1 Introduction

Many important problems involve decision making under uncertainty, including aircraft collision avoidance, wildfire management, and disaster response. When designing automated decision-making systems or decision-support systems, it is important to account for the various sources of uncertainty when making or recommending decisions. Accounting for these sources of uncertainty and carefully balancing the multiple objectives of the system can be very challenging. We will discuss these challenges from a computational perspective, aiming to provide the theory behind decision-making models and computational approaches. This chapter introduces the problem of decision making under uncertainty, provides some example applications, and outlines the space of possible computational approaches. The chapter then summarizes how various disciplines have contributed to our understanding of intelligent decision making and highlights areas of potential societal impact. We conclude with an outline of the remainder of the book.

1.1 Decision Making

An *agent* is something that acts based on observations of its environment. Agents may be physical entities, like humans or robots, or they may be nonphysical entities, such as decision support systems that are implemented entirely in software. As shown in figure 1.1, the interaction between the agent and the world follows an *observe-act cycle* or *loop*.

The agent at time t receives an *observation* of the world, denoted o_t. Observations may be made, for example, through a biological sensory process as in humans or by a sensor system like radar in an air traffic control system. Observations are often incomplete or noisy; humans may not see an approaching aircraft or a radar

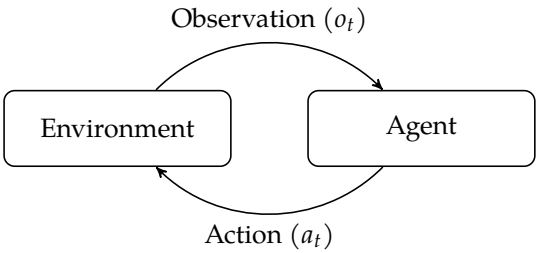

Figure 1.1. Interaction between the environment and the agent.

system might miss a detection through electromagnetic interference. The agent then chooses an action a_t through some decision-making process to be discussed later. This action, such as sounding an alert, may have a nondeterministic effect on the world.

Our focus is on agents that interact intelligently in the world to achieve their objectives over time. Given the past sequence of observations $o_1, \ldots, o_t$ and knowledge about the environment, the agent must choose an action a_t that best achieves its objectives in the presence of various sources of uncertainty, including:

1. *outcome uncertainty*, where the effects of our actions are uncertain,
2. *model uncertainty*, where our model of the problem is uncertain,
3. *state uncertainty*, where the true state of the environment is uncertain, and
4. *interaction uncertainty*, where the behavior of the other agents interacting in the environment is uncertain.

This book is organized around these four sources of uncertainty. Making decisions in the presence of uncertainty is central to the field of *artificial intelligence*[1] as well as many other fields, as outlined in section 1.4. We will discuss a variety of algorithms, or descriptions of computational processes, for making decisions that are robust to uncertainty.

[1] A comprehensive introduction to artificial intelligence is provided by S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson, 2021.

1.2 Applications

The decision making framework presented in the previous section can be applied to a wide variety of domains. This section discusses a few conceptual examples with real-world applications. Appendix F outlines additional notional examples that are used throughout this text to demonstrate the algorithms we discuss.

1.2.1 *Aircraft Collision Avoidance*

To help prevent mid-air collisions between aircraft, we want to design a system that can alert pilots to potential threats and direct them how to maneuver.[2] The system communicates with the transponders of other aircraft to identify their positions with some degree of accuracy. Deciding what guidance to provide to the pilots from this information is challenging. There is uncertainty in how quickly the pilots will respond and how strongly they will comply with the guidance. In addition, there is uncertainty in the behavior of the other aircraft in the vicinity. We want our system to alert sufficiently early to provide enough time for the pilots to maneuver the aircraft to avoid collision, but we do not want our system to alert too early and result in many unnecessary maneuvers. Since this system is to be used continuously worldwide, we need the system to provide an exceptional level of safety.

[2] This application is discussed in a chapter titled ''Collision Avoidance'' by M. J. Kochenderfer, *Decision Making Under Uncertainty: Theory and Application*. MIT Press, 2015.

1.2.2 *Automated Driving*

We want to build an autonomous vehicle that can safely drive in urban environments.[3] The vehicle must rely upon a suite of sensors to perceive its environment to make safe decisions. One type of sensor is lidar, which involves measuring laser reflections off of the environment to determine distances to obstacles. Another type of sensor is a camera, which, through computer vision algorithms, can detect pedestrians and other vehicles. Both of these types of sensors are imperfect and susceptible to noise and occlusions. For example, a parked truck may occlude a pedestrian that may be trying to cross at a crosswalk. Our system must predict the intentions and future paths of other vehicles, pedestrians, and other road users from their observable behavior in order to safely navigate to our destination.

[3] A similar application was explored by M. Bouton, A. Nakhaei, K. Fujimura, and M. J. Kochenderfer, ''Safe Reinforcement Learning with Scene Decomposition for Navigating Complex Urban Environments,'' in *IEEE Intelligent Vehicles Symposium (IV)*, 2019.

1.2.3 *Breast Cancer Screening*

Worldwide, breast cancer is the most common cancer in women. Detecting breast cancer early can help save lives, with mammography being the most effective screening tool available. However, mammography carries with it potential risks, including false positives, which can result in unnecessary and invasive diagnostic followup. Research over the years has resulted in various population-based screening schedules based on age in order to balance benefits and risks of the tests. Developing a system that can make personalized recommendations based

on personal risk characteristics and screening history has the potential to result in better outcomes for women.[4] The success of such a system can be compared to population-wide screening schedules in terms of total expected quality-adjusted life years, the number of mammograms, false-positives, and risk of undetected invasive cancer.

[4] Such a concept is proposed by T. Ayer, O. Alagoz, and N. K. Stout, "A POMDP Approach to Personalize Mammography Screening Decisions," *Operations Research*, vol. 60, no. 5, pp. 1019–1034, 2012.

1.2.4 *Financial Consumption and Portfolio Allocation*

Suppose we want to build a system that recommends to an individual how much of their wealth they should consume that year and how to allocate their investments.[5] The investment portfolio may include stocks and bonds with different levels of risk and expected return. The evolution of wealth is stochastic due to uncertainty in both earned and investment income, often increasing until near retirement and then steadily decreasing. The enjoyment that comes from the consumption of a unit of wealth in a year typically diminishes with the amount consumed, resulting in a desire to smooth consumption over the lifespan of the individual.

[5] A related problem was studied by R. C. Merton, "Optimum Consumption and Portfolio Rules in a Continuous-Time Model," *Journal of Economic Theory*, vol. 3, no. 4, pp. 373–413, 1971.

1.2.5 *Distributed Wildfire Surveillance*

A major challenge in fighting wildfires is situational awareness. The state of the fire evolves over time, influenced by factors such as wind and the distribution of fuel in the environment. Many wildfires span large geographic regions. One concept for monitoring a wildfire is to use a team of drones equipped with sensors to fly above the fire.[6] The sensing range of individual drones are limited, but the information from the team can be fused to provide a unified snapshot of the situation to inform resource allocation decisions. We would like the team to autonomously determine how to collaborate with each other to provide the best coverage of the fire. Effective monitoring requires deciding how to maneuver to cover areas where new sensor information is likely to be useful; spending time in areas where we are certain the fire is burning or not burning would be uninformative. Identifying important areas to explore requires reasoning about the stochastic evolution of the fire given only imperfect knowledge of the current state of the fire.

[6] This application was explored by K. D. Julian and M. J. Kochenderfer, "Distributed Wildfire Surveillance with Autonomous Aircraft Using Deep Reinforcement Learning," *AIAA Journal on Guidance, Control, and Dynamics*, vol. 42, no. 8, pp. 1768–1778, 2019.

1.3 Methods

There are many different methods for designing decision agents. Depending on the application, some may be more appropriate than others. They differ in the responsibilities of the designer and the tasks left to automation. This section briefly overviews a collection of these methods. The book will focus primarily on planning and reinforcement learning, but some of the techniques will involve elements of supervised learning and optimization.

1.3.1 Explicit Programming

The most direct method for designing a decision agent is to anticipate all the different scenarios the agent might find itself in and then explicitly program the agent to do what is desired. The explicit programming approach may work well for simple problems, but it places a large burden on the designer to provide a complete strategy. Various agent programming languages and frameworks have been proposed to make programming agents easier.

1.3.2 Supervised Learning

In some problems, it may be easier to show an agent what to do rather than to write a program for the agent to follow. The designer provides a set of training examples, and an automated learning algorithm must generalize from these examples. This approach is known as supervised learning and has been widely applied to classification problems. This technique is sometimes called behavioral cloning when applied to learning mappings from observations to actions. Behavioral cloning works well when an expert designer actually knows the best course of action for a representative collection of example situations. Although there exists a wide variety of different learning algorithms, they generally cannot perform better than human designers in new situations.

1.3.3 Optimization

Another approach is for the designer to specify the space of possible decision strategies and a performance measure to be maximized. Evaluating the performance of a decision strategy generally involves running a batch of simulations. The optimization algorithm then performs a search in this space for the optimal

strategy. If the space is relatively low dimensional and the performance measure does not have many local optima, then various local or global search methods may be appropriate. Although knowledge of a dynamic model is generally assumed in order to run the simulations, it is not otherwise used to guide the search, which can be important in complex problems.

1.3.4 Planning

Planning is a form of optimization that uses a model of the problem dynamics to help guide the search. A broad literature has arisen on planning problems, much of it focused on deterministic problems. For some problems, it may be acceptable to approximate the dynamics with a deterministic model. Assuming a deterministic model allows us to use methods that can more easily scale to high-dimensional problems. For other problems, accounting for future uncertainty is absolutely critical. This book focuses entirely on problems in which accounting for uncertainty is important.

1.3.5 Reinforcement learning

Reinforcement learning relaxes the assumption in planning that a model is known ahead of time. Instead, the decision-making strategy is learned while the agent interacts with the world. The designer only has to provide a performance measure; it is up to a learning algorithm to optimize the behavior of the agent. One of the interesting complexities that arises in reinforcement learning is that the choice of action impacts not only the immediate success of the agent in achieving its objectives but also the agent's ability to learn about the environment and identify the characteristics of the problem that it can exploit.

1.4 History

The theory of automating the process of decision making has roots in the dreams of early philosophers, scientists, mathematicians, and writers. The ancient Greeks began incorporating automation into myths and stories as early as 800 B.C. The word *automaton* was first used in Homer's *Iliad*, which contains references to the notion of automatic machines including mechanical tripods used to serve dinner guests.[7] In the seventeenth century, philosophers proposed the use of logic rules

[7] S. Vasileiadou, D. Kalligeropoulos, and N. Karcanias, "Systems, Modelling and Control in Ancient Greece: Part 1: Mythical Automata," *Measurement and Control*, vol. 36, no. 3, pp. 76–80, 2003.

to automatically settle disagreements. Their ideas created the foundation for mechanized reasoning.

Beginning in the late eighteenth century, inventors began creating automatic machines to perform labor. In particular, a series of innovations in the textile industry led to the development of the automatic loom, which in turn laid the foundation for the first factory robots.[8] In the early nineteenth century, the use of intelligent machines to automate labor began to make its way into science fiction novels. The word *robot* originated in Czech writer Karel Čapek's play titled *Rossum's Universal Robots* about machines that could perform work humans would prefer not to do. The play inspired other science fiction writers to incorporate robots into their writing. In the mid-twentieth century, notable writer and professor Isaac Asimov laid out his vision for robotics in his famous *Robot Series*.

[8] N. J. Nilsson, *The Quest for Artificial Intelligence*. Cambridge University Press, 2009.

A major challenge in practical implementations of automated decision making is accounting for uncertainty. Even at the end of the twentieth century, George Dantzig, most famously known for developing the simplex algorithm, stated in 1991:

> In retrospect it is interesting to note that the original problem that started my research is still outstanding—namely the problem of planning or scheduling dynamically over time, particularly planning dynamically under uncertainty. If such a problem could be successfully solved it could (eventually through better planning) contribute to the well-being and stability of the world.[9]

[9] G. B. Dantzig, "Linear Programming," *Operations Research*, vol. 50, no. 1, pp. 42–47, 2002.

While decision making under uncertainty still remains an active area of research, over the past few centuries, researchers and engineers have come closer to making the concepts posed by these early dreamers possible. Current state-of-the-art decision making algorithms rely on a convergence of concepts developed in multiple disciplines including economics, psychology, neuroscience, computer science, engineering, mathematics, and operations research. This section highlights some major contributions from these disciplines. The cross-pollination between disciplines has led to many recent advances and will likely continue to support growth in the future.

1.4.1 Economics

An understanding of individual decision making is central to economic theory, prompting economists to develop techniques to model human decision making. One such technique is utility theory, which was first introduced in the late

eighteenth century.[10] Utility theory provides a means to model and compare the desirability of various outcomes. For example, utility can be used to compare the desirability of various monetary values. In the *Theory of Legislation*, Jeremy Bentham summarized the nonlinearity in the utility of money:

> *1st.* Each portion of wealth has a corresponding portion of happiness.
> *2nd.* Of two individuals with unequal fortunes, he who has the most wealth has the most happiness.
> *3rd.* The excess in happiness of the richer will not be so great as the excess of his wealth.[11]

[10] G. J. Stigler, ''The Development of Utility Theory. I,'' *Journal of Political Economy*, vol. 58, no. 4, pp. 307–327, 1950.

[11] J. Bentham, *Theory of Legislation*. Trübner & Company, 1887.

By combining the concept of utility with the notion of rational decision making, economists in the mid-twentieth century established a basis for the maximum expected utility principle. This principle is a key concept behind the creation of autonomous decision making agents. Utility theory also gave rise to the development of game theory, which attempts to understand the behavior of multiple agents acting in the presence of one another to maximize their interests.[12]

[12] O. Morgenstern and J. von Neumann, *Theory of Games and Economic Behavior*. Princeton University Press, 1953.

1.4.2 *Psychology*

Psychologists also study human decision making, typically from the perspective of human behavior. By studying the reactions of animals to stimuli, psychologists have been developing theories of trial-and-error learning since the nineteenth century. Researchers noticed that animals tended to make decisions based on the satisfaction or discomfort they experienced in previous similar situations. Russian psychologist Ivan Pavlov combined this idea with the concept of reinforcement after observing the salivation patterns of dogs when fed. Psychologists found that a pattern of behavior could be strengthened or weakened using a continuous reinforcement of a particular stimulus. In the mid-twentieth century, mathematician and computer scientist Alan Turing expressed the possibility of allowing machines to learn in the same manner:

> The organization of a machine into a universal machine would be most impressive if the arrangements of interference involve very few inputs. The training of a human child depends largely on a system of rewards and punishments, and this suggests that it ought to be possible to carry through the organising with only two interfering inputs, one for 'pleasure' or 'reward' (R) and the other for 'pain' or 'punishment' (P).[13]

[13] A. M. Turing, ''Intelligent Machinery,'' National Physical Laboratory, Report, 1948.

The work of psychologists laid the foundation for the field of reinforcement learning, a critical technique used to teach agents to make decisions in uncertain environments.[14]

[14] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, 2018.

1.4.3 Neuroscience

While psychologists study human behavior as it happens, neuroscientists focus on the biological processes used to create the behavior. At the end of the nineteenth century, scientists found that the brain is composed of an interconnected network of neurons, which is responsible for its ability to perceive and reason about the world. Artificial intelligence pioneer Nils Nilsson describes the application of these findings to decision making as follows:

> Because it is the *brain* of an animal that is responsible for converting sensory information into action, it is to be expected that several good ideas can be found in the work of neurophysiologists and neuranatomists who study brains and their fundamental components, neurons.[15]

[15] N. J. Nilsson, *The Quest for Artificial Intelligence*. Cambridge University Press, 2009.

In the 1940s, researchers first proposed that neurons could be considered as individual ''logic units'' capable of performing computational operations when pieced together into a network. This work served as a basis for neural networks, which are used in the field of artificial intelligence to perform a variety of complex tasks.

1.4.4 Computer Science

In the mid-twentieth century, computer scientists began formulating the problem of intelligent decision making as a problem of symbolic manipulation through formal logic. The computer program *Logic Theorist*, written in the mid-twentieth century to perform automated reasoning, used this way of thinking to prove mathematical theorems. Herbert Simon, one of its inventors, addressed the symbolic nature of the program by relating it to the human mind:

> We invented a computer program capable of thinking non-numerically, and thereby solved the venerable mind/body problem, explaining how a system composed of matter can have the properties of mind.[16]

[16] Quoted by J. Agar, *Science in the 20th Century and Beyond*. Polity, 2012.

These symbolic systems relied heavily on human expertise. An alternative approach to intelligence, called connectionism, was inspired in part by developments in neuroscience and focuses on the use of artificial neural networks as a substrate for intelligence. With the knowledge that neural networks could be trained for pattern recognition, connectionists attempt to learn intelligent behavior from data or experience rather than the hard-coded knowledge of experts. The connectionist paradigm underpinned the success of AlphaGo, the autonomous program that beat a human professional at the game of Go, as well as much of the development of autonomous vehicles. Algorithms that combine both symbolic and connectionist paradigms remains an active area of research today.

1.4.5 Engineering

The field of engineering has focused on allowing physical systems, such as robots, to make intelligent decisions. World-renowned roboticist Sebastian Thrun describes the components of these systems as follows:

> Robotics systems have in common that they are situated in the physical world, perceive their environments through sensors, and manipulate their environment through things that move.[17]

To design these systems, engineers must address perception, planning, and acting. Physical systems perceive the world by using their sensors to create a representation of the salient features of their environment. The field of state-estimation has focused on using sensor measurements to construct a belief about the state of the world. Planning requires reasoning about the ways to execute the tasks they are designed to perform. The planning process has been enabled by advances in the semiconductor industry spanning many decades.[18] Once a plan has been devised, an autonomous agent must act on it in the real world. This task requires both hardware in the form of actuators and algorithms to control the actuators and reject disturbances. The field of control theory has focused on the stabilization of mechanical systems through feedback control.[19] Automatic control systems are widely used in industry, from the regulation of temperature in an oven to the navigation of aerospace systems.

[17] S. Thrun, ''Probabilistic Robotics,'' *Communications of the ACM*, vol. 45, no. 3, pp. 52–57, 2002.

[18] G. E. Moore, ''Cramming More Components Onto Integrated Circuits,'' *Electronics*, vol. 38, no. 8, pp. 114–117, 1965.

[19] D. A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. JHU Press, 2002.

1.4.6 Mathematics

In order to make informed decisions in uncertain environments, an agent must be able to quantify its uncertainty. The field of decision making relies heavily on probability theory for this task. In particular, Bayesian statistics plays an important role in this text. In 1763, a paper of Thomas Bayes was published posthumously containing what would later be known as Bayes' rule. His approach to probabilistic inference fell in and out of favor until the mid-twentieth century, when researchers began to find Bayesian methods useful in a number of settings.[20] Mathematician Bernard Koopman found practical use for the theory during World War II.

> Every operation involved in search is beset with uncertainties; it can be understood quantitatively only in terms of [...] probability. This may now be regarded as a truism, but it seems to have taken the developments in operational research of the Second World War to drive home its practical implications.[21]

Sampling-based methods (sometimes referred to as Monte Carlo methods) developed in the early twentieth century for large scale calculations as part of the Manhattan Project, made some inference techniques possible that would previously have been intractable. These foundations serve as a basis for Bayesian networks, which increased in popularity later in the twentieth century in the field of artificial intelligence.

1.4.7 Operations Research

Operations research is concerned with finding optimal solutions to decision-making problems such as resource allocation, asset investment, or maintenance scheduling. In the late nineteenth century, researchers began to explore the application of mathematical and scientific analysis to the production of goods and services. The field was accelerated during the Industrial Revolution when companies began to subdivide their management into departments responsible for distinct aspects of overall decisions. During World War II, the optimization of decisions was applied to allocating resources to an army. Once the war came to an end, businesses began to notice that the same operations research concepts previously used to make military decisions could help them optimize their business decisions. This realization led to the development of management science, as described by organizational theorist Harold Koontz:

[20] W. M. Bolstad and J. M. Curran, *Introduction to Bayesian Statistics*. Wiley, 2016.

[21] B. O. Koopman, *Search and Screening: General Principles with Historical Applications*. Pergamon Press, 1980.

> The abiding belief of this group is that, if management, or organization, or planning, or decision making is a logical process, it can be expressed in terms of mathematical symbols and relationships. The central approach of this school is the model, for it is through these devices that the problem is expressed in its basic relationships and in terms of selected goals or objectives.[22]

[22] H. Koontz, ''The Management Theory Jungle,'' *Academy of Management Journal*, vol. 4, no. 3, pp. 174–188, 1961.

This desire to be able to better model and understand business decisions sparked the development of a number of algorithms used today such as linear programming, dynamic programming, and queuing theory.[23]

[23] F. S. Hillier, *Introduction to Operations Research*. McGraw-Hill, 2012.

1.5 *Societal Impact*

Algorithmic approaches to decision making have transformed society and will likely continue to an even greater extent in the future. This section briefly highlights a few ways decision making algorithms can contribute to society as well as challenges that remain in ensuring broad benefit.[24]

[24] A much more thorough discussion is provided by Z. R. Shi, C. Wang, and F. Fang, ''Artificial Intelligence for Social Good: A Survey,'' 2020. arXiv: 2001.01818v1.

Algorithmic approaches have contributed to environmental sustainability. In the context of energy management, for example, Bayesian optimization has been applied to automated home energy management systems. Algorithms from the field of multi-agent systems are used to predict the operation of smart grids, design markets for trading energy, and predict rooftop solar-power adoption. Algorithms have also been developed to protect biodiversity. For example, neural networks are used to automate wildlife census, game-theoretic approaches are used to combat poaching in forests, and optimization techniques are employed to allocate resources for habitat management.

Decision making algorithms have found success in the field of medicine for decades. Such algorithms have been used for matching residents to hospitals and matching organ donors to patients in need. An early application of Bayesian networks, which we will cover in the first part of this book, was disease diagnosis. Since then, Bayesian networks have been widely used in medicine for diagnosis and prognosis of many diseases such as cervical cancer, breast cancer, and glaucoma. The field of medical image processing has been transformed by deep learning, and recently, algorithmic ideas have played an important role in understanding the spread of disease.

Algorithms have enabled us to understand the growth of urban areas and facilitate their design. Data-driven algorithms have been widely used to improve public infrastructure. For example, stochastic processes have been used to predict

failures in water pipelines, deep learning has improved the management of traffic, Markov decision processes and Monte Carlo methods have been employed to improve emergency response. Ideas from decentralized multi-agent systems have optimized travel routes, and path planning techniques have been used to optimize delivery of goods. A major application of decision making algorithms in transportation has been in the development of autonomous cars and improving the safety of aircraft.

Algorithms for optimizing decisions can amplify the impact of its users, regardless of the nature of their intention. If the objective of the user of these algorithms, for example, is to spread misinformation during a political election, then optimization processes can help facilitate this. However, similar algorithms can be used to monitor and counteract the spread of false information. Sometimes the implementation of these decision making algorithms can lead to downstream consequences that were not intended by their users.[25]

Although algorithms have the potential to bring significant benefits, there are also challenges associated with their implementation in society. Data-driven algorithms often suffer from inherent biases and blind spots due to the way data is collected. As algorithms become part of our lives, it is important to understand how the risk of bias can be reduced and how the benefits of algorithmic progress can be distributed in a manner that is equitable and fair. Algorithms can also be vulnerable to adversarial manipulation, and it is critical that we design algorithms that are robust to such attacks. It is also important to extend moral and legal frameworks for preventing unintended consequences and assigning responsibility.

[25] For a general discussion see B. Christian, *The Alignment Problem*. Norton & Company, 2020. See also D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané, "Concrete Problems in AI Safety," 2016. arXiv: 1606.06565v2.

1.6 Overview

This book is divided into five parts. The first part addresses the problem of reasoning about uncertainty and objectives in simple decisions at a single point in time. The second part extends decision making to sequential problems, where we must make a sequence of decisions in response to information that we gather along the way about the outcomes of our actions. The third part addresses model uncertainty, where we do not start with a known model and must learn how to act through interaction with the environment. The fourth part addresses state uncertainty, where we do not know the current state of the environment due

to imperfect perceptual information. The final part discusses decision contexts involving multiple agents.

1.6.1 *Probabilistic Reasoning*

Rational decision making requires reasoning about our uncertainty and objectives. This part of the book begins by discussing how to represent uncertainty as a probability distribution. Real-world problems require reasoning about distributions over many different variables. We will discuss how to construct these models, how to use them to make inferences, and how to learn their parameters and structure from data. We then introduce the foundations of *utility theory* and show how it forms the basis for rational decision making under uncertainty through the maximum expected utility principle. We then discuss how notions of utility theory can be incorporated into the probabilistic graphical models introduced earlier to form what are called decision networks.

1.6.2 *Sequential Problems*

Many important problems require that we make a series of decisions. The same principle of maximum expected utility still applies, but optimal decision making in a sequential context requires reasoning about future sequences of actions and observations. This part of the book will discuss sequential decision problems in stochastic environments. We will focus on a general formulation of sequential decision problems under the assumption that the model is known and that the environment is fully observable. We will relax both of these assumptions later. Our discussion will begin with the introduction of the *Markov decision process* (*MDP*), the standard mathematical model for sequential decision problems. We will discuss several approaches for finding exact solutions to these types of problems. Because large problems sometimes do not permit exact solutions to be efficiently found, we will discuss a collection of both offline and online approximate solution methods along with a type of method that involves directly searching the space of parameterized decision policies. Finally, we will discuss approaches for validating that our decision strategies will perform as expected when deployed in the real world.

1.6.3 Model Uncertainty

In our discussion of sequential decision problems, we have assumed that the transition and reward models are known. In many problems, however, the dynamics and rewards are not known exactly, and the agent must learn to act through experience. By observing the outcomes of its actions in the form of state transitions and rewards, the agent is to choose actions that maximize its long-term accumulation of rewards. Solving such problems in which there is model uncertainty is the subject of the field of *reinforcement learning* and the focus of this part of the book. We will discuss several challenges in addressing model uncertainty. First, the agent must carefully balance exploration of the environment with the exploitation of that knowledge gained through experience. Second, rewards may be received long after the important decisions have been made, so credit for later rewards must be assigned to earlier decisions. Third, the agent must generalize from limited experience. We will review the theory and some of the key algorithms for addressing these challenges.

1.6.4 State Uncertainty

In this part, we extend uncertainty to include the state. Instead of observing the state exactly, we receive observations that have only a probabilistic relationship with the state. Such problems can be modeled as a *partially observable Markov decision process* (*POMDP*). A common approach to solving POMDPs involves inferring a belief distribution over the underlying state at the current time step and then applying a policy that maps beliefs to actions. This part of the book begins by discussing how to update our belief distribution given a past sequence of observations and actions. It then discusses a variety of exact and approximate methods for solving POMDPs.

1.6.5 Multiagent Systems

Up to this point, there has only been one agent making decisions within the environment. This part expands the previous four parts to multiple agents. We begin by discussing simple games, where a group of agents simultaneously each select an action. The result is an individual reward for each agent based on the combined joint action from everyone. The *Markov game* (*MG*) represents a generalization of both simple games to multiple states and the MDP to multiple

agents. Consequently, the agents select actions that can stochastically change the state of a shared environment. Algorithms for MGs rely on reinforcement learning due to uncertainty about the policies of the other agents. A *partially observable Markov game* (*POMG*) introduces state uncertainty, further generalizing MGs and POMDPs, as agents now only receive noisy local observations. The *decentralized partially observable Markov decision process* (*Dec-POMDP*) focuses the POMG on a collaborative multiagent team where there is a shared reward among the agents. This part of the book presents these four categories of problems and discusses exact and approximate algorithms that solve them.