Exploratory Wargaming with a "Superhuman" Tactician in the Team: A Controlled Experiment

Nicholas Bell
Defence Science and Technology Laboratory
nbell@dstl.gov.uk

Joel Brynielsson
Swedish Defence Research Agency
joel.brynielsson@foi.se

Mika Cohen
Swedish Defence Research Agency
mika.cohen@foi.se

Simon Collander-Brown
Defence Science and Technology Laboratory
sjcbrown@dstl.gov.uk

Johan Elg
Swedish Defence Research Agency
johan.erik.elg@foi.se

Leon Ericsson
Swedish Defence Research Agency
leon.ericsson@foi.se

Johan Ivari
Swedish Defence University
johan.ivari@fhs.se

Christoffer Limér
Swedish Defence Research Agency
christoffer.limer@foi.se

Noah Mannberg
Swedish Defence Research Agency
noah.mannberg@gmail.com

Abstract

The paper introduces a prototype decision support system for rapid exploratory wargaming of ground combat together with a conceptual framework for human-machine teaming. An AlphaZero agent provides continuous, GPS-device-like advice on how blue/red should proceed from the current situation to meet/prevent the long-term mission objective for blue. A controlled experiment with 100+ senior officers provides concrete evidence that the utility of wargaming in a digital environment improves with an AlphaZero agent in the team of players.

Keywords: Wargaming, *AlphaZero*, AI-supported decision-making, human-machine teaming.

1 Introduction

In early 1941, when German submarines were destroying Allied shipping at a devastating rate, Churchill ordered the British Navy to "Find out what's going on and sink the U-boats" [1]. A new tactics development unit was created, the Western Approaches Tactical Unit (WATU), where staff simulated submarine attacks and developed countermeasures through wargaming. The rules of the wargames reflected known physical properties of merchant ships, escorts and submarines in terms of speed, turning circle, visibility, armament and so on, but the rules left tactical decisions about formation, etc. open to the players to choose freely. Experimenting with tactics, the staff arrived at the best formations and search patterns for protecting the convoys through a creative, iterative process of trial-and-error.

The exploratory wargaming at WATU played an important role for the development of the Battle of the Atlantic [2]. It is not unreasonable to assume that "Similar challenges in the future could be tackled even more quickly and effectively with the help of AI programs like *AlphaZero*" [3], thereby "blurring the boundary between wargaming, game theory and OA" [4].

This paper reports on a controlled experiment studying exploratory wargaming in the tradition of the wargaming at WATU, but in a digital environment with *AlphaZero* in the team of players.

1.1 AlphaZero

Introduced in 2018, *AlphaZero* [5] is a general AI for double-sided strategy games that learns to play a given rule set through massive amounts of self-play. As *AlphaZero* repeatedly plays the game against itself, its decision-making converges towards the game-theoretic optimal (Nash equilibrium).

AlphaZero is still today considered one of DeepMind's flagships. The algorithm has changed the understanding of classic strategy games such as chess and go, uncovering more effective tactics that have eluded centuries of human creativity [6] [7]. More recently, *AlphaZero* has discovered novel algorithms for ubiquitous computational tasks, such as search and matrix multiplication, that surpasses the existing state-of-the-art algorithms that had taken decades of creative fine tuning to create within computer science research [8, 9, 10].

1.2 WARGAMING AND ALPHAZERO

Today, *AlphaZero* and derivations of *AlphaZero* are used routinely by professional players within e.g. chess when preparing for a tournament, reliably suggesting options overlooked by the professional players. However, the applicability of *AlphaZero* and related forms of self-play AI to professional wargaming has been questioned. According to a recent study from the Alan Turing Institute [11], there is "limited evidence on the successes and failures" and "few real-world case studies offer concrete evidence of effectiveness". The study concludes that "Despite the abundance of ambitious ideas, there remain significant doubts about whether any of these **are** 1) feasible or 2) helpful in answering decisionmakers' questions".

1.3 EXPLORATORY WARGAMING

In contrast to the wargames at WATU, wargaming today can run the risk of becoming a performative exercise in a planning process focused more on ticking boxes and formal artifacts than on fostering a deeper, shared understanding. There have been multiple calls within war studies for more iterated, exploratory forms of wargaming [12] [13]. In particular, in the Set Based Approach [14], wargaming shifts focus from delivering a static product to a process that aims to 'marinate' the team in contextual understanding, expanding the *Cognitive Space of Possibilities* through iterative exploration and refinement of multiple potential futures.

Relatedly, there is research to suggest that the extra value offered by a multiple-options evaluation model compared to a single-option evaluation model may sometimes be low in naturalistic decision situations [15]. This underscores the importance of contextual, iterative approaches to decision-making and problem-solving that prioritize understanding and alignment over formal comparison of discrete alternatives as prescribed by the MDMP [13].

1.4 CONTRIBUTION

This paper explores the use of *AlphaZero* in exploratory wargaming. The main contributions are as follows:

- A conceptual framework, based on the Set Based Approach, for understanding human-machine teaming in exploratory wargaming and in planning more broadly.
- An *AlphaZero*-based decision support system for exploratory wargaming of ground combat in line with the conceptual framework. The *AlphaZero*-agent provides continuous advice on how blue units should move and fire to meet the long-term mission objective for blue, as well as continuous advice on how red units should move and fire to prevent the mission objective for blue.
- Computer simulations measuring the time and parallel hardware required for *AlphaZero* to learn an existing rule set from professional wargaming (addressing the concern about feasibility in the study from the Alan Turing Institute).
- A controlled experiment with 114 senior officers measuring *AlphaZero*'s ability to expand the Cognitive Space of Possibilities when wargaming in a digital environment (addressing the concern about utility in the study from the Alan Turing Institute).

1.5 ORGANIZATION OF PAPER

The rest of the paper is organized as follows. Section 2 introduces a conceptual framework for understanding human-machine teaming. Section 3 introduces and demonstrates *AlphaSTRIKE*, a digital wargaming environment with *AlphaZero*-support in line with the proposed conceptual framework. Section 4 describes the parallel implementation of *AlphaZero* in *AlphaSTRIKE* and studies how well it scales on GPU hardware. Section 5 reports on a controlled experiment with senior officers that measures the utility of advice from *AlphaZero* when wargaming a battle plan in a digital environment. Finally, Section 6 concludes.

2 EXPLORATORY WARGAMING: A CONCEPTUAL FRAMEWORK FOR HYBRID COGNITIVE SYSTEMS

This section establishes a conceptual foundation for understanding and designing human-machine teaming in wargaming. Rather than focusing solely on AI's ability to generate resilient Courses of Action (COAs), we examine how integrating AI shapes cognition, collaboration, and trust in hybrid human-machine teaming. Planning is conceptualized here as an adaptive, recursive meaning-making process under uncertainty, emphasizing iterative exploration to expand a team's *Cognitive Space of Possibilities*.

We integrate three complementary frameworks to reconceptualize planning in this way:

- **Set-based Approach (SbA)** [14]: Frames planning as the co-evolution of problems and solutions, expanding the *Cognitive Space of Possibilities* through iterative exploration and refinement rather than prematurely narrowing options.

- **Harmonization Emergence Model (HEM)** [16]: Explains how alignment and trust emerge from recursive, *Participatory Sense-Making*, where transient *Stimmigkeit* accumulates into a more resilient, fluid sense of *coalescing unity* (*Einheit*).
- **Boyd's OODA loop, interpreted via Luhmann's *double contingency*** [17, 18]: Highlights the recursive mutual adjustments of *Orientation* that underpin *Co-creation of Meaning* and trust in dynamic interactions.

These frameworks position AI not as a passive tool but as an active co-creator in sense-making and coordination.

Within this framework, AI can perform three core functions — **Cognitive Resonator**, **Disruptive Irritator**, and **Meaning-making Enabler** — that together frame AI as an active participant in distributed, recursive human–machine teaming for meaning-making (see Table 1 at the end of this section for an overview of the three AI functions).

The following subsections (2.1–2.3) describe how these functions support planning, trust, collaboration, and the *Co-creation of Meaning* (all *italicized* terms are explained in *Section 7, Glossary of Key Terms*).

2.1 Planning & Possibilities (SbA)

The *Set-based Approach* reframes planning as a dynamic, contingent, and recursive meaning-making process that continuously expands the *Cognitive Space of Possibilities*. Planning is not about producing a fixed set of COAs, but about enabling the co-evolution of problems and solutions, a phenomenon termed the *Problem–Solution Eclipse* [14].

Instead of narrowing options too early, SbA emphasizes cultivating a shared *Orientation* within a rich cognitive space, allowing exploration of multiple, interrelated futures. As Ivari and Nolan note:

> *[W]hen planning, the focus should shift from the product (the plan) to the process, where the purpose of planning is to marinate a team in contextual understanding* [14, p.1].

Temporal awareness prioritizes **Kairos***,* the opportune moment sensed through contextual attunement, over linear, speed-measured **Chronos**. Planning becomes akin to a high-stakes sport, requiring continuous adaptation, improvisation, and re-*Orientation* [14]. This aligns with Boyd's emphasis on tempo and harmony over raw speed [19].

Central to this process is the *Cognitive Team Schema* (**CTS**): a shared, dynamic *Orientation* that mitigates bias, reduces groupthink, and enhances collective adaptability. SbA's principle, "If you plan it, you run it" [14], highlights the CTS as the primary vehicle for agile adaptation, enabling the team to respond fluidly to the evolving environment rather than merely following a preset plan.

2.1.1 AI's Role: Augmenting the Cognitive Ecology

Within SbA, AI augments rather than replaces human judgment. It contributes to the distributed cognitive ecology[1] of people, artifacts, and interactions, supporting the formation and maintenance of the CTS. AI expands the *Cognitive Space of Possibilities*, surfaces novel patterns, and enhances narrative coherence across multi-actor operations. Its role is co-creative, contingent, and aligned with the recursive nature of adaptive planning. within which meaning unfolds.

2.2 Trust & Unity (HEM)

Trust and collaboration in hybrid teams emerge from *Participatory Sense-Making*, the *Co-creation of Meaning* through reciprocal, adaptive interaction [21]. *Stimmigkeit*, a transient alignment of *Orientation*, plays a crucial role: it enables sufficient coordination without demanding consensus, allowing productive dissonance to expand the *Cognitive Space of Possibilities* [16].

As *Stimmigkeit* accumulates, it fosters *Einheit* — a resilient and fluid unity that sustains coherent team *Orientation* amid uncertainty [16], emerging as a **constitutive constraint regime** as defined by Juarrero [22], which underpins the team's collaborative **metastability**.

This dynamic is captured by the Harmonization Emergence Model (HEM), illustrating how transient

[1] This concept builds on Gibson's ecological framework, proposing that cognition is not confined to the brain but distributed across individuals, artifacts, language, and the environment, focusing on how *affordances*, action possibilities offered by the environment, shape embodied and situated cognitive processes where *meaning* arises [20].

alignment among team members can maintain collaborative metastability even in the face of entropy and uncertainty (Figure 1) [16].

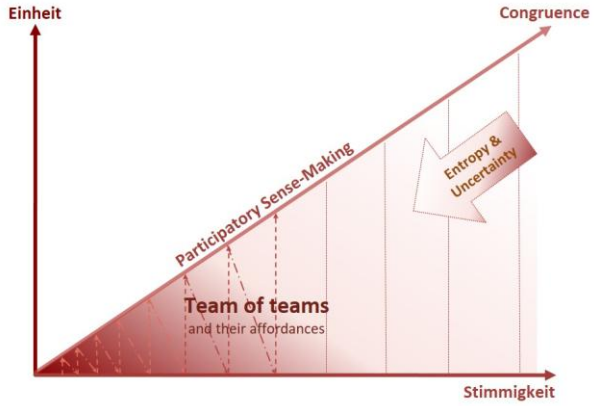


Figure 1: The Harmonization Emergence Model 2D[2].

2.2.1 AI's Role: Facilitating Trust and Collaboration

AI supports trust and collaboration by expanding the team's *Cognitive Space of Possibilities*. Through forward looking insights, recommendations, and contrasting perspectives, AI can enhance *Stimmigkeit* by aligning *Orientations* toward actionable congruent shared meaning. Misaligned or ungrounded inputs can disrupt this alignment, increasing uncertainty and reducing coordination (as illustrated by the Entropy & Uncertainty–labelled arrow in Figure 1). This highlights the fragility of trust and the need for careful calibration of AI support.

2.3 Co-creation of Meaning (OODA & Double Contingency)

Double contingency, a concept from Luhmann [18], describes the mutual dependence of actors' *Orientations*: each actor's choices depend not only on their own perceptions but also on their expectations of how others will perceive, interpret, and act. Meaning and trust, therefore, emerge recursively through continuous mutual adjustment rather than through linear, one-way transmission.

Boyd's **OODA loop** [17] (Figure 2), when interpreted through double contingency, highlights how *Orientation*, encompassing internal and external sensations, feed into *Orient*, shaping *Decide* and *Act*, while multiple feedback loops continually reshape our *Orientation*. Discrepancies between expectations and outcomes manifest as moments of **cognitive dissonance** [24] or, using Luhmann's term, as **Irritations** [23], which trigger continuous **re-*Observations*** and re-*Orientations* to reduce misalignment.

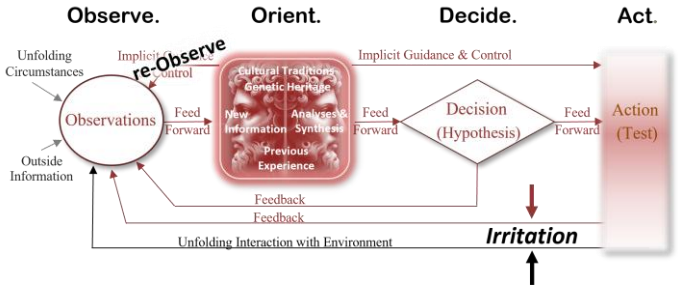


Figure 2: John Boyd's OODA Loop.

A **Janus face** is placed on *Orient* to symbolize the duality of mutual perception and the opacity between the external environment and internal models. As Janus looks inward and outward, each *Orientation* both reveals and conceals meaning. Irritations arise when internal models diverge from external feedback, highlighting the recursive adjustments through which meaning and trust are co-created [23, p. 3]. Extending this to multiple actors invokes Luhmann's concept of double contingency [18] (Figure 3), where each participant's OODA loop adapts continuously to others' actions. Meaning is therefore co-created, not transmitted.

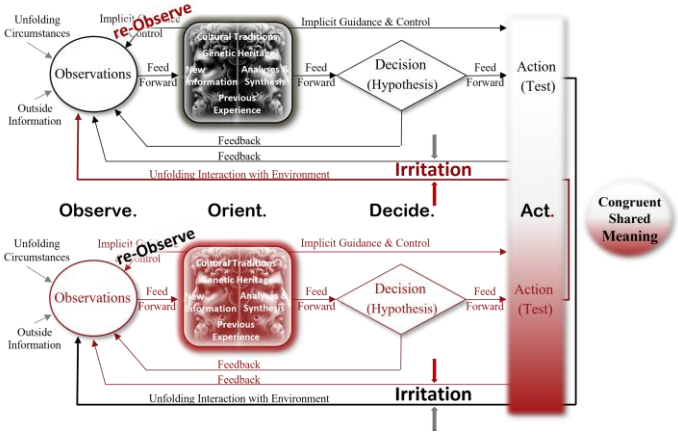


Figure 3: Interconnected OODA loops illustrating double contingency.

[2] In addition to the 2D Harmonization Emergence Model (HEM), a 3D version introduces a third axis to represent hierarchical aspects of harmonization within and between organizations, an important aspect when addressing multidomain operations.

2.3.1 AI's Role In Double Contingency And Wargaming

Within Luhmann's *double contingency* concept, AI functions as a co-adaptive human-machine teaming partner actively participating in the *Co-creation of Meaning* dependent on the emergence of trust. *AlphaSTRIKE* is not designed to replace human judgment but can operate as a ***Cognitive Resonator***, ***Disruptive Irritator***, and ***Meaning-making Enabler***, supporting recursive meaning-making processes.

By amplifying cognitive dissonances, moments when internal models diverge from observed feedback, AI generates productive Irritations that trigger re-observation and re-*Orientation*, enabling earlier detection of blind spots, expanding the *Cognitive Space of Possibilities*, and mitigating operational surprises in adversarial contexts. Through contingent monitoring and attentional control, AI helps detect emerging patterns that might otherwise go unnoticed. This aligns with Laamanen et al. [25], who emphasize AI's role in broadening organizational attention and accelerating feedback loops.

However, *Einheit* remains emergent and fragile. It must be continuously reproduced through *Participatory Sense-Making*. Persistent misalignments or excessive Irritations can overwhelm the team's cognitive and collaborative capacity, causing *Einheit* to deteriorate into fragmentation, a form of social entropy.

Where meaning emerges from *Orientation*, whether intuitively guiding *Action* or requiring deeper analysis through *Decide*, the traditional notions of humans being "in," "on," or "outside" the loop become obsolete. The defining feature is the relational dynamic of human–AI engagement in recursive cycles of Observation, *Orientation*, Action, and re-*Orientation*. Importantly, while AI augments cognition and attention, accountability remains with humans. Trust in AI must be understood operationally, as calibrated confidence in its contributions to the joint cognitive process, not ascribing it moral agency.

Building on these perspectives (Sections 2.1–2.3), we propose that AI in hybrid planning teams can fulfill three core functions, summarized in Table 1.

Table 1: Key AI Functions Enhancing Adaptive Meaning-Making and Collaborative Cognition in Planning

AI Function	Description
Cognitive Resonator	Expands the team's attentional reach by surfacing subtle cues and emerging patterns, broadening the shared *Cognitive Space of Possibilities*.
Disruptive Irritator	Creates productive Irritations by exposing divergences between internal models and external feedback, driving adaptive re-*Orientation*.
Meaning-making Enabler	Supports *Participatory Sense-Making* by aligning team Orientations, enabling the co-creation of congruent shared meaning.

3 AlphaSTRIKE: Decision Support for Iterated exploratory wargaming

This section introduces *AlphaSTRIKE,* a decision support system for iterated, *exploratory Wargaming* of mechanized warfare following the conceptual framework in Section 2. During a game, players can query the AI on how blue and red each ought to proceed. The advice from the AI continuously adapts to the current situation as the game evolves. The user interface makes it easy to ignore, reject or modify the advice from the AI. In addition to the AI-support, the user interface provides some customary forms of automation that may facilitate rapid wargaming: visualization of possible movement and possible fire, including the (stochastic) immediate effect of fire; visualization of lines of sight and fire ranges (Figure 4); the ability to replay events; the ability to rewind (with two different speeds) to an earlier state of the game (to try a different course of action, or simply in order to view in more detail the course of events recorded); bookkeeping (e.g. losses); etc.

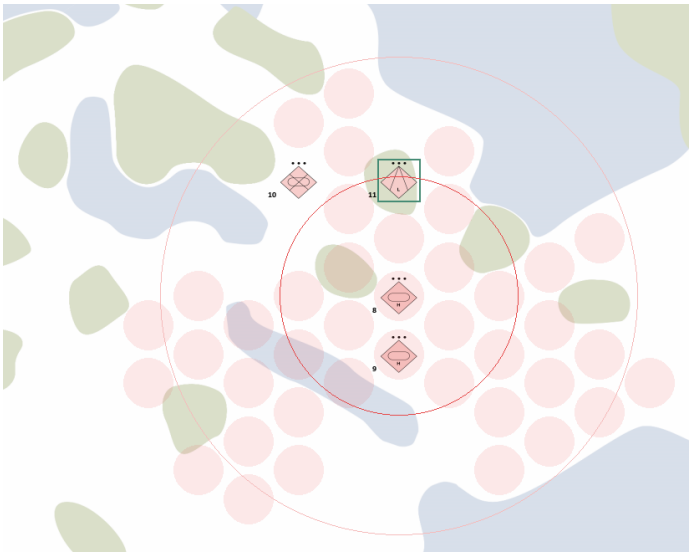


Figure 4: Lines of sight and fire range for a selected unit in AlphaSTRIKE. Inner red circle: effective range. Outer red circle: maximal range. Filled red circles: lines of sight.

The following example illustrates possible user interaction:

A combat group (blue) in a mechanized battalion is tasked with defending a high-value asset against an enemy (red) advancing from the east (Figure 5). The terrain is mostly open with sparse vegetation (white on the map), but occasionally interrupted by lakes (blue), sparse forest (green), and sparsely built-up terrain (gray). For the task, the combat group has three mechanized platoons with armored fighting vehicles (Ajax) and one dismounted infantry platoon with portable anti-tank missile systems (Javelin). The enemy has four mechanized platoons with old tanks (T55) and two mechanized platoons modern tanks (T90) in the area. How should the blue combat group position its units?

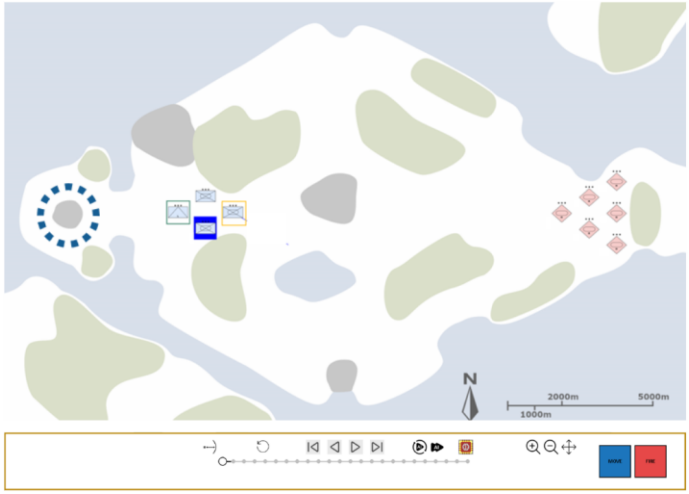


Figure 5: Example scenario in AlphaSTRIKE. The high-value asset is located to the west (dotted circle).

Asking the AI for advice (i.e., pressing the button in the toolbar at the bottom of the screen), the AI suggests that blue should move its units eastwards towards to the approaching enemy, away from the high-value asset, positioning the armored fighting vehicles among the buildings to the east and the infantry platoon in the forest close by (see Figure 6). The sparse buildings and trees provide some cover and may allow for opening fire as red approaches.

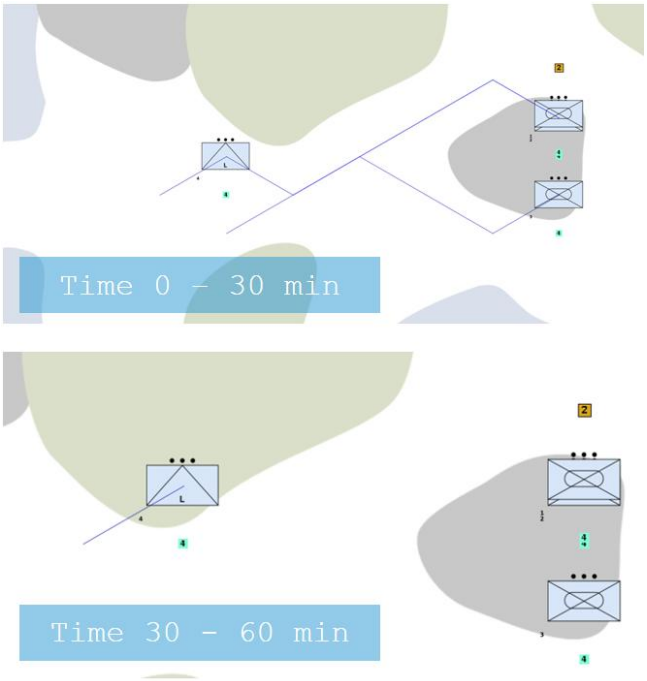


Figure 6: Opening positioning suggested by the AI.

Asking the AI how red should respond, (i.e., pressing), the AI suggests that red should approach more or less straight towards blue positions, with modern tanks ahead and older tanks in the rear (the latter having more limited fire range).

Could a slightly different opening positioning push the enemy to approach the high-value asset from the south instead? We rewind events (), direct the armored fighting vehicles to the northwest instead, and ask the AI () how red should respond. The AI still suggests that red should move towards the center – even though routes along the south are now beyond the reach of blue fire. However, approaching the high-value asset from the south would have left the enemy units more exposed later when they eventually closed in on the high-value asset from a narrow and exposed area.

We rewind (⏮) yet again and position blue units in the south this time – ignoring every aspect of the advised opening. We ask the AI to continue playing on its own, controlling both blue and red units (AI). The AI moves red units along a route in the north, out of reach of blue fire, making it all the way to the buildings and sparse forest in the northeast, just about within fire range to the high-value asset. To eliminate the threat to high-value asset, blue must now attack enemy units who have had the time to take hasty defensive positions among the sparse buildings and trees.

4 Design, implementation and scaling of AlphaZero in AlphaSTRIKE

AlphaSTRIKE generates advice to players by using a general AI for double-sided strategy games, *AlphaZero*, to explore options in a combat simulator that implements an existing professional wargaming rule set, the STRIKE Battlegroup Tactical Wargame. This section describes how *AlphaZero* is adapted to STRIKE and evaluates how the algorithm scales in STRIKE with increasing amounts of compute, providing some concrete evidence of feasibility asked for in the recent study from the Alan Turing Institute (see Section 1).

Section 4.1 gives a brief overview of the underlying STRIKE wargaming rule set. Section 4.2 describes the application of *AlphaZero* to STRIKE. Section 4.3 provides some technical details for readers familiar with deep reinforcement learning. Section 4.4 measures how well the algorithm scales in STRIKE with increasing amounts of computational resources. Section 4.5, finally, summarizes Section 4 and concludes.

4.1 STRIKE Battlegroup Tactical Wargame

The underlying simulator in *AlphaSTRIKE* implements STRIKE Battlegroup Tactical Wargame, a rule set for analytical wargaming of company/battalion/brigade level ground combat. Developed by the Defence Science and Technology Laboratory (Dstl) in the UK, STRIKE has been used in professional wargaming since 2017 [26].

In STRIKE, platoon units move and fire on an open (publicly visible) playing board. The effect of fire is given by stochastic combat results tables with numerous situation dependent modifiers and side conditions.

4.2 AlphaZero for STRIKE

The AI in *AlphaSTRIKE* is a parallelized implementation of *AlphaZero*, a game-theoretically sound AI for double-sided strategy games with open playing boards. During a game in *AlphaSTRIKE*, *AlphaZero* offers advice for red and blue players on how to continue based on the current state of the game and the task given. *AlphaZero* tries to compute the game theory optimal way forward with respect to the task (desired end state) given to the blue side; blue fire and movement is optimized to achieve the task given to blue, while red fire and movement is optimized to prevent the task given to blue. In a sense, *AlphaZero* can be viewed as continuously replanning against "the most dangerous enemy course of action" given the current situation [26].

For the optimization, the task given to blue is translated into a utility (reward, score) that measures the degree to which the desired end state has been met. E.g., the task "defend the high-value asset" above (Figure 5) translates (roughly) into a score that measures the amount of damage the high-value asset suffers; *AlphaZero* optimizes blue fire and movement so as to minimize the damage, while, conversely, optimizing red fire and movement so as to maximize the damage.

In order to provide players with continuous, real-time advice on the game-theory optimal course of action during wargaming, *AlphaZero* must first analyse the scenario being played (i.e., the map, the initial state, and the desired end state for blue) through a form of extensive trial and error, where *AlphaZero* plays blue and red against each other repeatedly. To reduce the time needed for this trial and error, *AlphaSTRIKE* implements the STRIKE rule set as a high-throughput simulation environment tailored for massive parallelization on GPU hardware.

4.3 Parallelization

This section provides technical details on the parallelization for readers familiar with deep reinforcement learning. The training pipeline is composed of two tightly coupled components: (i) a high-throughput simulation environment that implements the STRIKE rule set and (ii) a reinforcement-learning (RL) agent that jointly trains a policy and value network from scratch by repeatedly playing the game against itself following the

AlphaZero algorithm[3]. At every decision point the algorithm performs a batched Monte-Carlo Tree Search (MCTS) guided by the current network parameters; each move therefore requires thousands of forward passes, making environment-step throughput the principal bottleneck.

AlphaSTRIKE implements the underlying simulation in PGX, a JAX-native game-engine that supports fully GPU-accelerated, vectorized environments. Because the environment is fully JIT-composable, *AlphaSTRIKE* can vmap over tens of thousands of independent game states and execute them in parallel on the accelerator. On a single NVIDIA DGX H100 node, *AlphaSTRIKE* routinely sustains 40 000+ simultaneous self-play games without exhausting GPU memory.

The search component in *AlphaZero* is implemented with MCTX, DeepMind's open-source JAX implementation of MCTS, modified slightly to support stochastic games [28]. The homogeneous JAX stack (environment + MCTS + neural network) is compiled once with XLA, after which the entire forward-search/back-propagation loop executes inside the GPU with negligible Python overhead.

4.4 Scalability

To explore *AlphaZero*'s ability to scale to more complex scenarios in STRIKE, we assess to what extent the time needed for *AlphaZero* to learn a given scenario in STRIKE reduces with increased parallelization. To this end, we train a series of *AlphaZero* agents with progressively larger number of parallel game instances and evaluate the *AlphaZero* agents every 100 training steps against a baseline agent. The *AlphaZero* agents receive no reward shaping or other forms of heuristics.

We evaluate *AlphaZero* against two different base line agents. First, we evaluate against a strong rule-based ("heuristic") baseline agent with behaviour rules hand tailored to the specific scenario tested. Figure 7 plots the resulting performance curves. The y-axis indicates the average score (where 1 is maximum) for the *AlphaZero* agent as it plays against the rule-based baseline agent. The x-axis shows the initialization time (training time) given to the *AlphaZero* agent prior to meeting the baseline agent. The different curves represent *AlphaZero* agents that were trained with different degrees of parallelization (batch sizes). As can be seen, increased parallelization accelerates learning markedly. Scaling from 256 to 4096 concurrent self-play games yields a ≈12× reduction in the training time needed to reach a playing strength that consistently beats the baseline.

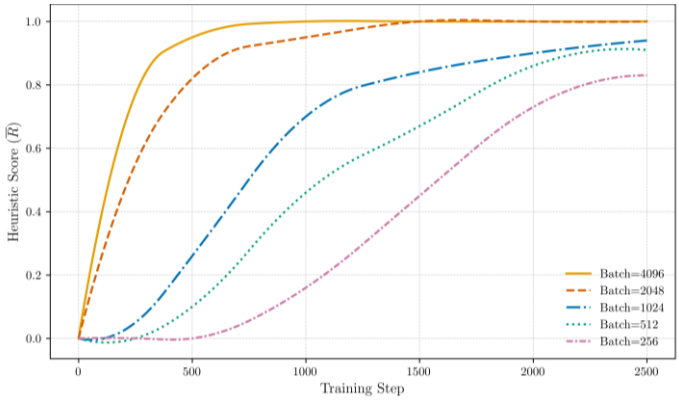


Figure 7: Performance of AlphaZero vs. rule-based baseline agent in STRIKE for different amounts of parallelization.

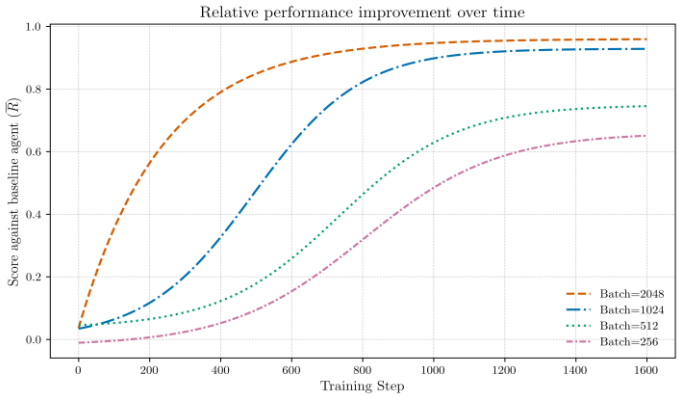


Figure 8: Performance of evolving AlphaZero vs. fixed AlphaZero in STRIKE for different amounts of parallelization.

Secondly, we repeat the evaluation but replace the rule-based baseline with a stronger opponent: a previously trained *AlphaZero* checkpoint. Specifically, we use the checkpoint taken at step 800 from a run with batch size 512. The resulting performance curves are shown in Figure 8. As expected, this checkpoint achieves an average score of 0.5 when evaluated against itself. Again, increased parallelization accelerates adaptation: an agent trained with batch size 2048 already surpasses the checkpoint baseline after roughly 175 steps.

[3] More specifically, the RL agent implements Gumbel AlphaZero [27], a recent more sample efficient variant of AlphaZero, adjusted slightly for stochastic environments [28].

The linear speed-up in Figures 7 and 8 stems from the parallel nature of environment stepping and neural inference, which map cleanly to large matrix operations that GPUs execute efficiently. The compile-once-run-many paradigm of JAX ensures that search overhead remains negligible even as parallelism increases.

All experiments used JAX 0.6, CUDA 12.6 and cuDNN 9. Automatic vectorisation (vmap) and multi-device data parallelism (pmap) lets *AlphaSTRIKE* scale to all eight H100 GPUs on a DGX node.

4.5 Discussion

The recent study from the Alan Turing institute [11] calls into question whether *AlphaZero* and related forms of reinforcement learning that we have seen master classical strategy games with such spectacular success would be able to transfer to professional wargames. The results in this section (Figures 7 and 8) indicate that the adaptation of *AlphaZero* to *STRIKE* scales to battalion level scenarios in *STRIKE* with a relatively modest amount of compute, which provides some "concrete evidence of feasibility" asked for in the study. The end-to-end JAX-based RL pipeline in *AlphaSTRIKE* exploits modern GPU clusters, achieving state-of-the-art training throughput without bespoke CUDA kernels or hand-tuned communication primitives.

To the best of our knowledge, *AlphaSTRIKE* represents the first time that *AlphaZero* – or any other general self-play reinforcement learning agent – is able to play an existing double-sided wargaming rule set that is in actual use in professional wargaming, achieving a playing strength beyond that of strong rule-based agents that have been hand tailored with expert domain knowledge to the specific scenarios tested (Figure 7).

5 Controlled experiment With AlphaSTRIKE

The study from the Alan Turing Institute called into question not only whether *AlphaZero* (and related forms of self-play reinforcement learning) would transfer from classical strategy games to professional wargames, but also whether such a transfer – if indeed possible – would be helpful to wargaming. Even if *AlphaZero* successfully computes a course of action which is optimal given a particular wargaming rule set, the computed course of action need not, it is argued, be feasible in the real combat that is being simulated. Therefore, the advice from *AlphaZero* might not be all that helpful to the players, even if the *AlphaZero* happens to have mastered the rule set being used.

This section reports on a controlled experiment studying the utility of *AlphaZero* for *exploratory Wargaming* of ground combat in a digital environment.

5.1 Experimental design

The experiment is designed to measure the degree to which advice from *AlphaZero* in *AlphaSTRIKE* expands the *Cognitive Space of Possibilities* for senior officers composing battle plans for mechanized combat groups. Experimental subjects are divided into planning teams of two, where some teams are given *AlphaSTRIKE* to explore possible courses of actions, and other teams are given *AlphaSTRIKE* with all *AlphaZero*-based functionality disabled (in effect a digitized version of STRIKE).

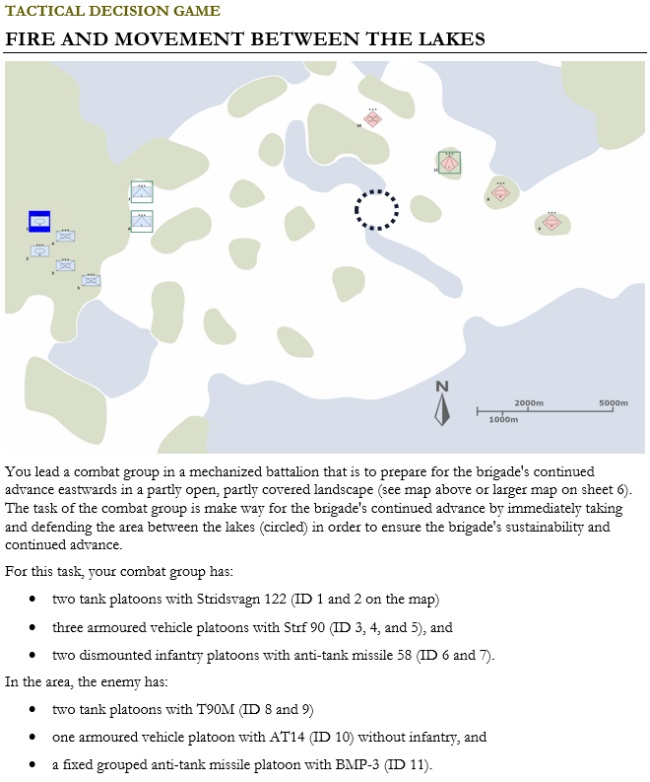
The experiment is run as follows:

1. Each planning team is handed a scenario for a combat group in a mechanized battalion (Figure 9) and asked to sketch an appropriate battle plan for the combat group (see Figure 10 for an example). Teams are given approximately 30 min. for this step.
2. Teams are given a cursory introduction to the user interface in *AlphaSTRIKE* (without any description or mention of either STRIKE or *AlphaZero*). The introduction takes approximately 3 min.
3. To get acquainted with the user interface in *AlphaSTRIKE*, the teams wargame a simpler warm-up scenario (Figure 5) with a different map, different units, and a different mission objective from the scenario in step 1. This step takes approximately 20 min.
4. Teams are asked to revisit their battle plans (from step 1) and wargame them in *AlphaSTRIKE*. Half of the teams are given an *AlphaSTRIKE* with advice from *AlphaZero* disabled. This step is given approximately 20 min.
5. Planning teams are asked if they now would like to revise their original battle plan (from step 1), and if so how. The instructions given emphasize that the goal of any revision should be to make the battle plan more effective in "real combat" (as judged by them), not in the simulated combat just played. This step is given approximately 5 min.
6. Participants answer the survey questions about their experience of planning with *AlphaSTRIKE* (not reported here).

The experiment is designed to minimize bias in favour of *AlphaZero*

1. Only the user interface in *AlphaSTRIKE* is presented in step 2, with only vague references to "an AI" and "a simulator" behind the user interface.
2. Experimental subjects construct a thorough battle plan (in step 1) prior to receiving advice from *AlphaZero* (in step 4).

Arguably, (1) and (2) both stack the odds against subjects accepting advice from *AlphaZero.* (1) makes the system opaque and, as a result, more difficult for subjects to trust and understand, while (2) makes subjects commit to a particular solution before receiving advice from *AlphaZero*.



TACTICAL DECISION GAME

FIRE AND MOVEMENT BETWEEN THE LAKES

You lead a combat group in a mechanized battalion that is to prepare for the brigade's continued advance eastwards in a partly open, partly covered landscape (see map above or larger map on sheet 6). The task of the combat group is make way for the brigade's continued advance by immediately taking and defending the area between the lakes (circled) in order to ensure the brigade's sustainability and continued advance.

For this task, your combat group has:

- two tank platoons with Stridsvagn 122 (ID 1 and 2 on the map)
- three armoured vehicle platoons with Strf 90 (ID 3, 4, and 5), and
- two dismounted infantry platoons with anti-tank missile 58 (ID 6 and 7).

In the area, the enemy has:

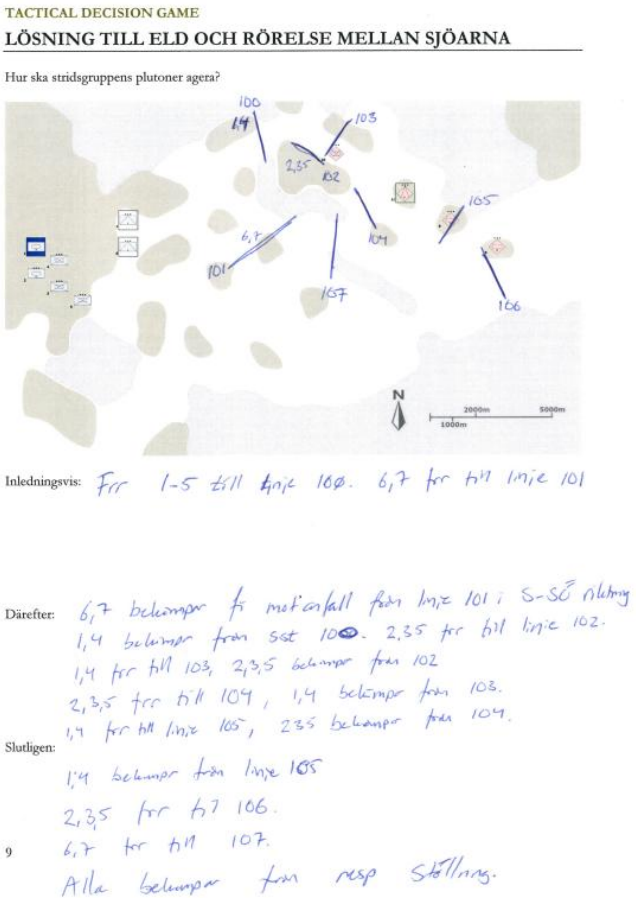
- two tank platoons with T90M (ID 8 and 9)
- one armoured vehicle platoon with AT14 (ID 10) without infantry, and
- a fixed grouped anti-tank missile platoon with BMP-3 (ID 11).

Figure 9: Scenario description handed to participants, first page (translated into English).

5.2 Experiment at Swedish Defence University

In the experiment reported in this paper, the participants consisted of 114 senior officers who performed the experiment during a two-hour class as part of their regular training within the framework of the Joint Advanced Command and Staff Programme at the Swedish Defence University.

The two-hour session had a somewhat broader experimental agenda than the experimental design described in Section 5.1, involving survey questions and other scenarios not reported in this paper.



TACTICAL DECISION GAME

LÖSNING TILL ELD OCH RÖRELSE MELLAN SJÖARNA

Hur ska stridsgruppens plutoner agera?

Inledningsvis: Fr 1-5 till linje 100. 6,7 fr till linje 101

Därefter: 6,7 bekämpar fr motanfall från linje 101; S-SÖ riktning
1,4 bekämpar från sist 100. 2,35 fr till linje 102.
1,4 fr till 103, 2,35 bekämpar från 102
2,35 fr till 104, 1,4 bekämpar från 103.
1,4 fr till linje 105, 2,35 bekämpar från 104.

Slutligen: 1,4 bekämpar från linje 105
2,35 fr 6,7 106.
9 6,7 fr till 107.
Alla bekämpar från resp ställning.

Figure 10: Example of a battle plan from participants at the Swedish Defence University.

5.3 Results

This section presents some results from the experiment at FHS. As explained in Section 5.2, the sessions at the Defence University involved other experiments not reported in this paper; their analysis is left to future work. It is also left to the future to analyse the actual battle plans produced by the participants, e.g. comparing the tactics used to that of *AlphaZero*.

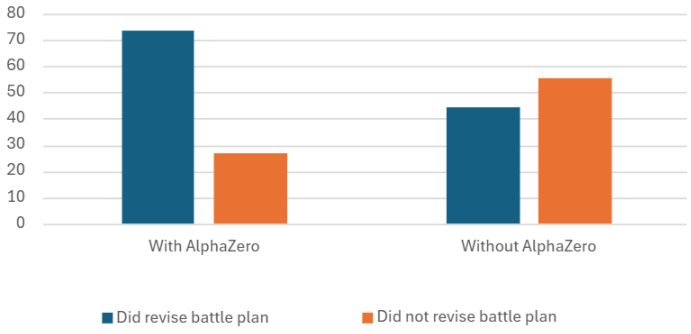


Figure 11: Percentage of revised battle plans. Left: Teams with AlphaZero. Right: Teams without AlphaZero.

Figure 11 shows the percentage of planning teams that revise their battleplans after wargaming in *AlphaSTRIKE*, differentiating between teams wargaming with advice from *AlphaZero* (left) and teams wargaming without advice from *AlphaZero* (right). As can be seen, 73% of teams revise their battle plans when wargaming with *AlphaZero*, while only 44% do so when wargaming without *AlphaZero*.

5.4 Discussion

The recent study from the Alan Turing Institute questioned whether *AlphaZero* (and related forms of AI playing strategy games) could be helpful in professional wargaming, calling for case studies and concrete evidence. This section reported on a controlled experiment studying the utility of *AlphaZero* for professional wargaming. The experiment measured the extent to which the advice given by *AlphaZero* expands the *Cognitive Space of Possibilities* for experienced officers composing battle plans for mechanized combat groups.

Results (Figure 11) from the controlled experiment indicate that wargaming a battleplan with the support of *AlphaZero* leads to 66% more revised battle plans than when wargaming the same battle plan without *AlphaZero*. As already emphasised, the participants were instructed to only add revisions that (according to their own judgement) would improve the effectiveness of the battle plan in "real combat".

While the effect of *AlphaZero* is significant, it is, perhaps, not unreasonable to assume that advice from *AlphaZero* would have an even greater effect without some of the artificial constraints imposed by the experiment. As explained in Section 5.1, the experiment was designed to stack the odds against *AlphaZero* rather than to reflect how *AlphaSTRIKE* ought to be used in practice. Handing a decision support system to a user without any explanations of the underlying assumptions and mechanics fly in the face of best practice for achieving trust. Moreover, having users commit to a particular solution prior to receiving advice from the decision support system likely biases the users against the advice ("commitment bias").

Because of external time constraints, the experimental subjects were given only very limited time for wargaming their battle plan (approximately 20 minutes), which is not conducive to the type of iterative, *exploratory Wargaming AlphaSTRIKE* is intended for. We leave it to future work to explore the effect of giving the planning more time.

6 Conclusion

The paper introduced a decision support system for rapid *exploratory Wargaming* based on a conceptual framework for human-machine teaming in an iterative planning process. In the prototype, a super-human tactician, *AlphaZero*, provides continuous GPS-device-like advice on how blue should move and fire to meet the long-term mission objective for blue, as well as continuous advice on how red should move and fire to prevent the mission objective for blue.

A controlled experiment with 114 senior officers at the Swedish Defence University measured the utility of *AlphaZero* for *exploratory Wargaming* in a digital environment. Results (Figure 11) from the experiment indicate that wargaming a battleplan with the GPS-device-like advice from *AlphaZero* leads to 66% more revisions than wargaming the same battle plan without advice from *AlphaZero*, suggesting that the utility of wargaming in a digital environment may improve with a self-play reinforcement learning agent such as *AlphaZero* in the team of players.

Simulation experiments (Figure 7 and Figure 8) provide some preliminary evidence that the implementation of *AlphaZero* scales to battalion level planning problems in the STRIKE wargaming rule set with no help of heuristics and only a modest amount of compute. To the best of our knowledge, this appears to be the first time in the literature that *AlphaZero* – or any other general self-play reinforcement learning agent – learns to master an existing double-sided ruleset used for professional wargaming.

The simulation experiments and the controlled user experiment at the Swedish Defence University provide some evidence that self-play reinforcement learning agents that we have seen revolutionize tactics in classical strategy games may transfer to wargaming-based planning.

7 Glossary of Key Terms

AlphaZero: A general AI for open, double-sided strategy games that learns to play a given rule set through self-play.

AlphaSTRIKE: An *AlphaZero*-based decision support system designed for rapid *exploratory Wargaming* in the STRIKE Battlegroup Tactical Wargame, exemplifying how AI can support planning as iterative exploration rather than product delivery.

Cognitive Space of Possibilities (CPS): The evolving field of cues, constraints, and options made meaningful through interaction; expands or contracts as dialogue and experience reshape what is seen as possible.

Cognitive Team Schema (CTS): A shared, dynamic *Orientation* shaped by co-created meaning. Develops through reciprocal irritation, enabling diverse perspectives to form adaptive capacity for action.

Co-creation of Meaning: The recursive negotiation of *Orientations* via *participatory sense-making*. Emergent shared understanding integrates differences, producing insights richer than any individual perspective.

Einheit: A resilient sense of unity arising from accumulated *Stimmigkeit*. It integrates differences into a higher-order whole, sustaining collaboration without continuous synchronization.

Exploratory Wargaming: A mode of wargaming that emphasizes iterative exploration, learning, and contextual understanding over formal plan validation. It echoes historical practices such as those of the WATU during WWII.

Harmonization Emergence Model (HEM): Describes how alignment and resilient collaboration emerge from recursive *participatory sense-making*. *Stimmigkeit* and *Einheit* capture shared understanding, from which trust and effective cooperation naturally emerge.

Orientation: A dynamic mental model synthesizing sensory input through culture, experience, and context. Guides action, enables shared meaning, and helps teams navigate planning limits while integrating diverse perspectives.

Participatory Sense-Making (PSM): *Co-creation of Meaning* through reciprocal interaction and mutual adaptation. Small *perturbations* trigger local adjustments that can accumulate into momentary *Stimmigkeit* and eventually *Einheit*.

Problem–Solution Eclipse: The co-evolution of problems and solutions, where planning unfolds as iterative enactment rather than linear problem-solving.

Set-based Approach (SbA): A cognitive planning methodology emphasizing the co-evolution of problems and solutions through iterative exploration. SbA introduces five principles that optimize planning time and reduce premature assumptions. Perturbations and reciprocal *irritation* trigger adaptive sense-making, enabling teams to integrate diverse perspectives into flexible action to navigate uncertainty effectively.

Stimmigkeit: A transient, moment-to-moment alignment of meaning arising dynamically through *participatory sense-making*. Integrates differences as generative sources, forming the foundation for *Einheit*.

References

[1] P. E. Strong, *Wargaming the Atlantic War: Captain Gilbert Roberts and the Wrens of the Western Approaches Tactical*, MORS Wargaming Special Meeting, Oct. 2017.

[2] S. Parkin, *A Game of Birds and Wolves: The Secret Game that Revolutionised the War*, London, UK: Hodder & Stoughton, 2020.

[3] E. Stringer, *Advancing the UK's Analytical Tools to Address Strategic Competition & Deterrence*, King's College Wargaming Lecture Series, 2019.

[4] P. Sabin, *What Strategic Wargaming Can Teach Us*, GIDS Statement 5/21, Hamburg, Germany: German Institute for Defence Studies, 2021.

[5] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, and D. Hassabis, "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play," *Science*, vol. 362, no. 6419, pp. 1140–1144, 2018. [Online]. Available: https://doi.org/10.1126/science.aar6404

[6] N. Tomasev, J. Schrittwieser, D. Silver, and D. Hassabis, "Reimagining Chess with AlphaZero," *Communications of the ACM*, vol. 65, no. 9, pp. 62–68, 2022. [Online]. Available: https://www.researchgate.net/publication/358268603_Reimagining_chess_with_AlphaZero

[7] M. Sadler, N. Regan, and G. Kasparov, *Game Changer: AlphaZero's Groundbreaking Chess Strategies and the Promise of AI*, New in Chess, 2019.

[8] A. Fawzi, A. Balog, D. Silver, and P. Kohli, "Discovering faster matrix multiplication algorithms with reinforcement learning," *Nature*, vol. 610, pp. 47–53, 2022. [Online]. Available: https://www.researchgate.net/publication/364188186

_Discovering_faster_matrix_multiplication_algorithms_with_reinforcement_learning

[9] D. J. Mankowitz, A. Fawzi, I. Kostrikov, et al., "Faster sorting algorithms discovered using deep reinforcement learning," *Nature*, vol. 618, pp. 266–273, 2023. [Online]. Available: https://doi.org/10.1038/s41586-023-06004-9

[10] Y. Chervonyi, A. Guez, T. Hubert, D. Silver, and D. Hassabis, "Gold-medalist performance in solving Olympiad geometry with AlphaGeometry2," *arXiv preprint* arXiv:2501.00001, 2025. [Online]. Available: https://arxiv.org/abs/2501.00001

[11] A. Knack and R. Powell, *Artificial Intelligence in Wargaming: An Evidence-Based Assessment of AI Applications*, CETaS Research Reports, 2023.

[12] P. Sabin, *Simulating War: Studying Conflict through Simulation Games*, London, UK: Bloomsbury Academic, 2014.

[13] Center for Army Lessons Learned, *Military Decision-Making Process*, Fort Leavenworth, KS: U.S. Army, Publication No. 23-07-594, Nov. 2023. [Online]. Available: https://api.army.mil/e2/c/downloads/2023/11/17/f7177a3c/23-07-594-military-decision-making-process-nov-23-public.pdf

[14] J. Ivari and A. Nolan, "A Set-Based Approach: Searching for the Problem-Solution Eclipse," in *Proc. 28th Int. Command and Control Res. and Technol. Symp. (ICCRTS)*, 2023. [Online]. Available: https://www.researchgate.net/publication/376649777_A_Set-based_Approach_searching_for_the_Problem-Solution_Eclipse

[15] P. Thunholm, "Decision making under time pressure: To evaluate or not to evaluate three options before the decision is made?," in *Military Decision Making and Planning: Towards a New Prescriptive Model*, Doctoral dissertation, Stockholm University. Edsbruk, Sweden: Akademitryck, 2003.

[16] J. Ivari and A. Nolan, "Team up for success: Harnessing participatory sense-making with the Harmonization Emergence Model," in *Proc. 28th Int. Command and Control Res. and Technol. Symp. (ICCRTS)*, 2023. [Online]. Available: https://www.researchgate.net/publication/376685554_Team_up_for_success_harnessing_Participatory_Sense-Making_with_the_Harmonization_Emergence_Model

[17] J. R. Boyd, *The Essence of Winning and Losing* [Briefing slides], Defense in the National Interest, 1996. [Online]. Available: https://slightlyeastofnew.com/wp-content/uploads/2010/03/essence_of_winning_losing.pdf

[18] N. Luhmann, *The Reality of Mass Media*, Cambridge, UK: Polity Press, 2007. (Original work published 2000.)

[19] J. R. Boyd, *Organic Design for Command and Control* [Briefing slides], Defense and the National Interest, 1987. [Online]. Available: https://fasttransients.files.wordpress.com/2010/03/organic_design5.pdf

[20] J. J. Gibson, *The Ecological Approach to Visual Perception*, Boston, MA: Houghton Mifflin, 1979.

[21] E. C. Cuffari, E. Di Paolo, and H. De Jaeger, "From participatory sense-making to language: There and back again," *Phenomenology and Cognitive Science*, vol. 14, pp. 1089–1125, 2015. [Online]. Available: https://rdcu.be/dbh2P

[22] A. Juarrero, *Context Changes Everything: How Constraints Create Coherence*, Cambridge, MA: The MIT Press, 2023.

[23] A. Nolan and J. Ivari, "Understanding linguistic diversity, a C2 enabler: Fostering harmonization in team dynamics through constructive-decoherence," in *Proc. 29th Int. Command and Control Res. and Technol. Symp. (ICCRTS*,* 2024. [Online]. Available: https://www.researchgate.net/publication/384324129_Understanding_Linguistic_Diversity_a_C2_Enabler_Fostering_Harmonization_in_Team_Dynamics_through_Constructive-Decoherence

[24] L. Festinger, H. W. Riecken, and S. Schachter, *When Prophecy Fails*. Minneapolis, MN: University of Minnesota Press, 1956.

[25] T. Laamanen, A.-K. Weiser, G. von Krogh, and W. Ocasio, "Artificial intelligence in adaptive strategy creation and implementation: Toward enhanced attentional control in strategy processes," *Long Range Planning*, vol. 58, 2025, Art. no. 102351. [Online]. Available: https://www.researchgate.net/publication/387061397_Artificial_Intelligence_in_Adaptive_Strategy_Creation_and_Implementation_Toward_Enhanced_Algorithmic_Attentional_Control_in_Strategy_Processes

[26] U.S. Navy, *Navy Planning (NWP 5-01)*, Navy Warfare Publication, 2013.

[27] I. Danihelka, A. Guez, T. Graepel, and N. Heess, "Policy improvement by planning with Gumbel," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2022. [Online]. Available: https://openreview.net/pdf?id=bERaNdoegnO

[28] I. Antonoglou, J. Schrittwieser, J. Hubert, D. Silver, and K. Simonyan, "Planning in stochastic environments with a learned model," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2021. [Online]. Available: https://openreview.net/pdf?id=X6D9bAHhBQ1