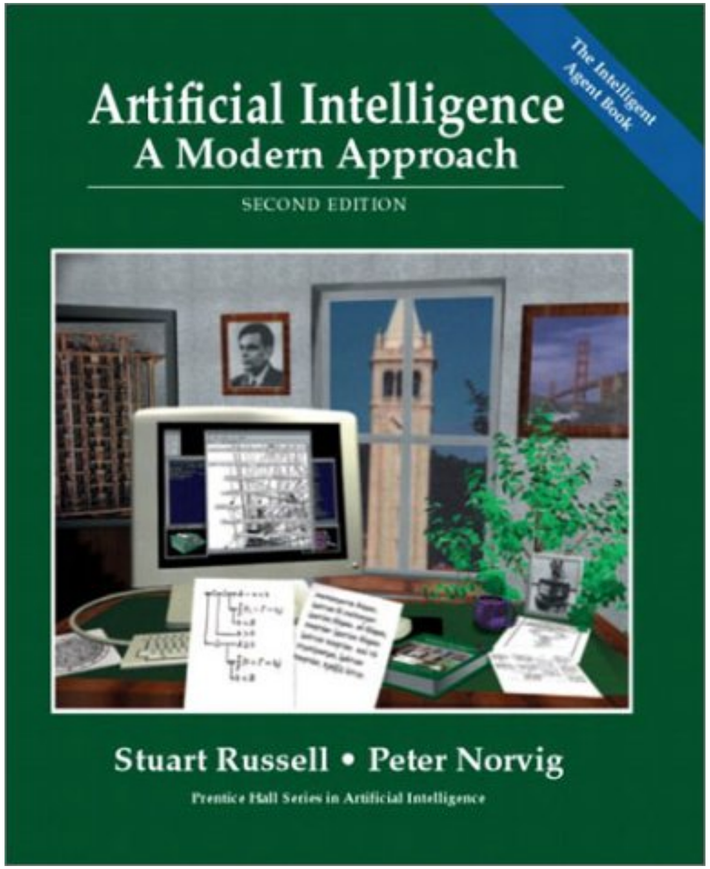
Intelligent Autonomous Agents

Agents and Rational Behavior

Lecture 7: Decision-Making under Uncertainty
Simple Decisions

Ralf Möller, Rainer Marrone
Hamburg University of Technology

Literature



- Chapter 16

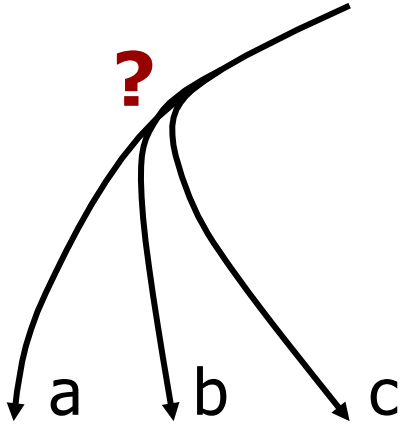
Material from Lise Getoor, Jean-Claude Latombe, Daphne Koller, and Stuart Russell

Decision Making Under Uncertainty

- Many environments have multiple possible outcomes
- Some of these outcomes may be good; others may be bad
- Some may be very likely; others unlikely

- What's a poor agent going to do??

Non-Deterministic vs. Probabilistic Uncertainty

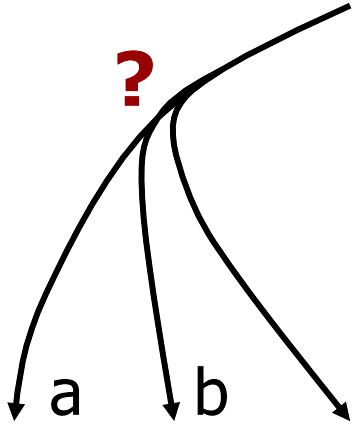


{a,b,c}

→ decision that is best for worst case

Non-deterministic model

~ Adversarial search



$\{a(p_a), b(p_b), c(p_c)\}$

→ decision that maximizes expected utility value

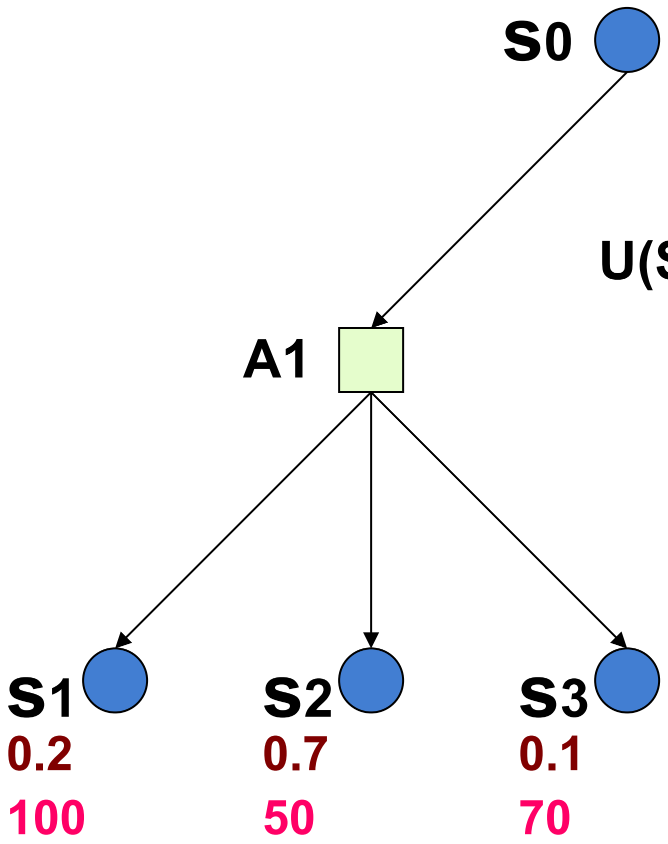
Probabilistic model

Expected Utility

- Random variable X with n values $x_1, \ldots, x_n$ and distribution $(p_1, \ldots, p_n)$
  E.g.: X is the state reached after doing an action A under uncertainty
- Function U of X
  E.g., U is the utility of a state
- The expected utility of A is

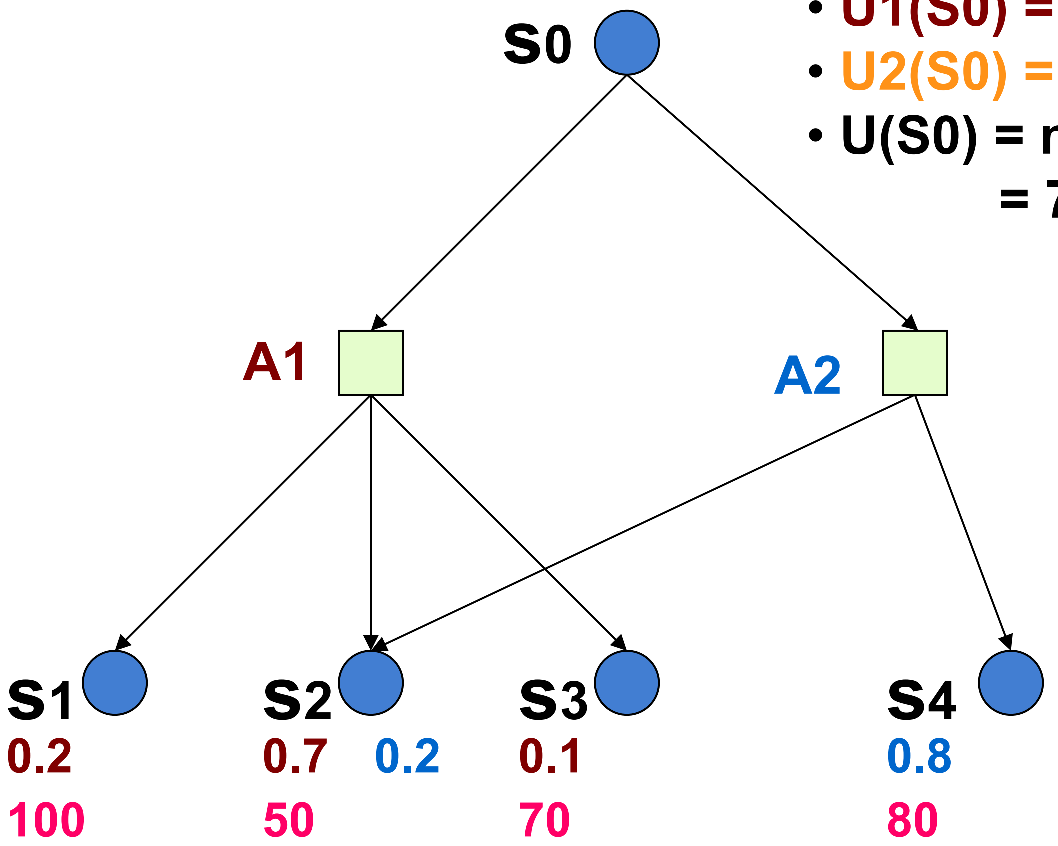
$$EU[A=a] = \Sigma_{i=1,\ldots,n}\, p(x_i|A=a)U(x_i)$$

One State/One Action Example



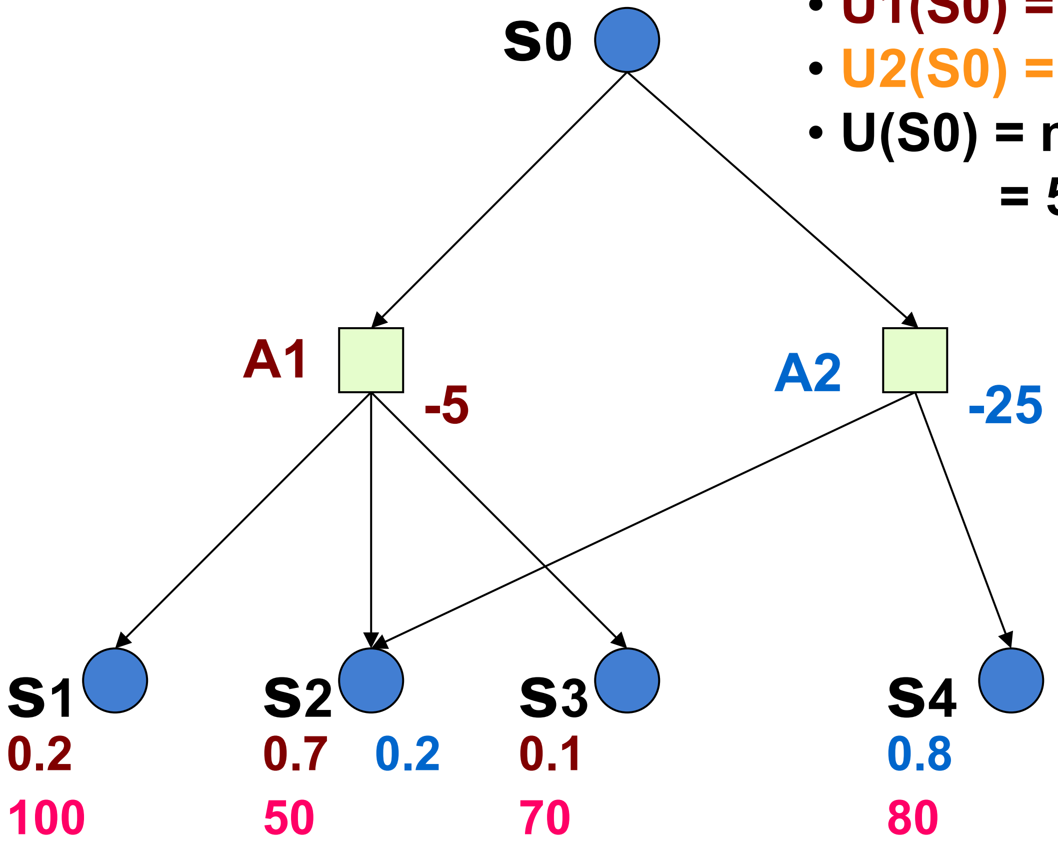
$$U(S0) = 100 \times 0.2 + 50 \times 0.7 + 70 \times 0.1$$
$$= 20 + 35 + 7$$
$$= 62$$

One State/Two Actions Example



- U1(S0) = 62
- U2(S0) = 74
- U(S0) = max{U1(S0),U2(S0)}
  = 74

Introducing Action Costs



- U1(S0) = 62 – 5 = 57
- U2(S0) = 74 – 25 = 49
- U(S0) = max{U1(S0),U2(S0)}
  = 57

MEU Principle

AI is Solved!!!

- A **rational agent** s… …ose the action t… …s expect…
- This is… …eld of **decisi…**
- The M… …ciple provides a **normative criterion** for rational choice of action

Not quite…

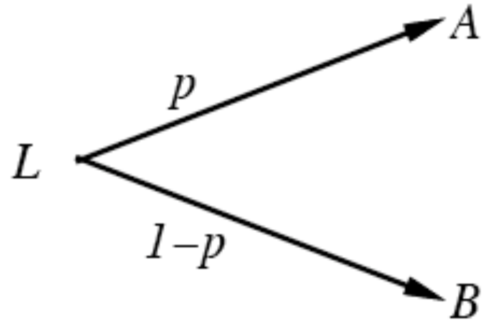
- Must have **complete** model of:
  - Actions
  - Utilities
  - States
- Even if you have a complete model, it might be computationally **intractable**
- In fact, a truly rational agent takes into account the utility of reasoning as well---**bounded rationality**
- Nevertheless, great progress has been made in this area recently, and we are able to solve much more complex decision-theoretic problems than ever before

We'll look at

- Decision-Theoretic Planning
  - Simple decision making (ch. 16)
  - Sequential decision making (ch. 17)

Preferences

An agent chooses among prizes (A, B, etc.) and lotteries, i.e., situations with uncertain prizes



Lottery $L = [p, A;\ (1-p), B]$

Notation:

$A \succ B$	A preferred to B
$A \sim B$	indifference between A and B
$A \succsim B$	B not preferred to A

A and B can be lotteries again: Prizes are special lotteries: [1, X; 0, not X]

Rational preferences

Idea: preferences of a rational agent must obey constraints.

Rational preferences $\Rightarrow$

behavior describable as maximization of expected utility

Rational preferences contd.

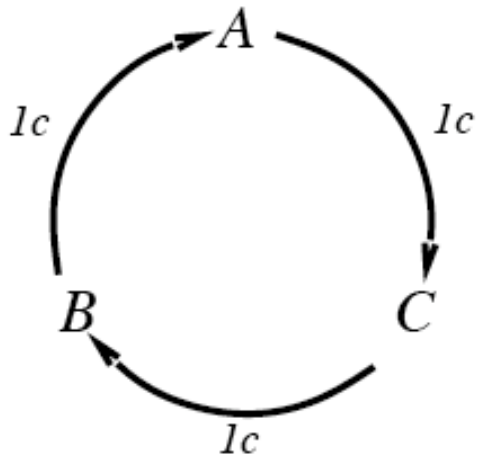
Violating the constraints leads to self-evident irrationality

For example: an agent with intransitive preferences can be induced to give away all its money

If $B \succ C$, then an agent who has C would pay (say) 1 cent to get B

If $A \succ B$, then an agent who has B would pay (say) 1 cent to get A

If $C \succ A$, then an agent who has A would pay (say) 1 cent to get C



Axioms of Utility Theory

- Orderability
  - $(A>B) \vee (A<B) \vee (A\sim B)$
- Transitivity
  - $(A>B) \wedge (B>C) \Rightarrow (A>C)$
- Continuity
  - $A>B>C \Rightarrow \exists p\ [p,A;\ 1-p,C] \sim B$
- Substitutability
  - $A\sim B \Rightarrow [p,A;\ 1-p,C]\sim[p,B;\ 1-p,C]$
- Monotonicity
  - $A>B \Rightarrow (p\geq q \Leftrightarrow [p,A;\ 1-p,B] >\sim [q,A;\ 1-q,B])$
- Decomposability
  - $[p,A;\ 1-p,\ [q,B;\ 1-q,\ C]] \sim [p,A;\ (1-p)q,\ B;\ (1-p)(1-q),\ C]$

Decomposability: There is no fun in gambling

And then there was utility

Theorem (Ramsey, 1931; von Neumann and Morgenstern, 1944):
Given preferences satisfying the constraints
there exists a real-valued function U such that

$$U(A) \geq U(B) \quad \Leftrightarrow \quad A \succsim B$$
$$U([p_1, S_1; \ \ldots \ ; \ p_n, S_n]) = \Sigma_i \ p_i U(S_i)$$

MEU principle:

Choose the action that maximizes expected utility

Note: an agent can be entirely rational (consistent with MEU) without ever representing or manipulating utilities and probabilities

E.g., a lookup table for perfect tictactoe

Utilities

Utilities map states to real numbers. Which numbers?

Standard approach to assessment of human utilities:

compare a given state A to a standard lottery L_p that has

"best possible prize" $\top$ with probability p

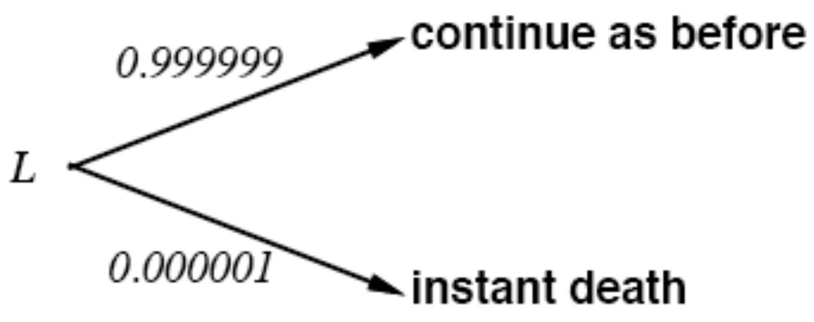
"worst possible catastrophe" $\bot$ with probability $(1-p)$

adjust lottery probability p until $A \sim L_p$

pay \$30 ~

-and-continue

-as-before



Utility scales

Normalized utilities: $u_\top = 1.0$, $u_\perp = 0.0$ U(pay $30...) = 0.999999

Micromorts: one-millionth chance of death
useful for Russian roulette, paying to reduce product risks, etc.

QALYs: quality-adjusted life years
useful for medical decisions involving substantial risk

Note: behavior is **invariant** w.r.t. +ve linear transformation

$$U'(x) = k_1 U(x) + k_2 \quad \text{where } k_1 > 0$$

With deterministic prizes only (no lottery choices), only ordinal utility can be determined, i.e., total order on prizes

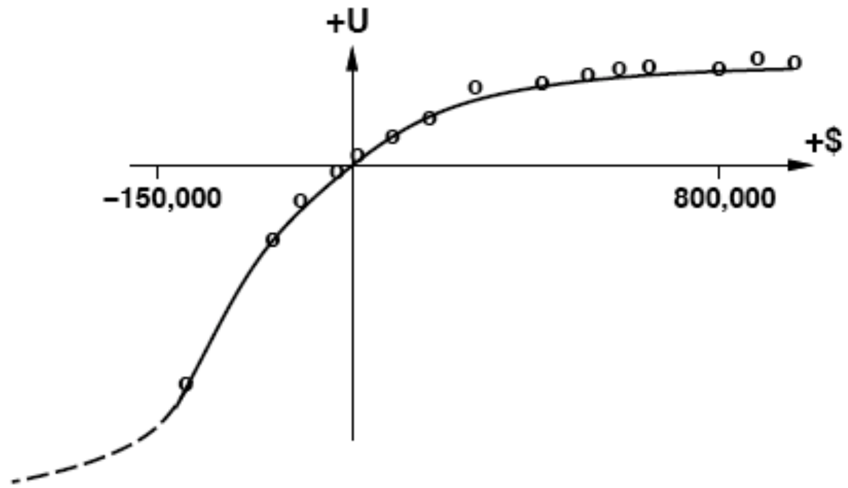
Money

Money does **not** behave as a utility function

Given a lottery L with expected monetary value $EMV(L)$,
usually $U(L) < U(EMV(L))$, i.e., people are risk-averse

Utility curve: for what probability p am I indifferent between a prize x and a lottery $[p, \$M;\ (1-p), \$0]$ for large M?

Typical empirical data, extrapolated with risk-prone behavior:



Money Versus Utility

- Money <> Utility
  - More money is better, but not always in a linear relationship to the amount of money
- Expected Monetary Value
- Risk-averse – $U(L) < U(S_{EMV(L)})$
- Risk-seeking – $U(L) > U(S_{EMV(L)})$
- Risk-neutral – $U(L) = U(S_{EMV(L)})$

Value Functions

- Provides a ranking of alternatives, but not a meaningful metric scale
- Also known as an "ordinal utility function"

- Remember the expectiminimax example:
  - Sometimes, only relative judgments (value functions) are necessary
  - At other times, absolute judgments (utility functions) are required

Multiattribute Utility Theory

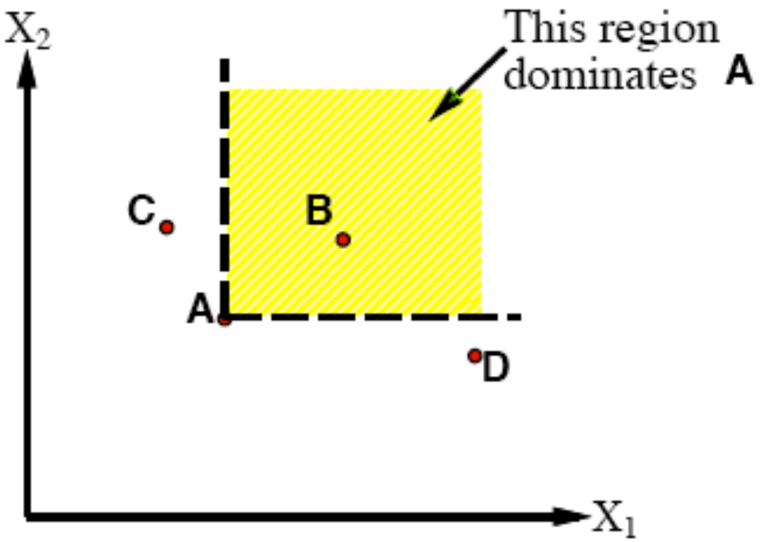
- A given state may have multiple utilities
  - ...because of multiple evaluation criteria
  - ...because of multiple agents (interested parties) with different utility functions

Strict dominance

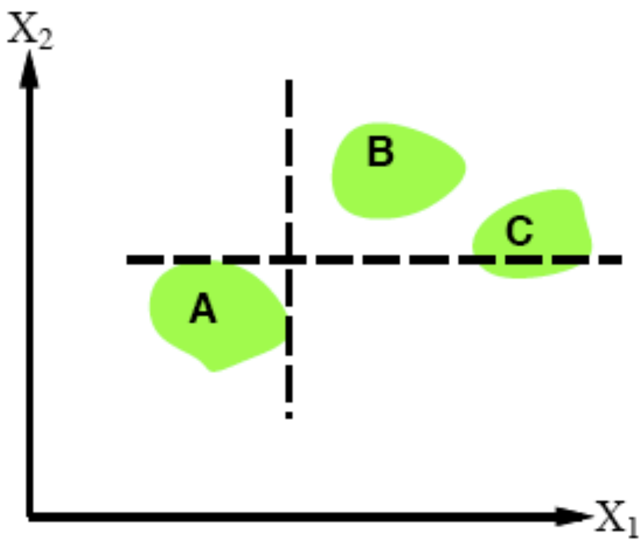
Typically define attributes such that U is monotonic in each

Strict dominance: choice B strictly dominates choice A iff

$$\forall i \;\; X_i(B) \geq X_i(A) \qquad \text{(and hence } U(B) \geq U(A)\text{)}$$

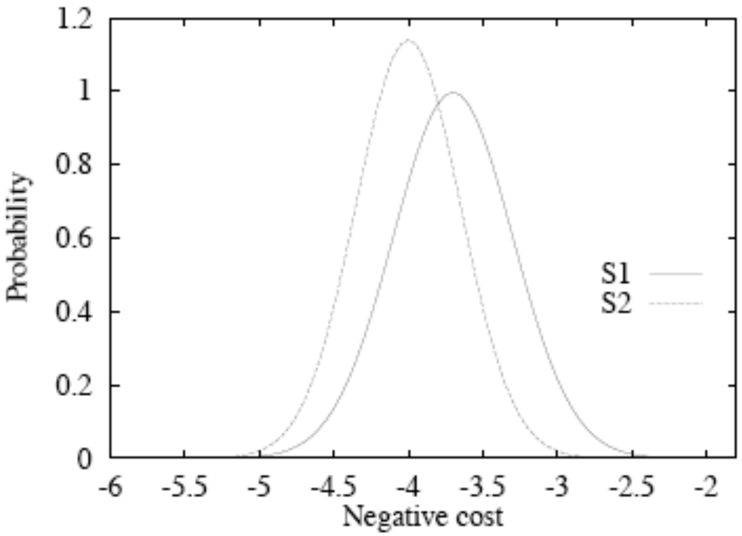


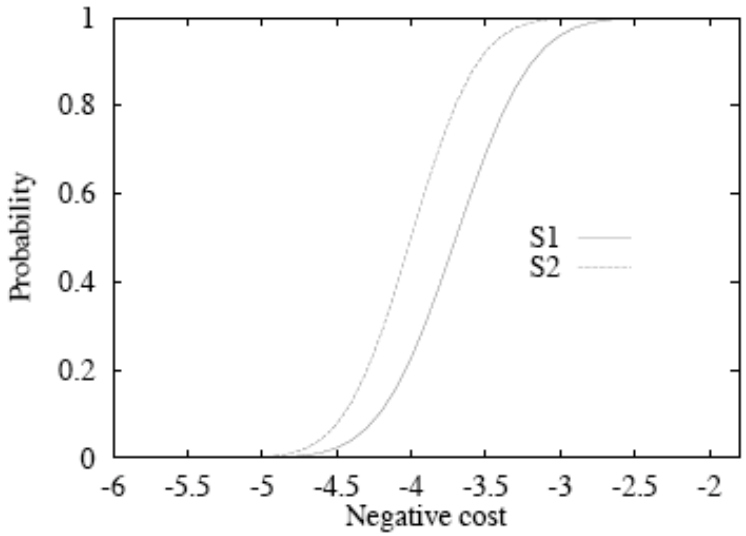
Deterministic attributes



Uncertain attributes

Stochastic dominance





Distribution p_1 stochastically dominates distribution p_2 iff

$$\forall t \quad \int_{-\infty}^{t} p_1(x)dx \leq \int_{-\infty}^{t} p_2(t)dt \qquad \forall x \,.\, P(X_1 >= x) >= P(X_2 >= x)$$

If U is monotonic in x, then A_1 with outcome distribution p_1 stochastically dominates A_2 with outcome distribution p_2:

$$\int_{-\infty}^{\infty} p_1(x)U(x)dx \geq \int_{-\infty}^{\infty} p_2(x)U(x)dx$$

Multiattribute case: stochastic dominance on all attributes $\Rightarrow$ optimal

Stochastic Dominance

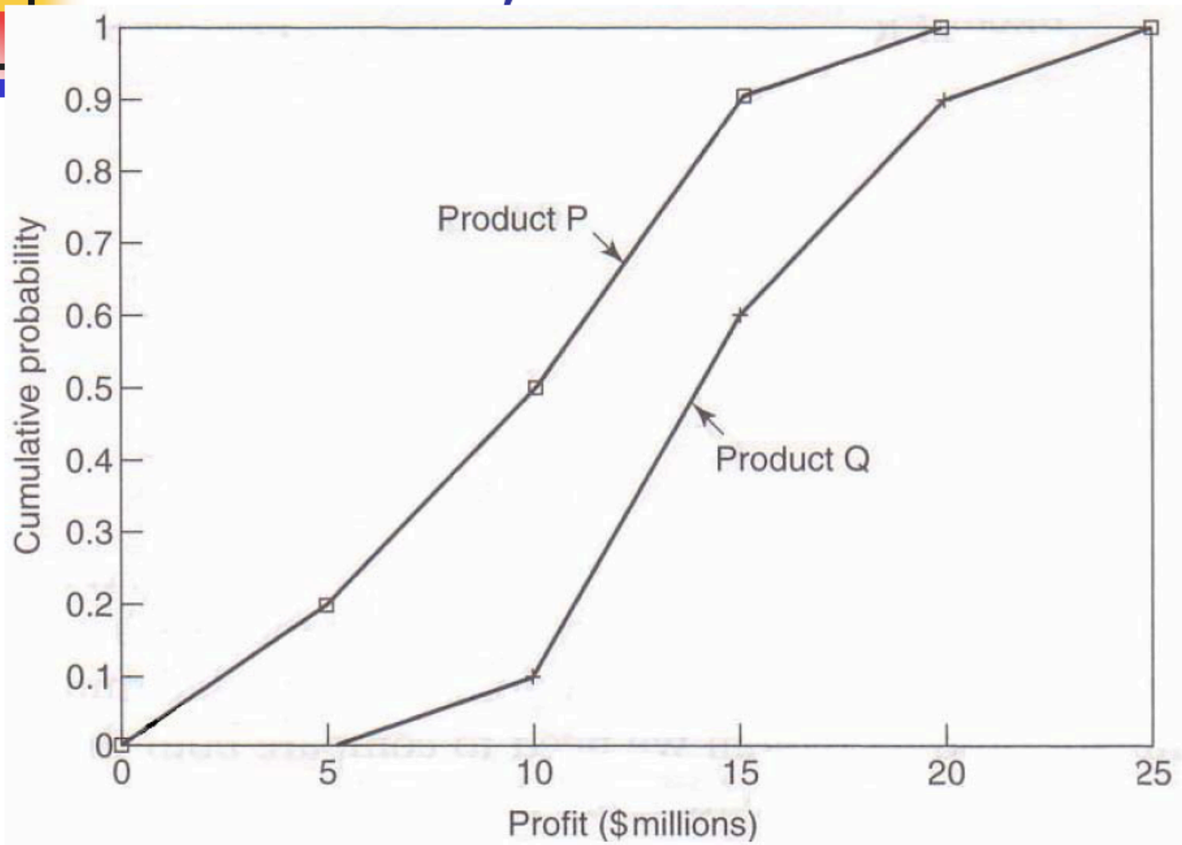
First order stochastic dominance, example

Profit ($m)	Product P Prob.
0 to under 5	0.2
5 to under 10	0.3
10 to under 15	0.4
15 to under 20	0.1

Profit ($m)	Product Q Prob.
0 to under 5	0
5 to under 10	0.1
10 to under 15	0.5
15 to under 20	0.3
20 to under 25	0.1

Stochastic Dominance

- First order stochastic dominance, illustration



Stochastic Dominance

First order stochastic dominance, remarks

- First order stochastic dominance does not imply that every possible return of the superior distribution is larger than every possible return of the inferior distribution
- If F(.) first order stochastically dominates G(.), the expected value of F(.) is higher than the expected value of G(.)
- The reverse is not necessarily true

Stochastic dominance contd.

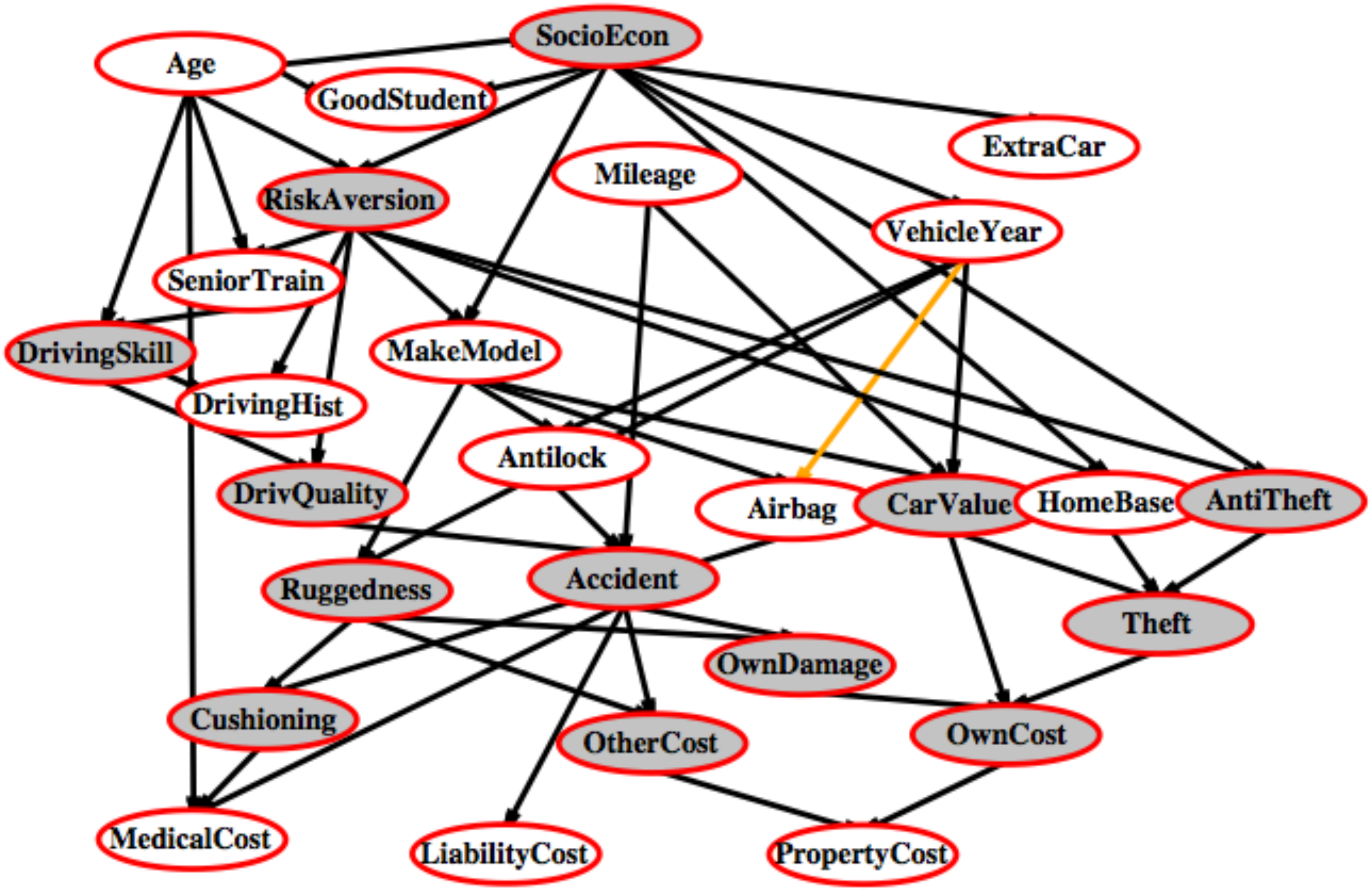
Stochastic dominance can often be determined without exact distributions using **qualitative** reasoning

E.g., construction cost increases with distance from city
S_1 is closer to the city than S_2
$\Rightarrow$ S_1 stochastically dominates S_2 on cost

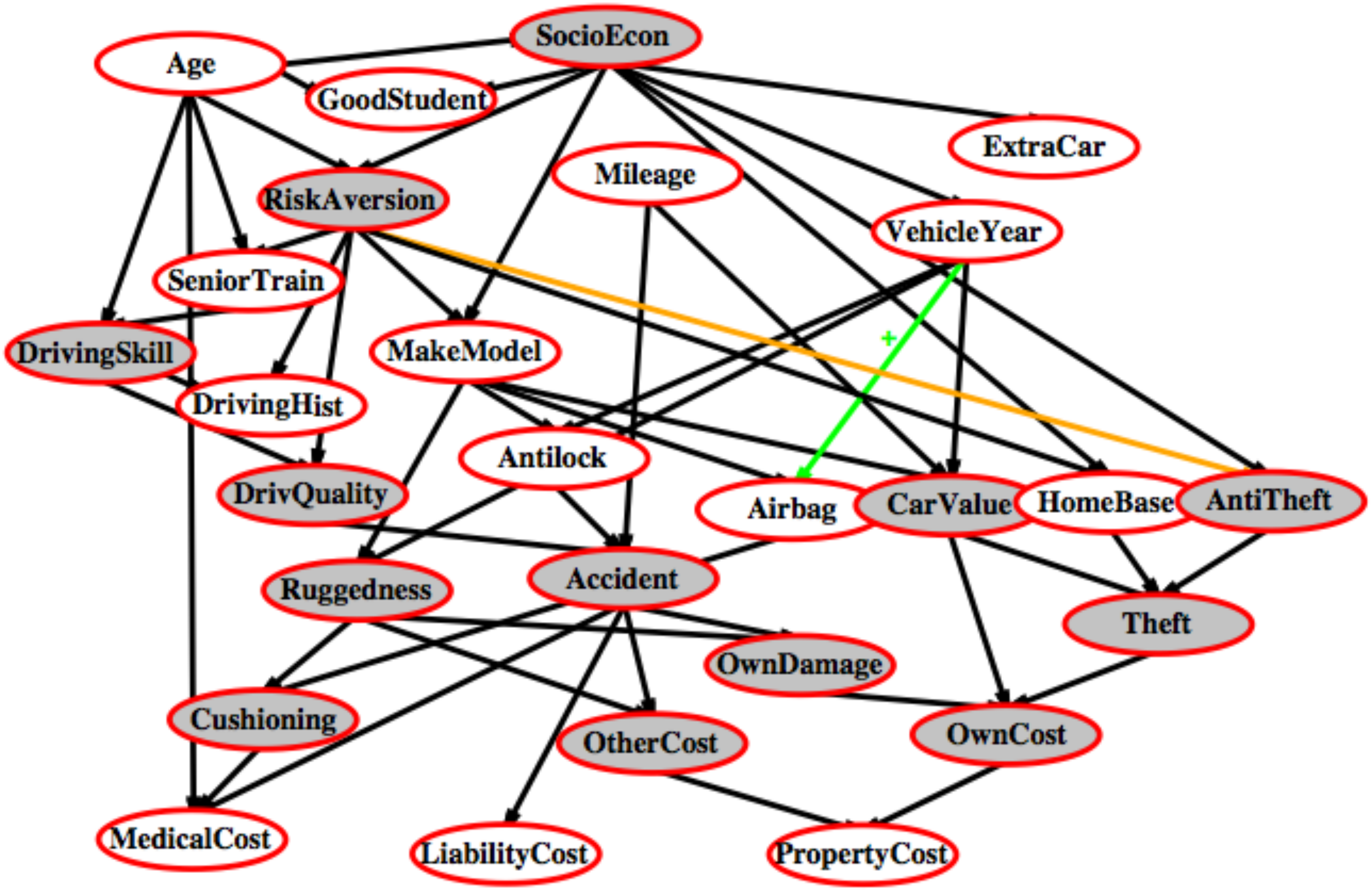
E.g., injury increases with collision speed

Can annotate belief networks with stochastic dominance information:
$X \xrightarrow{+} Y$ (X positively influences Y) means that
For every value $\mathbf{z}$ of Y's other parents $\mathbf{Z}$
$\forall\, x_1, x_2 \;\; x_1 \geq x_2 \Rightarrow \mathbf{P}(Y|x_1, \mathbf{z})$ stochastically dominates $\mathbf{P}(Y|x_2, \mathbf{z})$

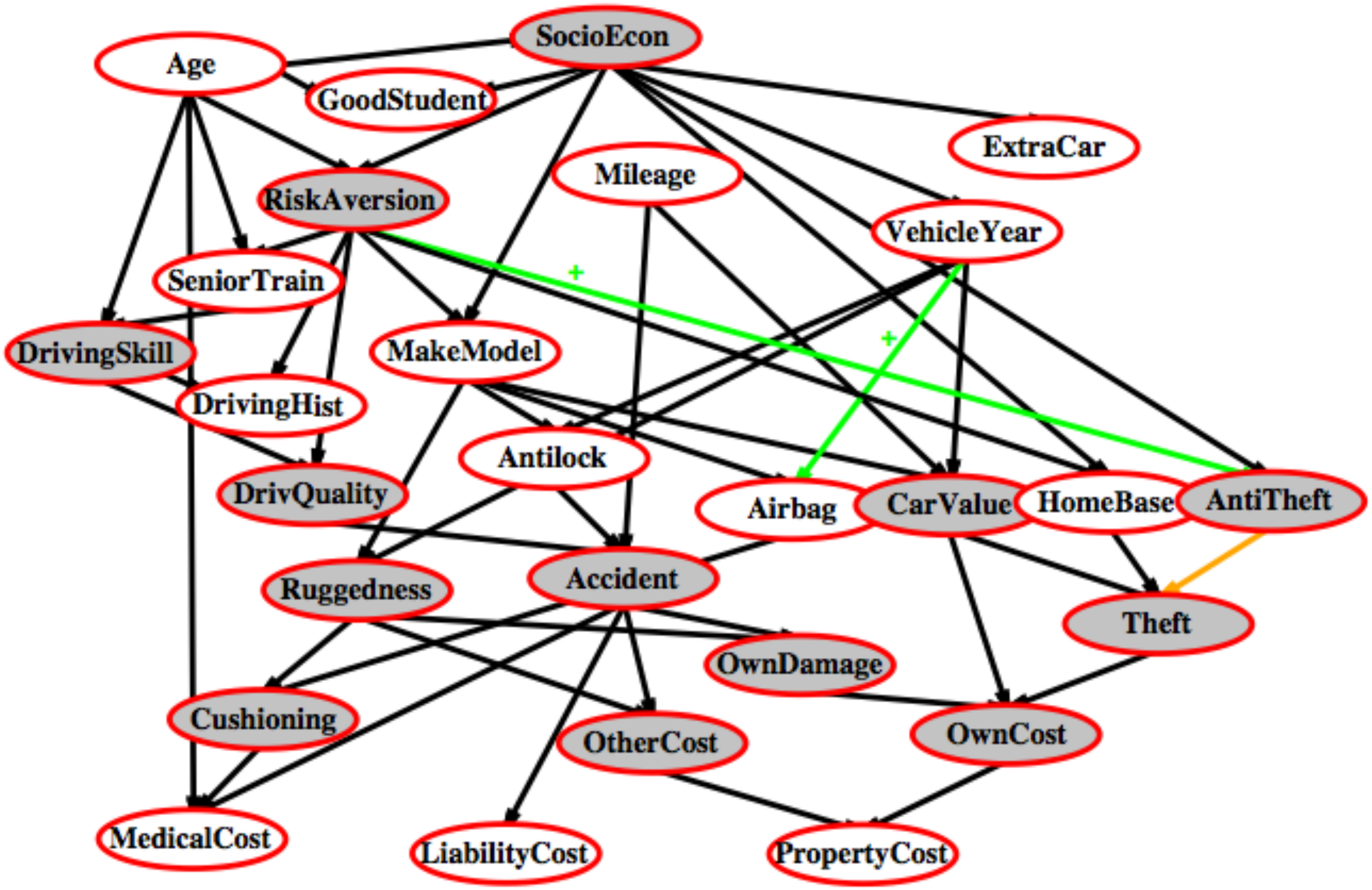
Label the arcs + or –



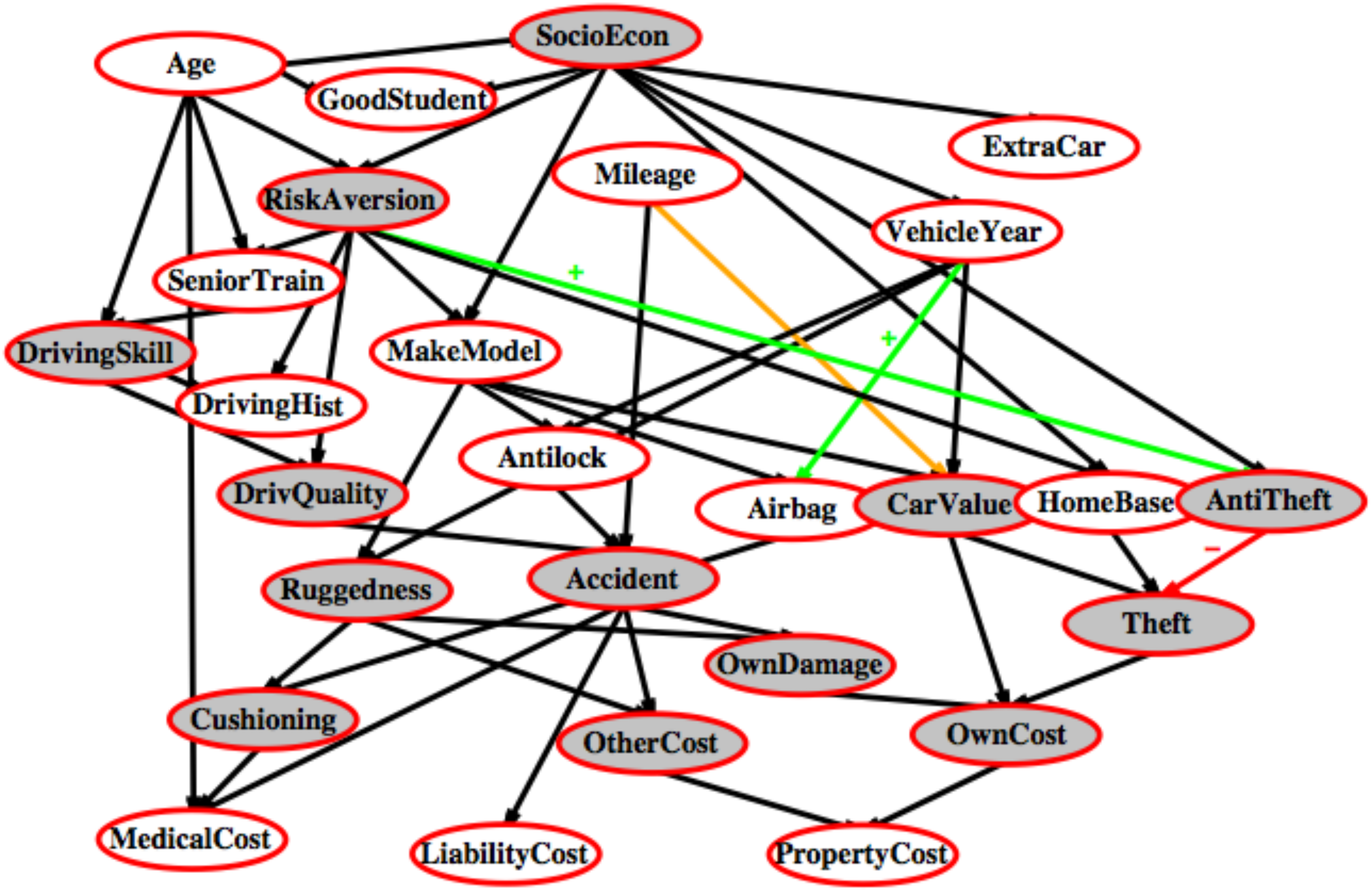
Label the arcs + or –



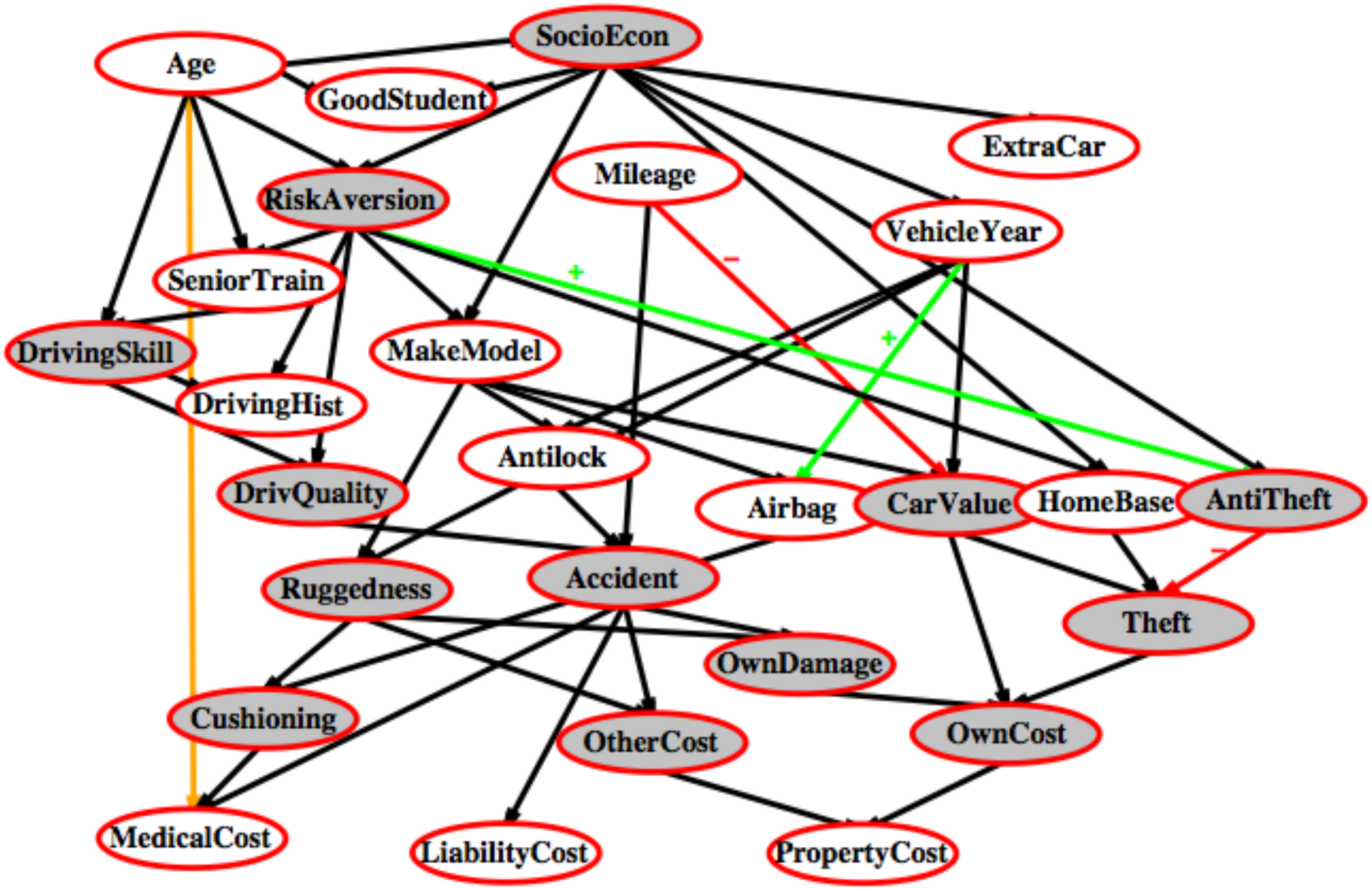
Label the arcs + or –



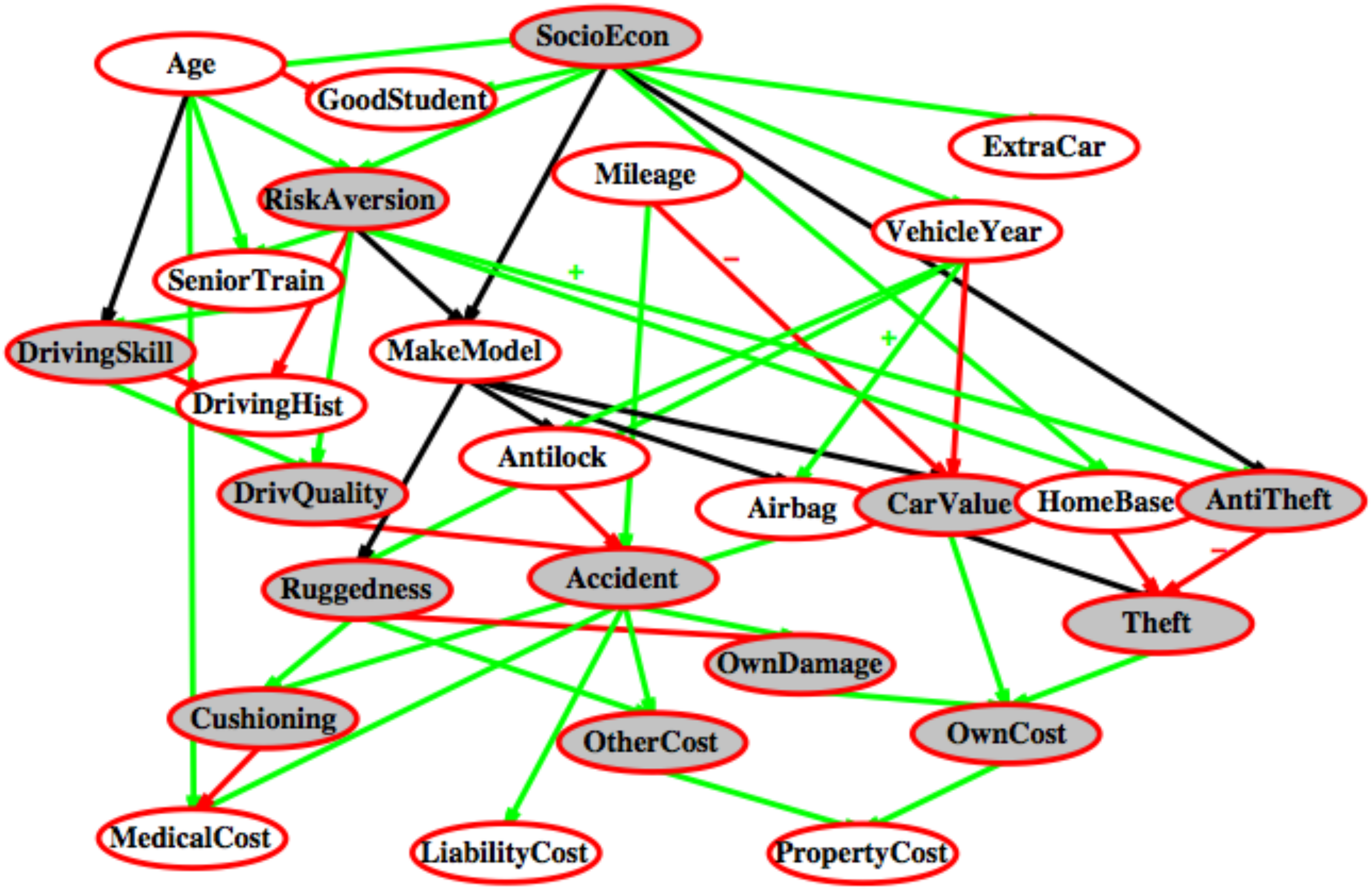
Label the arcs + or –



Label the arcs + or –



Label the arcs + or –



Preference structure: Deterministic

X_1 and X_2 preferentially independent of X_3 iff
preference between $\langle x_1, x_2, x_3\rangle$ and $\langle x'_1, x'_2, x_3\rangle$
does not depend on x_3

E.g., $\langle Noise, Cost, Safety\rangle$:
⟨20,000 suffer, \$4.6 billion, 0.06 deaths/mpm⟩ vs.
⟨70,000 suffer, \$4.2 billion, 0.06 deaths/mpm⟩

Theorem (Leontief, 1947): if every pair of attributes is P.I. of its complement, then every subset of attributes is P.I of its complement: mutual P.I..

Theorem (Debreu, 1960): mutual P.I. $\Rightarrow$ $\exists$ additive value function:

$$V(S) = \Sigma_i V_i(X_i(S))$$

Hence assess n single-attribute functions; often a good approximation

mpm - mortality prediction model

Preference structure: Stochastic

Need to consider preferences over lotteries:

$\mathbf{X}$ is utility-independent of $\mathbf{Y}$ iff

preferences over lotteries in $\mathbf{X}$ do not depend on $\mathbf{y}$

Mutual U.I.: each subset is U.I of its complement

$\Rightarrow$ $\exists$ multiplicative utility function:

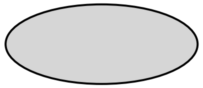
$$U = k_1U_1 + k_2U_2 + k_3U_3$$
$$+ k_1k_2U_1U_2 + k_2k_3U_2U_3 + k_3k_1U_3U_1$$
$$+ k_1k_2k_3U_1U_2U_3$$

Routine procedures and software packages for generating preference tests to identify various canonical families of utility functions

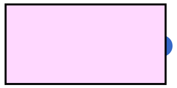
Decision Networks

- Extend BNs to handle actions and utilities
- Also called *influence diagrams*
- Use BN inference methods to solve
- Perform *Value of Information* calculations

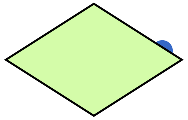
Decision Networks cont.



- Chance nodes: random variables, as in BNs

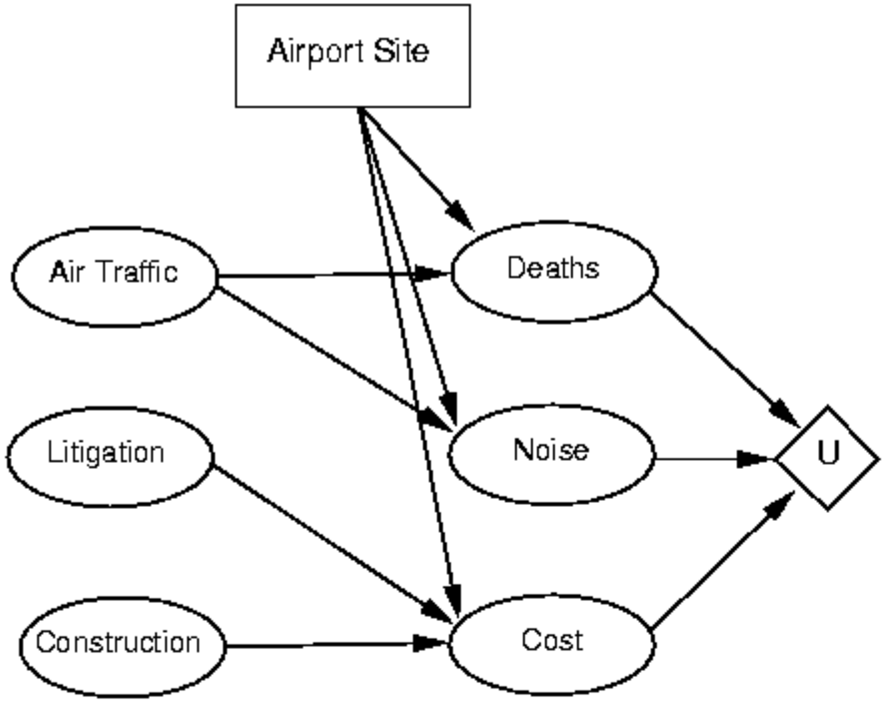


- Decision nodes: actions that decision maker can take



- Utility/value nodes: the utility of the outcome state.

R&N example

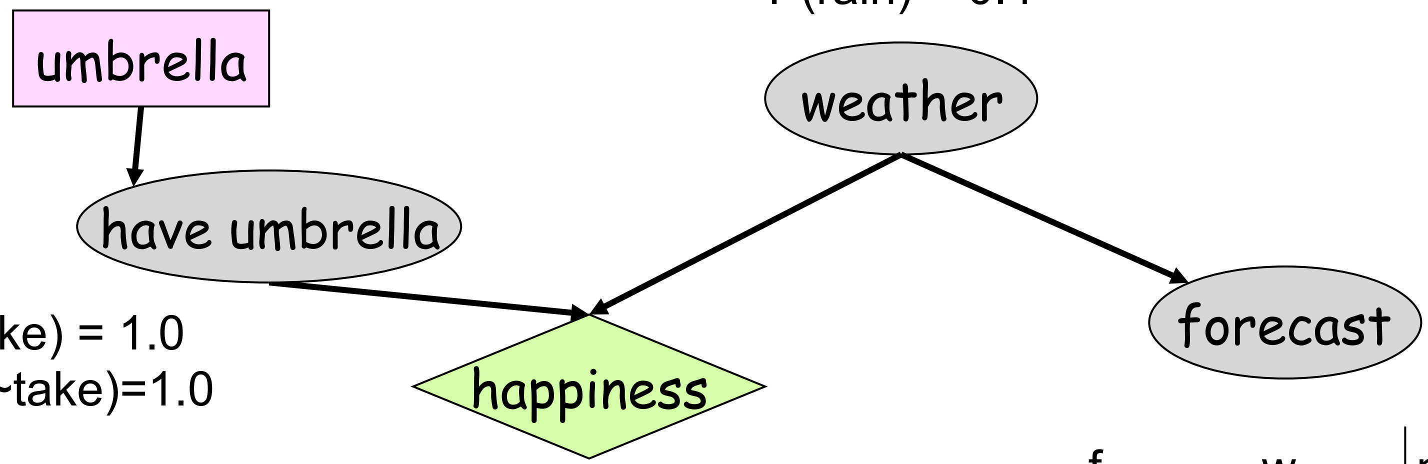


Airport Site
Air Traffic
Deaths
Litigation
Noise
U
Construction
Cost

Umbrella Network

take/don't take

P(rain) = 0.4



P(have|take) = 1.0
P(~have|~take)=1.0

U(have,rain) = -25
U(have,~rain) = 0
U(~have, rain) = -100
U(~have, ~rain) = 100

f	w	p(f|w)
sunny	rain	0.3
rainy	rain	0.7
sunny	no rain	0.8
rainy	no rain	0.2

Evaluating Decision Networks

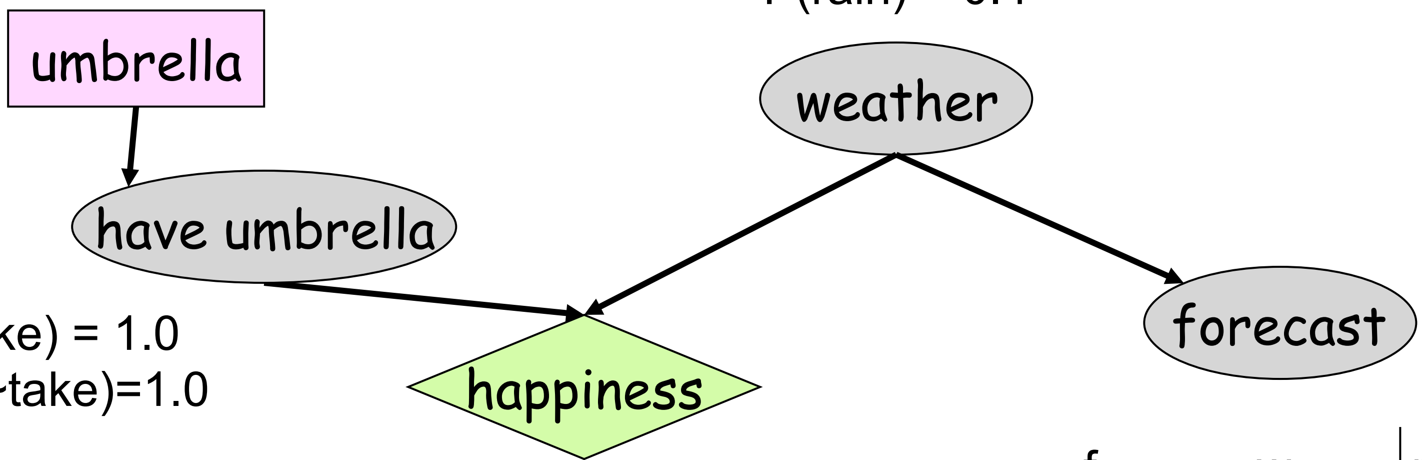
- Set the evidence variables for current state
- For each possible value of the decision node:
  - Set decision node to that value
  - Calculate the posterior probability of the parent nodes of the utility node, using BN inference
  - Calculate the resulting utility for action
- Return the action with the highest utility

Decision Making: Umbrella Network

Should I take my umbrella??

take/don't take

P(rain) = 0.4



P(have|take) = 1.0
P(~have|~take)=1.0

U(have,rain) = -25
U(have,~rain) = 0
U(~have, rain) = -100
U(~have, ~rain) = 100

f	w	p(f|w)
sunny	rain	0.3
rainy	rain	0.7
sunny	no rain	0.8
rainy	no rain	0.2

Value of information

Idea: compute value of acquiring each possible piece of evidence
Can be done **directly from decision network**

Example: buying oil drilling rights
Two blocks A and B, exactly one has oil, worth k
Prior probabilities 0.5 each, mutually exclusive
Current price of each block is $k/2$
"Consultant" offers accurate survey of A. Fair price?

Solution: compute expected value of information
= expected value of best action given the information
minus expected value of best action without information
Survey may say "oil in A" or "no oil in A", **prob. 0.5 each** (given!)
= [$0.5 \times$ value of "buy A" given "oil in A"
+ $0.5 \times$ value of "buy B" given "no oil in A"]
– 0

$$= (0.5 \times k/2) + (0.5 \times k/2) - 0 = k/2$$

General formula

Current evidence E, current best action α
Possible action outcomes S_i, potential new evidence E_j

$$EU(\alpha|E) = \max_a \Sigma_i \; U(S_i) \; P(S_i|E,a)$$

Suppose we knew $E_j = e_{jk}$, then we would choose $\alpha_{e_{jk}}$ s.t.

$$EU(\alpha_{e_{jk}}|E, E_j = e_{jk}) = \max_a \Sigma_i \; U(S_i) \; P(S_i|E, a, E_j = e_{jk})$$

E_j is a random variable whose value is *currently* unknown
$\Rightarrow$ must compute expected gain over all possible values:

$$VPI_E(E_j) = \left(\Sigma_k \; P(E_j = e_{jk}|E) EU(\alpha_{e_{jk}}|E, E_j = e_{jk})\right) - EU(\alpha|E)$$

(VPI = value of perfect information)

Properties of VPI

Nonnegative—in **expectation**, not **post hoc**

$$\forall j, E \;\; VPI_E(E_j) \geq 0$$

Nonadditive—consider, e.g., obtaining E_j twice

$$VPI_E(E_j, E_k) \neq VPI_E(E_j) + VPI_E(E_k)$$

Order-independent

$$VPI_E(E_j, E_k) = VPI_E(E_j) + VPI_{E,E_j}(E_k) = VPI_E(E_k) + VPI_{E,E_k}(E_j)$$

Note: when more than one piece of evidence can be gathered, maximizing VPI for each to select one is not always optimal

$\Rightarrow$ evidence-gathering becomes a **sequential** decision problem

Qualitative behaviors

a) Choice is obvious, information worth little
b) Choice is nonobvious, information worth a lot
c) Choice is nonobvious, information worth little