text{revenues}_t^i = U_t^i \cdot p_t^i \tag{5.23}$$

The service price, p, is taken as the agent's current service price, p_t^i; the updated value from the agent's action vector, p_{t+1}^i, is set after revenue computations as part of the state vector update. Costs at each epoch are computed as:

$$\text{costs}_t^i = \begin{cases} C_L + C_R^i = C_L + (M_t^i \cdot K) & \text{if} \quad b_t^i = 1 \\ C_R^i = M_t^i \cdot K & \text{otherwise} \end{cases} \tag{5.24}$$

where C_L is the fixed cost to launch an orbit plane, K is the recurring cost per active satellite, and M_t^i is the total number of active satellites for agent i, taken from the global state vector s_t. The recurring infrastructure cost $C_R^i = M_t^i \cdot K$ scales linearly with the number of active satellites agent i has in orbit. Note that C_L corresponds to the full deployment cost of an orbit plane, including the satellites themselves.

The individual reward function thus depends on agent i's components of the global state s_t and on agent i's decoded actions extracted from the joint action $\mathbf{a}_t$. The joint reward vector at epoch t is:

$$\mathbf{r}_t(s_t, \mathbf{a}_t) = \big(r_t^1(s_t, \mathbf{a}_t),\, r_t^2(s_t, \mathbf{a}_t),\, \ldots,\, r_t^n(s_t, \mathbf{a}_t)\big) \tag{5.25}$$

The funds update in the state transition (Section 4.2.4) follows directly from the individual reward:

$$f_{t+1}^i = f_t^i + r_t^i(s_t, \mathbf{a}_t) \tag{5.26}$$

5.2 MARL Customer Allocation

Each agent's customers can be represented as the summation of all the customers allocated to them by the environment based on their constellation design and current price. The customer behavior model is framed as a multi-attribute decision making (MADM) problem in which sets of customers, in a particular region, decide upon an internet service provider (agent) given a set of weighted preferences. This method of customer decision-making in the context of a reverse-bidding formulation is adapted from Section 3.2 of Cheng [111]. The general Technique for Order of Preference by Similarity to Ideal-Solution (TOPSIS) method, proposed by Tzeng and Huang [112] and outlined by Yang and Hung [113], is used to solve the MADM problem for each customer type, in each grid cell. It should be noted that while TOPSIS has been used for constellation design optimizations [114], it has not seen wide usage in the context of multi-agent customer/reward allocations. The TOPSIS method is given in detail:

Step 1. *Matrix Representation* The MADM problem can be written in matrix form with rows indicating competing alternatives (agents) and columns indicating the different attributes considered in the problem. Equation (5.27) presents a MADM problem formulation with n different attributes and m different alternatives, in which each entry of the matrix, x_{ij}, indicates the performance (raw value) of alternative (agent) i for attribute j. This matrix varies temporally and is updated once per virtual month.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{5.27}$$

Step 2. *Normalization of performance ratings*

Once a MADM matrix is formulated, the performance ratings of each attribute must be normalized so that they can be compared on an equal basis. Taking the approach of [112], attributes are categorized into cost and benefit attributes. From the virtual customer's perspective, cost attributes are those that should be minimized and benefit attributes are those that should be maximized. In the MADM problem

considered, internet price, total coverage gap, max coverage gap, and average altitude are all cost attributes, while mean coverage (average number of satellites in line-of-sight) is a benefit attribute. Equations (5.28) and (5.29) show the formulations of normalized cost and benefit attributes used in the TOPSIS method, respectively.

$$r_{ij} = \frac{\max_i \{x_{ij}\} - x_{ij}}{\max_i \{x_{ij}\} - \min_i \{x_{ij}\}} \tag{5.28}$$

$$r_{ij} = \frac{x_{ij} - \min_i \{x_{ij}\}}{\max_i \{x_{ij}\} - \min_i \{x_{ij}\}} \tag{5.29}$$

Through this normalization process, each attribute is expressed as a value in the interval [0,1]. These normalized values are irrespective of attribute type since the larger the r_{ij}, the more it satisfies the j-th attribute.

Step 3. *Weighting of attributes*

Weighting of attributes is an important component of the TOPSIS method since different customer types will have different priorities regarding attribute importance during their decision-making process. A weight vector is assigned to reflect each set of customers' preferences for each of the attributes in the MADM problem. This weighting vector is applied to the normalized matrix from the previous step and computed as:

$$v_{ij} = w_j r_{ij} \tag{5.30}$$

where w_j is defined as the individual weight (or importance) applied to the j-th attribute by the customer and v_{ij} is the weight-adjusted and normalized entry within the MADM matrix.

Step 4. *Identifying ideal and negative-ideal solutions*

The ideal solution, A^*, and negative-ideal solution, A^-, are defined by Equation (5.31) and Equation (5.32), respectively.

$$A^* = \left\{ \left(\max_i v_{ij} \mid j \in J_1 \right), \left(\min_i v_{ij} \mid j \in J_2 \right) \mid i = 1, \ldots, m \right\} = \{v_1^*, \ldots, v_n^*\} \tag{5.31}$$

$$A^- = \left\{ \left(\min_i v_{ij} \mid j \in J_1 \right), \left(\max_i v_{ij} \mid j \in J_2 \right) \mid i = 1, \ldots, m \right\} = \{v_1^-, \ldots, v_n^-\} \tag{5.32}$$

where J_1 is the set of all benefit attributes and J_2 is the set of all cost attributes. The ideal solution is therefore a theoretical vector that takes the best attribute values available from all alternatives (agents)

and the negative-ideal solution is the theoretical vector of the worst attribute values available from all alternatives (agents).

Step 5. *Distance calculation*

The distances from each alternative (agent) to the ideal solution and the negative-ideal solution are computed using the Euclidean norm and defined formally in Equation (5.33) and Equation (5.34), respectively.

$$S_i^* = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^*\right)^2}, \quad i = 1, \ldots, m \tag{5.33}$$

$$S_i^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^-\right)^2}, \quad i = 1, \ldots, m \tag{5.34}$$

Note that S^* and S^- are now vectors describing the ideal and negative-ideal distances for each of the m alternatives (agents), respectively.

Step 6. *Similarity calculation*

The similarity of each alternative is then computed using the following derivation:

$$R_i^* = \frac{S_i^-}{S_i^- + S_i^*}, \quad i = 1, \ldots, m \tag{5.35}$$

Note that the similarity of each alternative (agent) will always be in the interval [0,1] due to the nature of the derived distances.

Step 7. *Ranking and selection*

Finally, an *argmax* function is used on the vector of similarity scores, R^*, to obtain the alternative with the highest similarity metric. The agent chosen by the TOPSIS method is then assigned all the customers associated with that specific customer type in the evaluated region (winner take all model).

5.3 Results

With the complete POSG formulation, the environment was implemented in the Farama Foundation's PettingZoo framework [115]. This is the multi-agent standard and equivalent to the single-agent Gymnasium framework. The PettingZoo framework's ParallelEnv implementation allows agents to take simultaneous

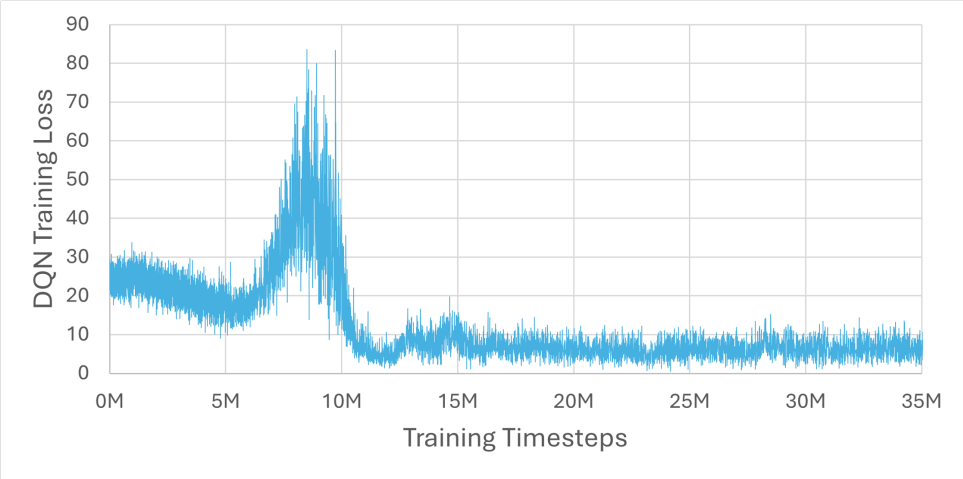

Figure 5.2: Training loss of the independent DQN algorithm shows convergence but does not guarantee optimal behavior.

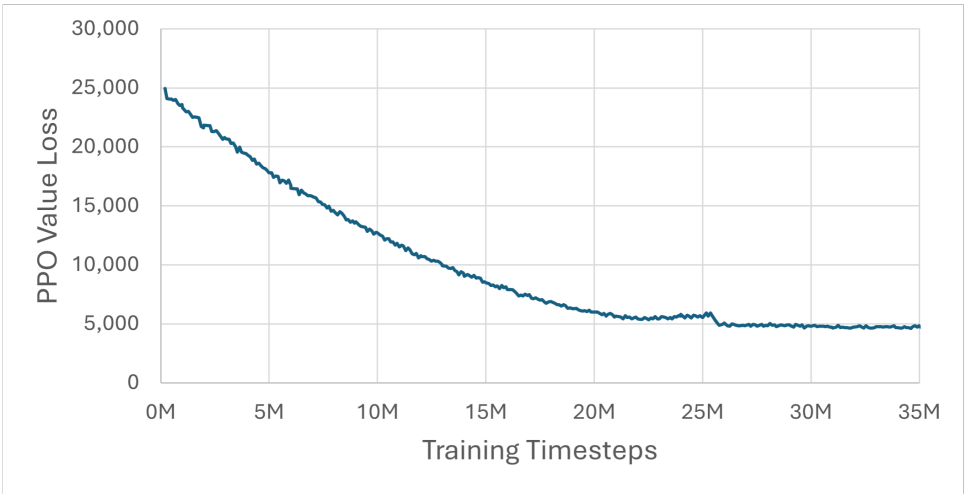

Figure 5.3: Value loss of the independent PPO algorithm shows convergence to a policy but does not provide any guarantees on the quality of that policy.

actions and has much the same functionality as Gymnasium. Following environment implementation, both IDQN and IPPO agents were trained on the environment with their convergence shown using training losses in Figures 5.2 and 5.3, respectively. While training seems to converge, additional analysis and detailed algorithm development are needed to establish any dominant strategies or Nash-equilibria.

Chapter 6

Conclusions and Future Work

This dissertation introduced a novel simulation-to-strategy method of exploring the complex dynamics of P-LEO satellite internet markets. Three research objectives were presented and different modeling techniques were used to address each objective. This chapter summarizes the contributions of this work with implications for constellation operators and space policy makers. It concludes with possible expansions on this work.

6.1 Summary & Contributions

For each research objective, specific formulations, results, insights, and contributions are offered: **Objective-1:** *Quantify the effects of satellite insurance and deorbit penalties on operators' constellation development strategies via an empirical study.*

1. A novel, multi-player modeling approach using the *Satellite Tycoon* tabletop board game was presented.

2. Using the board game as a surrogate simulation environment, a pilot RCT was conducted to explore the effects of two economic instruments designed to encourage sustainable usage of LEO: (1) optional satellite insurance and (2) mandatory deorbit penalties.

3. From limited empirical data, it is shown that players' general performance improved across multiple games and total debris created from derelict satellites decreased.

Objective-2: *Develop a parameterized environment of a single operator's performance in a dynamic P-LEO SATCOM market.*

1. Satellite deployment decisions, coverage computations, pricing strategies, and customer acquisitions were developed into a single RL environment to model strategic, sequential decision-making in P-LEO SATCOM markets.

2. The established framework enables systematic exploration of P-LEO SATCOM dynamics that are difficult to address using static analysis or combinatorial optimization. RL is shown to be a practical tool for evaluating complex tradeoffs and stress-testing constellation development strategies in unknown dynamics.

3. Learned policies exhibit satellite deployment and replenishment behavior, while parameter exploration reveals economic sensitivities from certain environment variables. Specifically, market size seems to be a strong factor in an operator's profitability calculation, with clear adoption thresholds required for market entry.

4. Longer time horizons and higher launch costs led to consistently lower returns, while the overall changes to compound annual growth rate remain stable across a wide array of parameter variations. Non-intuitive behavior also emerges as higher recurring costs lead to improvements in both net present value and compound annual growth rate.

5. Results from this environment demonstrate that economically sufficient policies can be learned.

Objective-3: *Develop a multi-agent framework for multiple P-LEO constellation operators to interact within the same competitive environment.*

1. The single-agent environment was extended into a MARL environment using POSG as the underlying mathematical model.

2. The unitary threshold-based utility function was replaced with a MADM model and solved using the TOPSIS method.

3. Using both independent DQN and independent PPO, agent performances converged during training; however, specific insights regarding Nash-equilibria and dominant strategies cannot be definitively determined without further ablation studies.

6.2 Implications for Constellation Operators

Results from the single-agent RL reveal that there are market adoption thresholds below which entry into the P-LEO market becomes infeasible. Additionally, it was found that launch cost is a primary barrier to profitability. This sets operators like SpaceX, who own their own launch vehicles and can give themselves higher priority in their launch vehicle rideshare programs, at a significant economic advantage. Learned policies as well as results from the RCT both show that early-movers who follow up with strategic replenishment are emergent behaviors: operators should plan for replenishment cadence at the same level at which they deorbit existing satellites, once established. Finally, the pilot RCT results suggest that even weak collusion among players led to improvements in profits and reductions of overall space debris. While self-policing is not practical, operators may benefit from coordination agreements where legal.

6.3 Policy and Insurance Mechanism Design

As shown in the RCT results, while derelict satellites were not a major source of debris, deorbit penalties showed a measurable reduction in the production of derelict satellites. This is a promising signal for financial penalties; however, enforcement of these penalties becomes an issue when international players are also in the domain. Optional insurance did not drive behavioral change so mandatory insurance at a lower price point may be needed to create meaningful space sustainability incentives. Additionally, collusion dynamics observed in the RCT suggest that industry coordination mechanisms such as self-regulatory bodies, spectrum sharing agreements, etc. could organically reduce debris if incentives are aligned. Finally, future policy design should consider flexible and dynamic economic instruments that can be adjusted based on the level of activity in the market.

6.4 Future Research Directions

While this dissertation establishes foundational frameworks for analyzing P-LEO satellite constellation development, several promising avenues of research remain open to exploration and extension. Both the fidelity and scope of each methodology can be extended. The following outlines future directions of enhancement to experimental game design, the introduction of stochasticity in the single-agent

environment and development of unique multi-agent algorithms to reveal strategies and interactions between agents in the MARL environment.

The RCT results obtained for Objective 1 can be further refined with statistical significance, given a follow-on study with a larger pool of participants. If participants were to play more games and longer-duration games, the conclusions in this dissertation could be further tested with new economic and policy instruments. Additionally, possible extensions to the underlying *Satellite Tycoon* board game could improve modeling fidelity and introduce new LEO-based use cases into the dynamics. Including different types of satellites, new regional tiles, and more mission-oriented objectives could expand the game into a tool for wargaming in the space domain.

While the single agent RL environment has already been expanded into a MARL environment, the introduction of stochasticity via collision probabilities could naturally expand orbit decay and satellite replenishment dynamics. Additionally, spectrum allocations and user terminal logistics may be added into the environment to give a more complete view of the P-LEO SATCOM market. Gateway availability and latency bottlenecks may also be incorporated into the environment as stochastic events that impact revenues and agent rewards. Finally, from the RL perspective, generalization of learned policies should be explored across different environment configurations and starting states.

The MARL environment itself should evolve with additional training to explore the converged market dynamics between different numbers and types of P-LEO constellation operators. This may capture emergent dynamics and equilibria that exist. Moreover, the TOPSIS-based customer model should be extended to include weights from heterogeneous consumer preferences within individual regions. Finally, adding a communication channel between agents may enable higher-level interactions such as collusion or monopolistic practices among agents.

Bibliography

[1] Jonathan C. McDowell. “The Low Earth Orbit Satellite Population and Impacts of the SpaceX Starlink Constellation”. In: *The Astrophysical Journal Letters* 892.2 (Apr. 2020), p. L36. ISSN: 2041-8205, 2041-8213. DOI: 10.3847/2041-8213/ab8016. URL: https://iopscience.iop.org/article/10.3847/2041-8213/ab8016 (visited on 04/07/2026).

[2] Marvin B. Lieberman and David B. Montgomery. “First-Mover Advantages”. In: *Strategic Management Journal* 9 (1988), pp. 41–58. ISSN: 01432095, 10970266. URL: http://www.jstor.org/stable/2486211 (visited on 08/07/2023).

[3] Abhipshito Bhattacharya and Marina Petrova. “Study on Handover Techniques for Satellite-to-Ground Links in High and Low Interference Regimes”. In: *2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*. Gothenburg, Sweden: IEEE, June 2023, pp. 359–364. ISBN: 9798350311020. DOI: 10.1109/EuCNC/6GSummit58263.2023.10188361. URL: https://ieeexplore.ieee.org/document/10188361/ (visited on 02/25/2024).

[4] Jonas Radtke, Christopher Kebschull, and Enrico Stoll. “Interactions of the space debris environment with mega constellations—Using the example of the OneWeb constellation”. en. In: *Acta Astronautica* 131 (Feb. 2017), pp. 55–68. ISSN: 00945765. DOI: 10.1016/j.actaastro.2016.11.021. URL: https://linkinghub.elsevier.com/retrieve/pii/S009457651630515X (visited on 03/23/2022).

[5] Yeluo Yu et al. “Analyses of Qianfan constellation coverage with respect to the ground users as of May 23, 2025”. In: *Journal of Physics: Conference Series* 3073.1 (Aug. 2025),

p. 012006. ISSN: 1742-6588, 1742-6596. DOI: 10.1088/1742-6596/3073/1/012006. (Visited on 04/07/2026).

[6] Paul Diaz et al. "Data-Driven Lifetime Risk Assessment and Mitigation Planning for Large-Scale Satellite Constellations". en. In: *The Journal of the Astronautical Sciences* 70.4 (July 2023), p. 21. ISSN: 2195-0571. DOI: 10.1007/s40295-023-00384-w. URL: https://link.springer.com/10.1007/s40295-023-00384-w (visited on 04/07/2026).

[7] Sofia Yang. "Conceptualizing thresholds for effective active debris removal in Low Earth Orbit". In: *Frontiers in Space Technologies* 7 (Feb. 2026), p. 1777020. ISSN: 2673-5075. DOI: 10.3389/frspt.2026.1777020. URL: https://www.frontiersin.org/articles/10.3389/frspt.2026.1777020/full (visited on 04/06/2026).

[8] Donald J. Kessler and Burton G. Cour-Palais. "Collision frequency of artificial satellites: The creation of a debris belt". en. In: *Journal of Geophysical Research: Space Physics* 83.A6 (June 1978), pp. 2637–2646. ISSN: 0148-0227. DOI: 10.1029/JA083iA06p02637. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/JA083iA06p02637 (visited on 04/06/2026).

[9] Garrett Hardin. "The Tragedy of the Commons: The population problem has no technical solution; it requires a fundamental extension in morality." en. In: *Science* 162.3859 (Dec. 1968), pp. 1243–1248. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.162.3859.1243. URL: https://www.science.org/doi/10.1126/science.162.3859.1243 (visited on 04/06/2026).

[10] Yu-Hsin Liu, Jeffrey Prince, and Scott Wallsten. "Distinguishing bandwidth and latency in households' willingness-to-pay for broadband internet speed". en. In: *Information Economics and Policy* 45 (Dec. 2018), pp. 1–15. ISSN: 01676245. DOI: 10.1016/j.infoecopol.2018.07.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167624517301609 (visited on 12/22/2025).

[11] Gaofeng Cui et al. “Latency Optimization for Hybrid GEO–LEO Satellite-Assisted IoT Networks”. In: *IEEE Internet of Things Journal* 10.7 (Apr. 2023), pp. 6286–6297. ISSN: 2327-4662, 2372-2541. DOI: 10.1109/JIOT.2022.3222831. URL: https://ieeexplore.ieee.org/document/9955992/ (visited on 09/24/2024).

[12] Michael Kan. “FCC Chair Encourages Satellite Internet Competition, Hints Starlink Is a Monopoly”. In: *PCMag* (). URL: https://www.pcmag.com/news/fcc-chair-encourages-satellite-internet-competition-hints-starlink-is-a?utm_medium=email&utm_source=rasa_io&utm_campaign=newsletter.

[13] Julien Guyot, Akhil Rao, and Sébastien Rouillon. “Oligopoly competition between satellite constellations will reduce economic welfare from orbit use”. In: *Proceedings of the National Academy of Sciences* 120.43 (2023), e2221343120. DOI: 10.1073/pnas.2221343120.

[14] T. P. Garrison et al. “Systems Engineering Trades for the IRIDIUM Constellation”. en. In: *Journal of Spacecraft and Rockets* 34.5 (Sept. 1997), pp. 675–680. ISSN: 0022-4650, 1533-6794. DOI: 10.2514/2.3267. URL: https://arc.aiaa.org/doi/10.2514/2.3267 (visited on 03/23/2022).

[15] Fred Dietrich. “The Globalstar satellite cellular communication system - Design and status”. en. In: *17th AIAA International Communications Satellite Systems Conference and Exhibit*. Yokohama,Japan: American Institute of Aeronautics and Astronautics, Feb. 1998. DOI: 10.2514/6.1998-1213. URL: http://arc.aiaa.org/doi/10.2514/6.1998-1213 (visited on 04/07/2026).

[16] M.A. Sturza. “The Teledesic satellite system”. In: *Proceedings of IEEE National Telesystems Conference - NTC ’94*. San Diego, CA, USA: IEEE, 1994, pp. 123–126. ISBN: 978-0-7803-1869-4. DOI: 10.1109/NTC.1994.316677. URL: http://ieeexplore.ieee.org/document/316677/ (visited on 03/23/2022).

[17] Andrew W. Lewin. “Low-Cost Operation of the ORBCOMM Satellite Constellation”. In: *Journal of Reducing Space Mission Cost* 1.1 (1998), pp. 105–117. ISSN: 13857479.

DOI: 10.1023/A:1009987231306. URL: http://link.springer.com/10.1023/A:1009987231306 (visited on 03/23/2022).

[18] Francesco Alessio Dicandia et al. “Space-Air-Ground Integrated 6G Wireless Communication Networks: A Review of Antenna Technologies and Application Scenarios”. en. In: *Sensors* 22.9 (Apr. 2022), p. 3136. ISSN: 1424-8220. DOI: 10.3390/s22093136. URL: https://www.mdpi.com/1424-8220/22/9/3136 (visited on 04/07/2026).

[19] Si-Yoon Kang et al. “Cost Effectiveness of Reusable Launch Vehicles Depending on the Payload Capacity”. en. In: *Aerospace* 12.5 (Apr. 2025), p. 364. ISSN: 2226-4310. DOI: 10.3390/aerospace12050364. URL: https://www.mdpi.com/2226-4310/12/5/364 (visited on 04/07/2026).

[20] Byeong-Un Jo and Jaemyung Ahn. “Optimal staging of reusable launch vehicles considering velocity losses”. en. In: *Aerospace Science and Technology* 109 (Feb. 2021), p. 106431. ISSN: 12709638. DOI: 10.1016/j.ast.2020.106431. URL: https://linkinghub.elsevier.com/retrieve/pii/S1270963820311135 (visited on 04/07/2026).

[21] Nils Pachler et al. “An Updated Comparison of Four Low Earth Orbit Satellite Constellation Systems to Provide Global Broadband”. In: *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*. 2021, pp. 1–7. DOI: 10.1109/ICCWorkshops50388.2021.9473799.

[22] Lu Jia and Yasheng Zhang. “Cost Estimation Model for Mega-Constellation Deployment Missions”. In: *IEEE Access* 9 (2021), pp. 156778–156788. DOI: 10.1109/ACCESS.2021.3130295.

[23] Jihao Li et al. “SkyCastle: Taming LEO Mobility to Facilitate Seamless and Low-latency Satellite Internet Services”. In: *IEEE INFOCOM 2024 - IEEE Conference on Computer Communications*. 2024, pp. 541–550. DOI: 10.1109/INFOCOM52122.2024.10621390.

[24] Barbara A. Cherry. “Technology transitions within telecommunications networks: Lessons from U.S. vs. Canadian policy experimentation under federalism”. In: *Telecommunications Policy* 39.6 (2015). Special Issue on ITS 2013 Florence, pp. 463–485. ISSN: 0308-5961. DOI: 10.1016/j.telpol.2015.02.001.

[25] Gian Luigi Somma, Hugh G. Lewis, and Camilla Colombo. “Sensitivity analysis of launch activities in Low Earth Orbit”. en. In: *Acta Astronautica* 158 (May 2019), pp. 129–139. ISSN: 00945765. DOI: 10.1016/j.actaastro.2018.05.043. URL: https://linkinghub.elsevier.com/retrieve/pii/S0094576517311244 (visited on 01/24/2022).

[26] Bassa, C. G. et al. “Bright unintended electromagnetic radiation from second-generation Starlink satellites”. In: *A&A* 689 (2024), p. L10. DOI: 10.1051/0004-6361/202451856.

[27] Ogutu B. Osoro and Edward J. Oughton. “A Techno-Economic Framework for Satellite Networks Applied to Low Earth Orbit Constellations: Assessing Starlink, OneWeb and Kuiper”. In: *IEEE Access* 9 (2021), pp. 141611–141625. DOI: 10.1109/ACCESS.2021.3119634.

[28] Davide Guzzetti et al. “Satellite Tycoon: Modeling Economic Competition in the Business of P-LEO Constellations”. In: *11th International Workshop on Satellite and Constellations Formation Flying*. Milan, Italy, June 2022.

[29] Rehman Qureshi et al. “Modeling and Gamification Framework of Business Competition Between P-LEO Constellations”. In: *2022 AAS/AIAA Astrodynamics Specialist Conference*. Charlotte, NC: AAS, Aug. 2022.

[30] Rehman Qureshi et al. “A Table-Top Game to Simulate Competition Between P-LEO Satellite Internet Constellations”. In: *2023 AAS/AIAA Astrodynamics Specialist Conference*. Big Sky, MT: AAS, Aug. 2023, pp. 1–17.

[31] Rehman Qureshi et al. “A Tabletop Game to Study Business Wargaming in the P-LEO SATCOM Marketplace”. In: *2024 IEEE Conference on Games (CoG)*. Milan, Italy: IEEE, Aug. 2024, pp. 1–8. ISBN: 979-8-3503-5067-8. DOI: 10.1109/CoG60054.

2024.10645581. URL: https://ieeexplore.ieee.org/document/10645581/ (visited on 03/24/2025).

[32] Ivan Oelrich, Paul Van Hooft, and Stephen Biddle. "Anti-satellite warfare, proliferated satellites, and the future of space-based military surveillance". In: *Journal of Strategic Studies* 47.6-7 (Nov. 2024), pp. 916–939. ISSN: 0140-2390, 1743-937X. DOI: 10.1080/01402390.2024.2379398.

[33] Zhuhang Li and Bodong Shang. "Fundamentals of Satellite-Maritime Communications: Downlink and Uplink Analysis". In: *IEEE Transactions on Communications* 73.4 (Apr. 2025), pp. 2191–2206. ISSN: 0090-6778, 1558-0857. DOI: 10.1109/TCOMM.2024.3466892.

[34] Yilei Wang et al. "Satellite–Aircraft Handover in Ultra-Dense LEO Satellite Networks". In: *IEEE Transactions on Vehicular Technology* 74.3 (Mar. 2025), pp. 4946–4961. ISSN: 0018-9545, 1939-9359. DOI: 10.1109/TVT.2024.3495658.

[35] Alexis Petit, Alessandro Rossi, and Elisa Maria Alessi. "Assessment of the close approach frequency and collision probability for satellites in different configurations of large constellations". en. In: *Advances in Space Research* 67.12 (June 2021), pp. 4177–4192. ISSN: 02731177. DOI: 10.1016/j.asr.2021.02.022.

[36] A. Rossi, A. Petit, and D. McKnight. "Short-term space safety analysis of LEO constellations and clusters". In: *Acta Astronautica* 175 (Oct. 2020), pp. 476–483. ISSN: 00945765. DOI: 10.1016/j.actaastro.2020.06.016.

[37] Jiaxin Hu et al. "A multi-objective optimization framework of constellation design for emergency observation". In: *Advances in Space Research* 67.1 (Jan. 2021), pp. 531–545. ISSN: 02731177. DOI: 10.1016/j.asr.2020.09.031.

[38] Inigo del Portillo, Bruce G. Cameron, and Edward F. Crawley. "A technical comparison of three low earth orbit satellite constellation systems to provide global broadband". In: *Acta Astronautica* 159 (June 2019), pp. 123–135. ISSN: 00945765. DOI: 10.1016/j.actaastro.2019.03.040. (Visited on 11/19/2021).

[39] Manuel Indaco and Davide Guzzetti. “Transformer-based anomaly detection in P-LEO constellations: A dynamic graph approach”. In: *Acta Astronautica* 218 (2024), pp. 177–194. ISSN: 0094-5765. DOI: https://doi.org/10.1016/j.actaastro.2024.02.019.

[40] S. Le May et al. “Space debris collision probability analysis for proposed global broadband constellations”. en. In: *Acta Astronautica* 151 (Oct. 2018), pp. 445–455. ISSN: 00945765. DOI: 10.1016/j.actaastro.2018.06.036. URL: https://linkinghub.elsevier.com/retrieve/pii/S0094576518304375 (visited on 01/24/2022).

[41] Hugh G. Lewis. “Evaluation of debris mitigation options for a large constellation”. en. In: *Journal of Space Safety Engineering* 7.3 (Sept. 2020), pp. 192–197. ISSN: 24688967. DOI: 10.1016/j.jsse.2020.06.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S2468896720300616 (visited on 01/24/2022).

[42] C. Pardini and L. Anselmo. “Assessing the Risk of Orbital Debris Impact”. In: *Space Debris* 1.1 (1999), pp. 59–80. ISSN: 13883828. DOI: 10.1023/A:1010066300520.

[43] Markus Guerster. “Revenue Management and Resource Allocation for Communication Satellite Operators”. PhD thesis. Massachusetts Institute of Technology, Sept. 2020. URL: https://dspace.mit.edu/handle/1721.1/129158.

[44] Harry W. Jones. “The Recent Large Reduction in Space Launch Cost”. In: *Proceedings of the 48th International Conference on Environmental Systems*. Albuquerque, New Mexico, July 2018, pp. 1–10.

[45] Akhil Rao, Matthew G. Burgess, and Daniel Kaffine. “Orbital-use fees could more than quadruple the value of the space industry”. In: *Proceedings of the National Academy of Sciences* 117.23 (2020), pp. 12756–12762. DOI: 10.1073/pnas.1921260117.

[46] Ogutu B. Osoro and Edward J. Oughton. “A Techno-Economic Framework for Satellite Networks Applied to Low Earth Orbit Constellations: Assessing Starlink, OneWeb and Kuiper”. In: *IEEE Access* 9 (Nov. 2021), pp. 141611–141625. ISSN: 2169-3536. DOI:

10.1109/ACCESS.2021.3119634. URL: https://ieeexplore.ieee.org/document/9568932/ (visited on 04/01/2022).

[47] Yinchien Huang, Qian Shi, and Cesare Guariniello. "A techno-economic framework for collaborative low Earth orbit satellite constellations". In: *2024 IEEE Aerospace Conference*. Big Sky, MT, USA: IEEE, Mar. 2024, pp. 1–11. ISBN: 979-8-3503-0462-6. DOI: 10.1109/AERO58975.2024.10521040. URL: https://ieeexplore.ieee.org/document/10521040/ (visited on 01/31/2025).

[48] Joshua F. Anderson, Michel-Alexandre Cardin, and Paul T. Grogan. "Design and analysis of flexible multi-layer staged deployment for satellite mega-constellations under demand uncertainty". In: *Acta Astronautica* 198 (2022), pp. 179–193. ISSN: 0094-5765. DOI: https://doi.org/10.1016/j.actaastro.2022.05.022.

[49] Richard Kim. "Stochastic Inventory Control Modeling for Satellite Constellations". In: *Journal of Spacecraft and Rockets* 57.3 (May 2020), pp. 612–620. ISSN: 0022-4650, 1533-6794. DOI: 10.2514/1.A34614.

[50] Leigha Capra et al. "SpaceNet Cloud: Web-based Modeling and Simulation Analysis for Space Exploration Logistics". en. In: *ASCEND 2021*. Las Vegas, Nevada & Virtual: American Institute of Aeronautics and Astronautics, Nov. 2021. ISBN: 978-1-62410-612-5. DOI: 10.2514/6.2021-4068. URL: https://arc.aiaa.org/doi/10.2514/6.2021-4068 (visited on 03/08/2022).

[51] Xuefeng Wang, Shijie Zhang, and Hongzhu Zhang. "The Optimal Deployment Strategy of Mega-Constellation Based on Markov Decision Process". In: *Symmetry* 15.5 (2023). ISSN: 2073-8994. DOI: 10.3390/sym15051024.

[52] Christopher J. Newman and Mark Williamson. "Space Sustainability: Reframing the Debate". en. In: *Space Policy* 46 (Nov. 2018), pp. 30–37. ISSN: 02659646. DOI: 10.1016/j.spacepol.2018.03.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S0265964617300462 (visited on 12/31/2025).

[53] Richard Crowther. “Space Junk–Protecting Space for Future Generations”. en. In: *Science* 296.5571 (May 2002), pp. 1241–1242. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1069725. URL: https://www.science.org/doi/10.1126/science.1069725 (visited on 04/05/2026).

[54] Nodir Adilov, Peter J. Alexander, and Brendan M. Cunningham. “An economic “Kessler Syndrome”: A dynamic model of earth orbit debris”. en. In: *Economics Letters* 166 (May 2018), pp. 79–82. ISSN: 01651765. DOI: 10.1016/j.econlet.2018.02.025. URL: https://linkinghub.elsevier.com/retrieve/pii/S0165176518300818 (visited on 09/26/2022).

[55] Nodir Adilov et al. “An estimate of expected economic losses from satellite collisions with orbital debris”. In: *Journal of Space Safety Engineering* 10.1 (2023), pp. 66–69. ISSN: 2468-8967. DOI: https://doi.org/10.1016/j.jsse.2023.01.002.

[56] Hugh G. Lewis and Vyara Yazadzhiyan. “Evaluation of low earth orbit post-mission disposal measures”. en. In: *Journal of Space Safety Engineering* 11.3 (Sept. 2024), pp. 526–531. ISSN: 24688967. DOI: 10.1016/j.jsse.2024.03.008. URL: https://linkinghub.elsevier.com/retrieve/pii/S2468896724000466 (visited on 04/06/2026).

[57] Richard S. Sutton and Andrew Barto. *Reinforcement learning: an introduction*. eng. Second edition. Adaptive computation and machine learning. Cambridge, Massachusetts London, England: The MIT Press, 2020. ISBN: 978-0-262-03924-6.

[58] Volodymyr Mnih et al. *Playing Atari with Deep Reinforcement Learning*. Version Number: 1. 2013. DOI: 10.48550/ARXIV.1312.5602. URL: https://arxiv.org/abs/1312.5602 (visited on 04/07/2026).

[59] David Silver et al. “Mastering the game of Go with deep neural networks and tree search”. en. In: *Nature* 529.7587 (Jan. 2016), pp. 484–489. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature16961. URL: https://www.nature.com/articles/nature16961 (visited on 04/07/2026).

[60] Oriol Vinyals et al. “Grandmaster level in StarCraft II using multi-agent reinforcement learning”. en. In: *Nature* 575.7782 (Nov. 2019), pp. 350–354. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-019-1724-z. URL: https://www.nature.com/articles/s41586-019-1724-z (visited on 04/07/2026).

[61] Hongwei Yang et al. “Reinforcement-Learning-Based Robust Guidance for Asteroid Approaching”. In: *Journal of Guidance, Control, and Dynamics* 47.10 (Oct. 2024), pp. 2058–2072. ISSN: 0731-5090, 1533-3884. DOI: 10.2514/1.G008085.

[62] Alessandro Zavoli and Lorenzo Federici. “Reinforcement Learning for Robust Trajectory Design of Interplanetary Missions”. en. In: *Journal of Guidance, Control, and Dynamics* 44.8 (Aug. 2021), pp. 1440–1453. ISSN: 1533-3884. DOI: 10.2514/1.G005794. URL: https://arc.aiaa.org/doi/10.2514/1.G005794 (visited on 10/25/2021).

[63] Deigant Yadava et al. “Attitude control of a nanosatellite system using reinforcement learning and neural networks”. In: *2018 IEEE Aerospace Conference*. Big Sky, MT: IEEE, Mar. 2018, pp. 1–8. ISBN: 978-1-5386-2014-4. DOI: 10.1109/AERO.2018.8396409.

[64] Charles E. Oestreich, Richard Linares, and Ravi Gondhalekar. “Autonomous Six-Degree-of-Freedom Spacecraft Docking with Rotating Targets via Reinforcement Learning”. en. In: *Journal of Aerospace Information Systems* 18.7 (July 2021), pp. 417–428. ISSN: 2327-3097. DOI: 10.2514/1.I010914.

[65] Anthony Aborizk and Norman Fitz-Coy. “Multiphase Autonomous Docking via Model-Based and Hierarchical Reinforcement Learning”. In: *Journal of Spacecraft and Rockets* 61.4 (July 2024), pp. 993–1005. ISSN: 0022-4650, 1533-6794. DOI: 10.2514/1.A35683.

[66] Andrew Harris et al. “Generation of Spacecraft Operations Procedures Using Deep Reinforcement Learning”. en. In: *Journal of Spacecraft and Rockets* 59.2 (Mar. 2022), pp. 611–626. ISSN: 0022-4650, 1533-6794. DOI: 10.2514/1.A35169.

[67] Adam Herrmann, Mark A. Stephenson, and Hanspeter Schaub. “Single-Agent Reinforcement Learning for Scalable Earth-Observing Satellite Constellation Operations”. In: *Journal of Spacecraft and Rockets* 61.1 (Jan. 2024), pp. 114–132. ISSN: 0022-4650, 1533-6794. DOI: 10.2514/1.A35736.

[68] Peng Mun Siew et al. “Space-Based Sensor Tasking Using Deep Reinforcement Learning”. In: *The Journal of the Astronautical Sciences* 69.6 (Nov. 2022), pp. 1855–1892. ISSN: 2195-0571. DOI: 10.1007/s40295-022-00354-8. (Visited on 02/23/2023).

[69] OpenAI et al. *Dota 2 with Large Scale Deep Reinforcement Learning*. 2019. DOI: 10.48550/ARXIV.1912.06680. URL: https://arxiv.org/abs/1912.06680.

[70] Lili Chen et al. “Decision Transformer: Reinforcement Learning via Sequence Modeling”. In: (2021). DOI: 10.48550/ARXIV.2106.01345.

[71] Josh Abramson et al. “Improving Multimodal Interactive Agents with Reinforcement Learning from Human Feedback”. In: (2022). DOI: 10.48550/ARXIV.2211.11602. URL: https://arxiv.org/abs/2211.11602.

[72] Stefano V. Albrecht, Filippos Christianos, and Lukas Schäfer. *Multi-Agent Reinforcement Learning: Foundations and Modern Approaches*. MIT Press, 2024. URL: https://www.marl-book.com.

[73] Dengyu Liao et al. “An Efficient Centralized Multi-Agent Reinforcement Learner for Cooperative Tasks”. In: *IEEE Access* 11 (2023), pp. 139284–139294. DOI: 10.1109/ACCESS.2023.3340867.

[74] Songyuan Zhang et al. “Solving Multi-Agent Safe Optimal Control with Distributed Epigraph Form MARL”. In: *Proceedings of Robotics: Science and Systems*. 2025.

[75] Pedro P. Santos et al. “Centralized training with hybrid execution in multi-agent reinforcement learning via predictive observation imputation”. In: *Artificial Intelligence* 348 (Nov. 2025), p. 104404. ISSN: 00043702. DOI: 10.1016/j.artint.2025.104404.

[76] Ryan Lowe et al. “Multi-agent actor-critic for mixed cooperative-competitive environments”. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS’17. Long Beach, California, USA: Curran Associates Inc., 2017, pp. 6382–6393. ISBN: 9781510860964.

[77] Christian Schroeder de Witt et al. *Is Independent Learning All You Need in the StarCraft Multi-Agent Challenge?* 2020. DOI: 10.48550/ARXIV.2011.09533.

[78] Kefan Su and Zongqing Lu. *A General Formulation of Independent Policy Optimization in Fully Decentralized MARL*. 2024. URL: https://openreview.net/forum?id=z1WiEHnjQs.

[79] Ruize Zhang et al. *A Survey on Self-play Methods in Reinforcement Learning*. 2024. DOI: 10.48550/ARXIV.2408.01072.

[80] Imre Gergely Mali. “Self-Play Meta-Reinforcement Learning in Multi-Agent Games”. In: *Acta Universitatis Sapientiae, Informatica* 18.1 (Feb. 2026), p. 5. ISSN: 2066-7760. DOI: 10.1007/s44427-026-00021-y.

[81] Lang Feng et al. *Bidirectional Distillation: A Mixed-Play Framework for Multi-Agent Generalizable Behaviors*. 2025. DOI: 10.48550/ARXIV.2505.11100.

[82] Charles Renshaw-Whitman et al. “Non-stationarity in multiagent reinforcement learning in electricity market simulation”. In: *Electric Power Systems Research* 235 (Oct. 2024), p. 110712. ISSN: 03787796. DOI: 10.1016/j.epsr.2024.110712.

[83] P. T. Grogan and O. L. de Weck. “Federated Simulation and Gaming Framework for a Decentralized Space-Based Resource Economy”. en. In: *Earth and Space 2012*. Pasadena, California, United States: American Society of Civil Engineers, Apr. 2012, pp. 1468–1477. ISBN: 978-0-7844-1219-0. DOI: 10.1061/9780784412190.156. URL: http://ascelibrary.org/doi/10.1061/9780784412190.156 (visited on 03/10/2022).

[84] Paul T. Grogan et al. “Multi-stakeholder interactive simulation for federated satellite systems”. In: *2014 IEEE Aerospace Conference*. Big Sky, MT, USA: IEEE, Mar.

2014, pp. 1–15. ISBN: 978-1-4799-5582-4. DOI: 10.1109/AERO.2014.6836253. URL: http://ieeexplore.ieee.org/document/6836253/ (visited on 03/08/2022).

[85] Paul T. Grogan and Olivier L. de Weck. "Interactive simulation games to assess federated satellite system concepts". In: *2015 IEEE Aerospace Conference*. Big Sky, MT: IEEE, Mar. 2015, pp. 1–13. ISBN: 978-1-4799-5380-6. DOI: 10.1109/AERO.2015.7119101. URL: http://ieeexplore.ieee.org/document/7119101/ (visited on 03/08/2022).

[86] Paul T. Grogan et al. "Bounding the value of collaboration in federated systems". In: *2016 Annual IEEE Systems Conference (SysCon)*. Orlando, FL: IEEE, Apr. 2016, pp. 1–7. ISBN: 978-1-4673-9519-9. DOI: 10.1109/SYSCON.2016.7490657. URL: https://ieeexplore.ieee.org/document/7490657/ (visited on 03/08/2022).

[87] Jay Kurtz. ""Business wargaming": simulations guide crucial strategy decisions". en. In: *Strategy & Leadership* 31.6 (Dec. 2003), pp. 12–21. ISSN: 1087-8572. DOI: 10.1108/10878570310505550. URL: https://www.emerald.com/insight/content/doi/10.1108/10878570310505550/full/html (visited on 02/29/2024).

[88] Daniel F. Oriesek and Jan Oliver Schwarz. *Business Wargaming*. en. 0th ed. Routledge, Apr. 2016. ISBN: 978-1-317-17041-9. DOI: 10.4324/9781315570648. URL: https://www.taylorfrancis.com/books/9781317170419 (visited on 02/29/2024).

[89] Jan Oliver Schwarz. "Business wargaming: developing foresight within a strategic simulation". en. In: *Technology Analysis & Strategic Management* 21.3 (Apr. 2009), pp. 291–305. ISSN: 0953-7325, 1465-3990. DOI: 10.1080/09537320902750590. URL: http://www.tandfonline.com/doi/abs/10.1080/09537320902750590 (visited on 02/17/2024).

[90] Haridimos Tsoukas and Jill Shepherd. “Coping with the future: developing organizational foresightfulness”. en. In: *Futures* 36.2 (Mar. 2004), pp. 137–144. ISSN: 00163287. DOI: 10.1016/S0016-3287(03)00146-0. URL: https://linkinghub.elsevier.com/retrieve/pii/S0016328703001460 (visited on 02/29/2024).

[91] D. H. Ingvar. “”Memory of the future”: an essay on the temporal organization of conscious awareness.” eng. In: *Human neurobiology* 4.3 (1985). Place: Germany, pp. 127–136. ISSN: 0721-9075.

[92] Jan Oliver Schwarz, Camelia Ram, and René Rohrbeck. “Combining scenario planning and business wargaming to better anticipate future competitive dynamics”. en. In: *Futures* 105 (Jan. 2019), pp. 133–142. ISSN: 00163287. DOI: 10.1016/j.futures.2018.10.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S0016328718300545 (visited on 02/17/2024).

[93] Jan Oliver Schwarz. “Business wargaming for teaching strategy making”. en. In: *Futures* 51 (July 2013), pp. 59–66. ISSN: 00163287. DOI: 10.1016/j.futures.2013.06.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S0016328713000864 (visited on 02/17/2024).

[94] Allan Young. “Hope & Despair: How Perceptions of the Future Shape Human Behavior”. en. In: *The Journal of Nervous and Mental Disease* 193.9 (Sept. 2005), pp. 636–637. ISSN: 0022-3018. DOI: 10.1097/01.nmd.0000177775.33360.16. URL: http://journals.lww.com/00005053-200509000-00012 (visited on 02/29/2024).

[95] Davide Guzzetti and Daniel Tauritz. *Modeling Economic Competition in the Business of Mega-Constellations*. Technical Report AD1163306. Auburn, AL: Auburn University, 2022.

[96] James Richard Wertz. *Mission geometry: orbit and constellation design and management: spacecraft orbit and attitude systems*. Space technology library 13. El Segundo,

Calif. : Dorfrecht ; Boston: Microcosm Press : Kluwer Academic Publishers, 2001. ISBN: 978-0-7923-7148-9.

[97] Manuel Ruiz-Pérez et al. “An institutional analysis of the sustainability of fisheries: Insights from FishBanks simulation game”. In: *Ocean & Coastal Management* 54.8 (2011), pp. 585–592. ISSN: 0964-5691. DOI: https://doi.org/10.1016/j.ocecoaman.2011.05.009.

[98] John D. Sterman. “Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment”. In: *Management Science* 35.3 (1989), pp. 321–339. URL: http://www.jstor.org/stable/2631975.

[99] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. 1st ed. Wiley Series in Probability and Statistics. Wiley, Apr. 1994. ISBN: 978-0-471-61977-2. DOI: 10.1002/9780470316887.

[100] Rehman Qureshi. “Modeling Environments for Exploration of Business Dynamics within P-LEO Constellation Markets”. MA thesis. Auburn, AL, USA: Auburn University, May 2023.

[101] Harry W. Jones. “The Impact of Reduced Space Launch Costs”. In: *AIAA AVIATION FORUM AND ASCEND 2025*. Las Vegas, Nevada: American Institute of Aeronautics and Astronautics, July 2025, pp. 1–9. ISBN: 978-1-62410-738-2. DOI: 10.2514/6.2025-4073.

[102] Gary Comparetto and Neal Hulkower. “Global mobile satellite communications - A review of three contenders”. In: *15th International Communicatons Satellite Systems Conference and Exhibit*. San Diego, CA, U.S.A.: American Institute of Aeronautics and Astronautics, Feb. 1994, pp. 1507–1517. DOI: 10.2514/6.1994-1138.

[103] Fabio Pardo et al. “Time Limits in Reinforcement Learning”. In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, July 2018, pp. 4045–4054. DOI: 10.48550/ARXIV.1712.00378.

[104] Richard E. Bellman. *Dynamic Programming*. Princeton University Press, Dec. 2010, pp. 81–83. ISBN: 978-1-4008-3538-6. DOI: 10.1515/9781400835386.

[105] Jonathan McDowell. *Satellite Constellation List*. Accessed Nov. 13, 2025. 2025. URL: https://planet4589.org/space/con/conlist.html.

[106] Center For International Earth Science Information Network-CIESIN-Columbia University. *Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11*. 2017. DOI: 10.7927/H49C6VHW.

[107] Dan Howdle. *Worldwide Broadband Price Research 2024*. Accessed Nov. 13, 2025. 2024. URL: https://bestbroadbanddeals.co.uk/broadband/pricing/worldwide-comparison/.

[108] Mark Towers et al. *Gymnasium: A Standard Interface for Reinforcement Learning Environments*. 2024.

[109] Antonin Raffin et al. "Stable-Baselines3: Reliable Reinforcement Learning Implementations". In: *Journal of Machine Learning Research* 22.268 (2021), pp. 1–8.

[110] Eric A. Hansen, Daniel S. Bernstein, and Shlomo Zilberstein. "Dynamic programming for partially observable stochastic games". In: *Proceedings of the 19th National Conference on Artifical Intelligence*. AAAI'04. San Jose, California: AAAI Press, 2004, pp. 709–715. ISBN: 0262511835.

[111] Chi-Bin Cheng. "Solving a sealed-bid reverse auction problem by multiple-criterion decision-making methods". en. In: *Computers & Mathematics with Applications* 56.12 (Dec. 2008), pp. 3261–3274. ISSN: 08981221. DOI: 10.1016/j.camwa.2008.09.011. URL: https://linkinghub.elsevier.com/retrieve/pii/S0898122108004896 (visited on 11/17/2021).

[112] Gwo-Hshiung Tzeng and Jih-Jeng Huang. *Multiple attribute decision making: methods and appliations*. CRC Press, Boca Raton, FL, 2012.

[113] Taho Yang and Chih-Ching Hung. “Multiple-attribute decision making methods for plant layout design problem”. In: *Robotics and Computer-Integrated Manufacturing* 23.1 (2007), pp. 126–137. ISSN: 0736-5845. DOI: https://doi.org/10.1016/j.rcim.2005.12.002. URL: https://www.sciencedirect.com/science/article/pii/S0736584506000044.

[114] Mikkel Søby Kramer et al. “A Simulation and TOPSIS Approach to the Satellite Constellation Design Problem”. In: *Aerospace* 13.3 (2026). ISSN: 2226-4310. DOI: 10.3390/aerospace13030284. URL: https://www.mdpi.com/2226-4310/13/3/284.

[115] J Terry et al. “Pettingzoo: Gym for multi-agent reinforcement learning”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 15032–15043.

[116] J.R. Wertz, Naval Center for Space Technology (U.S.), and Naval Research Laboratory (U.S.) *Mission Geometry: Orbit and Constellation Design and Management : Spacecraft Orbit and Attitude Systems*. Space technology library. Microcosm Press, 2001. ISBN: 9781881883074.

[117] Akhil Rao et al. *OPUS: An Integrated Assessment Model for Satellites and Orbital Debris*. 2023. DOI: 10.48550/ARXIV.2309.10252.

Appendix A

Right Ascension Coverage Mask Generation

Section 9.5.1.3 of Wertz [116] is used to estimate both the along-track and across-track coverage. These are used together to register the fraction of orbits with probable coverage over the right ascension coverage mask. This is used as an average-day estimate of coverage.

Algorithm 1 Right Ascension Coverage Mask Generation

for each Latitude Band $(\text{Lat}, \epsilon_{\min})$ **do**

  for each Orbit Plane (i, Ω, h, N_{sat}) **do**

    Step 0: Compute orbit period in days:

$$T_{\text{day}} = \frac{\left(1.658 \times 10^{-4} \times (R_E + h)^{3/2}\right)}{60 \times 24}$$

    Step 1: Compute access area ($\lambda_{\max}$):

$$\sin(\eta_{\max}) = \left[\frac{R_E}{R_E + h}\right] \cos(\epsilon_{\min}), \quad \lambda_{\max} = \frac{\pi}{2} - \epsilon_{\min} - \eta_{\max}$$

    Step 2: Across-Track coverage (P_{cov}):

    Percentage number of orbits per day that cross the access area

$$\cos(\phi_{1,2}) = \frac{\pm \sin(\lambda_{\max}) + \cos(i)\ \sin(\text{Lat})}{\sin(i)\ \cos(\text{Lat})}$$

    Coverage Case Determination:

Latitude Range	Coverage Regions (M)	Percent Coverage (P_{cov})
$\text{Lat} > \lambda_{\max} + i$	0	0
$i + \lambda_{\max} > \text{Lat} > i - \lambda_{\max}$	1	$\phi_1/180$
$i - \lambda_{\max} > \text{Lat} > 0$	2	$(\phi_1 - \phi_2)/180$

    ▷ Will differ for retrograde orbits and southern hemisphere latitude bands.

    Step 3: Compute Mean Transit Duration (ΔT):

$$\text{Lat}_{\text{pole}} = 90^\circ - i$$

$$\lambda_{\min} = \begin{cases} 0 & , \quad |A/B| \leq 1 \\ \min \left| \arcsin\big(A \pm B\big) \right|, & \text{otherwise} \end{cases}$$

    where $A = \sin(\text{Lat}_{\text{pole}})\sin(\text{Lat}), \quad B = \cos(\text{Lat}_{\text{pole}})\cos(\text{Lat})$

Algorithm 1 (continued)

$$\lambda_{\text{frac}} = \begin{cases} \dfrac{\cos(\lambda_{\max})}{\cos\left((\lambda_{\max} + \lambda_{\min})/2\right)}, & \lambda_{\max} > \lambda_{\min} \\ 0 \quad , & \text{otherwise} \end{cases}$$

$$\Delta T = \frac{T_{\text{day}}}{\pi} \arccos(\lambda_{\text{frac}})$$

Step 4: Along-Track coverage (n):

Fraction of orbit with probable coverage

$$n = \left(\frac{\Delta T}{T_{\text{day}}}\right) N_{sat}$$

Step 5: Register Across-Track coverage:

if $M \neq 0$ **then**

$$\sin(\Delta RA) = \frac{\tan(\text{Lat} - \lambda_{\min})}{\tan(i)}$$

$$RA_{LSP} = (\Omega + \Delta RA) \bmod 360$$

if $M = 1$ **then**

$RA_{start} = RA_{LSP} - 180 \cdot P_{\text{cov}}$ ▷ Start of single coverage region

$RA_L = 360 \cdot P_{\text{cov}}$ ▷ Length of coverage across right ascension mask

$\text{coverageMask}\left[(RA_{start} + i) \bmod 360\right] = n, \quad i = 0, 1, \ldots, RA_L - 1$

end if

if $M = 2$ **then**

$RA_{s1} = RA_{LSP} - 90 \cdot P_{\text{cov}}$ ▷ Start of first coverage region

$RA_{s2} = RA_{s1} + 180$ ▷ Start of second coverage region

$RA_L = 180 \cdot P_{\text{cov}}$ ▷ Length of coverages across right ascension mask

$\text{coverageMask}\left[(RA_{s1} + i) \bmod 360\right] = n, \quad i = 0, 1, \ldots, RA_L - 1$

$\text{coverageMask}\left[(RA_{s2} + i) \bmod 360\right] = n, \quad i = 0, 1, \ldots, RA_L - 1$

end if

end if

end for

end for

Appendix B

Additivity of Discrete Right Ascension Coverage Masks

Theorem B.1. *Let n orbit planes independently contribute to coverage over a fixed latitude. Define a discretized Right Ascension (RA) coverage mask for the i-th plane as:*

$$C_i = \{1, 2, \ldots, 360\} \to \mathbb{N} \cup 0 \tag{B.1}$$

which maps each index to a natural number of satellites from the i-th plane covering that RA at the given latitude. Then, the total coverage mask C_{total} is given by:

$$C_{total}(\alpha) = \sum_{i=1}^{n} C_i(\alpha) \quad \forall \alpha \in \{1, 2, \ldots, 360\} \tag{B.2}$$

Proof. For each orbit plane, i, let S_i be the set of satellites in that plane. We abstractly define a coverage function for each orbit plane, at a particular latitude, as:

$$C_i(\alpha) = \sum_{s \in S_i} \mathbb{1}_s(\alpha), \tag{B.3}$$

where the indicator function $\mathbb{1}_s(\alpha)$ is:

$$\mathbb{1}_s(\alpha) = \begin{cases} 1, & \text{if satellite } s \text{ provides coverage to RA } \alpha \\ 0, & \text{otherwise} \end{cases} \tag{B.4}$$

Note that the coverage function B.3 counts the total number of satellites from orbit plane i that provide coverage at each RA α.

Now, consider multiple orbit planes, indexed by $i \in \{1, 2, \ldots, n\}$, where each plane i has its own independent coverage function $C_i(\alpha)$ for the same latitude. We represent the total coverage function at

each RA α as the sum of the satellites from ***all*** orbit planes covering that RA α:

$$C_{total}(\alpha) = \sum_{s \in S_1 \cup S_2 \cup \cdots \cup S_n} \mathbb{1}_s(\alpha) \quad \forall \alpha \in \{1, 2, \ldots, 360\} \tag{B.5}$$

Since each satellite belongs to exactly one orbit plane, S_i, we can rewrite the total coverage function as a sum over orbit planes:

$$C_{total}(\alpha) = \sum_{i=1}^{n} \sum_{s \in S_i} \mathbb{1}_s(\alpha) \tag{B.6}$$

We may now use the definition of the coverage function for a single orbit plane, B.3, and substitute it into our total coverage function:

$$C_{total}(\alpha) = \sum_{i=1}^{n} C_i(\alpha) \tag{B.7}$$

Thus, the total coverage function is shown to be the sum of individual coverage masks across the entire discretized right ascension:

$$C_{total}(\alpha) = \sum_{i=1}^{n} C_i(\alpha) \quad \forall \alpha \in \{1, 2, \ldots, 360\} \tag{B.8}$$

□

Appendix C

Experimental Setup Parameters

The Baseline configuration defines the reference environment with appropriate start states used throughout the experiments. While a start state is not explicitly defined in the MDP formulation, we define one using publicly available data and informed estimates. Each episode begins from this start state given in Table C.1. The remaining environment parameters and training hyper-parameters, summarized in Table C.2, fully define the baseline environment configuration. Note that all training hyper-parameters not specified were default values from the Stable-Baselines3 implementation.

Table C.1: Start State ($\vec{s_0}$) of the agent

State Variable	**Value**	**Units**
Orbit Plane Lifespans ($\vec{C}$)	$\vec{\mathbf{0}} \in \mathbb{R}^N$	months
Funds (f)	1,000	$ million
Service Price (p)	0	$
Remaining Timesteps ($T - t$)	300	months
Total Active Satellites (M)	0	satellites

Table C.2: Environment parameters and hyper-parameters defining the training setup

Parameter	**Value**	**[Units] Notes**
Episode Length (T)	300	[months] Assumption
Orbit Plane Catalog Size (N)	1,139	From orbit shells collected
Number of Grid Cells (D)	2,592	Discretized $5^\circ \times 5^\circ$ cells
Max TCG Threshold (α)	40%	Percentage of coverage mask
Max Altitude Threshold (β)	1,000	[kilometers]
Max Satellite Lifespan (x_{max})	60	[months] Sourced from [117].
Launch Cost (C_L)	50	[$ million] Assumption
Recurring Cost Multiplier (K)	0.001	[$ m / satellite] Assumption
Service Population Multiplier (κ)	8×10^{-4}	Assumption
Minimum Elevation Angle ($\epsilon_{\min}$)	12°	Assumption
Discount Factor (γ)	1/1.07	Assuming 7% market rate.
Max Training Steps	100 million	[Timesteps] Sufficient for convergence
Min Training Steps	25 million	[Timesteps] Sufficient for convergence
Max No Improvement Steps	10 million	[Timesteps] Sufficient for convergence
Evaluation Frequency	1 million	Assumption
RL Algorithm	DQN	Chosen method
DQN Learning Rate	10^{-6}	Assumption