Revealing Robust Oil and Gas Company Macro-Strategies using Deep Multi-Agent Reinforcement Learning

Dylan Radovic\textsuperscript{a,1}, Lucas Kruitwagen\textsuperscript{a,b}, Christian Schroeder de Witt\textsuperscript{c}, Ben Caldecott\textsuperscript{d}, Shane Tomlinson\textsuperscript{e}, and Mark Workman\textsuperscript{f}

This manuscript was compiled on November 22, 2022 based on sources published at SSRN (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3933996) on Sept 30, 2021.

The energy transition potentially poses an existential risk for major international oil companies (IOCs) if they fail to adapt to low-carbon business models. Projections of energy futures, however, are met with diverging assumptions on its scale and pace, causing disagreement among IOC decision-makers and their stakeholders over what the business model of an incumbent fossil fuel company should be. In this work, we used deep multi-agent reinforcement learning to solve an energy systems wargame wherein players simulate IOC decision-making, including hydrocarbon and low-carbon investments decisions, dividend policies, and capital structure measures, through an uncertain energy transition to explore critical and non-linear governance questions, from leveraged transitions to reserve replacements. Adversarial play facilitated by state-of-the-art algorithms revealed decision-making strategies robust to energy transition uncertainty and against multiple IOCs. In all games, robust strategies emerged in the form of low-carbon business models as a result of early transition-oriented movement. IOCs adopting such strategies outperformed business-as-usual and delayed transition strategies regardless of hydrocarbon demand projections. In addition to maximizing value, these strategies benefit greater society by contributing substantial amounts of capital necessary to accelerate the global low-carbon energy transition. Our findings point towards the need for lenders and investors to effectively mobilize transition-oriented finance and engage with IOCs to ensure responsible reallocation of capital towards low-carbon business models that would enable the emergence of fossil fuel incumbents as future low-carbon leaders.

The strategies adopted by major oil and gas companies will help to determine their contribution to the transformation of the global energy system. Since the 2014 oil price collapse, declining performances of international oil companies (IOCs) have highlighted issues facing their business model (1) as evident with increasing upstream capital costs (2, 3) and declining returns (4, 5), dividend inflation (6) and waves of asset write-downs (7). The 9% drop in oil demand due to the COVID-19 pandemic has amplified these issues with significant losses (8), prompting cuts to capital expenditures (9) and further valuation declines (10). Furthermore, efforts to achieve the goals set out by the Paris Agreement (11) to combat climate change are gaining momentum as economies look to accelerate the low-carbon energy transition (12, 13).

The potential magnitude of energy transition risks (14) (e.g. demand shocks, impairments) are concerning investors (15) of fossil fuel companies. Coupled with IOCs’ potential capital mislocation (16), these risks could strand oil and gas assets (17–19), reduce industry revenues by potentially trillions of dollars (20), and create market instability as a result of reduced asset valuations (21–24). In response, peer-reviewed literature (1, 25–30) and industry reports (31–34) regarding oil and gas companies in the energy transition echo the same sentiment: the risks inherent to an energy transition could lead to the collapse of IOC business models, particularly the Majors, if they fail to adapt.

To mitigate downside risks, studies suggest several potential low-carbon business model opportunities and strategies as pathways towards transition- and climate-compatibility. Achieving a successful low-carbon transition requires IOCs to execute a challenging balancing act—that is, generate the necessary short-term returns for shareholders while investing in low-carbon businesses for future profitability. Although the Majors have recently announced pathways to cut carbon emissions and increase transition-oriented spending (35–38), criticisms have emerged due to likely incompatibility with the climate goals and insufficient low-carbon capital expenditures (39). The studies and criticisms (31, 33, 40–48) contributing to a widespread narrative that the Majors must change are, however, predicated on a range of energy futures scenarios with low hydrocarbon demand projections and widely varying and often not very transparent assumptions. Rigorous assessments of these energy futures and the robustness of their conclusions and what these mean for the future of the Majors remains lacking. As a result, analysis of risks and rewards of changing business models at different times and under a range of market conditions are largely missing from the literature and its low-carbon consensus.

This paper seeks to close this gap.

\textsuperscript{a}University of Oxford - Smith School of Enterprise and the Environment; \textsuperscript{b}University of Oxford - Institute for New Economic Thinking; \textsuperscript{c}University of Oxford - FLAIR, Department of Engineering Science; \textsuperscript{d}E3G; \textsuperscript{e}Imperial College London - Energy Futures Lab

Please provide details of author contributions here.

\textsuperscript{1}To whom correspondence should be addressed. E-mail: dylan.radovic@gmail.com

There are no competing interests here.
Here, we develop a data-driven approach to reveal and assess emergent IOC strategies robust\(^\text{1}\) to market and competitor uncertainty. To achieve this, we built a multi-agent system that solves an oil and gas majors wargame across the most recent collection of integrated assessment model (IAM) scenarios with deep reinforcement learning (Figure 1, Appendix Table A.2). Agents, acting as IOCs in market competition, were trained to compute an approximate best-response to varying market conditions and exploitative strategies along a 30-year time horizon. This work builds upon several studies regarding oil and gas companies in the context of climate-related risks and the energy transition as well as the utilization of multi-agent learning and deep reinforcement learning to explore emergent, robust agent behavior.

Early climate-related risk work explored the impact carbon budgets will have on fossil fuel companies (18, 49, 50). Institutions echoed these sentiments, calling for a massive reallocation of capital towards low-carbon solutions (51) and the disclosure of climate-related risks, physical and transitional, most pertinent to business activity (14). Simultaneously, energy pathways and scenarios using IAMs have been proposed, and are continually updated, to guide decision-makers on decarbonization strategies (52–55). Of significant importance to this work, studies exploring and quantifying transition risks with respect to the oil and gas industry that arise from these pathways have been elusive. This is largely due to the limitations of present scenario-analysis (56) as well as the policy insight shortcomings in the contexts of uncertainty (57). Recent studies have provided analyses on the state of oil and gas companies in the energy transition as well as suggested potential strategic responses (33, 34, 55, 58–60). Tangible upside and downside risks of their recommended strategies, however, are largely missing due to the studies’ linear assumptions and focuses on a singular energy future.

The 2 Degrees Pathways (2DP) wargaming tool61 sought to fill this research gap and inform stakeholder thinking around the macro-strategies oil and gas companies can take to become climate-compatible by simulating oil and gas companies in competition. Oil and gas competitive game theory simulations are used to enhance company strategic decision-making (62–67). Applying these conventional methods to discover effective company strategies, however, proves intractable due to the 2DP’s complexity as a high-dimensional continuous control problem (see Methods).

Advances in reinforcement learning have overcome game-theoretic challenges, successfully training agents to achieve superhuman-level performance in complex games such as Backgammon (68) and Go (69), StarCraft (70) and Dota (71). Of particular importance to this work, AlphaStar (StarCraft) and OpenAI Five (Dota) demonstrated that the combination of deep reinforcement learning and multi-agent learning can prove powerful in generating complex, robust agent behavior within high-dimensional continuous control environments. To the best of our knowledge, there is not yet a deep multi-agent reinforcement learning model that solves a wargaming tool relevant to the oil and gas industry in the energy transition until this work.

\(^\text{1}\)In this work, robust IOC strategies are defined as strategies that minimize downside risks that may arise from market uncertainty and competitor counter-strategies. A robust strategy is not necessarily one that leads to the greatest gains; multiple robust strategies may be present in a single game. Hence, the framing of 2DP as a general-sum game to allow for “win-win” scenarios (see Methods). The main IOCs training using the league mechanism described in Methods will boast robust strategies if they successfully mitigate downside risks, as indicated by the applied reward function (see Appendix Table A.6). Exploiter IOCs, on the other hand, do not yield robust strategies as they seek to optimize for opponent weaknesses, not energy futures and competitor uncertainty.

Solving a Wargame.

Core to this work, we use a variant of the 2DP wargaming tool61 game environment to simulate IOCs in competition within varying energy futures scenarios (Figure 1a). Solving 2DP to discover novel behavior, however, poses game-theoretic challenges due to the game’s non-linear payoffs, continuous high dimensional state and action space, planning horizon, and emergent counter-strategies. Furthermore, 2DP participants must balance economic decisions simultaneously given only incomplete and imperfect information of the game environment and rival players. To address these challenges and complexities, our model applies a combination of deep reinforcement learning and multi-agent learning. Details on the game environment as well as learning techniques employed are elaborated in Methods.

In the 2DP wargame, IOCs respond to varying energy futures scenario metrics, such as oil demand and low-carbon return on investment (ROI), by developing individual robust macro-strategies. These strategies emerge in the form of a combination of selected available actions: choose hydrocarbon production levels; invest in oil and gas exploration; invest in oil and gas development; invest in sustainable energy and low-carbon technologies or business models; allocate dividends; and access credit to adjust capital structure (Appendix Tables A.2,3.4, and 5). An IOC’s actions, as well as those of others, dictate the game’s core dynamics (Figure 1a). Winning the game requires an IOC to maximize shareholder value over its opponents via dividends payouts. 2DP serves as a suitable testbench to address our research questions as the creators balanced the game’s cogency and verisimilitude to enable a range of behaviors across potential energy futures and macro-strategies.

We use reinforcement learning, the computational approach to learning from environment interaction (72), to enable a 2DP IOC to discover scenario-robust strategies that maximize unlevered dividend payouts (see Methods). The learning IOC achieves this by computing approximate best-responses to limited game state information, to best mimic real-life market competition, and a real-valued reward signal—a value predicated on the efficacy of its developed strategy towards achieving its goal (see Appendix Table A.6). Our training method follows an independent learning (InRL) approach whereby a single learning IOC competes against non-learning IOCs (i.e. IOCs playing previously learned strategies) that are chosen by the multi-agent learning mechanism described below (Figure 1b). We equip our IOCs with deep neural networks to alleviate concerns regarding 2DP’s non-linearities, high dimensional state and action spaces, and planning horizons.

Discovering novel, robust strategies in 2DP is difficult due to its complex dynamics and game-theoretic challenges. Self-play reinforcement learning algorithms train a learning agent by simulating play against itself. Successful applications of self-play to achieve superhuman-level performance are seen in games such as Backgammon (68) and Go (69), StarCraft (70) and Dota (71). Despite resulting in emergent, complex behavior, agents trained with self-play are susceptible to forgetting (73) (i.e. improve against itself but fail to win against past versions) and training imbalances (74). Fictitious self-play (75, FSP) solves these issues by uniformly sampling opponents from past versions of the learning agent. DeepMind (75) extended this approach with prioritized fictitious self-play (PFSP) to train against a non-uniform mixture of opponents by focusing on the most difficult of agents. To address game-theoretic challenges and make our IOCs’ strategies more robust, we use a combination of the aforementioned self-
Fig. 1. The model and training setup. a, The 2DP wargame involves multiple players competing as IOCs, reminiscent of major oil and gas companies, within varying energy futures scenarios. IOCs select actions based on the global scenario metrics given to them, as well as information about themselves and their opponents, that dictate a year’s production, cash borrowing, trading and low-carbon asset auction bidding, and capital allocation (Appendix Tables A.1). The game advances to the next year after the allocation stage. Upon reaching the endgame, a new energy scenario is chosen to alter dynamics of the next game. b, We train our 2DP IOCs with deep reinforcement learning. The learning IOC collects game experience in the 2DP wargame via game state observations and a reward signal to update its strategy. Frozen (non-learning) IOCs are chosen by the multi-agent learning mechanism at play and compete against the learning IOC. c, We use a league training system involving two learning IOCs to address game-theoretic challenges. Main IOCs do not have strategy constraints but vary in initial asset distributions and are initially trained against copies of themselves. Exploiter IOCs, constrained to certain strategies (e.g. BAU, delayed transition 2030), play algorithms to create a league of learning IOCs, similarly used by DeepMind to discover novel strategies in a real-time strategy game (70).

Our 2DP-league mechanism involves two distinct learning IOCs that differ in the distribution of opponents they play against, when training parameters are reset, and the availability of actions (Figure 1c). Our main IOCs train as energy companies with the goal of learning robust, investor-attractive strategies utilizing a mixture of self-play and PFSP. Exploiter IOCs, on the other hand, train as energy companies with the goal of finding strategy weaknesses in the latest versions of the main IOCs. Unlike the main IOCs, constraints are imposed on the actions available to them so they become dominant with respect to the strategies they are forced to play. We focus on two constrained strategies: BAU and delayed transition. The former restricts an IOC from entering low-carbon markets thereby optimizing strategies in oil and gas markets, the latter does the same until a given year (e.g. 2030). This method encourages main IOCs, later trained against the exploiters, to adapt to such extreme strategies. Unlike the exploiter IOCs, training parameters for the main IOCs never reset, enabling continuous learning.

We train eight IOCs—six main, two exploiters (one BAU, one delayed transition)—in a six player 2DP game by training each IOC consecutively against five non-learning opponents as described by the multi-agent mechanism (see Methods). IOCs are initiated with the average 2DP-relevant balance sheet items drawn from six oil Majors’ annual reports (76) to best represent market conditions (Appendix Table A.8). The six main IOCs begin with differing asset distributions to demonstrate how an initial market dominance, or diversity, may affect strength of overall strategies. Our exploiter IOCs each begin with greater oil market share, optimizing either a BAU or delayed transition strategy. An IOC’s training iteration plays through 408 unique energy futures scenarios drawn from the Integrated Assessment Model Community (IAMC) and International Institute for Applied Systems Analysis (IIASA) 1.5°C explorer database (52) and the IEA’s Net-Zero report (55) (Appendix Table A.1). The IAMC/IIASA collection was chosen as it is the most recently updated ensemble of scenarios to span a range of energy futures within varying degree-warming scenarios, from below 1.5°C to above 4°C. Each main IOC is trained for five iterations.

Emergent, robust IOC strategies.

After training, we evaluated the emergent strategies of the main IOCs when competing amongst varying combinations of themselves and the two exploiter IOCs, sampled from the league. Doing so required testing across 11,424 games (28 unique opponent match-ups, 408 energy scenarios), with each IOC playing 8,568 games (21 unique opponent match-ups, 408 energy scenarios). Matches were played under similar conditions as seen in training yet with less scenario metric noise (Appendix Tables A.2, and 3). IOCs played mixed strategies as non-deterministic play best represents real-life market competition (see Methods). To assess performances between strategies, we compared their individual cumulative unlevered dividend payouts with the total amount of unlevered dividend payouts allocated by each IOC across all games. We observe whether greater initial market share, or diversity, will result in higher total dividend payouts (Figure 2a). IOCs initialized
Fig. 2. Emergent strategies from main IOCs. a-c. Resultant dividend payouts as compared to total dividend payouts within each game for (a) each main IOC by initial asset portfolio, (b) top oil producers, (c) top gas producers. d-f. Frequency of low-carbon strategy emergence, and their average resultant dividend payouts as compared to total dividend payouts within each game, with respect to (d) the first year in which an IOC’s low-carbon capital expenditures eclipse hydrocarbon capital expenditure, (e) the year an IOC’s low-carbon assets are valued higher than their hydrocarbon reserves, (f) the significance of credit used to facilitate a complete low-carbon transition as shown in (e). g-h, Emergent main IOC strategies with respect to (g) business model as determined by which market an IOC allocates capital to most, (h) the total cashflow distribution whereby the preceding market indicates the dominant source of income (e.g. Oil | Low-Carbon indicates oil cashflows are greater than low-carbon cashflows). i-j, Distribution of dividend payouts as compared to total dividend payouts within a given game separated by six degree-warming scenarios (1.5°C ± 4°C) for all emergent strategies demonstrating (i) low-carbon business models as a result of early low-carbon movement (E-LC), (j) oil-focused business models with early movement towards low-carbon (E-O), (k) low-carbon business models as a result of mid-term low-carbon movement (M-LC). Such strategy combinations were obtained by drawing insight from (d) and (e). Each E-LC and M-LC strategy combination boasted positive rewards signals, indicating such strategies successfully mitigated debt engulfment risks (see Appendix Table A.8).
with greater amounts of low-carbon assets, apart from Gas-LC, tend to perform better than those with initial portfolios focused on hydrocarbon markets. While the Oil-LC IOCs boast strategies that best its competitors, the greater market positioning in low-carbon markets proves more advantageous than equivalent positioning in oil as evidenced by the LC IOCs performance over its Oil IOC peer. Moreover, we find that IOCs initialized with competitive advantages in gas markets noticeably underperform IOCs with advantages focused on the low-carbon or oil markets.

To evaluate the IOCs' tactics across the oil, gas, and low-carbon markets, we examined several, distinct categories with mutually exclusive strategies (Figure 2b-f). With respect to oil markets, IOCs cutting production to quickly exiting the market risk underperformance. Increasing gas production, however, does not prove a favorable strategy for the IOCs, suggesting the carbon-intensive asset's low returns. Observing low-carbon market behavior, we find emergent IOC strategies demonstrate considerable movement towards low-carbon business models by 2024 with greater preferences towards allocating most capital towards low-carbon within the first two years of play (76%, Figure 2d). From a dividends perspective, IOCs that move into low-carbon early outperform mid-term movers. Considerable capital movement towards low-carbon continues as 99.99% of IOC strategies focus on completely transitioning to low-carbon business models—that is, when an IOC's low-carbon assets are valued higher than their hydrocarbon reserves—a majority of which occurring within the mid- to late-2020s (74%, Figure 2e). Moreover, we discover that the use of credit, coupled with hydrocarbon production returns, was instrumental in facilitating a complete, robust low-carbon transition as well as developing high-dividend payout strategies (Figure 2f).

The aforementioned energy market tactics resulted in substantial similarities across all strategies' endgame business models and total cash flow distribution (Figure 2g,h). Observing capital allocation behavior across the three markets, we find that 90.94% of emergent IOC strategies allocate considerable transition-oriented capital to facilitate transitions towards low-carbon business models by the endgame (Figure 2g). Oil-focused business models accounted for the remaining 9.06%. To fund these energy investments, the IOCs' main source of income is split between oil and low-carbon cashflows (Figure 2h). For 57.39% of emergent strategies, oil production generated a majority of an IOC's income with supporting revenues from the acquired low-carbon assets. Remaining strategies (42.61%) resulted in incomes with greater low-carbon revenues, outweighing both oil and gas cashflows.

We further examine the performance of emergent strategy combinations comprising of initial low-carbon movement (Figure 2d) and business model (Figure 2g) strategies for all IOCs across several degree-warming scenarios (Figure 2i-k). With 51, 408 opportunities (6 main IOCs, 21 matchups, 408 energy scenarios) to demonstrate trained, robust behavior, three distinct strategy combinations emerged: early and mid-term movement towards low-carbon both resulting in a low-carbon business model (E-LC, M-LC, respectively) and early movement towards low-carbon yet with total capital focused on sustaining oil business models (E-O). E-LC, M-LC, and E-O strategy combinations emerged across 75.36%, 15.38%, and 9.06% of the 51, 408 opportunities. On average, E-O strategy combinations tend to outperform their E-LC and M-LC peers across most degree-warming scenarios and variations of competing strategy combinations, including the exploiters. In 1.5°C and below degree-warming scenarios, however, E-LC IOCs tend to allocate the greatest share of dividends pointing towards the strategies' compatibility with climate policies. The performances of the E-O and M-LC IOCs increase with degree-warming scenarios, owing to their continued hydrocarbon production focuses that present greater upside risks with increasing demand curves.

We further examine the combination of low-carbon movement timing and business model strategies as they perform against varying opponent combinations of themselves, and exploiters, across different degree-warming scenarios (Figure 3). Exiting a market (e.g., oil, gas) to enter development into a new one (e.g., low-carbon) is a possible disadvantage for early movers as it creates exploitable gaps in the former market (77). This disadvantage is of particular concern to IOCs in the event of high, continuously increasing hydrocarbon demand. The equilibrium match, containing only the main IOCs, supports this notion as the average E-O IOCs outperform the average E-LC and M-LC IOCs in most degree-warming scenarios as IOCs focusing on oil development benefit from hydrocarbon market gaps (Figure 3a-c). Beyond equilibrium and involving exploitive strategies, however, E-LC IOCs outperform most variations of E-O, M-LC, Exp-B-O (BAU IOC), and Exp-D-O (delayed transition IOC) strategy combination matchups and degree-warming scenarios with (Figure 3d-s, Appendix Figure B.8). The E-LC strategy's robustness is most apparent in climate-compatible scenarios (1.5°C and below, 1.75°C and below, 2°C and below), apart from Match 2 and 3’s 1.5°C scenarios, whereby its dividend allocations remain higher than that of the next-best performer by margins ranging from 1.6 – 10.4%. Notably, performances of exploiter IOCs increase with levels of degreewarming yet fail to payout higher dividends than any of their peers. While E-LC IOCs suffer from underperformance in equilibrium, the early low-carbon movers effectively mitigate concerns of exploitation from Exp-B-O and Exp-D-O strategy combinations.

Oil price forecasts are critical for internal stress testing of hydrocarbon business models to oil price volatility (3) and are typically averaged to a single price for several years (44). We compare the upside and downside risks of the emergent strategy combinations within selected matches, in parallel with Figure 33, when exposed to varying oil prices. From a high-level, in equilibrium, E-LC IOCs are less exposed to the downside risks of oil price drops, yet are unable to attain the potential gains of higher oil prices similar to E-O IOCs (Figure 4a,b). The addition of exploiter IOCs, however, challenges the E-O IOCs strategies resulting in decreased possibility of gains with high oil prices and increased susceptibility to low oil prices (Figure 4e,i,q). E-LC IOCs successfully mitigate downside risks as well as maintain, and, in some cases, even increase, upside risks when compared to the E-O IOCs (Figure 4d,h,p). This is largely due to the oil-focused IOCs’ abilities to benefit from market gaps when exploiter IOCs are involved.

Examining the sensitivity of the exploiter IOC’s dividend payouts to oil prices, we find Exp-B-O strategies can maximize their upside risks, until the late 2020s, yet fail to mitigate downside risk exposure to low oil prices (Figure 4f,m,r). Moreover, Exp-B-O strategies’ dividend payouts fall sharply towards the endgame despite high oil prices as compared to its adversaries. The sudden fallout points towards the Exp-B-O IOC’s inability to replenish reserves and the main IOC’s buildup of high low-carbon returns as a result of earlier transition strategies. Exp-D-O IOCs are unable to balance the upside and downside risks as delaying movement into low-carbon risks failure to facilitate a transition and acquire desirable returns in a timely manner.

---

2We define ‘early’ (E) movers as IOCs that allocate most capital expenditures toward low-carbon within the first three years of play; ‘mid-term’ (M) movers are IOCs that exhibit this behavior after 2022, but before 2025.
Fig. 3. Average dividend payouts of initial low-carbon movement (Figure 2d) and business model (Figure 2g) strategy combinations within selected matches of similar or differing strategy combinations across degreewarming scenarios. a-c, Average dividend payouts as compared to total dividend payouts seen in Match 1 (main IOC approximate mixed-strategy Nash equilibrium) for the average (a) E-LC, (b) E-O, (c) M-LC strategy combinations. d-g, Average dividend payouts as compared to total dividend payouts seen in Match 2 for the average (d) E-LC, (e) E-O, (f) Exploiter-B-O, (g) M-LC strategy combinations. h-k, Average dividend payouts as compared to total dividend payouts seen in Match 3 for the average (h) E-LC, (i) E-O, (j) Exploiter-D-O, (k) M-LC strategy combinations. l-o, Average dividend payouts as compared to total dividend payouts seen in Match 4 for the average (l) E-LC, (m) Exploiter-B-O, (n) Exploiter-D-O, (o) M-LC strategy combinations. p-s, Average dividend payouts as compared to total dividend payouts seen in Match 4 for the average (p) E-LC, (q) E-O, (r) Exploiter-B-O, (s) Exploiter-D-LC. Most matches did not occur across all 408 scenarios due to sensitivity of strategies in response to different game scenario metrics (e.g. a single M-LC strategy combination seen in Match 1 initiated earlier lowcarbon movement in certain scenarios thereby becoming an E-LC strategy combination). See Appendix Figure B.8 for additional unique matches, with greater than 300 unique scenarios played, and their strategy combinations’ resulting dividend payout performance.
Fig. 4. The sensitivity of yearly dividend payouts to varying oil prices for strategy combinations within five unique matches. a-c, Sensitivity of yearly dividend payouts to varying oil prices seen in Match 1 for average (a) E-LC, (b) E-O, (c) M-LC strategy combinations. d-g, Sensitivity of yearly dividend payouts to varying oil prices seen in Match 2 for average (d) E-LC, (e) E-O, (f) Exploiter-B-O, (g) M-LC strategy combinations. h-k, Sensitivity of yearly dividend payouts to varying oil prices seen in Match 3 for average (h) E-LC, (i) E-O, (j) Exploiter-D-O, (k) M-LC strategy combinations. l-o, Sensitivity of yearly dividend payouts to varying oil prices seen in Match 4 for average (l) E-LC, (m) Exploiter-B-O, (n) Exploiter-D-O, (o) M-LC strategy combinations. p-s, Sensitivity of yearly dividend payouts to varying oil prices seen in Match 5 for average (p) E-LC, (q) E-O, (r) Exploiter-B-O, (s) M-LC. See Appendix Figure B.2 9 for additional unique matches.
Furthermore, we investigate the dividend dynamics of the main IOCs, particularly in the early game. E-LC and E-O IOCs noticeably cut dividend payouts within the first four years of play, regardless of the selected match up. M-LC IOCs, on the other hand, delay dividend cuts as the companies focus on business-as-usual strategies before reallocating capital towards low-carbon in later years. The early dividend cuts by the early low-carbon movers highlight the need too quickly reallocate capital towards building low-carbon economies that provide meaningful mid- and long-term returns. In the case of E-LC IOCs, cutting dividends over consecutive years in the early-game to elevate long-term returns risks displeasing investors in the near-term.

IOCs yielding the most winning strategies - that is, leading to highest dividend payouts across each of their 21 unique match ups - continue considerable low-carbon asset acquisitions beyond their early movement. We investigate this behavior as well as hydrocarbon tactics by observing the evolution of each IAMC/IIASA scenarios’ and the IEA’s Net-Zero scenario’s most winning IOCs’ asset portfolios (Figure 5).

In all cases, winning IOCs transition into low-carbon with endgame low-carbon asset holdings accounting for 80.52% (50.36 ± 96.08%) of its asset portfolio value, on average (minimum ± maximum). While E-O IOCs allocate most capital towards developing oil business models, the value of their endgame reserves never exceeds the value of their low-carbon assets. This is due to the IOC’s lower cost of low-carbon asset acquisition through optimal bidding as well as timely hydrocarbon reserve development and depletion. Focusing on the latter, we find that IOCs exhaust their initial hydrocarbon reserves - to 16.80% (3.63 ± 47.68%) and 2.68% (0.20 ± 15.64%) of its asset portfolio, on average (minimum ± maximum), for oil and gas, respectively - rather than replenishing them. The substantial decrease from an average 66% of initial oil reserves value to 9.21% supports earlier findings on the necessity of oil production to enable a robust low-carbon transition (Figure 2h).

Examining behavior in hydrocarbon markets, the E-LC IOCs boast the greatest performances by swiftly producing hydrocarbon assets. Oil reserve portfolio representation remains above 50% until 2023 (2021 – 2032), on average (minimum ± maximum), while gas reserve representation remains above 20% until 2024 (2023 – 2031), on average (minimum ± maximum). These rapid market exits are prioritized to enable larger acquisitions of low-carbon assets in the early game. Despite efforts to alleviate hydrocarbon dependency, IOCs risk stranding the carbon-intensive assets in low degree-warming scenarios (Figure 4a-c). In these climate-compatible scenarios, 47.5% of all most winning IOCs maintain endgame oil reserves that represent above 20% of their asset portfolio. The resulting forgone revenue, due to the scenarios’ lower demand and prices, limits an IOC’s ability to scale its low-carbon economies. This curb to further transition-oriented acquisitions is evidenced by the dissimilarities in endgame low-carbon asset portfolio representations amongst the most winning IOCs in lower degree-warming scenarios.

Similar strategies in diverging demand scenarios.

To examine how strategies develop more closely, we compare the most winning strategies for two diverging degree-warming, hydrocarbon demand, scenarios - IEA’s Net-Zero (1.5°C and below) and WITCH-GLOBIOM’s SSP5-Baseline (Above 4°C) (Figure 6). Despite the two scenario’s asymmetric oil and gas demands, each winning IOC, on average, converges towards a similar high-level robust strategy: scaling low-carbon business models predicated on early (2020 for both) low-carbon mover behavior (Figure 6a,b). However, we find that these strategies differ with respect to four characteristics: primary financial instruments deployed to scale low-carbon economies in the late-game; year in which low-carbon cashflow becomes the main source of income; endgame hydrocarbon reserves; and dividend policies.

Observing how energy assets are acquired, or developed, we
find that the most winning IOCs of the Net-Zero and SSP5-Baseline scenarios yield nearly identical capital allocation distributions (Figure 6c,d). Both IOCs boast considerable unlevered hydrocarbon development strategies yet avoid exploration activities in efforts to rapidly deplete reserves. The focus on development, therefore increasing production, allows for both IOCs to rely primarily on unlevered cashflows to finance their low-carbon acquisitions. These similar development and acquisition strategies accelerate the Net-Zero and SSP5-Baseline IOC’s low-carbon transition, occurring in 2026 and 2025, respectively. In the last decade of play, while both IOCs decrease hydrocarbon development activities, low-carbon acquisition tactics begin to diverge as the Net-Zero IOC becomes increasingly reliant on raising debt to bolster its low-carbon business model. These late-game leveraged acquisitions are a result of the Net-Zero scenario’s hydrocarbon demand shocks, thus diminishing respective revenues. Despite the different financial instruments deployed, IOCs continue to focus capital expenditures on low-carbon assets regardless of hydrocarbon scenario.

The benefits of early low-carbon movement, particularly in 2020 where low-carbon acquisitions accounted for at least 90% of capital expenditures, are seen as early as 2023 for both IOCs (Figure 6c-h). The continued acquisition of these assets result in low-carbon returns becoming the primary source of income for the Net-Zero and SSP5-Baseline IOCs by 2028 and 2039, respectively. Though the latter exhibits a quicker low-carbon transition, as noted previously, the SSP5-Baseline IOC continues to focus its income dependency on hydrocarbon returns, namely heavy oil, due to the high oil demand (Figure 6h). In addition to increased cashflow, the continued production of heavier oil assets allows the SSP5-Baseline IOC to deplete its oil reserves and minimize risks of asset stranding (Figure 6b). The Net-Zero IOC, despite its rapid exit efforts, however, bears these stranded costs as it is unable to reduce its oil reserves which represent 21% of its endgame portfolio value.

To effectively build low-carbon economies, both IOCs cut dividend payouts in the early-game and reallocate capital towards low-carbon acquisitions (Figure 6e,f). While the SSP5-Baseline IOC increases its dividends policy to allocate 50% of its unlevered cashflow by 2028, the Net-Zero IOC continues to keep dividend payouts below this level for an additional eight years. The climate-compatible IOC follows this delayed dividends policy due to its anticipation of lower hydrocarbon returns the following years as a result of decreased demand. Focusing unlevered cashflows on scaling low-carbon business models enables the IOC to benefit from stable, long-term returns. Moreover, it decreases risks of a spiraling cost of capital which allows the IOC to raise sufficient debt to acquire low-carbon assets in the late-game without adversely affecting its bottom-line (Figure 6c,e).

Recently, researchers (78) estimated the required annual low-carbon energy investment to achieve an energy transition that would remain consistent with the goals of Nationally Determined Contributions (NDCs) as well as of 2°C and 1.5°C degree-warming scenarios with respect to several IAM frameworks. We use these estimates as a proxy for 2DP’s Sustainable and Low-Carbon Energy Investment Available metric, available as purchasable assets in the sustainable and low-carbon energy auction mechanism, under the respective model and degree-warming scenario to observe the potential impact our IOC’s early low-carbon mover strategies have on the global energy transition (Appendix Table A.2).

Across all scenarios, we find the average E-LC, M-LC, and E-O IOC have the potential of playing an integral role in enabling
a climate-compatible energy transition (Figure 7). This is particularly true in the 2020s decade when global transition capital requirements begin to gain momentum and fewer potential demand shocks are present - granting IOCs stable hydrocarbon cash flows as well as credit access. On average, an E-LC IOC could provide 1.87% or 2.48% of required low-carbon supply-side investment to help achieve a 1.5°C or 2°C degree-warming scenario, respectively (Figure 6a). A M-LC IOC contributes slightly less capital at 1.72% or 2.26% for a 1.5°C or 2°C degree-warming scenario, respectively (Figure 6b). An E-O IOC allocates the least, albeit still significant, amount of capital at 1.65% or 2.10% for a 1.5°C or 2°C degree-warming scenario, respectively (Figure 6c). These figures decrease in the following decades yet remain non-negligible at 0.55% or 0.85% for the E-LC IOC, 0.49% or 0.84% for the M-LC IOC, and 0.48% or 0.65% for the E-O IOC in the 1.5°C or 2°C degree-warming scenarios.

Instrumental to these yearly low-carbon expenditures is the access to credit markets as highlighted previously (Figure 2d, Figure 6c,d). Within the 2020s decade of the 1.5°C degree-warming scenarios, levered low-carbon acquisitions account for 29.2%, 33.4%, and 22.4% of an E-LC, M-LC, and E-O IOC’s transition-oriented investments, on average, respectively. As the temperature rises, the reliance on credit markets decreases to a minimum of 24.9%, 27.9%, and 19.4% of an E-LC, M-LC, and E-O IOC, on average, respectively. In the following two decades, reliance on credit markets to finance low-carbon asset acquisition increases as IOCs are, then, able to raise significant amounts of debt without adversely affecting their bottom-line, as previously noted. With early-game dividend cuts and continued dependency on credit markets, our findings point to the potential roles lenders and investors play in mobilizing finance to scale the low-carbon transition and ensuring capital is responsibly allocated towards these endeavors.

**Discussion**

This paper finds that IOC strategies robust to market uncertainty and adversarial, including exploiter, strategies emerge in the form of low-carbon business models as a result of early transition-oriented movement. Our model alleviates concerns regarding global solution convergence guarantees by applying state-of-the-art deep multi-agent reinforcement learning algorithms and discovering no main IOC discovers a robust business-as-usual (BAU) or delayed transition strategy across all games evaluated (see Methods). Our results suggests that IOCs responsibly allocating capital towards low-carbon business models could benefit from, and accelerate, the energy transition to emerge as transition leaders.

Observing our IOCs’ emergent, robust behavior, we discover that 90.94% of strategies were predicated on scaling low-carbon business models, each resulting from movement towards low-carbon within the first five years of play (Figure 2). In addition to out-performing exploitative strategies across all energy futures scenarios, transition-focused IOCs intent on early movement towards low-carbon business models and rapid exits from hydrocarbon markets mitigate the downside risks of oil price shocks (Figures 3, 4, and 5). Comparing best-performing strategies of two diverging hydrocarbon demand scenarios, including a Net-Zero emissions scenario, we found each winning agent yields a similar high-level robust strategy: scaling low-carbon business models predicated on early movement towards low-carbon (Figure 6). However, notable differences with respect to future credit market dependency, hydrocarbon production, and investor payout policies are present (Figure 6c-f). We examine the roles these transition-oriented IOCs, as well as their lenders and investors, play in fulfilling low-carbon investment needs to achieve climate goals. With effective engagement from their lenders and investors, IOCs responsibly reallocating capital towards low-carbon have the potential to emerge as global low-carbon energy leaders, benefiting greater society.
As of late 2018, oil and gas Majors spent 1.4% of their capital expenditures towards sustainable energy and low-carbon technologies in the previous decade (79), on average. Although industry executives are planning on committing more capital to less carbon-intensive business models after COVID-19 (48), cost-cutting remains at the forefront of the IOC’s agendas and they are more willing to maintain, or increase, capital expenditures in the year ahead as compared to after the 2014 price crash. We find that it is of utmost importance to allocate such capital to scale low-carbon business models, as opposed to oil and gas exploration, activities to best maximize value and mitigate transition risks, namely asset stranding and low hydrocarbon prices reducing dividend allocations. Focusing on building low-carbon economies, with the sufficient leverage, situates IOCs in more robust financial positions than those who delay low-carbon movement as well as those who continue BAU practices in efforts to exploit possible future gaps in hydrocarbon markets. Our results show that the emergent transition-focused strategies outperform hydrocarbon-centric strategies as well as effectively balance returns to investors and ease debt pressures throughout a range of oil prices.

While climate pundits point to IOCs’ exposure to transition risks and incompatibility with climate-aligned scenarios (44), industry attempts to justify their hydrocarbon-focused strategies with increased demand projections. Our results show that, regardless of demand scenario, immediately reallocating capital from hydrocarbon reserve replenishment (exploration) and into low-carbon business models boasts a more favorable strategy in satisfying investors and mitigating debt engulfment. We do not expect cuts to hydrocarbon production, however, capitalizing on returns from current reserves would raise low-carbon spending as well as reduce the amount of assets susceptible to stranding. Rather, we find hydrocarbon production becomes gradually less relevant to a company’s cashflow. This pace is typically hastened in demand shock scenarios. The financial risks of pursuing such transition-oriented strategies, however, could create high financial risks, namely escalating cost of capital, as efforts to achieve climate-compatible economies (e.g. lower demand for hydrocarbons) accelerate. Our results demonstrate that immediately reallocating unlevered cashflows from hydrocarbon production mitigates such risks, even as raising debt to finance future low-carbon acquisitions becomes necessary. Thus, it is imperative for IOCs to adopt strategies centered on early movement towards a low carbon transition before the financial implications of transition risks are felt and magnified.

Lastly, our findings point towards the mutually beneficial outcomes of low-carbon movement shared between the transition-oriented IOC and the global energy transition. Despite reaching a new high in 2020 with over $500B (80), current global low-carbon energy investment trends still fall short in attaining a climate-compatible energy transition (78). IOCs adopting low-carbon strategies could help fill these investment gaps, especially in the current decade where transition-focused finance needs to take center stage (81). We note this impact, however, ultimately relies on the abilities of lenders to mobilize transition-oriented finance and investors engaging with IOCs to responsibly reallocate capital towards climate-compatible business models.

Conclusion

Our model reveals that IOCs have the potential to overcome uncertainty by emerging as low carbon energy leaders. To achieve this and mitigate downside risks, lenders and investors should effectively engage with IOCs to ensure responsible reallocation of capital that would enable timely shifts towards robust, low-carbon business models. Doing so will require IOCs to increase disclosure efforts regarding their hydrocarbon and low-carbon activities as well as understand their potential in enabling a global energy transition. While it is important to note that our model does not directly restrict hydrocarbon production through emissions targets, our findings reinforce the low-carbon consensus that IOCs must transition to maximize value and mitigate downside risks regardless of degree-warming scenario. To this end, we argue IOCs can, and should, emerge as low-carbon leaders to benefit its stakeholders as well as greater society.

This work provides a data-driven analysis to support literature’s low-carbon consensus that IOCs can and should transition. Our model complements energy futures research and applied game theory in many ways. From a broad view, the integration of deep multi-agent reinforcement learning addresses non-linearities inherent to IAM scenarios as well as opens opportunities for new ways of assessing uncertainty in energy scenario analysis. Specifically, our model allows for decision-makers to stress test core governance questions as well as explore the emergent behavior of intelligent agents exposed to a range of market designs. Oil and gas literature could bolster their qualitative findings by adding a game-theoretic evaluation to quantify risks and rewards of specific strategies with respect to a chosen energy future(s). The multi-agent league system could incorporate other competing entities to further assess robustness of an IOC’s optimized strategy.

Methods

2 Degree Pathways Wargame. The 2 Degree Pathways (2DP) wargame is a decision support tool developed by the Oxford Sustainable Finance Programme and E3G, a think tank, to “help inform company, investor, government and civil society thinking around the pathways the oil and gas majors can take to become 1.5°C/2°C-compatible by simulating oil companies in competition within varying transition scenarios (61). The tool was designed based on a range of wargaming literature and ex-oil and gas executive input. Two versions of 2DP were developed, one with IOCs and the other with the addition of NOCs. This work builds on the former. Wargame participants role play as fictitious IOCs, bearing resemblance to real IOCs, with the goal of maximizing shareholder value by responsibly allocating capital on a year-by-year basis. Players participate in eleven markets in 2DP: two exogenous oil and gas markets, with demand driven by the scenario at-play, and nine endogenous markets comprised only of game players, one for each for balance sheet asset. Game actions are continuous giving the game a high-dimensional state and action space. Details on this work’s 2DP variation, such as player setup and game stages, are further detailed below.

Player Setup. Players maintain eleven on-hand asset classes and sixteen pipeline assets, seven decision-making metrics and the ability to choose from 64 actions (Appendix Tables A.8, A.9, A.4, and 5, respectively). All players yield the same oil and gas capital costs (Appendix Table A.7). While equal in total cash value, players begin with differing asset distributions to represent how an initial market dominance, or diversity, may affect strength of overall strategies (Appendix Table A.8). These values were calculated based on the average respective asset holdings seen in six oil Majors’ annual reports (76) to best represent market conditions. A player’s level of assets accumulated and actions chosen throughout the game make up the respective player’s decision-making metrics for
a given year (Appendix Table A.9).

**Scenario Setup.** Global scenario metrics dictate the game environment’s dynamics, asset payoffs, and a player’s strategy. Upon game initiation, beginning in year 2020, these metrics are taken from the selected energy futures scenario, where each year of the scenario’s data represents the metrics for the respective in-game year. In this work, we acquire data to represent 408 energy futures scenarios from the Integrated Assessment Model Community (IAMC) and International Institute for Applied Systems Analysis (IIASA) $1.5^\circ C$ explorer database, IEA’s Net-Zero report (55), (author5) (78), OPEC historical trends (82), as well as BP (83) and Equinor (84) statements (Appendix Table A.9). Given multiple scenarios to test across, we incorporated a random energy future scenario generator to prevent agents from overfitting to a single scenario’s demand curve. Moreover, they are unable to play the same transition scenario twice without having played through each of the other scenarios at least once. At the end of each game, year 2050, the game resets and the generators select a new scenario for the following epoch.

**Game Stages.** 2DP’s model breaks one year of play into four trans-action stages: production, borrowing, trading, and allocation. Each stage is explored further below and their respective player actions found in Appendix Table A.4, and 5.

**Production** The production stage calculates the player’s net income for a given year. Net income include cash gained from oil production, gas production, low-carbon assets, and cash j used to pay off debt interest, if any. A key feature to this stage is oil price formation predicated on the current scenario’s oil and gas demand, OPEC & other’s production share global metric, and the sum of produced oil and gas assets players choose to produce in a given year. Therefore, players must appropriately produce oil and gas assets that provide sufficient returns each year as overproducing (underproducing) may lead to glut (missed returns). Returns from low-carbon assets are predicated on the current year’s sustainable energy and low-carbon asset return to investment.

**Borrowing** The borrowing stage allows a player to borrow acquire an amount of credit dependent on the health of their balance sheet. Players are only able to borrow additional cash if their current debt-equity ratio is below 200%. Player’s chosen borrowed amount is added to their cash assets and debt liabilities.

**Trading** Trading is split into two sub-mechanisms: one sustainable energy and low-carbon investment auction and one player-to-player trading platform. In the former, players simulate inorganic low-carbon growth by placing bids to purchase the respective assets in sealed-bid auction form. The low-carbon auction allows players to bid with cash and credit separately, designed to further explore transition finance behavior with respect to the two endogenous markets. While submitted independently, all cash and credit bids compete for the same amount of sustainable energy and low-carbon assets available of a given year. These auctioned assets are dictated by the scenario at hand (see Appendix Table A.2). Auction sales follow the order of the highest bid placed. A player with the highest bid maintains purchasing priority. Players with lower bids are at risk of being unable to purchase low-carbon assets due to the finite amount of assets held within the bank.

Post-auction, players have the choice to participate in trading amongst each other. All hydrocarbon and low-carbon on-hand assets are available to trade. Similar to the above, sale of assets follow the order of the highest bid placed for the respective asset. Multiple sales from the same player to others may occur if the volume of an asset up for sale is greater than the other players’ bidding volume for said asset. No trades occur if there is no buyer or no seller.

**Allocation** The allocation stage grants the ability to explore and/or develop oil and/or gas assets, pay off debt, pay dividends, and/or save cash into the next year. Similar to the low-carbon auction, allocation actions are split between credit-only and cash-only actions. The motivation for this is to restrict agents from borrowing significant amounts of cash only to pay as dividends, an unrealistic business model.

**Game Theory.** This work’s focus on revealing robust strategies to energy scenario uncertainty and adversarial entities is a game theoretic problem. Game theory is the study of strategic interactions between a set of agents, or players, in game form (85). Games are largely categorized by how its agents’ total losses and total gains are summed. We frame 2DP as a non-zero (general-) sum game - that is, a game in which strategic payoffs may sum to values greater than, or less than, zero - by centering each agent’s utility payoff on its allocated total dividends. Agents are said to be in Nash Equilibrium when strategy profiles converge to a point in which they cannot be improved upon by unilateral deviation (86). Of particular importance of this work is the concept of approximate mixed-strategy Nash equilibria whereby at least one player is playing a non-deterministic strategy and the conditions of a Nash equilibria are approximately satisfied.

Of particular importance of this work is the concept of approxi-mate mixed-strategy Nash equilibria whereby at least one player is playing a non-deterministic strategy and the conditions of a Nash equilibria are approximately satisfied.

Computing Nash equilibria of n-player, general-sum games using conventional game theoretic methods is impractical (87). While some problem formulation possibilities exits (e.g. approximating Nash using a sequence of linear complementarity problems approach or reverse the problem to optimize for a minimum) as well as applicable algorithms (e.g. n-player extensions of Scarf, Lewke-Howson, and the support-enumenation method), no method exists that could effectively solve for an infinitely-repeated game, such as 2DP. Moreover, modeling the 2DP wargame’s non-linear pay-offs as well as high dimensional state and action spaces proves analytically and computationally intractable.

**Deep Reinforcement Learning.** We use reinforcement learning, the computational approach to learning from environment inter-action (72), to have our IOC agents learn robust strategies under energy futures uncertainty. Reinforcement learning algorithms seek to solve problems defined as a Markov Decision Processes (MDPs) comprising of four key components: the environment and action space, the observation space, describing the game elements accessible to an agent, and reward function. Stochastic, or Markov, games extend MDPs to involve multiple agents whose actions impact their resulting rewards and next state. To best mimic real-life market competition, we frame the 2DP wargame continuous control problem as a partially observable stochastic game (POSG) which restricts an agent’s observation space to a limited set of information. POSGs rely on the same four general components used in MDPs. This work utilizes the aforementioned variant of 2DP and its player’s available actions as the game environment and action space. Agent state observations, following the POSG framework, are incomplete and imperfect (i.e. restricted to limited game and opponent information) to best mimic the partial observability of real-world market competition (Appendix Table A.3). The agent’s reward function, r, focuses on maximizing share-
The incorporation of the aforementioned reinforcement learning components make up each agent’s discounted value function $V_a(s)$ and policy function $\pi_a(a|s)$. The former is used to predict rewards in a given state $s$ typically discounted by some factor $\gamma$ in infinite horizon games, such as 2DP, while the latter represents the strategy an agent takes by playing action $a$ for a given state observation $s$. In the context of this work, we follow the actor-critic setting (88). Here, the agent’s actor seeks to converge at an optimal policy function $\pi_a(a|s)$ by finding parameters $\theta$ that maximize the performance measure $J(\theta)$, some loss objective function, of a policy through gradient descent. The agent’s critic, where learning takes place, computes a value function $V_a(s)$ to best predict $r$ by critiquing the actor’s policy function updates. These updates are applied as the agent generates new trajectories (i.e. training data) at every game time-step (i.e. year of 2DP play).

A challenging task for reinforcement learning is solving an environment of high-dimensional continuous state and action spaces, such as the 2DP wargame. The introduction of artificial neural networks as function approximators enables deep reinforcement learning algorithms to effectively address this challenge. In this work, we apply a combination of neural networks and actor-critic based reinforcement learning methods, detailed below, to improve agent performance.

**Algorithm.** We train agents with the advantage actor-critic (A2C) (89) and proximal policy optimization (PPO) algorithms (90), equipped with deep neural networks, due to their state-of-the-art performance on high-dimensional continuous control tasks (71, 90).

A2C follows the actor-critic method as described above but replaces the value function with an advantage function, measuring the improvement of taking an action $a_s$ over the average action $\pi_c$ in that state, to increase performance. Moreover, it enables the agent to engage in continuous state and action spaces by modeling $\pi_a(a|s)$ as a Gaussian distribution. In this work, we equip the actor and the critic with separate deep neural networks to serve as policy and value function approximators, respectively, allowing for the effective mapping of state-action pairs to expected rewards. A2C alone, however, is insufficient to solve a high-dimensional continuous control task due to its high sensitivity to hyperparameter tuning and susceptibility to training instability due to large policy updates.

Built on A2C, PPO responds to the issues evident in the actor-critic method, as well as in past policy-based learning methods, by modifying the policy update performed in the actor network with a more conservative policy update region via a clipped surrogate objective function $J_{\text{clip}}(\theta)$. With PPO, policy updates are performed by collecting a batch of trajectories (i.e. experiences from game play) and are optimized with mini-batch stochastic gradient descent to increase stability and convergence. Additionally, PPO addresses concerns regarding large number of samples and high variance present in high-dimensional continuous control problems by upgrading A2C’s advantage function with a generalized advantage estimator.

**Agent Training.** We use an independent learning (InRL) training method to decompose an $n$-agent multi-agent reinforcement learning (MARL) problem to a single learning agent problem whereby other (frozen) agents are treated as part of the localized environment. The application of InRL allows for the implementation of the 2DP league mechanism described below. Concerns of this InRL approach, however, arise with respect to its theoretical limitations that would result in learning instabilities, overfitting, and loss of convergence guarantees (91, 92). Despite these potential issues, the use of PPO, specifically the multi-agent variant Independent PPO, in InRL problems has been shown to match, and even outperform, state-of-the-art MARL algorithms in multi-agent settings (93).

**Multi-Agent Learning.** To address game-theoretic challenges and encourage more robust agent strategies during training, we introduce league training similar to the AlphaStar League (70). Central to our 2DP-League training setup is the incorporation of self-play and prioritized fictitious self-play (PFSP) variants. We iteratively populate the league with players that represent saved parameters from previously trained main and exploiter agents.

**Self-Play.** Self-play (74) is a training mechanism central to multi-agent learning. In self-play, a learning agent is trained by competing against itself. This mirrored battle provides sufficient amount of challenge such that an agent achieves superhuman-level performance and complex emergent behavior (68–71, 94). Our self-play algorithm updates all agents, learning and frozen, in the environment with the learning agent’s latest policy update every epoch so long as it is a winning policy (i.e. win-rate greater than or equal to 50%). While the same policy, an agent’s policy output may differ from its opponents’ policy outputs as we train agents playing mixed strategies to encourage convergence towards a mixed strategy Nash equilibrium. The motivation for this is the thinking that, for example, company $A$ should not, and cannot, expect company $B$ to play an expected, deterministic strategy. Training against deterministic policies would result in company $A$ overfitting its strategies towards a narrowed belief in its opponents’ strategies.

**Prioritized Fictitious Self-Play.** Overfitting is of concern with respect to multi-agent learning algorithms such as self-play (94). Fictitious self-play (FSP) 75 helps prevent this issue and the occurrence of strategy cycles by uniformly sampling from previous saved policies. Sampling from all previously saved policies, however, is computationally intensive and inefficient as many games would be played against extremely poor policies. To combat this inefficiency, AlphaStar proposed PFSP whereby a matchmaking system replaces the uniform sampling to grant a learning signal. In this work, we employ two variations of AlphaStar’s PFSP, PFSP-past and PFSP-opponent. We sample a frozen agent (PFSP-past), or group (PFSP-opponent), from frozen policy pool $C$ for the learning agent to play with the probability

$$\sum_{C \in C} e^{f_{\text{weight}}}$$

where $f_{\text{weight}}$ is some weighting function described by the PFSP mechanism employed.

In PFSP-past, we sample from past-selves of the learning agents initialized using the weighting function $f_{\text{past}}(x) = (1-x)^2$, where $x$ is the win-rate of the learning agent, to encourage play against the most difficult past-selves. We iterate through this sampling $N-1$ times, where $N$ represent the number of total players in a match. In the event the learning agent fails to learn (i.e. unable to beat the frozen agents for more than several consecutive epochs), we reset the matchmaking mechanism and use $f_{\text{past}}(x) = x(1-x)$ as the new weighting function to encourage play against frozen agents around the learning agent’s skill. Once learning has been sufficiently facilitated, we reset the matchmaking mechanism again to $f_{\text{hard}}$. 

Radovic et al.
When playing against opponent agents, we initiate the league with unique combinations of opposing player policies from the latest training iteration. We do not populate the league with all past opposing player policies as involving all such policies would be unattainable. Instead, the learning agent plays each combination once before allocating each $f_{\text{learn}}$ and prioritizing the most difficult combinations. Like in PFSP-past, if the learning agent is struggling, we reset the matching mechanism such that all opponents have an equal chance of being played, and back to $f_{\text{learn}}$ once a learning signal is created.

**League Participants.** We employ two distinct agents that differ in the evolving distribution of opponents they play against, when training parameters are reset, and the constraint on actions available. Main agents train as IOCs with no constraints on their available actions yet differ in their initiated asset levels. In the first training iteration, main agents are trained with probability of 80% self-play and 20% PFSP-past. The PFSP-opponent mechanism is introduced in subsequent iterations beginning at 35%, with 50% self-play and 15% PFSP-past, and scaling linearly by iteration to a maximum of 100%. Main agents’ parameters never reset. The trained policy at the end of each iteration is saved to later be used as the base parameter for the respective main agent’s next iteration, facilitating continuous learning. End of each iteration is saved to later be used as the base parameter for the respective main agent’s next iteration, facilitating continuous learning.

Exploiter agents are initiated with oil-dominant initial assets yet are constrained to a select group of actions based on their pre-determined strategy type. Exploiter agents are always trained with 100% PFSP-opponent against the main agents. The idea is to exploit weaknesses in the main agents’ strategies, thus making them robust to both energy scenario and opponent uncertainty. The exploiter agents’ trained parameters are added to the league after each training iteration. Their parameters are reset to encourage diversity in exploitation.

**Evaluation.** Agents were evaluated across all scenarios for all different combinations of opposing agents. We evaluated eight agents – six main and two exploiters - in a six-player game across 406 energy scenarios, totaling 11,421 unique games (each agent plays 8,568 games). Selected main agent policies were drawn from their respective agent’s latest saved policy. Exploiter agent policies were sampled from the league. Further more agents in league, as noted previously, in this way would require significant levels of compute.

---

1. P Stevens, International Oil Companies: The Death of the Old Business Model (2016).
2. I Markit, Energy Cost and Technology Indexes | IHS Markit (2020).
3. C Tracker, Sense and Sensitivity: Maximising Value with a 2D Portfolio, Technical report (2019).
4. R Fitz, M Abel, Winning Back Investors’ Trust (2019).
5. N Investor, Annual S&P Sector Performance • Novel Investor (2020).
6. K Hippie, C Williams-Derry, T Sarullo, IEFEA report: Oil majors live beyond their means – can’t pay for dividends, Backwards, Technical report (2020).
7. C Eaton, S McFarlane, 2020 Was One of the Worst-Ever Years for Oil Write-Downs - WSJ (year?).
8. J Hillier, Pandemic pushes Exxon to historic annual loss, $20 billion cut in share value. Reuters (2021).
9. OilGas, Global upstream investments set for 15-year low, falling to $383 billion in 2020 - Oil & Gas Middle East (2020) Section: DRILLING & PRODUCTION.
10. J Abraham, Seven top oil firms downgrade assets by $87bn in nine months. The Guard. (2020).
11. UNFCCC, The Paris Agreement | UNFCCC (2018).
12. K Larsen, J Larsen, PP Chaudhuri, JF Kirkegaard, L Wright, 2020 Green Stimulus Spending in the World’s Major Economies, Technical report (2021).
13. UNFCCC, Commitments to Net Zero Double in Less Than a Year | UNFCCC (2020).
14. TCFD, Task Force on Climate-related Financial Disclosures | TCFD - About (2016).
15. CA 100+, Climate Action 100+ – O2O PROGRESS REPORT, Technical report (2020).
16. N Landell-Mills, Are oil and gas companies overstretching their position? (2018).
17. C Tracker, Mind The Gap: the $1.6 trillion energy transition risk, Technical report (2018).
18. C Tracker, Wasted capital and Stranded Assets, Technical report (2013).
19. C McGlade, P Ekins, The geographical distribution of fossil fuels unused when limiting global warming to 2 °C. Nature 517, 187–190 (2015) Number: 7533 Publisher: Nature Publishing Group.
20. M Lewis, S Visisin, S Hahaz, S Mary, R Walker, Stranded assets, fossilised revenues, (Keppler Chevreux), Technical report (2014).
21. M Carney, Open letter on climate-related financial risks (2019).
22. P Spedding, K Mehta, N Robins, Oil & carbon revisited - Value at risk from unburnable reserves, (HSBC Global Research), Technical report (2021).
23. PRI, Preparing investors for the Inevitable Policy Response to climate change, Technical report (2020).
24. B Fatouh, R Poudineh, R West, Energy Transition, Uncertainty, and the Implications of Change in the Risk Preferences of Fossil Fuels Investors, Technical report (2019).
25. R West, B Fatouh, The Energy Transition and Oil Companies’ Hard Choices, (The Oxford Institute for Energy Studies), Technical report (2019).
26. M Boon, A Climate of Change? The Oil Industry and Decarbonization in Historical Perspective. Bus. Hist. Rev. 93, 101–125 (2019) Publisher: Cambridge University Press.
27. MJ Pickl, The renewable energy strategies of oil majors – From oil to energy? Energy Strateg. Rev. 26, 100370 (2019).
28. M Zhong, MD Bazilian, Contours of the energy transition: Investment by international oil and gas companies in renewable energy. The Electr. J. 31, 89–92 (2018).
29. P Stevens, The Geopolitical Implications of Future Oil Demand, (Chatham House), Technical report (2019).
30. RG Ecese, MP Kruzs, Implementing the Task Force on Climate-related Financial Disclosures Recommendations: An Assessment of Corporate Readiness. Schmalchenbach Bus. Rev. 71, 287–293 (2019).
31. IEA, The Oil and Gas Industry in Energy Transitions – Analysis (year?).
32. RJ Johnston, R Blakemore, R Bell, The role of oil and gas companies in the energy transition, (Public Council), Technical report (2019).
33. E Asmelash, R Gorini, International oil companies and the energy transition, (IRENA), Technical report (2021).
34. C Beck, J Kar, S Hall, D Clouston, D Bettone, How oil and gas is navigating the energy transition (McKinsey) (2021).
35. BP, BP sets ambition for net zero by 2050, fundamentally changing organisation to deliver | News and insights | Home, (BP), Technical report (year?).
36. TotalEnergies, Total adopts a new Climate Ambition to Get to Net Zero by 2050, (TotalEnergies), Technical report (2020).
37. Shell, Shell accelerates drive for net-zero emissions with customer-first strategy, (Shell), Technical report (2021).
38. UNI, Reducing greenhouse gas (GHG) emissions, (UNI), Technical report (2021).
39. C Tracker, Absolute Impact: Why oil majors’ climate ambitions fall short of Paris limits, (Carbon Tracker), Technical report (2021).
40. C Tracker, Absolute Impact: Why oil majors’ climate ambitions fall short of Paris limits, (Carbon Tracker), Technical report (2021).
41. E Shojaidtini, S Naimoli, S Ladiwala, M Bazilian, Oil and gas company strategies regarding the energy transition. Prog. Energy 1, 012019 (2019) Publisher: IOP Publishing.
42. L Fletcher, T Crocker, J Smyth, K Marcell, Beyond the cycle - Which oil and gas companies are ready for the low-carbon transition?, (CDP), Technical report (2018).
43. C Tracker, 2 degrees of separation – Transition risk for oil and gas in a low carbon world, (Carbon Tracker), Technical report (2017).
44. C Tracker, Fault Lines: How diverging oil and gas company strategies link to stranded asset risk, (Carbon Tracker), Technical report (2020).
45. C Tracker, Balancing the Budget: Why deflating the carbon bubble requires oil & gas companies to shrink, (Carbon Tracker), Technical report (2019).
46. C Tracker, Breaking the Habit - Why none of the large oil companies are “Paris-aligned”, and what they need to do to get there, (Carbon Tracker), Technical report (2018).
47. IEA, Oil and gas industry needs to step up climate efforts now - News, (IEA), Technical report (2020).
48. DNV, Turmoil and Transformation: The Insights for the Oil and Gas Industry in 2021 DNV, (DNV), Technical report (2021).
49. C Tracker, Unburnable Carbon: Are the World’s Financial Markets Carrying a Carbon Bubble?, (Carbon Tracker), Technical report (2019).
50. M Meinshausen et al., Gas emissions targets for limiting global warming to 2 °C. Nature 458, 1158–1162 (2009) Number: 7242 Publisher: Nature Publishing Group.
51. M Carney, Breaking the tragedy of the horizon - climate change and financial stability - speech by Mark Carney (2015).
52. D Huppmann et al., IAMC 1.5°C Scenario Explorer and Data hosted by IIAA (2019) Publisher: IOP Publishing.
53. M Boon, A Climate of Change? The Oil Industry and Decarbonization in Historical Perspective. Bus. Hist. Rev. 93, 101–125 (2019) Publisher: Cambridge University Press.
54. MJ Pickl, The renewable energy strategies of oil majors – From oil to energy? Energy Strateg. Rev. 26, 100370 (2019).
55. IEA, Net zero by 2050 – Analysis, (IEA), Technical report (2021).
56. S Patteod, Energy scenarios: the value and limits of scenario analysis. WIREs Energy Environ. 6, e242 (2017) _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wene.242.
57. M Workman, K Dooley, G Lomax, J Matlby, G Darch, Decision making in contexts of deep uncertainty - An alternative approach for long-term climate policy. Environ. Sci. & Policy 103, 77–84 (2020).
Table 1. Though 411 scenarios were assessed, only 409 scenarios provided oil and gas demand projects. Two of these scenarios boasted unrealistic, outlier oil and gas demand projects (starting at approximately 30% of current demand and falling).
Appendix A.2 - Global scenario metrics

| Field                                      | Value Range                        | Randomization Margin $^*$ | Source               |
|--------------------------------------------|------------------------------------|--------------------------|----------------------|
| Oil Demand (mbbl/d)                        | Dependent on IAMC/IASA scenario    | ±5%                      | 52                   |
| Gas Demand (bcf/yr)                        | Dependent on IAMC/IASA scenario    | ±5%                      | 52                   |
| OPEC & Others’ Production Share (%)        | [min, max]:                        |                          |                      |
|                                            | -1.72% for each player in a       |                          |                      |
|                                            | 2DP game$^*$                       |                          |                      |
|                                            | -2.5% 2.5%                         |                          | Majors’ cumulative   |
|                                            | 2.5% 5.0%                          |                          | production values in  |
|                                            |                                    |                          | 2019$^{52}$, 2010-2020 |
|                                            |                                    |                          | OPEC production      |
|                                            |                                    |                          | trends used to justify |
|                                            |                                    |                          | randomization margin$^{52}$ |
| Available Sustainable Energy and Low-Carbon Investment (SM) | Dependent on IAMC/IASA scenario related to McCollum et al. | ±5% | 78 |
| Sustainable Energy and Low-Carbon Return on Investment (%) | [min, max]:                        |                          |                      |
|                                            | 6.0% 8.0%                          |                          | BP$^{93}$, Equinor$^{94}$ |
|                                            | 0.0% 10.5%                         |                          |                      |
| Credit Access Limit (SM)                   | $15,000                            | -                        | 95                   |
| Debt-to-Equity Ratio Limit                 | 200%                               | -                        | 61                   |

Table 2. * Randomization margin represents the degree to which a value may deviate from its original value. This was introduced for several reasons: to prevent agents from overfitting to similar demand curves, prevent agents from learning to optimize with respect to the endgame (i.e. play as though it were a finite repeated game), encourage agents to discover further robust strategies with respect to any, reasonable scenario metrics. $^*$ The OPEC & Others’ Production Share played a considerable role in dictating our agents’ strategies. It is commonly noted (95, 96) that future oil prices are uncertain as trajectory is highly dependent on a given day’s geopolitical landscape - the OPEC influence and dilemma. Rather than suggest OPEC & Others’ production levels will follow a general trendline, we attributed the average production seen across the six oil Majors with respect to global production as the baseline production ratio for each agent. This way agents have equal hydrocarbon production ratios throughout the game. We introduce noise to this OPEC & Others’ Production Share metric that essentially increase, or decreases, an agent’s available hydrocarbon production as well as the global oil and gas prices. The idea is to make our agents’ strategies more robust with respect to an oil cartel and geopolitical uncertainty as seen in the real world.

Appendix A.3 - IOC game state observations

| Category                | Field                                      | Randomization Margin $^*$ |
|-------------------------|--------------------------------------------|--------------------------|
| Global Scenario Metrics | Oil Demand                                 | ±1%                      |
|                         | Gas Demand                                 | ±1%                      |
|                         | OPEC & Others’ Share Production            | ±0.5%                    |
|                         | Available Sustainable Energy and Low-Carbon Investment | ±2% |
|                         | Sustainable Energy and Low-Carbon Return on Investment | ±0.5% |
|                         | Credit Access Limit                        | -                        |
|                         | Debt-to-Equity Limit                       | -                        |
|                         | Balance Sheet Assets                       | -                        |
|                         | Pipeline Assets                            | -                        |
|                         | Decision-Making Metrics                    | -                        |
| Opposing Agents’ Data  | Balance Sheet Assets                       | -                        |

Table 3. * Further randomization was introduced into agent observations directly with the purpose of preventing agents from solving for the endgame, and instead play a game with an infinite horizon. The randomization here holds the same for each agent and, unlike the randomization introduced in Appendix Table A.2, does not affect the actual metric (e.g. oil demand) it is imposed upon.
### Appendix A.4 - IOC action space categorized by the game stage in which they are applicable

| Game Stage | Action | Value Range [low, high] | Description | Agent Types Available To |
|------------|--------|-------------------------|-------------|--------------------------|
| Production | Produce Oil – Low | [0.0, 1.0] | Portion of developed hydrocarbon balance sheet items looking to produce each year | All |
|           | Produce Oil – Medium | [0.0, 1.0] | | |
|           | Produce Oil – High | [0.0, 1.0] | | |
|           | Produce Gas | [0.0, 1.0] | | |
| Borrowing | Borrow Cash | [0.0, 1.0] | Portion of cash looking to borrow with respect to an agent’s equity and available debt according to their debt-to-equity ratio | All |
| Trading   | Low-Carbon Asset Auction Bid Cash | [1.0, 1.5] Credit: [1.0, 1.5] | Multiple willing to pay per low-carbon asset looking to purchase with cash (or credit) | Main, Delayed Transition |
|           | Share of Capital to Low-Carbon Asset Auction Cash | [0.0, 1.0] Credit: [0.0, 1.0] | Portion of total cash (or total credit) dedicated towards the low-carbon asset auction | Main, Delayed Transition |
|           | Capital to Player-to-Player Trading Cash | [0.0, 1.0] Credit: [0.0, 1.0] | Portion of total capital dedicated towards player-to-player trading | All |
|           | Asset A * Bidding Volume Cash | [0.0, 1.0] Credit: [0.0, 1.0] | Portion of total opponent players’ asset A an agent looks to purchase | All |
|           | Asset A * Bid Cash | [0.5, 1.5] Credit: [0.5, 1.5] | Multiple willing to pay per asset A looking to purchase | All |
|           | Asset A * Selling Volume Cash | [0.0, 1.0] Credit: [0.0, 1.0] | Portion of own balance sheet asset A an agent looks to sell | All |
|           | Asset A * Sale Price Cash | [0.5, 1.5] Credit: [0.5, 1.5] | Sale price multiple per asset A looking to sell | All |

Table 4. IOC action space categorized by the game stage in which they are applicable. IOC action space categorized by the game stage in which they are applicable. * Asset A represents any hydrocarbon or low-carbon balance sheet item. In total, there are nine balance sheet items available to trade.

| Game Stage | Action | Value Range [low, high] | Description | Agent Types Available To |
|------------|--------|-------------------------|-------------|--------------------------|
| Allocation | Explore for Oil Cash | [0.0, 1.0] Credit: [0.0, 1.0] | Portion of cash (or credit) dedicated to exploring undeveloped oil assets; if chosen, agents will explore total oil assets with respect to the cash they put in and the capital costs of exploring for oil. As a result, 1/6 of the explored oil will be undeveloped low, 1/3 undeveloped medium, and 1/2 undeveloped high. | All |
|           | Explore for Gas Cash | [0.0, 1.0] Credit: [0.0, 1.0] | Portion of cash (or credit) dedicated to exploring undeveloped gas assets; if chosen, agents will explore total oil assets with respect to the cash they put in and the capital costs of exploring for gas. | All |
|           | Develop Oil – Low Cash | [0.0, 1.0] Credit: [0.0, 1.0] | Portion of cash (or credit) dedicated to developing undeveloped low oil balance sheet items. The amount of oil developed is calculated with respect to the cash put in and the capital costs of developing an undeveloped low oil asset. | All |
|           | Develop Oil – Medium Cash | [0.0, 1.0] Credit: [0.0, 1.0] | Portion of cash (or credit) dedicated to developing undeveloped low oil balance sheet items. The amount of oil developed is calculated with respect to the cash put in and the capital costs of developing an undeveloped low oil asset. | All |
|           | Develop Oil – High Cash | [0.0, 1.0] Credit: [0.0, 1.0] | Portion of cash (or credit) dedicated to developing undeveloped low oil balance sheet items. The amount of oil developed is calculated with respect to the cash put in and the capital costs of developing an undeveloped low oil asset. | All |
|           | Develop Gas Cash | [0.0, 1.0] Credit: [0.0, 1.0] | Portion of cash (or credit) dedicated to developing undeveloped low oil balance sheet items. The amount of oil developed is calculated with respect to the cash put in and the capital costs of developing an undeveloped low oil asset. | All |
|           | Pay Debt Obligations Cash | [0.0, 1.0] Credit: [0.0, 1.0] | Portion of cash (or credit) dedicated to paying off any debt obligations. Agents are allowed to “return” borrowed cash. We allowed this behavior due to inefficiencies seen during training. | All |
|           | Pay Dividends Cash | [0.0, 1.0] | Portion of cash dedicated to paying dividends to investors. | All |
**Appendix A.5 - Reward function**

| Signal | Condition          | Real-Valued Reward | Justification                                                                                                                                 |
|--------|--------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Negative | Negative Net Income |   -10              | Negative net income values within the game indicate an agent has taken on too much debt given its current business models. Large quantities of leverage (accumulating debt) were often seen as a method of maximizing dividend payouts as early, and as much, as possible. This resulted in an agent becoming engulfed by debt before the endgame. To discourage such behavior, we introduced a negative reward given for agents that yield a negative net income for a given year. This forces agents to find realistic, robust business models. |
| Positive | Dividend Payouts   | \( \frac{\text{Dividend Payout}}{1e5} \) | Dividend payouts, scaled by a factor of 10,000 to optimize value function estimation and prevent unrealistic agent behavior, are counted as positive rewards if, and only if, an agent has a positive net income and pays out sufficient dividends. |

Table 6. IOC action space categorized by the game stage in which they are applicable (cont.).

**Appendix A.6 - Hydrocarbon capital costs**

| Exploration Costs ($/bbl or $/kcf)          | Source |
|--------------------------------------------|--------|
| Undeveloped Oil – Low                      | 5.0    |
| Undeveloped Oil – Medium                   | 12.5   |
| Undeveloped Oil – High                     | 20.0   |
| Undeveloped Gas                            | 1.375  |

| Development Costs ($/bbl or $/kcf)         |       |
|--------------------------------------------|-------|
| Undeveloped Oil – Low                      | 5.0   |
| Undeveloped Oil – Medium                   | 12.5  |
| Undeveloped Oil – High                     | 20.0  |
| Undeveloped Gas                            | 1.375 |

| Lifting Costs ($/bbl or $/kcf)             |       |
|--------------------------------------------|-------|
| Developed Oil – Low                        | 10.0  |
| Developed Oil – Medium                     | 25.0  |
| Developed Oil – High                       | 40.0  |
| Developed Gas                              | 2.75  |

Table 7
### Balance Sheet Distributions

| Asset Category | Asset                      | Average | Oil   | Gas   | Low-Carbon | Oil-Low-Carbon | Gas-Low-Carbon |
|----------------|----------------------------|---------|-------|-------|------------|----------------|----------------|
| Cash (SM)      |                            | 0       | 0     | 0     | 0          | 0              | 0              |
| Credit (SM)    |                            | 0       | 0     | 0     | 0          | 0              | 0              |
| Debt (SM)      |                            | 20000   | 20000 | 20000 | 20000      | 20000          | 20000          |
| Sustainable Energy and Low-Carbon Assets (SM) | | 4000   | 0     | 0     | 8000       | 6000           | 6000           |
| Undeveloped Oil – Low (mbbl) | | 800 | 666.7 | 800 | 933.3 | 800 | 933.3 |
| Undeveloped Oil – Medium (mbbl) | | 2200 | 2146.7 | 2200 | 2253.3 | 2200 | 2253.3 |
| Undeveloped Oil – High (mbbl) | | 3600 | 3566.7 | 3600 | 3633.3 | 3600 | 3633.3 |
| Developed Oil – Low (mbbl) | | 200 | 266.7 | 200 | 133.3 | 200 | 133.3 |
| Developed Oil – Medium (mbbl) | | 200 | 226.7 | 200 | 173.3 | 200 | 173.3 |
| Developed Oil – High (mbbl) | | 200 | 216.7 | 200 | 183.3 | 200 | 183.3 |
| Undeveloped Gas (bcf) | | 13600 | 13600 | 15054.5 | 13600 | 12872.7 | 14327.3 |
| Developed Gas (bcf) | | 3100 | 3100 | 3827.3 | 3100 | 2736.4 | 3463.6 |
| Undeveloped Oil – Low (mbbl) | | 0 | 0 | 0 | 0 | 0 | 0 |
| Undeveloped Oil – Medium (mbbl) | | 0 | 0 | 0 | 0 | 0 | 0 |
| Undeveloped Oil – High (mbbl) | | 0 | 0 | 0 | 0 | 0 | 0 |
| Developed Oil – Low (mbbl) | | 200 | 266.7 | 200 | 133.3 | 200 | 133.3 |
| Developed Oil – Medium (mbbl) | | 200 | 226.7 | 200 | 173.3 | 200 | 173.3 |
| Developed Oil – High (mbbl) | | 200 | 216.7 | 200 | 183.3 | 200 | 183.3 |
| Undeveloped Gas (bcf) | | 0 | 0 | 0 | 0 | 0 | 0 |
| Developed Gas (bcf) | | 9450 | 9450 | 9450 | 9450 | 9450 | 9450 |
| **Total Asset Value** (SM) | | 290,700 | 290,700 | 290,700 | 290,700 | 290,700 | 290,700 |

Table 8. IOC action space categorized by the game stage in which they are applicable (cont.).
### Appendix A.8 - IOC decision-making metrics

| Decision-Making Metric               | Equation                                      | Source |
|-------------------------------------|-----------------------------------------------|--------|
| Reserves-to-Production Rate         | \( \frac{\text{Reserves}_{\text{oil}}}{\text{Production}_{\text{oil}}} \) | 61     |
| Gas-to-Oil Reserves Ratio (R/G Ratio)| \( \frac{\text{Reserves}_{\text{oil}}}{\text{Reserves}_{\text{gas}}} \) | 61     |
| Equity\( ^\text{*} \)              | \((\text{Sum of Asset Valuations}) - \text{Debt}\) | Own analysis\( ^\text{a1} \) |
| Return on Assets (ROA)              | \( \frac{\text{Net Income}}{\text{Balance Sheet Assets}} \) | Own analysis\( ^\text{a6} \) |
| Return on Equity (ROE)              | \( \frac{\text{Net Income}}{\text{Equity}} \) | 61     |
| Debt-to-Equity Ratio (D/E Ratio)    | \( \frac{\text{Debt}}{\text{Equity}} \) | 61     |
| Cost of Capital                     | \( \begin{cases} 
20\%, D/E \text{ Ratio} = 2.0 \\
2.5\%, D/E \text{ Ratio} = 0.0 
\end{cases} \) | 61     |
| Weighted Average Cost of Capital (WACC)\( ^\text{1} \) | \( \frac{\text{Debt}}{\text{Debt} + \text{Equity}} \times \text{Cost of Capital} \times (1 - \text{Tax Rate}) + \frac{\text{Equity}}{\text{Debt} + \text{Equity}} \times \text{Cost of Equity} \) | Own Analysis |
| Market Capitalization               | \( \frac{\text{Unlevered Dividends Paid}}{\text{WACC}} \) | 61     |

Table 9. \( ^\text{*} \) The costs of oil and gas assets include costs associated with the total amount of developed and undeveloped reserves for an agent in a given year. These costs are only added to an agent’s Equity if the oil price remains above the asset’s lifting costs (e.g. developed high asset costs are included if that year’s oil price is at least \$40/bbl); the costs of low-carbon assets are simply the number of low-carbon assets it maintains in respective year, not the cost at which it paid for them via the low-carbon auction or through the player-to-player trading desk.

\( ^\text{1} \) A global, average corporate tax rate (24%) is used; cost of equity calculated using CAPM method.
Appendix B.1 - Additional selected matches to Figure 3

Fig. 8. Average dividend payouts of initial low-carbon movement (Figure 2d) and business model (Figure 2g) strategy combinations within selected matches of similar or differing strategy combinations across degree-warming scenarios. a-e. Average dividend payouts as compared to total dividend payouts seen in Match 6 for the average (a) E-LC, (b) E-O, (c) Exploiter-B-O, (d) Exploiter-D-O, (e) M-LC strategy combinations. f-g. Average dividend payouts as compared to total dividend payouts seen in Match 7 for the average (f) E-LC, (g) Exploiter-B-O strategy combinations. h-i. Average dividend payouts as compared to total dividend payouts seen in Match 8 for the average (h) E-LC, (i) Exploiter-D-O strategy combinations.
Fig. 9. The sensitivity of yearly dividend payouts to varying oil prices for strategy combinations within five unique matches. a-e, Sensitivity of yearly dividend payouts to varying oil prices seen in Match 6 for average (a) E-LC, (b) E-O, (c) Exploiter-B-O, (d) Exploiter-D-O, (e) M-LC strategy combinations. f-g, Sensitivity of yearly dividend payouts to varying oil prices seen in Match 7 for average (f) E-LC, (g) Exploiter-B-O strategy combinations. h-k, Sensitivity of yearly dividend payouts to varying oil prices seen in Match 8 for average (h) ELC, (i) Exploiter-D-O strategy combinations.