Why Reasoning Under Uncertainty Is Hard for Both Machines and People—and an Approach to Address the Problem

Chapter Ten

EDWARD GEIST, RAND CORPORATION

This chapter is extracted from *Adaptive Engagement for Undergoverned Spaces: Concepts, Challenges, and Prospects for New Approaches*, by Aaron B. Frank and Elizabeth M. Bartels, eds., RR-A1275-1, 2022 (available at www.rand.org/t/RRA1275-1).

Prepared for the Defense Advanced Research Projects Agency,
Defense Sciences Office
Approved for public release; distribution unlimited

RAND
CORPORATION

For more information on this publication, visit **www.rand.org/t/RRA1275-1**.

Published by the RAND Corporation, Santa Monica, Calif.

CHAPTER TEN

Why Reasoning Under Uncertainty Is Hard for Both Machines and People—and an Approach to Address the Problem

Edward Geist, RAND Corporation

Reasoning under uncertainty has always been central to decisionmaking. For example, under-governed spaces (UGS) involve uncertainty by definition. In conducting decisionmaking for UGS, actors lack complete knowledge of what is going on, who is doing what, and the effects of their own and others' actions. This is also true of all four parts of the Act-Sense-Decide-Adapt (ASDA) cycle, which require reasoning with uncertain knowledge. When we try to *sense* something, we have to account for the possibility of erroneous and misleading stimuli, and we hedge against uncertainty about the true state of the world. As Descartes fretted four centuries ago, a powerful, malevolent foe may have constructed an entire world out of illusions to deceive us. When we *decide* what action to take, we have to account for uncertainty about the likely effects of the available actions and even about what actions are available to us. When weighing the merits of these actions, we must also grapple with uncertainty about both our own and others' current and future preferences. It is very common to have no more than an educated guess about what a rival wants, but much of the time we also have difficulty specifying our own desires. When we *adapt,* we have to reason about the uncertain possible futures we are choosing between. An unnerving aspect of this is the possibility of being tripped up by what Donald Rumsfeld dubbed "unknown unknowns." And when we *act,* we have to juggle all these types of uncertainty at the same time. However, time is not a luxury that we have when acting, sensing, deciding, or adapting. When facing an intelligent adversary, we may have to act quickly—even if we have not been able to think through everything fully.

Because reasoning under uncertainty is central to creating machines that exhibit intelligent behavior, reasoning under uncertainty is one of the most-studied problems in artificial intelligence (AI). Unfortunately, despite all this effort, no entirely satisfactory way to reason under uncertainty has yet emerged; however, efforts to find one have yielded considerable theoretical insights into the problem, as well as a wide variety of experimental systems. These programs use a variety of alternative techniques with associated strengths and weaknesses. Those that excel in some respect, such as *expressiveness*—the ability to describe a large number of different situations—come with a weakness, such as prohibitive computational demands. Because AI researchers have tried to translate all these different approaches for reasoning about uncertainty into engineering, they arguably gleaned deeper insights into the challenges involved compared with less empirical investigators. Perhaps their most essential finding is that there is no single "best" or "right" way to reason under uncertainty; this is both because of trade-offs between characteristics, such as computational complexity and accuracy, and because there is more than one way to be uncertain. For example, sometimes one is uncertain about whether a proposition is true, but other times one is uncertain about the degree to which a proposition is true. To reason comprehensively about uncertainty, it is necessary to be able to account for many qualitatively different species of uncertainty simultaneously.

In this chapter, I first examine the challenges that AI researchers have encountered by using the approaches they have used historically. Then, I discuss the fundamental ontological challenges that these approaches face. This is followed by the implications these challenges have for national security decisionmaking. Given all these obstacles, I suggest a proposed research agenda going forward. The chapter ends with some concluding thoughts.

Challenges That AI Researchers Face in Making Machines Reason Under Uncertainty

Broadly speaking, AI researchers have developed two approaches to reason under uncertainty that can be classified into two paradigms: Bayesian and non-Bayesian. We discuss each in turn, along with challenges that have been encountered, followed by a discussion of ontological challenges common to both.

The Bayesian Paradigm and Its Challenges

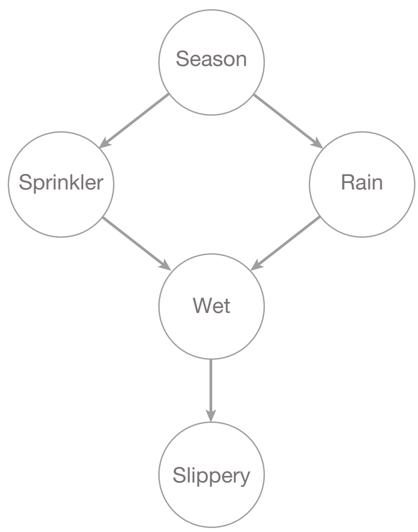
Bayesian approaches represent knowledge about a set of state variables as probabilities. These state variables can be either *discrete* (for instance, 50 percent confidence that a proposition is true) or *continuous* (such as a probability density function representing the likelihood that a variable takes a particular value). These variables are initialized to a prior (starting estimate) and then updated using Bayes' rule when new evidence is received. The basic version of Bayesian reasoning uses the full joint probability, accounting for possible correlations between all the variables. This fundamental approach is almost never used for nontrivial problems because combinatorial explosion rapidly inflates the size of the joint probability table to an unmanageable size.[1] Thankfully, in most use cases, the bulk of the variables are either weakly or totally uncorrelated, which enables systems to use a subset of the full joint probabilities in the form of conditional probabilities. A simplistic, but often useful, version of this is "naïve

[1] For a canonical discussion, see Stuart J. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach,* 2nd ed., Upper Saddle River, N.J.: Pearson Education, Inc., 2005, Ch. 13.

Bayes," which simply assumes that all variables are uncorrelated. But most of the time, a moderate number of significant correlations need to be accounted for to attain good results.[2]

Bayesian belief networks—often shortened to "belief nets"—emerged as the predominant solution for these more-complex problems. One of the major advances that came out of AI research during the 1980s, belief nets use an acyclic directed graph to represent the correlations between variables (see Figure 10.1). Not only does this scheme provide compact representations, it allows efficient inferences that consider only those conditional probabilities relevant for a particular query, while ignoring the remainder of the graph.[3] Processes that evolve over time, such as tracking, can be analyzed using a derivative method, the Dynamic Bayesian Network (DBN). Researchers have shown that many of the tools used in tracking and information fusion before the specification of belief nets in the 1980s—for instance, Kalman filters—are actually DBNs.[4] This allows for formal analysis of the computational complexity and tractability of these tools.

FIGURE 10.1
A Simple Bayesian Network (five variables)



SOURCE: Derived from Pearl and Russell, 2000.

---

2 For a useful comparison of naïve Bayesian methods with more-sophisticated derivatives, see Pedro Domingos, "A Few Useful Things to Know About Machine Learning," *Communications of the ACM*, Vol. 55, No. 10, 2012.

3 Judea Pearl and Stuart J. Russell, *Bayesian Networks*, Los Angeles, Calif.: University of California, November 17, 2000.

4 Vladimir Pavlovic, James M. Rehg, Tat-Jen Cham, and Kevin P. Murphy, "A Dynamic Bayesian Network Approach to Figure Tracking Using Learned Dynamic Models," in *Proceedings of the Seventh IEEE Interna-*

A limitation of Bayesian methods is that their underlying knowledge representation is inherently propositional—a particular variable can be true or false, or, if it is continuous, it is assumed to have one and only one true value. To reason about complex dynamic processes—for example, the surveillance of a region containing an unknown number of targets of interest—one must extend the Bayesian paradigm to consider multiple worlds.[5] For the target tracking case, these multiple worlds could include worlds with various numbers of targets in the observation area, each of which becomes, in turn, a proposition to be evaluated according to Bayes' rule.

Proponents argue that Bayesian models are more well specified, comprehensible, and correct than available alternatives, but for decades AI researchers hesitated to embrace Bayesian approaches. They had two reasons—one practical and the other theoretical. Until Judea Pearl introduced the belief net, there was no efficient way to conduct Bayesian inference and updating for a nontrivial problem. However, in the 1960s and 1970s, AI researchers mostly worked outside Bayesian models because of their observation that human reasoning seemed not to be based on probability. Instead, humans appeared to rely on simple heuristic methods, such as "default reasoning," which basically consists of assuming that a proposition is true until a seemingly more plausible one comes along, at which point it becomes the new default hypothesis. Observational psychology offered support for this idea, so AI researchers prototyped experimental systems based on it.[6]

The Non-Bayesian Paradigm and Its Challenges

Just because a means of reasoning about uncertainty is not based on probability does not mean it is necessarily non-rigorous. Lotfi Zadeh's possibility theory provides a concrete example of a non-probabilistic system for these purposes. Zadeh's fuzzy logic posits the existence of sets in which an object can be a partial member; an item can be half inside the set associated with a concept or conclusion. This is distinct from the probabilistic representation used in Bayesianism, in which a proposition can only be true or false.[7] One benefit of this alternative ontological commitment is that fuzzy sets can naturally represent concepts that are difficult to express using probability alone. For example, say we wish to determine whether a person is in England or in Scotland. In a Bayesian context, we can assign probabilities that the person is in England or in Scotland and update those probabilities when new evidence becomes available. However, if the person is straddling the border, then they are partially in both countries at the same time. This sort of partial set membership need not be associated with uncertainty:

tional Conference on Computer Vision, Vol. 1, IEEE, 1999.

[5] Russell and Norvig, 2005, pp. 519–522.

[6] See Raymond Reiter, "A Logic for Default Reasoning," *Artificial Intelligence*, Vol. 13, Nos. 1–2, 1980; and David Poole, "A Logical Framework for Default Reasoning," *Artificial Intelligence*, Vol. 36, No. 1, 1988.

[7] Lotfi A. Zadeh, "Fuzzy Logic and Approximate Reasoning," *Synthese*, Vol. 30, Nos. 3–4, 1975.

Perhaps we can see the person and know for certain what proportion of their body is on each side of the border.

Dempster-Shafer theory is another non-Bayesian approach for reasoning under uncertainty that has found substantial use for defense applications. Also known as the theory of belief functions, Dempster-Shafer theory originated as an attempt to introduce an interval-valued alternative to Bayes' rule. In an influential 1976 book, Glenn Shafer reinterpreted Dempster's original mathematics to represent what he dubbed "belief" and "plausibility" instead of bounds on an interval of Bayesian-like probability values.[8] Shafer argued that his approach transcended Bayesianism by providing a natural mechanism for representing the concept of "ignorance." The semantics of Dempster-Shafer theory and its relationship to the Bayesian paradigm are both controversial. Some critics argue that Shafer's characterization of these two values as "belief" and "plausibility" is misleading, irrespective of the soundness of the underlying mathematics. And while proponents tend to emphasize that Dempster-Shafer theory is equivalent to Bayes' rule when "belief" is equal to "plausibility," skeptics contend that the theory is qualitatively distinct from—and inferior to—the older Bayesian paradigm it sought to extend.[9] They are bolstered in this conclusion by the sometimes counterintuitive behavior of Dempster's rule, which can violate common sense when combining evidence from two sources that consider different possibilities, because the rule deletes any proposition that a source gives a probability of zero.[10] Proponents argue that these criticisms are overstated and that the use of Dempster-Shafer theory in practical systems suggests that it is still useful even if imperfect.

The Challenges of Using Such Approaches in Practice

Despite the theoretical distinction between Bayesian and non-Bayesian approaches for reasoning under uncertainty, real implemented systems tend to incorporate elements of both. A dominant technique for tracking multiple targets, the Multiple Hypothesis Tracker (MHT), exemplifies this. The MHT aims to associate every possible detection of a target with one and only one source. It does this by maintaining a collection of single-target recursive Bayesian filters (Kalman filters) and computing a value for each track. Each valid assignment of all detections to possible tracks constitutes a hypothesis, which explains the name of MHT.[11]

8 Glenn Shafer, *A Mathematical Theory of Evidence*, Princeton, N.J.: Princeton University Press, 1976.

9 Albena Tchamova and Jean Dezert, "On the Behavior of Dempster's Rule of Combination and the Foundations of Dempster-Shafer Theory," in *2012 6th IEEE International Conference Intelligent Systems*, IEEE, 2012.

10 Lotfi A. Zadeh, "Review of *A Mathematical Theory of Evidence*," *AI Magazine*, Vol. 5, No. 3, 1984.

11 Stefano Coraluppi, "Fundamentals and Advances in Multiple-Hypothesis Tracking," in *NATO STO IST-134 Lecture Series on Advanced Algorithms for Effectively Fusing Hard and Soft Information*, Paris, France: NATO Collaboration and Support Office, 2015. Although the introduction of the Kalman filter predated belief nets by over two decades, Kalman filters were later shown to be a variety of DBNs, and, today, a sizable literature exists analyzing them. See Pavlovic et al., 1999, and Kevin P. Murphy, *Switch-*

In this sense, MHTs are also DBNs in that they are made of many DBNs "hooked together." However, although in theory one can specify a fully Bayesian MHT as a hybrid DBN (that is, one with both discrete and continuous variables), such a system is not practical because its processor and memory requirements would rapidly balloon to astronomical dimensions. Practical MHTs use non-Bayesian mechanisms, such as pruning low-value hypotheses and gating (assuming that any detection beyond a certain distance from the expected position of a target being tracked cannot be associated with that target), to reduce the number of possible track-to-target associations that are considered and keep computational demands manageable.[12] Such mechanisms often work well in practice but tend to be designed on an ad hoc basis without a rigorous theoretical rationale. One consequence of these design compromises is that theoretical analysis of systems incorporating both Bayesian and non-Bayesian elements can be impractical.

The computational complexity of belief nets, by contrast, has been extensively studied by computer scientists—but their findings are very sobering. Theoretical studies have shown that exact inference in arbitrary belief nets is NP-hard relative to the size of the network.[13] This means that we cannot reasonably expect to be able to find the exact answer to a query of a Bayesian network of nontrivial size, even if we assume an arbitrarily powerful computer. This is not necessarily a showstopper in and of itself, because very effective algorithms exist to find approximate solutions to some NP-hard problems. The more damning finding is that accurate and efficient approximation algorithms for inference in arbitrary Bayesian networks of the classes of interest apparently cannot exist either. Approximate inference in Bayesian networks turns out to be NP-hard as well; furthermore, there is no guarantee that the resulting approximations will be close enough to the true values to be informative.[14] Perhaps this is to be expected, because in a cosmic sense, such an approximation would be too good to be true. Given the ubiquity of phenomena that can be stated as belief nets, that approximation would be applicable to a mind-boggling array of diverse problems and could make a revolutionary impact on both science and engineering. In the 1990s, theoretical computer scientists

ing Kalman Filters, Berkeley, Calif.: Department of Computer Science, University of California, Berkeley, August 1998.

[12] Samuel Blackman and Robert Popoli, *Design and Analysis of Modern Tracking Systems*, Norwood, Mass.: Artech House, 1999, pp. 1072–1075.

[13] The *NP* in *NP-hard* stands for *nondeterministic polynomial* and refers to a kind of hypothetical computer that computer scientists use to theorize about computational complexity. This nondeterministic computer could explore multiple branches of a search tree at the same time, which is of interest because it could find the answer to any query in the same amount of time it would take to check whether that answer was correct. *NP-hard* is a broad category of problems that are at least as hard as the hardest problems in NP and would be hard in the sense that even the physically unrealizable nondeterministic computer would take a long time to solve them (Gregory F. Cooper, "The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks," *Artificial Intelligence*, Vol. 42, Nos. 2–3, 1990).

[14] Paul Dagum and Michael Luby, "Approximating Probabilistic Inference in Bayesian Belief Networks Is NP-Hard," *Artificial Intelligence*, Vol. 60, No. 1, 1993.

set out for this particular El Dorado, only to come back with proofs that the fabled city of gold probably could not exist.[15]

This is not to say that approximate inference in Bayesian networks is impossible. If it were, then Bayesian networks would be practically useless. Instead, approximate inference in belief nets must make assumptions about the structure and/or state of the underlying problem to get good results with a feasible expenditure of theoretical resources. This restriction makes intuitive sense when one considers just how diverse the set of arbitrary Bayesian networks truly is: The graph structures that form the "backbones" of these belief nets are the set of all directed acyclic graphs. One cannot reasonably expect an efficient algorithm to exist that would work on all these structures. By contrast, when one restricts one's attention to certain classes of belief nets, it becomes evident that one should expect efficient algorithms to exist for some of them—for instance, those whose underlying graphs are singly connected (each node is connected to no more than one other node). Instead of relying on one unfailing "master algorithm," one is obligated to develop an endless series of tailored algorithms, making appropriate trade-offs for their use cases.

Ontological Challenges

Beyond the challenges already listed, there are also ontological ones. To reason about uncertainty, we need to represent it somehow—but how are we supposed to represent something we are uncertain about? The problem of choosing appropriate systems of knowledge representation (or *ontologies*, as they are termed by AI researchers) remains an unsolved one. A complete ontology encompasses both a knowledge representation language (symbolic in most historical systems, but often incorporating learned vector embeddings in modern ones) and a semantic interpretation mapping that language to external entities. The reason that researchers have advocated so many alternative approaches is that these have meaningful differences that make them more or less fit for specific use cases. One important distinction between some major approaches is that they make different ontological commitments—that is, "truth" actually does not mean the same thing in them. For example, in a Bayesian context, probabilities are maintained about whether a proposition is true, but that proposition is assumed to be either true or false, with no possibility that the proposition is simultaneously

---

[15] Uri N. Lerner and Ron Parr, *Inference in Hybrid Networks: Theoretical Limits and Practical Algorithms*, arXiv, 2013. The analysis in Lerner's 2002 dissertation is based on a kind of DBN called a Continuous Linear Gaussian (CLG), which uses both discrete variables and continuous variables, with restrictions, such as that the continuous variables must be Gaussian and that a discrete node cannot have a continuous parent. Lerner proved that unless P = NP, even approximate inference in CLGs is intractable, and, more surprisingly, no polynomial approximate inference algorithm could have an absolute error smaller than 0.5. CLGs can be embedded into the more general DBNs of interest for most defense applications, so these pessimistic complexity and approximability results should be expected to apply to them as well; see Chapter Four of Uri N. Lerner, *Hybrid Bayesian Networks for Reasoning About Complex Systems*, dissertation, Stanford University, 2002.

true and false.[16] By contrast, Lotfi Zadeh's fuzzy logic, which exists in a non-Bayesian context, allows variables to simultaneously be partially in more than one category. Sometimes the ability to represent that kind of uncertainty seems essential because of a need to reason about intermediate cases—like in the location example. Dempster-Shafer theory, meanwhile, aims to add a way to reason about ignorance, which its proponents assert is distinct from uncertainty per se. It seems that a truly comprehensive ontology would need to encompass all these forms of uncertainty and more.

Another, perhaps more important, reason that no single scheme for knowledge representation has found universal acceptance is that none has proved decisively superior in practical applications. Knowledge representation was one of the hottest areas of AI research during the expert systems boom of the 1970s and 1980s, resulting in some significant theoretical insights into the relevant problems. Unfortunately, one of the outcomes of this research was the discovery that it is not possible to find one ideal system of knowledge representation that is both representative enough to account for everything we would like while simultaneously being practical. Studies of simple knowledge representation languages showed that in any language capable of nontrivial representativeness, inference for some queries was computationally intractable.[17] The consequences of this are profound: Knowledge representation languages need to be well adapted for their particular use cases to ensure that these intractable inferences do not need to be made in practice. This has particularly tricky implications for reasoning about uncertainty because it suggests that it may be necessary to modify or replace the knowledge representation language dynamically to keep it performant as knowledge is updated.

A fruitful way to think about the challenge of choosing an appropriate ontology for reasoning about uncertainty is using the "many worlds" metaphor used in the Bayesian paradigm. When we reason about uncertainty, we are considering a set of possible worlds in which the evidence available to us has different implications. But even for some relatively mundane problems, this set of possible worlds is noncountably infinite. In the general case, we need to reason about countless possible worlds—but obviously this is impractical, because it would demand infinite computational resources. Somehow, we must select an efficacious subset of possible worlds and discard those that can be disregarded without too much risk. But how are we to do this when we do not know what it is that we do not need to worry about?

To revisit the target tracking example mentioned earlier, say we are tracking an unknown number of targets in a defined area. Obviously, we need to define variables about whether each potential detection is from an actual target of interest, which of those detections should

[16] In a continuous case, a probability density function describes the probability that the variable takes on a particular value—but the variable is assumed to have one and only one "true" value. Russell and Norvig, 2005, p. 524.

[17] Hector J. Levesque and Ronald J. Brachman, "A Fundamental Tradeoff in Knowledge Representation and Reasoning," in Ronald J. Brachman and Hector J. Levesque, eds., *Readings in Knowledge Representation*, Los Altos, Calif.: Morgan Kauffmann, 1985.

be assembled into track histories, and so on. But while it is straightforward to represent these "known unknowns," there are also "unknown unknowns" that could be critically important. For instance, there could be targets in the scene that are never detected, which might be moving along innumerable alternative trajectories through the surveillance region. Possibilities of this kind presumably should be weighted against one's confidence that such a target could exist without being detected. However, doing so is not straightforward because these "unknown unknowns" are difficult to count. Given the difficulties of merely tracking detected targets, most real-world tracking systems forgo any attempt to reason about undetected targets. However, these are not merely an academic problem but an acute threat on the contemporary battlefield. For example, cruise missiles are low-visibility airframes that are designed and operated with the aim of avoiding detection.

To overcome these challenges, a robust system for reasoning under uncertainty will need to be equipped with a dynamic ontology that can be modified and extended on the fly to learn new concepts and remain computationally efficient. This seems to be how humans grapple with uncertainty. As we gain familiarity with a novel situation, we often come to conceptualize it in very different terms than we did initially. Moreover, we often invent new concepts and heuristics to navigate this uncertain situation. Contrast this with the classic Bayesian formalism, in which all propositions must be known at the outset and nothing outside the support of the prior can ever be learned.[18] Prominent AI researchers, such as Douglas Hofstadter, suggested decades ago that creating machines with "general" intelligence will probably require endowing them with a similar ability to be introspective and self-modify their own ontologies.[19] Despite this, to date only a handful of AI systems have evinced even a token ability to do this, and it does not appear to be a target of much active research.[20]

Implications of the Challenges for National Security

These challenges suggest some far-reaching, and sometimes counterintuitive, implications for information fusion and other forms of reasoning under uncertainty in defense applications. For instance, they indicate that more sensors are not necessarily better; they may very well turn out to be worse. In the abstract, it seems intuitive that, all else being equal, more sensors should increase the likelihood of reconstructing the state of the environment accurately. This intuition is true in a cosmic sense but does not apply when we must account for

[18] Andrew Gelman and Cosma Rohilla Shalizi, "Philosophy and the Practice of Bayesian Statistics," *British Journal of Mathematical and Statistical Psychology*, Vol. 66, No. 1, 2013.

[19] Douglas R. Hofstadter, *I Am a Strange Loop*, New York: Basic Books, 2007.

[20] Perhaps the most famous historical example was Douglass Lenat's EURISKO heuristic concept discovery system, which in some versions was endowed with the ability to introspect and modify its own source code dynamically. Douglas B. Lenat and John Seely Brown, "Why AM and EURISKO Appear to Work," *Artificial Intelligence*, Vol. 23, No. 3, 1984.

the computational costs of reasoning under uncertainty.[21] Increasing the number of sensors increases the number of evidence variables, but the computational complexity of the task increases supralinearly with the total number of variables. This implies the existence of a turnover point, where adding more evidence variables ceases to make a marginal improvement in reasoning quality and actually starts to impair it. This point of diminishing returns can also be expected to have additional peculiar properties. It seems obvious that a small amount of high-quality evidence might be preferable to a large amount of noisy, low-quality evidence. However, accounting for the theoretical computational complexity of uncertain reasoning, it appears that a moderate amount of good-quality evidence could be better than a larger amount of good-quality evidence.

This is a particularly worrisome possibility given our theoretical understanding of the underlying computational problems: Some of them are known to belong to complexity classes potentially more imposing than the familiar NP.[22] That implies that the point of diminishing returns for additional evidence may present itself far sooner than we might expect. These considerations are particularly damning for the common vision of relying on networking and computers to turn data from cheap ubiquitous sensors into splendid situational awareness: From a theoretical standpoint, such a scheme appears somewhere between technically implausible and practically impossible.

These obstacles grow even more imposing when we consider the knowledge quality problems associated with the available evidence. In practice, we often do not know what, if any, of the evidence available to us is actually good quality. This prevents us from simply starting with the best evidence and incorporating more as time and resources permit. We must also consider the possibility that some of the evidence is not merely of poor quality but actively pernicious. While in contrived scenarios we can make convenient assumptions that evidence will have known or zero bias and regular noise, reality tends to be less felicitous. Adversary disinformation obviously falls into this category, but sometimes very misleading evidence results from natural processes. Even in the absence of adversary action, many organizations exhibit a tendency to process random noise as signal given biases in data collection and analysis.

Therefore, computational complexity and knowledge quality problems present imposing obstacles to quality reasoning under uncertainty. Because of the properties of the underlying computational problem, we cannot solve the problem by simply buying a bigger computer. Brute-force solutions would demand astronomical computational resources, and Moore's

---

[21] Paolo Braca, Stefano Marano, Vincenzo Matta, and Peter Willett, "Asymptotic Efficiency of the PHD in Multitarget/Multisensor Estimation," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 7, No. 3, 2013; Florian Meyer, Paolo Braca, Peter Willett, and Franz Hlawatsch, "Tracking an Unknown Number of Targets Using Multiple Sensors: A Belief Propagation Method," in *2016 19th International Conference on Information Fusion (FUSION)*, 2016.

[22] A much-cited 1996 paper found that approximate inference in Bayesian networks in #P-complete. #P is strictly harder than NP, but the relationship between #P-complete and NP-hard is less obvious (Dan Roth, "On the Hardness of Approximate Reasoning," *Artificial Intelligence*, Vol. 82, Nos. 1–2, 1996).

Law cannot be counted on to save us. Instead, we must seek shortcuts of various kinds: approximation algorithms that might sometimes give inaccurate answers and/or solutions that assume a simpler, more tractable underlying problem. If we cut the right corners, we may attain the results we seek with the informational and computational resources available to us. But to pull off this feat, we need to know which corners to cut—and we cannot be sure that we have the knowledge necessary to do this. The adversary gets a vote, and these simplifying assumptions may prove a highly vulnerable attack surface. If we are tricked into making the wrong assumptions, we may play right into the enemy's designs. As Edward Feigenbaum put it, "in the knowledge lies the power."[23]

A Prospective Research Agenda—Setting Realistic Expectations for Systems That Reason Under Uncertainty

Given the challenges discussed and their implications for national security, what should research focus on going forward? In science fiction, as well as in many visions of the future role of AI in defense, computers conquer uncertainty once and for all.[24] However, there are compelling reasons to believe that computers will not be dramatically better at reasoning under uncertainty than humans. Theoretical analysis shows that rigorous thinking about the unknown would require effectively infinite computational resources. The difficulty of reasoning under uncertainty is a key reason that we may not be able to get AI to do what we want—but what should we do about this?

First, we need to temper our expectations. Progress in computer technology cannot be expected to automatically bring about "dominant battlespace knowledge"; given the relative potential of automation for enhancing deception, our situational awareness of future battlespaces might be *worse* than we have experienced in recent conflicts, not *better*.[25] However, the difficulty of reasoning under uncertainty also presents opportunities that the United States and its allies could exploit to their advantage. If reasoning under uncertainty is a wicked problem, can we force or trick the adversary into trying to solve that problem? If we find the right approaches, perhaps we can make uncertainty work for us, not against us.

---

[23] Edward A. Feigenbaum, *Knowledge Engineering: The Applied Side of Artificial Intelligence*, Palo Alto, Calif.: Stanford University Department of Computer Science, 1980, p. 9.

[24] William A. Owens and Ed Offley, *Lifting the Fog of War*, Baltimore, Md.: Johns Hopkins University Press, 2001; Stuart E. Johnson and Martin C. Libicki, *Dominant Battlespace Knowledge: The Winning Edge,* Washington, D.C.: National Defense University Institute for National Strategic Studies, 1995; Christian Brose, "The New Revolution in Military Affairs: War's Sci-Fi Future," *Foreign Affairs*, Vol. 98, 2019; Keir A. Lieber and Daryl G. Press, "The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence," *International Security*, Vol. 41, No. 4, 2017.

[25] Edward Geist and Marjory Blumenthal, "Military Deception: AI's Killer App?" *War on the Rocks*, October 23, 2019.

To attain our defense objectives, we need to set realistic expectations for systems that reason under uncertainty. This goal requires comprehensive research. We have considerable empirical experience with various experimental systems and some relevant theoretical findings, but we have yet to integrate these into a method for predicting the real-world performance of operationally useful systems. The risks of failing to develop this capability could prove grave. Inflated expectations could have pernicious consequences that might ultimately culminate in defeat on the battlefield. During the past three decades, many analysts envisioned concepts of operations based on the assumption that information fusion and reasoning under uncertainty already were solved or would be solved within the foreseeable future. Research and development funds were allocated to systems that would exploit the possibilities of perfect situational awareness, not to attaining better situational awareness per se. These misconceptions also heavily distorted long-range planning. Most predictions of how AI and other emerging technology will eliminate uncertainty continue to be based on hope, not technical analysis, despite critiques of past debacles associated with this error (such as the Millennium Challenge ’02 exercise).[26] Military critics of these assumptions tend to fall back on Clausewitzian dictums about “the fog and friction of war” and intuitions that perfect situational awareness seems far too good to be true.[27] They are correct, but they see only part of the picture. AI not only cannot be expected to “lift the fog of war”; from what we know, it appears to be far better suited to *thicken* the fog of war.[28] We need a much better sense of what the future battlespace is likely to look like to calibrate our expectations and guide appropriate investment. In turn, this will enable us to direct resources so as to ensure sufficient battlefield performance and to attain our objectives. Doing so involves two thrusts of research: theoretical and practical.

Theoretical Research

The first thrust of research for setting realistic expectations for systems that reason under uncertainty is theoretical. Computer science has produced some tools to begin tackling this problem, but they must be further cultivated to bring them closer to practical systems. For instance, computational-complexity results for DBNs focus on the worst-case complexity of specific subclasses, but those subclasses may not be those used in practice.[29] It might be useful, for instance, to analyze theoretical analogues to those systems that combine Bayesian and non-Bayesian elements, such as MHTs. Non-Bayesian components, such as the pruning

[26] Micah Zenko, “Millennium Challenge: The Real Story of a Corrupted Military Exercise and Its Legacy,” *War on the Rocks*, November 15, 2015.

[27] Herbert R. McMaster, *Crack in the Foundation: Defense Transformation and the Underlying Assumption of Dominant Knowledge in Future War,* Carlisle Barracks, Penn.: Army War College, 2003.

[28] Geist and Blumenthal, 2019.

[29] Lerner, 2002, pp. 91–92.

and gating mechanisms in tracking systems, might be treated as oracles in such analyses, somewhat bridging the gap between our practical systems and our theoretical models.

A particularly significant insight that better theory might provide would be a means to predict the turnover point (where additional information or sensors would cease to be useful). As noted, the location of this point will depend on the quality of the evidence being considered, and a theoretical treatment of knowledge quality would be a major enabler of such analysis. The usual information-theoretic treatment of informational flaws as noise may be inadequate when confronting an intelligent adversary who can introduce carefully designed disinformation into the environment. It seems useful to draw a distinction between such disinformation, which is signal from the standpoint of information theory but is of negative value as knowledge, and generic, naturally occurring noise. Introducing a "knowledge value" component to analysis could overcome this issue and help define the turnover point for various situations by serving as the basis of cost-benefit analysis.

Practical Research

The other line of research needed to set realistic expectations for reasoning under uncertainty is practical. Empirical tests are needed to see whether real systems adhere to theoretical limits, as real-world implementations might outperform worst-case assumptions. To do this, we need to have a sense of what both the average case and the "planning case"—the most extreme case we expect to encounter in an adversarial environment—will be like. Such definitions are essential for informing theoretical research and understanding the insights of that research. Much empirical research can be accomplished with toy systems in contrived environments, such as simulations. There also may be opportunities for large-scale, cost-effective empirical analysis by piggybacking on existing practical systems, such as multi-target trackers. Such piggybacking may make it possible to carry out the necessary empirical studies with a minimum of additional expenditure.

Research to Instrumentalize Uncertainty

If we can set expectations about how reasoning under uncertainty will work in practice, we might be able to instrumentalize uncertainty to work for our interests. The risks of failing to explore these possibilities could be great. If we neglect the possibilities of instrumentalizing the hardness of reasoning under uncertainty, a rival might beat us to this capability and weaponize it against us. Even if no adversaries do this, we may still be depriving ourselves of a potent new capability. If we can make adversaries reason about uncertainty in circumstances of our choosing, we may be able to attain our objectives at much less cost in blood and treasure. However, we are obligated to study this space to develop defenses even if we decide not to exploit it ourselves. These defenses could involve taking active countermeasures and reducing and disguising our vulnerabilities. However, to identify these vulnerabilities and shrink our attack surface, we have to be able to perceive that attack surface.

For instance, imagine a scenario in which we were competing in an undergoverned space with a sophisticated, near-peer adversary. We would need to be confident that we were allocating our computational and other resources efficiently to best understand how and when the space is undergoverned. We can anticipate that the adversary will be trying to complicate this task for us. They might be doing this through classic ambiguity-increasing or ambiguity-decreasing deception tactics. They could also be trying to impose computational costs by introducing uncertainty that is optimized not to disguise the truth but to increase the amount of processing needed to ascertain it (making "known knowns" more expensive) or by sowing doubt about how well we even know how to describe what is going on in the undergoverned space (aggravating "unknown unknowns"). To recognize these tactics, we need to know what to look for—and without research, an adversary might subject us to such tactics without our being able to tell.

As with a research program to set expectations, understanding how uncertainty might be instrumentalized can be divided into a theoretical part and an empirical part.

Theoretical Research

Fortunately, formalisms such as DBNs provide us with a rich foundation to conduct theoretical research. The DBN formalism suggests some ways to categorize different kinds of uncertainty that an adversary might attempt to exploit. For the sake of discussion, assume that the defender has excellent self-awareness and knows all the values associated with the belief net describing itself—that is, its internal state variables and the probable evidence variables that an external observer might detect. The defender aims to complicate the rival's ability to reconstruct these values, but its actions could take very different forms depending on its goals. Perhaps the defender cares little if the rival learns the true values, but merely aims to impose costs by making the rival work harder to learn them.

An obvious way to do this is simply by increasing noise, but there might be subtler or more focused strategies, such as adding targeted spurious variables and focusing noise around selected true variables in a manner theoretical analysis suggests will increase the difficulty of reasoning about the problem, even if it is not guaranteed to mislead the observer in the end. Another obvious case would be when the defender hopes to mislead the rival about the state of a few selected true variables. Such deceptions could have a variety of characteristics: An observer might experience a certain kind of uncertainty—"The value is between 3.5 and 4.2 and probably on the high side of that"—instead of making a specific wrong conclusion—"I'm sure the value is 7.1."

A particularly important goal could be to prevent the rival from correctly inferring the graph structure describing the relationship of the defender's state variables, as opposed to the variables per se. In many cases, this structure is much more important than the momentary state of its constituent variables, because it can be exploited to reconstruct other parts of the state under previously unobserved conditions. Once again, the formalism suggests efficient ways to accomplish this: One can add new variables, but there is also the possibility of convincing the rival of the existence of spurious edges between real variables. Such deceptions could be designed to reduce the accuracy of inference or increase the computational cost of inference. Finally, instrumentalizing uncertainty is not just about making rivals uncertain. Sometimes, we want to make sure that a potential adversary is absolutely certain about something that is true. For instance, to avoid undesired escalation, it is imperative that the rival not perceive a possibility of an imminent attack that does not exist.

With the definition of an appropriate metric, we can design algorithms to optimize gambits such as these. A concept like the knowledge value suggested earlier could serve as the basis for metrics to measure the efficacy of uncertainty-manipulation methods. For analytical purposes, this might be defined as "the value to the defender of the action that a boundedly rational rival will take given how they perceive a particular state," with *boundedly rational* defined as "instrumentally rational subject to finite computational resources for approximate Bayesian reasoning."[30] This would mean that an agent attempts to use the most accurate approximations that it can compute for the probable state of the world and that it acts rationally to pursue its goals given those perceptions. As noted, such a metric requires assumptions about the rival's ontology and preferences to predict their computational complexity. However, this is an unavoidable aspect of formalisms of this type (consider algorithmic game theory).[31]

Practical Research

The empirical aspect of the research program to instrumentalize uncertainty would use simulations and practical experiments to test both uncertainty-manipulation techniques and the applicability of proposed metrics. Simulated sandboxes could be used for both simulated agents and human test subjects. Given the unresolved mystery of human reasoning about uncertainty, a critical consideration is whether humans are better at overcoming or detecting uncertainty manipulation than predictions based on such theoretical abstractions as boundedly rational Bayesian agents. Alternatively, observational tests might find that humans have specific cognitive vulnerabilities that theory fails to predict. It is well known that humans employ various cognitive shortcuts, and some researchers have long sought to formalize these heuristic mechanisms to simulate them on a computer. Methods from cognitive science and existing cognitive architectures might be adapted to assist these inquiries.[32] The resulting findings could, in turn, inform updated theories and metrics, as well as the design of experiments to test them.

---

[30] This is basically the same as Kenneth J. Arrow's definition of *bounded rationality* (Kenneth J. Arrow, "Is Bounded Rationality Unboundedly Rational? Some Ruminations," in Mie Augier and James G. March, eds., *Models of a Man: Essays in Memory of Herbert A. Simon,* Cambridge, Mass.: MIT Press, 2004, p. 48).

[31] Tim Roughgarden, "Algorithmic Game Theory," *Communications of the ACM*, Vol. 53, No. 7, 2010.

[32] Iuliia Kotseruba and John K. Tsotsos, "40 Years of Cognitive Architectures: Core Cognitive Abilities and Practical Applications," *Artificial Intelligence Review*, Vol. 53, No. 1, 2020.

Concluding Thoughts

Machines capable of efficient, rapid, and accurate reasoning under uncertainty could revolutionize both civilian and military affairs. The allure of these possibilities has compelled generations of AI researchers to attempt to create such systems. Over decades, they have pioneered a succession of different approaches to this goal. However, despite some real successes, they have yet to reach it. As AI researchers' theoretical understanding grew, it became apparent that this disappointment stems from the nature of reasoning under uncertainty. This problem turns out to be computationally complex and epistemologically fraught. There is no single correct or optimal way to reason about uncertainty, because there is more than one way to be uncertain. AI researchers have translated some of these alternative modes of uncertainty into algorithms. Notable examples of these are Pearl's Bayesian belief networks, Zadeh's possibility theory, and Dempster-Shafer theory.

Although we can envision ideal systems for reasoning under uncertainty, these require unobtainable computational resources. As a consequence, actual systems must make tradeoffs between speed, accuracy, and expressiveness. In essence, to reason about uncertainty using machines, it is necessary to weigh between a set of possible worlds consistent with the available evidence. However, for a nontrivial problem, these possible worlds are too numerous for a physical computer to represent and reason with. A real-world system must instead work with a smaller subset of possible worlds; in some cases, this shortcut can enable good performance, but, to make it work, one must possess accurate knowledge about which subset will be encountered in practice. As a consequence, computers and AI cannot be expected to eliminate uncertainty.

We must learn to live with uncertainty, but we can mitigate its hazards and perhaps even make it work for us. AI research on the problem of reasoning under uncertainty can serve as the foundation for investigations of these possibilities. First, we must set realistic expectations for reasoning about uncertainty. If we cling to ill-founded hopes that computers will slay the dragon of uncertainty for us, we are likely to misallocate resources and might even suffer battlefield defeat because we placed too much faith in flawed systems. A dual-pronged research program with both theoretical and empirical components could help demystify these issues for us. In particular, theoretical considerations suggest that there is probably a point of diminishing returns beyond which the computational costs of reasoning with more information are greater than the additional value that the information ends up providing. Second, we must confront the possibility that uncertainty might itself be turned into a wieldable instrument of state power. If reasoning under uncertainty is a hard problem, perhaps others can be compelled or fooled to try to solve those problems. Even if we decide not to solve the problem of reasoning under uncertainty, we need to study the problem for defensive purposes. The same sort of theoretical and empirical research needed to set expectations for reasoning under uncertainty could suggest not only possible ways that uncertainty could be instrumentalized but also prospective defenses against those possibilities.

Acknowledgments

I would like to thank the editors and reviewers of this report for their contribution to the development of this chapter.

Abbreviations

AI	artificial intelligence
ASDA	Act-Sense-Decide-Adapt
CLG	Continuous Linear Gaussian
DBN	Dynamic Bayesian Network
MHT	Multiple Hypothesis Tracker
UGS	undergoverned spaces

References

Arrow, Kenneth J., "Is Bounded Rationality Unboundedly Rational? Some Ruminations," in Mie Augier and James G. March, eds., *Models of a Man: Essays in Memory of Herbert A. Simon*, Cambridge, Mass.: MIT Press, 2004.

Blackman, Samuel, and Robert Popoli, *Design and Analysis of Modern Tracking Systems*, Norwood, Mass.: Artech House, 1999.

Braca, Paolo, Stefano Marano, Vincenzo Matta, and Peter Willett, "Asymptotic Efficiency of the PHD in Multitarget/Multisensor Estimation," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 7, No. 3, 2013, pp. 553–564.

Brose, Christian, "The New Revolution in Military Affairs: War's Sci-Fi Future," *Foreign Affairs*, Vol. 98, 2019, pp. 122–134.

Cooper, Gregory F., "The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks," *Artificial Intelligence*, Vol. 42, Nos. 2–3, 1990, pp. 393–405.

Coraluppi, Stefano, "Fundamentals and Advances in Multiple-Hypothesis Tracking," in *NATO STO IST-134 Lecture Series on Advanced Algorithms for Effectively Fusing Hard and Soft Information*, Paris, France: NATO Collaboration and Support Office, 2015.

Dagum, Paul, and Michael Luby, "Approximating Probabilistic Inference in Bayesian Belief Networks Is NP-Hard," *Artificial Intelligence*, Vol. 60, No. 1, 1993, pp. 141–153.

Domingos, Pedro, "A Few Useful Things to Know About Machine Learning," *Communications of the ACM*, Vol. 55, No. 10, 2012, pp. 78–87.

Feigenbaum, Edward A., *Knowledge Engineering: The Applied Side of Artificial Intelligence*, Palo Alto, Calif.: Stanford University Department of Computer Science, 1980.

Geist, Edward, and Marjory Blumenthal, "Military Deception: AI's Killer App?" *War on the Rocks*, October 23, 2019.

Gelman, Andrew, and Cosma Rohilla Shalizi, "Philosophy and the Practice of Bayesian Statistics," *British Journal of Mathematical and Statistical Psychology*, Vol. 66, No. 1, 2013, pp. 8–38.

Hofstadter, Douglas R., *I Am a Strange Loop*, New York: Basic Books, 2007.

Johnson, Stuart E., and Martin C. Libicki, *Dominant Battlespace Knowledge: The Winning Edge,* Washington, D.C.: National Defense University Institute for National Strategic Studies, 1995.

Kotseruba, Iuliia, and John K. Tsotsos, "40 Years of Cognitive Architectures: Core Cognitive Abilities and Practical Applications," *Artificial Intelligence Review*, Vol. 53, No. 1, 2020, pp. 17–94.

Lenat, Douglas B., and John Seely Brown, "Why AM and EURISKO Appear to Work," *Artificial Intelligence*, Vol. 23, No. 3, 1984, pp. 269–294.

Lerner, Uri N., *Hybrid Bayesian Networks for Reasoning About Complex Systems*, dissertation, Stanford University, 2002.

Lerner, Uri N., and Ron Parr, *Inference in Hybrid Networks: Theoretical Limits and Practical Algorithms*, arXiv, 2013.

Levesque, Hector J., and Ronald J. Brachman, "A Fundamental Tradeoff in Knowledge Representation and Reasoning," in Ronald J. Brachman and Hector J. Levesque, eds., *Readings in Knowledge Representation*, Los Altos, Calif.: Morgan Kauffmann, 1985, pp. 42–70.

Lieber, Keir A., and Daryl G. Press, "The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence," *International Security*, Vol. 41, No. 4, 2017, pp. 9–49.

McMaster, Herbert R., *Crack in the Foundation: Defense Transformation and the Underlying Assumption of Dominant Knowledge in Future War,* Carlisle Barracks, Penn.: Army War College, 2003.

Meyer, Florian, Paolo Braca, Peter Willett, and Franz Hlawatsch, "Tracking an Unknown Number of Targets Using Multiple Sensors: A Belief Propagation Method," in *2016 19th International Conference on Information Fusion (FUSION)*, 2016, pp. 719–726.

Murphy, Kevin P., *Switching Kalman Filters*, Berkeley, Calif.: Department of Computer Science, University of California, Berkeley, August 1998.

Owens, William A., and Ed Offley, *Lifting the Fog of War*, Baltimore, Md.: Johns Hopkins University Press, 2001.

Pavlovic, Vladimir, James M. Rehg, Tat-Jen Cham, and Kevin P. Murphy, "A Dynamic Bayesian Network Approach to Figure Tracking Using Learned Dynamic Models," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Vol. 1, IEEE, 1999, pp. 94–101.

Pearl, Judea, and Stuart J. Russell, *Bayesian Networks*, Los Angeles, Calif.: University of California, November 17, 2000.

Poole, David, "A Logical Framework for Default Reasoning," *Artificial Intelligence*, Vol. 36, No. 1, 1988, pp. 27–47.

Reiter, Raymond, "A Logic for Default Reasoning," *Artificial Intelligence*, Vol. 13, Nos. 1–2, 1980, pp. 81–132.

Roth, Dan, "On the Hardness of Approximate Reasoning," *Artificial Intelligence*, Vol. 82, Nos. 1–2, 1996, pp. 273–302.

Roughgarden, Tim, "Algorithmic Game Theory," *Communications of the ACM*, Vol. 53, No. 7, 2010, pp. 78–86.

Russell, Stuart J., and Peter Norvig, *Artificial Intelligence: A Modern Approach,* 2nd ed., Upper Saddle River, N.J.: Pearson Education, Inc., 2005.

Shafer, Glenn, *A Mathematical Theory of Evidence*, Princeton, N.J.: Princeton University Press, 1976.

Tchamova, Albena, and Jean Dezert, "On the Behavior of Dempster's Rule of Combination and the Foundations of Dempster-Shafer Theory," in *2012 6th IEEE International Conference Intelligent Systems*, IEEE, 2012, pp. 108–113.

Zadeh, Lotfi A., "Fuzzy Logic and Approximate Reasoning," *Synthese*, Vol. 30, Nos. 3–4, 1975, pp. 407–428.

Zadeh, Lotfi A., "Review of *A Mathematical Theory of Evidence*," *AI Magazine*, Vol. 5, No. 3, 1984, pp. 81–83.

Zenko, Micah, "Millennium Challenge: The Real Story of a Corrupted Military Exercise and Its Legacy," *War on the Rocks*, November 15, 2015.