An Approach to Integrating Human Knowledge into Agent-Based Planning

Leah Kelley, Michael Ouimet, Bryan Croft, Eric Gustafson, Luis Martinez
Space and Naval Warfare Systems Center Pacific
San Diego, California, United States
leah.kelley@navy.mil

Abstract

For humans and machines to collaboratively accomplish complex tasks, each has to influence the others' decision-making (planning) processes. Artificial intelligence (AI) planning algorithms are often developed for simplistic scenarios that assume the world is static during plan generation, which is not valid in many real-world scenarios. When applied to more realistic scenarios, the AI planning algorithms suffer from state explosion from the added complexity and are unable to generate plans quickly enough to be practical. This research develops a method for injecting human knowledge into the AI planning process so that together humans and machine agents can perform complex missions. This approach is demonstrated by applying it to an underwater mine hunt in which a single underwater unmanned system searches for mines with human operator assistance. Preliminary results show the human can successfully influence the agent's planning.

1 Introduction

While serving as Secretary of Defense for the United States, Chuck Hagel expressed a desire for a truly collaborative human-machine fighting force as part of what is now known as the 3rd Offset Strategy [Hagel, 2014, Work, 2015]. The current operating paradigm in the field for the U.S. Military is, at best, a single sailor or soldier provides supervisory control to a single machine. Many aerial drone systems require at least two pilots, as well as additional personnel on the ground to support their missions. There is a big push for development of intelligent systems which incorporate more autonomy and for development of human-machine interfaces and controllers that permit a single human supervisory control over multiple machine assets. However, true human-autonomy teaming is more than just a human supervising several machines. It is a synergy of human and machine intelligence and decision-making. Machines can perform numerical computations and can solve complex optimization problems much more quickly than humans can. However, machine perception doesn't come close to human perception, especially when data quality is poor [Ali and Itti, 2014].

Humans are much better than machines at rapidly synthesizing a wide variety of sensory input, inference based on common sense that is hard to encode explicitly, and at making quick decisions in the face of uncertain and/or incomplete information [Ahmed, *et al.*, 2012]. So, given these specializations, how does one synthesize human and machine talents to synergistically accomplish complex missions?

Consider the way a machine interacts with its environment. It can react to sensory input, the way a "personal assistant" such as Apple's Siri does when asked a question by a human, or such as a mobile robot does when it swerves to avoid an obstacle it detects in its path. Autonomy can also achieve higher-level goals, such as determining a feasible route from one location to another, assigning mission tasks to unmanned vehicles or to groups of humans based on information about their skills and capabilities, or scheduling machining jobs. To achieve the higher level goals, the machine agent must plan. Based on an internal model of the world that includes all available actions the agent can take and their effects, as well as an initial state and a goal state, the machine's planning algorithm determines the sequence of actions required to transition from the initial to the goal state.

2 Related Work

The artificial intelligence (AI) community has devoted decades of research to developing and improving various types of planning algorithms, many of which are domain independent [Russell and Norvig, 2010, Long and Fox, 2002]. To use the algorithms, one must model the planning problem in the appropriate form, which often requires simplification of the problem if the planning algorithm is to successfully generate a plan. Since the computational complexity for AI planning problems can range from constant time to NEXPTIME-complete, with increasing complexity corresponding to fewer restrictions [Ghallab, *et al.*, 2004], the problems tend to be formulated as well constrained scenarios. It is a challenge to apply these techniques to highly complex scenarios where the simplifying assumptions no longer hold. One of the key assumptions in traditional AI planning is that while the planner generates a plan, the world is static. This assumption holds even in complex games, such as Go, since in games

there are elements of turn taking so while the planning algorithm is generating a decision, the world/environment it reasons over remains static. In most naval-relevant scenarios the world/environment is constantly changing, so planning algorithms that do not account for such changes will not necessarily produce useable plans.

There are methods of representing and reasoning over uncertainty. Contingency planning explores world domains modeled as "And-Or" graphs through online planning using depth-based search, random walk, hill climbing and Learning Real-Time A* (LRTA*) [Russell and Norvig, 2010]. Markov Decision Processes (MDPs) are applicable to fully observable, stochastic environments in which uncertainties in the environment, sensory data and effects of agent actions can be quantified [Boutilier, *et al.*, 1995, Kaebling, *et al.*, 1998, and Russell and Norvig, 2010]. Optimal policies for MDPs can be generated using dynamic programming. Partially Observable Markov Decision Processes (POMDPs) extend MDPs to scenarios in which the agent does not necessarily know its state, and such problems can be solved using an extension of value iteration [Kaebling, *et al.*, 1998], branch-and-bound and gradient ascent methods [Meuleau, *et al.*, 1999], and backchaining techniques [Boutilier, *et al.*, 1995]. Inference reasoning algorithms can solve partially observable stochastic processes modeled as Bayesian Non-Parametric Models (BNPs) [Campbell and How, 2015].

The methods described in the preceding paragraph often require more complex models with large state representations. When considering large state spaces, planning algorithms require significant time to perform computations or, in some cases, cannot generate plans. Real-time strategy games for the computer such as StarCraft capture the complexity of Navy-relevant scenarios and offer an avenue for developing and testing more complex autonomy algorithms. These games include opposing players that must make actions in real-time based on local partial knowledge provided by their own mobile units and it requires long term planning (games take one hour and early actions allow/preclude later actions). Recent success with reinforcement learning has renewed interest in developing AI algorithms that play these games to compete against humans or other AIs yet no breakthroughs have occurred to date.

For a human to team with an autonomous agent, the human will affect the agent's planning, and vice versa. Zilberstein defines semi-autonomous systems as "systems that require some degree of human intervention in order to complete a task" [Zilberstein, 2015]. The focus on semi-autonomous systems is on human intervention during task execution. Research in mixed-initiative interaction has addressed the problem of joint human-machine planning. The term *mixed-initiative* has been defined as "methods that explicitly support an efficient, natural interleaving of contributions by users and automated services aimed at converging on solutions to problems" [Allen, *et al.*, 1999]. *Mixed-initiative interaction* is also defined as "a flexible interaction strategy in which each agent (human or computer) contributes what it is best suited at the most appropriate time" [Allen, *et al.*, 1999]. In early mixed-initiative planning, "a computer acts as a collaborating assistant to the human, anticipating needs, performing the tasks it is well suited for, and for leaving the remaining tasks to the human" [Ferguson, *et al.*, 1996]. In the TRAINS-95 system, the designers intentionally used a simple, highly limited AI planner to force the human user to interact with it [Ferguson, *et al.*, 1996]. The MAPGEN mixed-initiative planning system incorporates a constraint-satisfying planner capable of planning one Martian day's worth of activities for the Mars Rovers based on goals defined by human mission scientists, but the plans it produced were not intuitive to the scientists judging the plans, hence the humans preferred to incrementally plan sets of goals while the AI agents handle constraint satisfaction [Bresina, *et al.*, 2005]. Both humans and the AI agents contribute to plan generation and checking. Tools such as VAL also provide plan validation and checking: given a domain model, a problem description and a plan, VAL determines whether the plan will succeed or fail. If a plan fails, VAL informs the user "what conditions must be achieved in order to repair [the failed plan]. It does not indicate which actions might be applied to achieve those conditions or explore the interactions they might introduce into the plan if they are added to it" [Howey, *et al.*, 2004]. The focus of mixed-initiative planning "is on visualization tools that allow people to participate in the planning process itself" [Zilberstein, 2015]. Additionally, the use of mixed-initiative planning circumvents the need to use complex domain models and encode human preferences, since human participation in plan generation is tightly integrated in such systems.

The hypothesis in this research is that human intelligence and ways of reasoning about uncertain, dynamic environments can be injected into the AI planning algorithms, so a human can help an algorithm arrive at a solution. Unlike a semi-autonomous system that needs intervention for task execution, the AI agent here can complete tasks, but may not generate a plan. This does not limit the human intelligence injection to high-level goal setting alone. In fact, in dynamic, uncertain environments injection of human intelligence and situational awareness during mission/task execution will be necessary and will trigger plan revision. However, unlike in mixed-initiative planning, the focus is on translating the human preferences and knowledge into a form the AI planning algorithm can use, rather than on the tools needed for the human to generate plans, as in mixed-initiative planning. The method is not limited to applications of single-agent, single human teaming, but is extensible to multi-agent, multi-human missions as well. One example is Advanced Chess, where cooperative hybrid human and AI teams compete in chess. The human helps direct the faster, more precise analysis of the autonomy producing a team that can outcompete human-alone or autonomy-alone competitors.

The rest of this paper is organized as follows. Section 3 describes the approach taken to mature AI planning algorithms from highly simplistic to more realistic scenarios by removing simplifying assumptions and introducing human intelligence to the resulting more-realistic, yet more-complex, problems. Section 4 describes the application of

this approach to a naval-relevant scenario of an underwater mine hunt carried out by a single unmanned underwater vehicle (UUV) and a human operator. In this application, an MDP is used to represent the domain because it provides the UUV's AI agent the ability to plan under a semi-realistic world representation, yet still suffers from shortcomings when applied to the mine hunt mission without human assistance. Section 5 covers preliminary results of the methodology as applied to the mine hunt scenario, and discusses some of the limits discovered. Section 6 provides a summary and concludes with future directions the work will take.

3 Approach

As discussed above, AI planning algorithms are often developed for well-constrained problems that only require low-fidelity domain models and contain simplifying assumptions. In this research, the complexity of domain models are systematically increased by removing simplifying assumptions one at a time, as shown in Figure 1. Removal of a simplifying assumption results in a higher-fidelity representation of how the world behaves, including the agent within it, but at a cost of increased model complexity. As a consequence, an AI planning algorithm that successfully solves a planning problem using a simplified representation may not be able to generate a solution using a more-complex domain model, since the number of possible states drastically increases as complexity increases. In some instances, a planning problem posed using a more-complex domain may be solvable if a different algorithm is used. In other instances, this may not be possible. This research hypothesizes that human intelligence can be injected into an AI planning algorithm in these instances to assist it with solution generation.
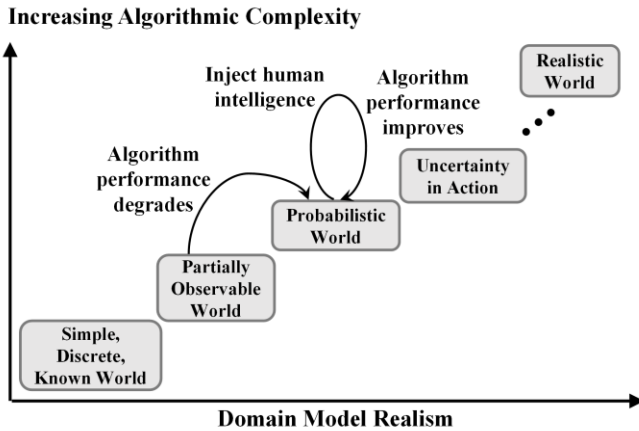

Increasing Algorithmic Complexity

Figure 1: Assumption removal process degrades algorithm performance; human intelligence facilitates solutions

In this paper, this method of systematically increasing model complexity to the point of AI planning algorithm failure, injection of human intelligence, and subsequent AI planning success is applied to a naval-relevant scenario, that of underwater mine hunting with an unmanned underwater

vehicle. Initially, all world and agent states are discrete, finite, known, observable and static during planning, and the planner's performance is assessed. A single assumption, e.g. locations of obstacles in an underwater map are known, is removed and the algorithm is applied to the planning problem in its slightly more-complex, more realistic representation. If the algorithm is still capable of solving the planning problem, another simplifying assumption is removed and complexity and realism are increased again. If not, and no other algorithm can solve the more-complex problem, human intelligence is injected to assist the planner. In this method, the human intelligence is essentially what is used to constrain the problem again, reducing the state space. Once the planner is able to solve the more-complex problem using human input, another assumption is removed, again increasing both realism and complexity of the domain representation. This process of assumption removal, degraded performance, human intelligence injection and algorithm improvement continues until the planner and human together are able to achieve complex missions using higher fidelity, more-realistic domain representations.

Here, planning scenarios are divided into two broad categories: navigation/detection and resource allocation/task scheduling/assignment. The approach presented here is applicable to both types of planning domains. Underwater mine hunting using teams of UUVs is an example of a planning scenario that includes both navigation/detection and task assignment. Here, the systematic process of removing assumptions to increase realism in domain representation is applied to a planar mine hunt at a level ocean depth, where a single UUV hunts for a single mine in a fixed region, while interacting with a human operator.

4 Application to UUV Mine Hunt

In this section, we consider a grid-based UUV mine hunt scenario. An agent maneuvers through an environment attempting to get to and disarm a mine while not colliding with obstacles, see Figure 2. In the simplest version, the UUV agent knows the locations of the mine and obstacles, its motion, localization, and sensing are perfect. Under these assumptions, a tree-based search algorithm such as A* will easily find the shortest path to the mine. Note, however, if one considered the state of the vehicle in 6 dimensional space (x, y, z, roll, pitch yaw), tree-based search algorithms would be unable to find optimal paths.

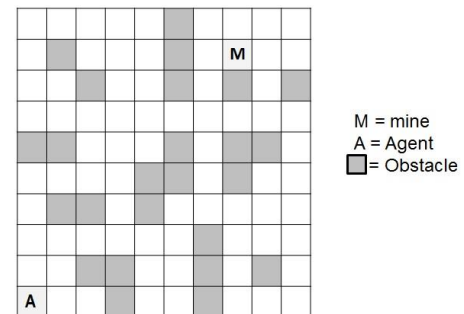M = mine
A = Agent
☐ = Obstacle

Figure 2: Mine-hunt grid world.

Several simplifying assumptions can be removed in this scenario without requiring human assistance, and the more complex representations can be managed by changing planning algorithms. When the agent has imperfect knowledge of its location in the map, this scenario becomes a simultaneous localization and mapping (SLAM) problem. Rather than inject human knowledge at this point, it makes sense to switch planning algorithms since the AI and robotics communities have already developed algorithms that solve this problem [Thrun, *et al.*, 2005]. Switching to a Markov Decision Process (MDP) representation of the problem enables removal of the following assumptions: known obstacle locations, perfect sensing, and perfect motion. This makes the representation more realistic, as one does not typically know where all the underwater obstacles are a priori, sensors are subject to noise and false readings, and unmodeled underwater currents affect the UUV's motion.

In this implementation, the underwater environment is still defined as a finite number of locations, as in the above grid world. The UUV agent moves between locations by choosing from a set of possible actions, although the location it ends up at given a specific action can be a random subset of the feasible locations. The agent's action choices are guided by a reward function that is defined over the world state: whether locations in the environment are free, obstacles, goals, unknown, and whether or not they have been visited. As the agent moves and takes sensor measurements, the world state is updated. The agent's overall goal is to find a policy that specifies which action to take at each state that maximizes the expected discounted reward the UUV receives when following the motion policy. MDPs assume that both the reward and transition probabilities at each state are known, and given that information, the MDP efficiently finds paths through the space to maximize reward subject to randomness. In this setup, the complete state space is too large to reason over, so the agent plans assuming a static reward function. As it moves and collects new measurements, it updates the world state, replanning with a new static reward function. This approach balances between fast execution and reasonable performance.

Since the reward for visiting unvisited locations is high, the initial reward function is uniformly large over the entire environment given that the agent has no prior knowledge of the search space. Thus, the agent wanders haphazardly through the environment. Here, a human operator has the opportunity to assist. A human can specify locations in the search environment that are of high or low interest by initially setting the rewards at such locations as high or low, respectively. Since the reward function is updated as the UUV agent moves, a human can also update the reward function values online during execution. For example, during a mine hunt, a human operator may receive reports via chat or telephone about suspicious activities at locations within the search environment that the UUV agent will not receive or cannot efficiently reason over. By increasing the reward for visiting some locations, the human operator guides the agent's motion choices without needing to take low-level control or teleoperate the UUV.

Central to human interaction with and manipulation of the UUV agent's reward function is the representation of that reward to the human. The team uses an existing research platform, the Intelligent UxV Planner with Adaptive Collaborative Control Technologies (IMPACT) [Lange and Reeder, 2017] as a visualization tool and human-operator interface. Here, two maps of the search region are displayed: an occupancy map of regions the agent has visited and a reward map depicting the agent's reward function. In the occupancy map, Figure 3, left, gray represents locations which the agent has not yet visited, black represents obstacles, and white represents traversable areas. In the reward map, Figure 3, right, a rainbow color gradient depicts reward values. Areas of high reward are red, and, at the other extreme, no-go areas (e.g. obstacles, user-defined keep-out regions) are dark blue.
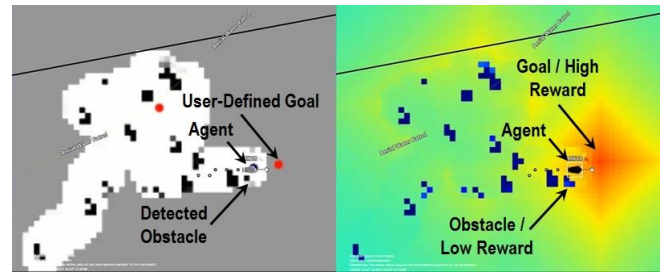

Figure 3: Occupancy map, left and reward function map, right.

A human operator manipulates the agent's reward function using a graphical user interface, Figure 4. To define a point of interest, the user can click on the color map and then on an "Increase Intensity" button. Similarly a low-priority location can be defined by clicking on the map and then on the "Decrease Intensity" button. The user is not required to understand the mathematical underpinnings of the agent's reward function in order to adjust it.
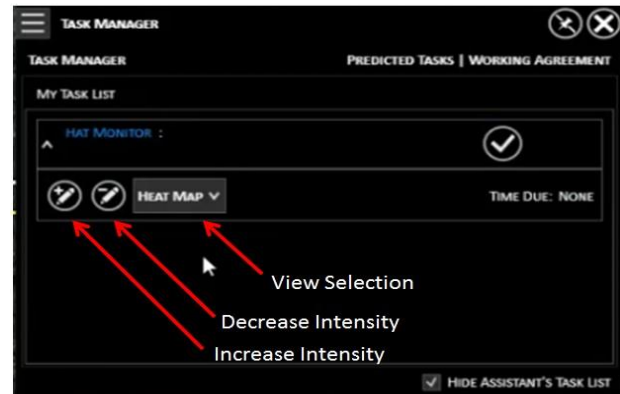

Figure 4: Controls for human operator.

The MDP representation has been extended to relax the assumption of perfect localization by transforming it into a Partially Observable Markov Decision Process (POMDP)

[Thrun, *et al.*, 2005]. Here, the agent reasons over a belief space based on local, noisy measurements and dynamics models to estimate where it is and where objects of interest are. Under this representation, the state space increases dramatically. Methods of handling this state explosion are discussed in the next section.

5 Preliminary Results and Analysis

Qualitative results of human teaming with the UUV agent under the MDP instance show that simply defining goal locations *a priori* forces the agent to explore those areas first and results in a much different search pattern, and demonstrates that human interaction can affect the agent's planning. When the human is able to change the weights during the search process, the agent will respond and change its search direction if the new location has higher value. The effectiveness of teaming with the human to find mines still has to be measured quantitatively. However, provided that the human's information is accurate (e.g. the mine is where the human believes it is), the UUV agent will encounter it more quickly than it would when following a fixed lawn-mower sweep pattern or by simply exploring the search environment in a haphazard way. Note that as such, this method is not robust to error in human judgement or misinformation.

One way to mitigate human error over time is to incorporate activity history into the planning agent's reasoning. The IMPACT system incorporates a provenance service. Provenance is "information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness" [Groth and Moreau, 2013]. In the UUV mine hunt scenario, the provenance service keeps track of the meta-data associated with events, such as the time a goal was defined by a human user, the confidence the human operator has in the goal location (based on the value the human placed on the goal), and the machine's confidence in the human operator (based on past results of human intelligence). The agent's reward function can be modified so that meta-data information is given weight, based on its recency and/or other factors. The reward function can be tailored to achieve a balance between trust in information gathered from the UUV's onboard sensors and trust in information provided by the human operator.

As described in Section 3, reformulating the UUV mine hunt into a POMDP planning problem to remove the perfect localization assumption results in a state explosion, which is not unexpected [Thrun, *et al.*, 2005]. Ways to overcome the state explosion include limiting the length of time that can be reasoned over, and changing the abstraction for what constitutes a 'state' and 'action'. In the second approach, consider how one might plan how to get dressed. It easier to realize that one should put on socks before shoes when the actions are put on shirt, put on shoes, put on socks, etc., rather than planning over the muscle firings that may achieve those outcomes. Similarly, with robotics, planning can be done at higher levels of abstraction, as long as lower level controls can be assumed to carry forth the high level plans.

With this idea in mind, the cooperative human-UUV mine hunt scenario can be decomposed into a hierarchical planning structure, which can then be expanded to address multi-agent mine hunting. The hierarchical planning structure and levels within the hierarchy are shown in Figure 5.
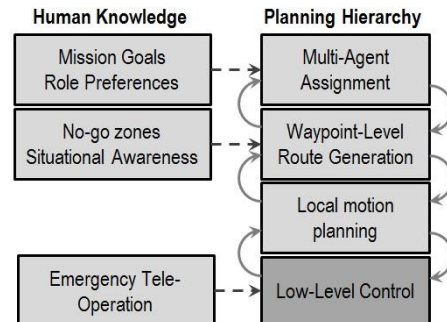

Figure 5: Hierarchical planning concept with human knowledge injection.

At the highest level, mission goals and role preferences expressed by the human operator are taken into consideration by the mission-level planner that performs multi-agent assignment. In a single UUV scenario, a mission goal is still defined. Human knowledge is also considered when generating overall routes. During mission execution, human knowledge can be injected at both upper levels, since other events (e.g. movement of a friendly vehicle into the search environment, reports of suspicious activity) may change both search goals and no-go zones. Except for emergency tele-operation, human interaction at the lower planning levels is impractical or may be impossible due to latency delays. Local motion planning, which governs reactive behaviors such as obstacle avoidance, occurs at high frequencies and is sufficiently small physical regions that human knowledge provides little benefit. Also, note that information is exchanged between planners throughout the hierarchy, so the knowledge injected by the human operators does influence all the planning, albeit indirectly at the lower levels.

6 Summary

In this research, cooperative human-machine (AI) solution generation to complex problems can be achieved by systematically removing assumptions and injecting human intelligence into the AI agent's planning process. The method is applied to a human-UUV cooperative underwater mine hunt scenario, and preliminary qualitative results show that a human can influence the AI agent's planning process. Results also show that removing assumptions to increase model realism can require changing algorithms and/or problem structure, though this does not omit the need for human knowledge injection. In the UUV-mine hunt example, removing the perfect localization assumption will be achieved by decomposing the planning strategy into a hierarchical structure. Quantitative results on the effectiveness of human assistance still need to be collected.

Future work in the human-UUV underwater mine hunt domain will expand the scenario to multi-agent search, and will incorporate a hierarchical planning structure. Human knowledge and assistance will be directly applied to higher levels of planning, yet will affect all planning levels. Additionally, incorporation of provenance data into the AI agent's planning process can be incorporated to mitigate human error.

Acknowledgments

References

[Ahmed, *et al.*, 2012] Nisar R. Ahmed, Eric M. Sample and Mark Campbell. Bayesian Multicategorical Soft Data Fusion for Human-Robot Collaboration. *IEEE Transactions on Robotics*, 29(1): 189-206. February 2013.

[Ali and Itti, 2014] Ali Borji and Laurent Itti. Human vs. computer in scene and object recognition. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 113-120. Columbus, OH, June 2014.

[Allen, *et al.*, 1999] James F. Allen, Curry I. Guinn, and Eric Horvtz. Mixed-initiative interaction. *IEEE Intelligent Systems and their Applications* 14(5): 14-23. September/October 1999.

[Boutilier, *et al.*, 1995] Craig Boutilier, Thomas Dean and Steve Hanks. Planning under uncertainty: Structural assumptions and computational leverage. In *Proceedings of the Second European Workshop on Planning*, pp. 157-171. September 1995.

[Bresina, *et al.*, 2005] John L. Bresina, Ari K. Jónsson, Paul H. Morris and Kanna Rajan. Mixed Initiative Planning in MAPGEN: Capabilities and Shortcomings. In the *Workshop on Mixed-Initiative Planning and Scheduling* at the *International Conference on Automated Planning and Scheduling (ICAPS 2005)*, pp. 54-61. Monterey, CA. 5-10 June 2005.

[Campbell and How, 2015] Trevor Campbell and Jonathan P. How.Bayesian nonparametric set construction for robust optimization. In *Proceedings of the 2015 American Control Conference (ACC)*, pp. 4216-4221. Chicago, IL, 1-3 July 2015.

[Ferguson, *et al.*, 1996] George Ferguson, James F. Allen and Bradford W. Miller. TRAINS-95: Towards a Mixed-Initiative Planning Assistant. In *Proceedings of the Third International Conference on Artificial Intelligence Planning Systems (AIPS 1996)*: pp. 70-77. AAAI, Edinburgh, Scotland, 29-31 May 1996.

[Ghallab, *et al.*, 2004] Malik Ghallab, Dana Nau and Paolo Traverso. *Automated Planning: theory and practice*. Morgan Kaufmann Publishers Inc., San Francisco, CA, 2004.

[Groth and Moreau, 2013] Paul Groth and Luc Moreau. PROV-Overview: An Overview of the PROV Family of Documents. Technical report. W3C Recommendation. World Wide Web Consortium, 2013.

[Hagel, 2014] Charles T. Hagel. "Reagan National Defense Forum Keynote." Reagan National Defense Forum. Ronald Reagan Presidential Library, Simi Valley, CA. Keynote Address. 15 November 2014.

[Howey, *et al.*, 2004] Richard Howey, Derek Long and Maria Fox. VAL: Automatic Plan Validation, Continuous Effects and Mixed Initiative Planning using PDDL. In *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2004)*: pp. 294-301. IEEE, Boca Raton, FL, 15-17 November 2004.

[Kaebling, *et al.*, 1998] Leslie Pack Kaelbling, Michael L. Littman and Anthony R. Cassandra. "Planning and acting in partially observable stochastic domains." *Artificial intelligence* 101(1): 99-134. May 1998.

[Lange and Reeder, 2017] Douglas S. Lange and John D. Reeder. IMPACT machine learning efforts. In *Proceedings of SPIE Unmanned Systems Technology XIX*. 11 April 2017.

[Long and Fox, 2002] Derek Long and Maria Fox. Progress in AI planning research and applications. *UPGRADE: The European Journal for the Informatics Professional,* 3(5):10–25. October 2002.

[Meuleau, *et al.*, 1999] Nicolas Meuleau, Kee-Eung Kim, Leslie Pack Kaelbling and Anthony R. Cassandra. 1999. Solving POMDPs by searching the space of finite policies. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pp. 417-426. Morgan Kaufman Publishers. Stockholm, Sweden, 30 July 1999.

[Russell and Norvig, 2010] Stewart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 3rd Ed. Prentice Hall, Upper Saddle River, New Jersey, 2010.

[Thrun, *et al.*, 2005] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic robotics*. MIT press, Cambridge, Massachusetts, 2005.

[Work, 2015] Robert Work. "The Third U.S. Offset Strategy and its Implications for Partners and Allies," Willard Hotel, Washington, D.C. 28 January 2015.

[Zilberstein, 2015] Shlomo Zilberstein. Building Strong Semi-Autonomous Systems. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pp. 4088-4092. Austin, TX, 29-30 January 2015.