

# Decision Making and Reinforcement Learning

## Module 1: Decision Making and Utility Theory

Tony Dear, Ph.D.

Department of Computer Science  
School of Engineering and Applied Sciences

# Topics

---

- Rational agents and environments
- Utilities and maximization of expected utility
- Preferences and axioms of utility theory
- Uncertain and multi-attribute utilities
- Value of perfect information

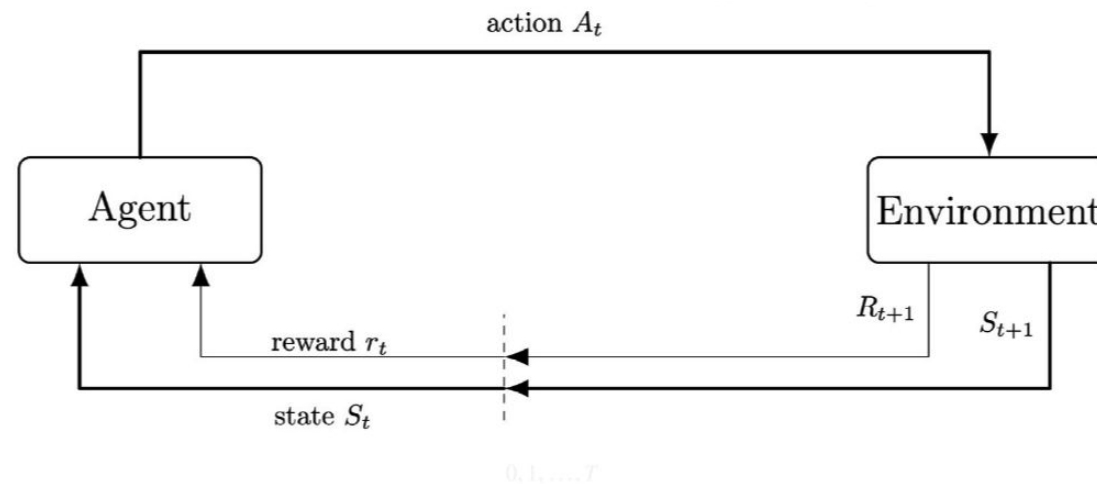
# Learning Objectives

---

- **Describe** rational decision-making as maximization of expected utility
- **List** and **understand** the axioms that govern rational preferences and existence of utility functions
- **Understand** properties of uncertain and multi-attribute utility functions
- **Compute** the value of perfect information in an information-gathering problem

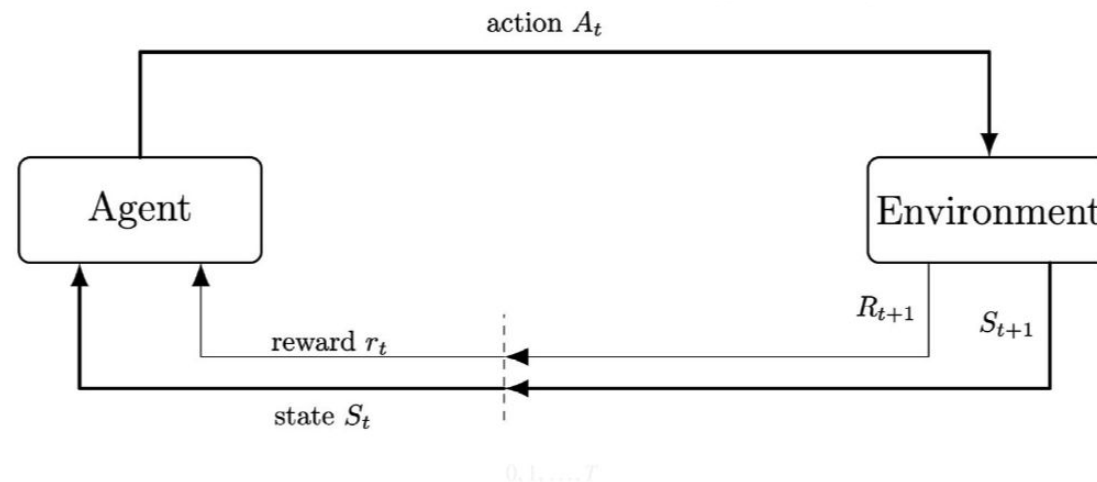
# Agent-Environment Interface

- *Decision-making* is the process in which an **agent** performs an *action* or set of actions in an **environment**



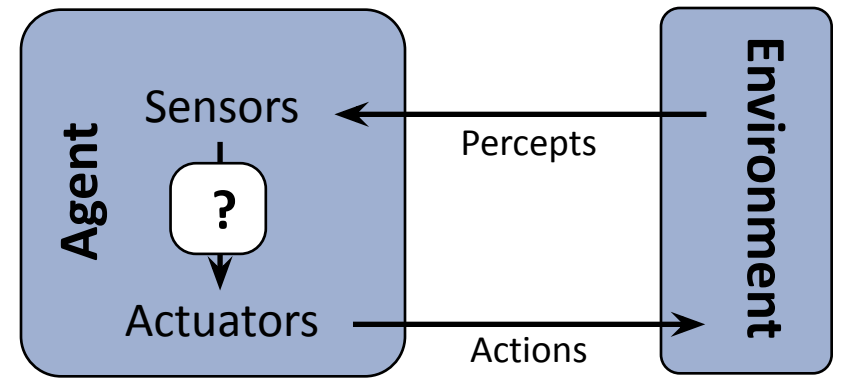
# Agent-Environment Interface

- *Decision-making* is the process in which an **agent** performs an *action* or set of actions in an **environment**
- The action may change the **state** of the agent and environment
- The agent can receive percepts from the environment, e.g., *rewards*



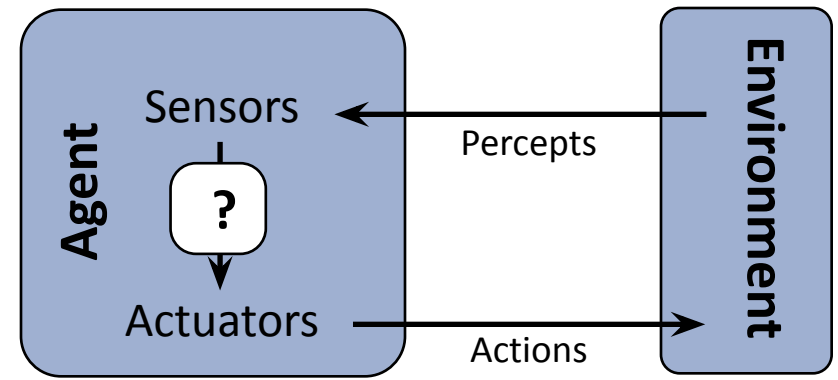
# Agent Functions

- What do we need to know about an agent to consider decision-making?
- An agent may have **sensors** and **actuators**



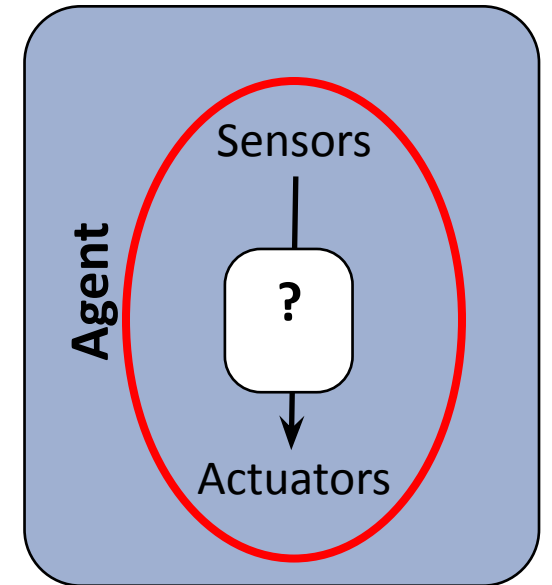
# Agent Functions

- What do we need to know about an agent to consider decision-making?
- An agent may have **sensors** and **actuators**
- Agent's actions depend on its percepts
- May even store an entire *percept sequence*
- An **agent function** maps percept sequences to action



# Agent Programs

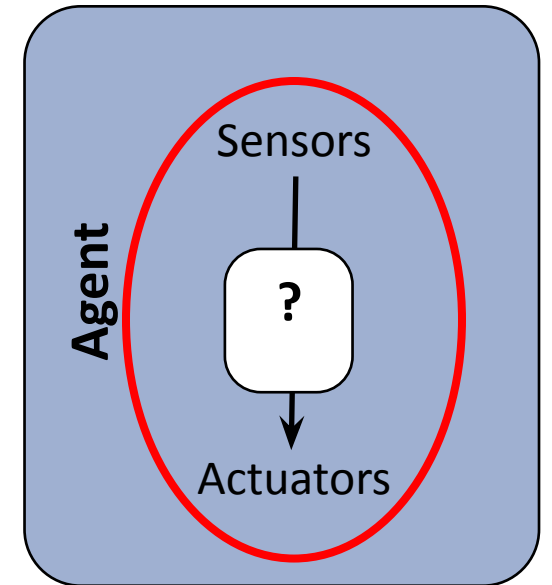
- **Agent programs** (percept to action) *implement* agent functions (percept sequence to action)
- One idea: Lookup table with all possible percept sequences





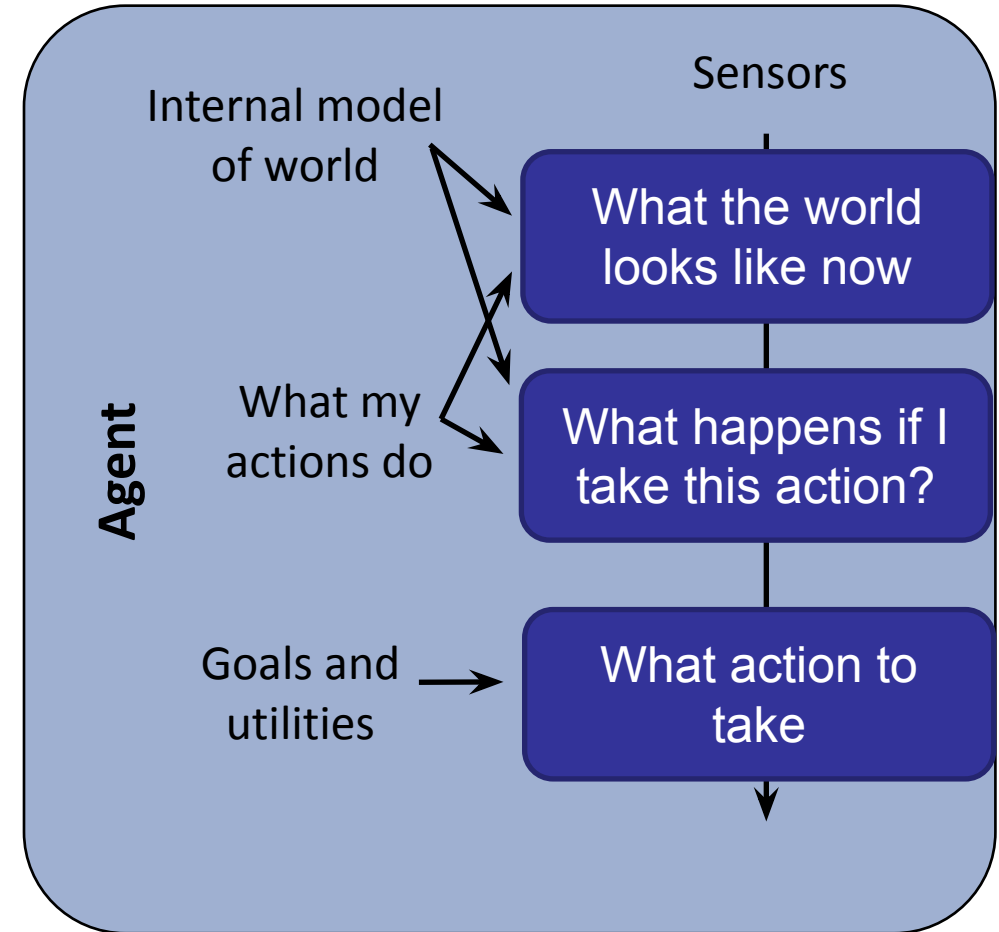
# Agent Programs

- **Agent programs** (percept to action) *implement* agent functions (percept sequence to action)
- One idea: Lookup table with all possible percept sequences
- Program usefulness depends on hardware, limitations
- E.g., may be impossible to implement a program to solve chess on a slow PC



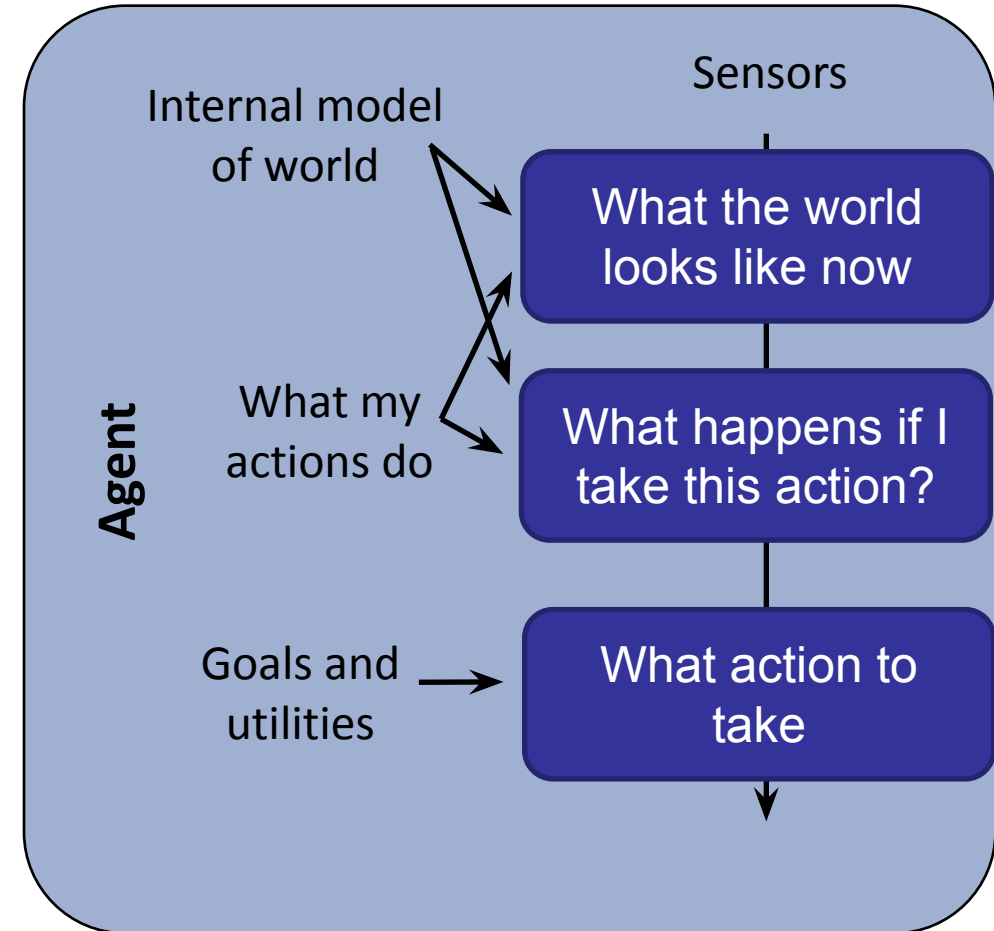
# Goals and Utilities

- An agent program may need to store and use internal models of the world
- These models can help the agent update its state and consider action consequences



# Goals and Utilities

- An agent program may need to store and use internal models of the world
- These models can help the agent update its state and consider action consequences
- Finally, we may have goals and utilities
- Decision-making is performed so as to achieve goals or maximize utilities



# Utilities

---

- **Utility function**  $U: S \rightarrow \mathbb{R}$ : Mapping from state to real numbers

# Utilities

---

- **Utility function**  $U: S \rightarrow \mathbb{R}$ : Mapping from state to real numbers
- Utilities describe preferences and goals, as opposed to behaviors
- Capture long-term consequences, as opposed to rewards

# Utilities

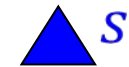
---

- **Utility function**  $U: S \rightarrow \mathbb{R}$ : Mapping from state to real numbers
- Utilities describe preferences and goals, as opposed to behaviors
- Capture long-term consequences, as opposed to rewards
- Principle of **maximum expected utility**: A rational agent chooses actions so as to maximize *expected* utility, given its knowledge

# Maximizing Expected Utility

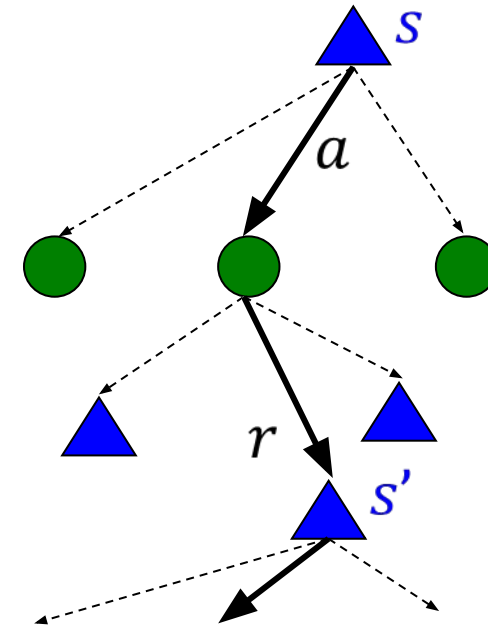
---

- MEU tells what an agent *should* do, but it doesn't solve the problem 😞
- Suppose our agent is currently in a state  $s$



# Maximizing Expected Utility

- MEU tells what an agent *should* do, but it doesn't solve the problem 😞
- Suppose our agent is currently in a state  $s$
- If it takes action  $a$ , it may end up in one of several possible *successor states*  $s'$  according to a *transition model*

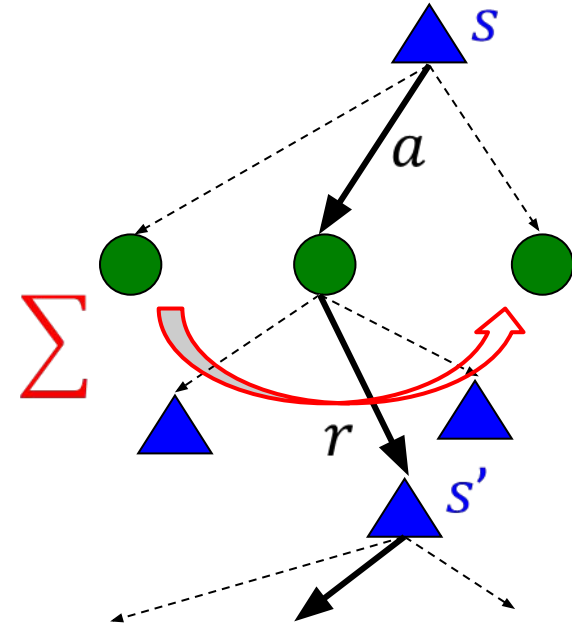




# Maximizing Expected Utility

- The *expected utility of an action  $a$*  is the weighted average utility over all  $s'$ :

$$EU(a) = \sum_{s'} \Pr(s') U(s')$$



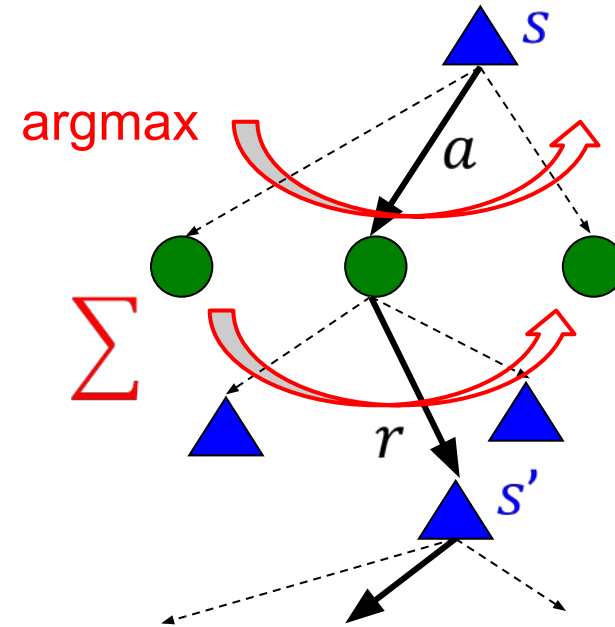
# Maximizing Expected Utility

- The *expected utility of an action  $a$*  is the weighted average utility over all  $s'$ :

$$EU(a) = \sum_{s'} \Pr(s') U(s')$$

- To *maximize  $EU$* , we choose the “best” action—easier said than done!

$$a^* = \operatorname{argmax}_a EU(a)$$



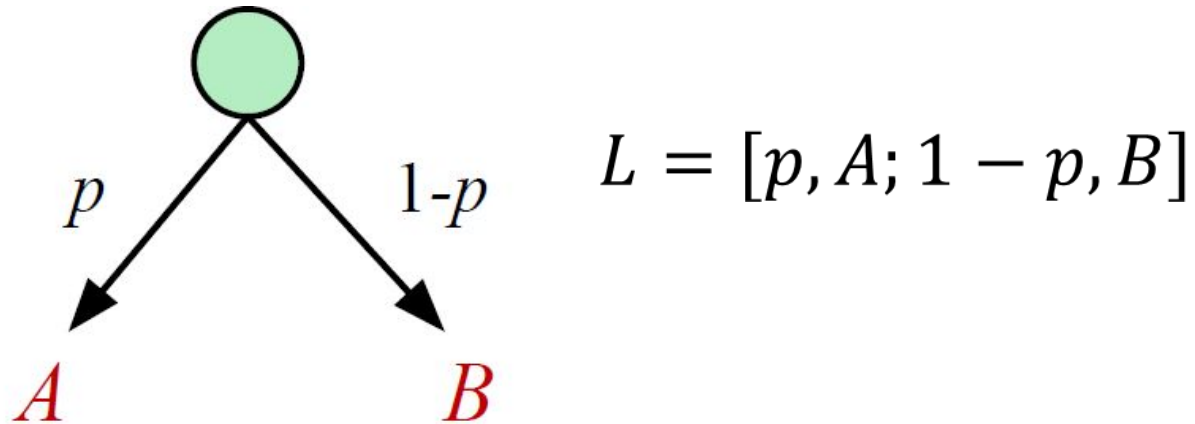
# Outcomes and Lotteries

---

- Utilities ultimately express an agent's *preferences* among different states
- We can also have uncertainty among multiple states or outcomes

# Outcomes and Lotteries

- Utilities ultimately express an agent's *preferences* among different states
- We can also have uncertainty among multiple states or outcomes
- A **lottery** is a set of possible outcomes with associated probabilities
- An agent has preferences over both definite outcomes and lotteries



# Rational Preferences

---

- We will use the following notation to express preferences:

--	--	--

# Rational Preferences

- We will use the following notation to express preferences:

--	--	--

- *Rational* preferences must satisfy certain axioms!
- Orderability:  $A \succ B$  or  $B \succ A$  or  $A \sim B$
- Transitivity:  $A \succ B$  and  $B \succ C$  implies  $A \succ C$
- Continuity:  $A \succ B \succ C$  implies  $\exists p [p, A; 1 - p, C] \sim B$

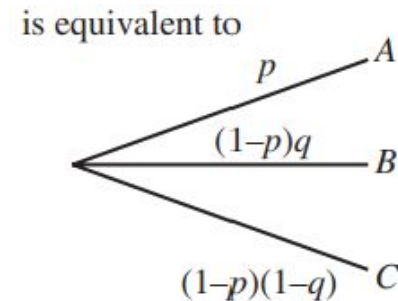
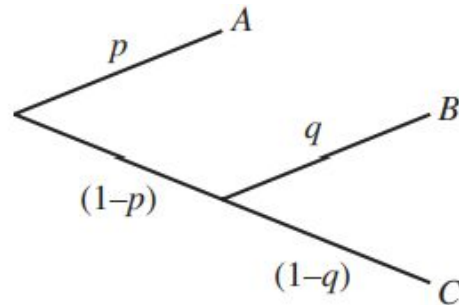
# Axioms of Utility Theory

---

- *Rational* preferences must satisfy certain axioms!
- Substitutability:  $A \sim B$  implies  $[p, A; 1 - p, C] \sim [p, B; 1 - p, C]$
- Monotonicity:  $A \succ B$  implies ( $p > q$  iff  $[p, A; 1 - p, B] \succ [q, A; 1 - q, B]$ )

# Axioms of Utility Theory

- *Rational* preferences must satisfy certain axioms!
- Substitutability:  $A \sim B$  implies  $[p, A; 1 - p, C] \sim [p, B; 1 - p, C]$
- Monotonicity:  $A \succ B$  implies ( $p > q$  iff  $[p, A; 1 - p, B] \succ [q, A; 1 - q, B]$ )
- Decomposability:  $[p, A; 1 - p, [q, B; 1 - q, C]] \sim [p, A; (1 - p)q, B; (1 - p)(1 - q), C]$





# Irrational Preferences

---

- Preferences that do not preserve all the previous axioms may yield irrational behavior
- Suppose that an agent has preferences among three goods:
  - $A \succ B, B \succ C, C \succ A$
- These are *intransitive* preferences
- Each one is incompatible with the other two

# Irrational Preferences

---

- If the agent has good C, it would trade it away for good B for \$X
- Now suppose it is offered good A; it would trade B away for A for \$X
- It would do the same to retrieve good C
- The agent is back to where it started, less \$3X!

# Existence of Utilities

---

- von Neumann and Morgenstern, 1944: Given a set of outcomes  $S_1, \dots, S_n$  satisfying the preceding axioms, there exists a *utility function*  $U$  such that

$$U(S_i) \geq U(S_j) \Leftrightarrow S_i \succeq S_j$$

$$U[p_1, S_1; \dots; p_n, S_n] = \sum_i p_i U(S_i)$$

# Existence of Utilities

- von Neumann and Morgenstern, 1944: Given a set of outcomes  $S_1, \dots, S_n$  satisfying the preceding axioms, there exists a *utility function*  $U$  such that

$$U(S_i) \geq U(S_j) \Leftrightarrow S_i \succeq S_j$$

$$U[p_1, S_1; \dots; p_n, S_n] = \sum_i p_i U(S_i)$$

- Values assigned by  $U$  preserve preferences over prizes and lotteries
- $U$  is not unique! Agent behaviors do not change if we replace  $U$  with a *positive affine transformation* of it:  $U'(S) = aU(S) + b, a > 0$

# Preference Elicitation

---

- Utility functions are guaranteed to exist, but how to come up with one?
- Suppose we have a *standard lottery with normalized utilities*:

$$[p, u_{\top}; 1 - p, u_{\perp}]$$

# Preference Elicitation

---

- Utility functions are guaranteed to exist, but how to come up with one?
- Suppose we have a *standard lottery* with *normalized utilities*:

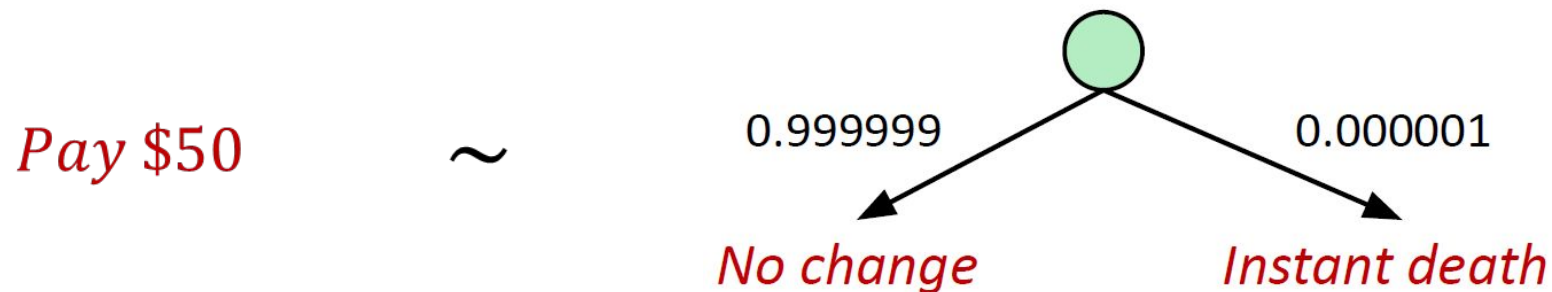
$$[p, u_{\top}; 1 - p, u_{\perp}]$$

- $u_{\top} = 1$  corresponds to “best possible prize”,  $u_{\perp} = 0$  to “worst possible outcome”
- The utility of a prize  $S$  is the value  $p$  s.t.  $S \sim [p, u_{\top}; 1 - p, u_{\perp}]$

# Preference Elicitation

$$S \sim [p, u_{\top}; 1 - p, u_{\perp}]$$

- Ex: Research has shown that people value a 1-in-a-million chance of death (a *micromort*) at about \$50
- Many activities have an associated micromort assignment



# Utility of Money

---

- Money typically does not behave exactly like a utility function
- Consider the lotteries  $L_1 = [0.5, \$2.1M; 0.5, \$0]$  vs  $L_2 = [1, \$1M]$
- $L_1$  has a higher expected monetary value than  $L_2$ , but which would you choose?



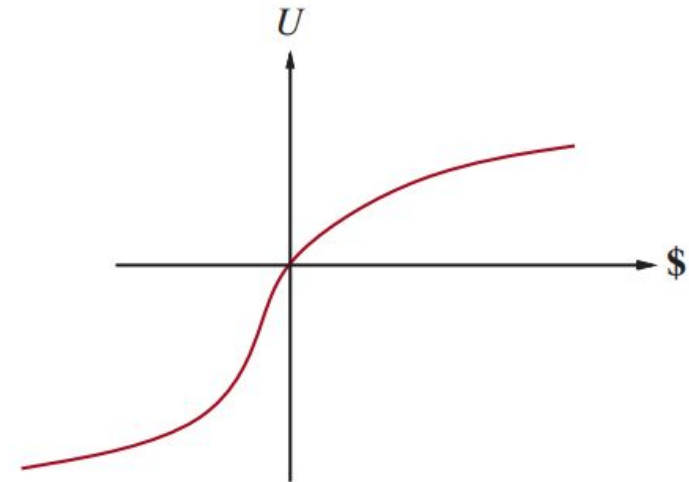
# Utility of Money

---

- Money typically does not behave exactly like a utility function
- Consider the lotteries  $L_1 = [0.5, \$2.1M; 0.5, \$0]$  vs  $L_2 = [1, \$1M]$
- $L_1$  has a higher expected monetary value than  $L_2$ , but which would you choose?
- Most people would choose  $L_2$  because they are *risk-averse*
- Utilities increase more slowly than dollar amounts

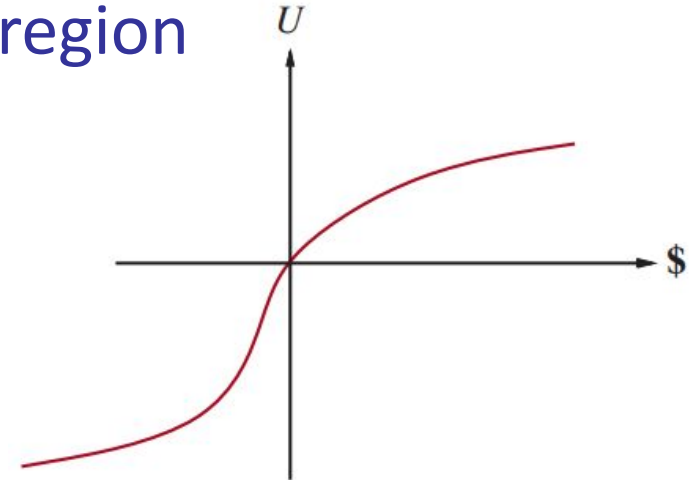
# Utility of Money

- A wealthier person may have a linear utility curve on a larger range
- Differences in risk acceptance give rise to insurance premiums



# Utility of Money

- A wealthier person may have a linear utility curve on a larger range
- Differences in risk acceptance give rise to insurance premiums
- People have concave utility curves for expensive products
- Curves for insurance companies are linear in the same region



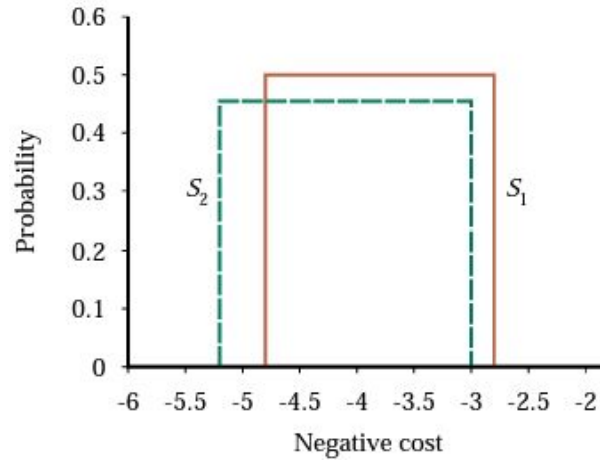
# Uncertain Utilities

---

- 
- We say that outcome  $S_1$  **strictly dominates** outcome  $S_2$  if  $U(S_1) > U(S_2)$

# Uncertain Utilities

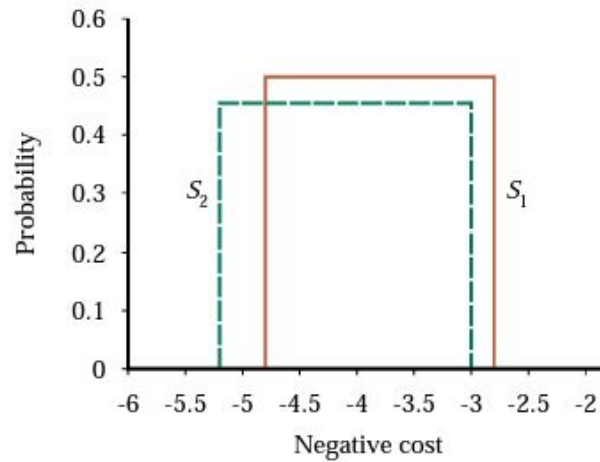
■  
Probability density  
function (pdf)



- We say that outcome  $S_1$  **strictly dominates** outcome  $S_2$  if  $U(S_1) > U(S_2)$
- What if the utilities are uncertain and described as probability distributions  $p_1(x)$  and  $p_2(x)$  over an attribute  $X$ ?

# Uncertain Utilities

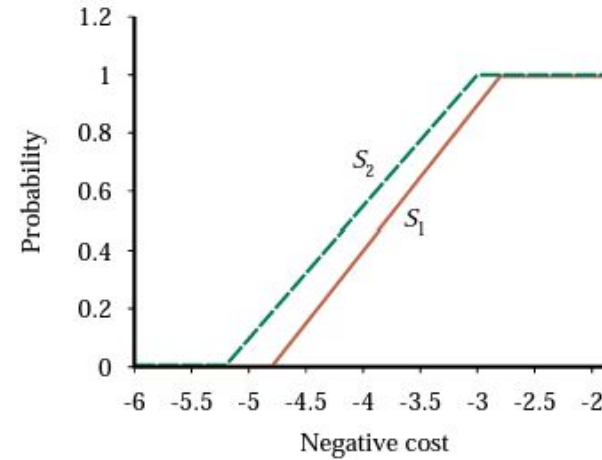
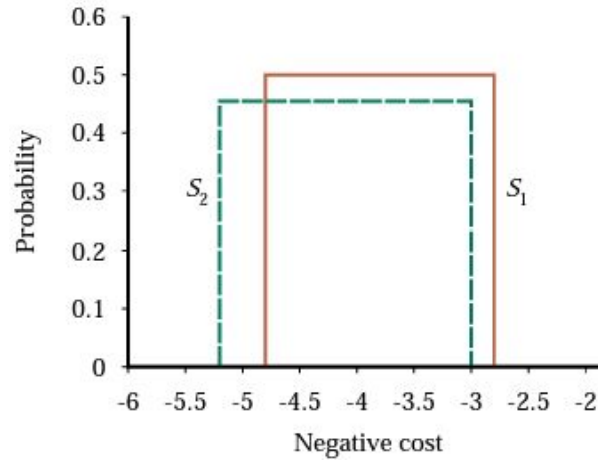
■  
Probability density  
function (pdf)



- $S_1$  **stochastically dominates**  $S_2$  if  $\Pr(S_1 \geq x) \geq \Pr(S_2 \geq x)$  for all  $x$

# Uncertain Utilities

■  
Probability density  
function (pdf)



Cumulative  
distribution  
function (cdf)

- $S_1$  **stochastically dominates**  $S_2$  if  $\Pr(S_1 \geq x) \geq \Pr(S_2 \geq x)$  for all  $x$
- The *cumulative distribution function* of  $S_1$  is smaller than or equal to that of  $S_2$  for all  $x$

# Multi-attribute Utilities

---

- An outcome may be described by *multiple* attributes  $\mathbf{X} = X_1, \dots, X_n$
- E.g., job A: \$150k salary, 2 wks vacation; job B: \$130k salary, 4 wks vacation



# Multi-attribute Utilities

---

- An outcome may be described by *multiple* attributes  $\mathbf{X} = X_1, \dots, X_n$
- E.g., job A: \$150k salary, 2 wks vacation; job B: \$130k salary, 4 wks vacation
- An outcome is preferable to another if it is stochastically dominant across all attributes
- Otherwise, we may need a **multi-attribute utility function**  $U(x_1, \dots, x_n)$

# Multi-attribute Utilities

---

- The size of multi-attribute utility functions can grow *exponentially*
- If we have  $n$  attributes with  $d$  values each, this function must be defined for  $d^n$  values!

# Multi-attribute Utilities

---

- The size of multi-attribute utility functions can grow *exponentially*
- If we have  $n$  attributes with  $d$  values each, this function must be defined for  $d^n$  values!

- Special case: If attributes are **additive independent**, then we can write

$$U(x_1, \dots, x_n) = \sum_{i=1}^n k_i U_i(x_i)$$

- Uncertain attributes with weaker forms of independence may lead to *multiplicative* utility functions

# Value of Information

---

- Suppose current best action for a problem is  $\alpha$ :  $EU(\alpha) = \max_a EU(a)$
- Suppose we can learn new information that may change  $\alpha$

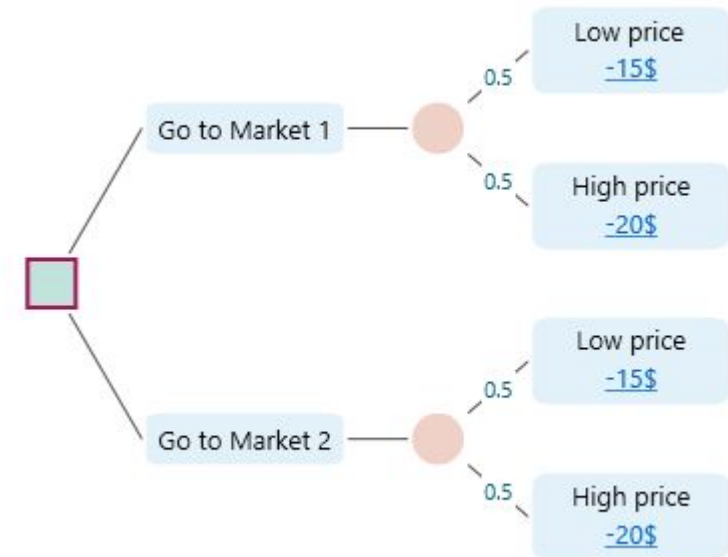
# Value of Information

---

- Suppose current best action for a problem is  $\alpha$ :  $EU(\alpha) = \max_a EU(a)$
- Suppose we can learn new information that may change  $\alpha$
- We might learn something about a *random variable*  $E$  with possible outcomes  $e_i$
- We may change our action and thus utility depending on  $e_i$

# Value of Information

- For each value  $E = e_i$ , the (possibly new) best action is  $\alpha_i$ :  $EU(\alpha_i|e_i) = \max_a EU(a|e_i)$



# Value of Information

- For each value  $E = e_i$ , the (possibly new) best action is  $\alpha_i$ :  $EU(\alpha_i|e_i) = \max_a EU(a|e_i)$
- **Value of perfect information (VPI)** is the *expected* improvement in expected utility:

$$VPI(E) = \left( \sum_e \Pr(E = e) EU(\alpha_e|e) \right) - EU(\alpha)$$
$$VPI(E) = \left( \sum_e \Pr(E = e) EU(\alpha_e|e) \right) - EU(\alpha)$$

# Example: VPI

---

- An oil company is trying to choose one of  $n$  possible drilling sites
- Each site may contain oil with probability  $\frac{1}{n}$
- If the net profit of finding oil is  $C$ , then the  $EU$  of drilling in any site is  $C/n$



# Example: VPI

---

- An oil company is trying to choose one of  $n$  possible drilling sites
- Each site may contain oil with probability  $\frac{1}{n}$
- If the net profit of finding oil is  $C$ , then the  $EU$  of drilling in any site is  $C/n$
- A seismologist offers to survey one site and definitively find out if it contains oil
- What are the possible outcomes?

# Example: VPI

---

- If oil is found ( $p = 1/n$ ), best action is to drill there to obtain utility  $C$
- If oil is *not* found ( $p = \frac{n-1}{n}$ ), best action is to *not* drill there to obtain expected utility  $\frac{C}{n-1}$

# Example: VPI

- If oil is found ( $p = 1/n$ ), best action is to drill there to obtain utility  $C$
- If oil is *not* found ( $p = \frac{n-1}{n}$ ), best action is to *not* drill there to obtain expected utility  $\frac{C}{n-1}$
- The new  $EU$  is thus  $\frac{1}{n} \times C + \frac{n-1}{n} \times \frac{C}{n-1} = \frac{2C}{n}$
- The VPI is the *difference* between new and old  $EU$ :  $VPI = \frac{2C}{n} - \frac{C}{n} = \frac{C}{n}$

# Properties of VPI

---

- Similar analysis can be applied in any information gathering scenario
- Ex: Should a doctor order more tests to be done on a patient?
- Ex: Should an investment firm hire a consultant to better understand the market?

# Properties of VPI

---

- Similar analysis can be applied in any information gathering scenario
- Ex: Should a doctor order more tests to be done on a patient?
- Ex: Should an investment firm hire a consultant to better understand the market?
- Theorem: VPI is non-negative; it is never disadvantageous to acquire more information
- Maximization of VPI can be used to design an *information-gathering agent*

# Summary

---

- An agent's underlying preferences must satisfy certain axioms in order to be considered rational
- Rational preferences lead to utility functions; we maximize expected utility
- Utilities may be uncertain or may be described by multiple attributes
- Value of information can be quantified by expected gain in expected utility