Revealing the Oil Majors' Adaptive Capacity to the Energy Transition with Deep Multi-Agent Reinforcement Learning

Dylan Radovic
Lucas Kruitwagen
Christian Schroeder de Witt

Outline

Introduction

Solving a Wargame

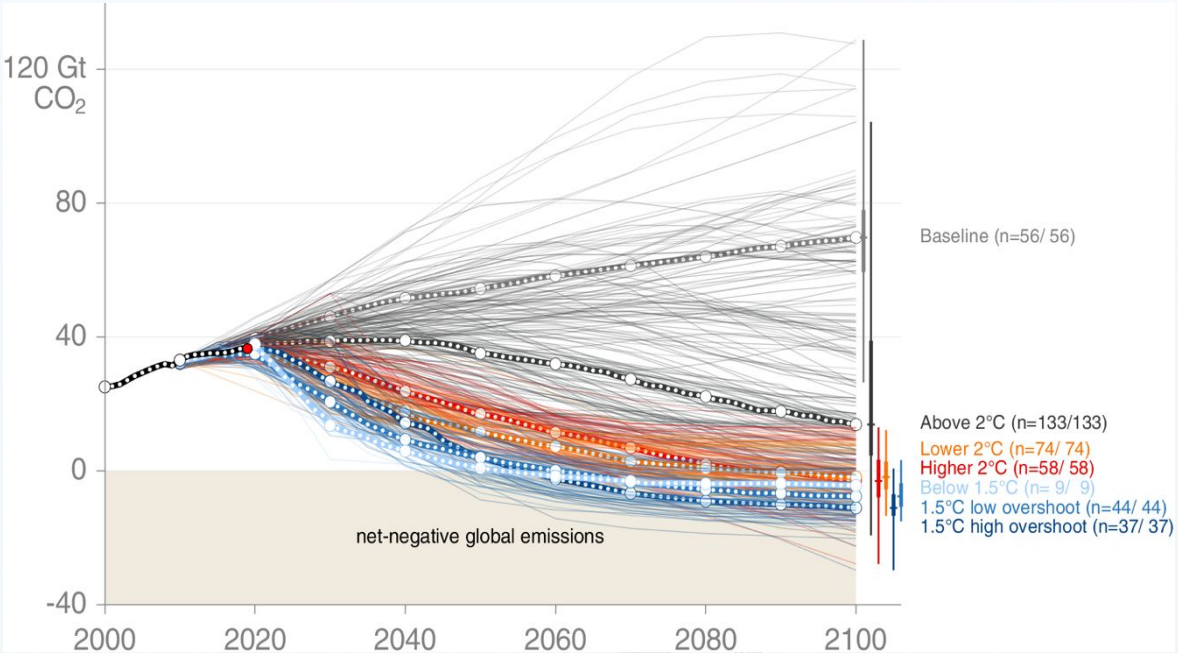
2DP-MARL

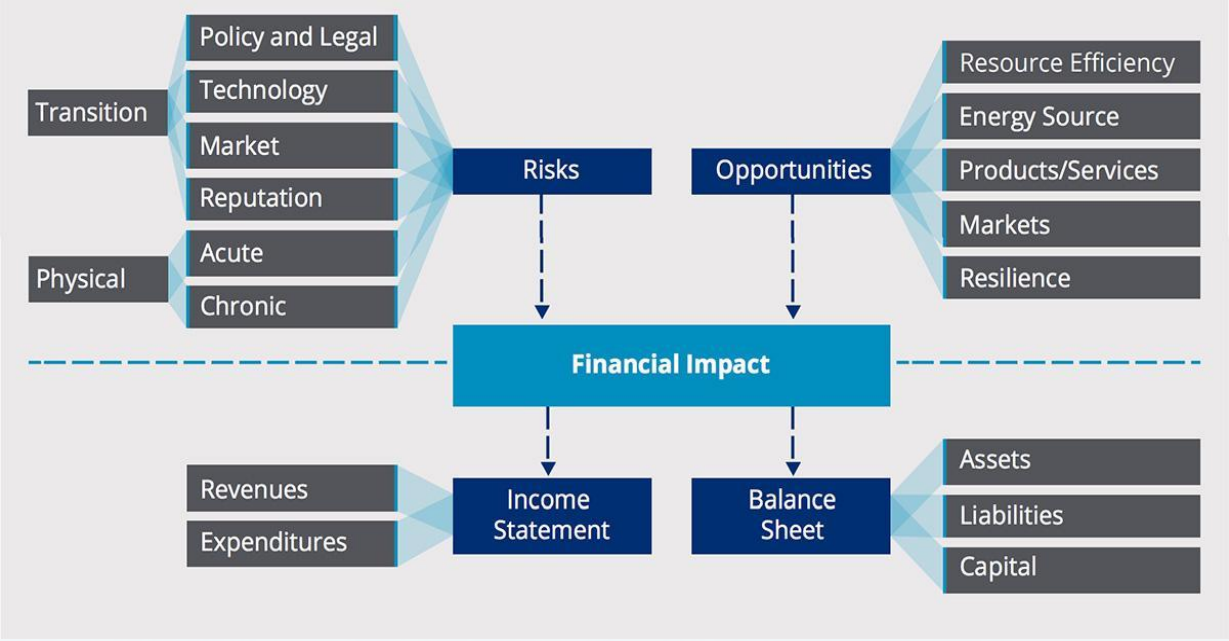
Results & Analysis

Impact

Future Work

Introduction



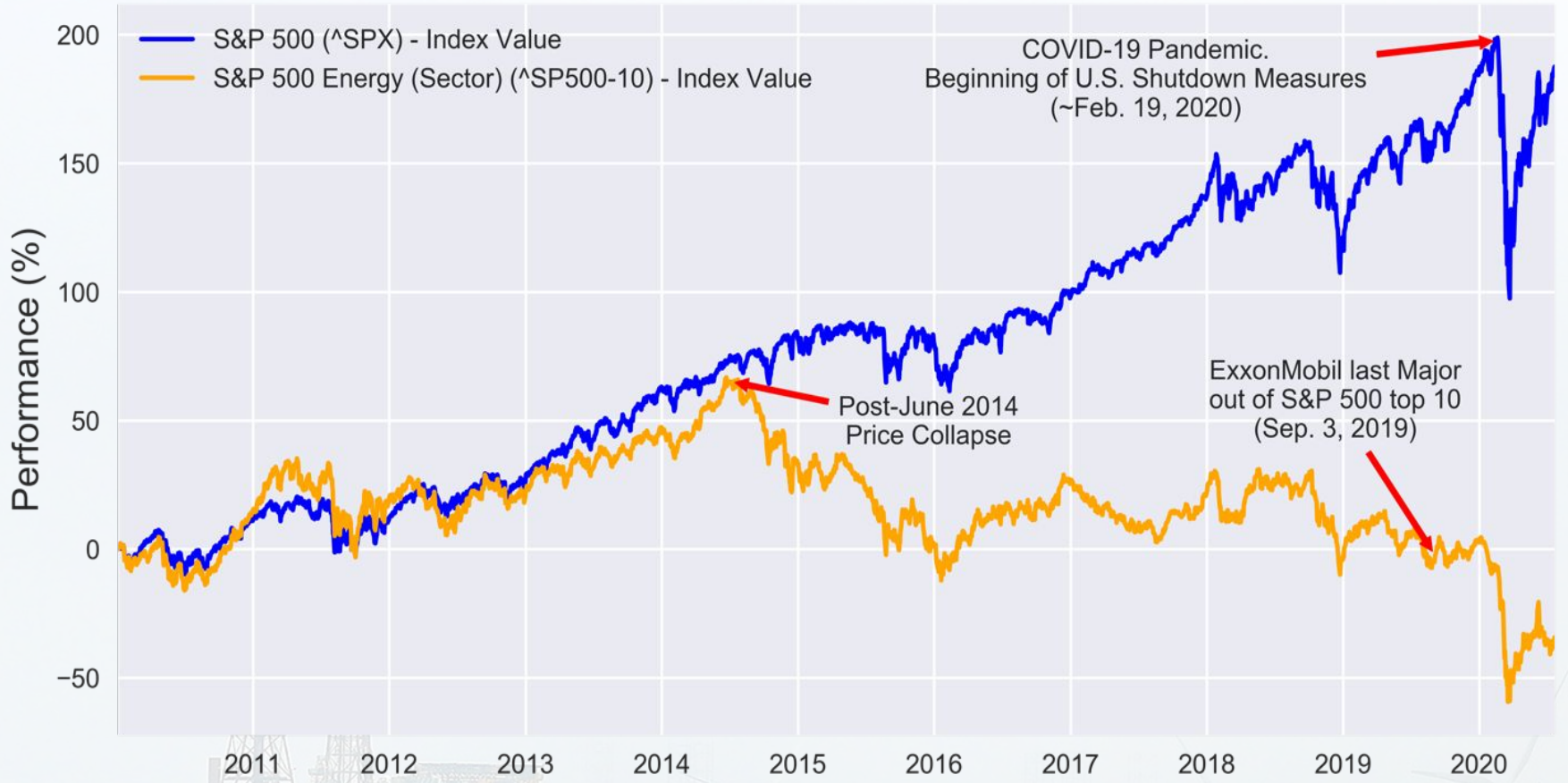


Energy transition manifesting in response to climate change

- Though pathways vary, all scenarios predicated on low-carbon energy production
- Deep decarbonization requires massive reallocation of capital

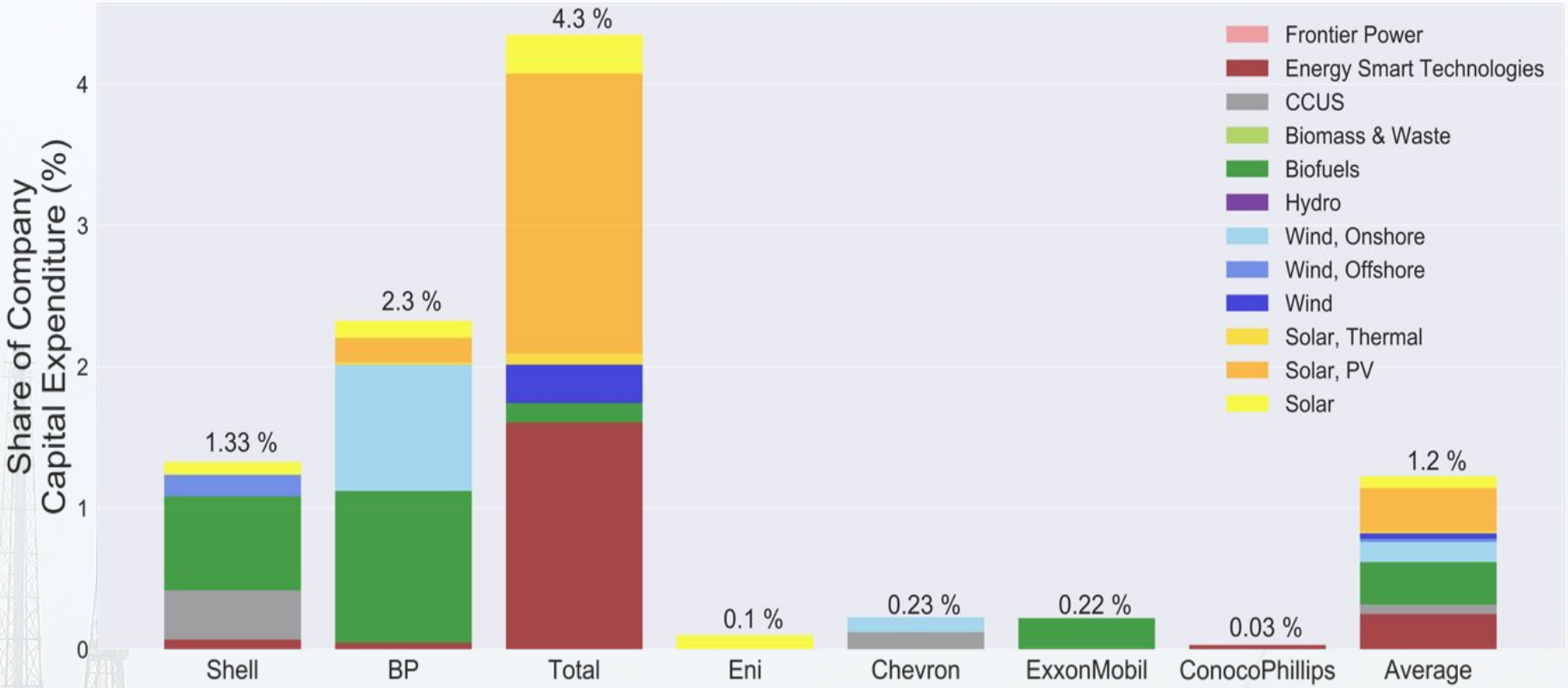
Transition risks will arise, however, so will opportunities

- Rapid low-carbon transition poses an existential threat for the fossil fuel industry, particularly the oil Majors
- Adaptation may prove financially favorable



The Majors have been in decline since 2014, COVID-19 has accelerated it

- Oil & gas companies have consistently underperformed despite economic growth
- Financial recovery post-COVID through BAU becomes increasingly unlikely as the impeding energy transition unfolds



Low-carbon actions speak louder than net-zero words

- Majors have set out plans for a full-scale decarbonization of their business models
- Throughout the previous decade, the Majors' low-carbon efforts proved minimal
- Majors have yet to make significant moves into low-carbon business models as the upside and downside risks of doing so remain unclear

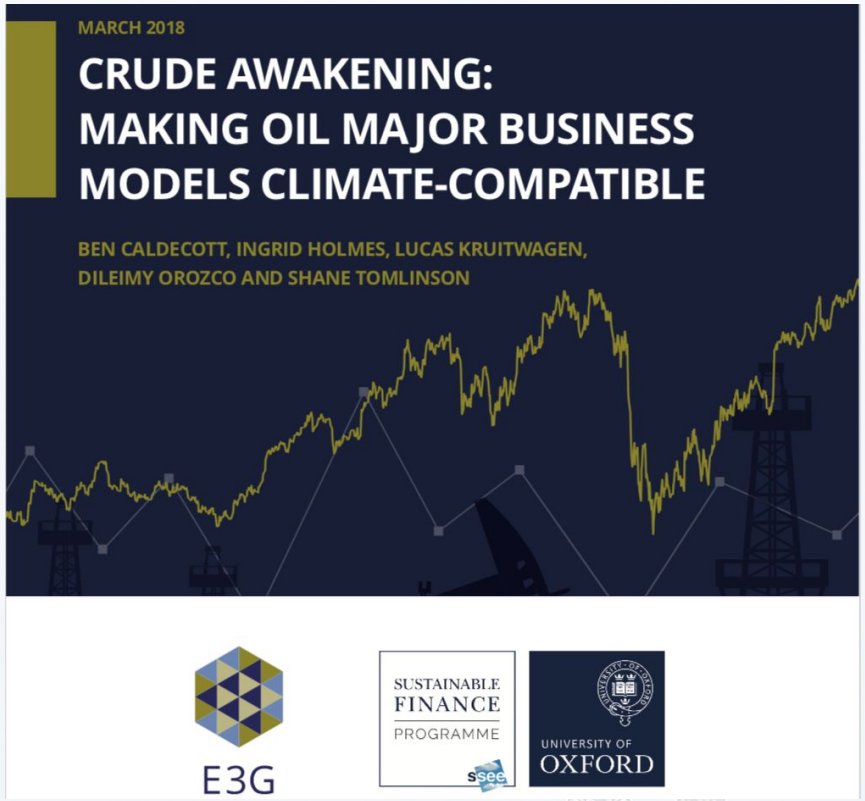
Aim

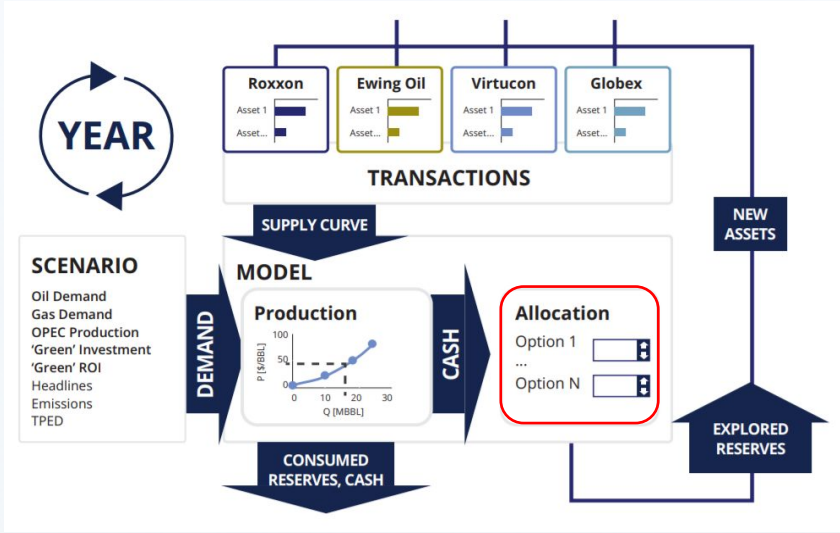
Provide tangible insights into the Majors' adaptive capacity to the energy transition by

Exploring upside and downside risks of a first low-carbon mover

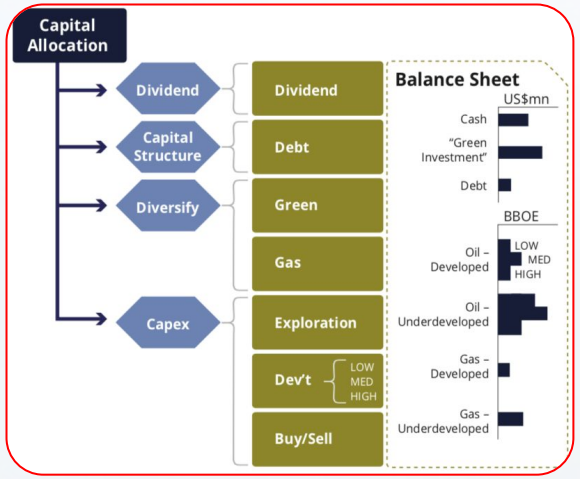
Solving for robust business pathways amidst energy transition uncertainty

Solving a Wargame





2DP Wargame Schematic



Player Allocation Options

2 Degrees Pathways Wargame – testbed for oil & gas transition

- The 2 Degrees Pathways (2DP) wargame created to

"To help inform company, investor, government, and civil society thinking around pathways the oil and gas majors can take to become 1.5C/2C-compatible"

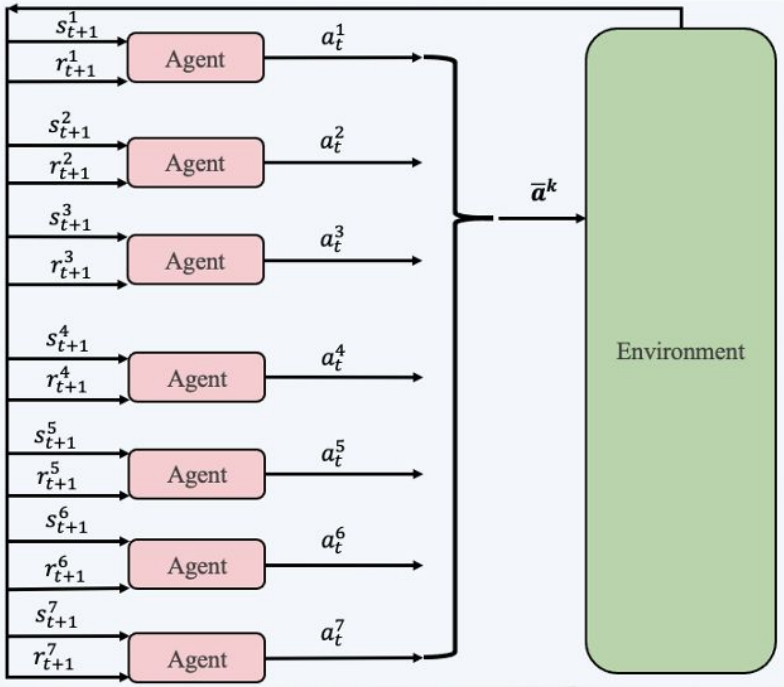
- Originally played with human players, revealing more about human bias than robust, climate-compatible pathways

Partially Observable Stochastic Game

$\langle \mathcal{I}, \mathcal{S}, \{b^0\}, \{A_i\}, \{\mathcal{O}_i\}, \mathcal{P}, \{\mathcal{R}_i\} \rangle$

Notation	Description
$\mathcal{I}$	Finite set of agents indexed 1, …, n
$\mathcal{S}$	Finite set of states $\{s^1, \ldots, s^N\}$
$b^0 \in \Delta(\mathcal{S})$	Initial state distribution
A_i	Finite set of actions available to agent i and $\vec{A} = \times_{i \in \mathcal{I}} A_i$ is the set of joint actions (i.e. action profiles), where $\vec{a} = \langle a_1, \ldots, a_2 \rangle$ denotes the joint action
$\mathcal{O}_i$	Finite set of observations for agent i and $\vec{\mathcal{O}} = \times_{i \in \mathcal{I}} \mathcal{O}_i$ is the set of joint observations where $\vec{o} = \langle o_1, \ldots, o_2 \rangle$ denotes the joint observation
$\mathcal{P}$	Set of Markovian state transition and observation probabilities, where $\mathcal{P}(s', \vec{o} \mid s, \vec{a})$ denotes the probability that taking joint action $\vec{a}$ in state s results in a transition to state s' and joint observation $\vec{o}$
$\mathcal{R}_i: \mathcal{S} \times \vec{\mathcal{A}}$	Reward function for agent i

Multi-Agent Reinforcement Learning



Framing 2DP as a Partially Observable Stochastic Game

- A continuous control problem with multiple competing entities, 2DP is best characterized as a Partially Observable Stochastic Game (POSG)–a Markovian framework resemblant of real-life market competition

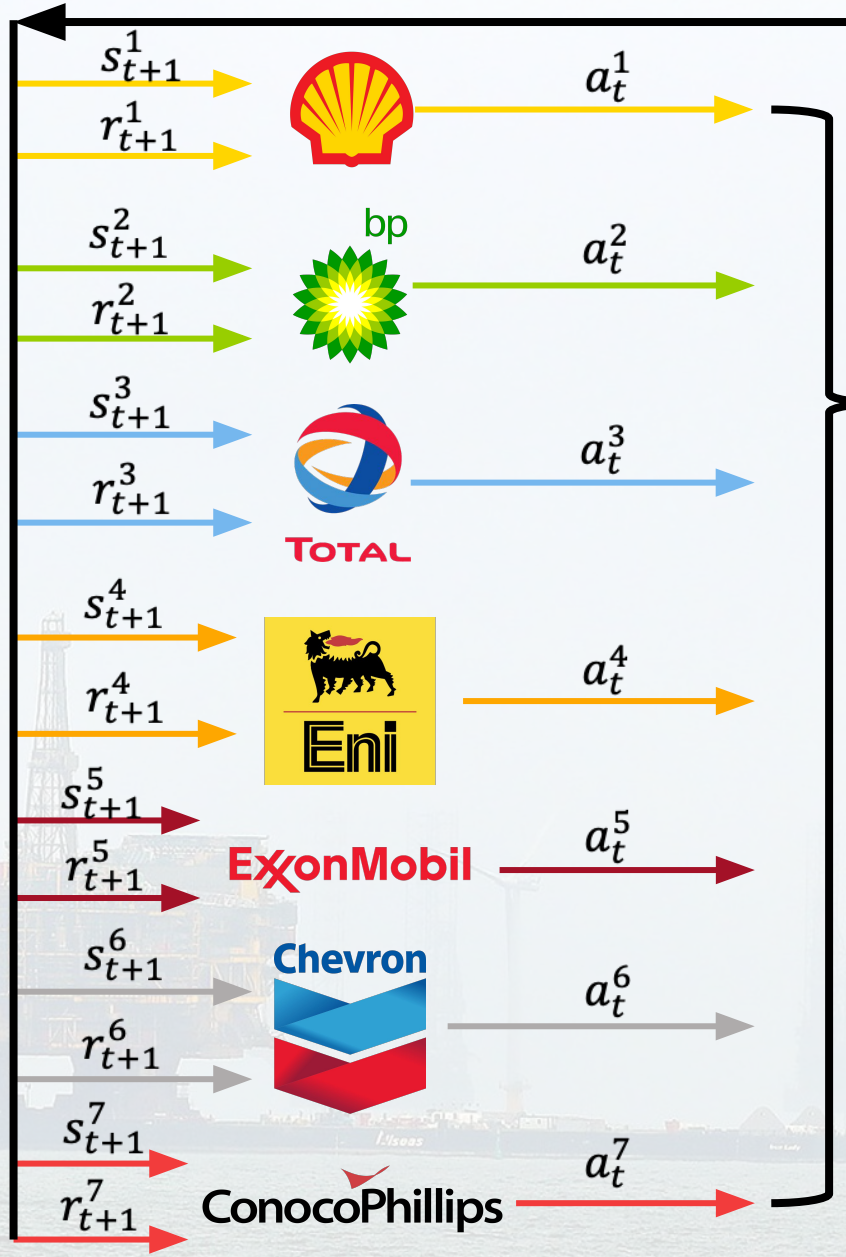
Solving with Deep Multi-Agent Reinforcement Learning

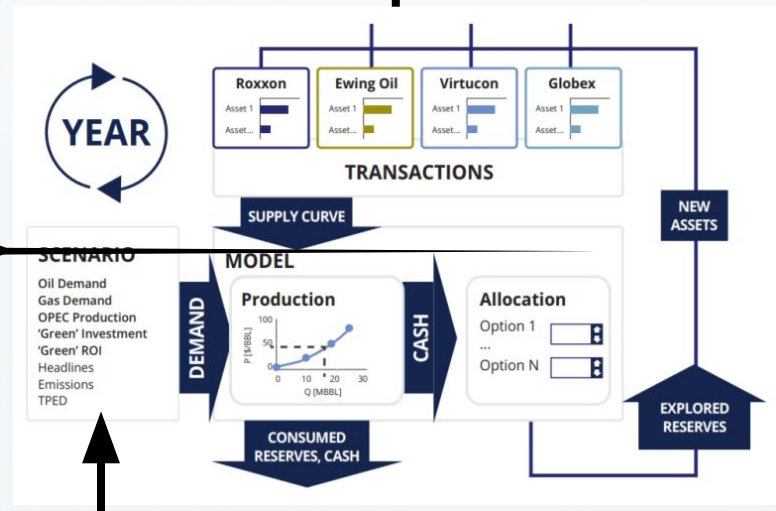
- Advances in Deep Multi-Agent Reinforcement Learning (MARL) have achieved superhuman-level performance in POSGs of high-dimensional settings

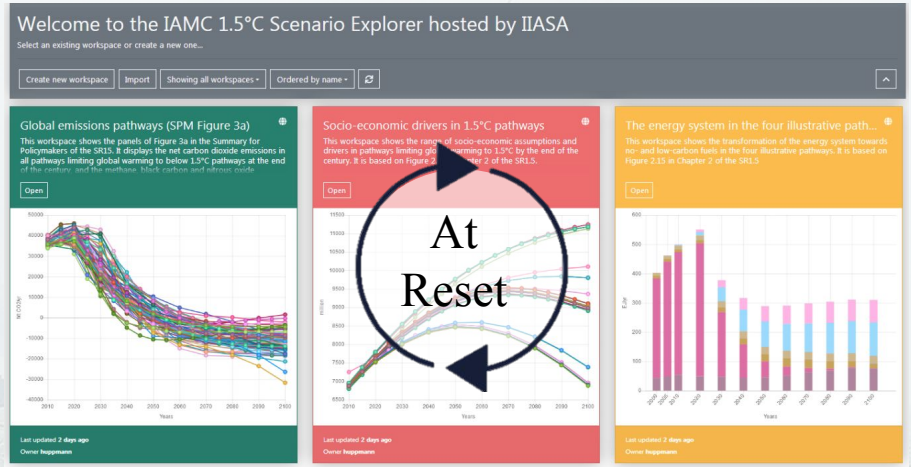
2DP-MARL

A novel microeconomics model relevant to the energy transition discussion

- 2DP-MARL combines wargaming and advances in Deep MARL to help guide the Majors' stakeholders in assessing climate-compatible business models

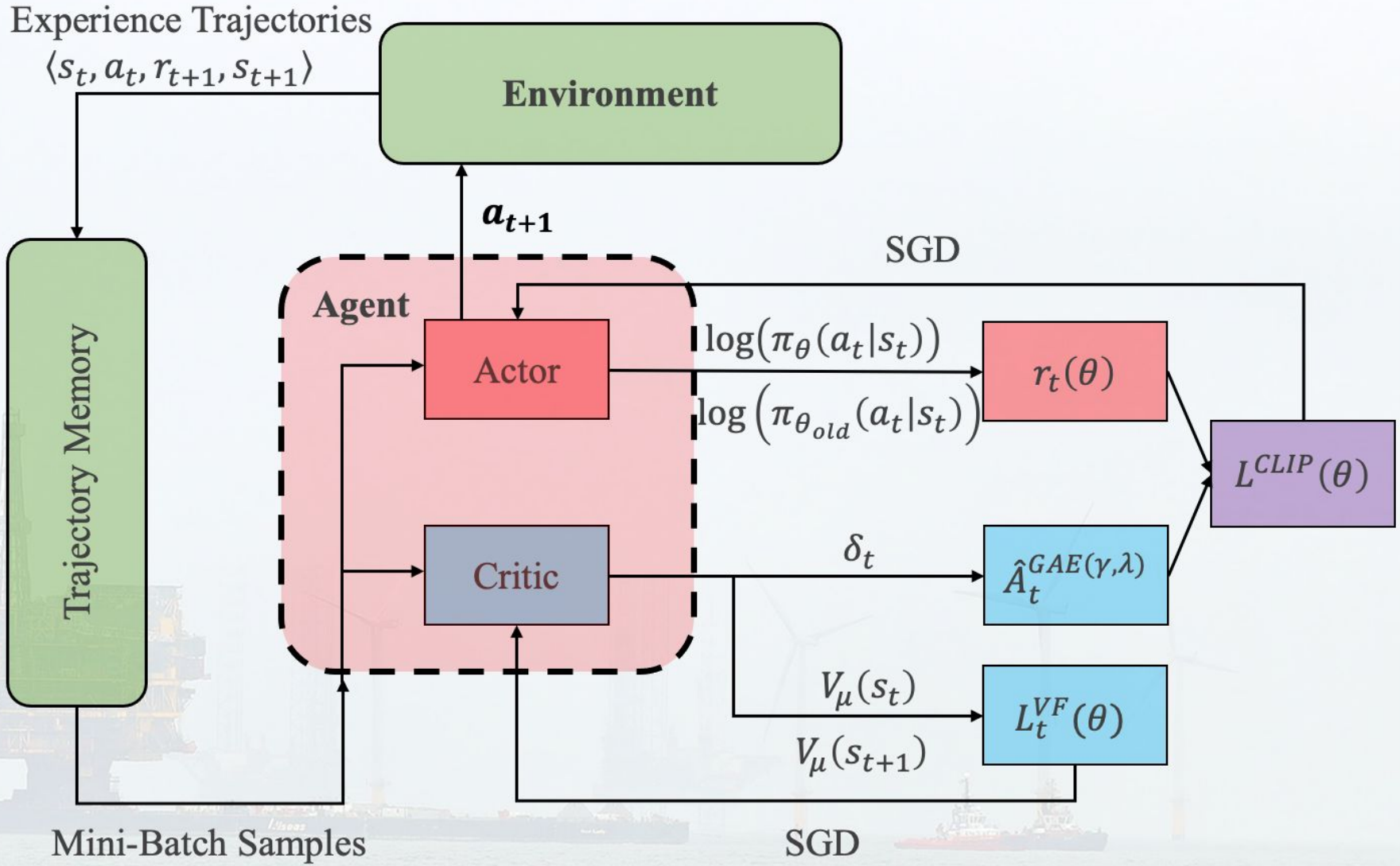






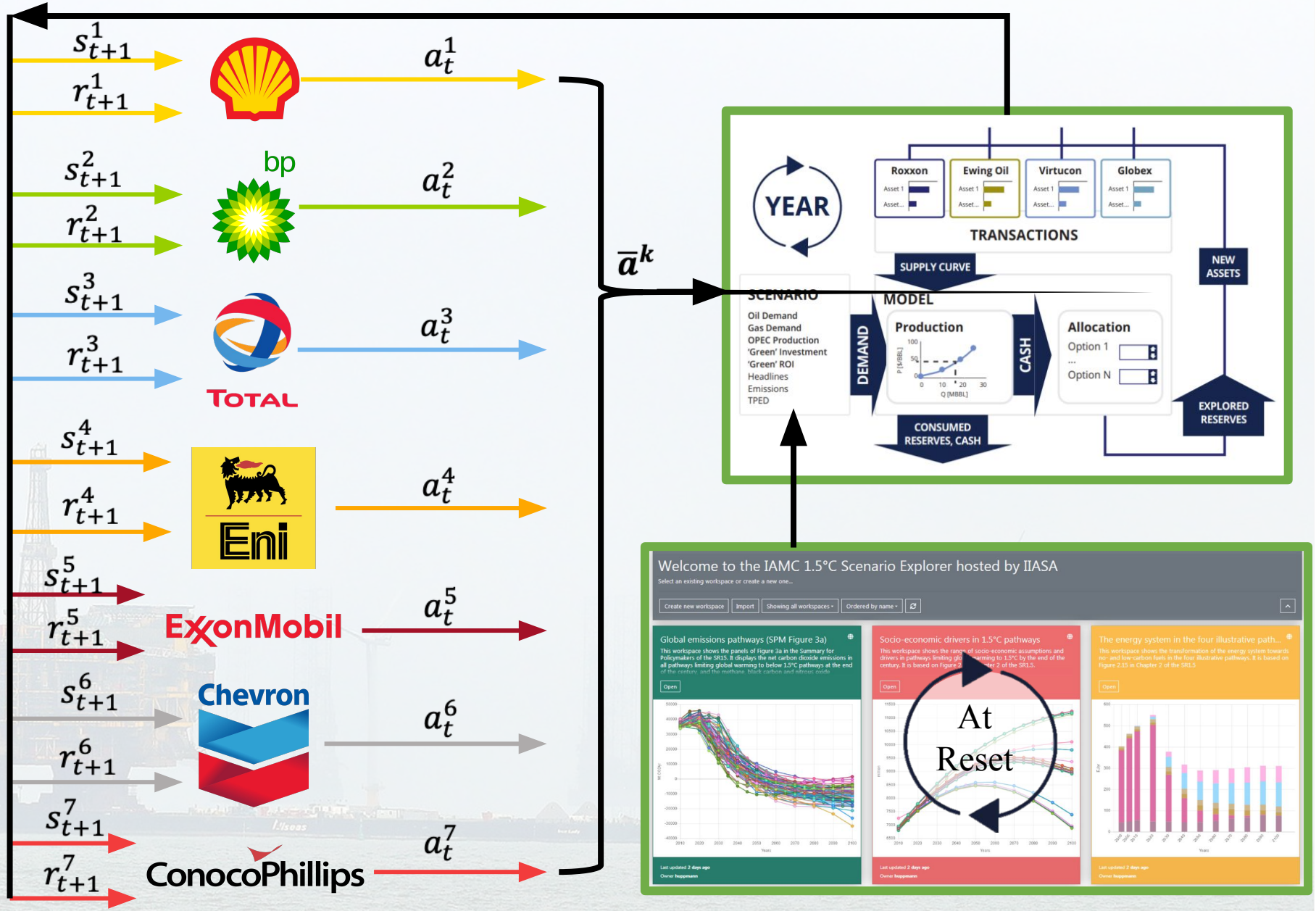
A2C/PPO as foundational algorithms

- Algorithms employed take from Advantage Actor-Critic and Proximal Policy Optimization to solve a game in a high-dimensional, continuous space



Environment

- Original 2DP Wargame, 2020-2040
- Climate-compatible scenario data generated from Integrated Assessment Modeling Consortium (IAMC) and International Institute for Applied Energy Systems (IIASA) ensemble

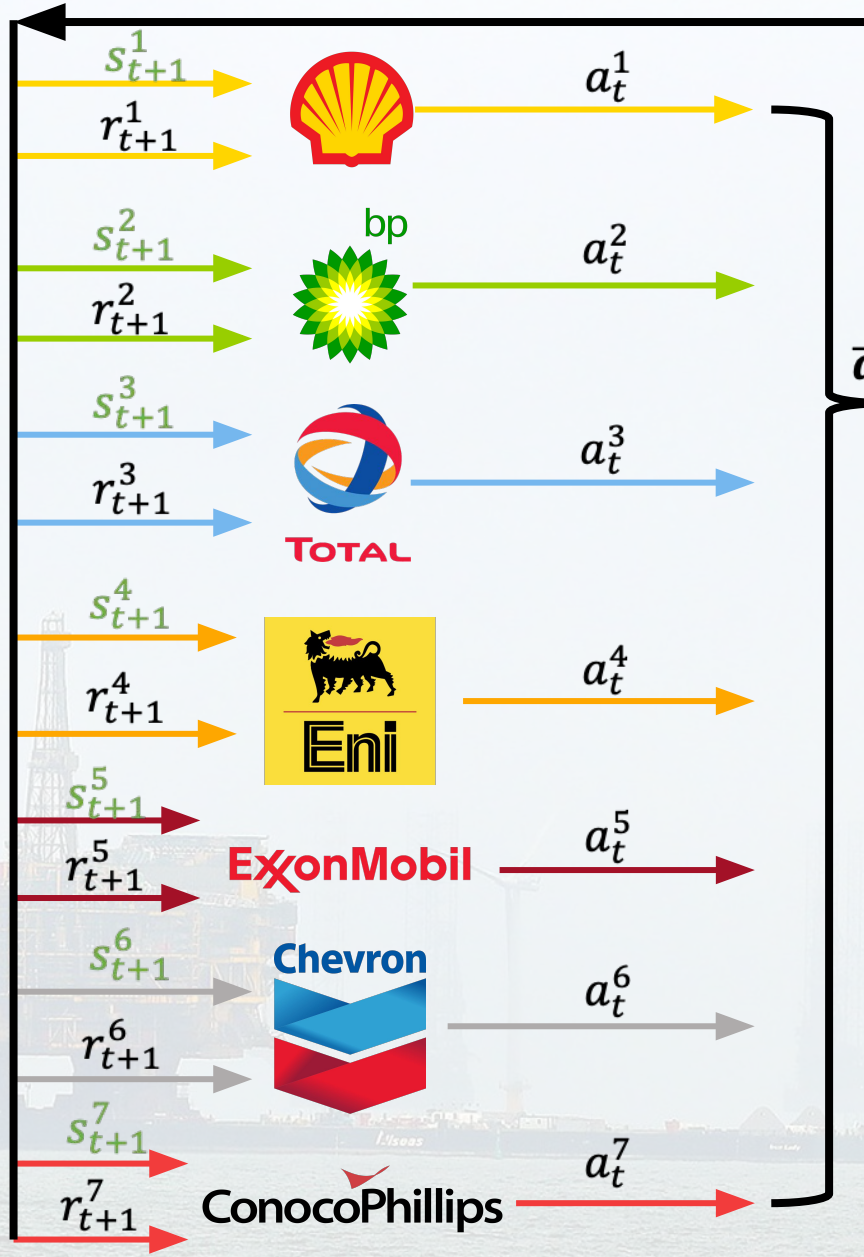


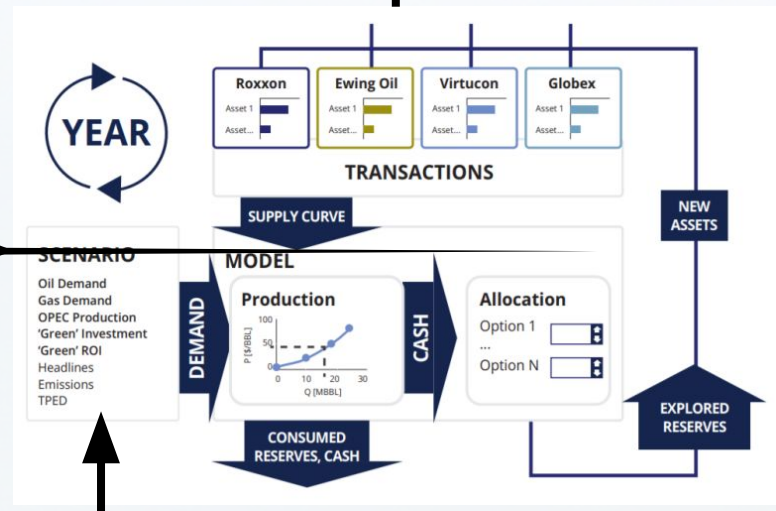
Environment

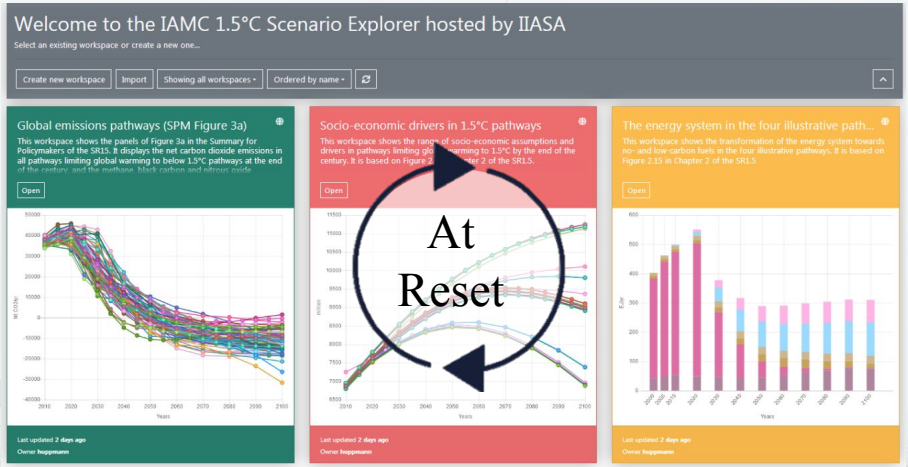
- Original 2DP Wargame, 2020 – 2040
- Climate-compatible scenario data generated from IAMC/IIASA ensemble

Observations

- Incomplete information to mimic real-life competition (i.e. players may only view some of their adversaries' assets)







Environment

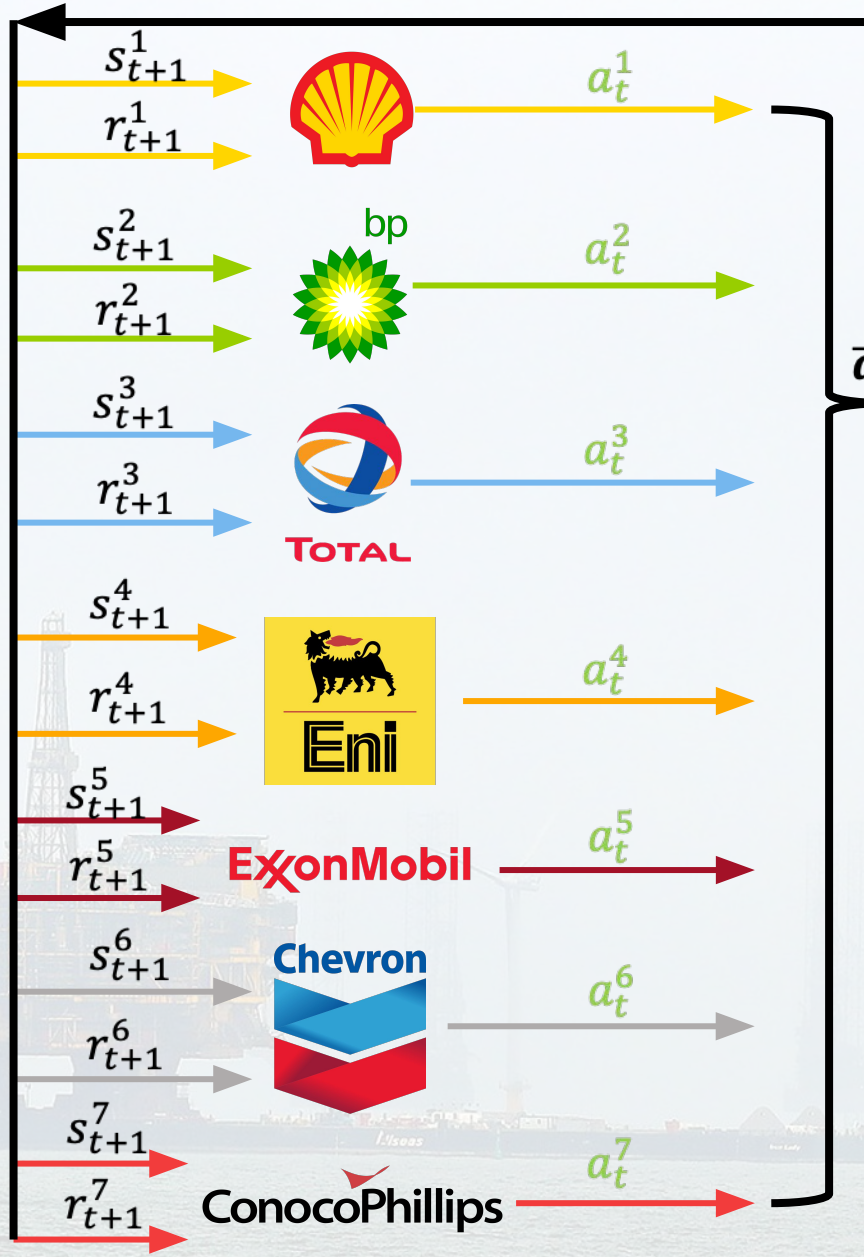
- Original 2DP Wargame, 2020 – 2040
- Climate-compatible scenario data generated from IAMC/IIASA ensemble

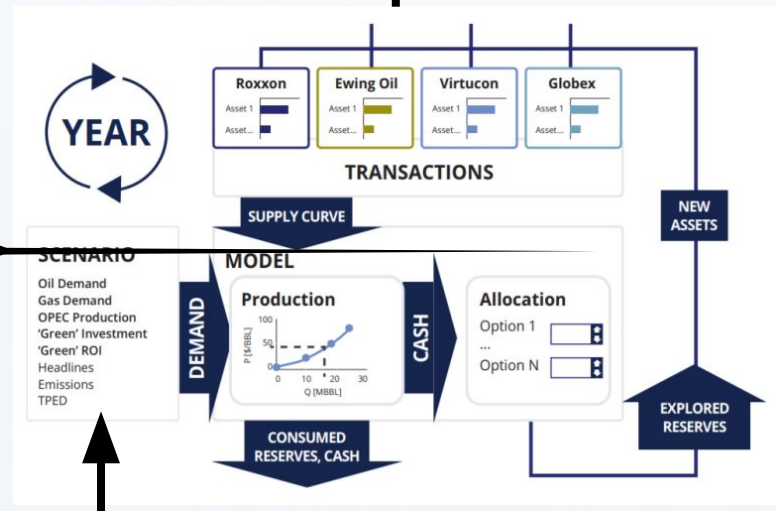
Observations

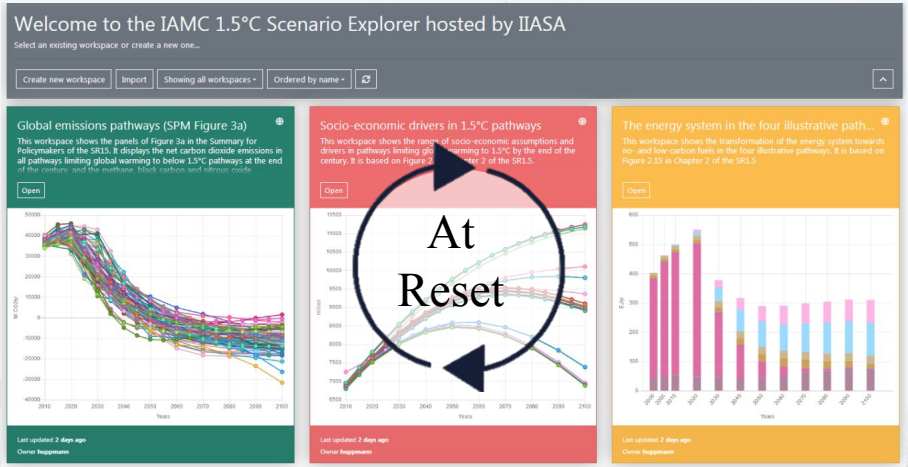
- Incomplete information to mimic real-life competition (i.e. players may only view some of their adversaries' assets)

Actions

- Similar to 2DP wargame capital allocation options, includes player-to-player trading and 'green' auction house







Environment

- Original 2DP Wargame, 2020 – 2040
- Climate-compatible scenario data generated from IAMC/IIASA ensemble

Observations

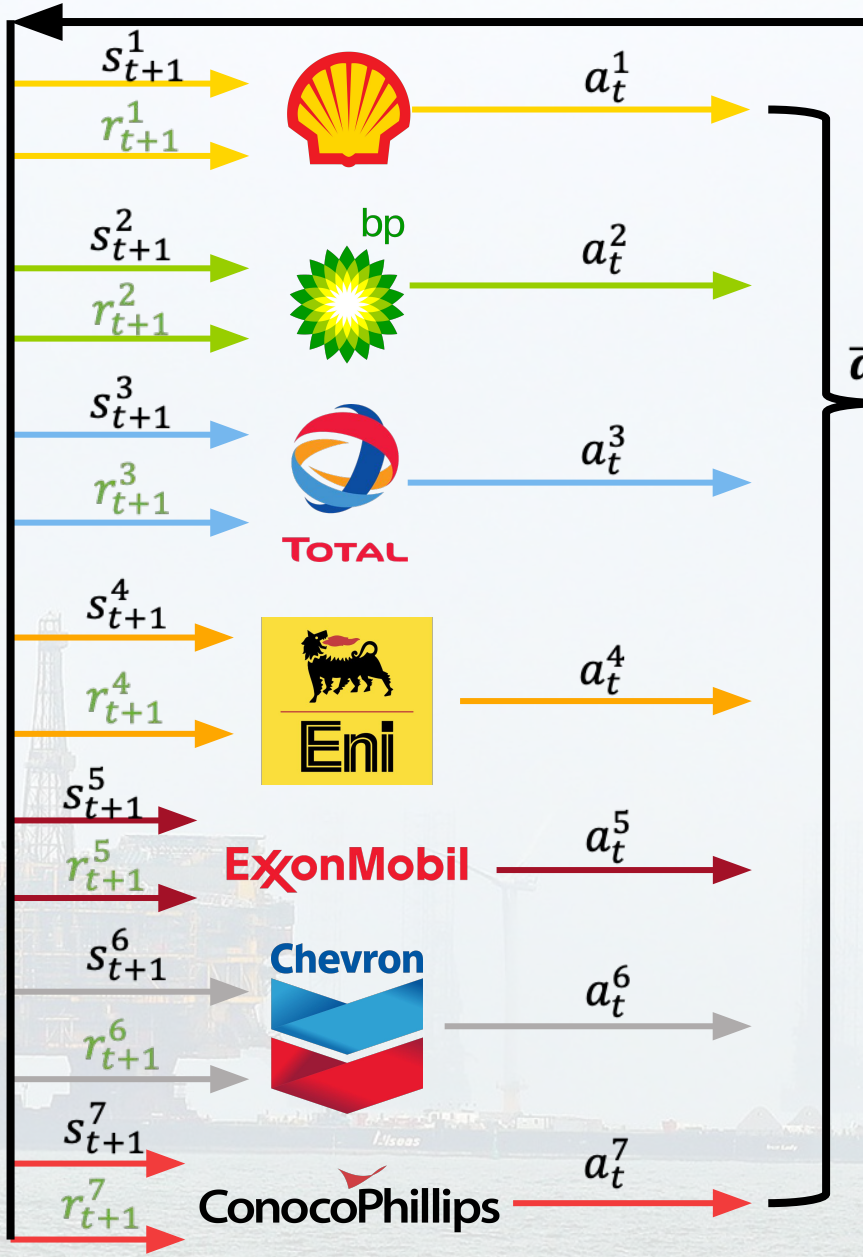
- Incomplete information to mimic real-life competition (i.e. players may only view some of their adversaries' assets)

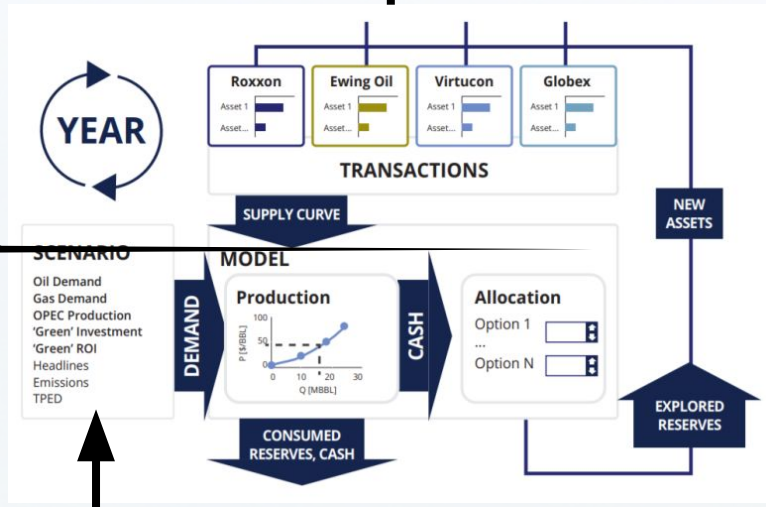
Actions

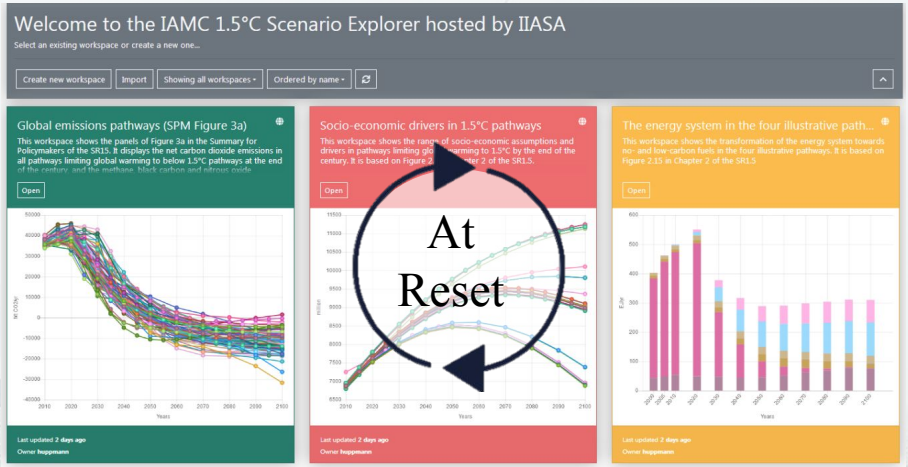
- Similar to 2DP wargame capital allocation options, includes player-to-player trading and 'green' auction house

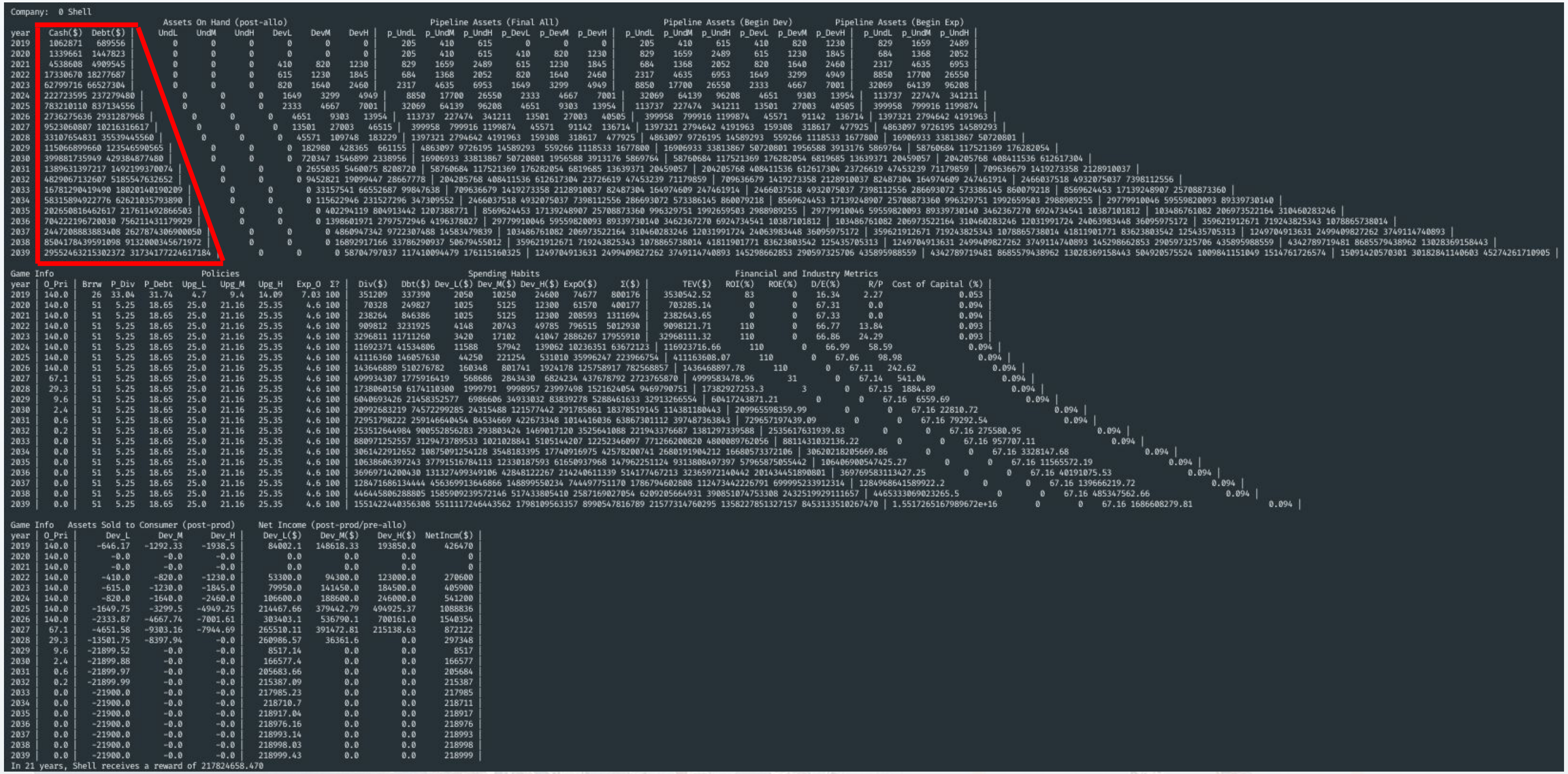
Rewards

- Focus on maximizing shareholder value through dividends
- Negative rewards to enforce realistic behavior





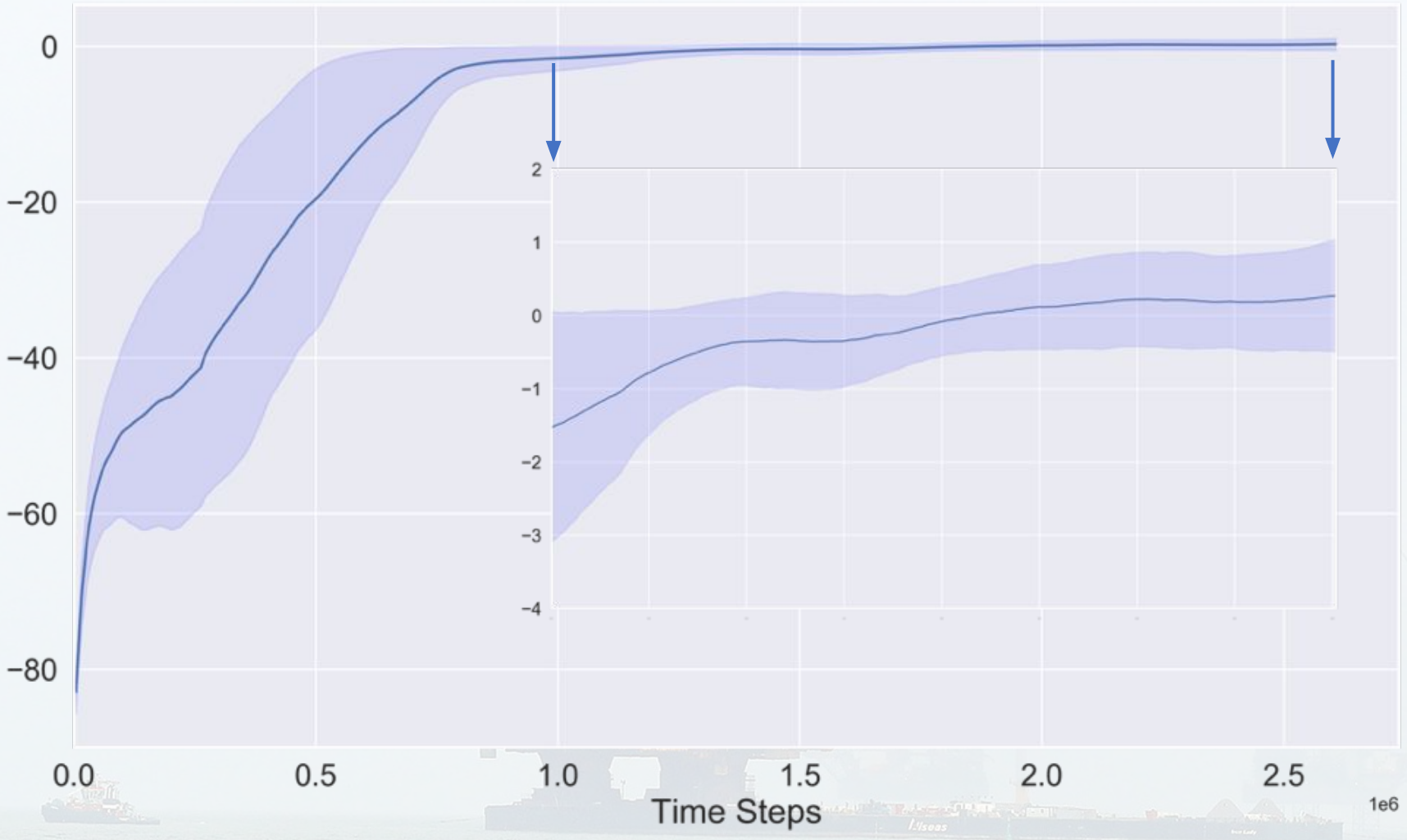




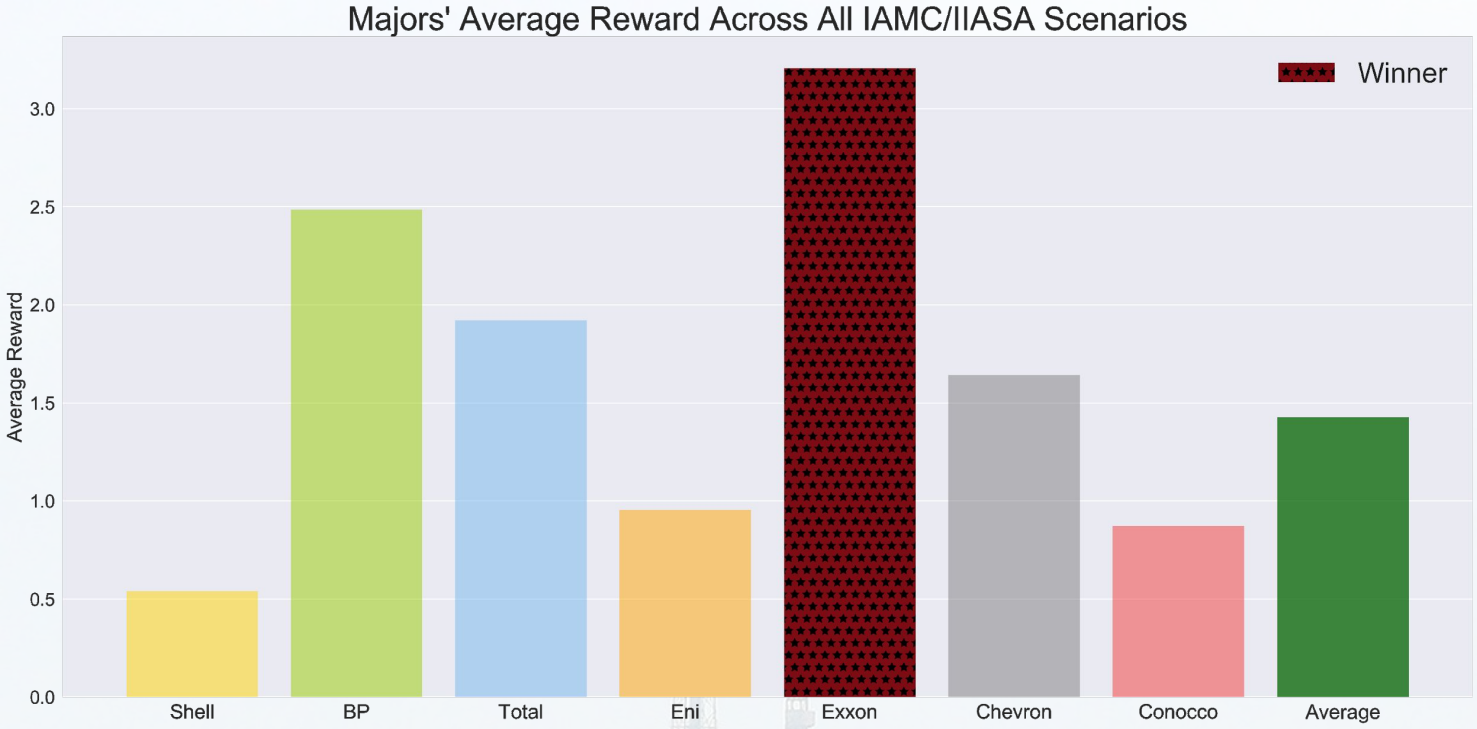
Reason for Negative Rewards:

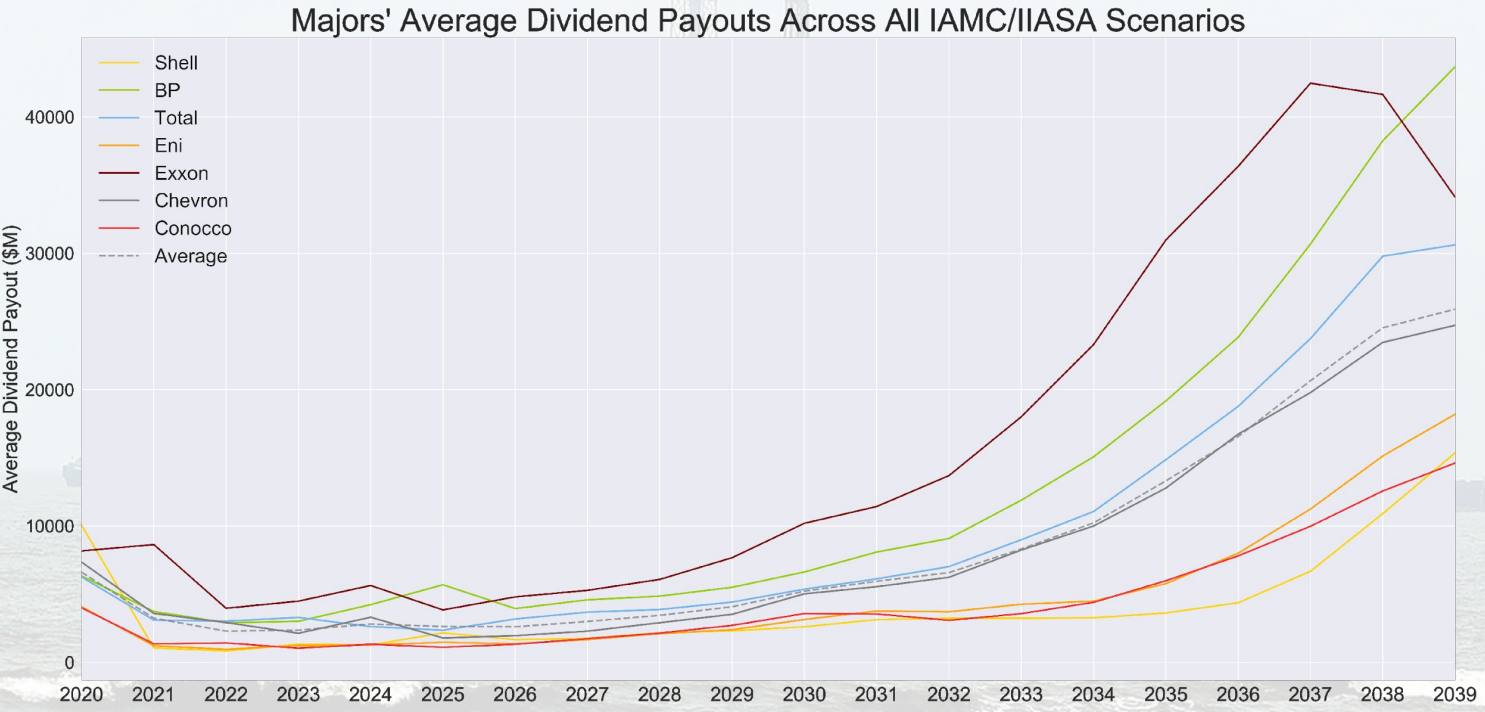
Intelligent agents love to find loopholes!

Results & Analysis

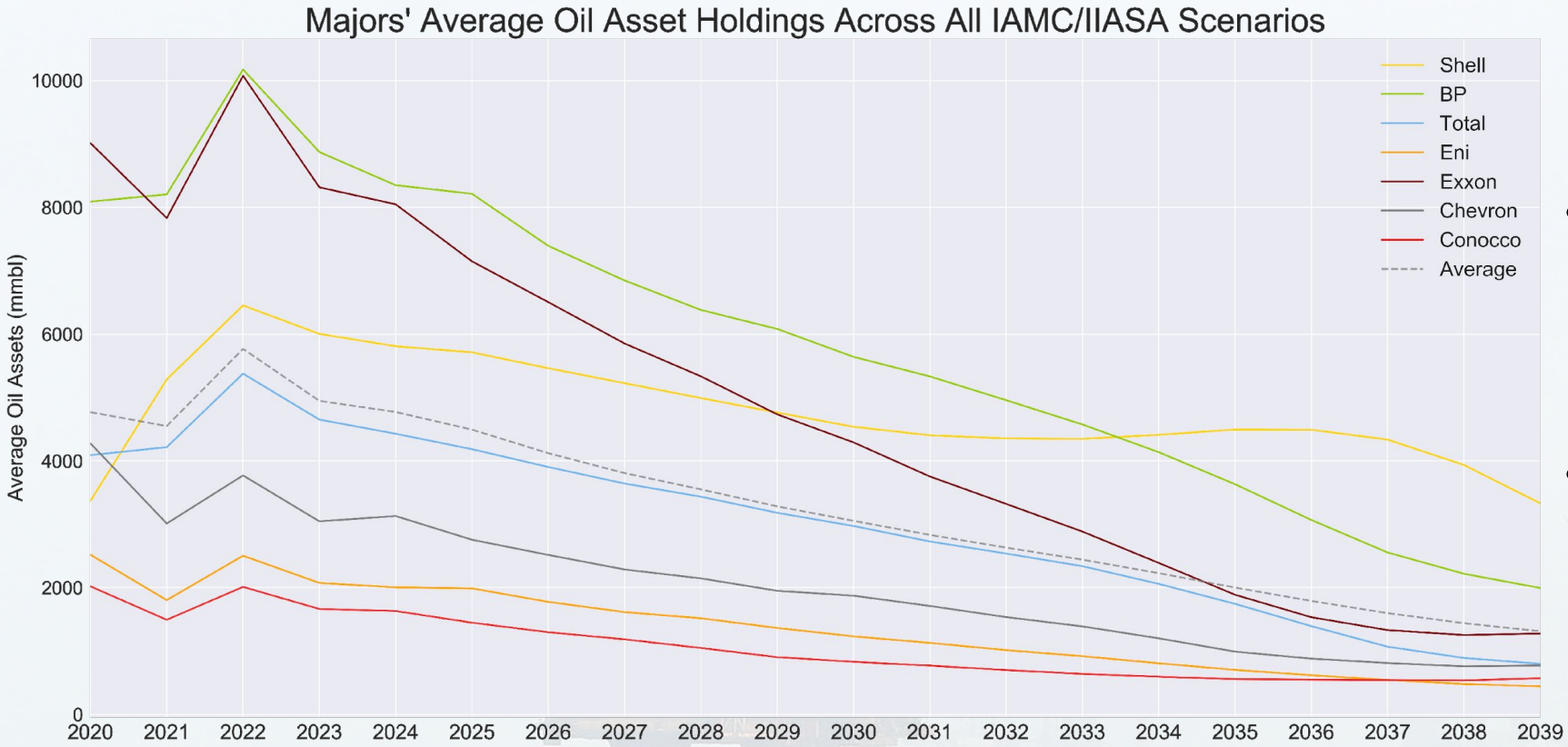


- Convergence towards a scenario-robust solution is reached after ~2.6M timesteps, or 320 pass throughs each scenario (~130,000 games played)
- Convergence towards a positive reward indicates agents achieved realistic, robust strategies to maximize dividend payouts
- Oil and gas, ‘green’ and debt holdings are explored to evaluate the Majors’ busines models

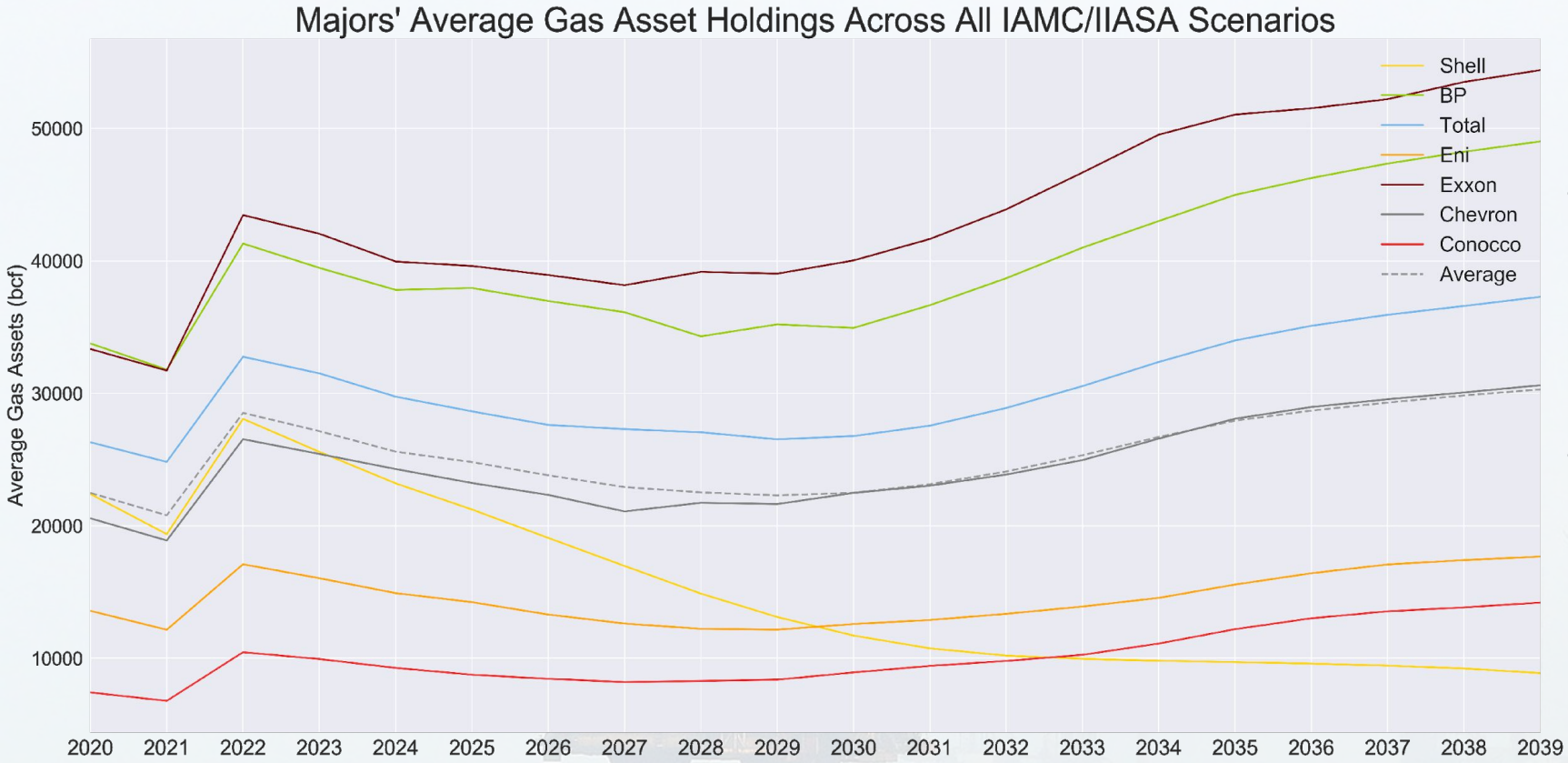




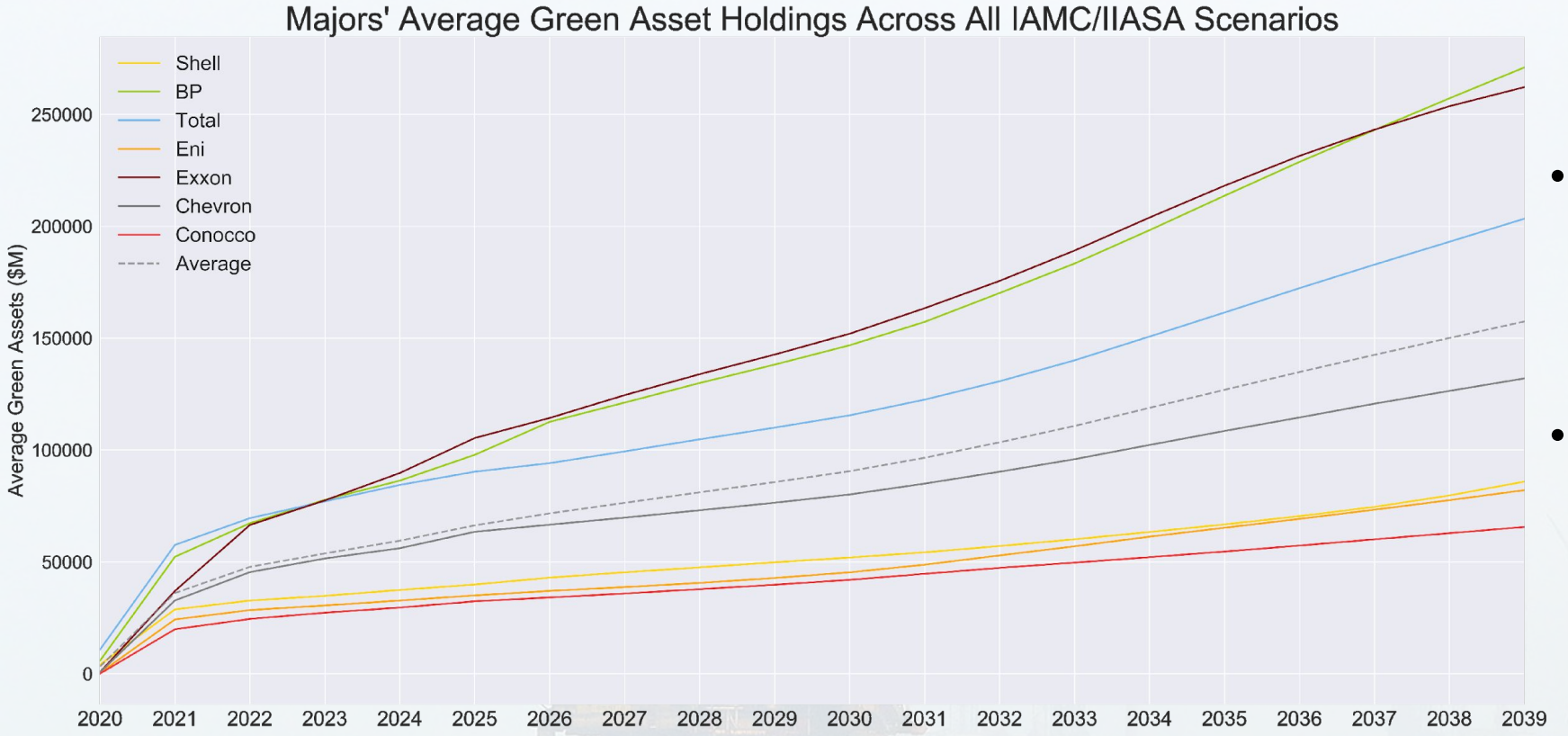
- Across all scenarios, Exxon pays out the most dividends while maintaining a realistic strategy pathway
- Average yearly dividend payouts take an initial dip before noticeably increasing in the latter half of the game



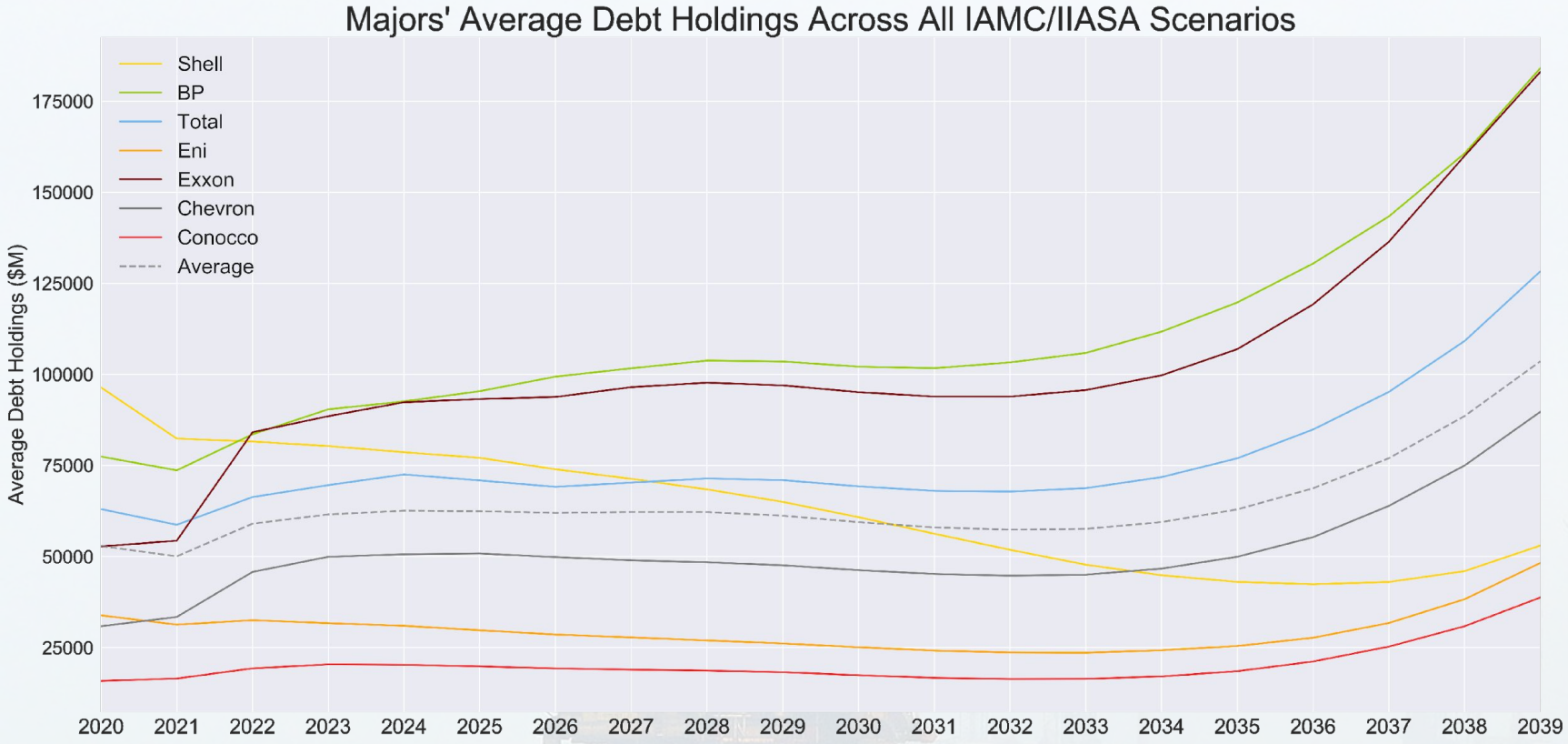
- Majors heavily divest from oil assets early, continue divestment throughout
- Shell and BP maintain the greatest oil market share at the end game



- Majors gradually increase gas asset holdings, suggesting the carbon asset's net income stability
- Gas market shares remain similar, however, Shell tends to let its market share slip



- Majors heavily invest in ‘green’ assets; display similar ‘first-mover’ behavior and long-term trends
- Exxon and BP act as leaders of the ‘green’ movement primarily due to their large, initial balance sheets



- Hints of a leverage transition (i.e. borrowing cash to buy ‘green’ assets) undertaken by BP and Total, Exxon and Chevron.
- Levels of debt increase towards the end game due to Majors’ new debt resilience

Impact

Moving First, Benefits Outweigh Costs

Moving first into ‘green’ assets allows Majors to accumulate higher returns long-term and maintain stable levels of debt

A Robust, ‘Green’ Strategy

Going ‘green’ while diminishing net income reliance on oil assets proves a robust response to uncertainty in a 1.5C world

Decarbonization of Emission Leaders

Providing a financial case for the decarbonization of the Oil Majors helps stakeholders understand the potential these carbon emission leaders have to become net-zero vanguards

Future Work

Capital cost sensitivity analysis

Modeling National Oil Companies

Expanding ‘green’ asset allocation

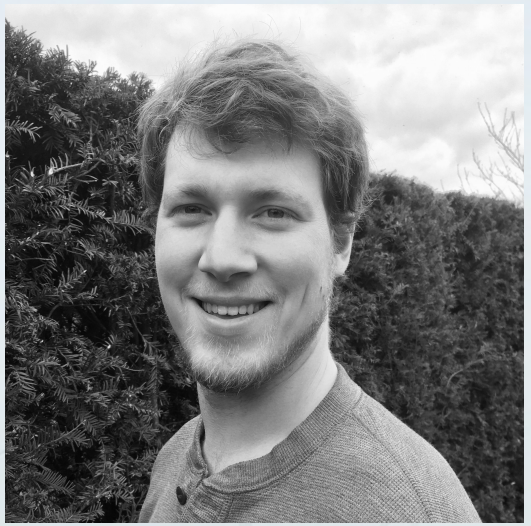
Focus on corporate governance questions

Mass rollout as 1-player game

Thank You!

Lucas Kruitwagen
University of Oxford
lucas.kruitwagen@smithschool.ox.ac.uk

Dylan Radovic
Imperial College London
dylan.radovic19@imperial.ac.uk



Christian Schroeder de Witt
University of Oxford
christian.schroeder@stcatz.ox.ac.uk