A challenging case for Artificial Intelligence: evaluation of the benefits of AI-enabled socio-technical concept solutions for Operational-level planning

Patrick Turner
QinetiQ
pjturner1@qinetiq.com

Andrew Leggatt
Trimetis
andrew.leggatt@trimetis.co.uk

Holly Roberts
QinetiQ
hvroberts1@qinetiq.com

Richard Ellis
RKE Consulting
richard.ellis@rkeconsulting.co.uk

Simon Attfield
Trimetis
simon.attfield@trimetis.co.uk

Amy Jones
QinetiQ
amjones1@mail.qinetiq.com

Martin McMillan
Faculty AI
martin.mcmillan@faculty.ai

Rachel Asquith
Faculty AI
rachel.asquith@faculty.ai

Albert Forsey
Faculty AI
albert.forsey@faculty.ai

Abstract

The UK Defence Science and Technology Laboratory's Machine Speed Command and Control (MSC2) project aims to support the transformation of C2 capability through the adoption of artificial intelligence (AI) and exploitation of data. MSC2 has developed several functional concepts for Operational-level planning, each of which translates into one or more socio-technical concept solutions, comprising humans and AI agents performing purposeful planning functions collectively and through interactions, supported by data. This paper describes the formative evaluation of two such concept solutions: the Stakeholder Mapper (a way of developing understanding of multiple actor perspectives and potential responses) and Support to Operational Design (a way of generating Decisive Conditions as part of Operational Design). The multi-disciplinary research team, comprising C2 researchers, ex-military planners and AI developers, employed socio-technical evaluation methods, to enable early, formative evaluation of potential C2 benefits of the concept solutions and thereby shape requirements for AI development, before major development resources have been committed. We summarise the C2 benefits of the concept solutions and draw conclusions for the parent concepts, the evaluation approach and the broader goal of MSC2 to support the transformation of C2 capability.

1 Introduction

The UK Defence Science and Technology Laboratory (Dstl) Machine Speed C2 (MSC2) project (2021-24) aimed to support the transformation of Command and Control (C2) capability – specifically that of Operational-level planning – through the development and validation of the C2 Human Agent Collective (C2 HAC). The C2 HAC is a prototype socio-technical system comprising humans, agents, and data within a tailored architecture. It supports the representation and evaluation of C2 concepts, demonstration of novel ways of working in C2 between humans and AI agents and exploration of data requirements and architectural requirements to enable such ways of working.

Within MSC2, the research task denoted 'Realising C2 Improvements in a Human-Agent Collective' developed several novel concepts for Operational-level planning that are enabled by Artificial Intelligence (AI) and data (see Figure 1 on page 4 for examples). This task also developed concept solutions, based upon functional design and the allocation of functions to planners, AI agents or both. These solutions were brought to life in collaboration with other MSC2 tasks, including the 'Agent Development and Research' task, which produced proof-of-concept agents, and 'Data Collection to enable & Shape Machine Speed C2 in a HAC', which generated data to support agent development.

The 'Realising C2 Improvements' task also designed and orchestrated socio-technical evaluation activities to create the environment within which concept solutions were put to use and evidence for their C2 benefits was developed. Such evidence is shaping ongoing development activity and thereby supporting the transformation of Operational-level planning capability.

This paper begins by describing key challenges faced by Operational-level planning, setting out the drivers for our research and the range of research perspectives we adopted. We then summarising the approach to concept development and evaluation, and thereafter introduce two concepts – Auto-Piggery and the Systems Approach – and the associated concept solutions that we brought to life and put to use within socio-technical evaluation activities. We summarise the C2 benefits of the concept solutions, draw conclusions from the research and look ahead to next steps.

2 Operational-level planning challenges

The Operational Level is defined in UK military doctrine as "the level at which major operations are planned and effects are created to achieve strategic objectives" [1]. A key purpose of Operational-level Headquarters (HQs) is to plan operations and thereafter manage the execution of such operations. Practically, the management of execution also involves planning, albeit focused upon shorter timeframes and typically adapting plans previously generated. Hence Operational-level planning is not confined to the pre-operations phase and is conducted in the context of change in the operating environment.

What distinguishes Operational-level planning from planning at the Strategic or Tactical levels is that it serves to translate strategic objectives into tactical activity and thereby produces a framework for the execution and orchestration of such activities in support of the attainment of such objectives. The Operational level is also characterised by a high degree of complexity and change, both in the Operating Environment itself and across the so-called Complex of Actors who act to bring about change in that environment. Whilst there is typically less time-pressure on Operational-level planning as compared with Tactical-level planning, Operational-level commanders and planners must grapple with enormous challenges, for example:

- The impact of complexity and change within a contested environment. This includes the manifestation of such complexity in terms of uncertainty, equivocality and ambiguity in information, and the manifestation of change in terms of unforeseen events and nonlinear shifts in operational dynamics.
- Plurality of intent, motivations, allegiances and perspectives within the Complex of Actors.

3 Research drivers and perspectives

Our research into realising C2 improvements was essentially about the development and evaluation of new ways of conducting Operational-level planning, supported by novel capabilities. There were two main drivers for this research. The first driver was a desire to explore the potential benefits of Artificial Intelligence (AI) in Operational-level planning (and C2), realising the promise of machine-speed processing, greater access to data and the potential to address human biases. The second was to address the challenge of complexity, as described above.

The research was informed by a range of perspectives:

- C2 is a socio-technical system.
- Complexity demands new approaches to Operational-level planning.
- Operational-level planning is a 'challenging case' for AI.
- The C2 HAC is a design pattern for future C2 capability.
- Operational-level planning should strive for utility rather than accuracy.
- AI provides a supporting role in human-agent interaction within Operational-level planning.

3.1 C2 is a socio-technical system

We consider C2 as a socio-technical system, comprising data-rich interactions between people, processes, structures and technology. This message is consistent with the UK Development, Concepts and Doctrine Centre (DCDC) Joint Concept Note (JCN) 2/17, The Future of Command and Control [2]. This perspective applies equally to planning, which is a function of C2, and necessarily applies at every level, including the Operational level.

3.2 Complexity demands new approaches to Operational-level planning

JCN 2/17 [2] made the case that C2 must change to cope with complexity in the operating environment. This demands new approaches to Operational-level planning across C2 as a socio-technical system, including new ways of thinking and organising. Whilst AI can be part of solutions, we need concepts that embody the required change. We therefore approached the research from this perspective and did not simply identify change with the introduction of AI to current Operational-level planning activities.

3.3 Operational-level planning is a 'challenging case' for AI

Operational-level planning is a challenging case for AI because it involves macrocognitive functions such as sensemaking, design, planning and decision-making in real-world settings [3]. Such functions rely heavily upon human expertise, tacit knowledge, judgment and creativity, are far from procedural, are difficult to codify and are therefore difficult for AI to support. The effectiveness of such functions is critical to operational outcomes, however, and improvements in their effectiveness could offer decisive advantage. AI may be a source of such improvements. We therefore believe that AI research *should* be focused upon Operational-level planning as a work domain. This level of ambition requires AI research to develop advanced technologies and thereby realise the promise of future AI.

3.4 The C2 HAC is a design pattern for future C2 capability

The C2 HAC is based upon the idea that C2 practitioners and AI agents work purposefully together as part of a collective. Whilst current Operational-level HQs exploit technology and data, they are essentially human collectives (teams of teams) that exploit information, knowledge, expertise, judgment and creativity to carry out macrocognitive functions (e.g. sensemaking, design, planning, decision-making). The C2 HAC offers a socio-technical 'design pattern' for future HQs, within which AI and data play a more prominent and integrated role.

3.5 Operational-level planning should strive for utility rather than accuracy

Operational-level planning must grapple with complexity and there are no 'correct' or 'incorrect' ways of formulating complex problems or solutions. The macrocognitive functions of C2 (e.g. sensemaking, design and planning) translate into a range of activities, from the 'routine' to 'non-routine', and the latter type is more common and typifies the 'challenging case' that our research was focused upon. Non-routine activities do not deal in absolute truths; instead, they develop understanding, designs and plans that are 'better' or 'worse', and this can only be determined in retrospect. Accuracy is, therefore, generally not a useful measure of merit for Operational-level planning. By extension, it is also unreasonable to expect AI to produce accurate outputs in support of the same activities. What is more useful is that Operational-level planning develops plausible insight and foresight, and logically sound and compelling designs and plans, offering utility to the commander in grappling with complex circumstances. If AI can stimulate human thought and improve planning outputs from the human-agent collective, then it offers benefit.

3.6 AI provides a supporting role in human-agent interaction within Operational-level planning

Our research was essentially starting from the beginning, with no human-agent practice in Operational-level planning to build upon. Further, given that most Operational-level planning activities grapple with complexity, are non-routine and rely heavily upon human expertise, judgment and creativity, we decided upon a common design philosophy that placed human planners in a lead role, with AI agents in a supporting role. Further, we determined that any outputs from agents must be explained to planners and must not overwhelm them cognitively.

4 Concept Development & Evaluation

Our core research approach was concept development and evaluation. We define a concept for Operational-level planning as a way of carrying out some aspect of planning activity. We developed new concepts to both address future challenges – not least the need to grapple with complexity – and exploit the potential in AI and data.

We generated sixty-five concepts, each of which implied a degree of change for Operational-level planning, ranging from the adaptation of current practice (grounded in current doctrine) to transformation of practice for future C2 contexts. The expected degree of benefit to C2 of these concepts varied, broadly, with the implied degree of change. Eight concepts were down-selected through dialogue with Dstl and UK Defence Futures[1]. The concepts are depicted in Figure 1, below.

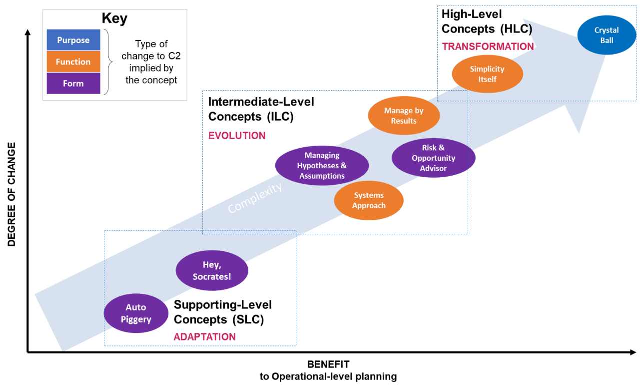

Figure 1: Map of down-selected concepts

We then developed concepts into concept solutions. A concept solution is a functional design and design philosophy for an implementation of the concept. It shapes which activities are carried out, whether they are carried out by humans, AI agents or both (through interaction) and what data is exploited. The design philosophy essentially embodies non-functional requirements. As described above, all concept solutions were based upon a common philosophy.

Each concept was translated into multiple concept solutions. This enabled both alternative solutions and partitioned solutions to be developed. Partitioned solutions were developed wherever concepts were too broad to be implemented in full within the research task and required dividing up sequentially. Partitioning was done by planning activity, thereby ensuring that each concept solution had a defined purpose and well-bounded functionality. 'Support to Operations Design' is an example of a partitioned concept solution (for the Systems Approach concept).

Each concept solution was then brought to life through the design, development and integration of:

- A proof-of-concept (PoC) functional AI agent.
- An interaction AI agent to broker/manage interactions between humans and PoC functional agents.
- Activity workflows for planners and agents, as appropriate.
- Human-agent interactions, supported by wireframes and human-machine interfaces (HMI).

Integration of the above prototype elements rendered each concept solution as a socio-technical system, albeit one more tightly bounded, less mature and less shaped by real-world constraints than any practical instantiation of Operational-level planning.

Design, development and integration activities were iterated with formative evaluation activities. This ensured that learning shaped the form of the concept solutions during the research, to enhance their benefit to Operational-level planning. We employed a range of socio-technical evaluation methods, appropriate to the low maturity level of the concept solutions. Whilst these methods are described and compared at length in Leggatt et al [4], an important way of distinguishing them is whether they accommodate 'functioning code' or earlier representations of agent functionality and human-agent interactions, either described and dissected through dialogue (design briefs) or mimicked by humans (Wizard of Oz).

This paper focuses upon evaluation activities that featured proof-of-concept agents themselves, because these activities generated the highest level of evidence available from the research and thereby enabled us to make more specific statements about the C2 benefits of the concept solutions. These evaluation activities were conducted in laboratory settings, with two ex-Operational-level planners (members of the research team), a representative scenario (the Allied invasion of Sicily in 1943, Operation HUSKY) and representative data (drawn from both documents about Operation HUSKY and understanding/planning products developed by the team[2]). In each evaluation, the planners worked through specified planning activities in the Operation HUSKY scenario and were supported through interaction with the PoC AI agent, as mediated by both the interaction agent and the PoC agent HMI.

The aim of each evaluation was to develop evidence for the C2 benefits of the concept solution. Data collection was framed by a set of research questions that elicited the character of these benefits and their sources. This enabled the research team, through observation and structured dialogue with the planners, to develop a benefits map for each concept solution and reason about whether the benefits originated from agent functionality, human-agent interactions and/or human activity.

Over the following pages, we introduce the concepts and

[1] Previously known as the UK Development, Concepts and Doctrine Centre (DCDC).

[2] The products were developed during a Campaign Planning week for Operation HUSKY, run by the MSC2 Data Collection task.

concept solutions that were developed and evaluated within the research:

- The Auto-Piggery concept and the concept solution taken forward to evaluation: Stakeholder Mapper
- The Systems Approach concept and the concept solution taken forward to evaluation: Support to Operational Design

5 Concept: Auto-Piggery

The name of this concept comes from a systems thinking technique that produces a so-called 'pig diagram'. Figure 2, below, depicts the canonical pig diagram, which features a pig and how four people, or 'stakeholders' – see it.

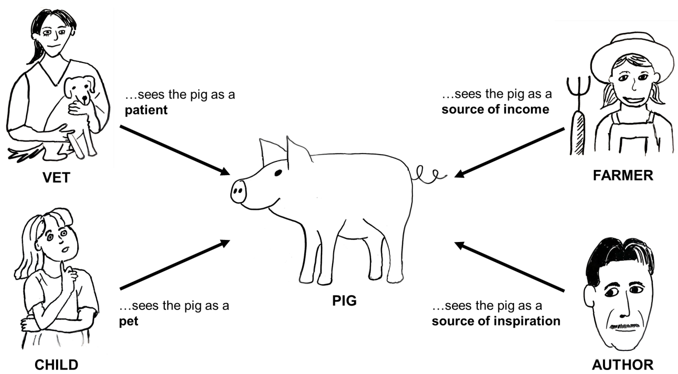

Figure 2: Pig diagram

The child sees the pig as a pet, the vet sees it as a patient, the farmer sees it as a source of income and the author sees the pig as a source of inspiration. This type of thinking is useful within Operational-level planning because planners face the challenge of developing understanding of stakeholders in the Operating Environment. They need this understanding to design Operations that take into consideration a multitude of stakeholder interests in a range of issues. Failure to do so may lead to Operational surprise when a stakeholder reacts in an unforeseen manner.

Typically, planners need to develop such understanding early in the planning process and update it as they go. It is particularly useful within Mission Analysis, in support of counterfactual analysis, i.e. reasoning about what interests stakeholders *might* have in future events or phenomena in the operating environment, and how they *might* react. We denote these events or phenomena as 'critical issues'. The planners on the research team indicated that understanding likely stakeholder reactions to potential blue intervention options is of huge value, as it helps to frame operations that serve to transform the complex operating environment towards the desired End-State. Other potentially relevant types of critical issue include key 'plays' by other stakeholders and key events e.g. elections.

The concept of Auto-Piggery involves interaction between planners and an AI agent – to aid planners' understanding of stakeholders' interests in, and responses to, critical issues.

6 concept solution: Stakeholder Mapper

The Stakeholder Mapper was one of three alternative concept solutions developed for Auto-Piggery. This concept solution was defined as:

A socio-technical system in the Operational-level planning domain to develop understanding of multiple stakeholders' interests in, and potential reactions to, a critical issue, through the interactions between planners and an AI agent, and based upon data derived from planners' dialogue, information and intelligence (i2) products and a Large Language Model, in order to both contribute to planners' understanding of the operating environment and their ability to conceptualise a response that exploits this multi-perspective understanding.

The role of the Stakeholder Mapper PoC functional agent was to identify stakeholders and thereby analyse and visualise their interests in critical issues, based upon external data sources and HQ understanding products. Table 1 describes the PoC agent functionality.

Stakeholder Mapper PoC Agent functionality
Based upon a choice of a closed-system, locally hosted Large Language Model (LLM) (Falcon 40B) or proprietary LLM accessible via an API (GPT-3.5)[3].
Provided with a small corpus of relevant open-source information about Operation HUSKY and a transcript of dialogue from the Mission Analysis stage of the Campaign Planning week, conducted by the 'Data Collection' task.
Includes a graphical / text-based Human-Machine Interface (HMI) through which a critical issue is entered and outputs are presented.
Exploits Named Entity Recognition to identify stakeholders with an interest in the Critical Issue. Such interests are produced using the LLM with Retrieval Augmented Generation (RAG) using the information described above. This both reduces hallucinations and controls what information is used to form responses[4].

[3] Only the Falcon 40B variant was evaluated (see Section 9).
[4] This recognises that pre-trained LLMs can become out-of-date quickly, especially in changing operational circumstances.

Isolates the specific parts of the source documentation that were used as context for producing stakeholder interests[5].

Table 1: Stakeholder Mapper PoC Agent functionality. © Faculty AI

To access PoC agent output, the planners were presented with a visualisation, based upon the LLM and corpus of information, of relevant stakeholders who had an interest in the critical issue (see Figure 3, below).

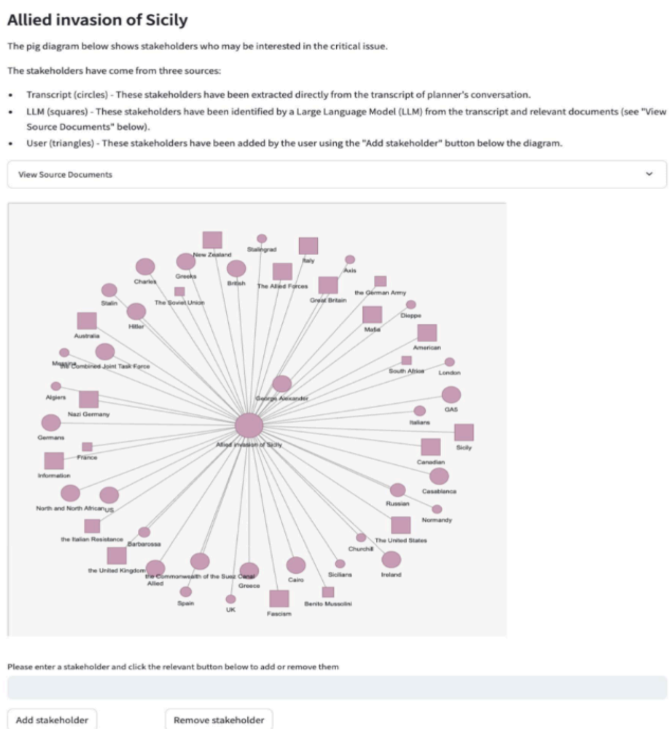

Figure 3: Example graphical component of the Stakeholder Mapper PoC agent HMI. © Faculty AI

Planners could also scroll to view a text-based description of each stakeholder's perspective on the critical issue (see Figure 4) together with a summary of the source document text that contributed to this description.

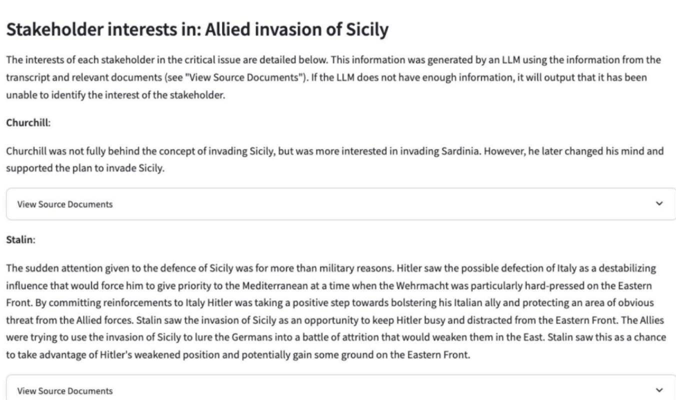

Figure 4: Example textual component of the Stakeholder Mapper PoC agent HMI. © Faculty AI

There was a second agent, denoted Tinman[6], which essentially managed interactions between planners and the Stakeholder Mapper PoC agent. As this agent was not fully functional at the time of the evaluation (January 2024), some functionality was mimicked by members of the research team, denoted 'wizards' [4].

The planning context for the socio-technical concept solution was Mission Analysis (and generally those Mission Analysis activities through which planners make sense of the operating environment, through the lens of the mission.) The workflow for the concept solution was:

- Planners' default activity was sensemaking in the context of the mission, conducted through human-human dialogue (part of Mission Analysis).
- The Stakeholder Mapper PoC agent was prompted to generate a pig diagram. This human-agent interaction was mediated in two alternative ways during the evaluation: 1. Initiated by planners verbally, at a time of their choosing, interpreted and confirmed by Tinman[7] via Microsoft Teams; 2. Critical issue suggested to planners by Tinman[8] via Microsoft Teams, based upon analysis of planner dialogue, with planners accepting, refining or rejecting this suggestion.
- The critical issue was passed to the Stakeholder Mapper PoC agent HMI by Tinman[9].
- Planners were alerted by Tinman when output from the Stakeholder Mapper PoC agent was available. (This took 10-15 minutes, therefore planners returned to their default activity between prompting and receiving the alert.)
- Planners chose when, how and for how long to interact with that output using the Stakeholder Mapper PoC agent HMI (e.g. to inspect the visualised pig diagram, to analyse stakeholder interests from the text-based output or to view relevant sections of source documents upon which agent outputs were based).

[5] Thereby aiding explainability.

[6] Developed by another MSC2 task, 'Human Agent Language and Interaction Styles' (HALIS).

[7] Wizard-enabled.

[8] Wizard-enabled.

[9] Wizard-enabled. Practically, the wizards entered the critical issue through the Stakeholder Mapper PoC agent HMI.

- Planners decided when to complete Mission Analysis and which sensemaking outputs to take forward as part of this planning step, i.e. there was no dependency upon the agent to complete the activity and planners were free to exploit agent output in any way they wanted in finalising outputs from the activity.

7 Concept: The Systems Approach

A key challenge in Operational-level planning is that commanders and planners need to develop deep understanding of the operating environment and design effective Operations. The type of understanding being discussed here is qualitatively different from that generated through Auto-Piggery because it concerns the essence of the operational problem, not individual stakeholders' interests. As argued in Firth et al [5], Operational-level planning based upon systems thinking is more appropriate given the inherent complexity in the operating environment. This approach is founded on a way of thinking about that environment as a complex adaptive system (CAS). Concepts of Operational Design take on a systemic character as compared with their current use within doctrine, e.g. NATO AJP-5 [6]. For example:

- **Centre of Gravity** (CoG). Rather than developing multiple Centres of Gravity (one per key stakeholder), a systemic Centre of Gravity is developed that describes the crux of the operational problem and provides a focal point for Operational Design. A systemic Centre of Gravity will typically describe some system-level idea, pattern or key relationship that, if supported or opposed, with enable the End State to be achieved – and thereby intended systemic change to be effected. CoG Analysis – the development of candidate CoGs and the identification of their Critical Capabilities (CCs), Critical Requirements (CRs) and Critical Vulnerabilities (CVs) – is used in both approaches.
- **Decisive Condition** (DC). This is defined in NATO AJP-5 [6] as "a combination of circumstances, effects, or a specific key event, critical factor, or function that, when realized, allows commanders to gain a marked advantage over an opponent or contribute materially to achieving an operational objective." Within a systems-based approach, an alternative definition, from Firth et al [7] is "an effect that constitutes a fundamental change in the behaviour or physical state (character) of a system. It is created by a number, or series of actions and events, deliberate or otherwise, and usually transient in nature (otherwise it risks becoming expressed as a finite objective). The aggregate effect of the framework of Decisive Conditions creates the systemic condition described by the End-State." Whilst the AJP-5 definition links DCs to opponents and the achievement of operational objectives, the systems-based definition links DCs to systemic change and eschews operational objectives because their 'fixed point' nature is less appropriate for describing intended change within a complex system. The generation of DCs exploits CoG Analysis, especially the CVs of the selected CoG(s) and hence the character of the DCs generated is dependent upon the type of CoG(s) developed – stakeholder-centric or systemic. As a consequence, within a systems approach, DCs are focused upon systemic effects rather than effects for specific stakeholders.

The Systems Approach concept is grounded in the systemic thinking outlined above and therefore represents an evolution from AJP-5. It includes human-agent activities that generate a deep understanding of the operating environment and a systemic approach to Operational Design. With respect to Operational Level Planning, this concept maps to Mission Analysis and, specifically, to the following constituent activities [6]:

- Centre of Gravity Analysis (generation and analysis of the Centre of Gravity, including the identification of Critical Requirements, Critical Capabilities and Critical Vulnerabilities).
- Operational Design (generation and analysis of Decisive Conditions, development of the Operational Design that links DCs to the Operational CoG and thereby the Operational End-State, and generation, analysis and linking of Operational Effects).

8 Concept solution: Support to Operational Design

Support to Operational Design was the third of four concept solutions developed for the Systems Approach concept. These four solutions were partitioned and sequenced, to ensure coverage of the concept:

- System Appreciation. (Sensemaking about elements, relationships, trends etc. in the operating environment.)

- System Understanding & Support to CoG Analysis. (Sensemaking to understand the essence of the essence of the operational problem.)
- Support to Operational Design (Initiating the Operational Design.)
- Support to Effects Development. (Elaborating the Operational Design.)

Whilst the first two concept solutions deepen understanding of the operating environment, Support to Operational Design maps to that part of the planning process where planners pivot from sensemaking to design. This is cognitively challenging as it demands creativity. It was therefore of great interest to the research. How can AI support such an activity?

The Support to Operational Design concept solution is defined as:

A socio-technical system in the Operational-level planning domain to develop an Operational Design, including the generation, description, analysis and representation of a set of Decisive Conditions that address the Operational Centre of Gravity and realise the Operational End-State. This includes the development and representation of themes or other relationship that serve to convey how DCs combine to form the Operational Design. These functions are achieved through interactions between planners and an AI agent, supported by data derived from planners' dialogue and previous planning products, underlying models of 'good practice' in Operational Design (e.g. guidelines) and external data sources that may provide conceptual ideas and/or challenge to DC generation. The narrower purpose of this socio-technical system is to provide a conceptual framework for subsequent planning activities; the broader purpose is to provide a framework for Operations.

The role of the Support to Operational Design PoC functional agent was to generate Decisive Conditions for planners' consideration, based upon external data and HQ understanding and planning products. Table 2 describes the agent functionality and Figure 5 illustrates its Human-Machine Interface (HMI).

Support to Operational Design PoC Agent functionality
Based upon a closed-system pre-trained Large Language Model (LLM).
Provided with a small corpus of relevant open-source information about Operation HUSKY and planning data generated during the Operation HUSKY Campaign Planning week (Operational End-State, Operational Centre of Gravity and associated CoG Analysis – including CCs, CRs and CVs), conducted as part of the 'Data Collection' task.
Includes a text-based Human-Machine Interface (HMI) through which outputs are presented.
Exploits Retrieval Augmented Generation (RAG) to generate a long list of proposals for Decisive Conditions, including title and description.
Isolates the specific parts of the source documentation and planning data that were used as context for producing DCs.
Analyses DCs it has produced by applying guidelines developed through expert elicitation conducted by the 'Realising C2 Improvements in a Human-Agent Collective' team [7].

Table 2: Support to Operational Design PoC Agent functionality. © Faculty AI

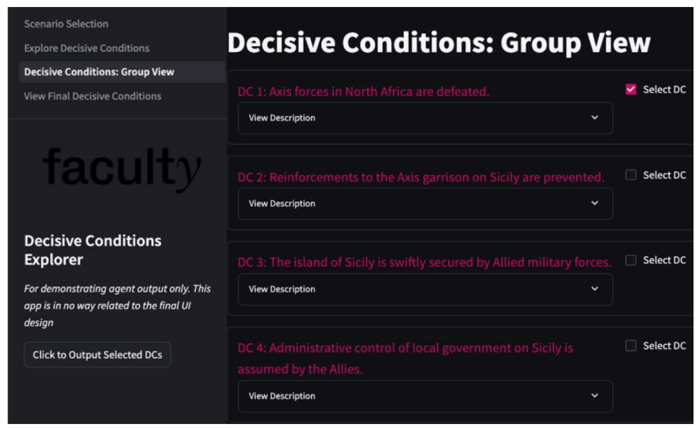

Figure 5: example textual component of the Support to Operational Design PoC agent HMI. © Faculty AI

With respect to HMI, planners were presented with candidate DCs, with both titles and descriptions available. Further, for each DC, planners could also inspect a summary of the source document text that contributed to this description and view the results of the DC analysis tool.

As with Auto-piggery, there was a second agent, denoted Tinman[10], which essentially managed interactions between planners and the Support to Operational Design POC agent. As Tinman was not fully functional at the time of the evaluation (March 2024), some functionality was mimicked by members of the research team, denoted 'wizards' [4].

The planning context for the socio-technical concept solution was Mission Analysis (specifically the generation of DCs, as the first part of the activity to initiate the

[10] Developed by another MSC2 task, Human Agent Language and Interaction Styles (HALIS).

Operational Design). The workflow for the concept solution was:

- Planners' default activity was the generation of DCs, conducted through human-human dialogue.
- The Support to Operational Design POC agent was prompted to generate DCs.
- When planners wanted AI support to aid DC generation, they prompted Tinman.
- Tinman initiated the Support to Operational Design POC agent.
- Planners were alerted by Tinman when output from the Support to Operational Design PoC agent was available.
- Planners chose when, how and for how long to interact with that output using the Support to Operational Design PoC agent HMI (e.g. to inspect the DC descriptions, view relevant sections of source documents upon which agent outputs were based or to view DC analysis).
- Planners decided when to complete DC generation and which DCs to take forward, i.e. there was no dependency upon the agent to complete the activity and planners were free to exploit agent output in any way they wanted in finalising outputs from the activity.

9 Evaluations

9.1 Purpose and approach

Each of the concept solutions (Stakeholder Mapper and Support to Operational Design) was evaluated during the course of the research.

The purpose of each evaluation was to develop qualitative evidence for the C2 benefits of the respective concept solution. The level of evidence for such benefits was expected to be low, for the following reasons:

- Each concept solution was brought to life and put to use for the first time during the course of the research – and prior evidence for its benefits did not exist.
- The evaluation settings, although sufficiently realistic to create the conditions for Operational-level planning activity to be undertaken, did not impose all of the real-world constraints that would increase ecological validity (e.g. HQ structures, time pressure and, perhaps most importantly, an Operational-level scenario that had previously not been seen by the planners). We note that these settings were entirely appropriate given the lack of prior evidence.
- The evaluation method did not impose tight controls on human activity or human-agent interaction. Structure was provided through pre-defined workflows, which were adhered to, but planning activity was – on occasion – paused to clarify aspects of Tinman-mediated human-agent interaction, PoC agent output or human interaction with the PoC agent HMI.

9.2 Data Collection

Data were collected through both observation of planning activity and structured dialogue with the planners, post-activity. Data collection was shaped by a set of research questions that were grounded in common set of categories

- C2 benefit
- Human activity
- Human-agent interaction.
- Agent functionality.

These research questions are summarized below. The C2 benefit questions were specific to the concept solution (and are indicated as such) whereas the questions against the other categories were concept solution-agnostic.

9.2.1 C2 benefit

- Stakeholder Mapper: In what ways does the Stakeholder Mapper concept solution develop understanding of stakeholders' interests in critical issues?
- Support to Operational Design: In what ways does the Support to Operational Design concept solution generate Decisive Conditions that have utility to Operational-level planning?
- Support to Operational Design: Given that DC generation, as a human activity, existed before the PoC agent was developed, how did the effectiveness of the concept solution change when the PoC agent was introduced?

9.2.2 Human activity

- In what ways did the performance of human activities adhere to the workflow in the concept solution?
- What were the potential reasons for this performance?

9.2.3 Human-agent interaction

- In what ways did the performance of human-agent interaction meet the requirements of the concept solution?
- What were the potential reasons for this performance?

9.2.4 Agent functionality

- In what ways did the performance of the agent, including the quality of its outputs, meet the requirements of the concept solution?
- What were the potential reasons for this performance?

10 Evidence for C2 benefits

10.1 Stakeholder Mapper

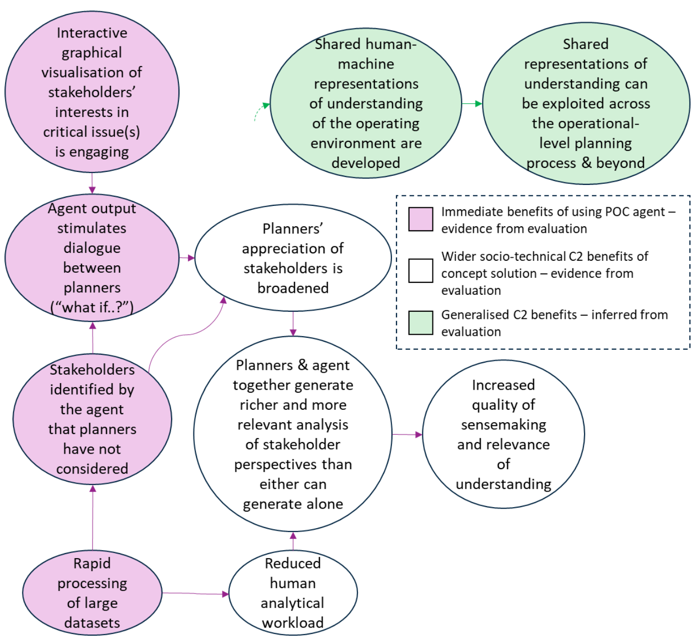

Figure 6: Stakeholder Mapper Benefits Map

Figure 6 shows a linked set of benefits associated with the Stakeholder Mapper concept solution, identified through evaluation. There are three levels of benefit:

- Immediate benefits to planners of using the PoC agent (pink ellipses). These are based directly upon evaluation evidence.
- Wider benefits of the concept solution, i.e. benefits at the level of the socio-technical system itself (white ellipses). These are also based directly upon evaluation evidence.
- Generalised benefits of the concept solution (green ellipses). These are inferred from the evaluation evidence.

The PoC agent offered rapid processing of large datasets, which revealed stakeholders that the planners had not considered or had no awareness of. Planners found the interactive graphical visualization of stakeholders' interests in critical issue(s) to be engaging and this, coupled with the fact that the visualisation included information about new stakeholders, both stimulated and mediated a great deal of dialogue between them, encouraging counterfactual thinking ("what if…?" thinking) about stakeholders' potential actions or reactions in the context of the critical issue(s).

PoC agent processing reduced the analytical workload on the planners because they did not have to search source documents to identify new stakeholders. This freed up more time to focus on sensemaking, including counterfactual thinking about more stakeholders. This broadened planners' appreciation of stakeholders in the operating environment. Consequently, planners and the PoC agent together generated richer and more relevant analysis of stakeholder perspectives than either could generate alone. This led to another benefit: increased quality of sensemaking and relevance of understanding.

With respect to generalized and inferred benefits, we argue that Auto-Piggery is a type of concept through which planners and agents generate and develop a shared (human-agent) representation of understanding of the operating environment. Auto-Piggery is concerned with stakeholders, perspectives and critical issues. More general representations could include other artefacts of sensemaking, such as drivers and trends. We reason that such shared representations are reusable, human-interpretable and machine-readable data, which could be exploited in any other activity that is concerned with either developing or exploiting understanding of the same operating environment, elsewhere in the Operational-level planning process, or beyond. The development of Operational Effects and Operational Actions clearly benefit from such understanding as they invariably need to be directed towards supporting or opposing stakeholder interests.

10.2 Support to Operational Design

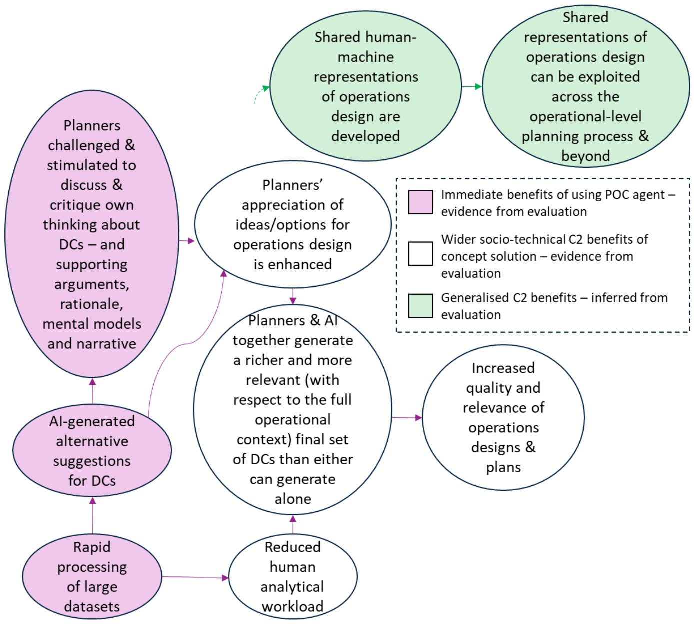

Figure 7: Support to Operational Design Benefits Map

Figure 7 shows a linked set of benefits associated with the Support to Operational Design concept solution, identified through evaluation. As with Figure 6, three levels of benefit are shown. There is also a similar pattern of benefits as for the Stakeholder Mapper.

The PoC agent offered rapid processing of large datasets, which supported the generation of DCs that provided alternative suggestions to the planners' own thinking. Whilst planners questioned the relevance and quality of *some* of the agent output, they also remarked that it included novelty and this novelty was useful. It was also clear that, in using the agent, the planners were challenged and stimulated to discuss and critique their own thinking about DCs, including supporting arguments, rationale, mental models and narrative.

As with the Stakeholder Mapper, the Support to Operational Design PoC agent reduced the analytical workload on planners because it generated outputs based upon previous planning products (Operational End-State and Operational Centre of Gravity) for planners' consideration, rather than planners *necessarily* having to do that themselves. This, however, is a conditional benefit because:

- The PoC agent did not generate DCs in exactly the same way as the planners do[11].
- This discrepancy manifested in *some* of the agent-generated DCs appearing – to the planners – to be of a lower quality and less relevant.
- Planners consequently spent additional time making sense of the agent-generated DCs, reasoning about their origins (aided by agent explainability) and selecting useful output to use as the basis for challenging their thinking.

Planners' appreciation of ideas for Operational Design (i.e. DCs) was enhanced through their interactions with the PoC agent. The central benefit of the Support to Operational Design concept solution was that planners & AI together generate a richer and more relevant a richer and more relevant final set of Decisive Conditions than either can generate alone. This led to another benefit: increased quality and relevance of plans.

With respect to generalized and inferred benefits, we argue that Support to Operational Design is a type of concept through which planners and agents generate and develop a shared (human-agent) representation of the developing plan – beginning with the Operational Design. understanding of the operating environment. Here, again, we have reusable, human-interpretable, machine-readable data. Such a representation is fundamental to any other activity that must exploit and / or extend the Operational Design. This includes the development of Operational Effects, Actions and Resources within the plan. Naturally, such a representation is also useful to other planning processes within the chain of command.

11 Conclusions and Reflections

11.1 Addressing the challenge

Operational-level planning is a challenging case for AI because it:

- Grapples with complexity – which is supported by systems thinking (as opposed to reductionist, linear thinking).
- Includes macrocognitive functions such as sensemaking and design, which rely upon experience, judgment and creativity and are extremely difficult to understand and codify.

These characteristics place cognitive demands on planners and, in the case of a human-agent collective,

[11] This was due to a data issue. The agent should have used the Critical Vulnerabilities (CVs) of the Operational CoG as the primary basis for generating DCs but there was insufficient contextual data on CVs to support this. The agent therefore worked primarily from the Operational End-State data.

these demands translate to functional and non-functional requirements for AI (both now and in the future).

Of the concepts and concept solutions that we have developed:

- Auto-piggery – and the Stakeholder Mapper concept solution – supports sensemaking.
- The Systems Approach supports both sensemaking and design; the Support to Operational Design concept solution specifically supports design.

Within each concept solution, the functional requirements for AI extended to the generation of new ideas for exploiting within Operational-level planning:

- The role of the Stakeholder Mapper PoC functional agent was to identify stakeholders and thereby analyse and visualise their interests in critical issues, based upon external data sources and HQ understanding products.
- The role of the Support to Operational Design PoC functional agent was to generate Decisive Conditions for planners' consideration, based upon external data and HQ understanding and planning products.

Non-functional requirements were embodied in the design philosophy shared by both concept solutions. These requirements included the following:

- Planners are in the lead and the functional agents are in a supporting role.
- The agent should explain its outputs to planners.

For both concept solutions, we exploited generative AI (Large Language Models) with Retrieval Augmented Generation to meet the functional requirements for AI and isolated those parts of the source documents that had contributed to the output to support explainability. With respect to the first non-functional requirement, human and agent roles were established and supported by the design of human-agent interactions, supported by voice and HMIs, and also mediated by the Tinman interaction agent.

When the research began, it was unclear whether (current) LLMs could be exploited to support Operational-level planning. We successfully developed working concept solutions, however, and thereby made progress in addressing this 'challenging case'.

11.2 C2 Benefits

The evidence indicates that both concept solutions offer benefits to Operational-level planning and therefore to C2. In summary, planners and agents were, collectively, more effective than either would have been alone. Given the design philosophy of 'planners in the lead, AI supporting', this translates to the PoC agents adding value to human activities as carried out by experts.

Where did this added value come from? It was, in part, due to the PoC agents' capabilities, not least the rapid processing of large datasets to generate outputs for consideration by planners. A key finding from the research is that even when some outputs were questioned for their relevance and quality (in the case of the Support to Operational Design concept solution), most outputs were considered useful, and a sufficient proportion were also novel. This was evidenced by the way planners' thinking and dialogue was stimulated and shaped by such outputs. Reflecting on the earlier discussion about utility vs. accuracy in Operational-level planning (in complex environments), it was clear that planners did not reject PoC agent outputs on the basis of accuracy; indeed, the concept of accuracy is less meaningful when making sense of, and designing responses within, complex situations. This calls into question the standards that we hold AI to in different work domains. In Operational-level planning, it should perhaps be lower than in higher-risk domains where the impact of inaccuracy is critical. It also reinforces a fundamental requirement for planners to understand this argument, not to expect or assume accuracy from AI where it is unreasonable to do so and, critically, to challenge outputs from AI in order to understand them.

The explainability feature in both concept solutions helped the planners to understand the derivation of outputs and thereby gain confidence in them. The specific challenge for Support to Operational Design is that it was intended to mirror the thought processes of planners in deriving DCs primarily from the CVs (and potentially CRs) of the Operational CoG, relying upon a rich and deep understanding of the operating environment as a CAS; the Operational End-State is a secondary input used to 'fit' DCs. The research team made excellent progress in eliciting guidelines for DC generation from the planners [7], to inform agent development. The key problem, however, was a lack of rich data from CoG Analysis (i.e. CCs, CRs and CVs) and, to counter this, the agent derived DCs primarily from the Operational End-State, for which there was richer textual data available.

This data issue can easily be overcome through more detailed elicitation of not just outputs from CoG Analysis but also the argumentation and rationale associated with this activity – as would likely be available in the planners' dialogue rather than the agreed outputs. The key learning point, however, concerns sources of trust and confidence in AI, from an expert planner's perspective. The DCs generated by the PoC agent were, according to the planners – too deductive – as one might expect if the primary input was, essentially, a goal itself (the End-State) rather than a key idea to 'unlock' this goal (the CoG). We also note that some DCs were less trusted because they were framed at the Tactical, rather that Operational, level – and this issue was attributed to a Tactical focus within at least some of the source documentation.

We must also keep open the possibility that a future Support to Operational Design PoC agent derives 'good' DCs through completely novel approaches. That is, there is no specific requirement for the agent to mirror human cognition – rather, the requirement is for output that is both useful and explainable.

There were fewer issues of trust and confidence in the outputs from the Stakeholder Mapper PoC agent. This may be because its functionality did not mirror a process that the planners themselves work through. That is, as a team member, the PoC agent was representing an analyst rather than a planner, whereas in the Support to Operational Design, the PoC agent was representing another planner.

The generalized benefits, whilst inferred and therefore less strongly grounded in the evidence, indicate that the two concept solutions are each an example of two much broader classes of solutions that generate, build, maintain and exploit shared human-machine representations of both understanding and design/planning. These are exactly the types of representations that the C2 HAC is based upon and the research described in this paper thereby provides a stepping stone to user-centred development of that wider socio-technical system.

11.3 Next steps

Our research generated novel concepts for Operational-level planning and brought them to life as AI-enabled concept solutions. These concept solutions were necessarily developed at an initial level of maturity and a goal of evaluation was to establish C2 benefit. The evaluation methods we used were:

- Exploratory – valuing open reflection and dialogue, especially about design, rather than control of the setting within which evaluation was conducted.
- Formative – focused upon understanding why concept solutions were (or were not) effective, thereby informing further development, rather than provided a pass/fail mark.

Further, the setting we created (which was identical for both evaluations) was representative of an Operational-level planning setting but at a relatively low level of ecological validity. That is, it was free of many of the real-world constraints that planners experience in Operational-level HQs. This enabled us to identify, qualify and reason about a wide range of potential C2 benefits. We saw this as an important first step – rather than to jump straight to hypothesis testing as part of a limited-objective experiment [8].

Further, the planners who participated in the evaluations were also members of the research team and led the Operation HUSKY Campaign Planning week. The planners therefore knew the evaluation scenario extremely well and had generated a degree of the input data. This naturally meant that the planners had excellent context for understanding the PoC agent outputs (especially in the case of the Support to Operational Design concept solution). We note that this is both a strength (because the planners were able to distinguish more useful agent output from less useful) and a weakness (because the planners were less reliant upon the agent in their work).

Having established evidence for C2 benefits for these low-maturity concept solutions, in laboratory settings, the natural next steps are to continue user-centred design and development, including the generation and curation of relevant data, and evaluate in more ecologically valid settings with methods that generate higher levels of evidence. Possible options for increasing ecological validity include:

- Introducing a novel scenario
- Working with participants who have had no previous exposure to the scenario
- Working with participants who have differing levels of expertise in Operational-level planning
- Introducing real-world constraints such as time-pressure and sparse data.

By increasing ecological validity, we generate richer opportunities to determine whether the concept solutions both generate C2 benefits and *disbenefits*, i.e. do they degrade any aspect of C2 performance and effectiveness? This is critical to understanding the readiness of any such concept solutions for implementation in Operational-level Headquarters.

Acknowledgements

This work was carried out as part of the Serapis Research Programme. The research reported on in this paper was funded by the UK MOD Machine Speed Command and Control (MSC2) project. This project was part of the UK Defence Science and Technology Laboratory's (Dstl) AI Programme, with the intent to transform C2 by enabling more timely and effective C2 processes across all environments, domains and levels of command, so the Defence enterprise can anticipate and adapt more successfully than adversaries.

This paper is one of six presented at the 29th ICCRTS that document different aspects of the MSC2 project, which explored the feasibility of a Human Agent Collective (HAC) that combines human insight with machine speed AI agents employing shared digital artefacts, shifting C2 from human teams to human-machine teams, where humans and AI work together.

References

[1] UK Development, Concepts and Doctrine Centre (2022). Joint Doctrine Publication 0-01: UK Defence Doctrine (6th Edition). MOD, UK.

[2] UK Development, Concepts and Doctrine Centre (2017). Joint Concept Note 2/17 The Future of Command and Control. MOD, UK.

[3] Klein, G. & Wright, C. (2016). Macrocognition: From Theory to Toolbox. Frontiers in Psychology. 7. 10.3389/fpsyg.2016.00054.

[4] Leggatt, A., Turner, P., Asquith, R., Attfield, S., Roberts, H. & Ellis, R. (2024). Assessing novel C2 socio-technical AI enabled planning concepts: What we learnt using the Wizard of Oz and other evaluation methods. 29th International Command and Control Research and Technology Symposium (in preparation).

[5] Firth, A., Poole, J. & Turner, P. (2024). Opportunity Knocks. 29th International Command and Control Research and Technology Symposium (in preparation).

[6] Allied Joint Publication-5: Allied Joint Doctrine for the Planning of Operations, Edition A Version 2, 2019, NATO Standardization Office (NSO), 2019.

[7] Firth, A., Attfield, S. & Poole, J. (2024). A Set of Guidelines for Designing Decisive Conditions in Military Operational Planning. 29th International Command and Control Research and Technology Symposium (in preparation).

[8] Turner, P., Leggatt, A. & Smith, G. (2022). Bridging the gap between desktop-based C2 research and large-scale C2 experimentation. 27th International Command and Control Research and Technology Symposium.