REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
03/29/2020

2. REPORT TYPE
Technical Report

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
Simulating the Operational Impact of Artificial Intelligence-Enabled Systems

5a. CONTRACT NUMBER
FA8702-20-C-0001

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Richard D Flournoy; Robert L Wittman Jr.; Bianica S Pires; Robert G Layden

5d. PROJECT NUMBER
0720A90AMA

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102

8. PERFORMING ORGANIZATION REPORT NUMBER
PRS-21-0660

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
United States Army (ACC-APG)

10. SPONSOR/MONITOR'S ACRONYM(S)
ACC-APG

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report documents an investigation to explore current AI simulation capabilities and recommend a path forward in this area. The investigation was focused on AI-enabled system concepts of interest in these areas: (1) automated target detection and recognition, (2) the impact of AI-generated data on strategic decision making and related timelines, and (3) strategies for managing unmanned vehicle maneuvers and tactics ranging from human-controlled through completely autonomous operations.

15. SUBJECT TERMS
Modeling and Simulation; Artificial Intelligence; autonomous systems; intelligence modeling; artificial intelligence; conceptual frameworks; mission level simulation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Susan Carpenito
				13	19b. TELEPHONE NUMBER *(Include area code)* 781-271-7646

SIMULATING THE OPERATIONAL IMPACT OF AI-ENABLED SYSTEMS

R. Douglas Flournoy
Dr. Bianica Pires
Dr. Rob Wittman Jr.
Dr. Rob Layden
MITRE Corporation
7525 Colshire Drive
McLean, VA, USA
{ rflourno, bpires,
rwittman, rlayden}@mitre.org

ABSTRACT

Artificial Intelligence (AI) technologies are advancing rapidly, and interest in leveraging these technologies to improve operational outcomes is high. This is leading to an increasing number of questions about AI-enabled systems and their impact on battlefield operations. This report documents an investigation to explore current AI simulation capabilities and recommend a path forward in this area. The investigation was focused on AI-enabled system concepts of interest in these areas: (1) automated target detection and recognition, (2) the impact of AI-generated data on strategic decision making and related timelines, and (3) strategies for managing unmanned vehicle maneuvers and tactics ranging from human-controlled through completely autonomous operations. The traditional, community-preferred simulation tools we examined in this investigation are useful for representing some aspects of the AI-enabled systems analysis problem space. However, none as currently constituted offers all the capabilities necessary to cover the entire problem space. One of the key areas identified for strengthening current mission-level simulations was the modeling of intelligent behavior. We studied current behavior modeling techniques and found conceptual framework approaches to be well suited for quickly creating transparent, theory-based, credible models of both artificial and human intelligence. The widely used Beliefs, Desires, and Intentions (BDI) conceptual framework is particularly well suited for simulating AI-enabled systems because of its goal-oriented approach. We had initial success implementing BDI in a lightweight, agent-based mission-level simulation. We recommend the use of conceptual frameworks in general, and BDI in particular, for building the new models of intelligent behavior necessary to analyze mission-level impacts of AI-enabled systems.

Keywords: artificial intelligence, autonomous systems, intelligence modeling, conceptual frameworks, mission level simulation

1 INTRODUCTION

Artificial Intelligence (AI) technologies are advancing rapidly, and interest in leveraging these technologies to improve operational outcomes is high. This naturally leads to questions when considering enabling a system with a specific application of AI, such as:

- How much improvement can we expect to see in our mission outcomes?
- Are vulnerabilities introduced, and how might these play out in actual operations?
- How will operational procedures and timelines change?
- During initial fielding, could there be adoption and trust issues that mitigate impact?
- Should the AI platform or component look to a human for guidance, control, override, or acknowledgement, or be completely invisible to the human participant?

- Should the AI platform provide recommendations or make independent decisions?
- Will the platform or component's performance improve over time (will it learn?)

To answer questions about proposed system enhancements like these, for decades the military analysis community has employed battle simulations to facilitate mission-level analyses. But because many AI solutions are relatively new, these simulations typically don't represent the effects of AI out-of-the-box. The purpose of this paper is to generate guiding principles for the military analysis community as they anticipate analysis work that involves assessing the mission impacts of AI-enabled systems. As part of this charter, this paper also examines the suitability of current analysis tools to capture the effects of AI-enabled systems appropriately at the mission level.

The number of ways AI can be applied as a system-level aid is growing rapidly. Moreover, with the advent of Machine Learning (ML) technologies, the AI "solution space" is expanding beyond AI approaches of previous generations. For the purposes of this paper, we focus only on specific AI methods and mission foci. These areas include: (1) AI-enabled target detection & recognition, (2) AI-enabled strategic decision support, and (3) AI-enabled tactical behaviors. Noting future concept development projects that indicate a high level of military interest in AI-enabling Unmanned Air Systems (UASs), we focus on AI-enabled UASs. Highlighting scenarios involving AI-enabled UASs ties back to the three concept areas above in these ways:

1. ML-enabled UAS imaging/video systems speed target detection and recognition timelines.
2. Impacts of AI-enabling UASs on engagement decision-making at more strategic levels are realized through increased engagement opportunities and decreased decision timelines.
3. Both ML and Rule-based AI on UASs enable the orchestration of autonomous maneuvering (swarming, avoiding detection/jamming, timing of communications, etc.).

In the remainder of this paper, we explain the key motivating concepts and considerations in the areas of (1) AI and autonomy, and (2) and modeling of intelligence (Section 2), examine the capabilities for simulating AI-enabled systems in candidate battle simulations (Section 3), and provide conclusions and recommendations (Section 4).

2 REPRESENTING AI-ENABLED SYSTEMS: CONCEPTS AND CONSIDERATIONS

This section covers advanced concepts and considerations foundational to the study effort. Section 2.1 briefly explains the relationship between the fields of AI and Autonomous Systems. Section 2.2 provides an overview of behavior modeling approaches, and indicates how the selective use of these approaches can aid in simulating AI-enabled systems for analysis.

2.1 AI and Autonomy

Considering the use of AI to enhance UASs and their impact on the battlespace brings into focus the related field of Autonomous Systems. UASs can be designed to operate across the spectrum of autonomy, from completely operator-controlled to completely self-controlled. Many questions that arise about UAS operations center on this theme: what is the most effective level of autonomy for a particular use case? So it is useful to consider the relationship between AI and autonomy as it affects different facets of UAS operations and how they are modeled.

Just as "knowledge" and "control" are different concepts, AI and Autonomous Systems are distinct scientific fields. One can imagine an AI algorithm that is not part of an autonomous system—for example, any knowledge-driven recommender algorithm such as an ML-based pattern recognition algorithm that is used to support a human decision-making process. Conversely, early gyroscope-based autopilot systems are examples of autonomous systems that are not supported by AI. But with technological advances the complementary nature of AI and Autonomous Systems is becoming increasingly evident and is particularly relevant to UAS design and operations.

Looking back at the 3 numbered AI-enabled use case areas listed at the end of Section 1, consider how the level of autonomy involved in each case changes the dynamics and impact of the AI-enabled system:

1. For AI-enabled target detection and recognition, computer vision that suggest targets to a human operator is a much different dynamic than automatically recognizing targets and feeding them into a tactical-level autonomous effects chain. But both approaches, in addition to the traditional case where there is no computer vision support, may be appropriate in different situations, and we need to be able to model and analyze them all. For AI-enabled vs. human operator target identification, changing the target recognition timeline, throughput, and probability of correctness parameters may be sufficient. Or if there is a particular detail of the computer vision algorithm's performance or vulnerability (e.g., cyber) that needs to be explicitly represented in the analysis, in these extreme cases integrating the actual system software into the simulation (Software-in-the-Loop, or SIL) could be a desired approach.

2. Where AI-enabled UASs are generating data that affects strategic-level engagement decisions, we need to be able to model different target processing throughputs, along with any differences in confidence or trust decision-makers have between AI-provided vs. human-generated target data.

3. In the case of autonomous vs. traditional human-controlled maneuvers, the model needs to take into account any differences between the rules triggering and governing autonomous movement vs. the movement behavior a human operator would be likely to induce. These differences could involve representing in behavior models how an autonomous UAV could possibly react quicker than a human operator to locally available sensor data, but might be missing a "big picture" perspective that could impact a human operator's control decisions. Or in the case of human-controlled operations, comms are required which can expose the UAS to interference or spoofing. Also, controlled operations can limit the effective range of small UAS that need to maintain this direct comms link back to the human operator. Again, in extreme cases that border more on system-level software testing as opposed to mission-level analysis, an SIL approach could be beneficial.

In the (a) and (b) cases, we noted that the analysis space is not just either/or when it comes to autonomy or human control. There are also in-between states to consider, which we also need to be able to model, that involve different levels of autonomy and potential conditional triggers for consciously deciding to change the level of autonomy of platforms mid-mission.

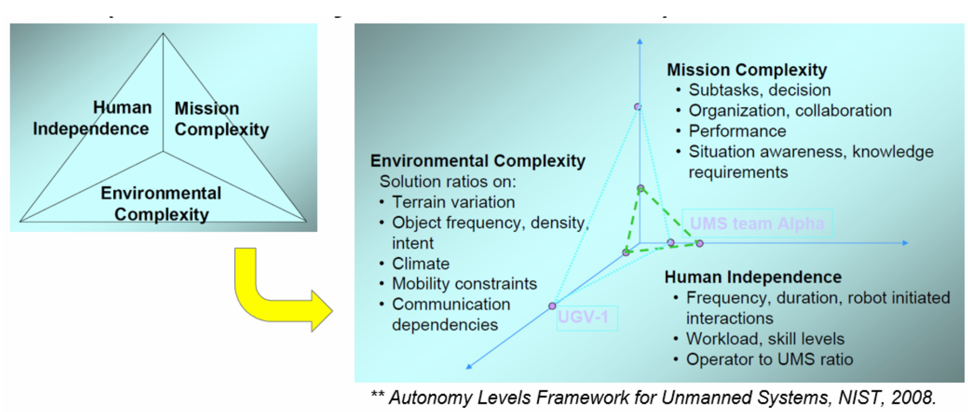

Figure 1. Autonomy Levels Framework for Unmanned Systems (NIST, 2008)

To help us grapple with these possibilities, we searched the literature for methods others have designed or applied in the past to think about levels of autonomy. We found an Autonomy Levels Framework for Unmanned Systems (ALFUS), conceived by a working group in 2008 under the direction of the National Institute for Standards and Technology (NIST). Figure 1, above, provides 2 key conceptual model diagrams from ALFUS. The triangular diagram on the left represents the ALFUS Contextual Autonomy Capability (CAC) Model, illustrating how autonomy of an unmanned system (Human Independence axis) needs to be determined in context with the Mission Complexity and the Environmental Complexity surrounding the system's operational objectives. The ALFUS framework defines metrics along these 3 axes (some of which are listed on the right-hand graph of Figure 1). Further detail in the ALFUS documentation shows how they envision using this construct and supporting metrics to facilitate articulating, communicating, evaluating, and documenting unmanned system requirements and capabilities (NIST, 2008). To represent AI-enabled systems properly, our analysis tools must be flexible enough to allow modeling different levels of operational autonomy, in addition to capturing how behavior at different levels of autonomy maps into the Mission and Environmental Complexity of the analysis use case.

2.2 Modeling Intelligent Behavior

Concepts associated with modeling behaviors are central both to our ability (1) to represent AI-enabled systems and (2) to model human (not AI-enabled) behaviors realistically in comparison to AI-enabled systems. As part of this investigation, we conducted a literature search on approaches used to model behaviors in agent-based models. While there is not necessarily one "right" approach (or combination of approaches) for modeling any certain behavior, there are certain considerations one should take into account in selecting an approach. These include:

1. The purpose for the model (e.g., exploratory, policy guidance).
2. The application (domain) of the model (e.g., combat, infectious disease, transportation).
3. The types of behavior exhibited by the system (AI-enabled or not), or components of the system, to be modeled (e.g., the UASs multi-layer architecture includes operational, tactical, and strategic layers, each of which has unique behaviors).
4. The level of abstraction (or complexity) of the environment, the mission, and human independence of the systems to be modeled (see Figure 1).

The model's purpose in conjunction with domain are arguably the most important considerations in deciding a suitable approach for modeling behavior. They will drive how and at what level we abstract AI-enabled (and not AI-enabled) systems, the environment, and the mission, and in turn the approach or combination of approaches suited for representing behaviors in the model. Furthermore, we must consider the representation of the components of the AI-enabled (and not AI-enabled) system: from a platform-centric representation that would be limited in its representation of independent sensors, weapons, basic load (munitions), and other resources to explicitly detailed representation of each component of the system.

The literature categorizes the approaches for modeling behavior into four main areas: mathematical approaches, graphical approaches, conceptual frameworks, and cognitive architectures. A brief description of each is provided in Table 1. These approaches are often complementary and can be combined when appropriate; the use of a hybrid approach will be discussed briefly at the end of this section.

Table 1: Four main approaches for modeling intelligent behavior in agents.

Approach for representing behaviors in M&S	**Description**	**Alignment with AI-enabled behaviors**	**Benefits**	**Limitations**
Mathematical approaches	Mathematical simplifications such as	Operational-level decisions related to the control of UASs	Easy to implement	May lack transparency, particularly when used to

	if-then statements and threshold calculations			simulate complex behaviors
Graphical approaches	Graphical structure visualized in one or more diagrams. Behavior often represented as "states"	Operational- and tactical-level behaviors	Rigorous structure for designing and implementing behavioral rules	Does not provide guidance on how humans or AI-enabled systems actually behave
Conceptual frameworks	Guiding frameworks that integrate the varied components of behavior	A general approach for simulating behaviors and their interactions at different levels of detail	Transparency and explainability; provides guidance and clear structure; allows for modeling higher-level behaviors such as beliefs	Many implementation decisions still left to the modeler; may be too cumbersome for implementing simple behaviors
Cognitive architectures	Architectures of the "mind"; mechanisms and structures that underlie human cognition	Strategic-level decisions related to a UAS's mission	Clear guiding structure; validated and well-established approach for simulating decision-making	Captures lower level processes that are typically beyond the scope of most applications; computationally intensive; challenging to scale

Mathematical approaches. Mathematical approaches are methods that produce agent behavior through the use of mathematical simplifications (Kennedy, 2012). These include informal (ad hoc) custom coding of behaviors through mathematical simplifications such as if-then statements and threshold calculations and formal mathematical models such as game theory and machine learning.

Reactive behaviors (reaction to some stimuli) are often implemented through informal custom coding given the simplicity of the behavioral rules. As an example, a simple traffic model could implement if-then statements to program agents (cars) to slow down if they "see" another agent in front or to speed up if no agent is in sight (Wilensky and Rand, 2015). Reactive behaviors are the most basic of the behaviors; and depending on the model's purpose and application this may be where we stop model development, or it may provide a good starting point for implementing more complex behaviors (Wilensky and Rand, 2015). While the purpose of this traffic model would be exploratory, similar reactive agents in a traffic scenario that represent a real-world location, such as through the use of geographic information systems (GIS), may be appropriate for understanding varying traffic scenarios in a real-world city or region for purposes of policy guidance. Mathematical approaches may also be well-suited for representing mission-level effects of certain AI applications. For example, it may be sufficient to put a threshold rate on the number of potential targets a human operator can identify over time from UAS-provided images or videos, then represent AI-enabled recognition with a higher rate parameter. While these behaviors are complex, they can be broken down into simpler threshold calculations. The vast majority of processes that underly decision-making (e.g., knowledge encoding, perception, reasoning) do not need to be simulated when the behaviors of interest can be reasonably specified through the simplifications discussed here.

Formal mathematical models can provide more rigorous grounding for agent decision-making over an ad hoc, custom approach. Game theoretic approaches, for instance, provide a formal (mathematical) model of behavior in situations of cooperation or conflict. "Players" must make strategic decisions based on a set of potential payoffs. Game theory has been used to simulate a variety of military scenarios, including cyber-attacks, sensor networks, and combat (James et al., 2017). Statistical models have also been used to drive decision-making behaviors. An agent may apply statistical methods (e.g., logistic regression, convolutions, Bayesian networks) to make inferences about its environment. For instance, a UAS may want to move in such a way that places it near the adversary, but not so close it risks being shot down. A convolution matrix can assess the "fitness" of locations within the UASs vision and determine the location that best meets this criteria. This directly impacts UASs' decisions around combat and tactical maneuvering in the battle space.

These examples demonstrate that the use of mathematical approaches does not necessarily imply model simplicity or even behavior simplicity, but instead, has a wide range of applicability. These approaches, however, still lack an overarching general structure suited for different types of behaviors. The decision to use a mathematical approach may be driven largely by a modeler's understanding of the rules of behavior governing the AI-enabled system, whether simple or complex. For a more general approach for modeling behaviors, conceptual frameworks (discussed later) may provide better building blocks for a modeler to quickly use and effectively capture the behaviors.

Graphical approaches. Graphical approaches aid in the design and development of behavior modeling. They provide a structure that can be easily visualized in one or more diagrams. The most widely known graphical approach and one that has gained wide acceptance in the agent-based modeling community is Unified Modeling Language (UML) diagramming. UML is a family of standardized graphical notations used for describing and designing object-oriented software systems. State charts, which describe behaviors through a set of states (e.g., move, attack, defend) and conditional transitions, have been shown to be particularly useful for describing agent behaviors. While a strength of UML diagrams is their ability to represent a variety of ideas and behaviors, this format lacks the rigor necessary for computational execution of the model. Another approach that has had less notoriety in the agent-based modeling community but that more rigorously specifies agent behaviors is Discrete Event Specification (DEVS). As with UML diagrams, DEVS consists of a family of diagrams for specifying the model's architecture, including state charts, sequence diagrams that specify how agents interact and exchange messages over time, and coupled diagrams that specify how coupled components in the model interact. The rigor inherent in DEVS is what allows for the transformation of these graphical representations into an executable simulation model (Martin et al., 2007). Multi-agent software and programming languages such as AnyLogic and Python, for instance, have built-in tools and packages for converting state charts directly to computer code.

Conceptual frameworks. Conceptual frameworks integrate the varied components of agent decision-making processes (e.g., Caillou et al., 2017; Pires and Crooks, 2017) and help guide the modeler in its implementation. Such frameworks include "fast and frugal", Physical conditions, Emotional state, Cognitive capabilities, and Social status (PECS), and the Beliefs, Desires, and Intentions (BDI) framework (Kennedy, 2012). These frameworks are typically flexible enough to model a broad-spectrum of behaviors, from simple, stimulus-response behaviors to proactive, goal-oriented behaviors. While programmatically, the computational implementation of the framework may consist of a combination of mathematical formulations, utility functions, and/or simple rules designed to achieve some goal, a framework provides guidance for how one might implement the different components of decision-making and their interactions.

Fast and frugal (Gigerenzer and Todd, 1999) provides a computationally inexpensive option for implementing decision-making processes in agents. It models behavior using a simple decision tree format, but does not account well for higher-level behaviors, such as emotion, acts of will, and social belonging. PECS, on the other hand, views agents as a psychosomatic unit with cognitive capabilities residing in a social environment (Schmidt, 2002). The PECS framework is flexible due to its ability to model higher-level behaviors, including intricate reflective behaviors, which requires that agents be fully aware of their internal model. However, it is not as widely known and there are few examples to draw from.

BDI (Beliefs, Desires, and Intentions) represents the information, motivational, and deliberative components of the system (Rao and Georgeff, 1995). Beliefs are the agent's knowledge about the environment; desires contain information about the priorities and payoffs associated with the current objectives; and intentions represent the chosen course of action. As with "fast and frugal", BDI agents use a decision tree process but rely on payoff and utility maximizing functions to select goals and to determine the optimal action sequence by which to achieve those goals. In a military context, beliefs may represent the situational awareness of an AI-enabled system, desires are its mission goals (objective), and intentions represent the course of action used to achieve these goals. Figure 2 provides a high-level depiction of the BDI framework in the context of a military entity.

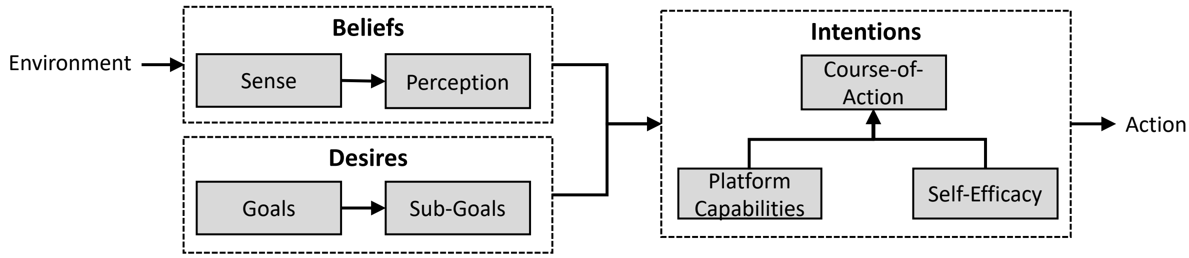

Figure 2. BDI framework in the context of a military entity. (Source: MITRE)

BDI is the most widely used architecture for modeling behavior (Malleson et al., 2012), and has been used in a variety of applications, from VR animation to UAV applications. It is especially useful for modeling goal-directed and autonomous behaviors, which is of particular relevance to the military and AI-enabled systems. Mualla et al. (2019) found that it was the predominant architecture used for simulating civilian UAV behaviors in agent-based models, suggesting that it is a promising paradigm for modeling more autonomous UAVs. Because behavior is driven over well-structured concepts (e.g., beliefs, goals), the cause-effect relationship can be clearly depicted and explained (Alzetta et al., 2020).

Cognitive architectures. Cognitive architectures are a "theory of the fixed mechanisms and structures that underlie human cognition" (Lehman et al., 2006). They are motivated by an in-depth understanding of the human brain from biological and neurological perspectives (Mualla et al., 2019). We can categorize cognitive architectures into symbolic and sub-symbolic learning. Two well-known cognitive architectures are ACT-R and Soar. Soar follows the symbolic paradigm, while ACT-R has both symbolic and sub-symbolic components, although it tends to be binned into the symbolic AI architectures. Purely sub-symbolic (connectionist) architectures include deep learning, neural networks, and genetic algorithms.

ACT-R provides a framework for representing knowledge and skill acquisition and recollection within agents. It is composed of a number of modules that capture various aspects of cognitive capabilities, including working memory, learning, perception, and action. ACT-R was originally developed in 1990 (Anderson, 1990) and has evolved since into a comprehensive cognitive architecture that closely matches human performance data (Anderson, 2007). The focus of the architecture is on low-level cognitive processes operating over very short periods of time. It can represent individual differences in knowledge acquisition, capacity, and knowledge recollection from simple knowledge gathering (declarative knowledge) to more complex task performance (procedural knowledge) (Orr et al., 2018). ACT-R does not support higher-level concepts such as beliefs, desires, and intentions; a conceptual framework or application-specific framework would be needed. In addition, ACT-R was not originally designed for scaling to populations (whether human or AI-enabled). Some progress has been made in modeling multiple cognitive agents within an agent-based model (Lebiere et al., 2010). ACTR-UP, for example, applies what they term "Accountable Modeling" to model only those components of cognition that are specified and supported by theoretical and empirical evidence (Reitter and Lebiere, 2010). Having said that, these scaling techniques are still an emerging research area and as such, simulating thousands of ACT-R agents concurrently may pose a computational challenge.

Soar is the basis for Allen Newell's Unified Theories of Cognition (Newell, 1994) - a set of general assumptions (microtheories) that together produce the full range of cognition. It is an AI system designed to meet or exceed human performance in problem-solving tasks. Thus, it differs from ACT-R which emphasizes modeling human cognition. Like ACT-R, however, scaling models to thousands of "Soar agents" would pose a significant challenge in computing power. The main process for decision-making in a Soar agent involves the interaction between procedural memory (the agent's learned behaviors) and working memory (how it perceives the current environment). It's representation of the world is stored as "states" while procedural memory is represented as if-then statements (Lehman et al., 2006). Soar "could be considered to be an implementation of a BDI architecture in that it maintains an internal representation of the world, is always working to solve a goal, and has available internal state variables" (Kennedy, 2012).

Sub-symbolic approaches seek to simulate neurobiological processes at the level of the neuron. By representing neurons, activations, and their associated network of connections and interactions, a deep (neural network) algorithm can learn given many training examples. For example, a convolutional neural network (CNN) could be implemented to perform image processing within a UAS agent in a simulation. This would allow the UAS agent to not only process static images but learn based on past experience, thus more closely aligning perception with ground truth over time. How this type of technology impacts the battlefield could then be simulated. While such deep learning approaches provide a data-driven method for knowledge representation with noteworthy success across a variety of domains, it is also a "black box" in that one cannot trace back how the algorithm derived any given answer. This may be particularly limiting when explainability and transparency in the AI-enabled system is sought.

Cognitive architectures are beneficial when seeking to model and understand the underlying processes associated with cognitive phenomena (e.g., learning, planning, memory, perceptions) for purposes of complex decision making, such as determining mission goals and directives. In particular, if capturing decision-making under information overload or other duress is critical to a particular AI-enabled system analysis, a cognitive architecture could give the modeler a better toolset for their modeling tasks. The implementation detail associated with using a cognitive architecture, however, is a significantly "heavier lift" than the conceptual frameworks, mathematical simplifications, and graphical approaches discussed. If the purpose of the model and the desired abstraction levels of the agents (AI-enabled or otherwise) warrant the use of a cognitive architecture, then the toolset is available, is widely accepted in the fields of cognitive science and artificial intelligence, and provides a strong theoretical grounding for modeling such behaviors.

Hybrid Approaches. In considering an approach to use, one should not overlook the possibility for combining one or more of the approaches discussed. For instance, we may want to combine simple, reactive behaviors (e.g., speed up if no vehicle is in front; slow down otherwise) with proactive, goal-oriented behaviors (e.g., plan the best route from my current location to a destination). Reactive behaviors can typically be implemented through mathematical simplifications. Proactive behaviors, on the other hand, may be better suited for a decision tree format, such as BDI or Fast and Frugal. Conceptual frameworks can provide the structure for organizing various types of behaviors, simple or complex. If we decompose the implementation of the reactive behavior, for example, one may find that it uses a mathematical approach; but that it is structured within a larger framework. This may be beneficial when transparency and model explainability is important. Moreover, conceptual frameworks can provide a meta-framework (sometimes called a macro-architecture) to organize knowledge and skill content in respect to a cognitive architecture (e.g., West et al., 2017). As such, cognitive architectures are not necessarily a replacement for conceptual (meta-) frameworks; the two are fundamentally complementary (Orr et al., 2018). This is because a cognitive architecture by itself may not be able to represent the full range of desired behaviors. Should a combination of higher-level behaviors (e.g., beliefs) combined with the cognitive detail of an architecture be needed, combining the two approaches may be useful.

As an example of a hybrid solution, we might consider how the perception (Beliefs) component of BDI is responsible for translating the state of the world (sensing) to an entity's perception of that state. A UAS performing an ISR mission will sense ground and air combat at varying degrees of uncertainty. If capturing that uncertainty (i.e., the difference between ground truth and perception) across platform components in a way that is reflective of an actual UAS is of particular importance to the model's purpose, then a hybrid solution that integrates BDI with a cognitive architecture may be appropriate. For example, we could envision training a neural network to perceive the world at a level of uncertainty that is representative of the UAS simulated. Moreover, we could allow the neural network to learn in a way that uncertainty decreases with time on the battlefield. Results from the neural network, i.e., what the UAS has perceived, would then inform the UAS's desires (goals) as defined in the decision tree structure, whether that might be to continue ISR or to approach for attack.

Transparency and approaches for modeling behavior. In selecting an AI-effects modeling approach we should also consider the transparency of the approach, to promote (1) understanding the implementation, (2) establishing validity, and (3) gaining trust in the model and its results.

Building a complicated behavior algorithm through math/logic statements can be very hard to understand and debug for simulation developers and users, let alone explain to the analysis community. On the other hand, if something complex is being represented by a very simple implementation, this needs to be made clear (e.g., the probability of target identification increases from 50% to 90%; time delay to target ID decreases from 90 seconds to 10 seconds). State machines, decision trees, and other behavior modeling constructs that are found in conceptual and graphical frameworks can make implementations much more transparent. For instance, BDI is seen as a promising approach for building explainable agents (Mualla et al., 2019). This is because "the core concepts of the [BDI] agent framework map easily to the language people use to describe their reasoning and actions in everyday conversations" (Norling, 2004). Cognitive architectures such as ACT-R and Soar, on the other hand, are computationally complex but offer a theoretically grounded, principled, and validated approach for modeling human cognition. They are arguably the most principled approach for simulating cognitive phenomena and behaviors. While ACT-R's complexity poses a challenge for explainability, it is transparent in the sense that every component (module) and interaction has been thoroughly documented and validated in publications over the course of the last 40+ years. For example, if a model is said to use the procedural learning module in ACT-R, one could easily reference multiple publications that describe this module in detail. On the other hand, deep learning approaches, while perhaps more representative of the human brain, lack this type of transparency and explainability.

3 SIMULATING AI-ENABLED SYSTEMS: CAPABILITIES OF CANDIDATE BATTLE SIMULATIONS

For decades, the military community has been using simulations—not only for analysis purposes as is the focus of this paper, but also to support training, experimentation, testing, and other activities. Often no existing simulation meets all the requirements for an upcoming study, test, or operator-in-the-loop event. So a simulation that meets many of the requirements is chosen as a starting point and development work is done to the simulation to address the remaining requirements. Over time, simulations that are strong in certain areas evolve to become ever more capable in areas not too far removed from these strengths. In this manner, several mission-level simulations have become popular within their own subsets of the military community for the capabilities and behaviors they model well and/or the warfighting domains they cover.

With the advent and impending adoption of new AI techniques in military systems, simulations will be called on to help determine (1) which techniques are worth deploying and on which systems, and (2) how these techniques can best be incorporated in operations ranging from tactical information gathering to strategic decision making. Based on the unmanned systems-related operational scope of this study and the AI modeling concepts and considerations discussed in Section 2, the characteristics that make a simulation a strong candidate to be extended to represent AI-enabled systems for mission-level analysis include:

1. **Domain strength.** A primary driver for selecting a simulation to support an AI-related analysis will be the strength of the models it offers in the warfighting domain of interest. For example, if an analysis will emphasize low flying AI-enabled UASs supporting ground operations, the simulation should feature strong ground warfare models and terrain representation.
2. **General modeling flexibility, including tools to support intelligent behavior modeling.** Simulations will be called on to represent future systems and tactics that include AI effects and autonomy logic. Where intelligent behaviors are needed, constructs such as the conceptual

modeling frameworks outlined earlier will aid in efficiently creating new, understandable models of both human and AI-enabled tactics and decisions.

3. **Transparency** in the way both new and existing models are represented, so that the effects of the newly modeled behaviors and how they interact with existing systems and decision makers can be clearly understood and their impacts can be mapped to system-level measures of performance and mission-level measures of effectiveness.

4. **Right-fidelity/abstraction match** to the analysis use case. A lower-fidelity simulation can be appropriate for a mission-level analysis where much detail can be abstracted away, but higher fidelity might be needed for an analysis focused more at the system or engagement level. However, often a mix of fidelities for different platforms within the simulation will be the ideal case, where certain featured platforms are modeled with more detail than background platforms. The ability for a simulation to allow the user to tailor the level of detail needed in selected areas is key.

5. **Software-in-the-loop interface capability.** This requirement may be out-of-scope for most mission-level analyses. However, for testing or analysis at the system- or engagement-level, or for certain detailed software vulnerabilities, the ability to "play" the actual AI/ML algorithm—or even the entire system software stack—in a simulated context may be useful or necessary.

Given enough time, effort, and development expertise, many of the existing battle simulations could be extended to sufficiently represent the mission impacts of AI-enabled system concepts. For this study, we chose to examine two such traditional simulations, each with its own substantial history and user base within the military analysis community, to see how they "measure up" to the tenets above. The traditional simulations have strong, validated models in their domain areas of specialty. Especially in the physics-related aspects of military operations, they tend to model systems and engagements at a high level of fidelity (sometimes higher than needed or desired). But, as expected, there is a lack of built-in models and parameters to represent decision-making and other human and intelligent systems behaviors. The simulations vary in the tools they offer analysts to develop these models themselves—one offers a scripting language that a user can learn to use, given time, to develop new models. The other traditional simulation offers a defined set of prepackaged, parameterizable models of traditional systems, but if new capabilities beyond the scope of those models are required, the user must build them at the source code level.

We also examined a third simulation that uses a more streamlined model development philosophy than the traditional battle simulations. This third simulation is developed in NetLogo, a multi-agent programmable modeling environment developed at Northwestern University (Wilensky, 1999). This Netlogo-based simulation adds a military agents and capabilities layer to NetLogo's agent, network, experimentation, and simulation environment. This NetLogo-based simulation does not at this time offer high-fidelity physics-based models of system performance and engagements. It does, however, include a BDI conceptual framework implementation and other agent-based behavior modeling building blocks. This simulation shows promise for analyses that require efficient and transparent modeling of AI-enabled vs. human-driven pattern recognition, tactical behaviors, and decision making. The NetLogo-based simulation has already been used to represent AI-enabled target recognition effects, autonomous tactical operations, and different levels of UAS control vs. autonomy.

4 CONCLUSIONS & RECOMMENDATIONS

Concepts associated with modeling behaviors are central both to our ability to (1) represent AI-enabled systems and (2) model human (not AI-enabled) behaviors realistically in comparison to AI-enabled systems. The selected approach for modeling behaviors should align with the purpose of the model and desired abstraction level and account for the benefits and limitations of each approach. We conducted a literature search on current intelligent behavior modeling approaches and found they conformed to four main categories: mathematical simplifications, graphical approaches, conceptual frameworks, and cognitive architectures; with hybrid combinations of these approaches also possible. Of these, conceptual frameworks are particularly attractive in that they are typically (a) flexible enough to model a broad spectrum of behaviors while (b) offering helpful constructs that guide the modeler to a transparent implementation based on accepted theory and practice. The BDI conceptual framework is particularly well suited for simulating AI-enabled systems in a military context because of its goal-oriented approach. BDI is the most widely used architecture for modeling behaviors and the predominant approach for simulating civilian UASs. Because BDI has been used across a variety of non-military and military domains, there exists open source literature, models, and built-in libraries and packages across different simulation platforms that can be leveraged. BDI's simple decision tree structure promotes model transparency and explainability. Finally, BDI can provide an overarching structure (meta-framework) to integrate cognitive architectures (symbolic or sub-symbolic) and mathematical approaches for hybrid solutions (simple or complex). For these reasons we recommend the use of conceptual frameworks in general, and BDI in particular, for building the new models of intelligent behavior necessary to analyze mission-level impacts of AI-enabled systems.

The simulation tools we examined in this investigation have all been shown useful for representing subsets of this analysis space. However, none as currently constituted offers the unique combination of strengths necessary to cover the entire problem space. The two traditional simulations we examined offer strong physics-based system representations in their areas of domain specialty, but they vary in the tools they offer to the user to develop new models—and neither offers a Conceptual Framework for modeling the behavior of human and/or artificial intelligence. A third, NetLogo-based simulation we examined is a relatively new, lightweight simulation. It does not offer high-fidelity physical models of system performance, but this third simulation provides a scripting language and NetLogo's open source libraries for rapidly modeling new systems at desired low-to-medium fidelity levels. It also includes a BDI implementation for building transparent, theory-backed models of intelligent behavior. And this NetLogo-based simulation has been used for several AI-enabled autonomous systems projects.

Regarding the application and future development of these three tools to simulate and analyze the operational impact of AI-enabled systems, we offer the following recommendations:

- A traditional simulation should be chosen for analyses where comprehensive, validated, and/or high fidelity physical warfare models are required. In future development efforts, use conceptual frameworks (e.g. BDI) to build in support for AI-enabled systems modeling as resources permit.

- Consider the NetLogo-based simulation where rapid creation of models of interactive systems and intelligent behaviors is needed. Continue using this simulation as a prototyping-level testbed for learning how to best model the range of human-to-autonomous behaviors in AI-enabled systems.

REFERENCES

Alzetta, F., Giorgini, P., Najjar, A., Schumacher, M. I., & Calvaresi, D. (2020). In-Time Explainability in Multi-Agent Systems: Challenges, Opportunities, and Roadmap. In International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems (pp. 39-53). Springer, Cham.

Anderson, J. R. (1990). The adaptive character of thought. Psychology Press.

Anderson, J.R. (2007). How Can the Human Mind Occur in the Physical Universe? Oxford University Press.

Brousse, O. (2020). "AI for Ardupilot Developers: Neural Networks," presentation to the 2020 Ardupilot Developer Conference, 26 March 2020.

Caillou, P., Gaudou, B., Grignard, A., Truong, C.Q., Taillandier, P.: A simple-to-use BDI architecture for agent-based modeling and simulation. In: Jager, W., Verbrugge, R., Flache, A., de Roo, G., Hoogduin, L., Hemelrijk, C. (eds.) Advances in Social Simulation (2015). AISC, vol. 528, pp. 15–28.

E. Norling, Folk psychology for human modelling: Extending the BDI paradigm, in: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1, IEEE Computer Society, 202-209, (2004).

Gigerenzer, G., Todd, P.M., (1999). Fast and Frugal Hueristics: The Adaptive Toolbox, in: Simple Heuristics That Make Us Smart, Edited by G. Gigerenzer, P. Todd, and The ABC Group. Oxford University Press, Oxford, UK, pp. 3–34.

James, A., Pierowicz, J., Moskal, M., Hanratty, T., Tuttle, D., Sensenig, B., Hedges, B.: Assessing Consequential Scenarios in a Complex Operational Environment Using Agent-Based Simulation. U.S. Army Research Laboratory Technical Report (2017).

Kennedy, W.G., (2012). Modelling Human Behaviour in Agent-Based Models, in: Agent-Based Models of Geographical Systems, Edited by A. Heppenstall, A. T. Crooks, L. M. See, and M. Batty. Springer, New York, NY, pp. 167–179.

Lebiere, C. J., Stocco, A., Reitter, D., & Juvina, I. (2010). Scaling up high-fidelity cognitive modeling to real-world applications. Technical report, DTIC Document.

Lehman, J.F., Laird, J., Rosenbloom, P., (2006). A Gentle Introduction to Soar, An Architecture for Human Cognition: 2006 Update.

Malleson, N., See, L., Evans, A., Heppenstall, A.: Implementing comprehensive offender behaviour in a realistic agent-based model of burglary. Simulation 88(1), 50–71 (2012).

Mualla, Y.`, Najjar, A., Daoud, A., Galland, S. (2019). Agent-Based Simulation of Unmanned Aerial Vehicles in Civilian Applications: A Systematic Literature Review and Research Directions. Future Generation Computer Systems.

National Institute for Standards and Technology (NIST) (2008). "Autonomy Levels for Unmanned Systems Framework (ALFUS)," version 2.0, NIST Special Publication 1011-I-2.0.

Newell, A. (1994). Unified theories of cognition. Harvard University Press.

Orr, M. G., Lebiere, C., Stocco, A., Pirolli, P., Pires, B., & Kennedy, W. G. (2018, July). Multi-scale resolution of cognitive architectures: A paradigm for simulating minds and society. In International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (pp. 3-15).

Pires, B., Crooks, A.T.: Modeling the emergence of riots: a geosimulation approach. Comput. Environ. Urban Syst. 61, 66–80 (2017).

Rao, A.S., Georgeff, M.P., (1995). BDI Agents: From Theory to Practice, in: Proceedings of the First International Conference on Multiagent Systems, Edited by L. Gasser and V. Lesser. San Francisco, CA, pp. 312–319.

Reitter, D. & Lebiere, C. (2010). Accountable modeling in ACT-UP, a scalable, rapid-prototyping ACT-R implementation. In Proceedings of the 10th international conference on cognitive modeling, Philadelphia, PA.

Risco-Martın, J. L., Mittal, S., Zeigler, B. P., & Jesús, M. (2007, September). From UML state charts to DEVS state machines using XML. In Proceedings of the workshop on multi-paradigm modeling: concepts and tools, Nashville, TN.

Schmidt, B., (2002). The Modelling of Human Behaviour: The PECS Reference Models, in: Proceedings 14th European Simulation Symposium. SCS-Europe BVBA, Dresden, Germany.

West, R., Nagy, N., Karimi, F., Dudzik, K.: Detecting macro cognitive influences in micro cognition: using micro strategies to evaluate the SGOMS macro architecture as implemented in ACT-R. In: Proceedings of the 15th International Conference on Cognitive Modeling, pp. 235–236 (2017).

Wilensky, U. (1999). "NetLogo," website available at http://ccl.northwestern.edu/netlogo/ (accessed 2020-08-14).

Wilensky, U., & Rand, W. (2015). An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo. Mit Press.

Zeigler B., Muzy A., Yilmaz L. (2009). "Artificial Intelligence in Modeling and Simulation." In: Meyers R. (eds) Encyclopedia of Complexity and Systems Science. Springer, New York, NY.

AUTHOR BIOGRAPHIES

R. DOUGLAS FLOURNOY is a Principal Simulation Engineer for the Simulation, Experimentation, and Gaming Department at The MITRE Corporation. He holds an MS in Operations Research from the George Washington University. His email address is *rflourno@mitre.org*.

BIANICA PIRES is a Lead Modeling & Simulation Engineer for the Simulation, Experimentation, and Gaming Department at The MITRE Corporation. She received her PhD in Computational Social Science from the George Mason University. Her email address is *bpires@mitre.org*.

ROB WITTMAN JR. is the Department Manager of the Simulation, Experimentation, and Gaming Department at The MITRE Corporation. He received his PhD in Industrial Engineering from the University of Central Florida. His email address is *rwittman@mitre.org*.

ROB LAYDEN is a Project Leader and Group Leader at The MITRE Corporation. He received his PhD in Electrical Engineering from the University of Dayton. His email address is *rlayden@mitre.org*.