Improving Tactical Decision-Making through Multi-Objective Contrastive Explanations

Michelle Blom*, Ronal Singh‡, Tim Miller§, Liz Sonenberg*,
Kerry Trentelman†, Adam Saulwick†, Steven Wark†
*The University of Melbourne, Melbourne, Australia
{michelle.blom, l.sonenberg}@unimelb.edu.au
†Defence Science and Technology Group, Adelaide, Australia
{kerry.trentelman, adam.saulwick, steven.wark}@defence.gov.au
‡CSIRO's Data61, Melbourne, Australia
ronal.singh@data61.csiro.au
§The University of Queensland, Brisbane, Australia
timothy.miller@uq.edu.au

***Abstract*—We consider the effectiveness of multi-objective counterfactual explanations (MOCE) in helping individuals learn *tactics*, or rules of thumb, to apply when required to select a course of action in a specific context. In this setting, a counterfactual explanation compares one course of action against another. A MOCE presents this comparison by highlighting how the two options differ across a range of objectives or metrics. We conduct a study in which participants are presented with various scenarios alongside courses of action that could be implemented in those scenarios. Counterfactual explanations, including those involving multiple objectives, are used to identify the positive and negative aspects of the provided options. Participants were then required to identify the best course of action in various contexts. Participants trained with MOCE outperformed those given no explanations in seven of eight scenarios and those given single-objective explanations (SOCE) in four. SOCEs gave participants an aggregated outcome (expected rewards) without breaking these into specific objectives. MOCE improved tactic learning, but participants provided with SOCE or no explanation performed better in multi-tactic scenarios. These findings suggest that MOCE enhances tactical decision-making, but further research is needed for multi-tactic integration.**

***Index Terms*—Explainable agency, Multi-objective contrastive explanation, Tactical decision-making**

I. Introduction

Explainable AI (XAI) has emerged from a growing demand for transparency in the behaviour of automated or AI tools that make decisions. XAI uses *explanation* to convey the underlying reasoning or inner workings of such tools to a user. While the focus has largely been on explaining machine learning models, there is a growing body of research on explaining sequential decision-making tasks, such as explainable reinforcement learning [1, 2] and planning [3, 4].

In sequential decision-making, Multi-Objective Counterfactual Explanation (MOCE) is a contrastive explanation that highlights the trade-offs and underlying factors influencing decisions in multi-objective contexts, helping users understand how different objectives affect outcomes [5–9]. In contrast, single-objective counterfactual explanations (SOCE) provide explanations using an aggregate of the objectives (such as the expected rewards) without breaking them down into specific objectives. SOCE allows us to investigate the value of exposing the objectives' trade-offs instead of simply informing the user about the expected rewards when helping users learn tactics. However, research shows that learning rules/tactics is difficult if people are only exposed to outcomes [10, 11].

Prior studies use MOCE for improving transparency, showing how agents balance trade-offs [5], reasons behind their actions [6], objective function [12], and priorities [13]. Users could also make informed adjustments to achieve desired behaviours [7, 14]. These works show that MOCE enhances users' understanding of how agents trade off objectives.

We claim that MOCE helps users acquire and apply tactics by breaking aggregate scores into multiple objectives. We acknowledge that 'learning' here refers to the immediate uptake of tactics, not their formal philosophical definitions. A tactic is a heuristic for choosing immediate actions aligned with long-term goals. A tactical decision in gaming optimises immediate actions to maximise end-game rewards [15]. In a military setting, it involves organising forces to win engagements [16]. Broadly, we view a tactic as a problem-solving heuristic [17].

Presenting users with information identifying the trade-offs between objectives may help people recognise or discover tactics that support the realisation of long-term goals or strategies. When users grasp these trade-offs, they can identify which objectives are most critical in specific contexts, allowing them to prioritise their actions accordingly. Even with a few key examples, the detailed feedback MOCE provides can enable better tactical understanding, helping users make informed tactical decisions and modify their short-term actions to manage competing objectives. This is useful to allow people to learn tactics discovered by intelligent agents, without such tactics having to be explicitly extracted and formalised a priori.

In contrast, SOCE, which presents only an aggregate score, may lead participants to focus on optimising short-term results, potentially overlooking the trade-offs between objectives and long-term outcomes. We extend the literature by empirically testing MOCE's effectiveness in improving tactical decision-making, a gap existing studies have not explicitly addressed.

We investigate the effectiveness of MOCE over no explanations and SOCE in helping users 1) identify the best course of action to pursue in a given situation and 2) learn tactics or rules of thumb for identifying the best course of action. We evaluate the effectiveness of MOCE using a multi-turn resource-management game. This game was chosen because it is a sequential decision-making task with clear trade-offs between short-term gains and long-term goals, allowing us to test the presence of different tactics. In our setup, a scenario comprises information provided to participants at a specific turn (a discrete decision-making round) within the seven-turn game (a sequence of rounds), such as the arrangement of units and contrast behaviours used to guide decision-making. In this setting, our research questions are:

- **R1** *Do MOCE improve a player's ability to identify an effective course of action in a given scenario (vs SOCE)?*
- **R2** *Do MOCE improve a player's ability to identify desired tactics (vs SOCE)?*

Our hypotheses are:

- **H1** Participants who receive MOCE will be more effective in identifying the best course of action in a given scenario than those who receive no explanation or SOCE.
- **H2** Participants who receive MOCE will be more effective in identifying a desired set of tactics than those who receive no explanations or SOCE.
- **H3** Participants who receive MOCE will be more confident in their decisions and will rate their satisfaction higher on Hoffman's satisfaction scale [18].

Our study links tactics learning to the experimental manipulation of explanation types and decision-making effectiveness. By revealing trade-offs between objectives, MOCE is hypothesised to improve action identification (R1) and tactic formation (R2) compared to SOCE.

II. Related Work

The use of XAI for decision support has gained much interest recently [19–23], although it dates back to the days of expert systems [24]. This resurgence is in response to issues such as warranted and unwarranted distrust [25], meaning that AI systems are often deployed and largely ignored [26].

XAI is one mechanism to mitigate uncalibrated trust and reliance [25]. Decision makers can aim to understand why an algorithm reached a particular recommendation, allowing them to determine whether it should be relied upon. Recent research has investigated XAI for a range of tasks, such as document matching [27], financial advice [28], question answering [29], medical diagnosis and treatment planning [30–32], and even mushroom picking [22]. Empirical studies show that explainability tools can positively impact decision-making [20, 22, 23, 33], however, in many cases, it has minimal effect [34–38]. Studies like the one in this paper are required to help the decision-making and XAI community understand when and why explainability is useful for decision making.

Explainability can largely be broken into four main techniques: (1) feature-based explanation, where the importance of each feature is presented [39]; (2) prototypes, which show an example that is most similar to the inputs being explained, and its ground truth [40]; (3) contrastive explanations, which show an example similar to the input being explained, but which has some features changed, for example, such that the model output becomes a different class [41, 42]; and (4) rules, which summarise the behaviour of the model [20]. In this paper, we use a combination of contrastive and feature-based explanations: we treat objectives similar to the way features are used, and then provide a counterfactual contrastive explanation [41] for a model decision.

While most of the work on XAI has focused on explaining machine learning tasks, such as classification and regression, there is also a body of work that looks to explain sequential decision-making tasks, including explainable reinforcement learning (RL) [1, 2, 43], explainable planning [3, 4], and more generally summarising agent behaviour [44].

However, there is limited work in the XAI literature explicitly comparing multi-objective and single-objective explanations for sequential decision-making, which is the focus of this paper. Existing research has explored the value of conveying underlying factors, trade-offs, and tipping points to explain the desirability of an option [8]. This information helps decision-makers identify the most robust option in a given situation. Prior studies on MOCE have focused on improving transparency by understanding trade-offs [5, 6], communicating objectives [12], understanding agent priorities [13], and guiding decision-makers [7] (see Table I).

Sukkerd et al. [5] investigate using MOCEs to explain the decisions of a Markov Decision Process (MDP) planning agent that communicates its trade-off rationale in terms of objectives. Participants were provided with preference profiles identifying which objectives the agents should promote, and then required to judge whether the agent was following a preference profile. To help the participants decide, the investigators provided two explanations: one with objectives only; and the other with contrastive objectives. Their study showed that MOCEs significantly improved the users' understanding and confidence in their understanding of the agent's trade-off rationale.

Juozapaitis et al. [6] use reward decomposition to expose the individual components of an RL agent's overall reward. These *reward types* are measurements related to the agent's performance and are used to explain its behaviour. Two types of explanations are developed to explain why the agent preferred one action over another in a given state. Reward difference explanations expose all positive or negative reasons for the preference, where a reason expresses an advantage or disadvantage with respect to a reward type. Minimal sufficient explanations expose sets of "critical" positive and negative reasons. No human behavioural study was conducted to evaluate the effect of these explanations on decision making.

Septon et al. [13] conduct a user study examining the use of reward decomposition in conjunction with policy summaries (video highlights of example states) to explain the behaviour of

TABLE I
DIFFERENT WAYS IN WHICH MOCE HAS BEEN USED

Usage	Description
Understand trade-offs	Sukkerd et al. [5]: Investigated MOCEs to explain decisions in a MDP planning agent. The study focuses on how the agent balances multiple objectives, enhancing users' understanding of the trade-offs involved in decision-making.
Understand Reasons for Actions	Juozapaitis et al. [6]: Utilised reward decomposition to break down the overall reward into its constituent components, helping users understand the reasons behind a reinforcement learning agent's actions.
Understand Decision-making Criteria	Huang et al. [12]: Used MOCEs to communicate a robot's objective function through informative examples, helping users grasp decision-making criteria, highlighting trade-offs, and showing how different objectives lead to varying behaviours.
Conveying Preferences	Septon et al. [13]: Combined reward decomposition with policy summaries to explain RL agent behaviour, demonstrating that MOCEs could convey the agent's preferences, aiding users in understanding the agent's priorities and decision rationale.
Guiding Decision-Makers	Misitano et al. [7]: By providing insights into how different preferences affected solutions, MOCEs guided decision-makers in making informed adjustments to achieve their goals, emphasising the implications of balancing multiple objectives.

an RL agent. The ability of a user to rank the priorities of the agent in various states, given a reward decomposition and/or a policy summary explanation, was tested. Participants were asked to rank the agent's priorities and rate their confidence level in their answers. The provision of policy summaries over the reward decomposition was found to have a 'limited contribution' to participant performance, noting that reward decomposition was 'highly effective in conveying agent preferences' in the chosen domains.

Anderson et al. [45] examine different visualisations of an agent's reward, including *reward bars*, and their effectiveness in helping users predict agent behaviour. Reward bars depict the contribution of different reward components to the agent's overall reward. While Anderson et al. [45] found that the most effective visualisation varied according to the agent's situation and behaviour, participants who received reward decompositions 'had the most insight into nuanced concepts'.

Our study goes beyond existing studies and investigates whether MOCE can help users learn pre-defined tactical heuristics by contrasting example behaviours in which the tactics have been used, against those in which they have not. The example and contrast behaviours have been chosen such that the behaviour that does not use the tactic(s) results in a poorer game outcome. Reflecting on the concept of algorithmic teaching by Huang et al. [12], we present informative examples contrasting behaviours that follow the tactics with those that do not, aiming to convey the effectiveness of these tactics. Using MOCE to support user learning provides a structured framework for understanding and applying effective tactics in multi-objective settings, thereby enhancing users' ability to make well-informed tactical decisions.

III. METHODS

The participants in our study are exposed to a game called *The Island of Joadia* (see Figure 1), a variation of *Disaster at the Joadia Islands* [46, 47]. In this game, a tsunami has hit the tiny, fictional nation of Joadia Island.

The participant is part of a team that heads up a Joint Task Force to supply food, water and medical treatment and to evacuate as many refugees as possible from the island. Participants control the behaviour of several units across multiple turns, tasking the units with supplying food or rescuing civilians. Their goal is to maximise their end-of-game score, defined in terms of the total number of civilian deaths and rescues.

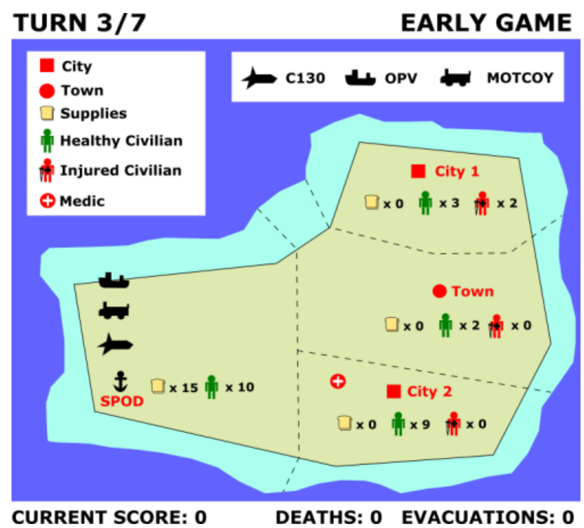

Fig. 1. Map of Joadia Island with four territories and four units (agents).

A. The Game

Joadia Island has four territories. The left-most territory, the SPOD (seaport of debarkation), is the location of the Joint Task Force base. This base contains supplies that can be distributed to the island's disaster-affected regions.

As a player, each participant controls the behaviour of four units: a C130 (aircraft); a MOTCOY (motorised company); an OPV (offshore patrol vessel); and a MEDIC. These units are located at the SPOD at the start of the game. A player can select a single activity for each unit to perform or leave it at its current location for the turn duration. Each unit's set of available activities depends on where the unit is located at the start of the turn. On any given turn, units perform their assigned behaviours in the order listed below. In Turn 1, players assign initial activities to their units.

- **C130** When located at the SPOD at the start of a turn, the C130 may evacuate up to 8 civilians located at the SPOD. Alternatively, the C130 may deliver 5 supplies from the SPOD to a city or town, ending the turn at its destination. When starting a turn at a city or town, the C130 may transport up to 5 healthy (not injured) civilians from that location to the SPOD.
- **OPV** When starting the turn at the SPOD, the offshore patrol vessel can deliver 3 supplies from the SPOD to

a city or town. When starting the turn at a city or town, the OPV may transport up to 3 healthy civilians from that location to the SPOD.

- **MOTCOY** The motorised company can deliver 2 supplies from the SPOD to a city or town, or transport up to 2 healthy civilians from a city or town to the SPOD.
- **MEDIC** The MEDIC may travel to one city or town on its turn, and heal up to 2 injured civilians.

The Island of Joadia is a resource management game. On any turn, a player assigns specific actions to each unit. Consider Figure 1. On this turn, one possible course of action is to use the C130 to deliver five supplies to City 2, the OPV to deliver three supplies to City 1, the MOTCOY to deliver two supplies to Town 1, and to use the MEDIC to heal the two injured civilians in City 1.

B. Environment

While the player controls the behaviour of game units, environmental events also occur. A *consumption* phase happens at the end of every second turn, where one supply will be consumed in each city and town for every three civilians (healthy and injured) at that location (rounded up). If there are insufficient supplies to meet demand, one healthy civilian at that location will die. One injured civilian from each city and town also dies, irrespective of the number of supplies present at the location. At the end of every third turn, the SPOD receives an additional 15 supplies in a *resupply* phase.

C. Game Score and Objectives

There are seven turns in the game. A player's score is composed of the number of evacuations performed, and civilian deaths that have occurred, over these seven turns. The score at any point in the game equals *2 × the total number of evacuated civilians - the total number of deaths*. The objective of the game is to maximise this score. The maximum possible end-of-game score equals *2 × the total number of civilians* present in the game scenario (i.e., all civilians are evacuated, and no civilians die). The score may be negative if too few civilians are evacuated, resulting in deaths.

An AI simulates the outcome of action choices on any given turn, using Monte Carlo Tree Search (MCTS) [48], and provides analysis in terms of domain-specific objectives. In our study, we define the following set of objectives or metrics. The first two communicate what the AI believes is likely to happen at the end of the game, while the latter two indicate what is likely to happen over the next one or two turns.

- *Expected evacuations*: expected number of evacuations the player will be able to perform by the end of the game according to the AI's simulations.
- *Expected deaths*: expected number of civilians that will die by game end according to the AI's simulations.
- *Civilians transported to the SPOD in the next turn*: number of civilians that the AI simulates will be transported from the cities and towns to the SPOD in the next turn. This metric highlights how effectively a player has distributed units to locations to maximise the number of civilians being relocated to the SPOD.
- *Expected number of deaths in the next consumption phase*: the number of civilians that the AI simulates will die in the next consumption phase. This metric highlights how effectively the player has used the MEDIC, and distributed supplies to cities and towns.

D. Game Tactics

There are many tactics that a player could employ throughout the game. Some of these are applicable in early turns, and some at later stages of the game. These tactics advise how to assign actions to units throughout the game to maximise a player's final score. We focus our study on three tactics to help players make good decisions in certain situations.

- **Max Capacity:** Each location has a different number of healthy civilians available for transport to the SPOD, and a varying need for supplies to sustain the civilians left behind. In general, players should send the vehicle with the largest capacity to the location with the most healthy civilians, the vehicle with the second largest capacity to the location with the second most number of healthy civilians, and so on. This will maximise the number of civilians brought back to the SPOD on the following turn, and distribute supplies most fairly.
- **Max MEDIC:** On a consumption turn, one healthy civilian in each location with insufficient supplies, and one injured civilian at each location, will die. A player can minimise deaths on these turns with careful use of the MEDIC. This tactic requires players to prioritise healing injured civilians at locations where the MEDIC can significantly impact survival rates. If one location has one injured civilian and another has three, healing the single civilian ensures only one death overall. However, if the player heals two of the three injured civilians, two will die, one at each location.
- **C130 Use:** The C130 has a dual role in the game. It can evacuate civilians located at the SPOD from the island, giving a player two points per evacuated civilian. It can also deliver supplies to regions, and transport healthy civilians from those regions back to the SPOD. During early turns, it is strategically best to use the C130 to deliver supplies, and maximise the number of civilians brought to the SPOD for later evacuation. In later turns, the C130 should be used solely for evacuation. It requires two turns for the C130 to travel to a region, and return to the SPOD. These are two turns where the C130 will not be able to evacuate civilians from the SPOD.

Each tactic's timing depends on the game phase and objectives. Max MEDIC is a single-turn tactic executed during consumption turns (2, 4, 6). Max Capacity takes two turns, best in early to mid-game (Turns 1–3) for SPOD preparation. C130 Use spans two turns: supply delivery early, evacuation later (Turns 4–7). One could imagine a larger set of tactics of relevance in this game. We focus on a small subset of low to moderate complexity to keep the study manageable. We selected these tactics based on our experience with the game and expected intuitiveness, with the *Max Capacity* and *Max MEDIC* tactics being the most (assigning the largest vehicle to the most civilians is a simple, logical strategy for maximising transport), and least (which injured civilians need to heal requires considering direct and indirect effects; immediate effects and overall survival across locations in future turns), intuitive, respectively.

E. Scenario Design

For each of the three tactics of interest, participants are presented with two training scenarios, followed by two questions in which they must select the best action assignments for each unit from three to four options. Finally, two training and test scenarios were presented that required the use of two or more of these tactics to make the best decision.

Participants are presented with a game state in a graphical, tabular, and textual format in each scenario. The graphical state takes the form of Figure 1, highlighting the locations of each unit; the number of healthy and injured civilians present at each location; and the number of supplies present at each location. The participants are told the current turn number, and the number of evacuations, deaths, and civilians healed thus far in the game. Alongside this information, participants are provided with an example behaviour for that state. The details of the example behaviour are defined in tabular and text formats. The state of the game after this behaviour is performed is shown in graphical, tabular, and text formats.

Participants not in the No Explanation group saw either a MOCE or SOCE. Both present a *contrast* behaviour that differs from the example. The contrasting behaviour highlights an alternative course of action that, according to the AI's simulations, will lead to a *different* end-of-game score and *different* values for the objectives defined in Section III-C. Each type of explanation expresses the differences between the two behaviours differently.

Each scenario is designed so that the difference in outcomes between the chosen and contrast behaviour reflects one or more of the tactics described in Section III-D. For example, the chosen behaviour may maximise the use of the MEDIC, which the contrast does not (or vice versa). Contrast behaviours are not necessarily the 'next-best' options, but are selected to best highlight the difference in game scores and objectives that arise when the tactic being taught is not followed.

F. Independent Variables – Explanation Types

To investigate the effectiveness of MOCE, we manipulate the *type* of explanation provided to participants during each training phase. We have three types of explanations, that is, three conditions. In all conditions, participants get the turn number in each training scenario, and the current game state information showing the locations of the units and the state of each territory. All participants are given an example action allocation to each unit, and shown the game state after the actions have been executed (i.e., the end of turn state). The three conditions manipulate the information given to participants. These conditions are:

1) **No Explanation:** As a baseline, we do not explain. All participants, including those in the No Explanation condition, receive the game state at the start and end of a turn and actions performed by a player. Participants in this condition do not see a contrast behaviour.
2) **SOCE:** Participants receive all information shown in Condition 1 and they are shown a chosen and contrast (alternate) behaviour that is compared in terms of expected end-of-game score, as determined by the AI using MCTS (Figure 2).
3) **MOCE:** Participants receive all information shown in Condition 2 and the four domain-specific objectives (Section III-C): expected deaths; expected evacuations; the number of civilians transported to the SPOD in the next turn; and the expected number of deaths in the next consumption phase (Figure 2).

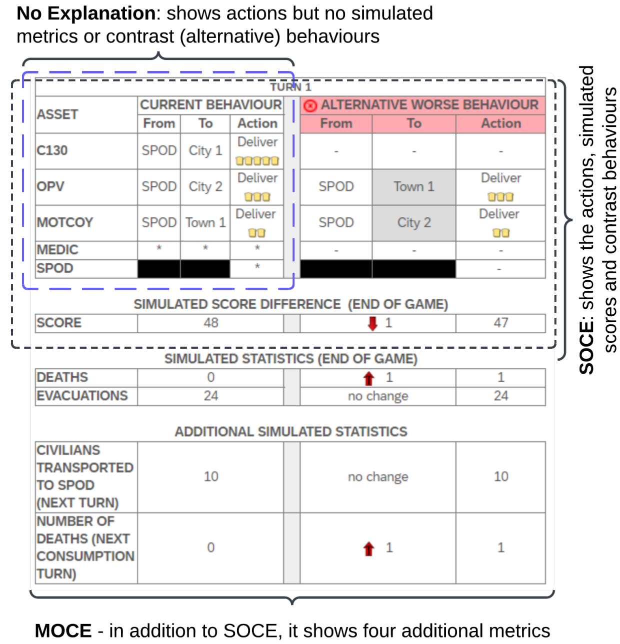

Fig. 2. Explanations provided to participants during training.

G. Dependent Variables

We collect quantitative and qualitative information to compare the impact of providing participants with different types of explanations during a training phase on a subsequent test phase. We use the following quantitative measures:

- **Accuracy:** Measures how accurately participants choose the best actions. During testing, participants evaluate scenarios with three to four behaviours, and accuracy is measured by how often they select the optimal action.
- **Confidence:** measures the participant's confidence that their selected choice will result in the highest possible score by the end of the game on a scale of 0% to 100%.
- **Satisfaction:** Measures how well participants feel they understood the explanations, rated on a 5-point Likert scale from *I strongly disagree* to *I strongly agree*.

We also asked the participants whether they had used any tactics by asking: *Please state the reason(s) for your chosen course of action above. For example, did you use any tactics or rules of thumb in your selection?*. We coded these and mapped

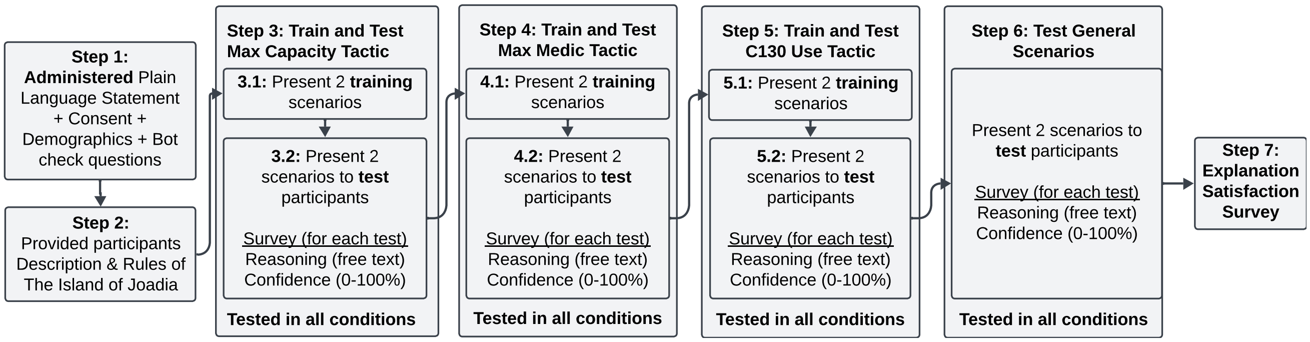

Fig. 3. A flowchart of the study procedure and survey administration. There were 14 scenarios in total; 6 training and 8 test.

them against the relevant tactics for the scenario. Recall that the scenarios (current state, actions, and resulting state) are designed to highlight a single tactic or combination of tactics in the contrast between behaviours.

H. Procedure

We conducted the study on Amazon MTurk, a popular platform for human-subject experiments, as participants had no prior experience with the task, making MTurk ideal for recruitment. The experiments were administered via a Qualtrics survey, with ethics approval from our institution. Participants were compensated at USD $15 per hour.

Figure 3 outlines the study procedure[1]. Participants received a plain language statement and provided consent to proceed, followed by logical questions to filter out automated respondents. Participants provided their Amazon MTurk WorkerID and filled in a demographics questionnaire, which had questions about how many games a participant plays weekly and what games they played. We used this question to distribute the participants almost equally between the conditions to avoid the gaming experience confounding performance.

We set up the study so the participants focused on learning one tactic at a time. We focused on one tactic at a time to reduce participants' cognitive load. Before training and testing each tactic, we provided specific instructions, such as for the *C130* tactic: *The C130 can perform different tasks at different times. In this phase, we will* ***train/now test*** *you to use the C130 effectively.* Similar instructions were given for other tactics. This ensured that participants understood the training's purpose and could concentrate on mastering each tactic.

We presented two scenarios related to a tactic in a training phase, and immediately after training, we tested the participants' understanding through two test scenarios in a testing phase. In this manner, the study alternated between training and testing phases for each of the three tactics of Section III-D. Finally, we included two general test scenarios that combined several objectives and required participants to integrate multiple tactics without explicit cues, allowing us to assess whether MOCE supports the transfer of learned heuristics to novel, more complex decision problems. All conditions were tested identically on general scenarios, with only the explanation types differing. We randomly allocated participants to one of the three conditions–No Explanation; SOCE; and MOCE–and randomised the ordering of scenarios within each training and test phase.

In each testing phase, we provided participants with two scenarios containing multiple-choice questions, where the choices were a combination of one optimal and two to three sub-optimal choices. The sub-optimal options were chosen to cover a broad range of possible end-of-game scores, from clearly sub-optimal to near-optimal. Each test scenario provided participants with the game state at the start of a specific turn, in both a graphical and textual format. The participants were required to select the best course of action from three to four options, state how confident they were that the chosen course of action would result in the highest score at the end of the game and whether they used any tactics. The order of the three to four options was randomised.

IV. Results

Sixty-six MTurkers participated in the study. We had 21 participants in the No Explanation condition, 23 in the SOCE condition and 22 in the MOCE condition. Per week, 8 participants played board/video games for more than 15 hours, 6 between 10 and 15 hours, 16 between 5 and 10 hours, 22 less than 5 hours, 12 rarely played, and 1 did not play. There were 28 participants aged 35-44, 22 aged 45-54, 13 aged 55-64, 18 aged 25-34, and 2 undisclosed; 25 have bachelor's degrees, 13 associate degrees, 11 some college, 9 high school diplomas, 4 master's degrees, 2 professional degrees, 2 less than high school, and 2 undisclosed.

All statistical tests were conducted using R. We conducted Fisher's Exact Test to detect significant differences between conditions concerning optimal choices, pairwise. We used the Mann-Whitney U test for other data, such as participant confidence ratings and Hoffman's Satisfaction Scale. We designed the survey (using branching in Qualtrics) to distribute participants almost equally between the conditions using their level of game-playing experience. A mixed-effects linear model confirmed no significant bias in participants' gaming experience across conditions.

For **Max Capacity**, which focused on how to assign vehicles to locations, we expected that most participants would follow the desired tactic, by distributing units and resources to locations in order of need, even without being provided with training or explanations, as we felt this behaviour was the most

[1]Example survey at: https://github.com/singhrr/tactical-xai-moce-study

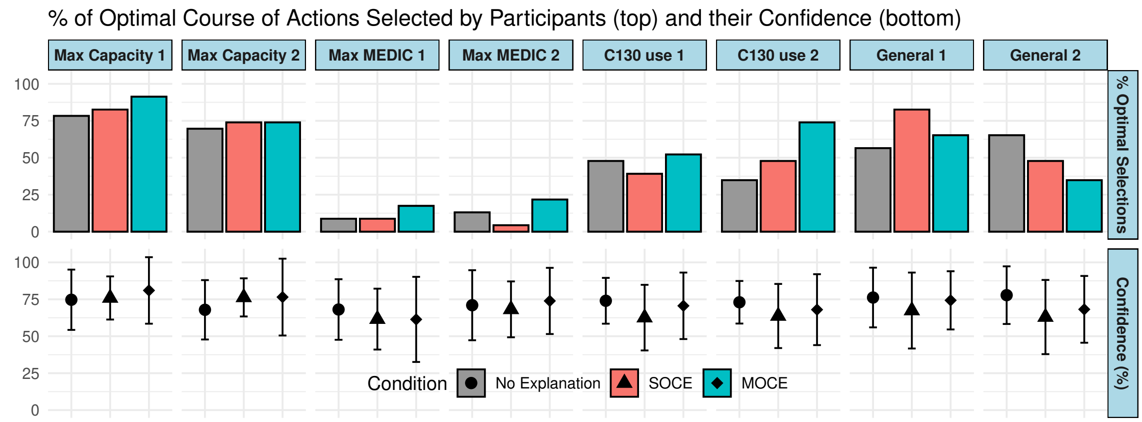

Fig. 4. The top figure shows the percentage of participants in the respective condition who selected the correct option and their mean confidence (bottom).

intuitive. We expected participants across the three explanation conditions to perform similarly in the testing phase.

For **Max MEDIC**, which focused on how to best use the MEDIC on a consumption turn, we expected that most participants, who did not receive training with explanations, would send the MEDIC to locations with the highest number of injured civilians, rather than to locations where the MEDIC can heal *all* injured civilians present. MOCEs indicate how many civilians will die in the next consumption phase. We expected that participants provided with MOCEs would perform better in the testing phase, as they highlighted that what might be the most intuitive choice, according to our experience and assessment, is not necessarily the best.

For **C130 Use**, we expected that most participants who did not receive explanations would use the C130 to evacuate civilians from the SPOD whenever possible. This greedy choice yields an immediate increase in points. We expected that participants who had received explanations would outperform those who did not during testing–choosing to evacuate in later turns while using the C130 for supply delivery and civilian transport to the SPOD in early turns. As the number of evacuations forms the largest component of a player's score, it is unclear how informative any additional metrics would be for participants across these scenarios.

The scenarios *Max Capacity 1* and *Max Capacity 2* tested participants on the *Max Capacity* tactic. *Max MEDIC 1* and *Max MEDIC 2* tested *Max MEDIC*, and *C130 use 1* and *C130 use 2* tested the *C130 Use* tactic. Two scenarios *General 1* and *General 2* tested combinations of all three tactics.

A. Performance: Number of times best action was chosen

Figure 4 shows the percentage of correct (optimal) choices made by participants in each condition across the test questions for each tactic and the two general tests.

Participants receiving explanations during training showed an advantage over no explanations. For C130 Use 2, Fisher's test revealed weak support for the difference between the No Explanation and MOCE conditions with $p = 0.047$ (an odds ratio of 0.096 (95% CI: 0.002 to 1.325)). However, General 2 showed a detrimental influence on receiving explanations. Fisher's test revealed a significant difference between the No Explanation and MOCE conditions with $p = 0.006$ (an odds ratio of 0.04 (95% CI: 0.006 to 0.578)). No significant differences were noted otherwise.

MOCE outperforms the No Explanation condition in seven of the eight tests and SOCE in five tests. SOCE provides some improvement but is generally less effective than MOCE, outperforming the No Explanation condition in only four tests. No Explanation results in lower performance across most tasks, highlighting the value of providing explanations. While MOCE excelled in most tasks, simpler SOCE and No Explanation conditions sometimes performed better in combined tasks (general tests), possibly due to reduced cognitive load and easier tactic integration. These findings suggest that further research is needed to optimise MOCE to better support users in integrating tactics during decision-making. Given the varied performance in general tests and lack of significant differences, our results partially support H1.

B. Understanding of Tactics

For each tactic, we quantified the behaviour in each test scenario if the tactic was employed, observed whether participants used it, and then analysed their qualitative responses to gain further insight into their cognitive process. We concluded that participants learnt a tactic if they correctly answered both of its tests and provided textual comments that expressed the tactic in some form. Our results partially support H2 due to varying performance and qualitative analysis (see below).

1) Tactic 1: Max Capacity: Across all three conditions, the same number of participants correctly answered both test questions in this category. Most participants had learnt the correct tactic whether or not they received explanations during training. Where a participant answered incorrectly, they typically opted to send multiple vehicles to a single location––the location with the greatest number of civilians-–thinking they could get supplies to any unvisited location on the next turn. They did not understand how many turns are required to deliver supplies and the timing of the consumption phase.

2) Tactic 2: Max MEDIC: Participants found tests in this category to be difficult. One participant appeared to have learned the correct tactic in the No Explanation condition, based on their textual responses. In the SOCE group, no participants appeared to learn the right tactic, while 4 participants in the MOCE group appeared to understand the tactic well.

3) Tactic 3: C130 Use: Across the No Explanation, SOCE, and MOCE groups, three, one, and six participants, respectively, demonstrated they had learnt the correct tactic based on their textual responses. As was the case across the MEDIC tests, more MOCE participants identified the correct tactic.

4) General Tests: The general tests were the last ones presented to participants. Of the two general tests, the first required participants to use the C130 to evacuate civilians and the second for supply delivery. Participants in the MOCE group generally prioritised evacuation across both tests, suggesting that they may not have truly understood the link between the game turn and stage to the ideal behaviour for the C130. Mental workload may have influenced results, as explanation groups, especially MOCE, processed more information.

TABLE II
ESTIMATED PARTICIPANTS PER CONDITION WHO LEARNT THE TACTIC.

	No Exp	SOCE	MOCE
Max Capacity	16 (76.2%)	16 (69.6%)	16 (72.7%)
Max MEDIC	1 (4.8%)	0 (0.0%)	4 (18.2%)
C130 Use	3 (14.3%)	1 (4.3%)	6 (27.3%)

C. Player confidence and Satisfaction

We did not observe any statistically significant differences in how confident participants were in their answers to each test question, across the three conditions. Participants who received MOCEs tended to be slightly more confident in their answers than those receiving SOCEs across most test scenarios. No Explanation participants, however, were the most confident, on average, $m = 72.8$ $(sd = 19.4)$ vs SOCE, $m = 67.3$ $(sd = 21.0)$ and MOCE, $m = 71.8$ $(sd = 23.9)$, especially in the *Max MEDIC 1*, *C130* tests, and *General* tests (see Figure 4 for a breakdown by test scenario).

While MOCE often results in more correct selections, participants' confidence in receiving MOCEs was not always the highest. In contrast, participants in the No Explanation group tended to have higher confidence despite not always having the best performance. This discrepancy suggests that explanations might be causing participants to reassess their confidence, making them more critical of their decisions, leading to lower confidence despite better performance.

Figure 5 shows the Hoffman's Satisfaction survey results. There were no significant differences in ratings for MOCEs and SOCEs. While our results do not support H3, they show that MOCE ($m = 3.98$ $sd = 1.06$) explanations are generally favoured over SOCE ($m = 3.75$ $sd = 1.09$), especially for providing sufficient detail and improving understanding. Although the differences are small, MOCE consistently scores higher, suggesting it offers greater overall satisfaction.

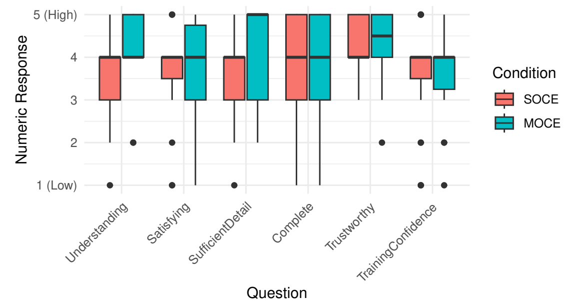

Fig. 5. Hoffman's explanation satisfaction ratings.

D. Qualitative Analysis

Table II reports the estimated number of participants who appeared to have learnt the correct tactics based on their performance across both tests associated with each tactic and their textual responses. Explanations during training were not needed for participants to grasp Max Capacity, with slightly more participants in no explanation conditions answering both questions correctly and providing textual responses demonstrating that they understood the tactic as intended. For Max MEDIC and C130 Use, the value of MOCEs becomes clearer. More participants demonstrated they understood these tactics based on their textual responses and correct answers.

V. Discussion and Conclusion

This study explored the effectiveness of different explanation methods (No Explanation, SOCE, and MOCE) in supporting participants' learning of tactical decision-making in a multi-turn resource management game. Our findings demonstrate that MOCE was the most promising of the three, enhancing participants' ability to learn and apply desired tactics in most scenarios (R1, R2). MOCE improved outcomes in 7 of 8 tests, though not significantly, indicating potential benefits that warrant further study.

No Explanation or SOCE often performed better when scenarios required integrating multiple tactics. We hypothesise this may be due to reduced cognitive load, as participants without explanations avoided processing detailed trade-offs, unlike those with explanations, especially MOCE, making it harder for them to integrate multiple tactics effectively. Future studies could explore this further with focused tactic training.

The correlation analysis revealed mixed results between confidence and performance. Experimental studies [49, 50] suggest most participants do not cognitively engage with explainability tools and those who do often overestimate their understanding, even with uninformative explanations [49, 50]. Therefore, we need better tools to engage participants cognitively rather than relying solely on explanations.

Despite MOCE's effectiveness, satisfaction ratings were similar across all conditions, though MOCE scored slightly higher for providing sufficient detail and improving understanding. Additionally, while this study explicitly identified tactics for experimental purposes, the broader goal is to leverage MOCE explanations to reveal tactics learned by an agent without intermediate intervention, which could potentially facilitate tactic discovery and experiential learning.

A. Limitations and Future Work

Beyond self-report, future work could use interviews or think-aloud protocols for deeper insight into participants' mental models. Focusing on individual tactics, randomising training, or expanding scenarios could improve assessment. Also, adding more objectives may add new information (such as different timescales when actions and objectives become relevant), suggesting that MOCE's success may involve multiple factors, not just the objectives alone, warranting further research. Joadia offers baseline insights, but its simplicity limits generalisability; future work should explore more complex tactical environments. We define a tactic as a heuristic, optimising short-term actions for long-term goals. Future work will refine this definition within broader decision-making frameworks.

While effective, our method followed the standard approach of providing a recommendation with an explanation. As noted, studies [49, 50] show participants often fail to engage with explainability tools and overestimate their understanding. Recent discussions recommend engaging decision-makers by using tools that provide feedback [51], encouraging interaction friction [52], or temporarily withholding AI recommendations [35]. To support tactic learning in future work, an approach like Miller's [51] could present the top five recommendations, along with their pros and cons, based on multiple objectives.

References

[1] R. Dazeley, P. Vamplew, and F. Cruz, "Explainable Reinforcement Learning for Broad-XAI: A Conceptual Framework and Survey," Aug. 2021.

[2] X. Peng, M. O. Riedl, and P. Ammanabrolu, "Inherently explainable reinforcement learning in natural language," *Advances in neural information processing systems*, vol. abs/2112.08907, Dec. 2021.

[3] B. Krarup, S. Krivic, D. Magazzeni, D. Long, M. Cashmore, and D. E. Smith, "Contrastive Explanations of Plans through Model Restrictions," *Journal of Artificial Intelligence Research*, vol. 72, pp. 533–612, Oct. 2021.

[4] T. Chakraborti, S. Sreedharan, Y. Zhang, and S. Kambhampati, "Plan explanations as model reconciliation: Moving beyond explanation as soliloquy," in *IJCAI 2017*, 2017, pp. 156–163.

[5] R. Sukkerd, R. Simmons, and D. Garlan, "Tradeoff-Focused contrastive explanation for MDP planning," in *RO-MAN*, 2020, pp. 1041–1048.

[6] Z. Juozapaitis, A. Koul, A. Fern, M. Erwig, and F. Doshi-Velez, "Explainable reinforcement learning via reward decomposition," in *IJCAI/ECAI Workshop on Explainable AI*, 2019.

[7] G. Misitano, B. Afsar, G. Lárraga, and K. Miettinen, "Towards explainable interactive multiobjective optimization: R-ximo," *Autonomous Agents and Multi-Agent Systems*, vol. 36, no. 2, p. 43, 2022.

[8] M. S. Pfaff, G. L. Klein, J. L. Drury, S. P. Moon, Y. Liu, and S. O. Entezari, "Supporting complex decision making through option awareness," *Journal of Cognitive Engineering and Decision Making*, vol. 7, no. 2, pp. 155–178, 2013.

[9] J. Gajcin and I. Dusparic, "Redefining counterfactual explanations for reinforcement learning: Overview, challenges and opportunities," *ACM Comput. Surv.*, vol. 56, no. 9, pp. 1–33, Apr. 2024.

[10] H. J. Einhorn and R. M. Hogarth, "Confidence in judgment: Persistence of the illusion of validity." *Psychological review*, vol. 85, no. 5, p. 395, 1978.

[11] E. Soyer and R. M. Hogarth, "Learning from experience in nonlinear environments: Evidence from a competition scenario," *Cognitive Psychology*, vol. 81, pp. 48–73, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010028515000602

[12] S. H. Huang, D. Held, P. Abbeel, and A. D. Dragan, "Enabling robots to communicate their objectives," *Autonomous Robots*, vol. 43, pp. 309–326, 2019.

[13] Y. Septon, T. Huber, E. André, and O. Amir, "Integrating policy summaries with reward decomposition for explaining reinforcement learning agents," in *PAAMS*. Springer, 2023, pp. 320–332.

[14] S. Nguyen, G. O'Keefe, S. Arisian, K. Trentelman, and D. Alahakoon, "Leveraging explainable AI for enhanced decision making in humanitarian logistics: An adversarial CoevoluTION (ACTION) framework," *International Journal of Disaster Risk Reduction*, vol. 97, p. 104004, Oct. 2023.

[15] J. A. Miller, B. Horn, M. Guthrie, J. Romano, G. Geva, C. David, A. R. Sterling, and S. Cooper, "How do players and developers of citizen science games conceptualize skill chains?" *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CHI PLAY, pp. 1–29, Oct. 2021.

[16] B. A. Friedman, "Tactics," in *Understanding the US Military*. Routledge, 2022, pp. 242–258.

[17] G. POLYA, "Methodology or heuristics, strategy or tactics ?" *Archives de Philosophie*, vol. 34, no. 4, pp. 623–629, 1971.

[18] R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman, "Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-ai performance," *Frontiers in Computer Science*, vol. 5, p. 1096257, 2023.

[19] O. Lammert, B. Richter, C. Schütze, K. Thommes, and B. Wrede, "Humans in XAI: increased reliance in decision-making under uncertainty by using explanation strategies," *Frontiers in Behavioral Economics*, vol. 3, 2024.

[20] J. van der Waa, E. Nieuwburg, A. Cremers, and M. Neerincx, "Evaluating XAI: A comparison of rule-based and example-based explanations," *Artificial intelligence*, vol. 291, p. 103404, Feb. 2021.

[21] F. Cabitza, C. Natali, L. Famiglini, A. Campagner, V. Caccavella, and E. Gallazzi, "Never tell me the odds: Investigating pro-hoc explanations in medical decision making," *Artificial intelligence in medicine*, vol. 150, p. 102819, Apr. 2024.

[22] B. Leichtmann, A. Hinterreiter, C. Humer, M. Streit, and

M. Mara, "Explainable Artificial Intelligence improves human decision-making: Results from a mushroom picking experiment at a public art festival," Sep. 2022.

[23] P. Madumal, T. Miller, L. Sonenberg, and F. Vetere, "Explainable Reinforcement Learning through a Causal Lens," in *AAAI*, vol. 34, no. 03, 2020, pp. 2493–2500.

[24] W. R. Swartout and J. D. Moore, "Explanation in second generation expert systems," in *Second Generation Expert Systems*. Springer, 1993, pp. 543–585.

[25] A. Jacovi, A. Marasović, T. Miller, and others, "Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI," in *FAccT*, 2021.

[26] V. Sivaraman, L. A. Bukowski, J. Levin, J. M. Kahn, and A. Perer, "Ignore, Trust, or Negotiate: Understanding Clinician Acceptance of AI-Based Treatment Recommendations in Health Care," in *CHI*, 2023.

[27] J. S. Kim, V. Chen, D. Pruthi, N. B. Shah, and A. Talwalkar, "Assisting Human Decisions in Document Matching," Feb. 2023.

[28] A. Bertrand, J. R. Eagan, and W. Maxwell, "Questioning the ability of feature-based explanations to empower non-experts in robo-advised financial decision-making," in *FAccT*, 2023, pp. 943–958.

[29] M. Lee, M. Srivastava, A. Hardy, J. Thickstun, E. Durmus, A. Paranjape, I. Gerard-Ursin, X. L. Li, F. Ladhak, F. Rong *et al.*, "Evaluating human-language model interaction," *arXiv preprint arXiv:2212.09746*, 2022.

[30] L. Famiglini, A. Campagner, M. Barandas, G. A. La Maida, E. Gallazzi, and F. Cabitza, "Evidence-based XAI: An empirical approach to design more effective and explainable decision support systems," *Computers in biology and medicine*, vol. 170, p. 108042, Feb. 2024.

[31] F. Cabitza, A. Campagner, and L. M. Sconfienza, "Studying human-AI collaboration protocols: the case of the Kasparov's law in radiological double reading," *Health information science and systems*, vol. 9, no. 1, p. 8, 2021.

[32] M. H. Lee and C. J. Chew, "Understanding the Effect of Counterfactual Explanations on Trust and Reliance on AI for Human-AI Collaborative Clinical Decision Making," *Proc. ACM Hum.-Comput. Interact.*, vol. 7, no. CSCW2, pp. 1–22, Oct. 2023.

[33] H. Vasconcelos, M. Jörke, M. Grunde-McLaughlin, T. Gerstenberg, M. Bernstein, and R. Krishna, "Explanations Can Reduce Overreliance on AI Systems During Decision-Making," in *CHI*, 2023, pp. 1–38.

[34] K. Z. Gajos and L. Mamykina, "Do People Engage Cognitively with AI? Impact of AI Assistance on Incidental Learning," in *Proceedings of IUI*, 2022, pp. 794–806.

[35] Z. Buçinca, M. B. Malaya, and K. Z. Gajos, "To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance in AI-assisted Decision-making," *Proc. ACM Hum.-Comput. Interact.*, vol. 5, pp. 1–21, 2021.

[36] B. Green and Y. Chen, "Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments," in *FAccT*, 2019, pp. 90–99.

[37] F. Poursabzi-Sangdeh, D. G. Goldstein, J. M. Hofman, J. W. Vaughan, and H. Wallach, "Manipulating and Measuring Model Interpretability," in *CHI*, 2021, pp. 1–52.

[38] M. Nourani, C. Roy, J. E. Block, D. R. Honeycutt, T. Rahman, E. Ragan, and V. Gogate, "Anchoring Bias Affects Mental Model Formation and User Reliance in Explainable AI Systems," in *Proceedings of IUI*, 2021, pp. 340–350.

[39] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[40] C. Chen, O. Li, A. Barnett, J. Su, and C. Rudin, "This looks like that: deep learning for interpretable image recognition," *Advances in neural information processing systems*, pp. 8928–8939, Jun. 2018.

[41] T. Miller, "Contrastive explanation: A structural-model approach," *The Knowledge Engineering Review*, vol. 36, p. e14, 2021.

[42] G. Warren, R. M. J. Byrne, and M. T. Keane, "Categorical and Continuous Features in Counterfactual Explanations of AI Systems," in *Proceedings of IUI*, 2023, pp. 171–187.

[43] A. Verma, V. Murali, R. Singh, P. Kohli, and S. Chaudhuri, "Programmatically interpretable reinforcement learning," in *ICML*, 2018, pp. 5045–5054.

[44] O. Amir, F. Doshi-Velez, and D. Sarne, "Summarizing agent strategies," *Autonomous Agents and Multi-Agent Systems*, vol. 33, pp. 628–644, 2019.

[45] A. Anderson, J. Dodge, A. Sadarangani, Z. Juozapaitis, E. Newman, J. Irvine, S. Chattopadhyay, A. Fern, and M. Burnett, "Explaining reinforcement learning to mere mortals: An empirical study," in *IJCAI*, 2019, pp. 1328–1334.

[46] P. Williams, *Disaster at the Joadia Islands - Rulebook*. Defence Science and Technology Group, 2019.

[47] M. Wong, D. Kamenetsky, and P. Williams, "Towards an intelligent agent for a humanitarian disaster relief wargame," in *International Congress on Modelling and Simulation*, 2019.

[48] G. Chaslot, S. Bakkes, I. Szita, and P. Spronck, "Monte-carlo tree search: A new framework for game AI," in *Proceedings of the AAAI Conference on AIIDE*, vol. 4, no. 1, 2008, pp. 216–217.

[49] M. Chromik, M. Eiband, F. Buchner, A. Kruger, and A. Butz, "I think I get your point, AI! the illusion of explanatory depth in explainable AI," in *Proceedings of IUI*, 2021.

[50] M. Eiband, D. Buschek, A. Kremer, and H. Hussmann, "The Impact of Placebic Explanations on Trust in Intelligent Systems," in *Extended Abstracts of CHI*, 2019, pp. 1–6.

[51] T. Miller, "Explainable AI is dead, long live explainable AI! hypothesis-driven decision support using evaluative AI," in *Proceedings of FAccT 2023*, 2023, pp. 333–342.

[52] F. Cabitza, A. Campagner, L. Famiglini, C. Natali, V. Caccavella, and E. Gallazzi, "Let Me Think! Investigating the Effect of Explanations Feeding Doubts About the AI Advice," in *Machine Learning and Knowledge Extraction*, 2023, pp. 155–169.