Causality and Decision Analysis for Risk Analysts

How to (probably) achieve what we want

Course Summary

- For most risk management decisions, all we really want is:

  $\Pr(c \mid a) = \Pr(\textit{consequence} \mid \textit{act})$
- Quantitative risk assessment depends on credible causal models of $\Pr(c \mid a)$
  - Learn $\Pr(c \mid a)$ from data if possible
  - Use knowledge, expertise, Bayesian model-averaging (BMA) if necessary
  - Or, adaptively optimize $\Pr(a \mid \textit{information})$

Course Summary

- Even rough quantitative models greatly improve decisions (usually)
  - Quantitative causal modeling, although imperfect, typically improves decisions
    - “Improve” = make preferred outcomes more likely
    - Fitting simple quantitative models to one’s *own judgments* can improve decisions!
- Quantitative risk assessment modeling is practical, even for complex and uncertain systems… and is too valuable not to use!

Course Summary

Getting there:

- Decision analysis (DA) and risk analysis
  - Risk profiles, F-N curves, frequency & severity
  - Expected utility theory, stochastic dominance
- Causal modeling ("knowledge representation")
  - Fault trees, decision trees, influence diagrams
  - Bayesian networks, conditional independence
  - Simulation models of dynamic systems
- Optimization modeling
  - Reinforcement learning, on-line algorithms, simulation-optimization, decision optimization

Introduction

Types of risky decisions

- One-shot
  - Buy, sell, accept/reject; optimize allocation
  - Choose “best” of alternative acts or prospects
- Multiple simultaneous choices
  - Subset selection/portfolio optimization
  - Budget allocation/diversification
- Repeated choices
  - Learn from experience, minimize regret
- Sequential (plan ahead)
  - Optimal stopping; optimal attack; R&D, etc.

Q: Why use decision analysis?

A: Real decisions are often dominated

Artificial example:

A = gain $240 for sure *vs.* B = 25% to win $1,000 (else 0)

C = lose $750 for sure *vs.* D = 75% to lose $1,000 (else 0)

- You must choose *one* prospect from the pair (A, B) *and* one prospect from (C, D).
- *What do you choose?* (A, C), (A, D), (B, C), or (B, D)?
- Many prefer (A, D) to (B, C)
- (A, D) gives 25% chance of $240, 75% chance of -$760
- (B, C) gives 25% chance of $250, 75% chance of -$750
- So, (B, C) dominates (A, D). Everyone *should* prefer (B, C)!

Q: Why use decision analysis?

- Get more of what we want (“gains”) and less of what we don’t want (“losses”) from available opportunities.
  - Avoid making dominated choices
- Identify opportunities to do better than intuitive decision-making
  - Overcome artificial segregation of problems; enable “portfolio thinking”
  - Opportunites to do better are hard to spot intuitively (by definition)

Decision Analysis Essentials

1. DECISION = choice among risk profiles (“acts”) (consequences and probabilities), one per act.
2. CONSEQUENCES: Numerical values on one or more outcome scales (“attributes”)
3. UNCERTAINTIES: Consequence probabilities
   - Often from published models
   - Models = influence diagrams/BNs/simulations
4. The *best* decision maximizes expected utility
   - Utilities reflect *risk attitudes* & *preferences* for consequences
   - Axiomatic justification. Prescriptive, not descriptive.
5. LEARNING = *conditioning* on data (via Bayes’ Rule and model-based likelihood functions.)

A “simple” example of DA: Buying a used car

One-shot decision: Buying used car

- Ingredients
  - Acts, states, consequences
  - Preferences, values, utilities for consequences
  - Probabilities
- *Optimal act* maximizes expected utility

	Buy	*Don't*
Good 0.8	*1*	*0.2*
Bad 0.2	*0*	*0.2*

One-shot decision: Buying used car

- Ingredients
  - Acts, states, consequences
  - Preferences, values, utilities for consequences
  - Probabilities
- *Optimal act* maximizes expected utility
- Q: What are the acts, states, utilities?

	Buy	*Don't*
Good 0.8	*1*	*0.2*
Bad 0.2	*0*	*0.2*

One-shot decision: Buying used car

- Ingredients
  - Acts, states, consequences
  - Preferences, values, utilities for consequences
  - Probabilities
- *Optimal act* maximizes expected utility
- Q: What are the acts, states, utilities?

	Buy	*Don't*
Good *0.8*	*1*	*0.2*
Bad *0.2*	*0*	*0.2*

One-shot decision: Buying used car

- Ingredients
  - Acts, states, consequences
  - Preferences, values, utilities for consequences
  - Probabilities
- *Optimal act* maximizes expected utility
- Q: What are the acts, states, utilities?

	Buy	*Don't*
Good 0.8	*1*	*0.2*
Bad 0.2	*0*	*0.2*

One-shot decision: Buying used car

- Ingredients
  - Acts, states, consequences
  - Preferences, values, utilities for consequences
  - Probabilities
- *Optimal act* maximizes expected utility
- Q: What are the acts, states, utilities?

	Buy	*Don't*
Good 0.8	*1*	*0.2*
Bad 0.2	*0*	*0.2*

One-shot decision: Buying used car

- Ingredients
  - Acts, states, consequences
  - Preferences, values, utilities for consequences
  - Probabilities
- *Optimal act* maximizes expected utility
- Q: Find *expected utility* of each act.

	Buy	*Don't*
Good 0.8	*1*	*0.2*
Bad 0.2	*0*	*0.2*

One-shot decision: Buying used car

- Ingredients
  - Acts, states, consequences
  - Preferences, values, utilities for consequences
  - Probabilities
- *Optimal act* maximizes expected utility
- Q: Find *expected utility* of each act.

	Buy	*Don't*
Good *0.8*	*1*	*0.2*
Bad *0.2*	*0*	*0.2*

One-shot decision: Buying used car

- Ingredients
  - Acts, states, consequences
  - Preferences, values, utilities for consequences
  - Probabilities
- *Optimal act* maximizes expected utility
- Q: Find *expected utility* of each act.

	Buy	*Don't*
Good *0.8*	*1*	*0.2*
Bad *0.2*	*0*	*0.2*
EU	0.8	0.2

One-shot decision: Buying used car

- Ingredients
  - Acts, states, consequences
  - Preferences, values, utilities for consequences
  - Probabilities
- *Optimal act* maximizes expected utility
- Q: Find *expected utility* of each act.

	Buy	*Don't*
Good *0.8*	*1*	*0.2*
Bad *0.2*	*0*	*0.2*
EU	**0.8**	0.2

One-shot decision: Buying used car

- Ingredients
  - Acts, states, consequences
  - Preferences, values, utilities for consequences
  - Probabilities
- *Optimal act* maximizes expected utility
- Q: Find *expected utility* of each act.

	Buy	*Don't*
Good *0.8*	*1*	*0.2*
Bad *0.2*	*0*	*0.2*
EU	0.8	**0.2**

One-shot decision: Buying used car

- Ingredients
  - Acts, states, consequences
  - Preferences, values, utilities for consequences
  - Probabilities
- *Optimal act* maximizes expected utility
- Q: Find *expected utility* of each act.

	Buy	*Don't*
Good 0.8	*1*	*0.2*
Bad 0.2	*0*	*0.2*
EU	**0.8**	**0.2**

One-shot decisions: Summary

- *Solution method:* Choose act with greatest EU.
  - *Calculation:* EU(*a*) = sum over *s* of Pr(*s*)*u*[*c*(*a*, *s*)].
    - *c*(*a*, *s*) is *consequence* of taking act *a* if state is *s*.
    - *u*[*c*(*a*, *s*)] = *utility* of consequence *c*(*a*, *s*) (between 0 and 1)
    - Pr(*s*) = *probability* of state *s*.
    - EU(*a*) = *expected utility* of act *a* = Σ_sPr(*s*)*u*[*c*(*a*, *s*)]
  - *Why it works:* Axioms relate preferences (utilities) for *consequences* to preferences for *acts*.
- *Inputs:*
  - *Preferences* for consequences, *u*(*c*)
  - *Beliefs:* Pr(*s*); *consequence model, c*(*a*, *s*)
- *Outputs:* Recommended *a*; Pr(*c* | *a*) (*risk profile*)

Why Maximize EU?

- Q: Why not consider variance of utility, probability of gain, probability of obtaining a target level, minimizing maximum loss, etc.?
- A: EU is the *unique* decision principle satisfying axioms of:
  - *Reduction:* Only consequence probabilities matter
  - *Weak Order* (transitive, reflexive, complete) for the “is not preferred to” relation
  - *Independence:* aRb iff (a p c)R(b p c), $0 < p < 1$
  - *Continuity:* (a P b) & (b P c) implies (a p c) is indifferent to b for some p

Why Else Maximize EU?

- EU is the *unique* decision principle satisfying *dynamic consistency* axioms…
  - … and it is the only procedure satisfying various other (coherent updating, substitution-of-equivalents, etc.) axioms.
  - It is also *one of* the few procedures that is guaranteed to avoid dominated decisions, sensitivity to logically irrelevant information, and various types of incoherence.
- Need EU (or something very similar) to *evaluate* alternative risk profiles.

Risk profiles: Ideas and critique

Developing risk profiles

- Two inputs:
  - Pr(s) = state probabilities
  - c(a, s) or Pr(c | a, s) (consequence models)
- Output: Pr(c | a)
- Calculation algorithms
  - Influence diagram: $a \rightarrow c \leftarrow s$
  - Formula: $\Pr(c \mid a) = \sum_s \Pr(c \mid a, s)\Pr(s)$
  - Monte Carlo: Sample from Pr(s), Pr(c | a, s), build up Pr(c | a)

Comparing risk profiles

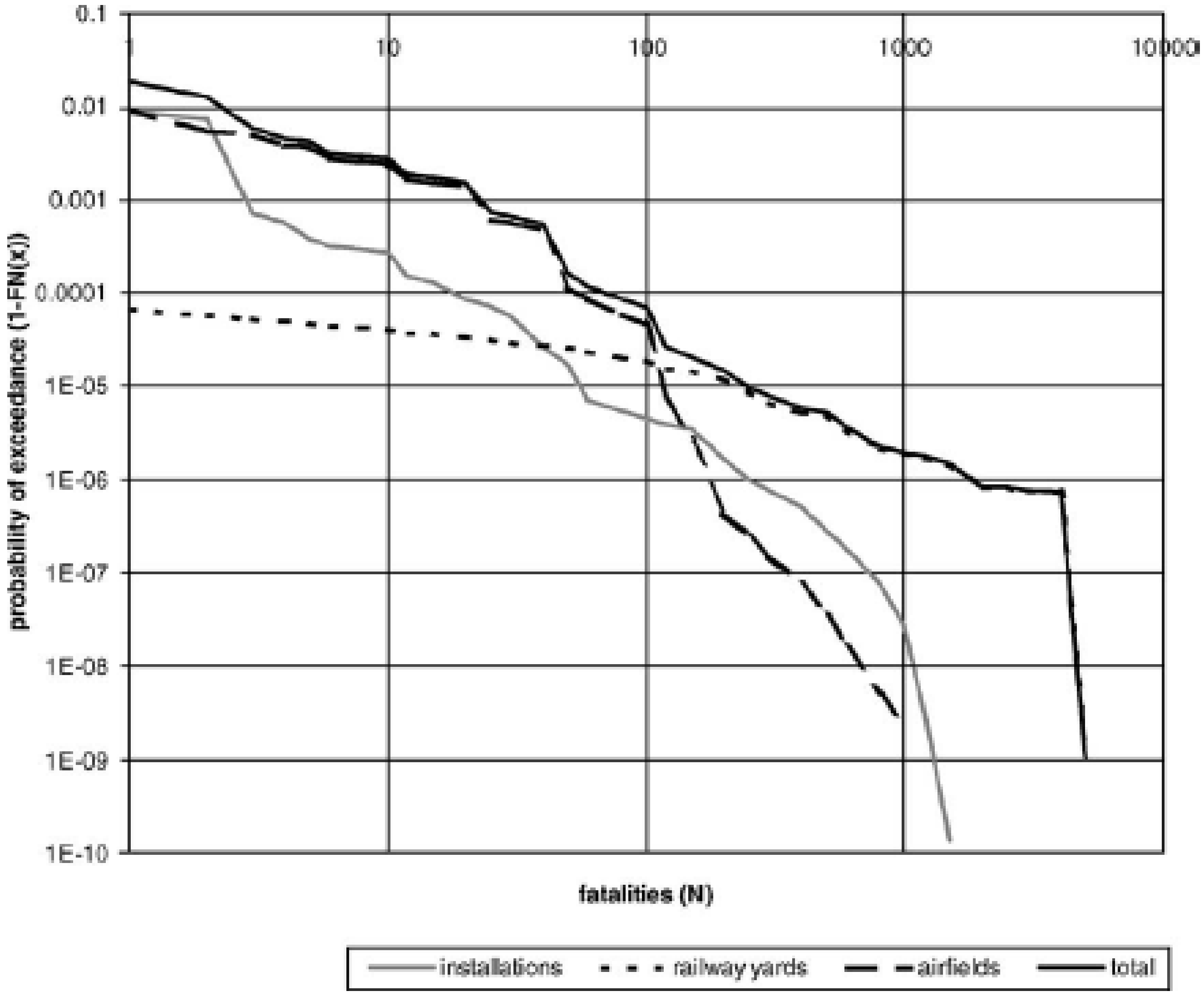


Fig. 5. FN-curve for the risks of various activities in The Netherlands in 1999 (source: RIVM, NL).

Comparing risk profiles

Is this a good way to make decisions?

10 S.N. Jonkman et al. / Journal of Hazardous Materials A99 (2003) 1–30

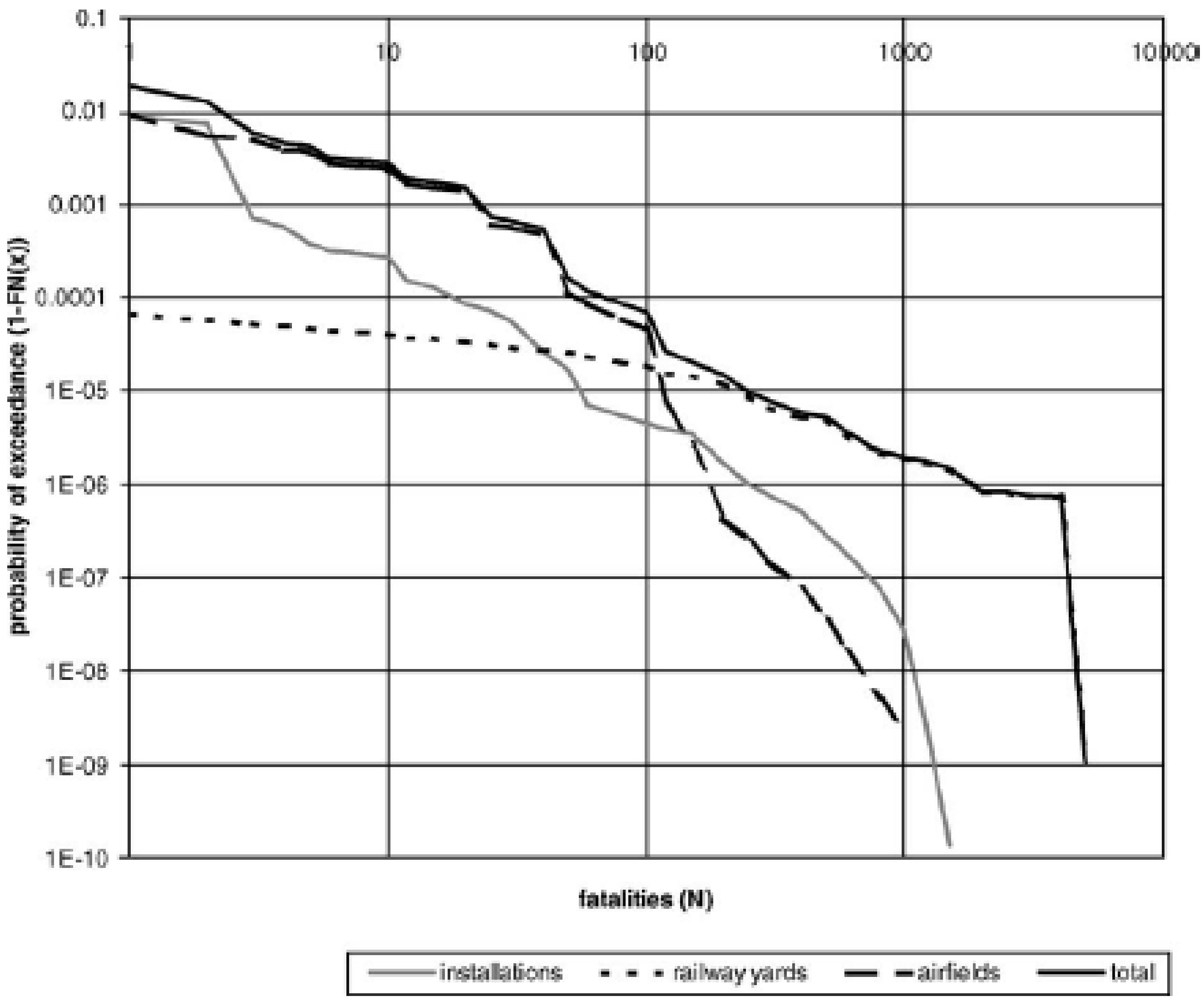
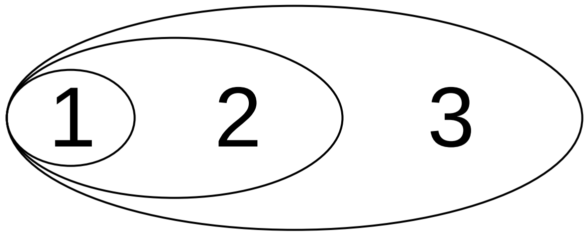
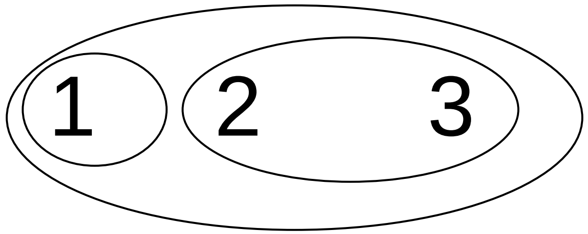


Fig. 5. FN-curve for the risks of various activities in The Netherlands in 1999 (source: RIVM, NL).

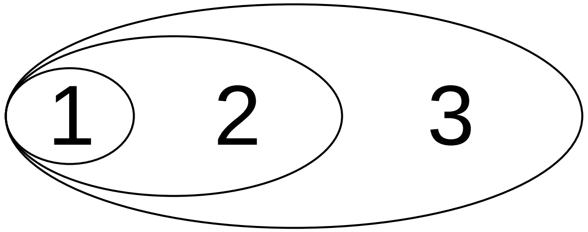
Which is preferable?

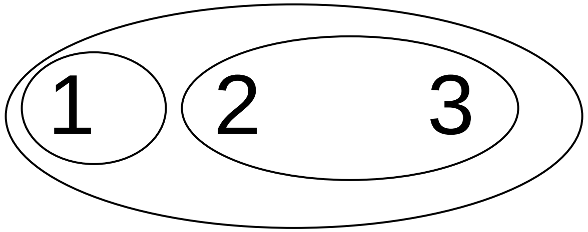
- Accident A 

- Accident B 

  - A kills {1}, {1, 2}, or {1, 2, 3}, Pr = 1/3 each
  - B kills {1}, {2, 3}, or {1, 2, 3 }, Pr = 1/3 each
  - *FN curves are identical,* but individual risks are not: (1, 2/3, 1/3) vs. (2/3, 2/3, 2/3).
  - A preference for risk equity favors B over A.

Which is preferable?

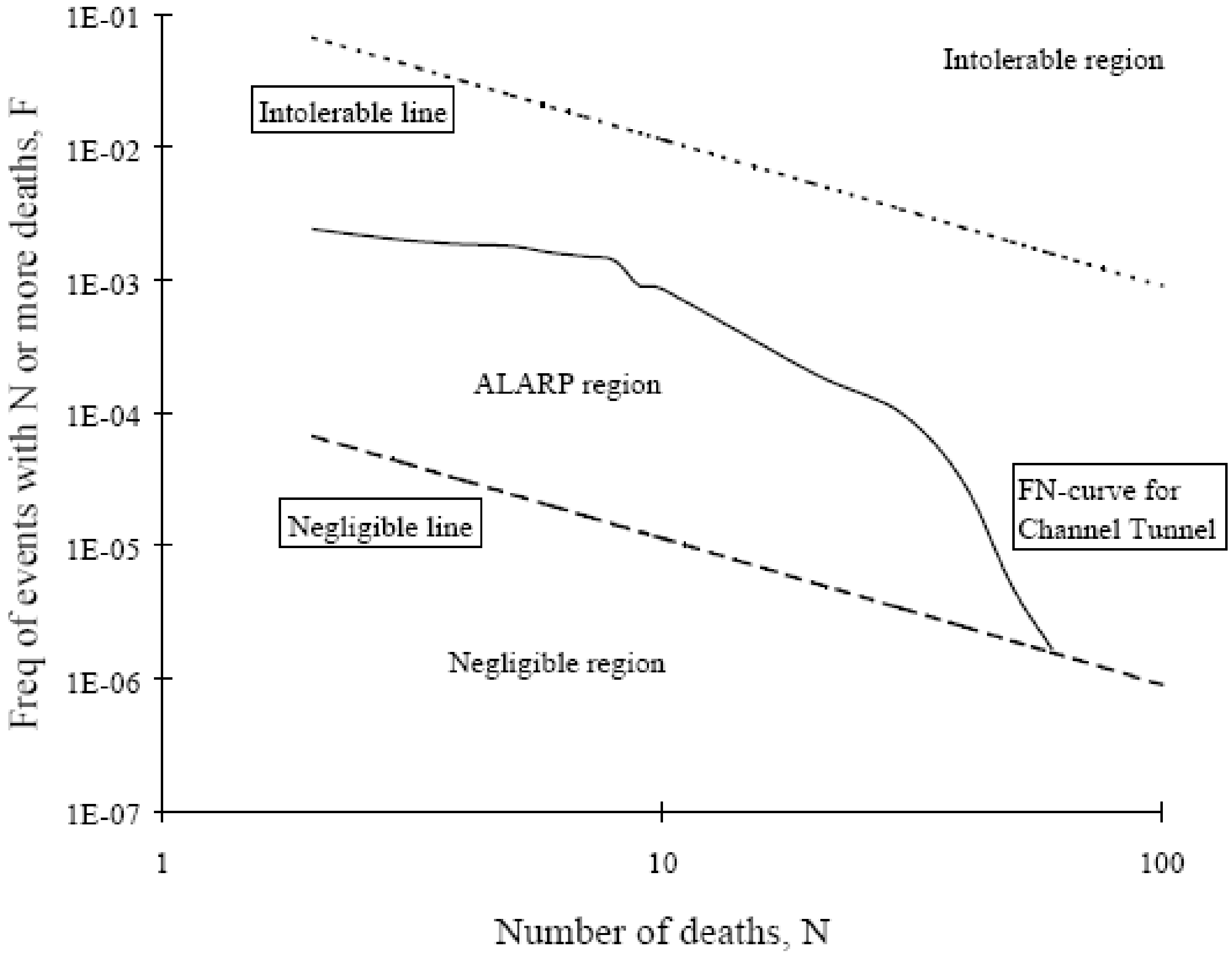
- Accident A 

- Accident B 

- If *either* is preferable, then FN curve does not capture everything needed to make decisions.

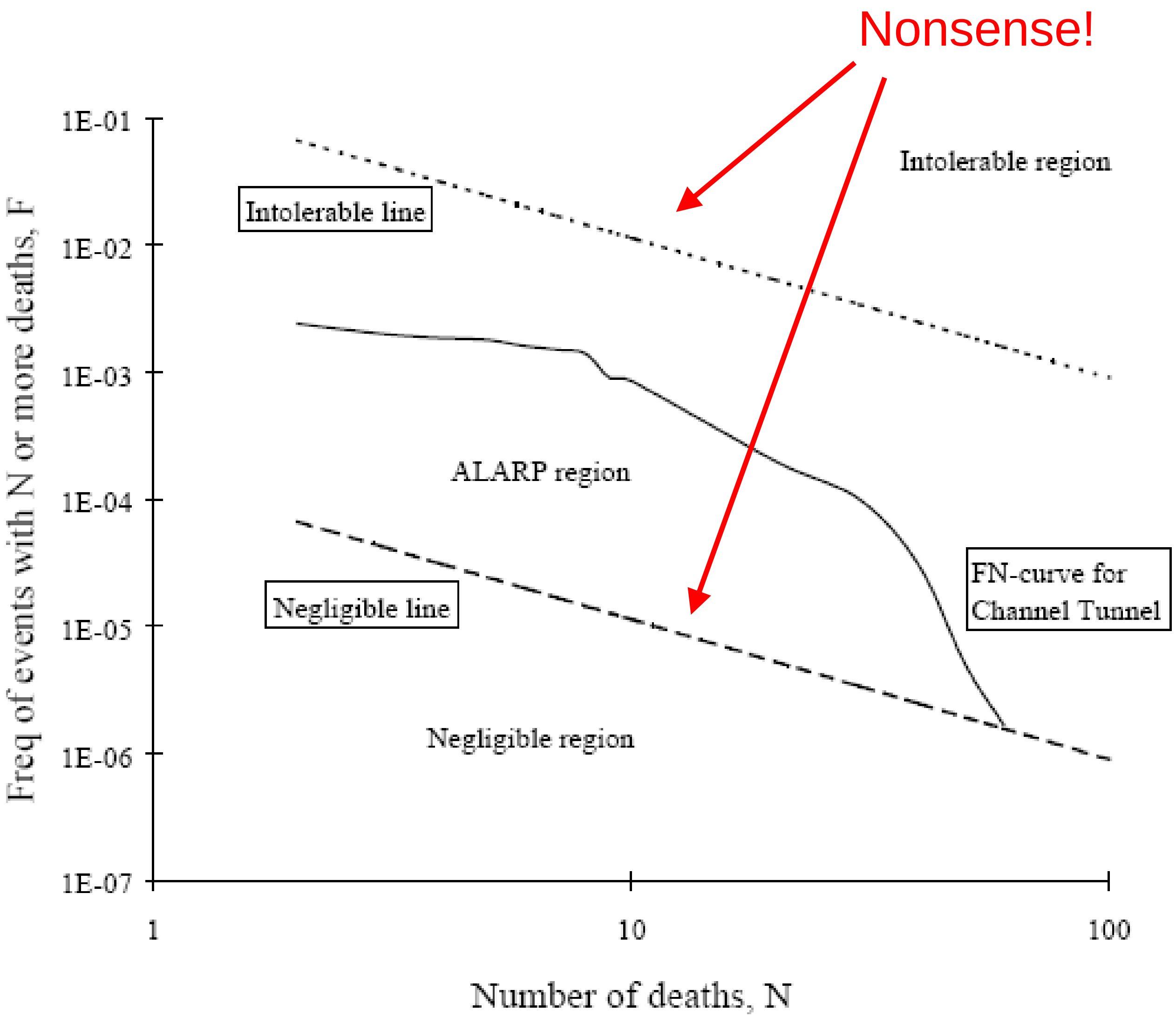
Evaluating risk profiles

http://www.dft.gov.uk/pgr/aviation/safety/thirdpartyrisknearairportsan2989



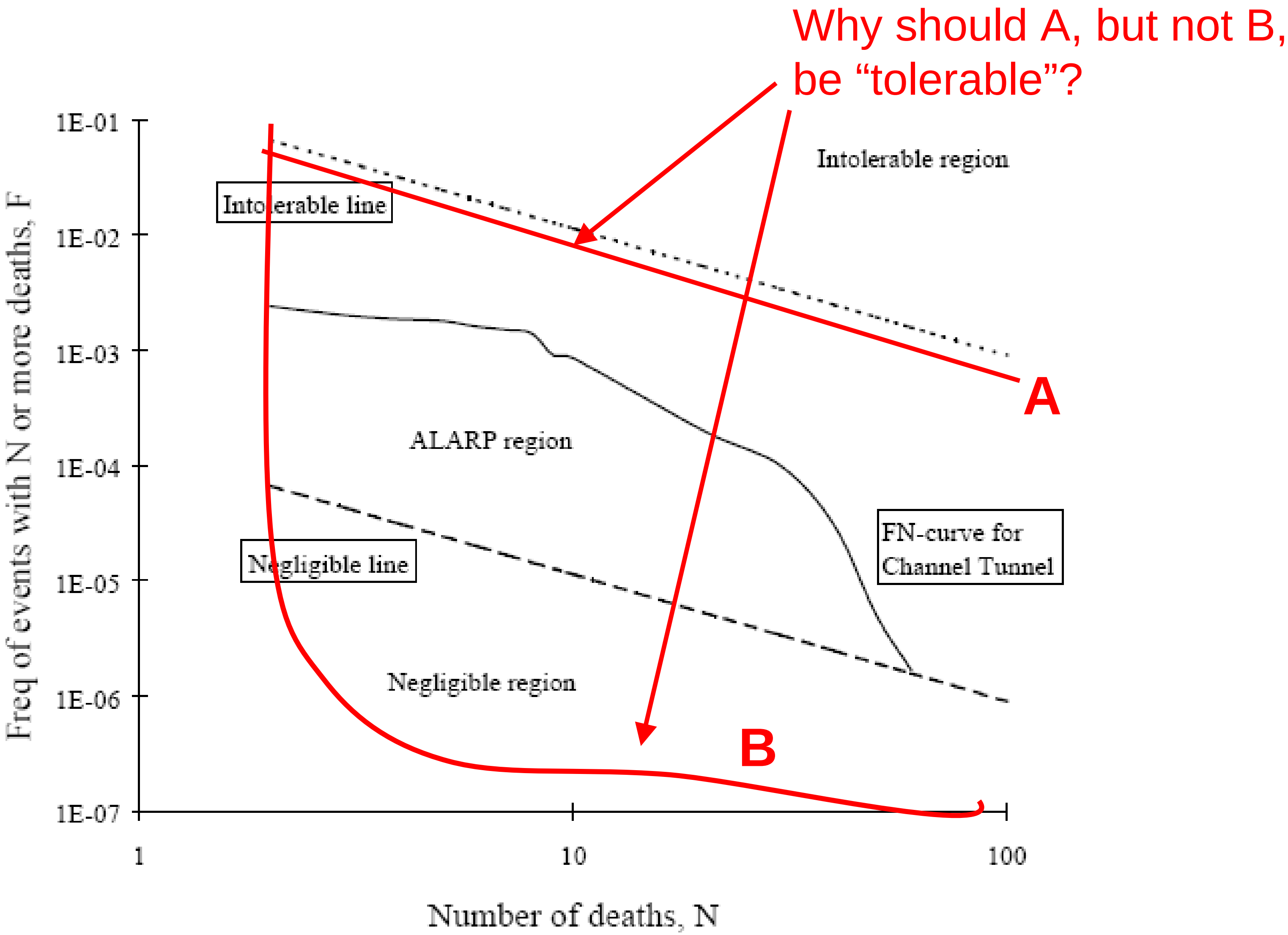
Evaluating risk profiles

http://www.dft.gov.uk/pgr/aviation/safety/thirdpartyrisknearairportsan2989



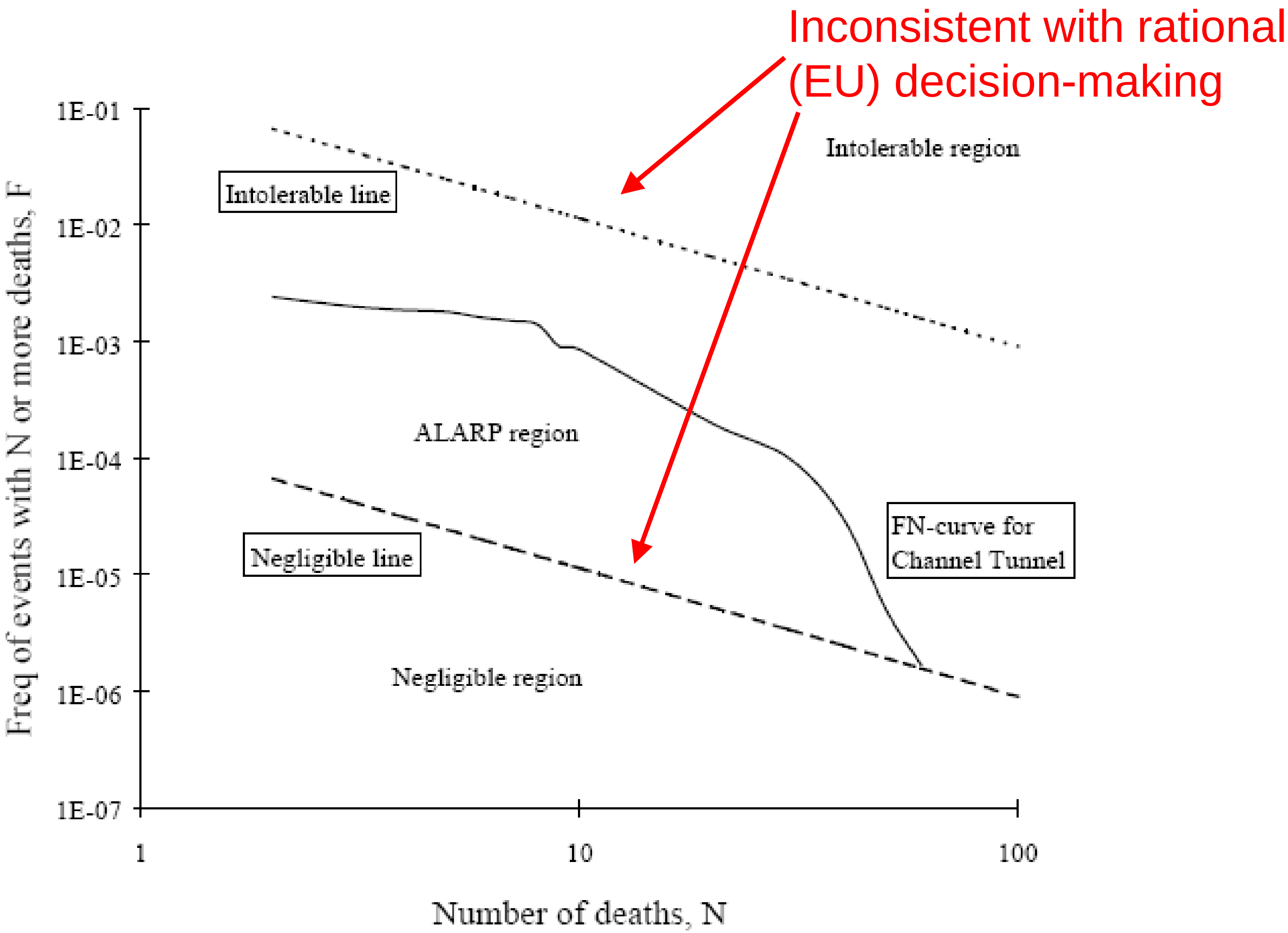
Evaluating risk profiles

http://www.dft.gov.uk/pgr/aviation/safety/thirdpartyrisknearairportsan2989



Evaluating risk profiles

http://www.dft.gov.uk/pgr/aviation/safety/thirdpartyrisknearairportsan2989



Lower frequency of adverse consequence does not necessarily imply lower risk (!)

	A. Exponential time to failure μ *= 66.67 years*
$1/\mu$ = average annual frequency	0.015
Probability of accident in 1 yr.	0.015

Lower frequency of adverse consequence does not necessarily imply lower risk (!)

	A. Exponential time to failure *μ = 66.67 years*	*B. Uniform time, U[0, 100]* *μ = 50 years*
$1/\mu$ = average annual frequency	0.015	?
Probability of accident in 1 yr.	0.015	?

Lower frequency of adverse consequence does not necessarily imply lower risk (!)

	A. Exponential time to failure μ *= 66.67 years*	*B. Uniform time, U[0, 100]* μ *= 50 years*
$1/\mu$ = average annual frequency	0.015	0.02
Probability of accident in 1 yr.	0.015	0.01

Which has lower frequency? (Which is preferable?)

A: Time to failure ~ U[2, 10], mean = 6 yr.

B: Time to failure ~ U[4, 6], mean = 5 yr.

- *Either may be preferable,* depending on required duration of a mission
  - *A* is preferable if mission is 8 years
  - *B* is preferable if mission is 3 years.
- No definition of “frequency” that ranks one over the other (or ties them) is satisfactory.

Conclusion on F-N curves

- “Frequency” is not well-defined in general
  - No general definition of “frequency” should be used to make risk management decisions
  - *Exception:* Poisson arrivals
- “Severity” omits information needed for risk equity comparisons.
- Something better is needed!

Expected Utility Theory

EU theory avoids such problems

- Evaluating risk profiles by expected utility (EU) avoids all such difficulties
  - At least in theory
- *Only* EU theory avoids them all!
  - Traditional axiomatic rationale for EU:
    - Make “avoiding difficulties” our axioms
    - Prove that EU representation results
    - EU is unique up to choice of origin and scale
- But, decision-maker must specify $u(c)$, $\Pr(s)$
  - May not exist, or may be hard to specify

Practical challenges for EU

- $EU(a) = \Sigma_s Pr(s)u[c(a, s)]$
- Where do beliefs, $Pr(s)$, come from?
  - Suppose they are unknown/uncertain?
  - Suppose states are too complex to describe?
- Where do utilities, $u(c)$ come from?
  - Hypothetical lotteries? You must be kidding!
- Where does consequence model, $c(a, s)$ or $Pr(c \mid a, s)$, or $Pr(c \mid a)$, come from?
  - Risk assessment models!

Practical challenges for EU

Need:

- $\Pr(s)$ = state probabilities
- $u(c)$ = utilities for consequences
- $\Pr(c \mid a, s)$ = consequence model (causal)
  - At least need $\Pr(c \mid a)$

Introduction to Causal Modeling

Reasoning with causality and conditionals is often surprisingly hard

- Consider a deck of cards. Each has a letter on one of its sides and a number on the other.
- *Claim:* “If a card has a vowel on one side, then it has an even number on the other side”

A	**D**	**4**	**7**

Q: Which card(s) must be turned over to determine whether the claim is true?

(“Wason selection task”, 1970)

Reasoning with causality and conditionals is often surprisingly hard

- Consider a deck of cards. Each has a letter on one of its sides and a number on the other.
- *Claim:* “If a card has a vowel on one side, then it has an even number on the other side”

A	**D**	**4**	**7**

Q: Which card(s) must be turned over to determine whether the claim is true?

Usual answers: “A only” or “A and 4”

Reasoning with causality and conditionals is often surprisingly hard

- Consider a deck of cards. Each has a letter on one of its sides and a number on the other.
- *Claim:* "If a card has a vowel on one side, then it has an even number on the other side"

A	**D**	**4**	**7**

Q: Which card(s) must be turned over to determine whether the claim is true?

Correct answers: "A and 7"

Wason selection task in practice

- http://www.psych.ucsb.edu/research/cep/socex/wason.htm

David planted a lovely garden with flowers of every color. He has not been able to enjoy it, though, because deer from the forest nearby have been nibbling on his plants, killing some of them.

He would like to keep the deer out of his garden. His grandmother said that in the old days, she kept deer away by spraying an herbal tea -- lacana -- in her garden. She said:

"If you spray lacana tea on your flowers, deer will stay out of your yard.“

This sounded dubious. So David convinced some of his neighbors to spray their flowers with lacana tea, to see what would happen. You are interested in seeing whether any of the results of this experiment violate Grandma's rule.

The cards below represent four yards near David's house. Each card represents one yard. One side of the card tells whether or not lacana tea was sprayed on the flowers in a yard, and the other side tells whether or not deer stayed out of that yard.

sprayed with lacana tea	not sprayed with lacana tea	deer stayed away	deer did not stay away

Which of the following cards would you definitely need to turn over to see if what happened in any of these yards violated Grandma's rule:

"If you spray lacana tea on your flowers, deer will stay out of your yard."

Don't turn over any more cards than are absolutely necessary.

sprayed with lacana tea	not sprayed with lacana tea	deer stayed away	deer did not stay away

Which of the following cards would you definitely need to turn over to see if what happened in any of these yards violated Grandma's rule:

"If you spray lacana tea on your flowers, deer will stay out of your yard."

Don't turn over any more cards than are absolutely necessary.

sprayed with lacana tea	not sprayed with lacana tea	deer stayed away	deer did not stay away

Lessons from psychology

- Most people (including scientists) tend to seek *confirming evidence* for causal theories, but disregard or ignore potential d*isconfirming* evidence.
  - “Confirmation bias”
- Effective risk analysis requires drawing sound causal inferences from data… and designing studies to collect needed data.
  - We are not naturally good at doing these things
  - Quantitative modeling can help

Practical challenges for EU

Need:

- Pr(s) = state probabilities
- $u(c)$ = utilities for consequences
- **Pr(c | a, s) = consequence model** (causal)
  - At least need Pr(c | a)

Practical challenges for EU

Need:

- Pr(*s*) = state probabilities
- *u*(*c*) = utilities for consequences
- **Pr(*c* | *a*, *s*) = consequence model** (causal)
  - At least need Pr(*c* | *a*)
  - *Causal and statistical models may differ*
    - Pr(*Death* | *Exposure*) = 0.2 + 0.4**Exposure* does <u>not</u> necessarily imply that *Exposure* increases risk!

Causal *vs.* statistical models

Let

- $c = 1$ if subject dies of heart attack, else 0.
- $a = 1$ if subject takes baby aspirin, else 0
- $s = 1$ if subject has high risk of heart attack
- *If* the causal model for $\Pr(c \mid a, s)$ is:
  - $\Pr(c = 1 \mid a) = E(c \mid a) = 0.2 + 0.6s - 0.2a$
  - Exposure model: $a = s$
- *Then* the statistical model for $\Pr(c \mid a)$ is:
  - $\Pr(c = 1 \mid a) = E(c \mid a) = 0.2 + 0.4a$

Causal *vs.* statistical models

- So, statistical relation between exposure a and effect c may be positive, even if causal relation is negative (protective).
    - c = 1 if subject dies of heart attack, else 0.
    - a = 1 if subject exposed, else 0
    - s = 1 if subject has high risk of heart attack
- Causal model: $E(c \mid a) = 0.2 + 0.6s - 0.2a$
    - One causal model for each value of s
- Statistical model: $E(c \mid a) = 0.2 + 0.4a$
    - "Structural" *vs.* "reduced" model

Causal modeling methods

- *Old* (1920s-1970s): Path analysis, structural equations models (SEMs)
  - Correlation-based, good for observational data
- *New* (1980s-present): Causal graphs, conditional independence tests
  - Information-based
- Intervention-based (experimental)

Structural Equation Models (SEMs)

- Explain left-hand variables by right-side variables.
  - *Interpretation:* Changing right-side variables causes changes in left-side variables
    - $E(c \mid a) = 0.2 + 0.6s - 0.2a$
  - Causal partial-ordering (Simon)
    - Some subsets of variables determine themselves *and* other variables.
  - SEMs are well developed in econometrics.
    - LISREL learns SEMs from data on partial correlations and multivariate linearity assumptions.

Conditional Independence Tests

- How to use data to identify a causal model?
  - *Data:* (*Exposure*, *Lifestyle*, *Response*) observations for many individuals (*Example:* ETS and lung cancer)
- Alternative hypotheses/models:

Model 1: *Exposure* $\rightarrow$ *Response* $\leftarrow$ *Lifestyle*
Model 2: *Exposure* $\leftarrow$ *Lifestyle* $\rightarrow$ *Response*
Model 3: *Lifestyle* $\rightarrow$ *Exposure* $\rightarrow$ *Response*

- These different hypothesized models imply different *conditional independence* (CI) relations.
  - Model 3: (*Response* CI *Lifestyle* | *Exposure*)
  - These implications are *testable* by statistics programs

X is a potential direct cause of Y if

- X is *informative* about Y
  - $I(X ; Y) > 0$; $\Pr(Y \mid X) \neq \Pr(Y)$
- Information that X provides about Y cannot be fully *explained away* by any other variables
  - No W for which $I(X ; Y \mid W) = 0$; $\Pr(Y \mid X, Z) \neq \Pr(Y \mid Z)$
  - Not $X \leftarrow W \rightarrow Y$ or $W \rightarrow X \rightarrow Y$ or $W \rightarrow Y \rightarrow X$
- Composition, ordering, conditional independence implications are all satisfied
- *Past* of X is informative about *future* of Y, given past of Y ("Granger-Sims" causality)

These are all information (vs. intervention) criteria

Statistical tests for potential causality

Different approaches:

- Time series:
  - Granger causality
  - Intervention analysis
  - Change point analysis
  - http://support.sas.com/rnd/app/examples/ets/granger/
- Systems of equations
  - Simon causal ordering
- Correlations: SEMs/path analysis
- Conditional independence and causal graphs
  - Classification trees

Classification Trees

- A powerful, popular method for data mining
- Background/software on classification trees:

  http://www.statsoftinc.com/textbook/stcart.html;

  http://www.stat.wisc.edu/~loh/quest.html
- *Idea:* "Always ask the most informative question" for reducing uncertainty (conditional entropy) about the dependent variable.
- Partitions a set of cases into groups or clusters (the leaf nodes) with similar conditional probabilities for dependent variable.
- Download applet from: http://www.cs.ubc.ca/labs/lci/CIspace/dTree/

Summary on Causal Modeling

- Potential causes can be identified from sufficiently large, diverse data sets
  - Efficient causal model-building algorithms are now available
  - New criteria (information-based) are far more effective than traditional criteria
- Building causal models from data is still a very active research area
  - Multiple plausible models: Bayesian model averaging and alternatives
  - Latent variables, Hidden Markov Models, DBNs, etc.

Practical challenges for EU

Need:

- **Pr(s) = state probabilities** $\rightarrow$ How to get?
- $u(c)$ = utilities for consequences
- $\Pr(c \mid a, s)$ = consequence model (causal)
  - At least need $\Pr(c \mid a)$

Obtaining Pr(*s*)

- Expert elicitation
  - Elicitation, calibration, combination
- Calculate from lower-level data
  - “Recursive decomposition”
  - Fault trees
  - Decision trees, event trees, game trees
  - Influence diagrams, Bayesian networks
- Simulate from a model

Obtaining Pr(*s*)

- Option 1: Elicit from experts
- Problems:
  - Experts are often wrong!
    - Consensus opinions & group-think
    - Premature closure, poor use of information
    - Sensitivity to irrelevant information
    - Confirmation bias
  - Experts may not know
    - …but still have opinions that can be elicited
  - Can easily elicit nonsense.
    - Pr(Ruritania will deploy nuclear power by 2015 if crude oil prices remain elevated through 2010.)

Example: Consensus *vs.* Reality

- System will work if at least one of its three components, A, B, C, works
- Each component has prior probability 0.5 of working.
- Expert 1 inspects A; expert 2 inspects B; expert 3 inspects component C.
- Each expert reports that system has 75% probability of working. (Consensus!)
- *Q: What is probability system will work?*

Example: Consensus *vs.* Reality

- System will work if at least one of its three components, A, B, C, works
- Each component has prior probability 0.5 of working.
- Expert 1 inspects A; expert 2 inspects B; expert 3 inspects component C.
- Each reports that system has 75% probability of working. (Consensus!)
- *A: Probability system will work = 0%*

Obtaining Pr(*s*) (Cont.)

- Option 2: Calculate from lower-level data
  - Divide-and-conquer ("recursive decomposition")
    - Partition possibilities into *scenarios*, *E*
    - Estimate $\Pr(s \mid E)$ and $\Pr(E)$ for each scenario, *E*.
  - Calculate $\Pr(s) = \Sigma_E \Pr(s \mid E)\Pr(E) = \Sigma_E \Pr(s \,\&\, E)$
- Option 3: Simulate $\Pr(s)$ from $\Pr(s \mid E)$, $\Pr(E)$
  - Simulate scenarios (randomly generated)
  - Evaluate $\Pr(s) = \Sigma_E \Pr(s \mid E)\Pr(E)$ via simulation
    - Rare-event simulation methods

Do we really need Pr(*s*)?

- In theory, decision-makers should act as if they had well-defined Pr(*s*) and *u*(*c*).
  - Subjective expected utility (SEU) theory
    - Coherence axioms $\rightarrow$ SEU representation
    - Ramsey (1928); Savage (1950s), Raiffa (1960s)
  - Interpret Pr(*s*) as willingness-to-bet
- In practice, they don't (and perhaps can't).
  - Allais, Ellsberg, Fischhoff, Kahneman &Tversky, Kunreuther, Loewenstein, Thaler,…

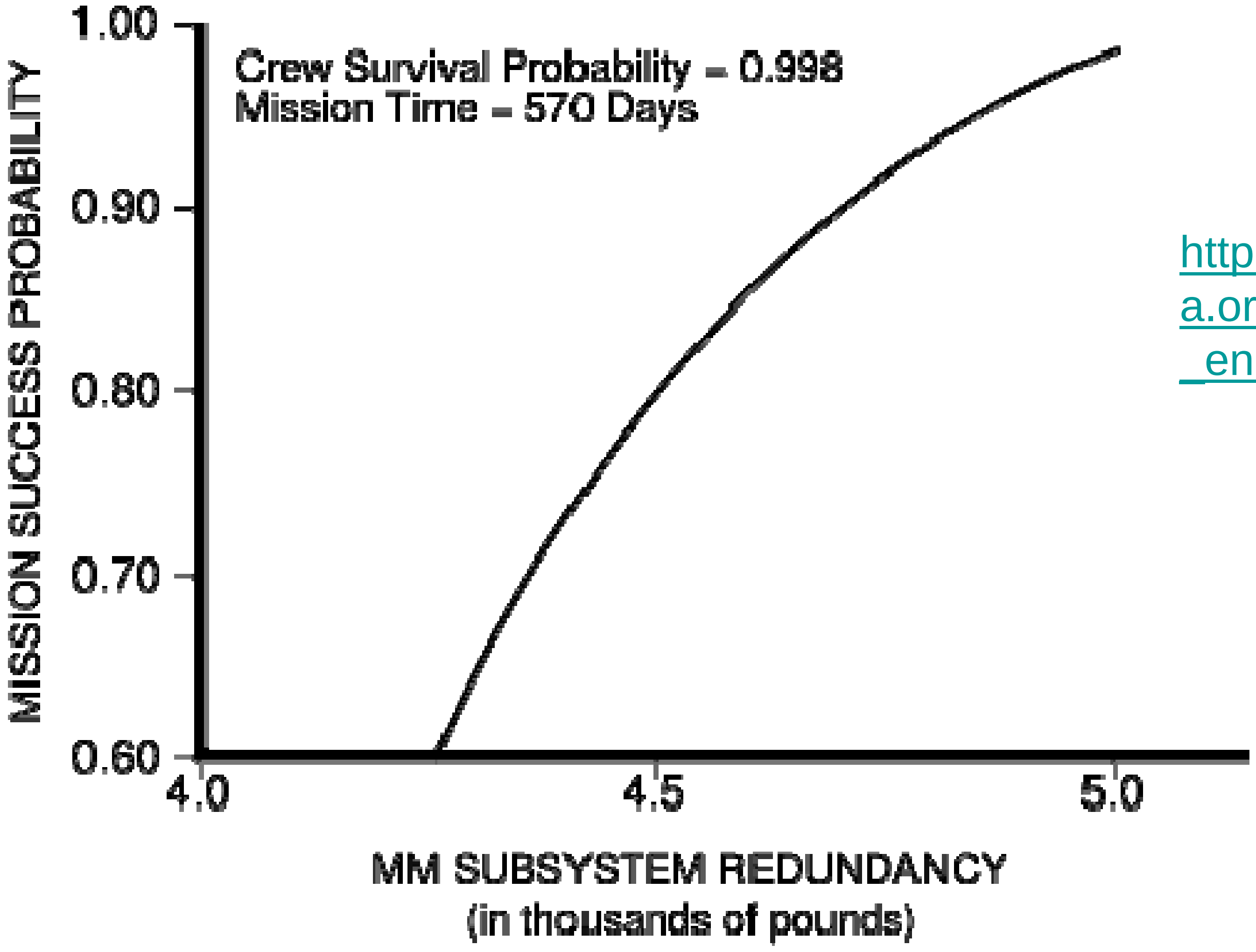
Doing without Pr(*s*)

- Use conservative values for uncertain Pr(*s*)
    - Ambiguity aversion for uncertain probabilities (Ellsberg) $\rightarrow$ "multiple priors" and minimax EU (Gilboa-Schmeidler)
        - www.tark.org/proceedings/tark_mar22_92/p143-gilboa.pdf
- Assess causal Pr(*c* | *a*) directly
    - Learn from experience
    - Predict from statistical models (carefully!)
    - Simulate using causal models
- Learn best act (or decision rule) by trial-and-error: Pr(*a* | *situation*) $\rightarrow$ *reward*
    - Reinforcement learning, simulation-optimization
    - Minimum-regret learning
    - Adaptive stochastic optimization

Example: RM without Pr(*s*)

- *Ski rental problem:* It costs $10/day to rent, $150 to buy.
  - Will ski for unknown number of days.
    - Can't assess useful probabilities, Pr(*s*)
- *What to do?*
  - *On-line algorithm* solution: Rent for 15 days, then buy (if still there).
  - Cost $\leq 2 \times$ "omniscient" minimized cost
    - No other decision rule always does better
    - http://everything2.com/title/Ski%2520Rental%2520Problem

Living without Pr(*s*) by using Pr(*c* | *a*)



http://en.wikipedia.org/wiki/Safety_engineering

Eliminating *s* by summing ("Marginalizing out")

- $\Pr(c \mid a) = \sum_s \Pr(c \mid a, s)\Pr(s) = \sum_s \Pr(c \,\&\, s)$
  - "Law of total probability"
- $E(c \mid a) = \sum_s E(c \mid a, s)\Pr(s)$
  - "Linearity of expected values"
- $EU(c \mid a) = \sum_s EU(c \mid a, s)\Pr(s)$

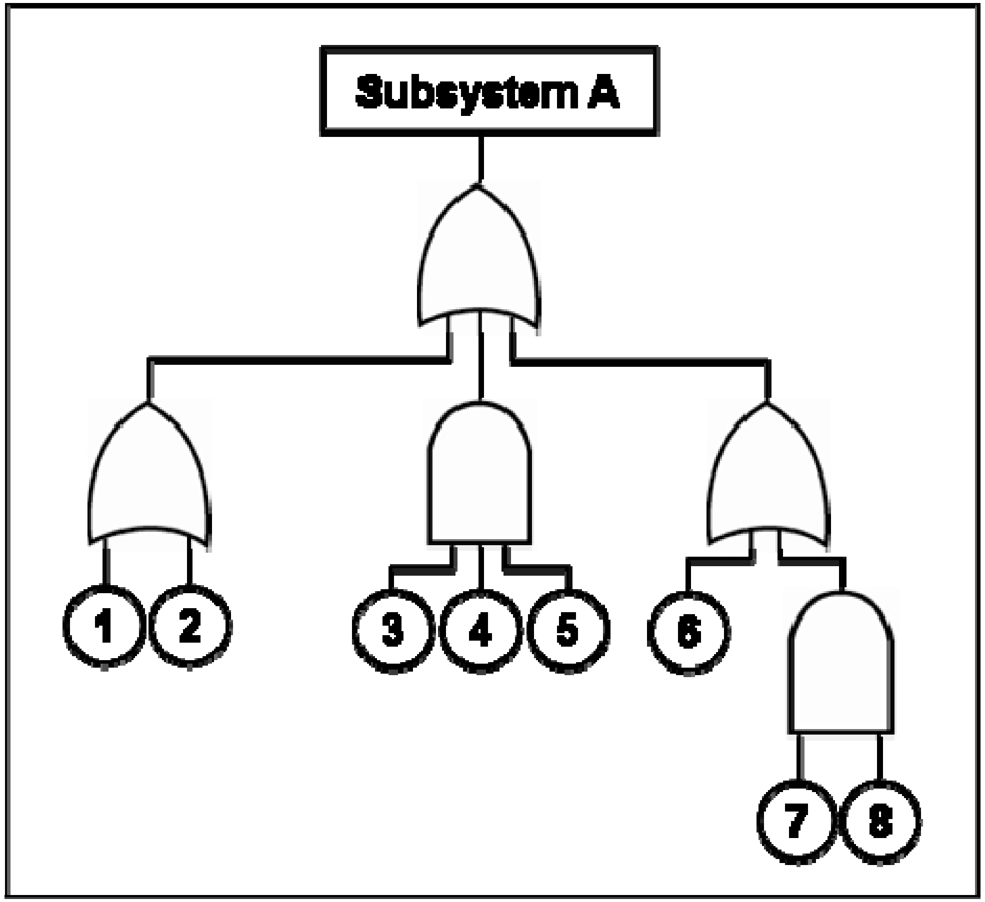
  $= \sum_s u[c(a, s)]\Pr(s)$, for deterministic model
- *So, if Pr(s) is known, we can eliminate it!*

Summary on state probabilities

- A useful means to an end
- Not needed for final decision-making
- Can get rid of them, once they are known
  - Marginalize out (sum over) states to get causal $\Pr(c \mid a)$
- But, still need to get them, before we can get rid of them!

Fault Trees and Decision Trees: Developing Pr(*s*) from sub-models

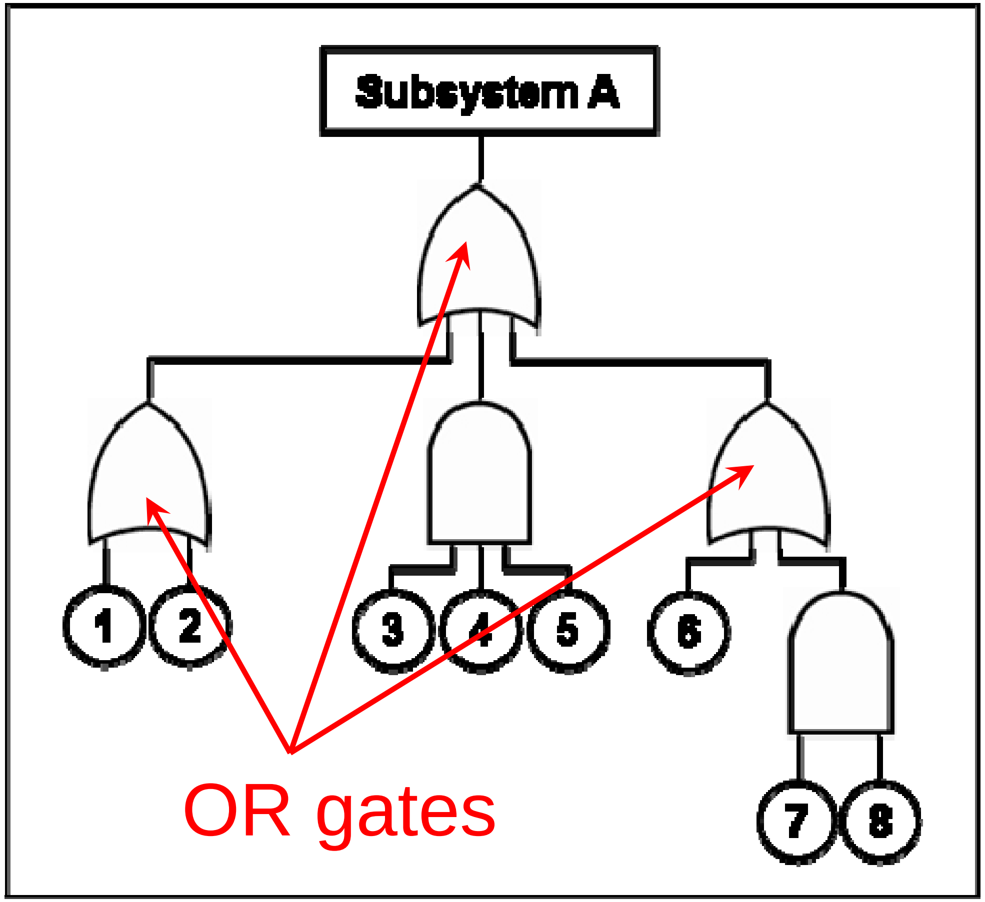
Obtaining Pr(*s*) from Fault trees



- Calculate
- Approximate
- Simulate

  http://en.wikipedia.org/wiki/Image:Fault_tree.png
- http://www.weibull.com/basics/fault-tree/index.htm

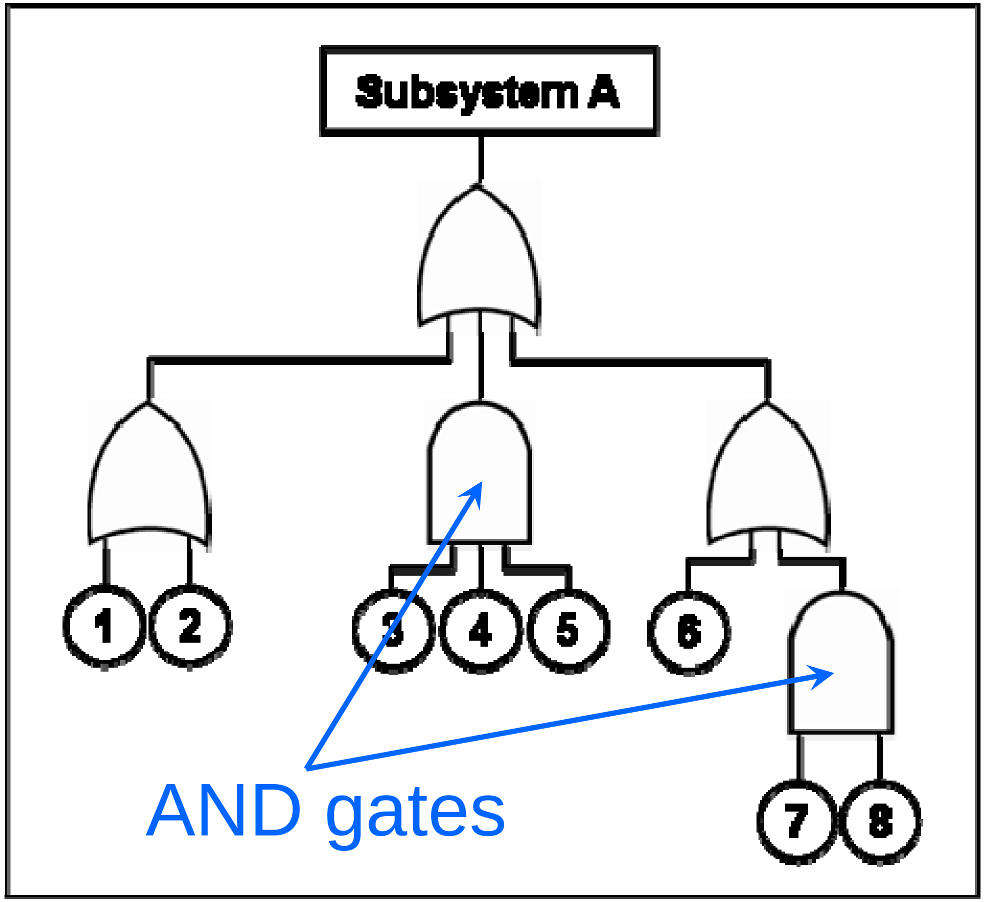
Obtaining Pr(*s*) from Fault trees



- Calculate
- Approximate
- Simulate

  http://en.wikipedia.org/wiki/Image:Fault_tree.png
- http://www.weibull.com/basics/fault-tree/index.htm

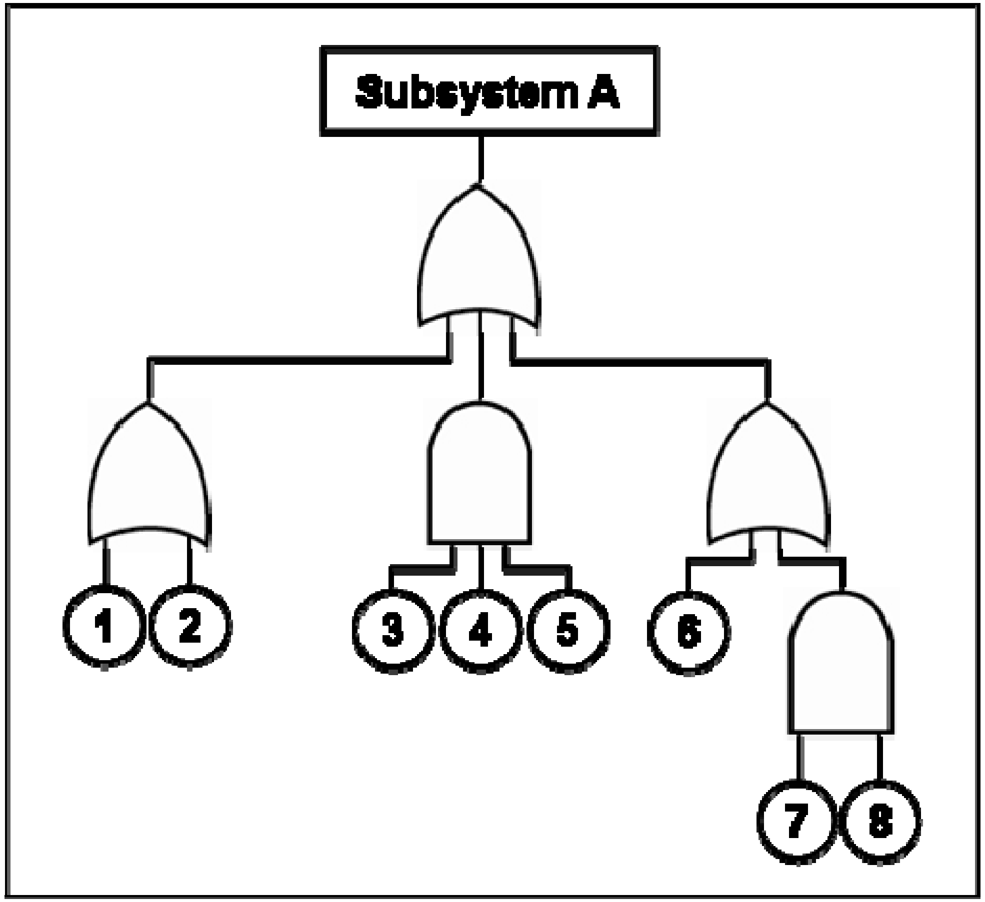
Obtaining Pr(*s*) from Fault trees



- Calculate
- Approximate
- Simulate

  http://en.wikipedia.org/wiki/Image:Fault_tree.png
- http://www.weibull.com/basics/fault-tree/index.htm

Obtaining Pr(*s*) from Fault trees

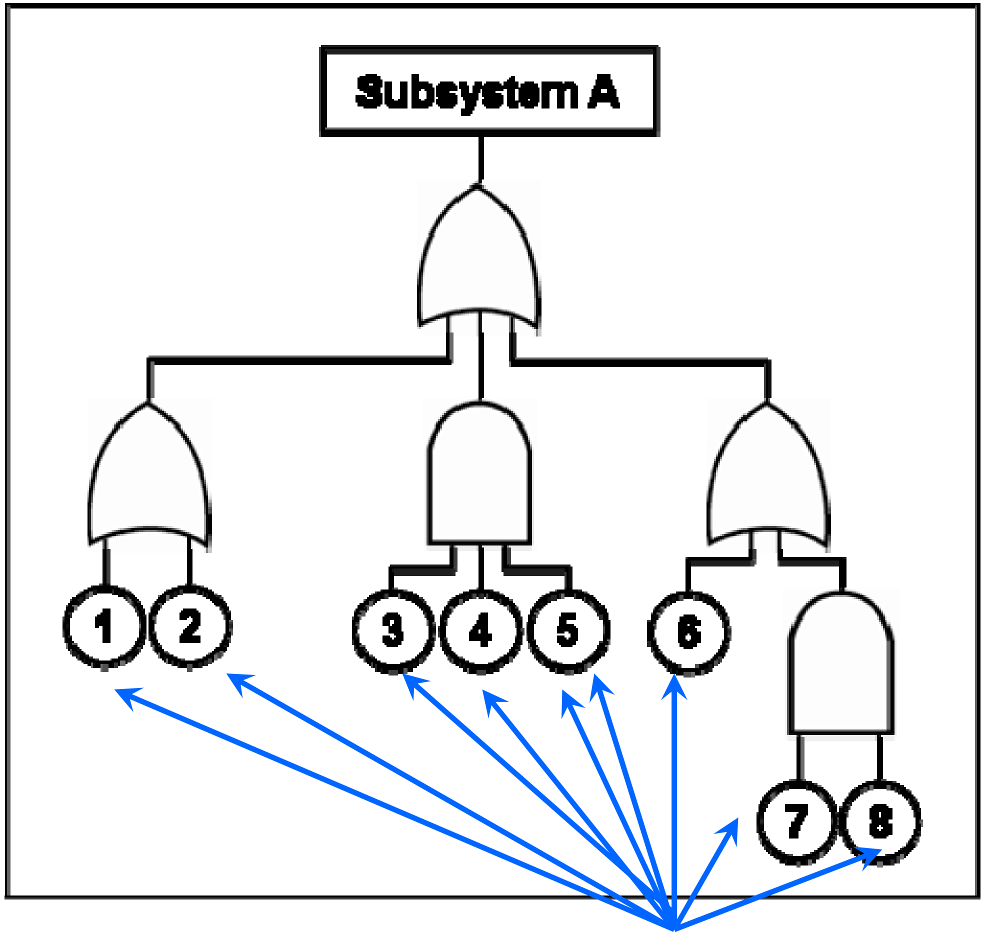


- Calculate
- Approximate
- Simulate

  http://en.wikipedia.org/wiki/Image:Fault_tree.png
- http://www.weibull.com/basics/fault-tree/index.htm

Causal modeling via **decomposition**

Obtaining Pr(*s*) from Fault trees



- Calculate
- Approximate
- Simulate

  http://en.wikipedia.org/wiki/Image:Fault_tree.png

- http://www.weibull.com/basics/fault-tree/index.htm

Causal modeling via **decomposition**

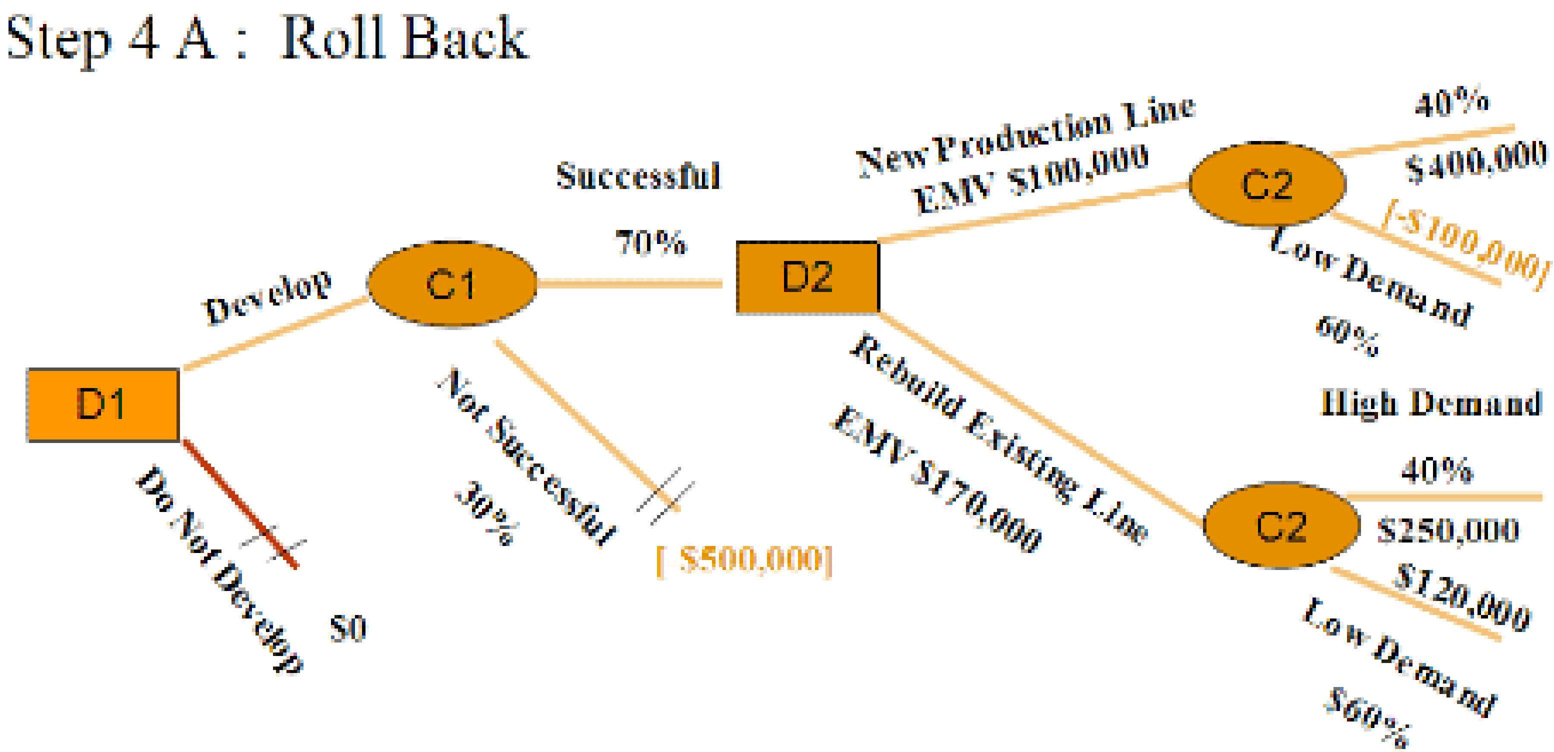
Still need data on basic event probabilities

Summary on fault trees

- Resolve the probability of a “top event” (root of tree) into a logical combination of probabilities of “basic events” (tips of tree)
  – Are probabilities of basic events trustworthy?
- Use approximation, simulation, or exact calculations to deduce probability of top events from probabilities of basic events.
- Many challenges remain
  – Common causes, interdependencies, failures of imagination, simulation of rate events, risk estimates with zero observed cases, etc.

Obtaining Pr(*s*) from Decision trees

http://www.eogogics.com/talkgogics/tutorials/decision-tree



Decision 1: Develop or Do Not Develop

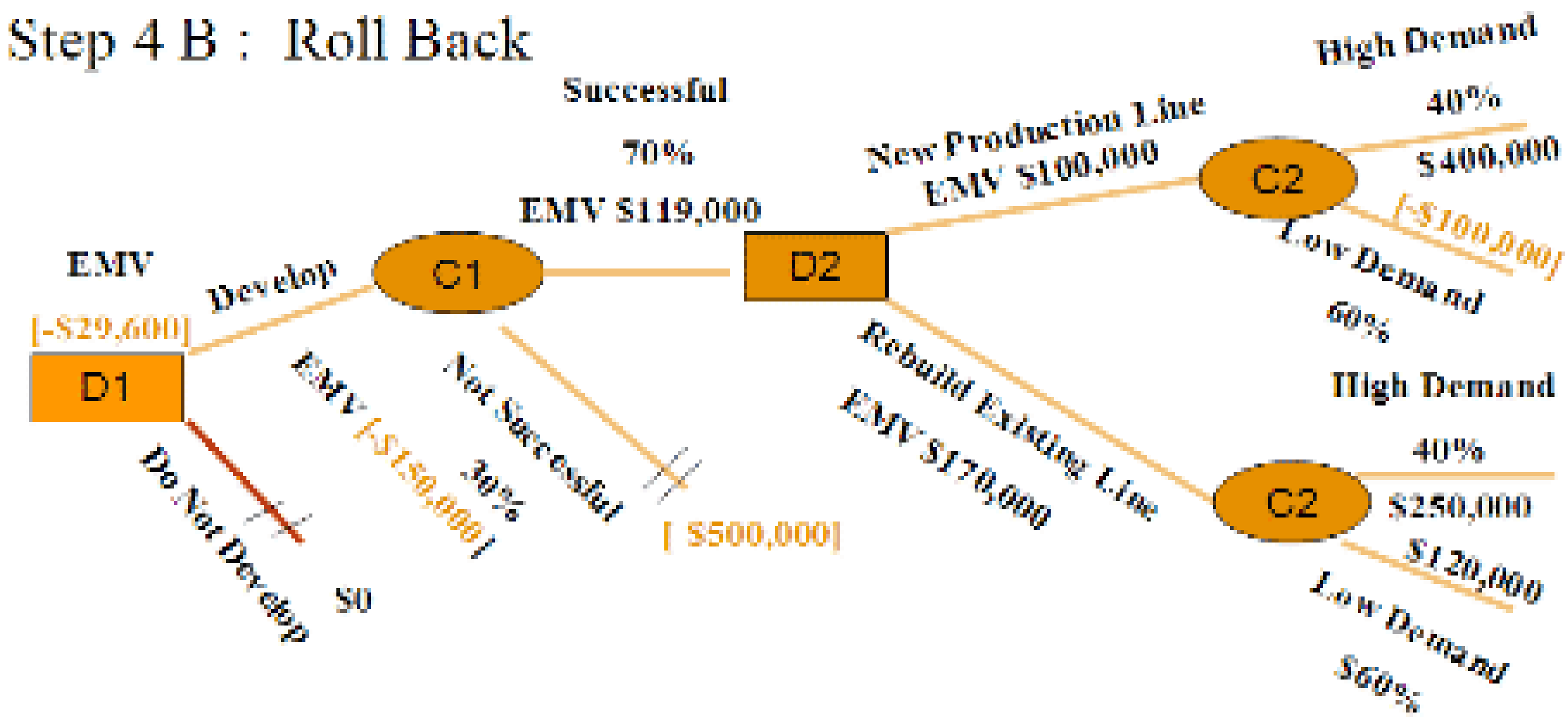
Development Successful + Development Unsuccessful

(70% X $172,000) (30% x (- $500,000))

$120,400 + (-$150,000)

Obtaining Pr(*s*) from Decision trees

http://www.eogogics.com/talkgogics/tutorials/decision-tree



Decision 1: Develop or Do Not Develop

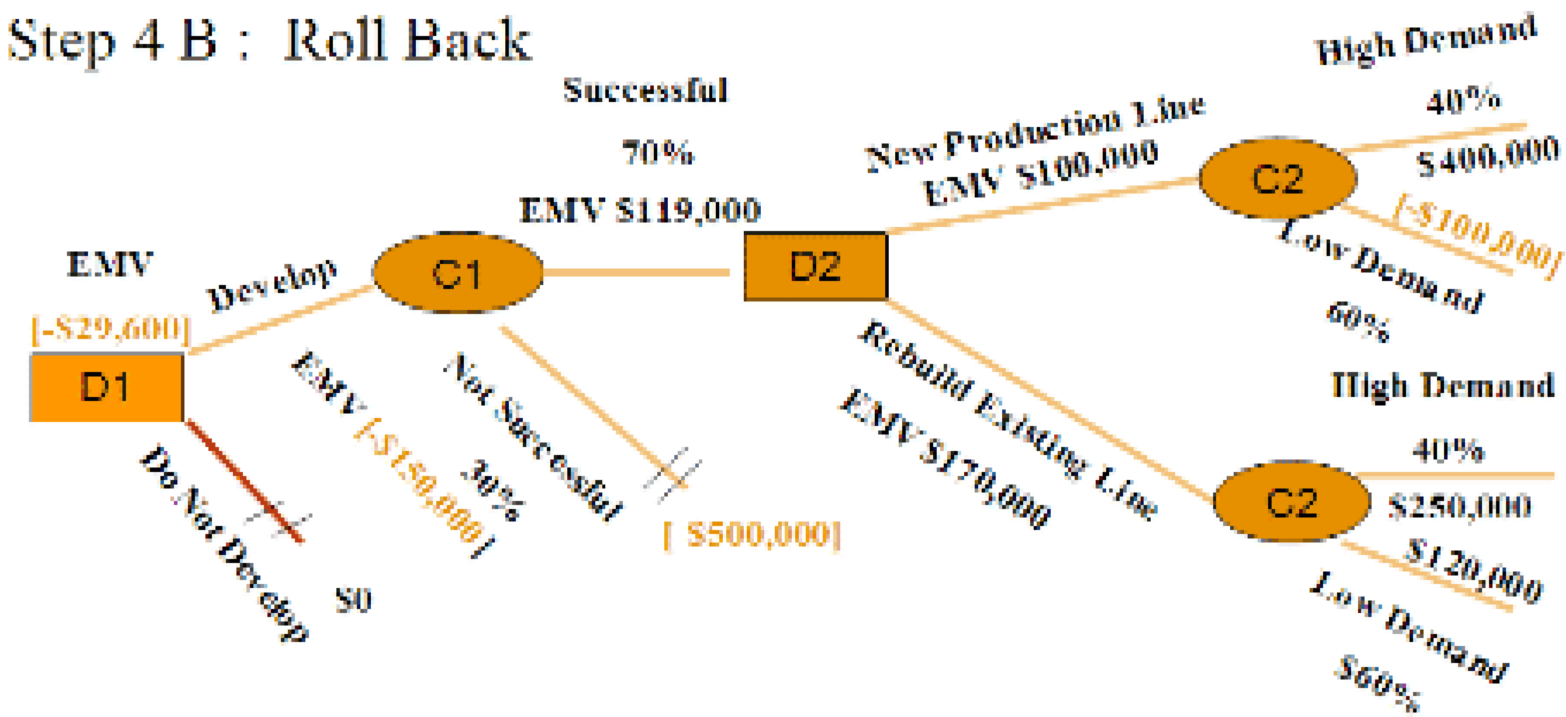
Development Successful + Development Unsuccessful

(70% X $172,000) (30% x (- $500,000))

$120,400 + (-$150,000)

What happened to act *a* and state *s*?

http://www.eogogics.com/talkgogics/tutorials/decision-tree



Decision 1: Develop or Do Not Develop

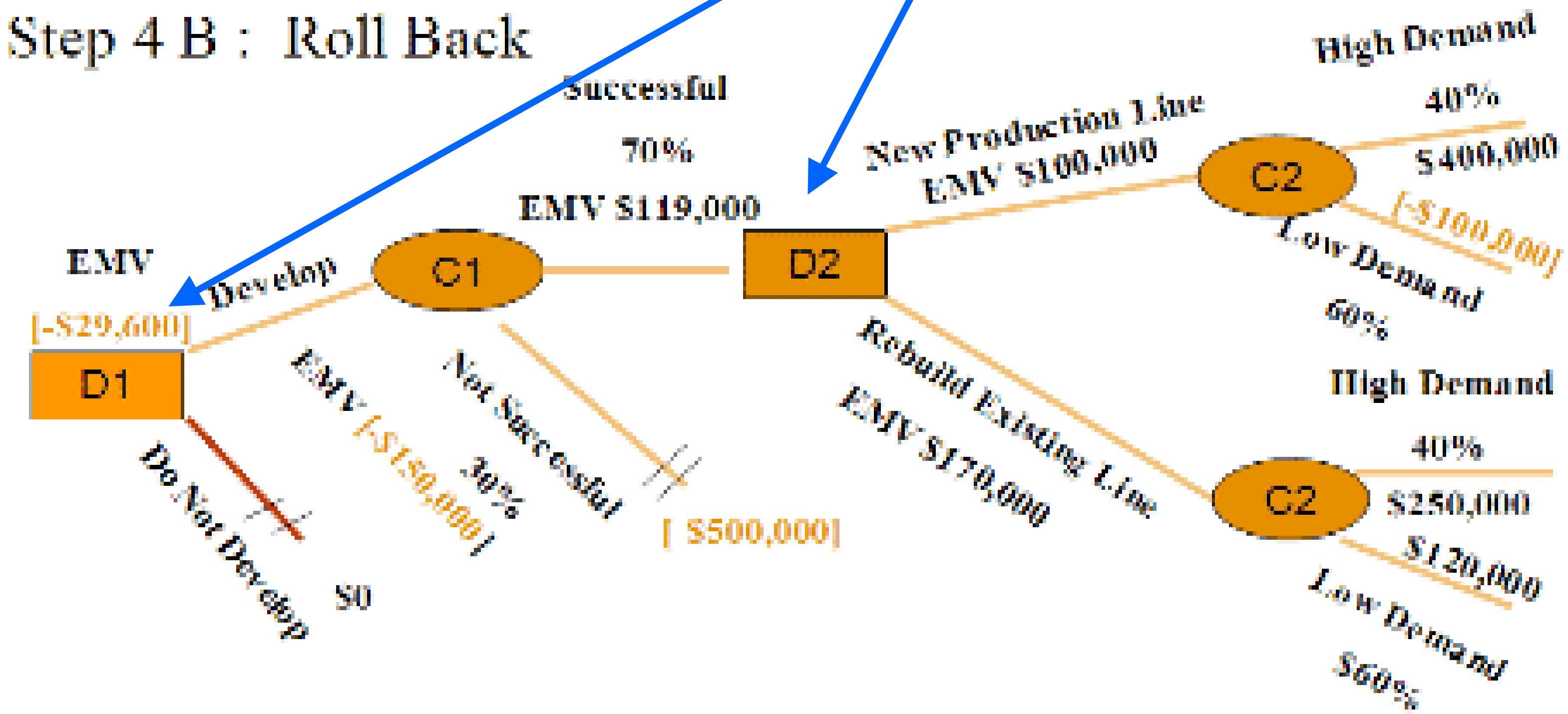
Development Successful + Development Unsuccessful

(70% X $172,000) (30% x (- $500,000))

$120,400 + (-$150,000)

What happened to act *a* and state *s*?

http://www.eogogics.com/talkgogics/tutorials/decision-tree



Decision 1: Develop or Do Not Develop

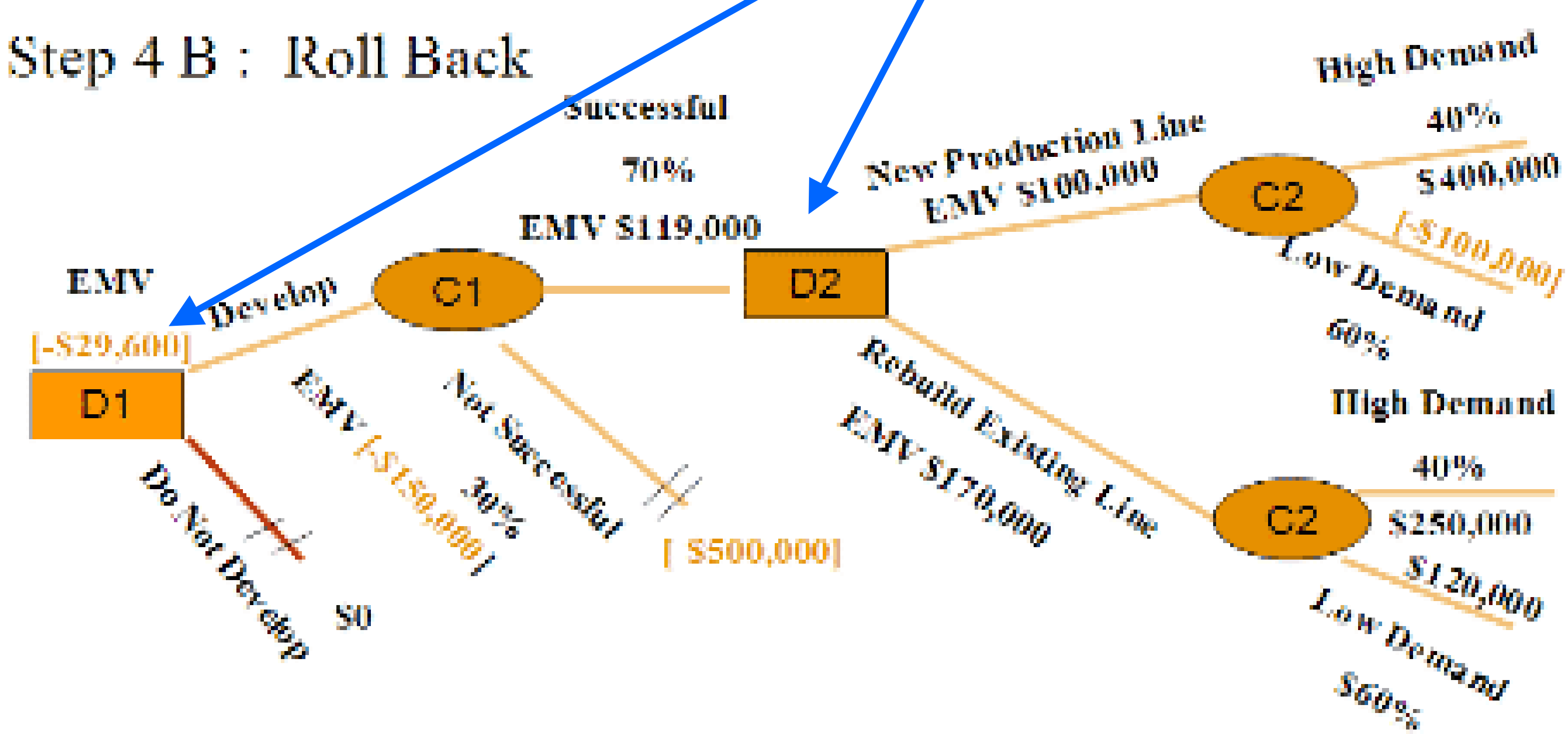
Development Successful + Development Unsuccessful

(70% X $172,000) (30% x (- $500,000))

$120,400 + (-$150,000)

What happened to act *a* and state *s*?

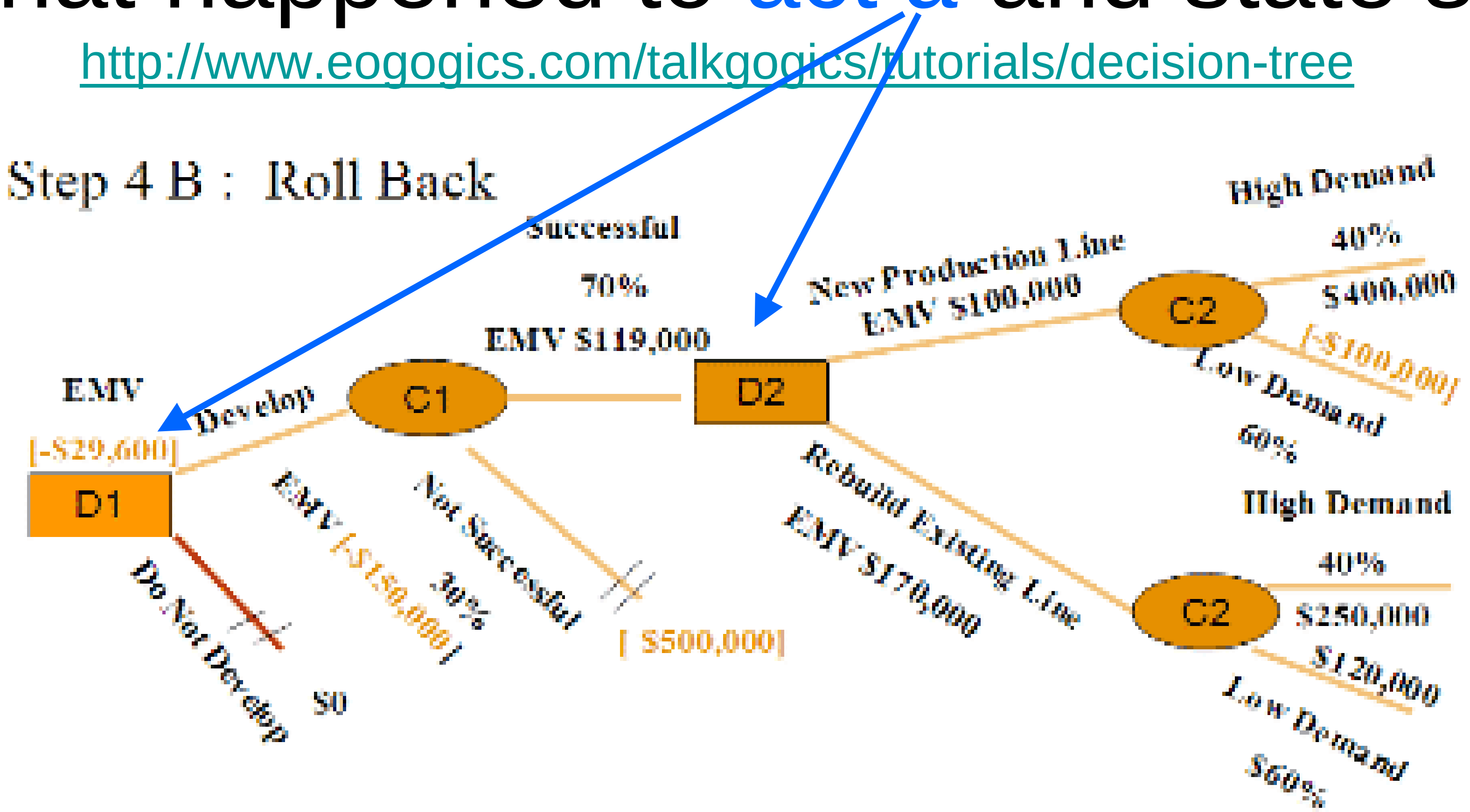
http://www.eogogics.com/talkgogics/tutorials/decision-tree



What are the 3 possible acts in this tree?

What happened to act *a* and state *s*?

http://www.eogogics.com/talkgogics/tutorials/decision-tree

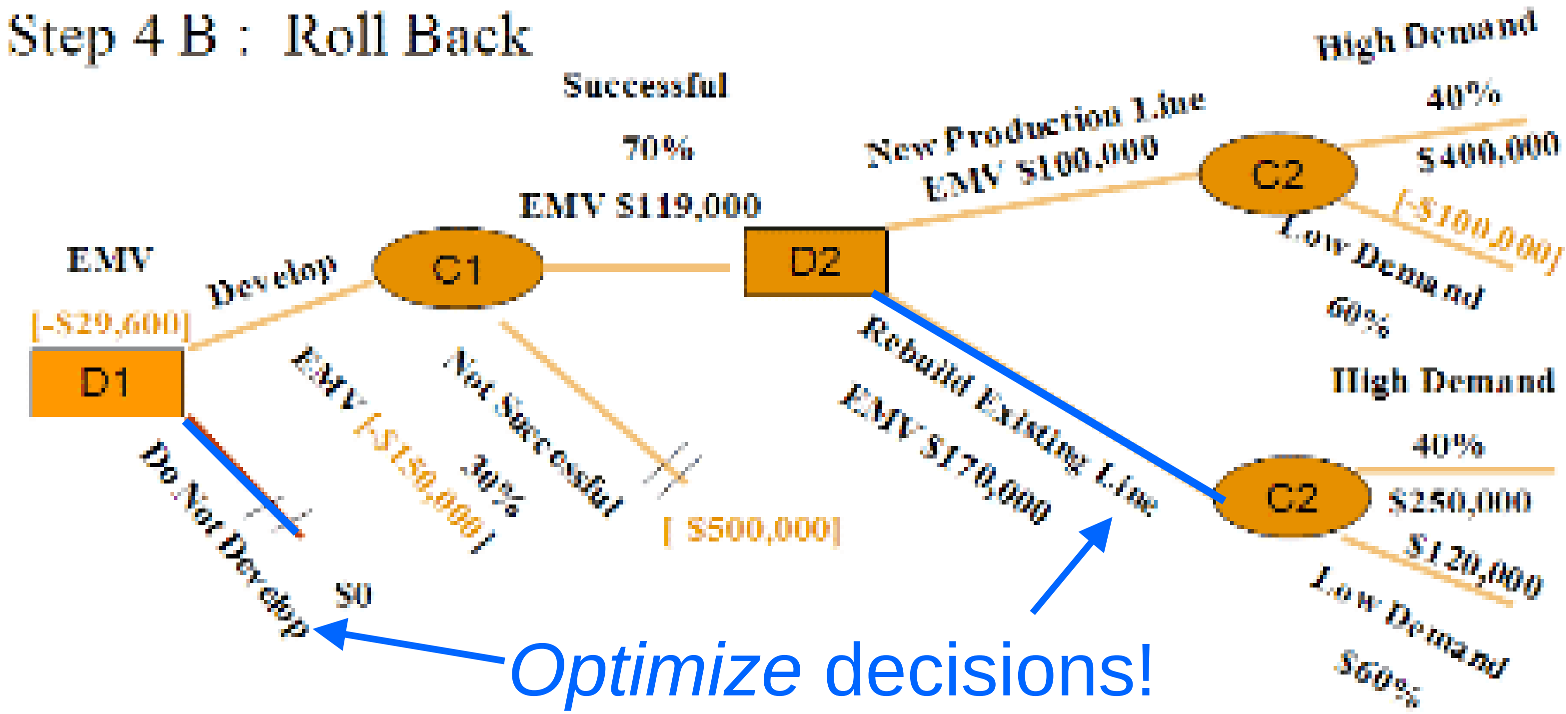


What are the 3 possible acts in this tree?

(a) Don't develop; (b) Develop, then rebuild if successful; (c) Develop, then new line if successful.

What happened to act *a* and state *s*?

http://www.eogogics.com/talkgogics/tutorials/decision-tree



What are the 3 possible acts in this tree?

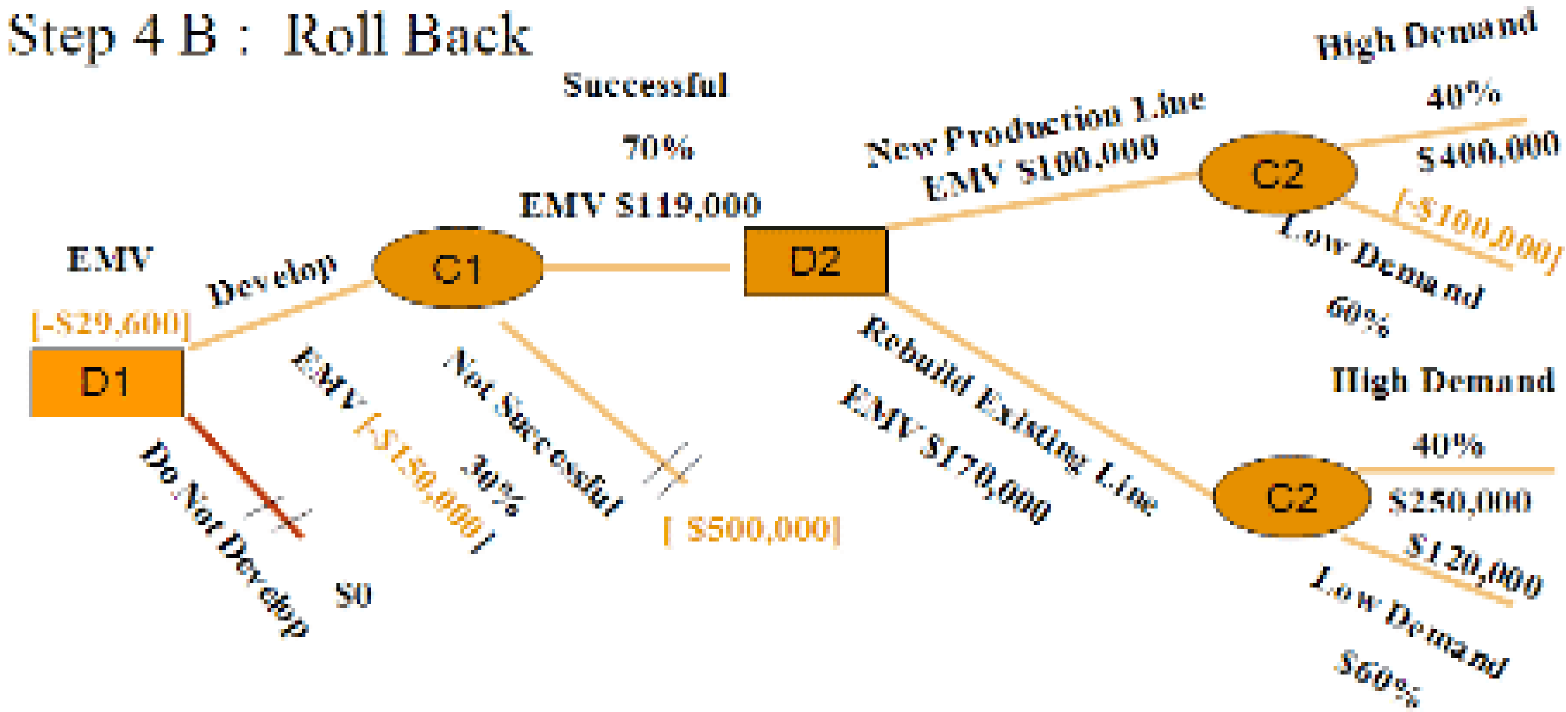
(a) Don't develop; (b) Develop, then rebuild if successful; (c) Develop, then new line if successful.

Key points

- Solving decision trees (with decisions) requires embedded *optimization*
  - Make future decisions optimally, given the information available when they are made
- *Event trees* = decision trees with no decisions
  - Can be solved, to find outcome probabilities, by forward Monte-Carlo simulation, or by multiplication and addition
- In general, sequential decision-making cannot be modeled well using event trees.
  - Must include (optimal choice | information)

What happened to state *s*?

http://www.eogogics.com/talkgogics/tutorials/decision-tree

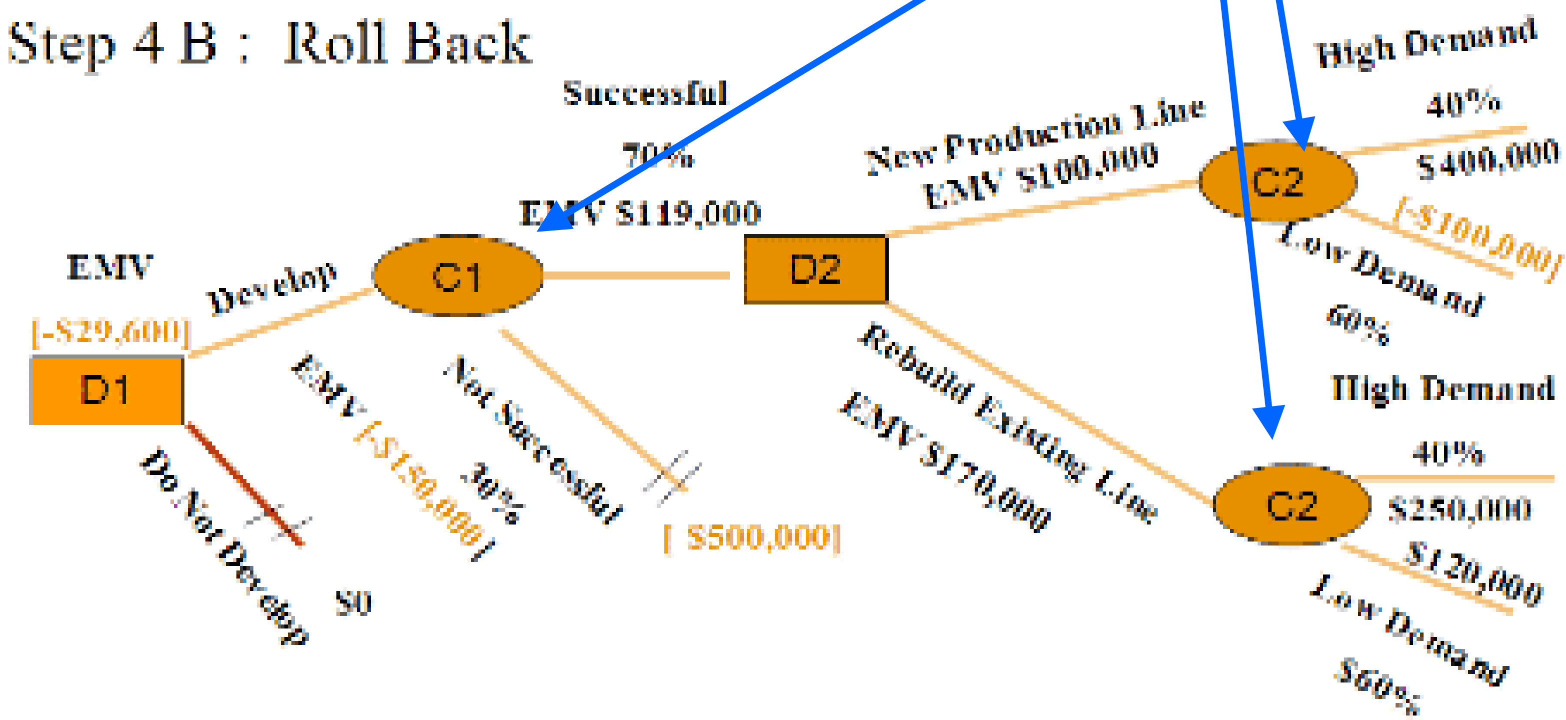


What are the 3 possible acts in this tree?

(a) Don't develop; (b) Develop, then rebuild if successful; (c) Develop, then new line if successful.

What happened to state s?

http://www.eogogics.com/talkgogics/tutorials/decision-tree

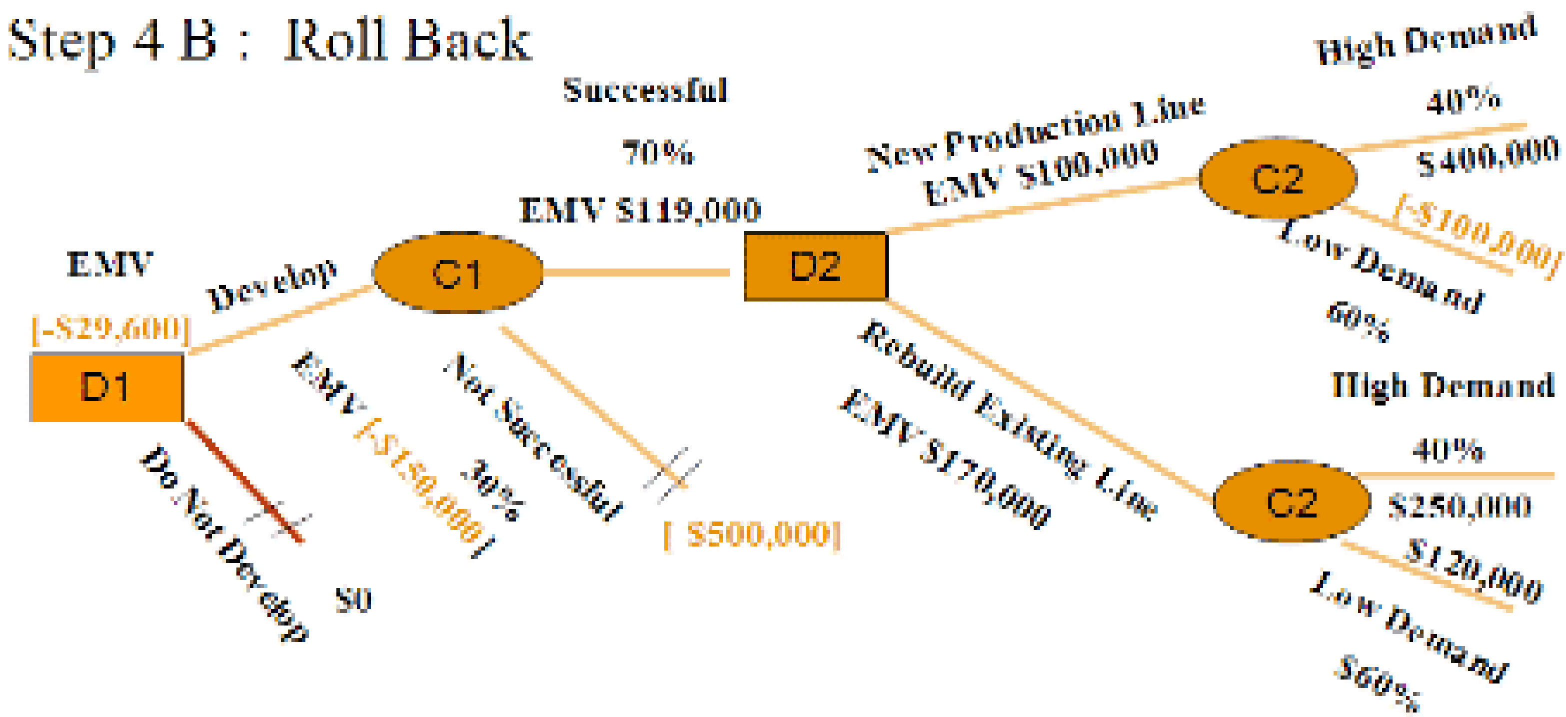


What are the 4 possible states?

C1 can succeed or not; C2 can be high or low demand

Acts and states cause consequences

http://www.eogogics.com/talkgogics/tutorials/decision-tree



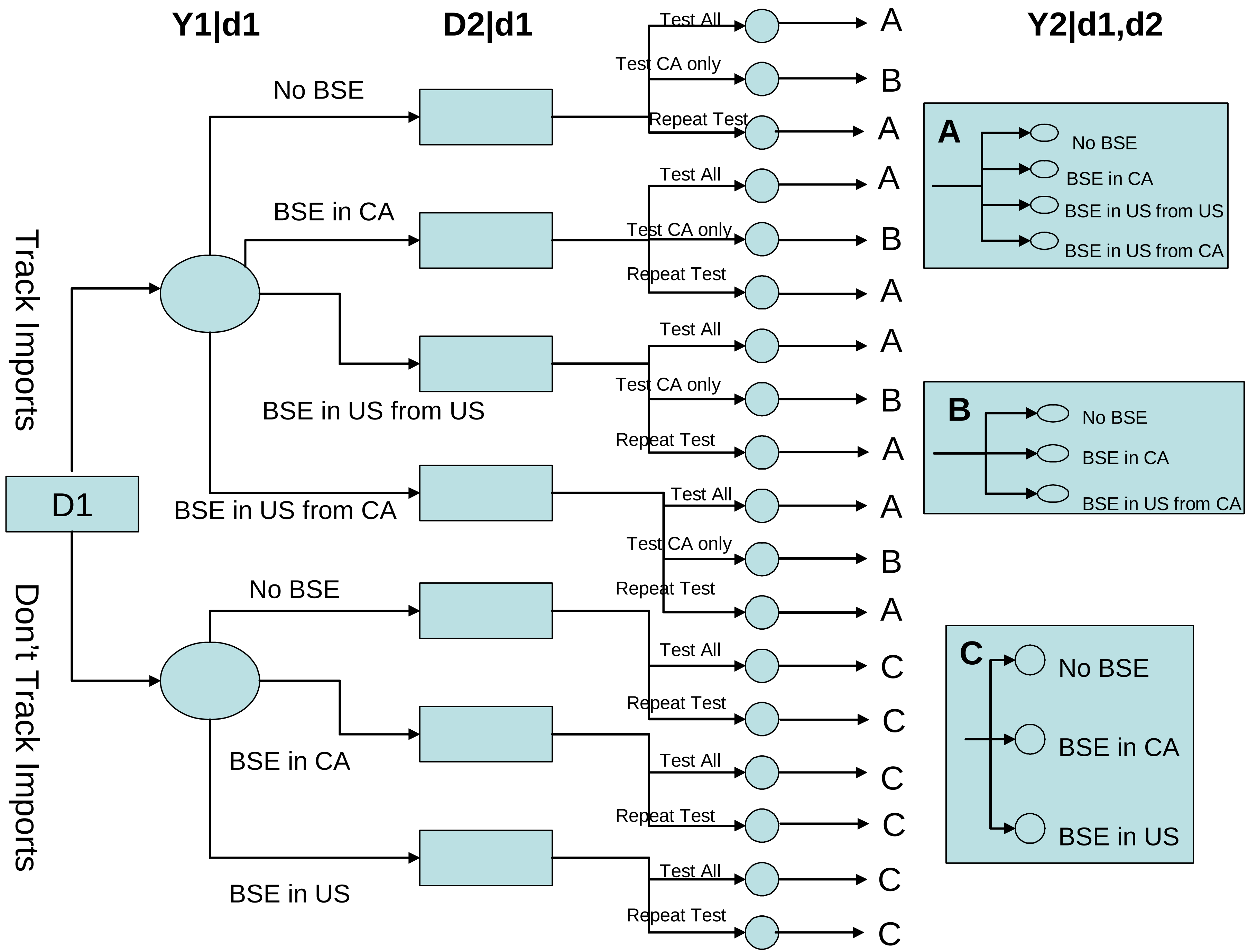
Key theoretical insight

- A complex decision model can be viewed as a simple $c(a, s)$ model.
  - s = selection of branch at each chance node
  - a = selection of branch at each choice node
    - *Decision rule* or *policy*
  - c = outcome at terminal node for (a, s)
- Other complex decision models can also be interpreted as $c(a, s)$ or $\Pr(c \mid a, s)$ models
  - s = system state & information signal
  - a = *decision rule* (information $\rightarrow$ act)
  - c may include changes in s and in possible a.

Other complex decision models

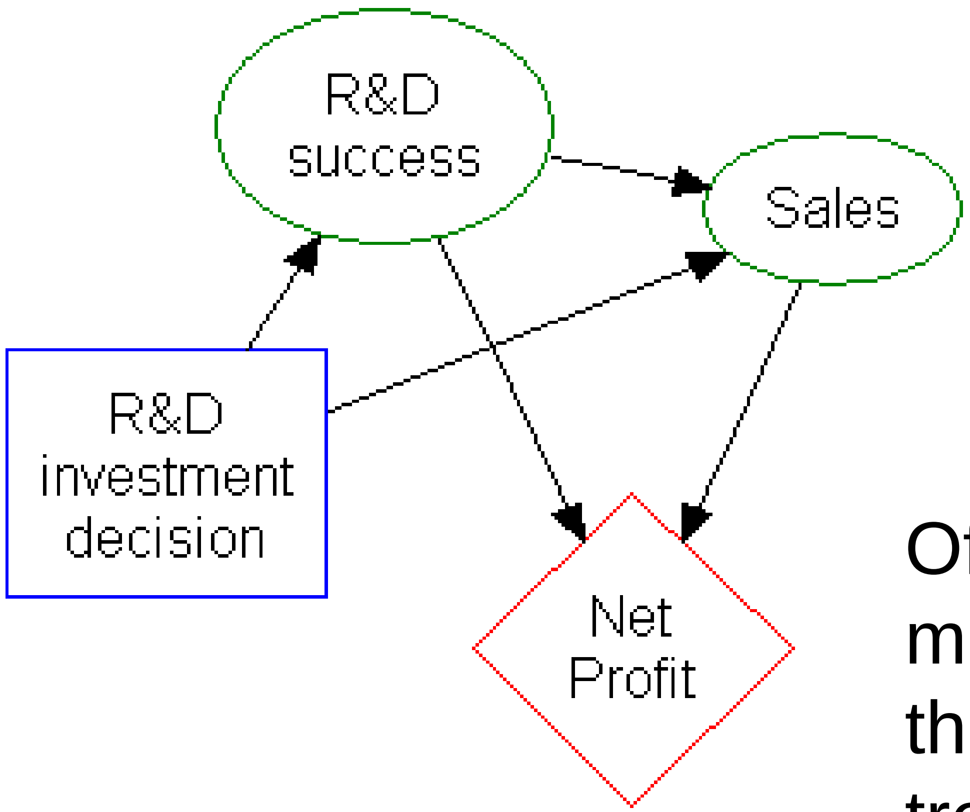
- Markov decision process (MDP)
  - Pr(*next state* | *current state*, *current act*)
  - *Reward(current state, current act*)
- Partially observable MDP (POMDP)
  - Pr(*signal* | *current state*)
- Dynamic Bayesian Network (DBN)
- Stochastic optimal control system
  - *a* = feedback control law

Real decision trees can quickly become “bushy messes” (Raiffa, 1968) with many duplicated sub-trees


Y1|d1
D2|d1
Y2|d1,d2
Track Imports
Don't Track Imports
D1
No BSE
BSE in CA
BSE in US from US
BSE in US from CA
No BSE
BSE in CA
BSE in US
Test All
Test CA only
Repeat Test
A
B
C
A
No BSE
BSE in CA
BSE in US from US
BSE in US from CA
B
No BSE
BSE in CA
BSE in US from CA
C
No BSE
BSE in CA
BSE in US

Obtaining Pr(*s*): Influence Diagrams

http://en.wikipedia.org/wiki/Decision_tree



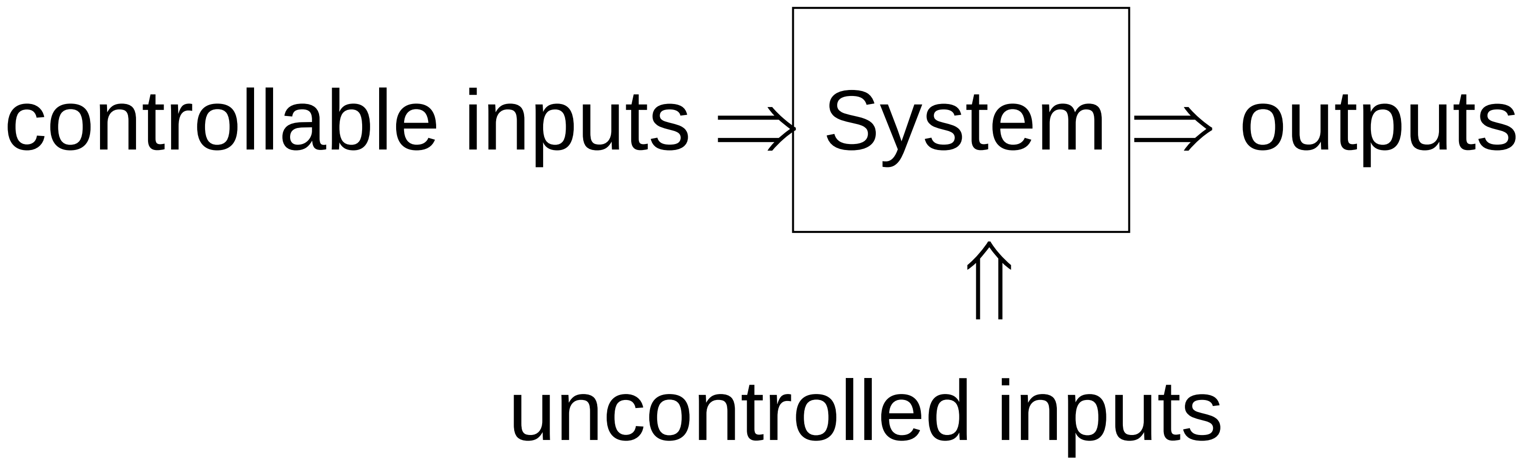
Often much more compact than decision trees

Summary on decision trees

- Decision trees show sequences of choices, chance nodes, observations, and final consequences.
  - Mix observations, acts, optimization, causality
- Good for very small problems; less good for medium-sized problems; unwieldy for large problems $\rightarrow$ use IDs instead
- Can view decision trees and other decision models as simple $c(a, s)$ models
  - But need good optimization solvers!

Dynamic systems, causality, repeated decisions

Systems, causality, decisions



Systems, causality, decisions

engineering system

industrial system, food chain

living system (organism, ecosystem)

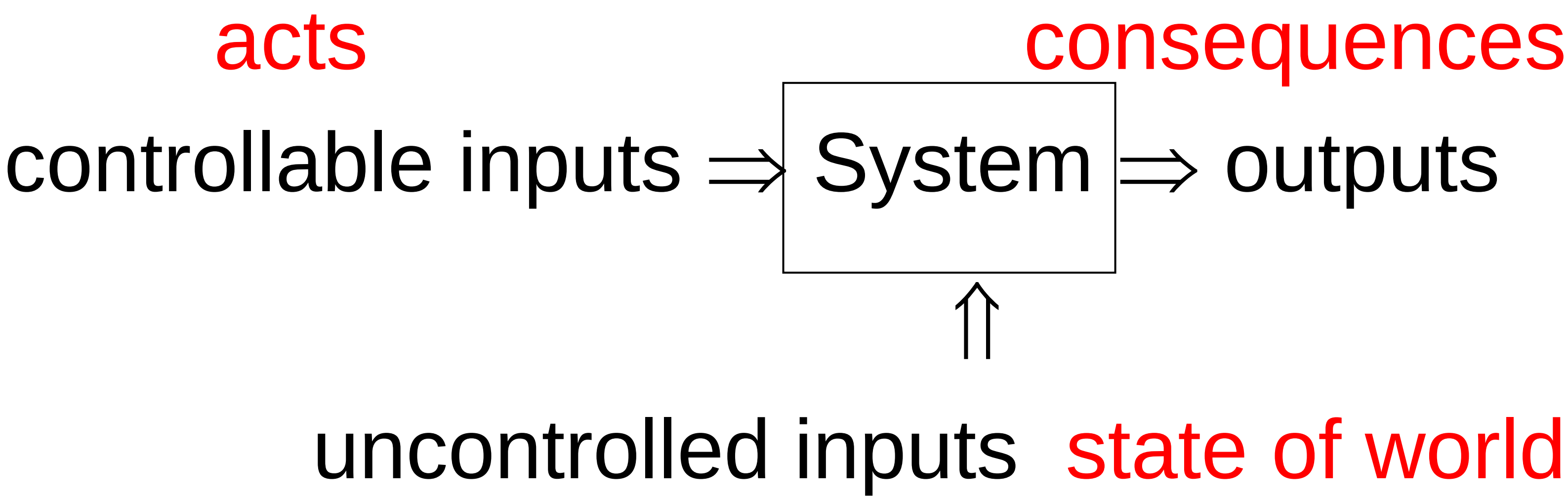
economic/business/social system

controllable inputs ⇒ System ⇒ outputs

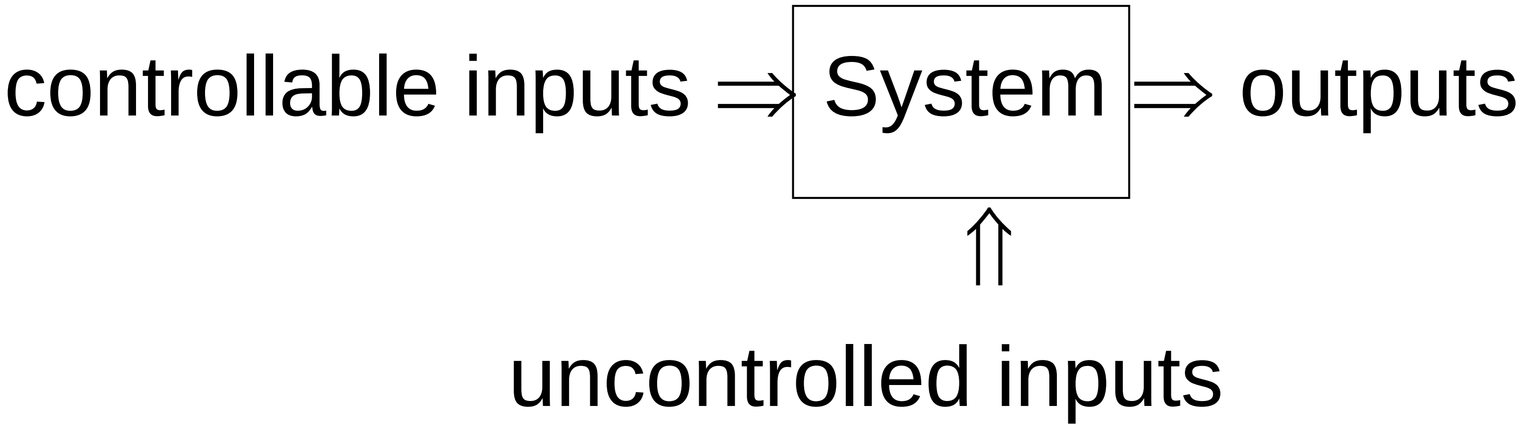
⇑

uncontrolled inputs

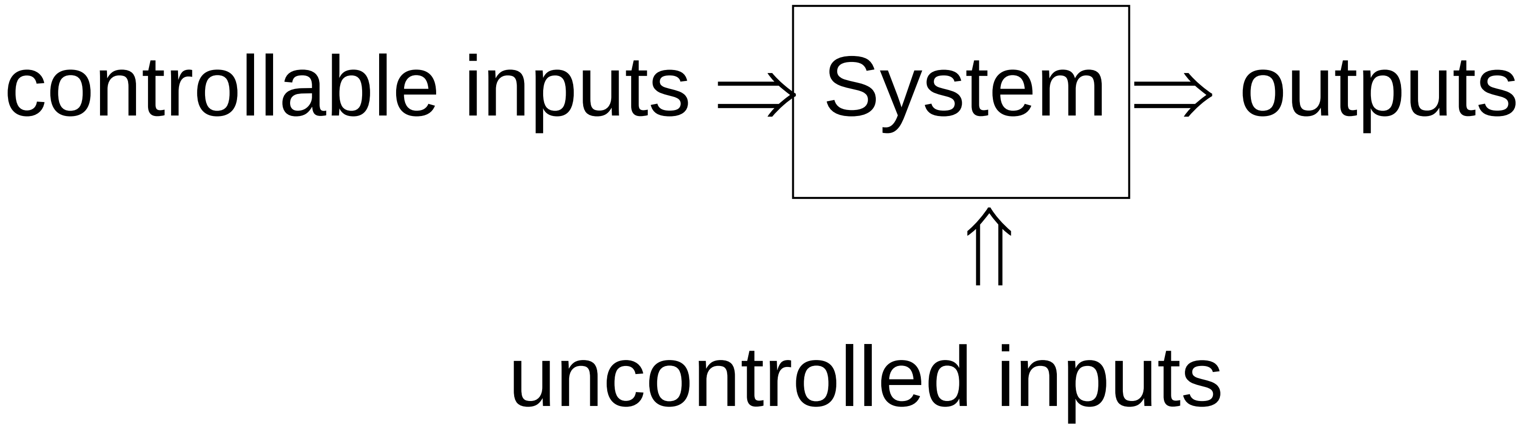
Systems, causality, decisions



How to predict Pr(*outputs* | *inputs*)?



How to predict Pr(*outputs* | *inputs*)?
How to choose controllable inputs?



Risk assessment $\rightarrow$ Pr(outputs | inputs)

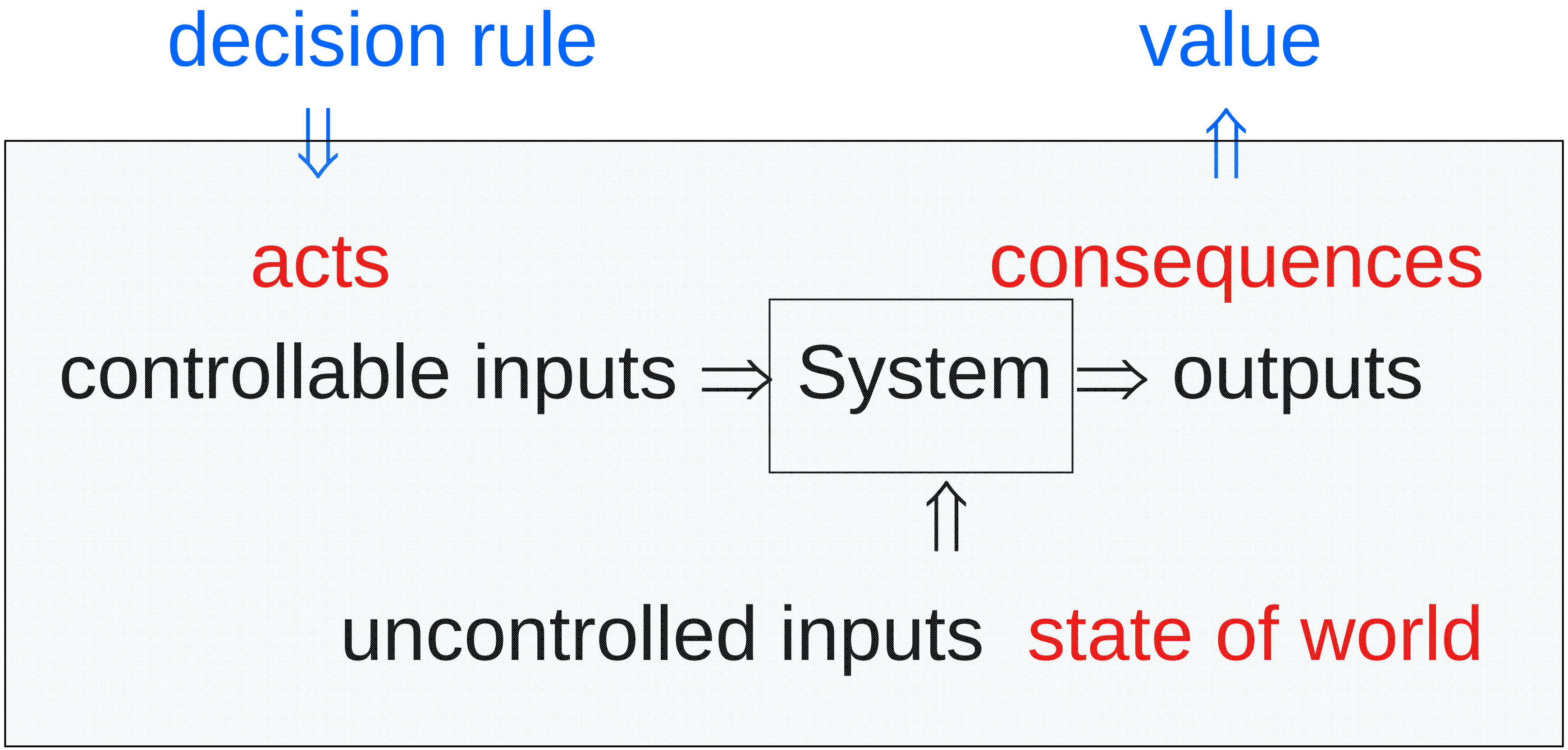
Risk management $\rightarrow$ choose inputs

controllable inputs $\Rightarrow$ System $\Rightarrow$ outputs

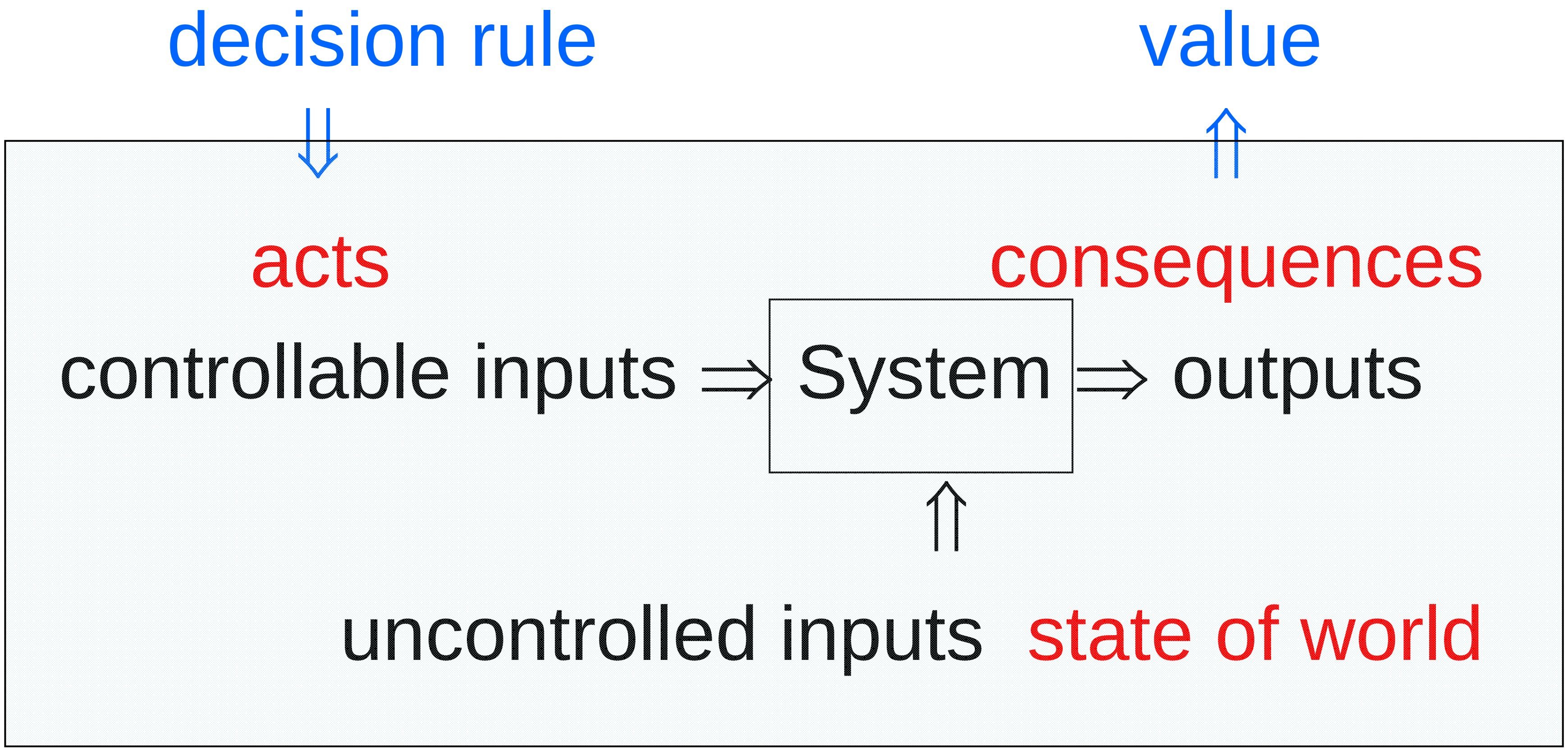
$\Uparrow$

uncontrolled inputs

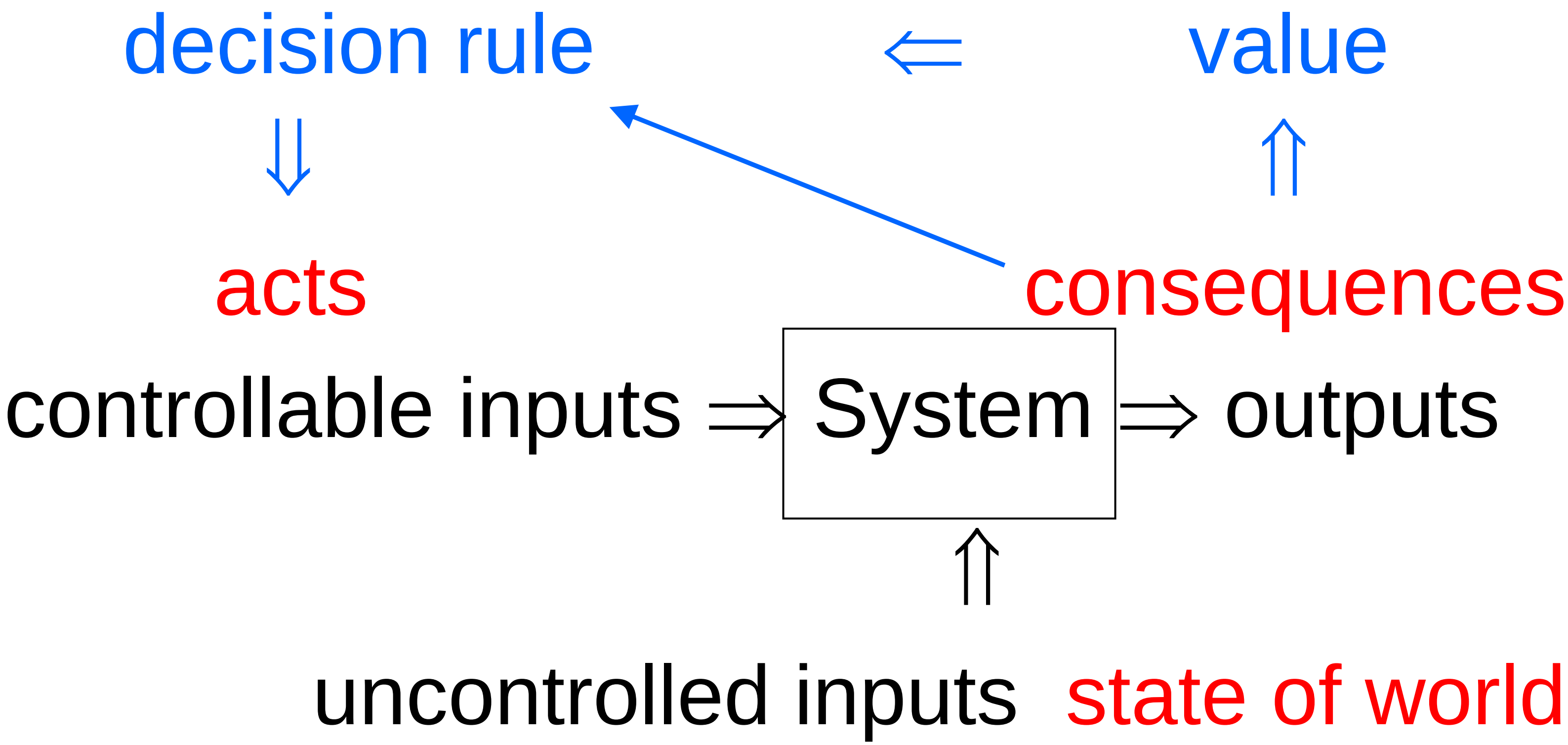
Systems, causality, decisions



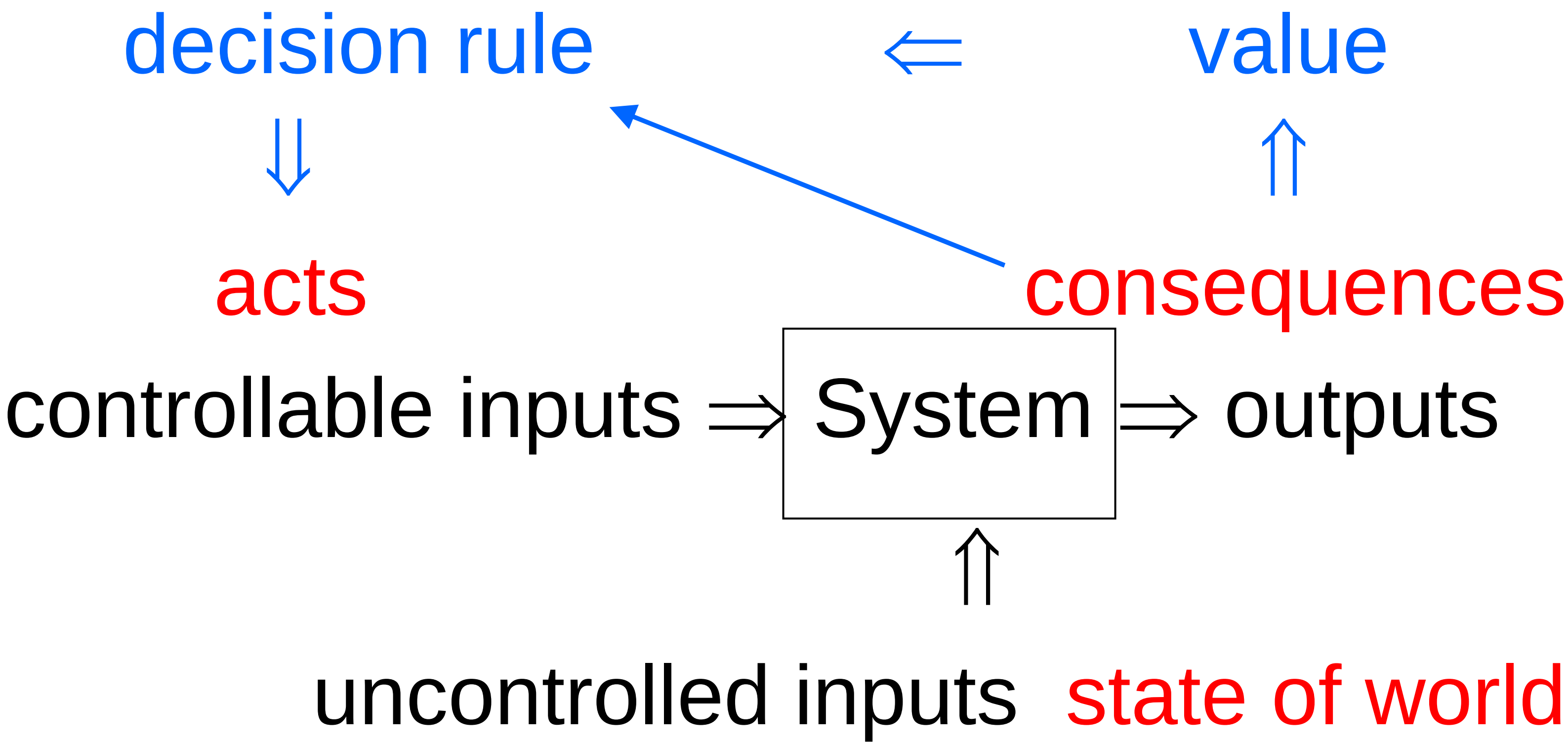
How to optimize acts/increase value?



1. Reinforcement learning

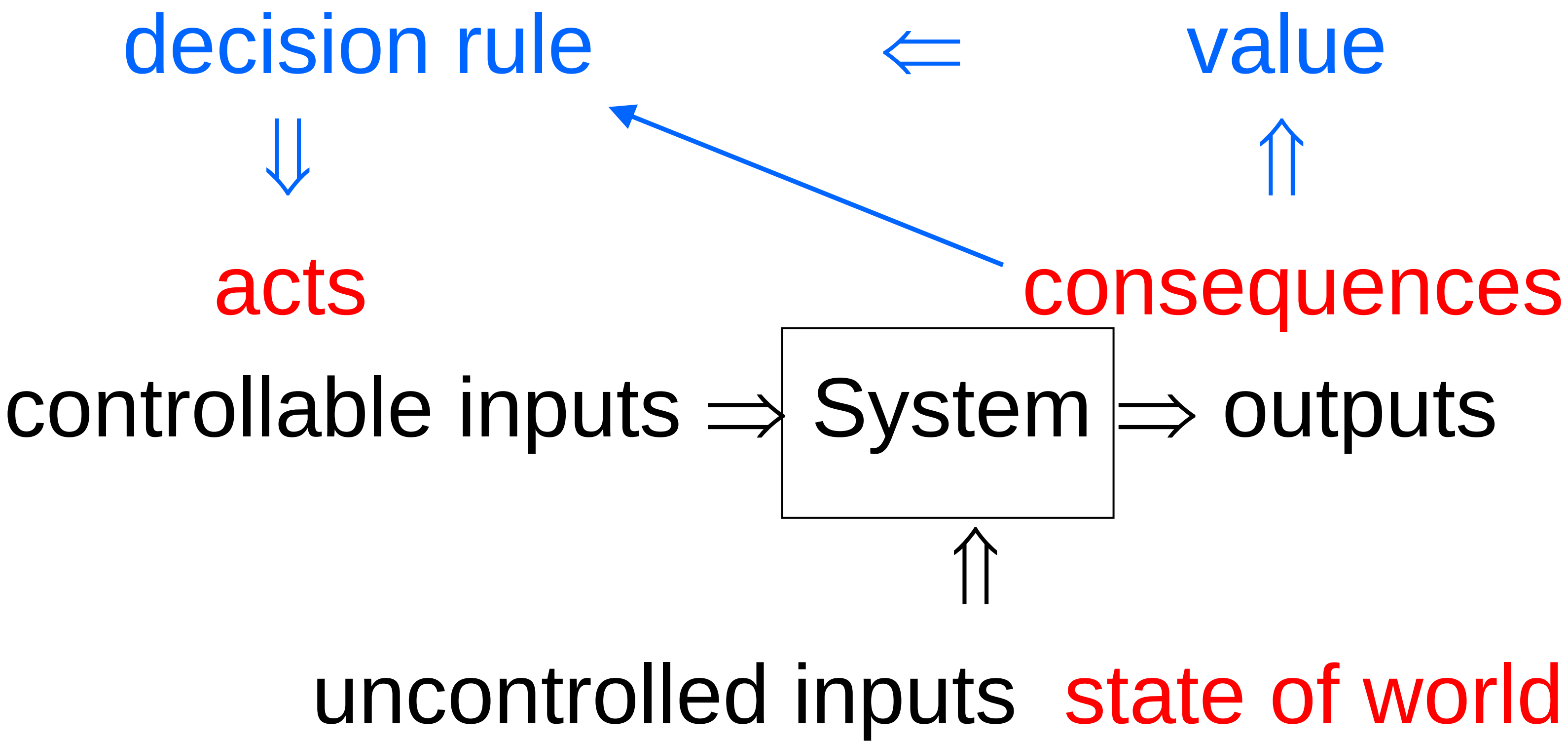


1. Reinforcement learning



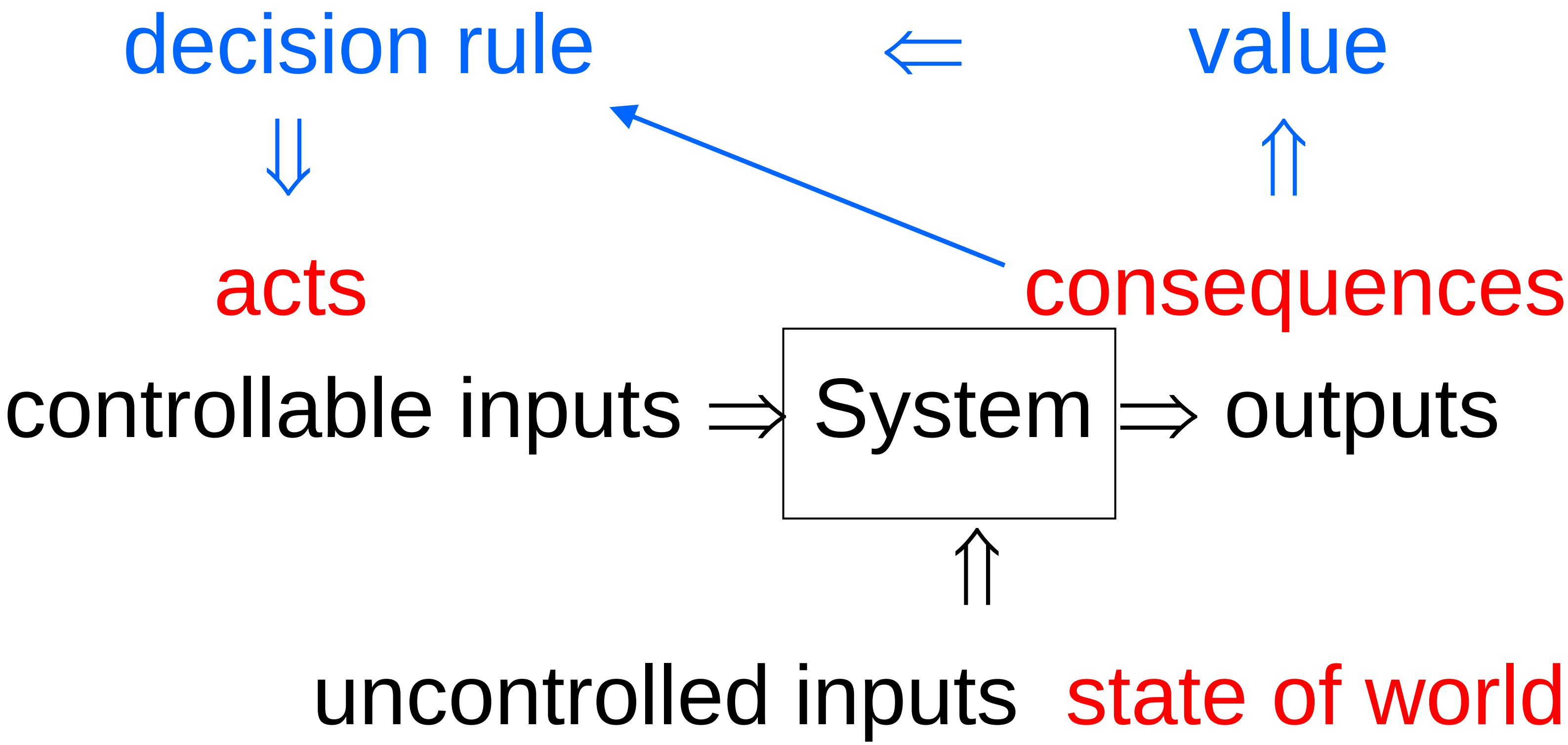
Good for *repeated trials* (or slowly varying).

Zero-regret learning



Choose $\Pr(a) = \exp(\alpha R(a))/\Sigma_x \exp(\alpha R(x))$,
where $R(x)$ = *average past reward* for act x

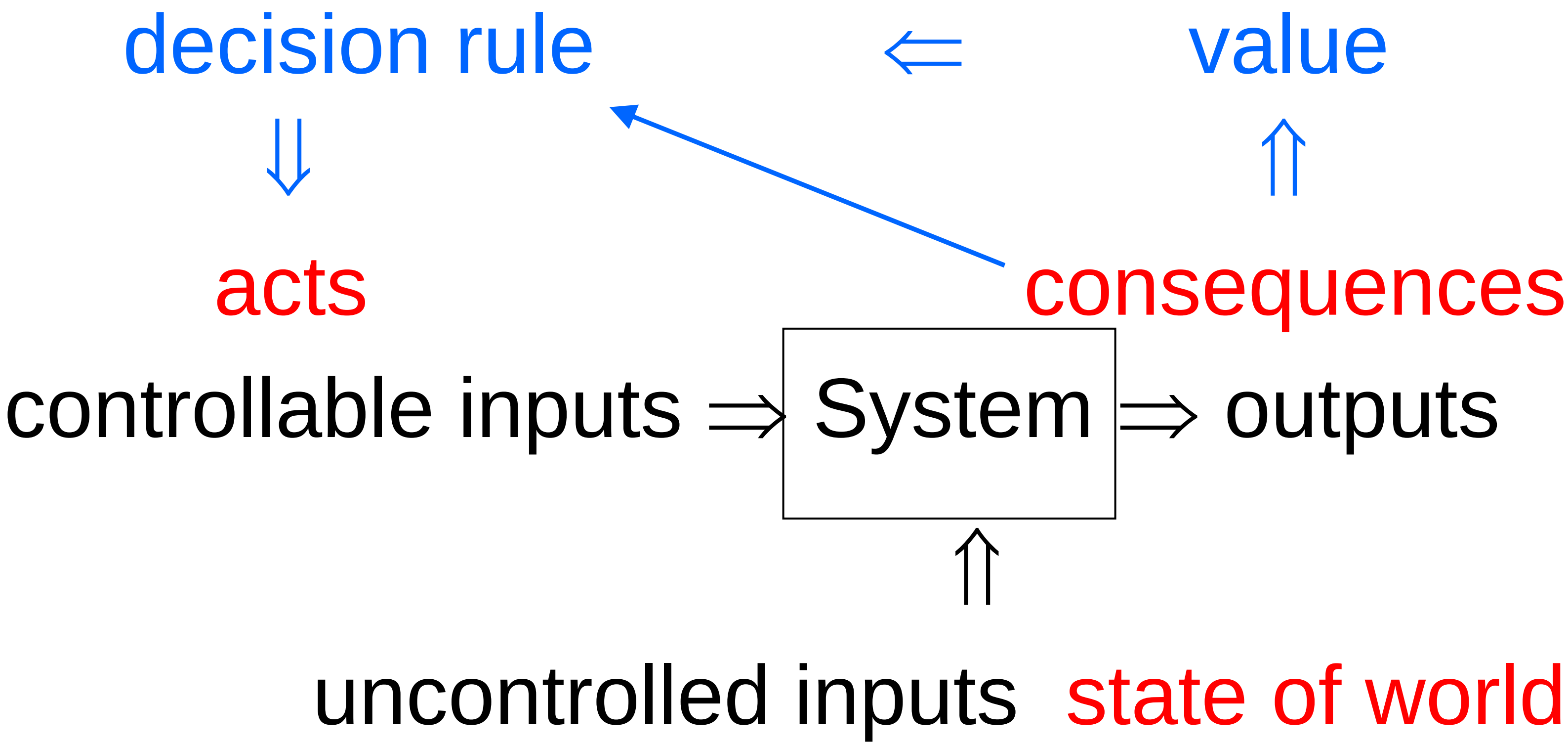
Zero-regret learning



Choose $\Pr(a) = \exp(\alpha R(a))/\Sigma_x \exp(\alpha R(x))$

Choose $\alpha = \sqrt{(8 \ln N / n)}$ for finite horizons

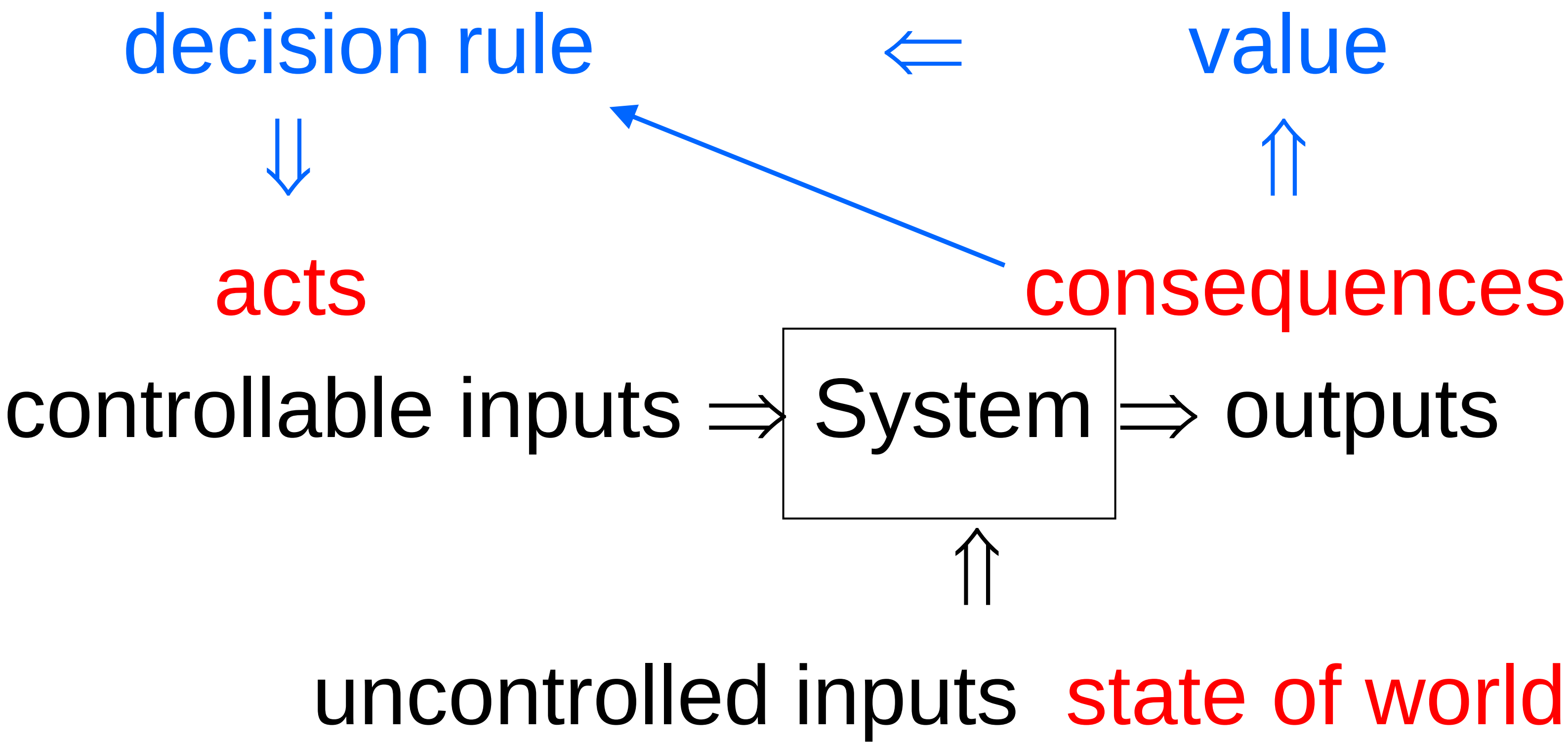
Zero-regret learning



Choose $\Pr(a) = \exp(\alpha R(a))/\Sigma_x \exp(\alpha R(x))$

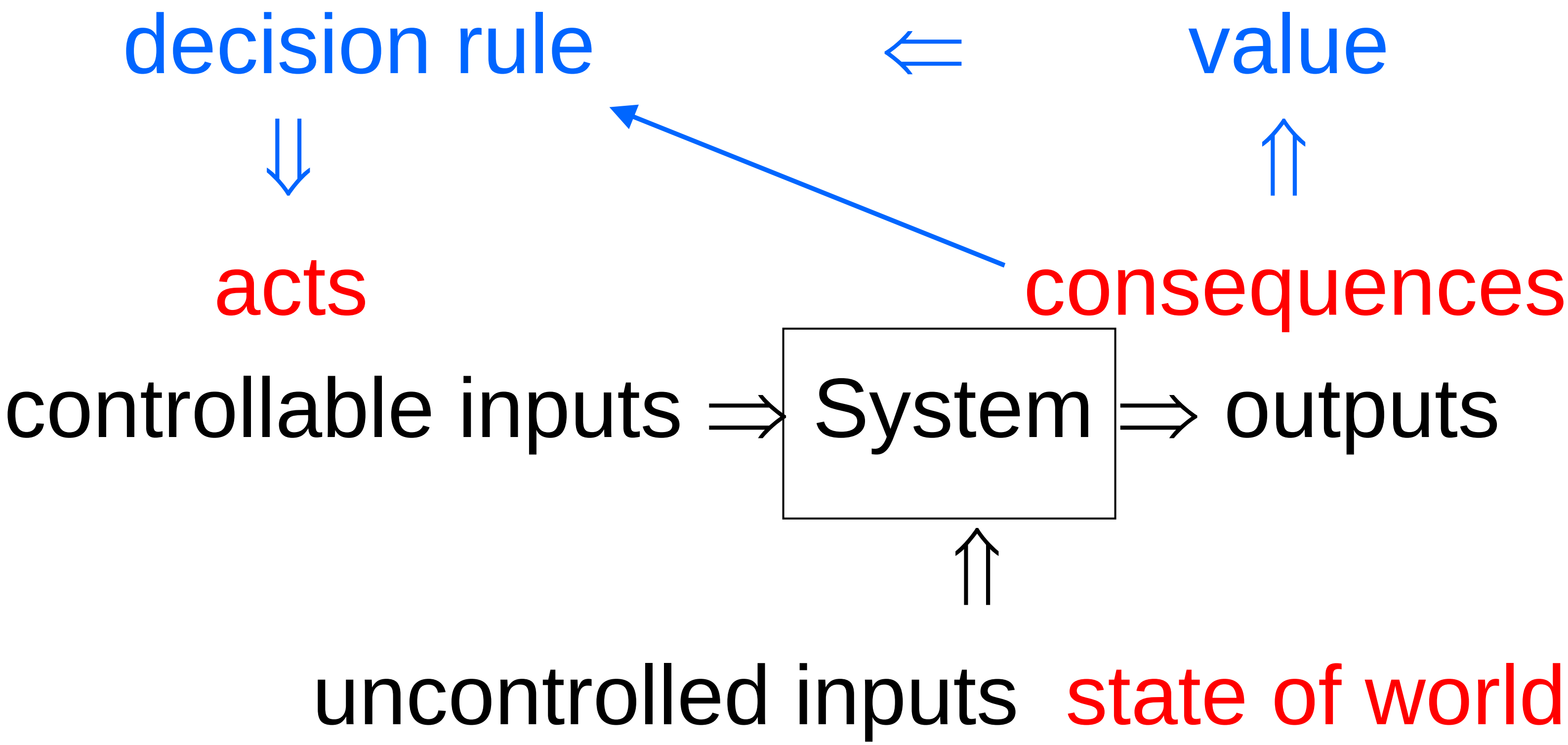
Then regret $\rightarrow$ 0 as $\sqrt{[\ln N/(2n)]}$ (!)

Zero-regret learning



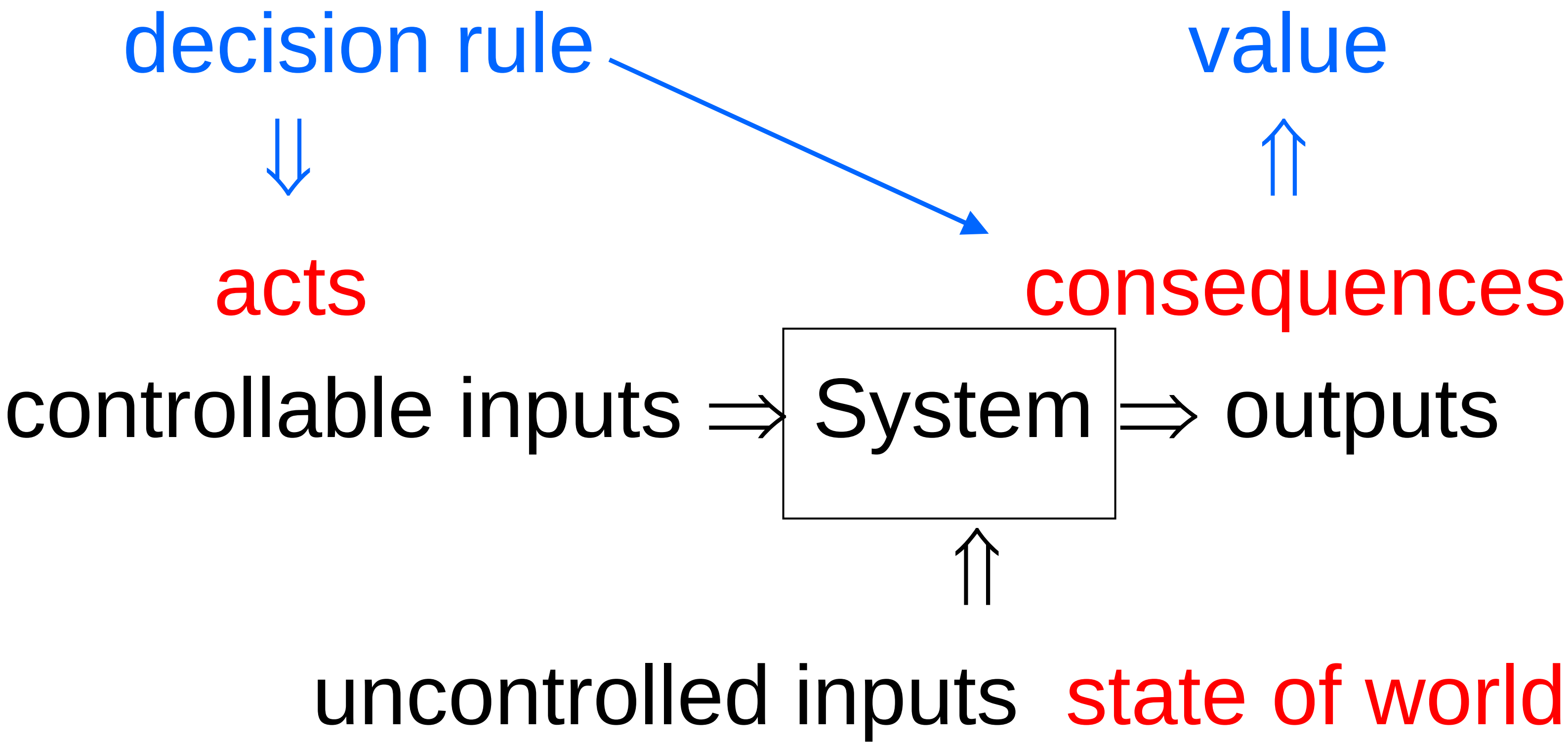
Increasing the probability of well-rewarded acts soon leads to optimal decision rule.

Zero-regret learning

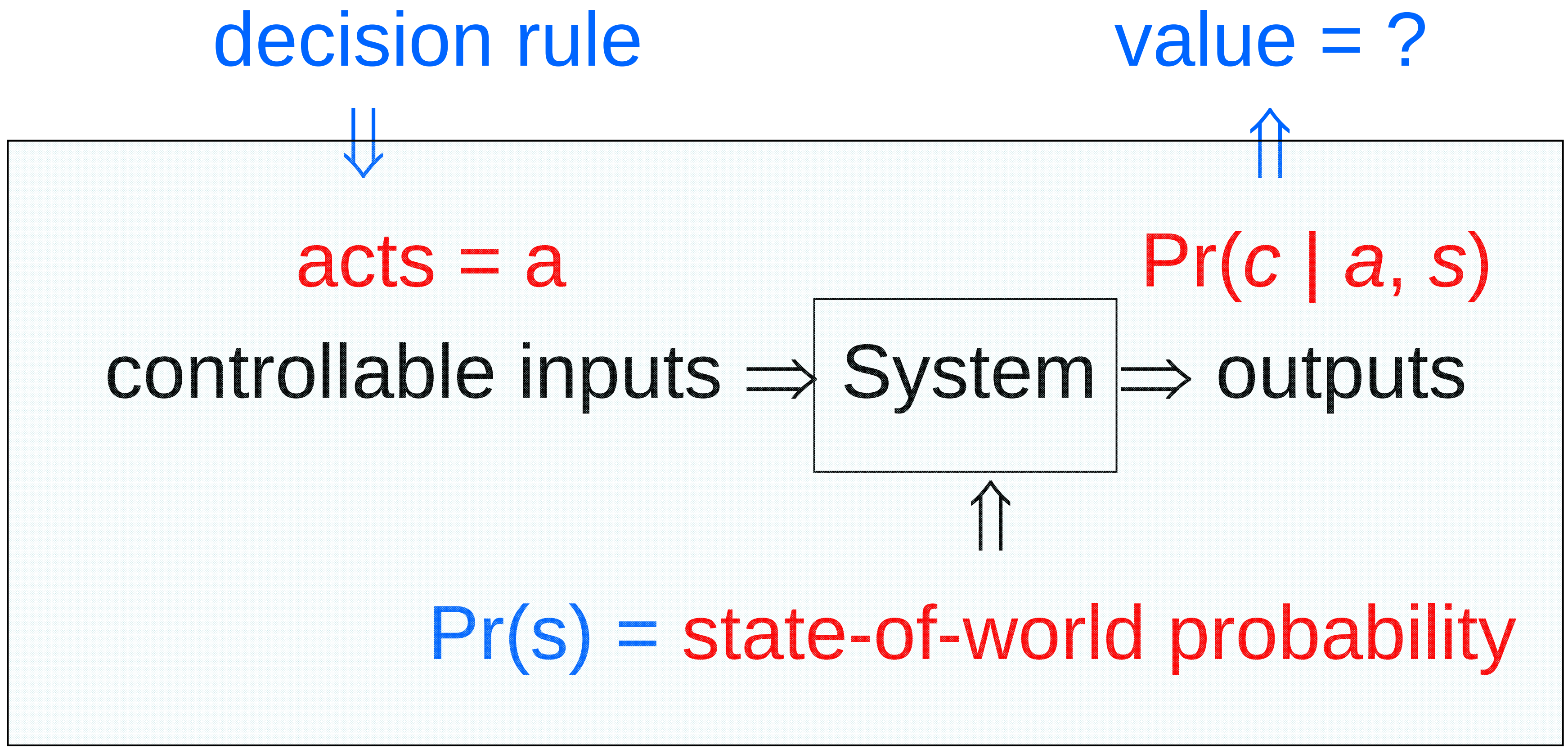


Lugosi *et al.* (2008), August *Math of OR*, consider imperfect monitorung.

2. Prediction and optimization



2. Prediction and optimization



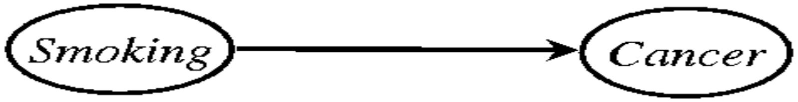
Bayesian Networks (BNs) and Influence Diagrams (IDs)

Smoker BN Example: Storing a PDF in a DAG

(*Source:* http://research.microsoft.com/~breese/)

Bayesian Networks

$S \in \{no, light, heavy\}$



$P(S{=}no)$	0.80
$P(S{=}light)$	0.15
$P(S{=}heavy)$	0.05

$C \in \{none, benign, malignant\}$

Smoking=	*no*	*light*	*heavy*
$P(C{=}none)$	0.96	0.88	0.60
$P(C{=}benign)$	0.03	0.08	0.25
$P(C{=}malig)$	0.01	0.04	0.15

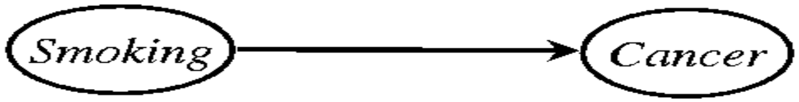
© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.

Smoker BN Example: Storing a PDF in a DAG

(*Source:* http://research.microsoft.com/~breese/)

Bayesian Networks

$S \in \{no, light, heavy\}$



$P(S=no)$	0.80
$P(S=light)$	0.15
$P(S=heavy)$	0.05

$C \in \{none, benign, malignant\}$

Smoking=	no	light	heavy
$P(C=none)$	0.96	0.88	0.60
$P(C=benign)$	0.03	0.08	0.25
$P(C=malig)$	0.01	0.04	0.15

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved.

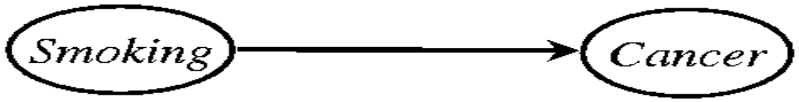
Probabilistic model of causation

Smoker BN Example: Storing a PDF in a DAG

(*Source:* http://research.microsoft.com/~breese/)

Bayesian Networks

$S \in \{no, light, heavy\}$



$C \in \{none, benign, malignant\}$

P(S=no)	0.80
P(S=light)	0.15
P(S=heavy)	0.05

Smoking=	no	light	heavy
P(C=none)	0.96	0.88	0.60
P(C=benign)	0.03	0.08	0.25
P(C=malig)	0.01	0.04	0.15

© 1997 Jack Breese, Microsoft Corporation and Daphne Koller, Stanford University. All rights reserved. 15

Real world: Get from classification trees

Solved Problem

Q: *What is the probability that someone is a heavy smoker (S = heavy), given that she is diagnosed with malignant lung cancer (C = malignant)?*

More generally, risk analysts (and litigators) often want to know: *What is the probability that plaintiff was exposed to / harmed by source s, given that adverse condition c has occurred?*

Manual Solution via Bayes' Rule

- *Q: What is the probability that someone is a heavy smoker (S = heavy), given that she is diagnosed with malignant lung cancer (C = malignant)?*
- A: $\Pr(s \mid c) = \Pr(c \mid s)\Pr(s) / [\Sigma_v \Pr(c \mid v)\Pr(v)]$
- Pr(heavy | malignant) =

  Pr(malignant | heavy)Pr(heavy)/Pr(malignant)

  = (0.15)*(0.05)/[(.01)(.8) + (.04)(.15) + (.15)(.05)]

  = 0.3488

Another way: Pr(s | c) = Pr(s, c)/Pr(c) = 0.0075/0.0215

= 0.3488

Solution via BN Solver

- DAG model: "State $\rightarrow$ Observation"
- Store marginal probabilities, Pr(*s*), at *input nodes* (having output arrows only)
- Store *conditional probability tables* (CPTs) at all other nodes.
- Make observations
- Enter query
  - Solver calculates *conditional* probabilities

Input requirements for BN

To fully *specify* (quantify) a BN, we need:

- **Prior probabilities** at all input nodes
- **Conditional probability tables** at all other nodes, for all possible combinations of values of their direct parents.
  - *Challenge:* How to learn BN structure and parameters (priors and conditionals) from data?
  - *Answer:* Use classification tree algorithms

To *use* the fully specified BN, we need:

- **Observations/evidence** (e.g., cancer type)

So what?

- A big advantage of Bayesian networks (BNs) is that they are as easy to solve with *multiple* nodes and observations as with only 2 nodes and 1 observation.
  - Gibbs sampling, MCMC
- BNs are closely related to *causal models*.
  - Show *conditional independence* relations.

    $X \rightarrow Y \rightarrow Z$ only if $\Pr(Z \mid Y, X) = \Pr(Z \mid Y)$

Applications of Bayes in RA

- *Bayes Rule:* From model $X \rightarrow Y$ to inference $\Pr(X \mid Y) = \Pr(\textit{input} \mid \textit{output})$
  - X = unobserved, Y = observed data
- **Diagnosis:** X = disease, Y = symptom
- **Hazard identification:** X = cause, Y = effect
- **Exposure assessment:** X = true exposures, Y = measured (e.g., sample) values
- **Dose-response modeling:** X = true dose-response model, Y = observed dose-response data
- **Risk characterization:** X = true model and exposure distribution (and risk), Y = data

Example of a BN for Risk Assessment

act → exposure → response → consequence → **QALYs**

↑ ↑ ↑

behavior → susceptibility → treatment

a → x → r → c → **Q** All CPTs and Pr(b) given.

↑ ↑ ↑

[b → s → m]

- Factor joint probability Pr(b, s, m, x, r, c, Q) given a:
  Pr(b)Pr(s | b)Pr(m | s)Pr(x | b, a)Pr(r | s, x)Pr(c | r, m)Pr(Q | c)
- BN technology lets us find the PDF of *any* unobserved variable(s) *given* observed values of any others
- Monte Carlo (MC) *composes* components into whole

Monte Carlo Prediction for X $\rightarrow$ Y

- Rule: Pr(Y = y) = Σ_xPr(Y = y | X = x)Pr(X = x)
- Suppose that Pr(Y = y | X = x) is hard to calculate but easy to simulate (e.g., it is a large, complex spreadsheet simulation model). What to do?

Answer: Sample x values from Pr(X = x).

  – For univariate X, partition unit interval into sub-intervals of lengths Pr(x), then use uniform U[0, 1] random number generator

- For each x, calculate/simulate a corresponding value of y from Pr(Y = y | X = x)
- Fraction of sampled Y values equal to y $\approx$ Pr(Y = y)
  – Adaptive importance sampling and variance reduction techniques (e.g., Latin hypercube sampling) improve the efficiency of this basic idea.

Gibbs Sampling for BNs

```
a → x   →   r  →  c  → Q      All CPTs and Pr(b) given
    ↑       ↑     ↑
    b   →   s  →  m
```

- Factor joint PDF for all variables:
  - Pr(b)Pr(s | b)Pr(m | s)Pr(x | b, a)Pr(r | s, x)Pr(c | r, m)Pr(Q | c)
- Gibbs sampling: Sample b from Pr(b), then s from Pr(s | b), then… then Q from Pr(Q | c)
- Repeat many times to obtain a random sample from the joint PDF!!! (Special case of MCMC)
- Can condition on observed values (with more or less efficiency) to obtain samples from posterior joint PDF of all unobserved values. (WinBUGS)

Remaining Challenges

- How to learn (and objectively validate) the causal *structure* of a BN from data?
  - *Answer:* Conditional independence tests
  - *Software:* Classification trees, BayesiaLab™, Bayesian Network Toolbox, etc.
- How to learn/estimate a BN's CPTs from data?
- Once these things are known, WinBUGS, Bayes Applet, etc. can be used to draw inferences *within* the BN about risk, exposure, etc.

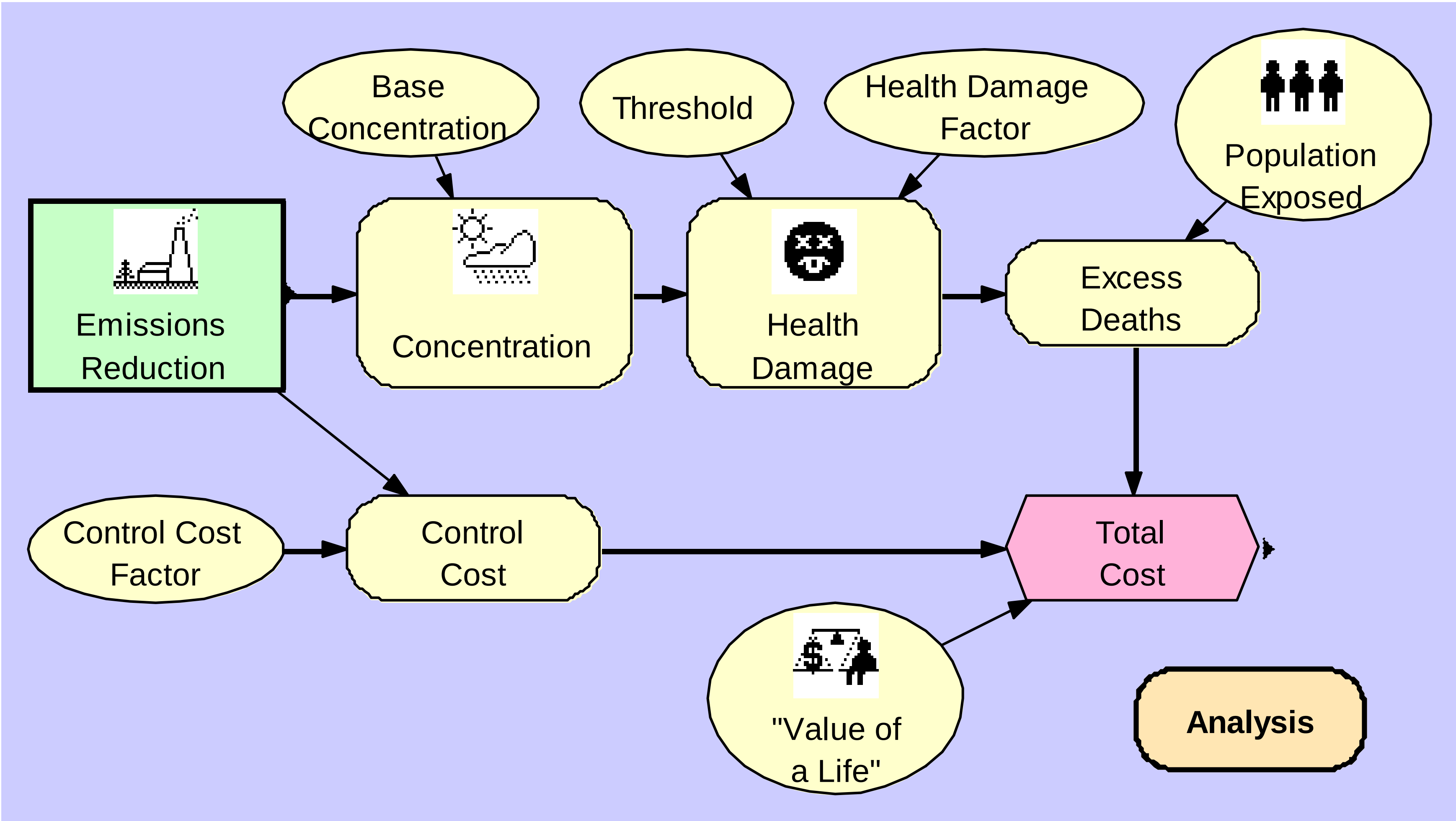
Example of a commercial ID program: *Analytica*

Influence diagrams and BNs

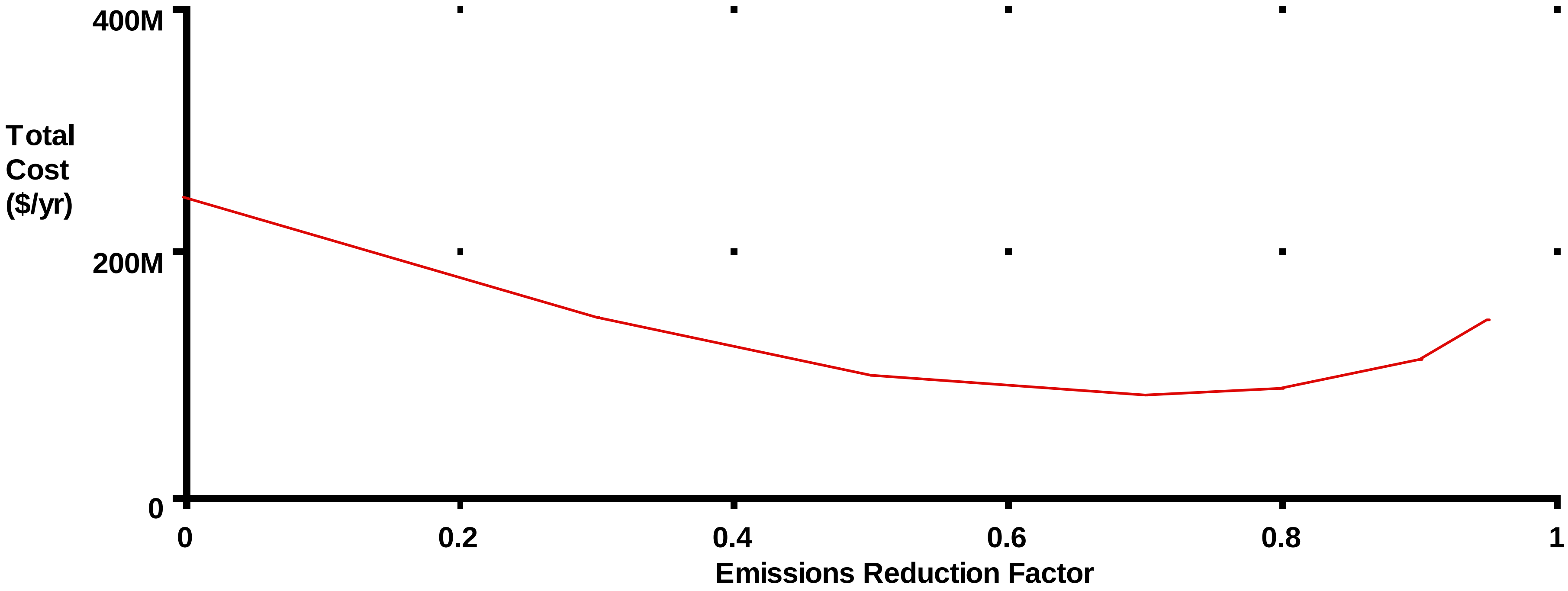
- *Influence diagram* software (e.g., Hugin, Analytica) allows models with decisions and value nodes to be quickly assembled and managed *if* the required CPTs and input probabilities are known.
- *Every influence diagram can be automatically converted to an equivalent Bayesian Network* and solved
  - Uniform distributions at decision nodes, utilities coded as probabilities at value nodes

EXAMPLE: *Analytica* Influence Diagram

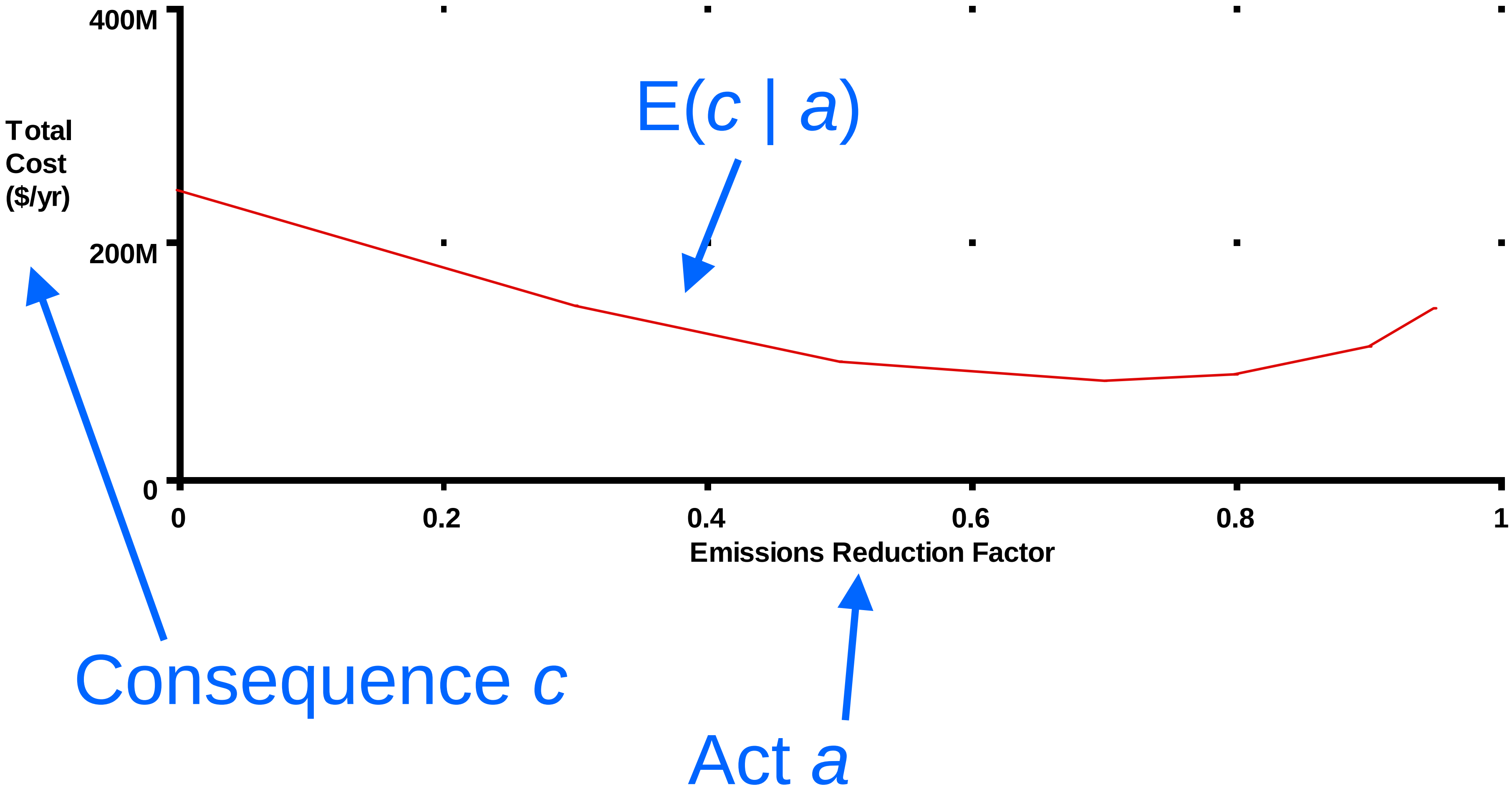
Description: Total cost to society = value of the lives saved - cost of Emissions Reduction required to save those lives.



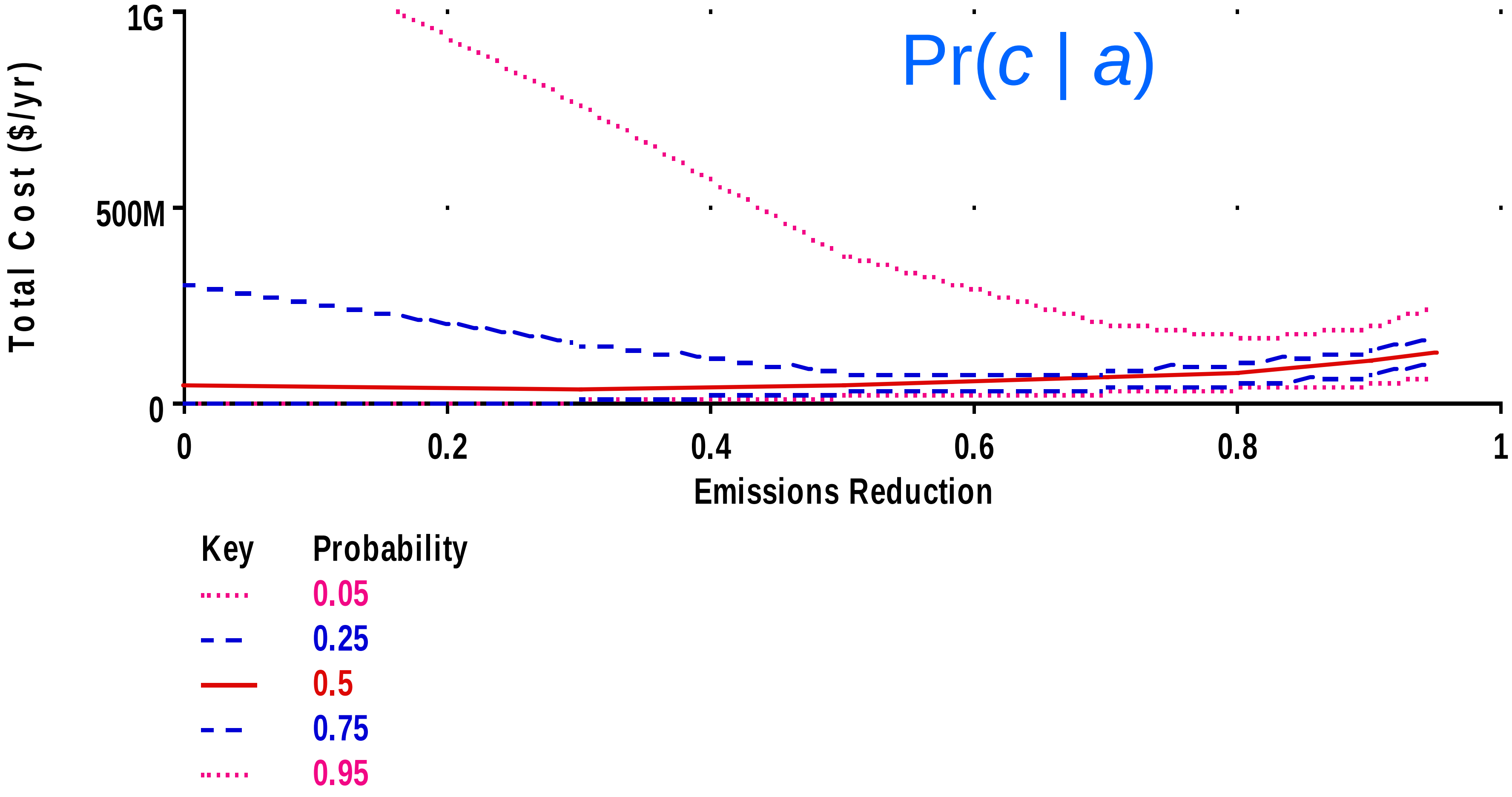
Clicking on “Total Cost” value node and selecting mean value yields:



Clicking on “Total Cost” value node and selecting mean value yields:



Selecting “Probability Bands” yields:



Example Analytica Model Wrap-Up

- Run time = ~1 sec.
  - Method = Forward Monte-Carlo propagation of input probability distributions through influence diagram.
- Conclusion: An emissions reduction factor of 0.7 is a robust optimum for this decision problem, given the probabilistic model (DAG influence diagram model) mapping inputs to output probabilities.

Relating different decision aids

- *Fault trees* can be converted to *influence diagrams*
- *Decision trees* can be converted to *influence diagrams*
- *Influence diagrams* can be converted to *Bayesian networks* (ID with no decisions)
  – Decision rules make choices chance nodes
- Bayesian networks (BN) can be solved!

Many ways to solve BNs

- Forward Monte-Carlo (Analytica)
- Bayesian Monte Carlo
  - Condition on observations
    - “Probabilistic logic sampling” (Henrion)
    - Gibbs sampling, WinBUGS
- Exact and approximate algorithms (Applet)
  - Junction trees, moral graphs, message-passing
  - Head-tail graph recursion, arc reversal
  - Variable elimination, bucket elimination
    - http://citeseer.ist.psu.edu/zhang98probabilistic.html;
    - http://www.cs.ubc.ca/~murphyk/Papers/gr03.pdf

Summary on IDs and BNs

- Good commercial-quality software is readily available for solving IDs/BNs.
  - Can use BN software to solve ID problems
- The main remaining problem is how to create valid causal models and populate them with data/estimates.
  - Learning models from data with classification trees and conditional independence tests
- Figuring out what to optimize is also a challenge in some applications.

Practical challenges for EU

Need:

- $\Pr(s)$ = state probabilities
- **$u(c)$ = utilities for consequences**
- $\Pr(c \mid a, s)$ = consequence model (causal)
  - At least need $\Pr(c \mid a)$

Utilities, $u(c)$

Obtaining $u(c)$

- Suppose that consequences, c, are measured on an *intrinsic value scale*.
  - Twice as many deaths, QALYs lost, etc. are counted as being twice as bad
  - "Measurable value" scale, Dyer, J. S., R. K. Sarin. 1979. Measurable multiattribute value functions. *Operations Res.* 27 (4) 810-822.
- Then we get an *exponential utility function*:

$$u(c) = 1 – \exp(-kc)$$

$k > 0$ for risk-averse decision-maker

k is coefficient of absolute risk-aversion

Certainty Equivalents

- *Certainty equivalent* of a prospect (act) X with normally distributed consequences, mean = μ, variance = σ^2, is:

$$CE(X) = \mu - (k/2)\sigma^2$$

  – CE is defined as: $u(CE(X)) = EU(X)$
    - For exponential utility, EU = moment generating fn.

  – *Every* risk-averse decision-maker prefers X to Y if X has higher mean and smaller variance than Y (and both are normally distributed)

  – “Efficient frontier” analysis $\rightarrow$ Don’t need k
    - Not true for some non-normal distributions

Application: Comparing uncertain consequence severities

Accidents *A*, *B*, and *C* have the same frequencies, but different consequence severity distributions:

- *A* has mean 1, variance 0: *A* ~ N(1, 0)
- *B* has mean 2, variance 2: *B* ~ N(2, 2)
- *C* has mean 3, variance 4: *C* ~ N(3, 4)

Rank A, B, and C by decreasing "severity".

Solution using CE(X) = $\mu - (k/2)\sigma^2$

- $A \sim N(1, 0)$, CE(A) = 1
- $B \sim N(2, 2)$, CE(B) = $2 - k$
- $C \sim N(3, 4)$, CE(C) = $3 - 2k$

Ranking based on CE depends on k:

For $k = 0$, ranking is $A < B < C$

For $k = 1$, ranking is $A = B = C$

For $k = 2$, ranking is $A > B > C$

Consequence rating is subjective.

Solution using CE(X) = $\mu - (k/2)\sigma^2$

- $A \sim N(1, 0)$, CE(A) = 1
- $B \sim N(2, 2)$, CE(B) = $2 - k$
- $C \sim N(3, 4)$, CE(C) = $3 - 2k$

Ranking based on CE depends on k:

For $k = 0$, ranking is $A < B < C$

For $k = 1$, ranking is $A = B = C$

For $k = 2$, ranking is $A > B > C$

"Severity" ratings mix objective (μ, σ^2) and subjective (k) information.

Implication

- It is not clear what severity rankings (or ratings, e.g., as “High”, “Medium”, “Low”) mean when consequences are uncertain.
  - How much of a rating is driven by objective information?
  - How much is driven by subjective (unstated) risk attitude?
  - How different could other rankings be?
- Objective part: Curve of CE *vs. k*

Application: Risk scores *vs.* optimal portfolios

- Suppose that there are three types of risk-reducing opportunities, with the following uncertain values:
  - Type A ~ N(1, 1) CE(A) = 1 – (k/2)
  - Type B ~ N(1, 2) CE(B) = 1 - k
  - Type C ~ N(1, 2) CE(C) = 1 - k
- If we can afford 2 interventions, what type(s) should we buy?

Application: Risk scores *vs.* optimal portfolios

- Suppose that there are three types of risk-reducing opportunities, with the following uncertain values:
  - Type A ~ N(1, 1) CE(A) = 1 – (k/2)
  - Type B ~ N(1, 2) CE(B) = 1 - k
  - Type C ~ N(1, 2) CE(C) = 1 - k
- If we can afford 2 interventions, what type(s) should we buy?
- *Risk score:* Buy two Type As

Application: Risk scores *vs.* optimal portfolios

- Suppose that there are three types of risk-reducing opportunities, with the following uncertain values:
  - Type A ~ N(1, 1) CE(A) = 1 – (k/2)
  - Type B ~ N(1, 2) CE(B) = 1 - k
  - Type C ~ N(1, 2) CE(C) = 1 - k
- If we can afford 2 interventions, what type(s) should we buy?
- *Risk portfolio:* It depends…

Application: Risk scores *vs.* optimal portfolios

- Suppose that there are three types of risk-reducing opportunities, with the following uncertain values:
  - Type A ~ N(1, 1) CE(A) = 1 – (k/2)
  - Type B ~ N(1, 2) CE(B) = 1 - k
  - Type C ~ N(1, 2) CE(C) = 1 - k
- Suppose that the As are i.i.d., but Bs and Cs can be paired, with C = 2 – B.
- *Risk portfolio:* Then buy 1 B and 1 C.

Application: Risk scores *vs.* optimal portfolios

- Suppose that there are three types of risk-reducing opportunities, with the following uncertain values:
  - Type A ~ N(1, 1) CE(A) = 1 – (k/2)
  - Type B ~ N(1, 2) CE(B) = 1 - k
  - Type C ~ N(1, 2) CE(C) = 1 - k
- With C = 2 – B, buying a BC pair yields N(1, 0), CE = 1 $\rightarrow$ Greater CE value!
- *Risk portfolio:* Then buy 1 B and 1 C.

Implication

- Scoring one item at a time, without considering correlations among uncertain values, may overlook risk-free returns, or other opportunities to improve risk-reducing value obtained for resources spent.
  - Hazardous waste sites, web server upgrades, terrorism countermeasures, customers, etc.
- Optimal risk management often requires considering *portfolios* of interventions.

Multiattribute Utility Theory (MAUT)

- Suppose that $c = (c_1, c_2, \ldots, c_n)$ is a vector of consequence attributes (e.g., lives, dollars, environmental impacts in different time periods, etc.)
- Then use indifference curves or other value trade-off models to assess $v(c)$, the value of consequence vector c.
- If (and only if) *preferential independence* conditions hold, simple formulas result.

Multiattribute Utility Theory (MAUT)

- $c = (c_1, c_2, \ldots, c_n)$
- *Additive independence* $\rightarrow v(c) = \Sigma_i v_i c_i$.
- *Mutual preferential independence*, weak difference independence, etc. $\rightarrow v(c) =$ product of terms $(1 + v_i c_i)$. (Dyer, Sarin 1979)
  - SMART and other practical methods assess and use simple value functions.
- Treat value, *v*, as a single attribute!
  - $u(c) = 1 - \exp(-kv(c))$

Importance of independence conditions in practice

Avoid foolish risk scores:

- RAMCAP™ gives the same score to:
  - 100% probability of 0 fatalities (5 + 0 = 5)
  - 20% probability of 100 fatalities (3 + 2 = 5)
- Bioterrorism risk priority scoring system gives the same score to:
  - Preventable but untreatable disease (0 + 2)
  - Unpreventable but treatable disease (2 + 0)
    - MacIntyre CR, Seccull A, Lane JM, Plant A. Development of a risk-priority score for category A bioterrorism agents as an aid for public health policy. *Mil Med.* 2006 Jul;171(7):589-94.

Conclusion on MAUT

- Although multicriterion decision making (MCDM) methods for iterative elimination of dominated choices and elicitation of local value trade-offs can be very useful in some practical applications…
- … we will just reduce multiattribute problems to single attribute problems.
  - The single attribute is “measurable value”
  - Export value trade-offs to value modelers

Empirical evidence on EU

- Coherent preferences (and hence utilities) for consequences often do not exist
  - "99% survival" vs. "1% mortality"
  - Framing effects, endowment effects, presentation effects, elicitation effects
  - Heuristics and biases
- Even when clear preferences do exist for consequences, many real decisions and preferences violate EU theory

Example: Certainty Effect

- Many people prefer $3000 with certainty to an 80% chance at $4000 (else $0).
- Many people also prefer a 20% chance at $4000 (else nothing) to a 25% chance at $3000 (else nothing).
- *Pop Quiz:* Is it consistent with EU theory ("rational") to hold both preferences?

Example: Certainty Effect

- Many people prefer \$3000 with certainty to an 80% chance at \$4000 (else \$0).
- Many people also prefer a 20% chance at \$4000 (else nothing) to a 25% chance at \$3000 (else nothing).
- *Pop Quiz:* Is it consistent with EU theory ("rational") to hold both preferences?
- *Answer:* No. If $u(3000) > 0.8u(4000)$, then not $0.25u(3000) < 0.2u(4000)$.

Empirical evidence on EU

- Coherent preferences (and hence utilities) for consequences often do not exist
  - “99% survival” vs. “1% mortality”
  - Framing effects, endowment effects, presentation effects, elicitation effects
  - heuristics and biases
- Even when clear preferences do exist for consequences, many real decisions violate EU theory

Empirical evidence on EU

- Coherent preferences (and hence utilities) for consequences often do not exist
  - “99% survival” vs. “1% mortality”
  - Framing effects, endowment effects, presentation effects, elicitation effects
  - heuristics and biases
- Even when clear preferences do exist for consequences, many real decisions violate EU theory
- …which may be why we need EU theory!

Empirical evidence on EU

- Coherent preferences (and hence utilities) for consequences often do not exist
  - “99% survival” vs. “1% mortality”
  - Framing effects, endowment effects, presentation effects, elicitation effects
  - heuristics and biases
- Even when clear preferences do exist for consequences, many real decisions violate EU theory
- …or, maybe we need a better theory?

Some alternatives to EU

- Normative theories with weaker axioms
  - Weaken (aRb) $\rightarrow$ (a,p,c)R(b,p,c) $\forall$ a, b, c, p
  - Generalized utility theory (Machina)
  - Lottery-dependent utilities
  - Rank-dependent expected utility (RDEU)
- More descriptive theories
  - Prospect theory/cumulative prospect theory
    - Probability weighting function, separate evaluation of gains and losses (“mental accounting”)
  - Many semi-empirical theories (Luce, Weber)

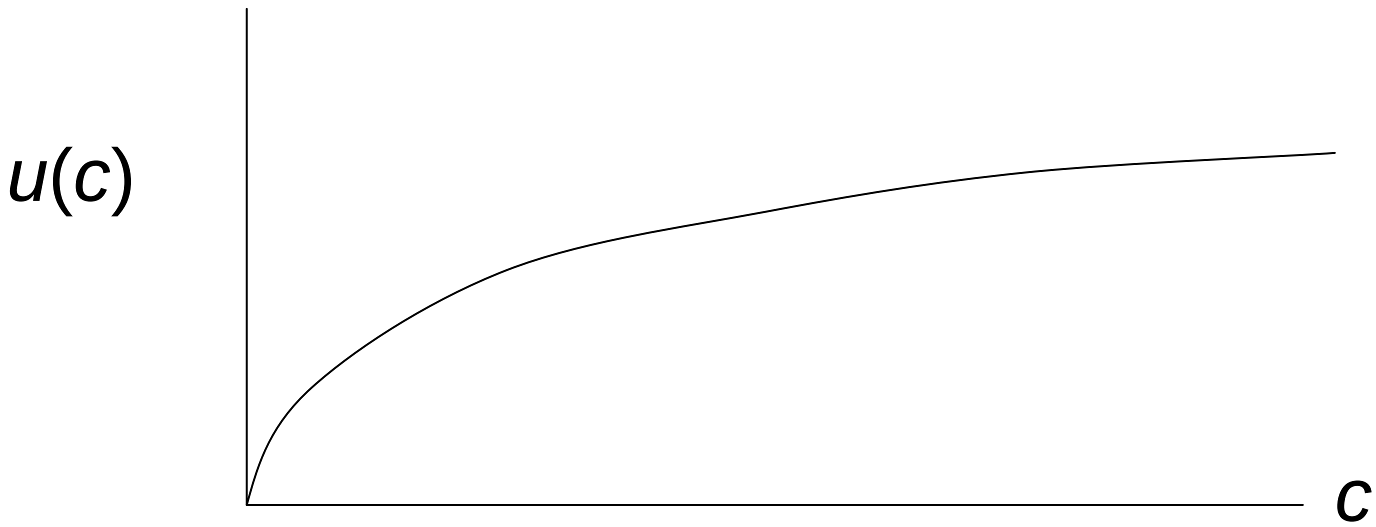
Living with a subset of EU

- *First-order stochastic dominance (FSD):*

  If $\Pr(X \geq c) > \Pr(Y \geq c)$ for *every* value of *c*, then should prefer *X* to *Y* (if *c* is desirable).
  - No utility function or CEs needed!
  - *Example:* For Poisson processes,10 expected deaths/yr. is worse than 5.
- *SSD:* Every risk-averse decision-maker should prefer normal distributions with higher means & smaller variances for gains.

So, what’s left?

- *Risk-aversion* is still useful



- Implies diversification in many applications

Some applications of EU to risk scoring and priority ramkings

Example: Hazardous Waste Site Cleanup

- Two types of sites, A and B; many of each
  - Type A is mostly “type 1” asbestos
  - Type B is mostly “type 2” asbestos
    - Cleaning up a *mix* of both types reduces lung cancer & mesothelioma risks around site
    - Could be that A is dangerous, or B, or both.
  - For simplicity, suppose potencies for (A, B) are either (0, 1) or (1, 0) – we don’t know which.
- What to do? How to allocate clean-up resources between type A and type B sites?

Example: Hazardous Waste Site Cleanup

- Two types of sites, A and B; many of each
  - For simplicity, suppose potencies for (A, B) are (0, 1) or (1, 0) – we don’t know which.
- Suppose we can afford to clean $N = 10$ sites (all equally costly).
- Expected utility from cleaning x type A sites and $N - x$ type B sites is:

$$EU(x) = [0.5*0 + 0.5*u(N - x)] + [0.5*u(x) + 0]$$

$$= [u(N - x) + u(x)]/2$$

Example: Hazardous Waste Site Cleanup

- Two types of sites, A and B; many of each
  - For simplicity, suppose potencies for (A, B) are (0, 1) or (1, 0) – we don't know which.
- Expected utility from cleaning x type A sites and $N - x$ type B sites is:

$$EU(x) = [u(N - x) + u(x)]/2$$

- This is maximized when $EU'(x) = 0$

$$du(N - x)/d(N - x) = du/dx \rightarrow N - x = x \rightarrow$$

$$x^* = N/2.$$

Example: Hazardous Waste Site Cleanup

- Two types of sites, A and B; many of each
  - For simplicity, suppose potencies for (A, B) are (0, 1) or (1, 0) – we don't know which.
- EU from cleaning x type A sites and $N - x$ type B sites is maximized for $x^* = N/2$.
- Example: $u(x) = \sqrt{x}$

x:	1	2	5*	8	10
$u(x)$:	1	1.4	2.2	2.8	3.2
$u(10 - x)$:	3	2.8	2.2	1.4	0
Total :	4	4.2	**4.4**	4.2	3.2

Example: Hazardous Waste Site Cleanup

- Two types of sites, A and B; many of each
  - – For simplicity, suppose potencies for (A, B) are (0, 1) or (1, 0) – we don't know which.
- EU from cleaning x type A sites and $N - x$ type B sites is maximized for $x^* = N/2$.
- *<u>Any</u> priority ordering of A and B is bad!*

x:	1	2	5*	8	10
$u(x)$:	1	1.4	2.2	2.8	3.2
$u(10 - x)$:	3	2.8	2.2	1.4	0
Total :	4	4.2	**4.4**	4.2	**3.2**

Implication

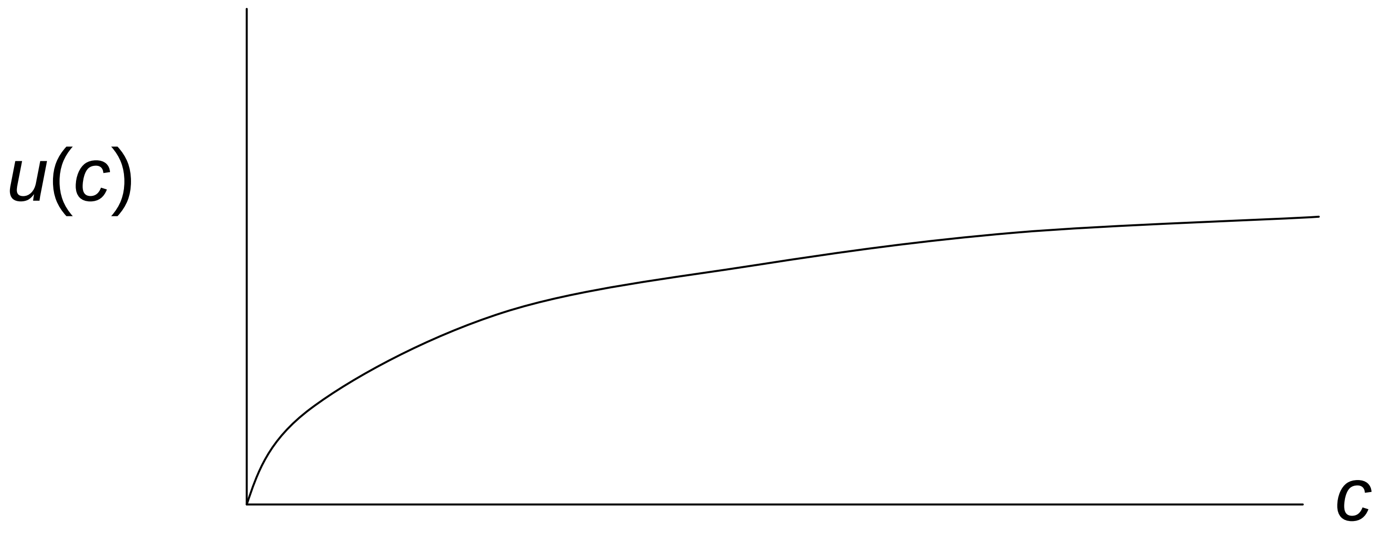
- Risk-scoring, ranking, and rating systems can *increase* risks!

Implication

- Risk-scoring, ranking, and rating systems *increase* risks by encouraging consistent investment in one type of intervention
  - This can *maximize risks* and *minimize value* of risk-reduction investments.
- Diversification is essential for effective risk reduction when effects of interventions are uncertain.

So, what's left?

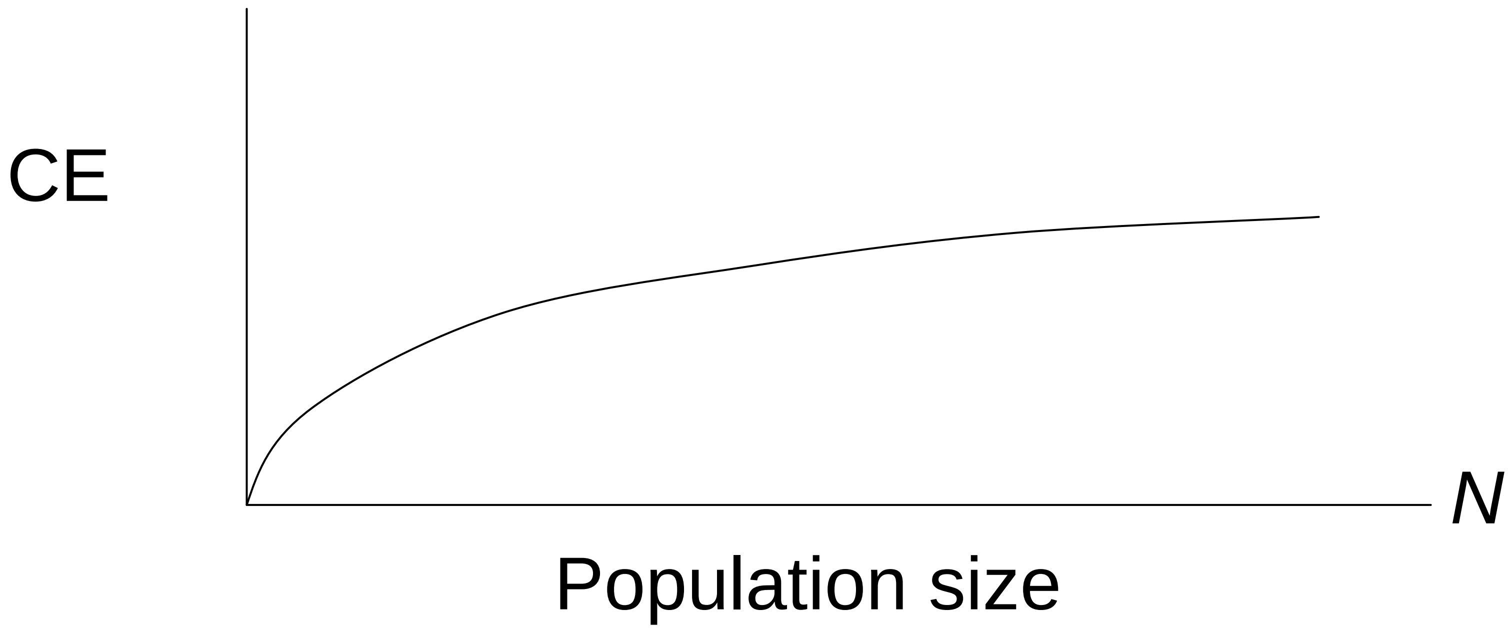
- *Risk-aversion* is still useful



- Implies diversification, in many applications

So, what's left?

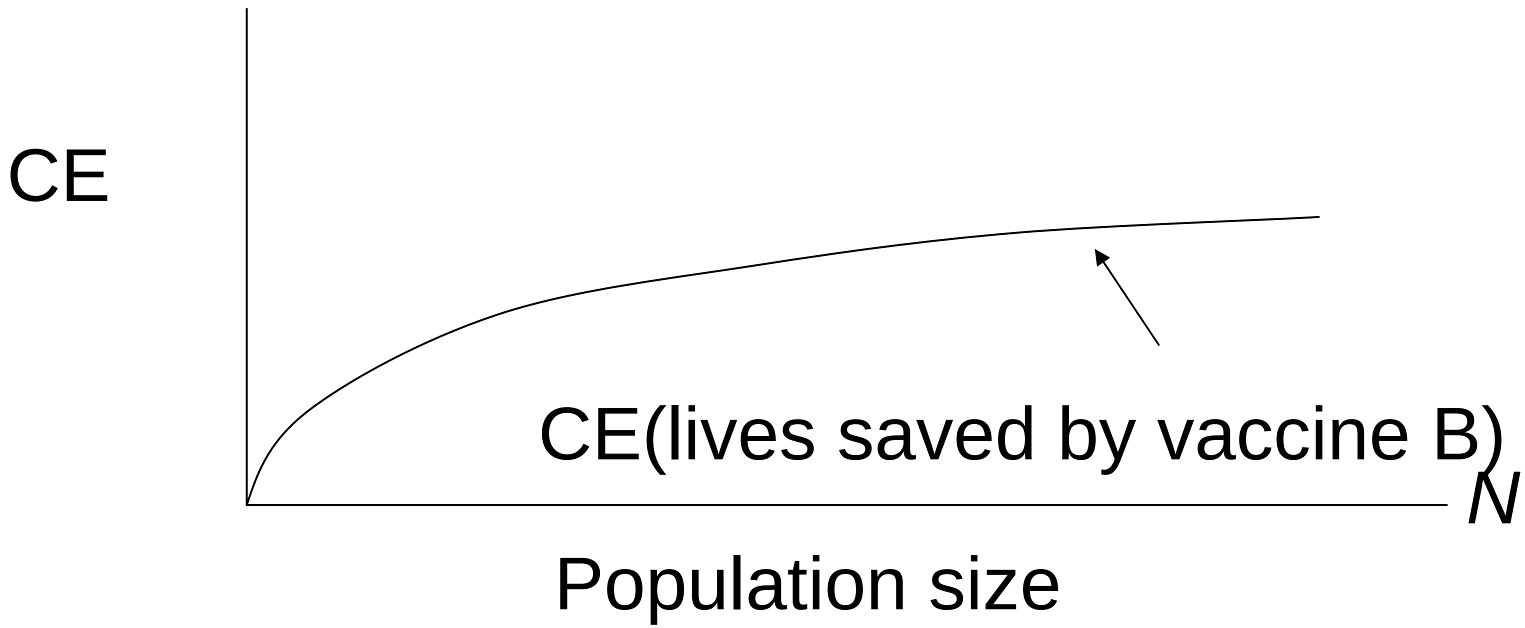
- *Risk-aversion* is still useful



- Implies aversion to excessively large bets on interventions with uncertain effects

Example

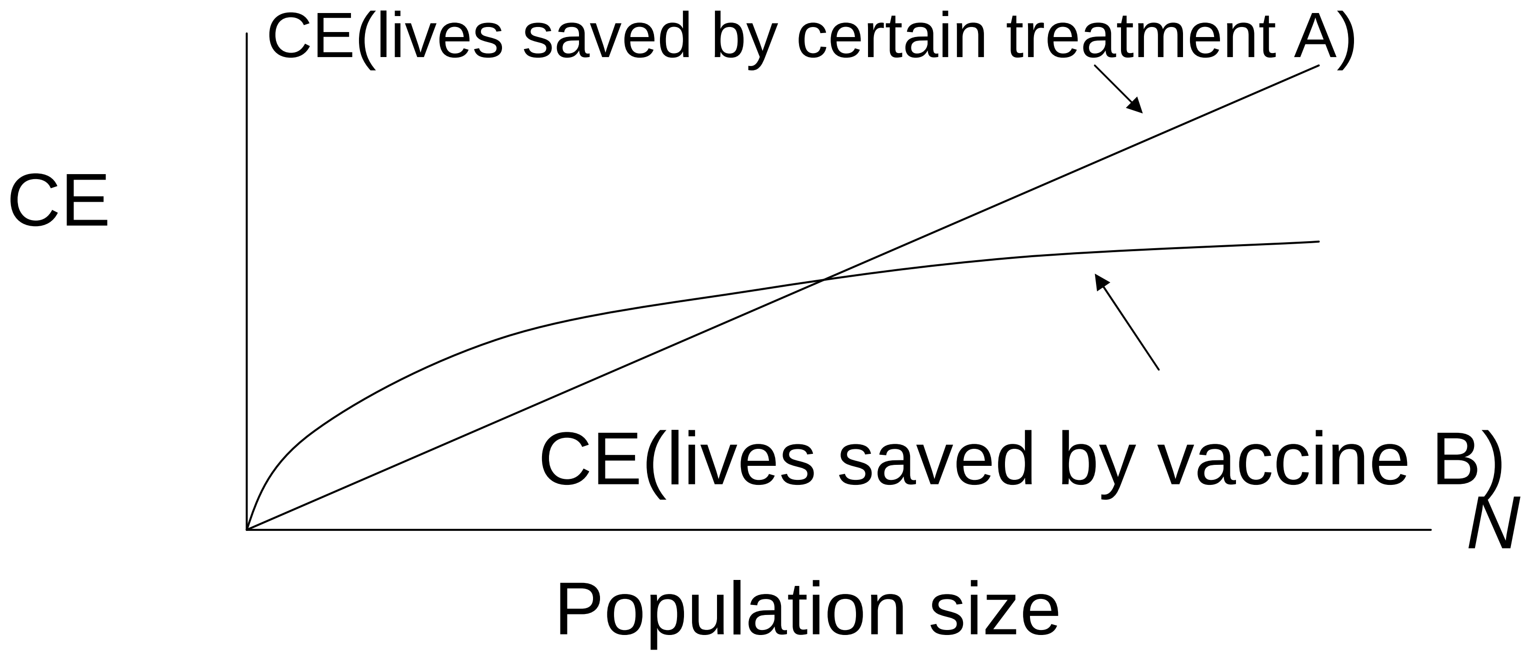
- *Risk-aversion* is still useful



- Implies aversion to excessively large bets on interventions with uncertain effects

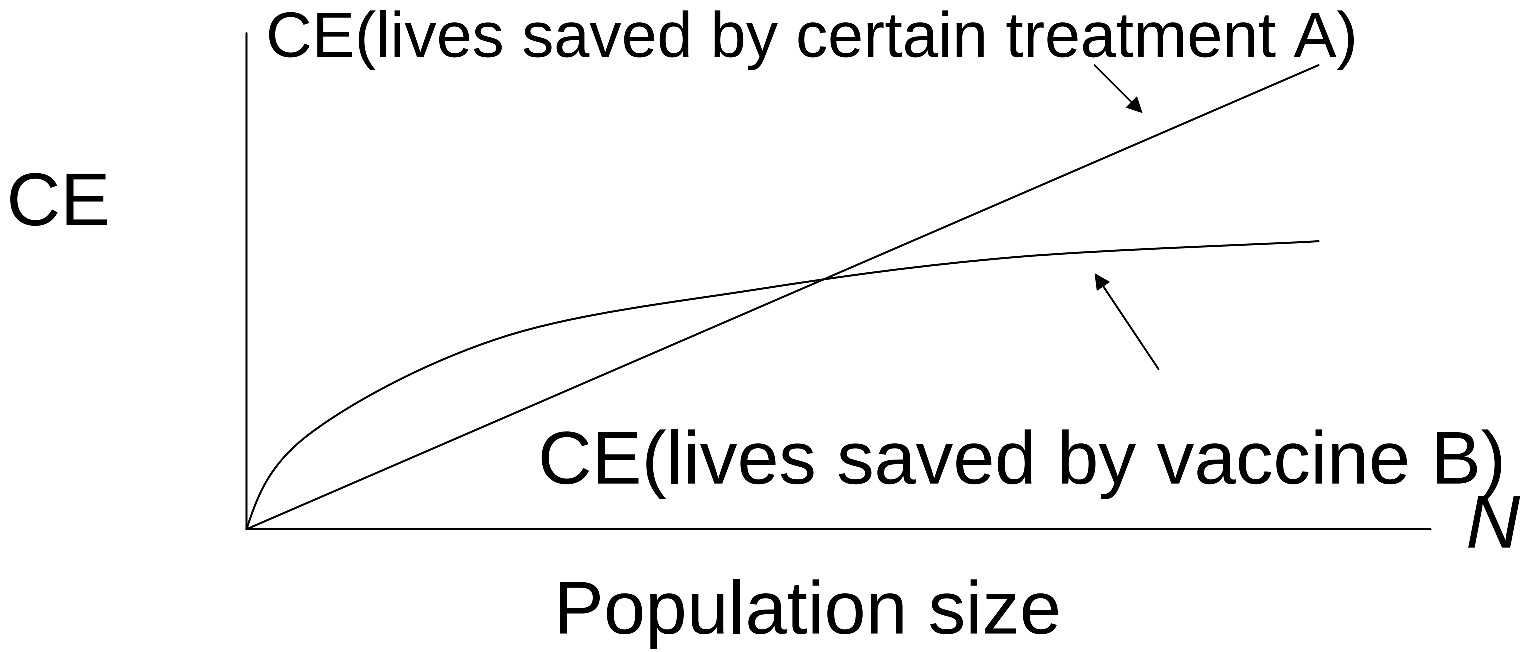
Example

- *Risk-aversion* is still useful



Example

- *Risk-aversion* is still useful



- Best intervention for small *N* may not be best bet for larger *N*.

Simple Example

- Treatment A reduces mortality risk of disease from 0.2 to 0.15.
- Vaccine B either reduces mortality risk from 0.2 to 0 (if it works) or leaves it unchanged at 0.2 (if it does not work). The probability that it works is 0.5.
- *Which is preferable, A or B?*

Simple Example

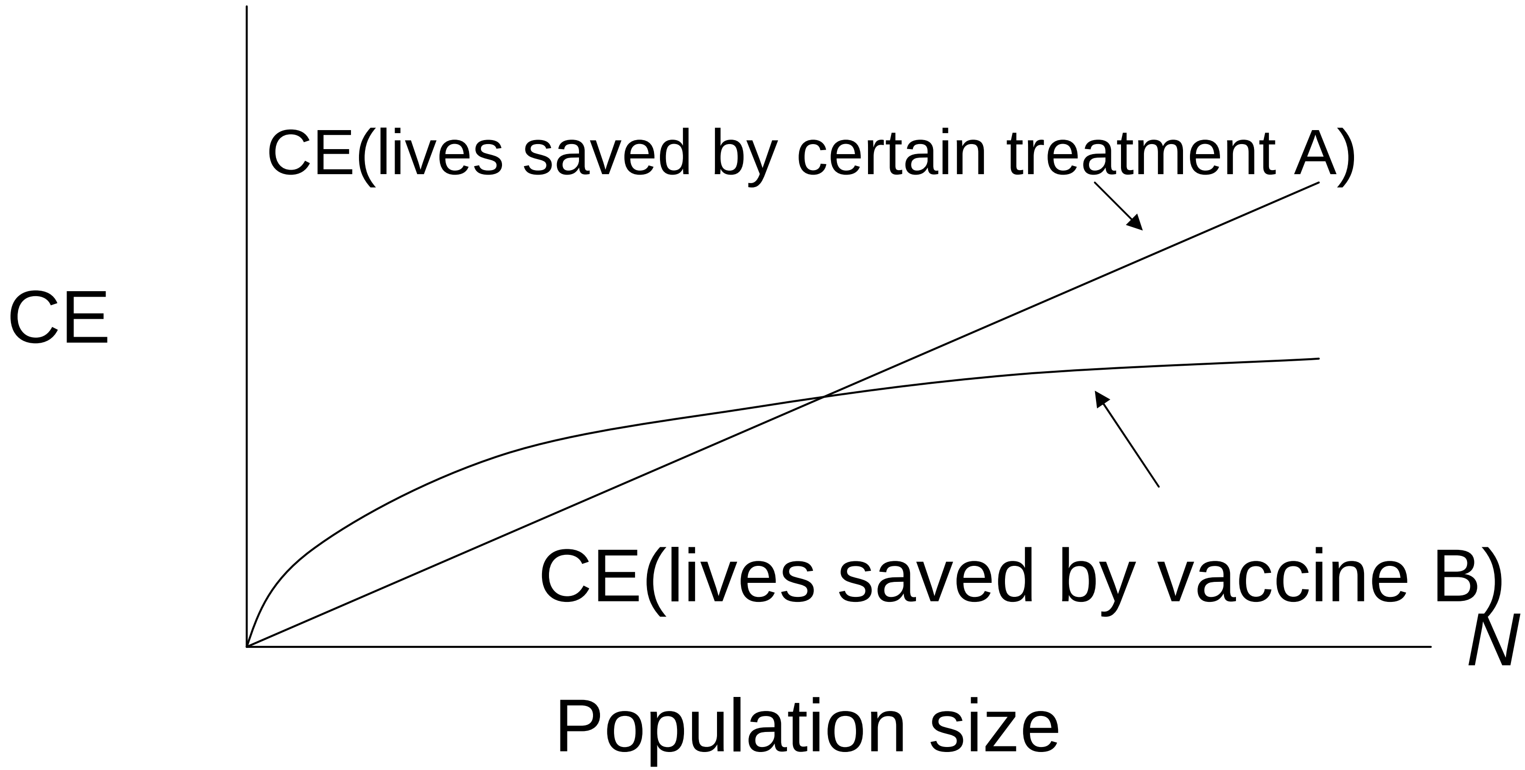
- A reduces mortality risk from 0.2 to 0.15.
- B either reduces mortality risk from 0.2 to 0 or leaves it unchanged at 0.2 (if it does not work). The probability that it works is 0.5.

• *Which is preferable, A or B?*

• Answer depends on *N,* number of subjects

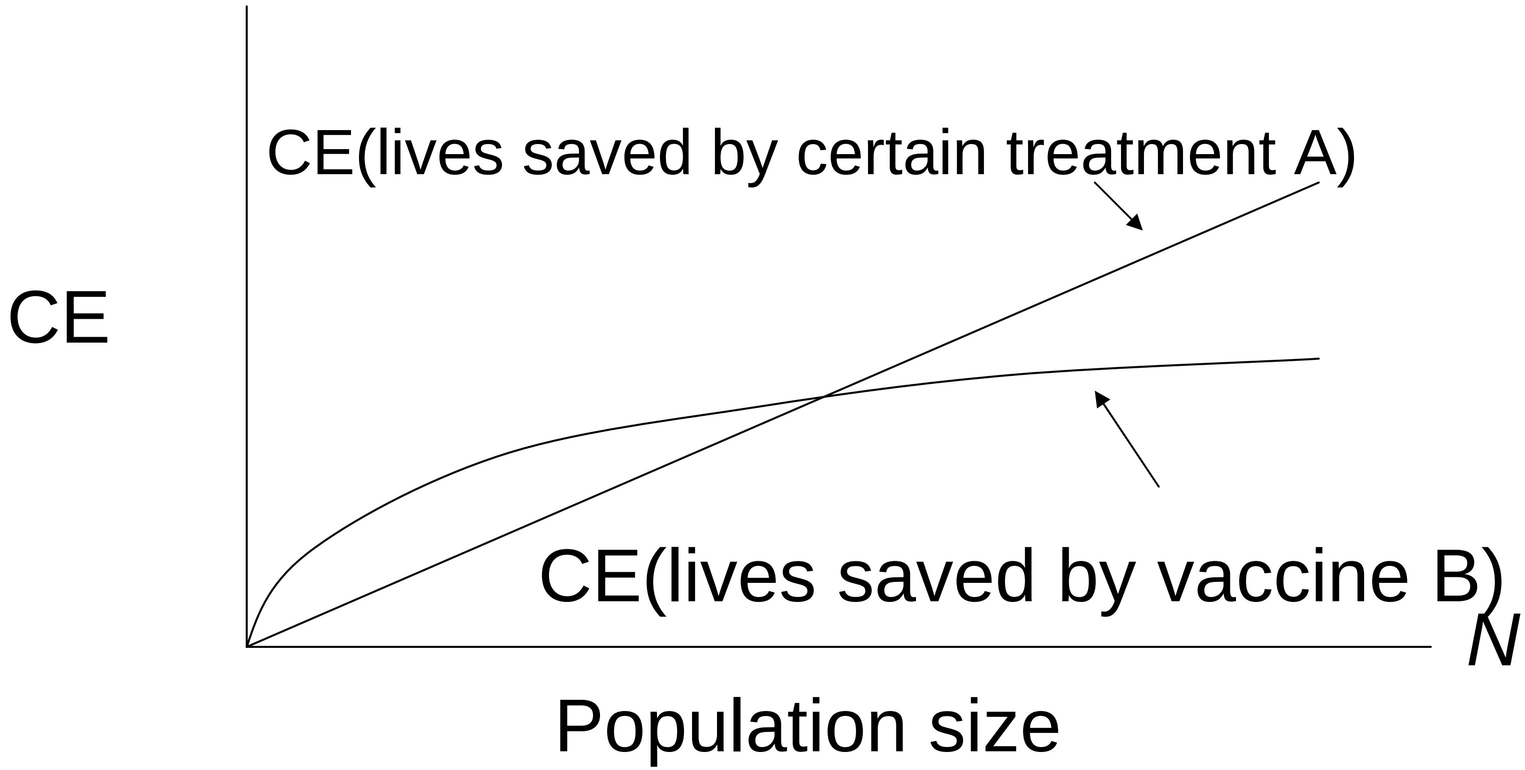
- For *N* = 1 individual, B reduces mortality risk to 0.5(0 + 0.2) = 0.1. *B is preferable to A.*
- For N = 1,000,000 individuals, *A may be preferable to B.*

Implication



- Best intervention depends on *N*.

Implication



- Applying risk scores one case at a time may lead to large risk exposures!

Summary on risk scoring

- Myopic or distributed risk management based on scoring risk management acts for each of many (*N*) cases, and choosing the “best” (top-scoring) act in each case, may lead to unintentionally large risk exposures.
- To do better, the whole portfolio of correlated risks must be evaluated.

Ranking interventions

Challenge: Prioritize 3 risk-reducing options.

- A reduces risk from 100 to 80, costs $30.
- B reduces risk from 50 to 10, costs $40.
- C reduces risk from 25 to 0, costs $20.

Q: What priorities maximize risk reduction?

Ranking interventions

Challenge: Prioritize 3 risk-reducing options.

- A reduces risk from 100 to 80, costs $30.
- B reduces risk from 50 to 10, costs $40.
- C reduces risk from 25 to 0, costs $20.

Q: What priorities maximize risk reduction?

Answer depends completely on budget!

- If budget = $49: Do B only (top priority).
- If budget = $50: Do all but B (bottom priority).

Implication

- *No* priority ranking (or scoring) of risk-reducing options can allocate resources effectively without considering budget (and dependencies).
- Effective resource allocation requires solving combinatorial optimization (capital budgeting) problem.
  - No way to do this in scoring systems

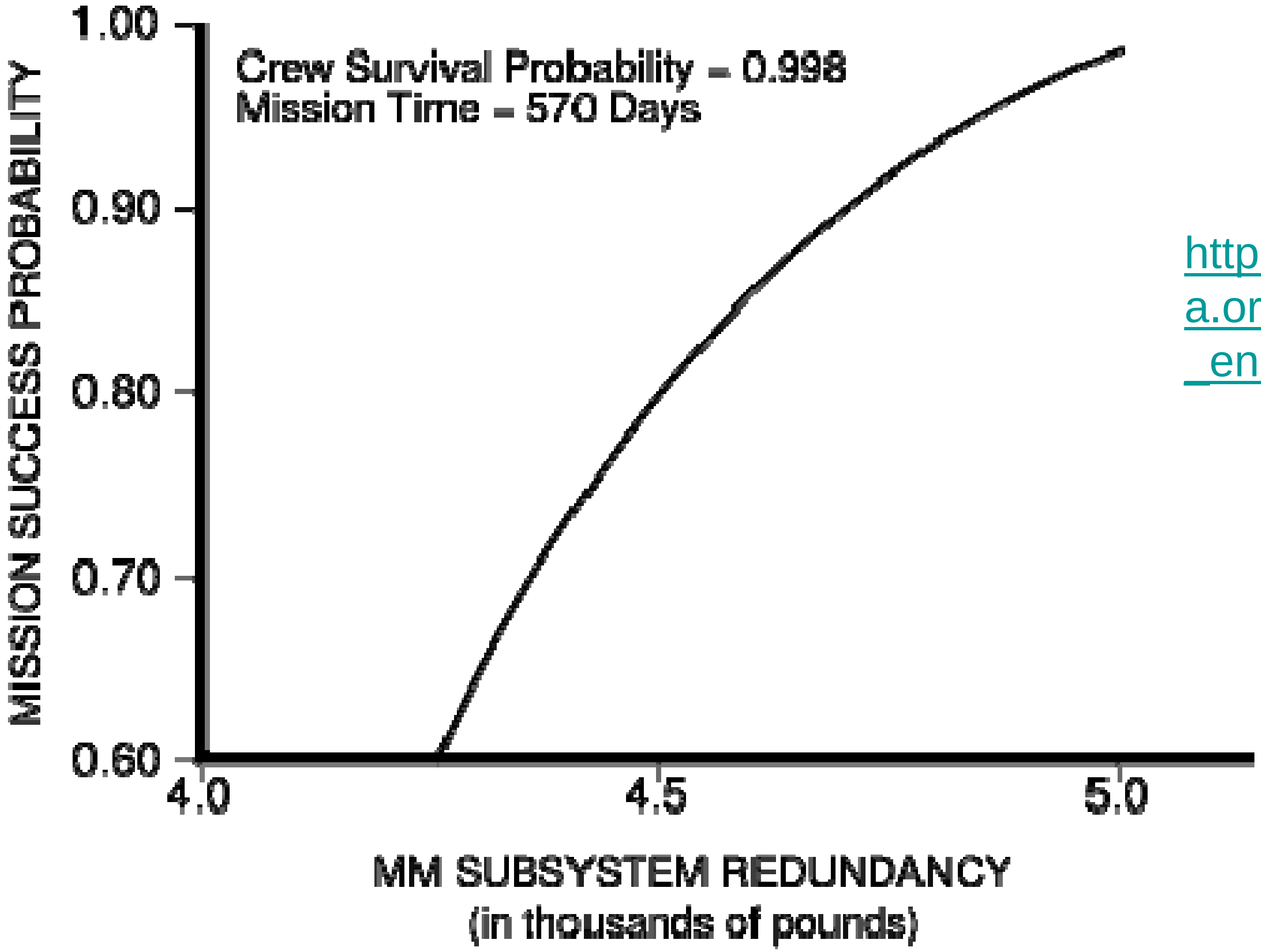
Summary on risk ranking systems

- Don’t use them!
  – Even though they are popular, incorporated into standards and regulations, etc.
- Use subset (or portfolio) *optimization* instead
  – Achieves larger risk-reduction value for same resources spent

Managing risk without utilities

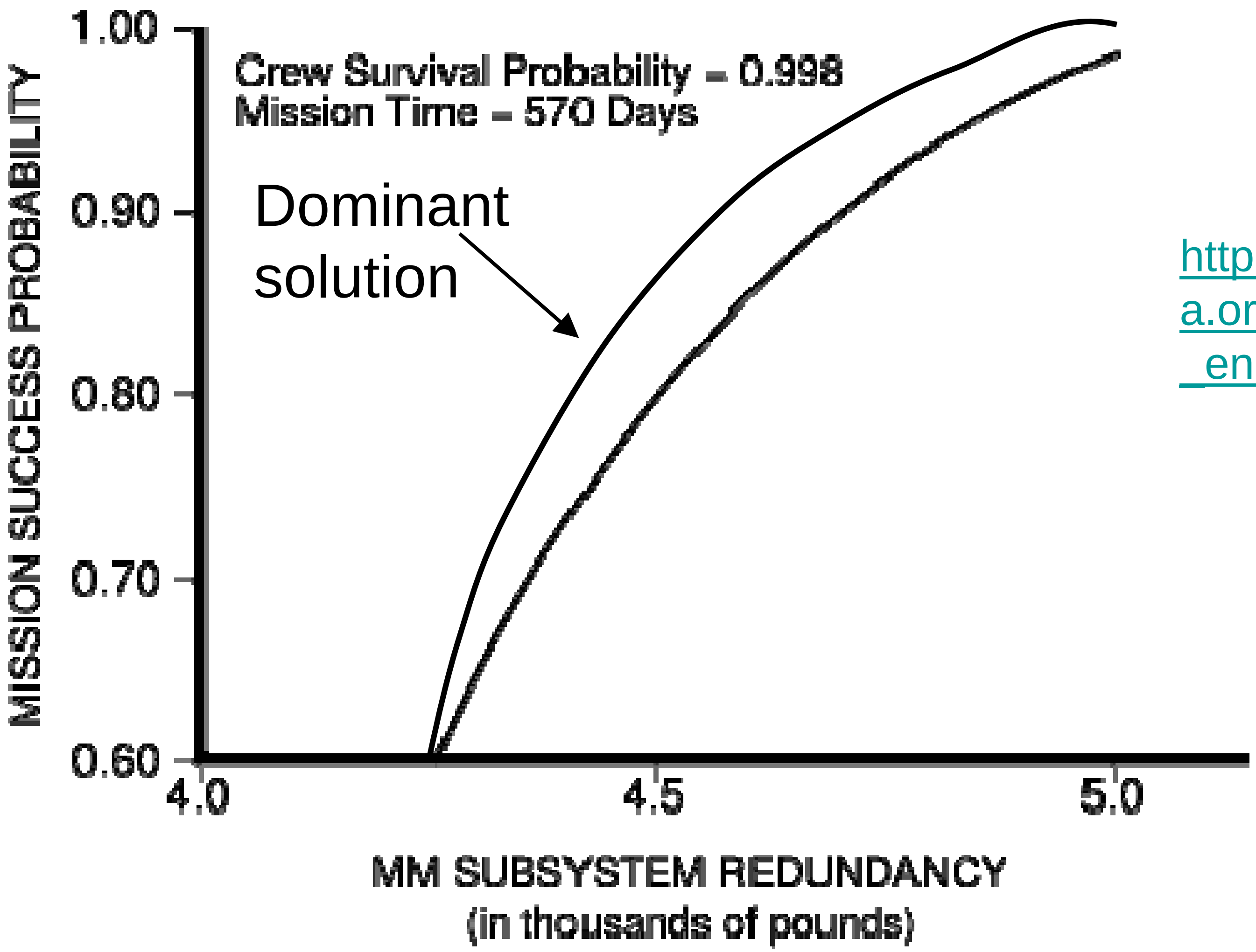
- Although detailed value and utility models cannot necessarily be justified (because coherent preferences may not exist), finding *dominant solutions* is practical.
- All risk-averse decision-makers should prefer certain risk management decisions
  – Diversify
  – Avoid excessively large bets
  – Exploit negative correlations among outcomes

Living without *u*(*c*) using FSD



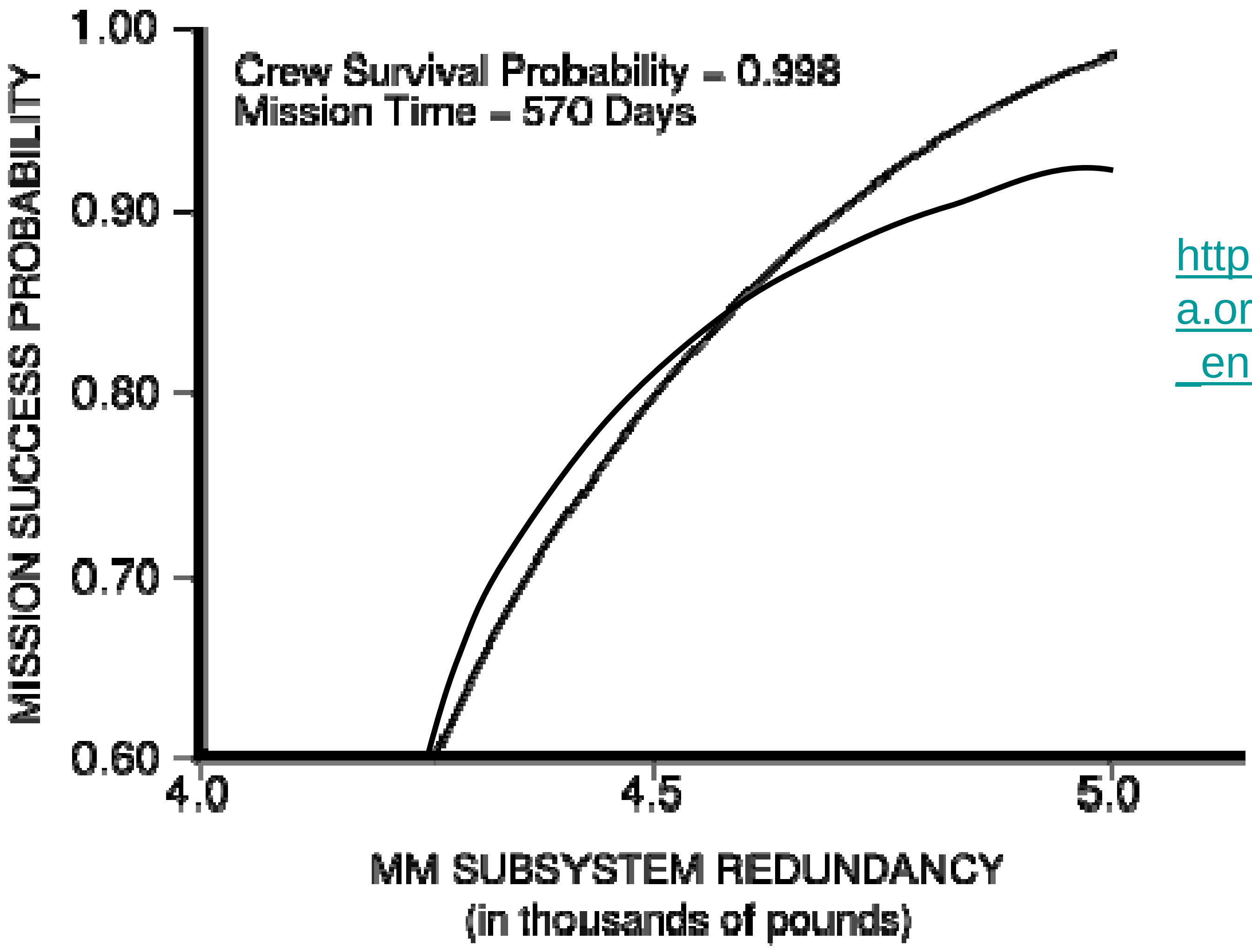
http://en.wikipedia.org/wiki/Safety_engineering

Living without *u*(*c*) using FSD



http://en.wikipedia.org/wiki/Safety_engineering

Living without *u*(*c*) using FSD



http://en.wikipedia.org/wiki/Safety_engineering

Conclusions

Course Summary

- For most risk management decisions, all we really want (and can trust) is:

  $$\Pr(c \mid a) = \Pr(\textit{consequence} \mid \textit{act})$$

  – Eliminate Pr(s) by marginalizing out

  – Eliminate $u(c)$ using FSD

- Quantitative risk assessment depends on credible causal models of $\Pr(c \mid a)$

  – Learn from data if possible

  – Use knowledge, expertise, BMA if necessary

  – Or, adaptively optimize $\Pr(a \mid \textit{information})$

Course Summary

Techniques:

- Decision analysis (DA)
  - Risk profiles, F-N curves
  - Expected utility theory, stochastic dominance
- Causal modeling
  - Fault trees, decision trees, influence diagrams
  - Bayesian networks, conditional independence
    - Gibbs sampling, classification trees, applets
- Optimization
  - Reinforcement learning, on-line algorithms

Course Summary

- Even rough approximate quantitative models Pr(*c* | *a*) greatly improve decisions (usually)
  - Even fitting quantitative models to one's own judgments can improve decisions!
- Quantitative causal modeling, even if very imperfect, typically improves decisions
  - Reduces confirmation bias and other biases
  - Makes better use of relevant information

Course Summary

- Quantitative risk assessment modeling is practical now, even for complex and uncertain systems…
- … and is too valuable not to use!

Thanks!