CMSC 471
Fall 2012

Class #20

Tuesday, November 6, 2012

Partial Order Planning and Probabilistic Reasoning

Kevin Winner, winnerk1@umbc.edu

Today's Class

- HW5 out
- Project Design Document
- Grades/Extra Credit Assignment
- Classical Planning
  - State-Space vs Plan-Space Planning
- Partial Order Planning (POP)
- Probabilistic Reasoning
  - Bayes Rule
  - Bayesian Inference

State-Space Planning

- We initially have a space of situations (where you are, what you have, etc.)
- The plan is a solution found by "searching" through the situations to get to the goal
- A **progression planner** searches forward from initial state to goal state
- A **regression planner** searches backward from the goal
  - This works if operators have enough information to go both ways
  - Ideally this leads to reduced branching: the planner is only considering things that are relevant to the goal

Plan-Space Planning

- An alternative is to **search through the space of *plans***, rather than situations.
- Start from a **partial plan** which is expanded and refined until a complete plan that solves the problem is generated.
- **Refinement operators** add constraints to the partial plan and modification operators for other changes.
- We can still use STRIPS-style operators:

  Op(ACTION: RightShoe, PRECOND: RightSockOn, EFFECT: RightShoeOn)

  Op(ACTION: RightSock, EFFECT: RightSockOn)

  Op(ACTION: LeftShoe, PRECOND: LeftSockOn, EFFECT: LeftShoeOn)

  Op(ACTION: LeftSock, EFFECT: leftSockOn)

could result in a partial plan of

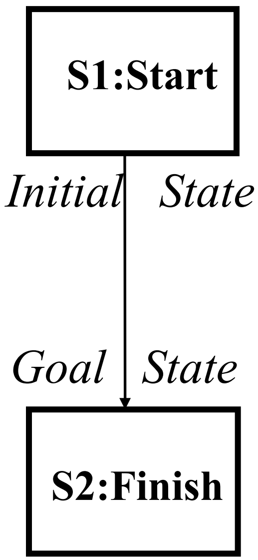
  [RightShoe, LeftShoe]

Partial-Order Planning

- A **linear planner** builds a plan as a **totally ordered sequence** of plan steps
- A **non-linear planner (aka partial-order planner)** builds up a plan as a set of steps with some temporal constraints
  - constraints of the form S1<S2 if step S1 must comes before S2.
- One **refines** a partially ordered plan (POP) by either:
  - **adding a new plan step**, or
  - **adding a new constraint** to the steps already in the plan.
- A POP can be **linearized** (converted to a totally ordered plan) by topological sorting

The Initial Plan

Every plan starts the same way



Trivial Example

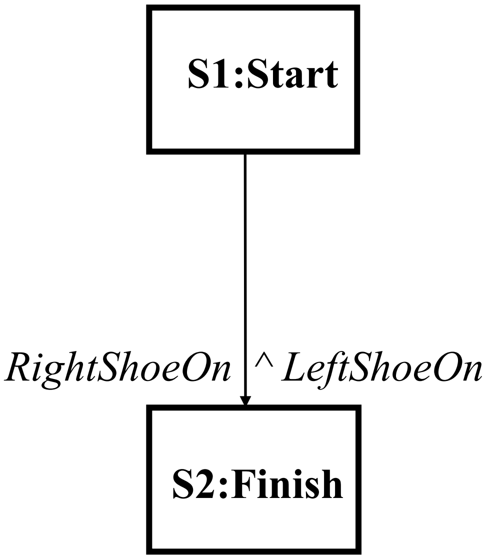
Operators:

Op(ACTION: RightShoe, PRECOND: RightSockOn, EFFECT: RightShoeOn)

Op(ACTION: RightSock, EFFECT: RightSockOn)

Op(ACTION: LeftShoe, PRECOND: LeftSockOn, EFFECT: LeftShoeOn)

Op(ACTION: LeftSock, EFFECT: leftSockOn)



Steps: {S1:[Op(Action:Start)],

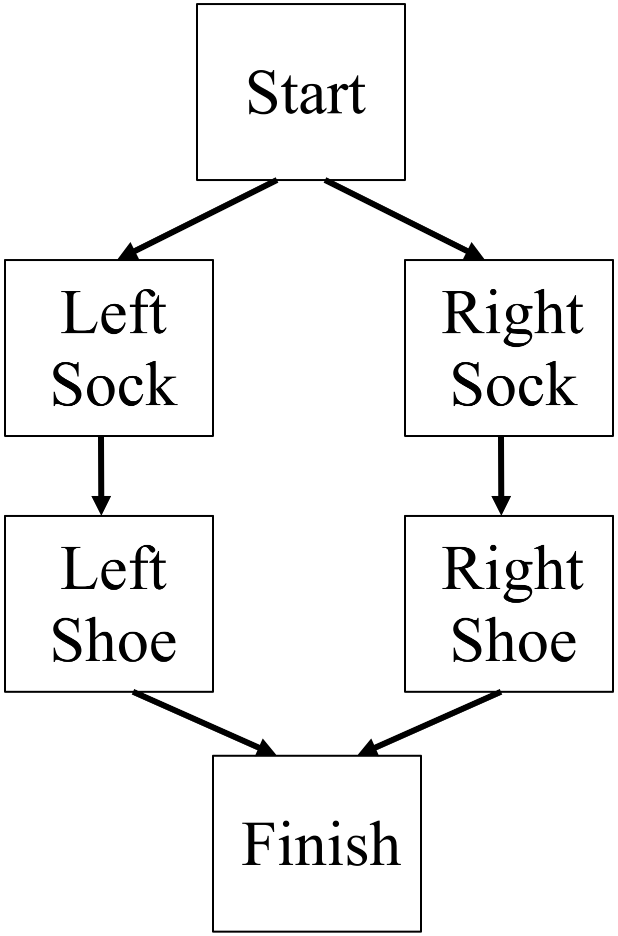
S2:[Op(Action:Finish,

Pre: RightShoeOn^LeftShoeOn)]}

Links: {}

Orderings: {S1<S2}

Solution



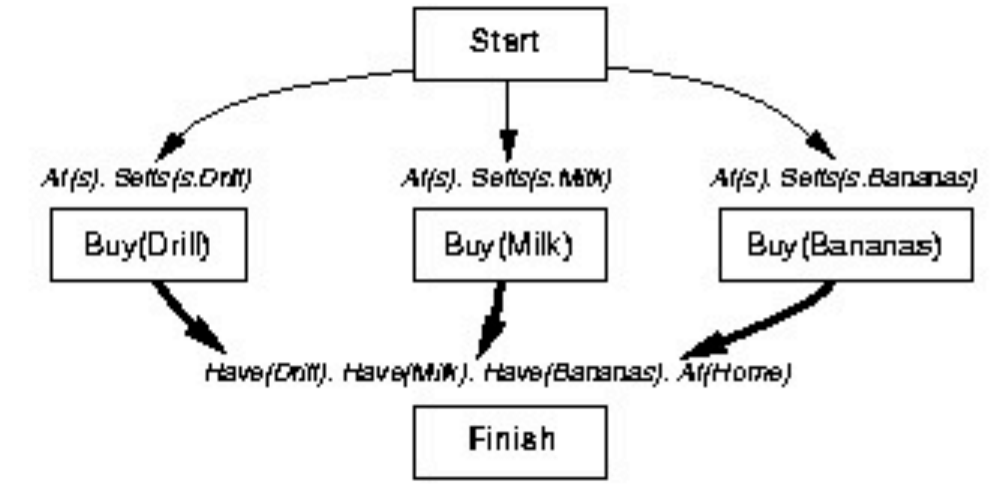
POP Constraints and Search Heuristics

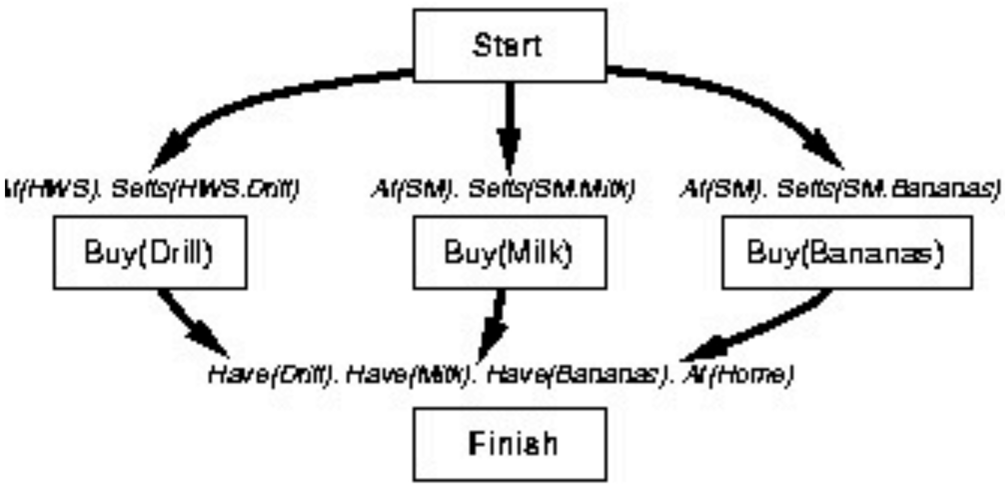
- Only add steps that achieve a currently unachieved precondition
- Use a least-commitment approach:
  - Don't order steps unless they need to be ordered
- Honor causal links $S_1 \xrightarrow{c} S_2$ that **protect** a condition *c*:
  - Never add an intervening step S_3 that violates *c*
  - If a parallel action **threatens** *c* (i.e., has the effect of negating or **clobbering** *c*), resolve that threat by adding ordering links:
    - Order S_3 before S_1 (**demotion**)
    - Order S_3 after S_2 (**promotion**)

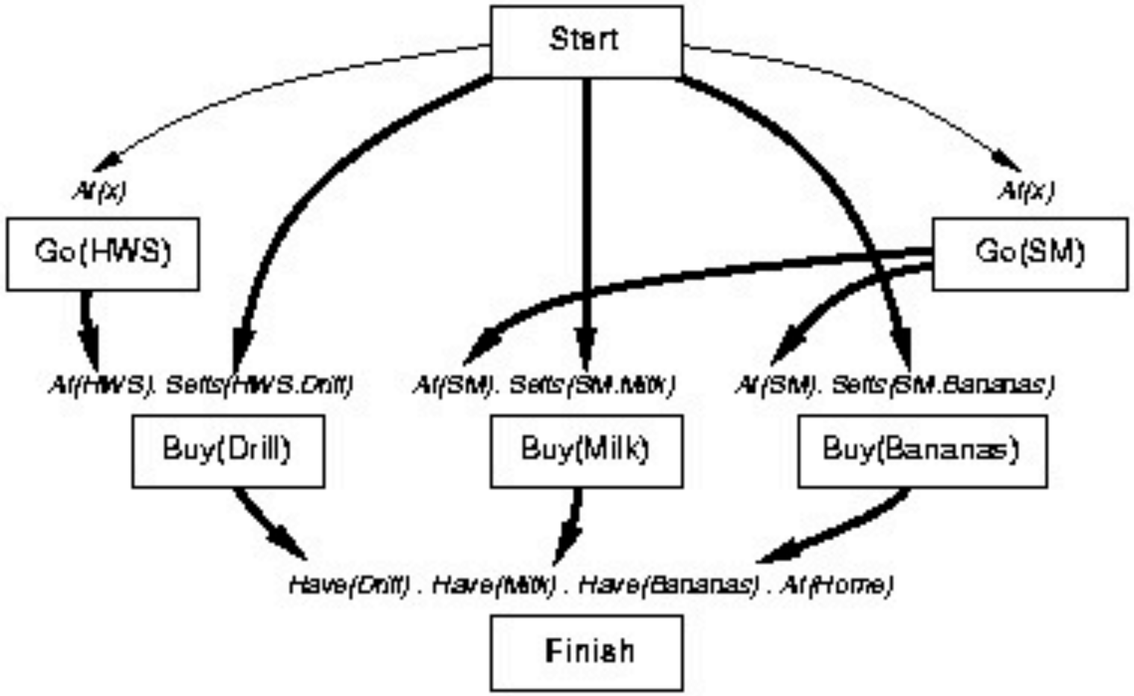
Partial-Order Planning Example

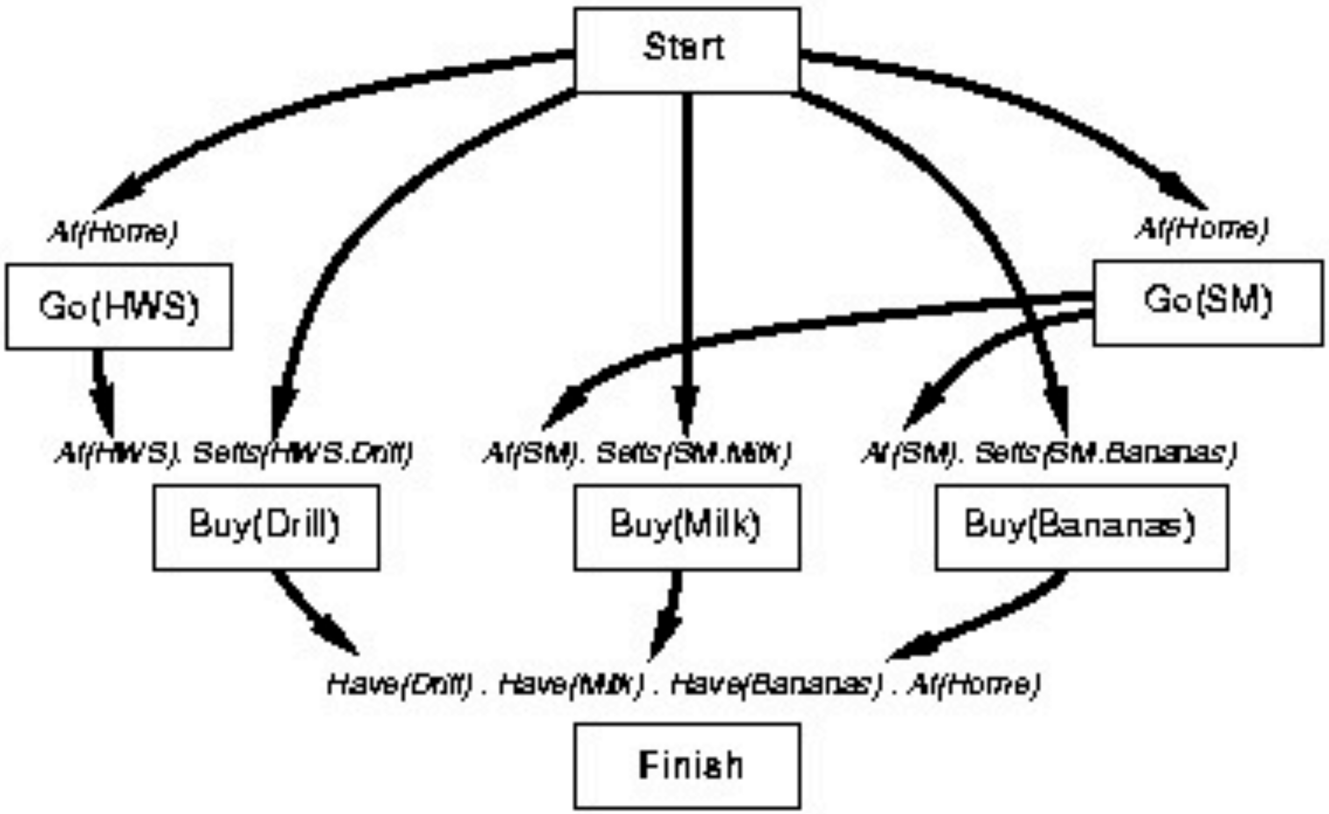
- Goal: Have milk, bananas, and a drill

Start
At(Home) Sells(SM,Banana) Sells(SM,Milk) Sells(HWS,Drill)
Have(Drill) Have(Milk) Have(Banana) At(Home)
Finish

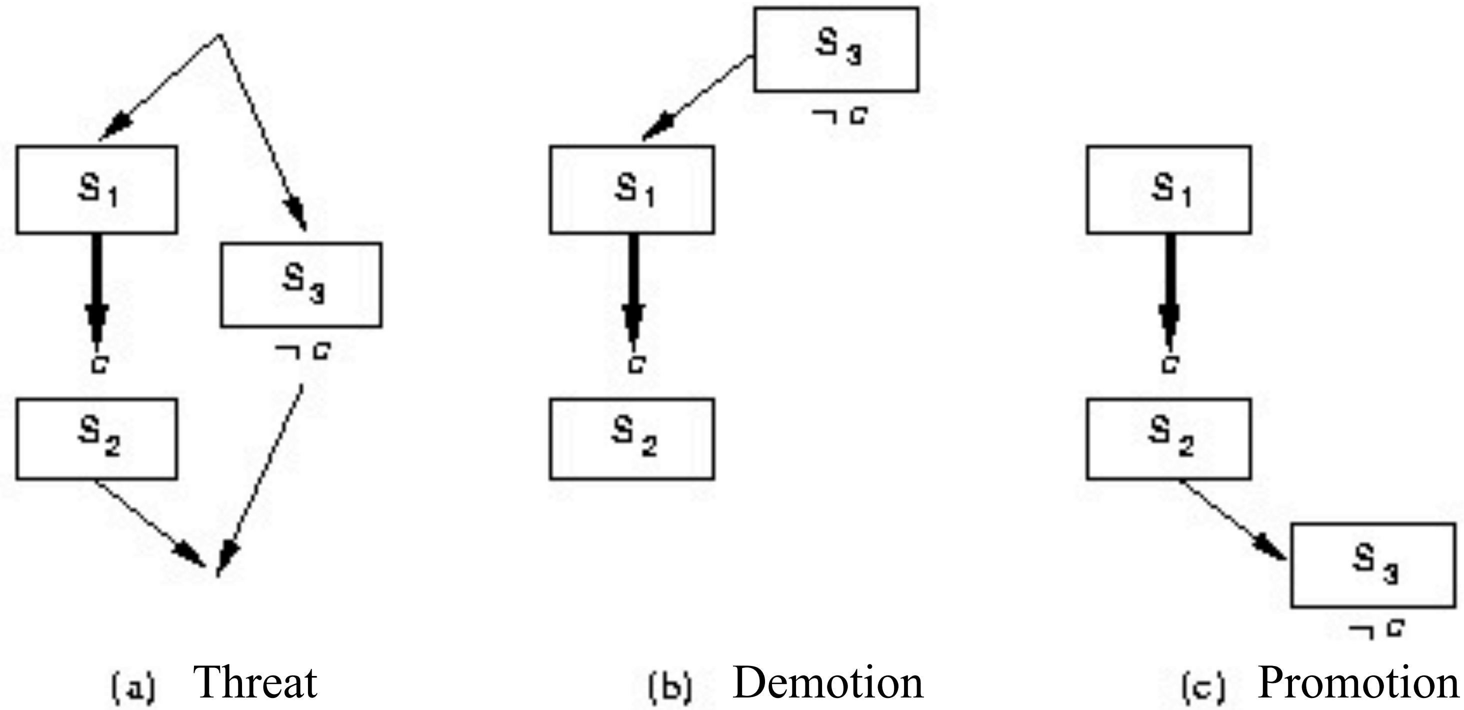

Start
At(s), Sells(s,Drill)
At(s), Sells(s,Milk)
At(s), Sells(s,Bananas)
Buy(Drill)
Buy(Milk)
Buy(Bananas)
Have(Drill), Have(Milk), Have(Bananas), At(Home)
Finish

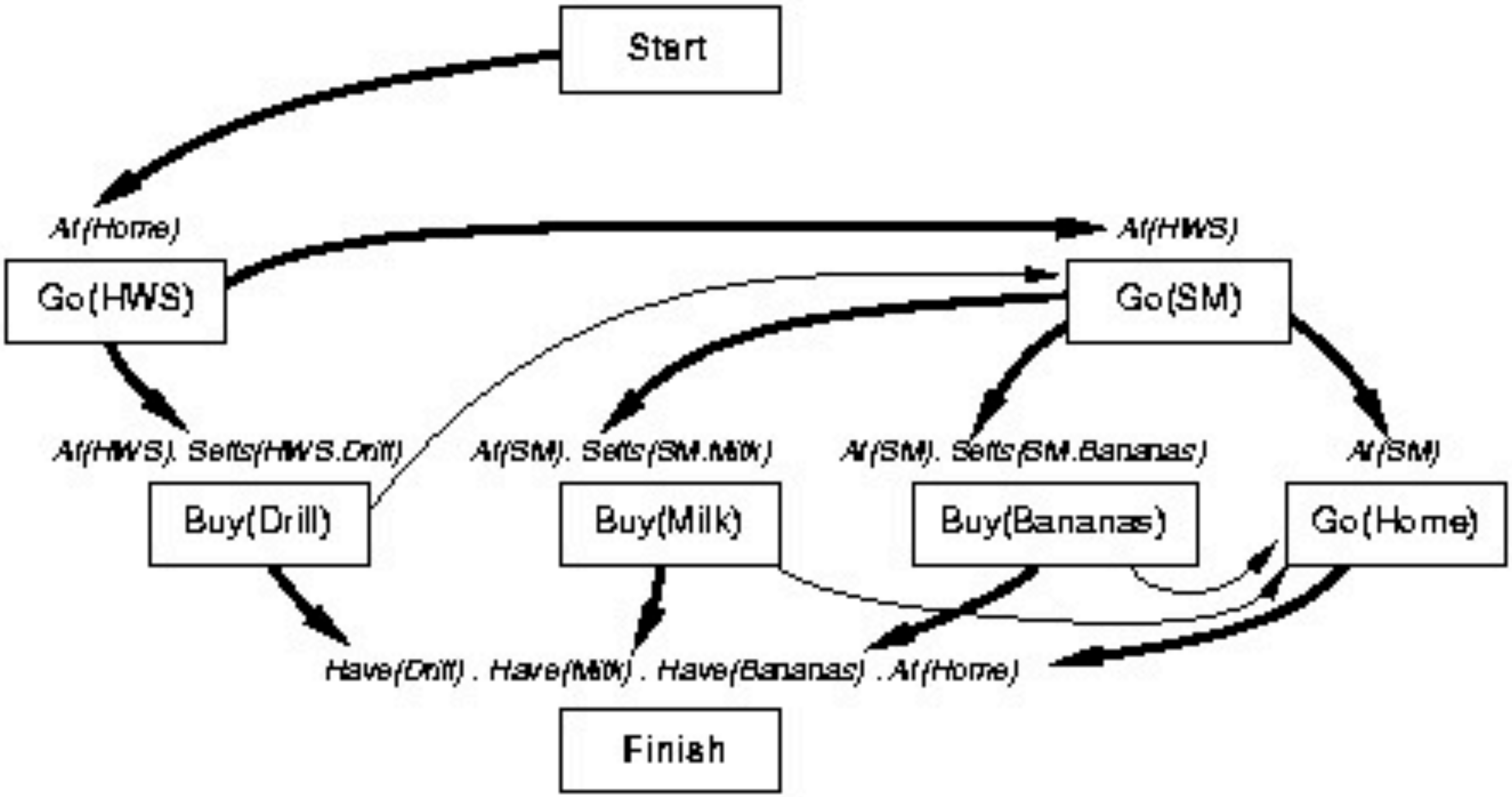

Start
t(HWS), Sells(HWS,Drill)
At(SM), Sells(SM,Milk)
At(SM), Sells(SM,Bananas)
Buy(Drill)
Buy(Milk)
Buy(Bananas)
Have(Drill), Have(Milk), Have(Bananas), At(Home)
Finish

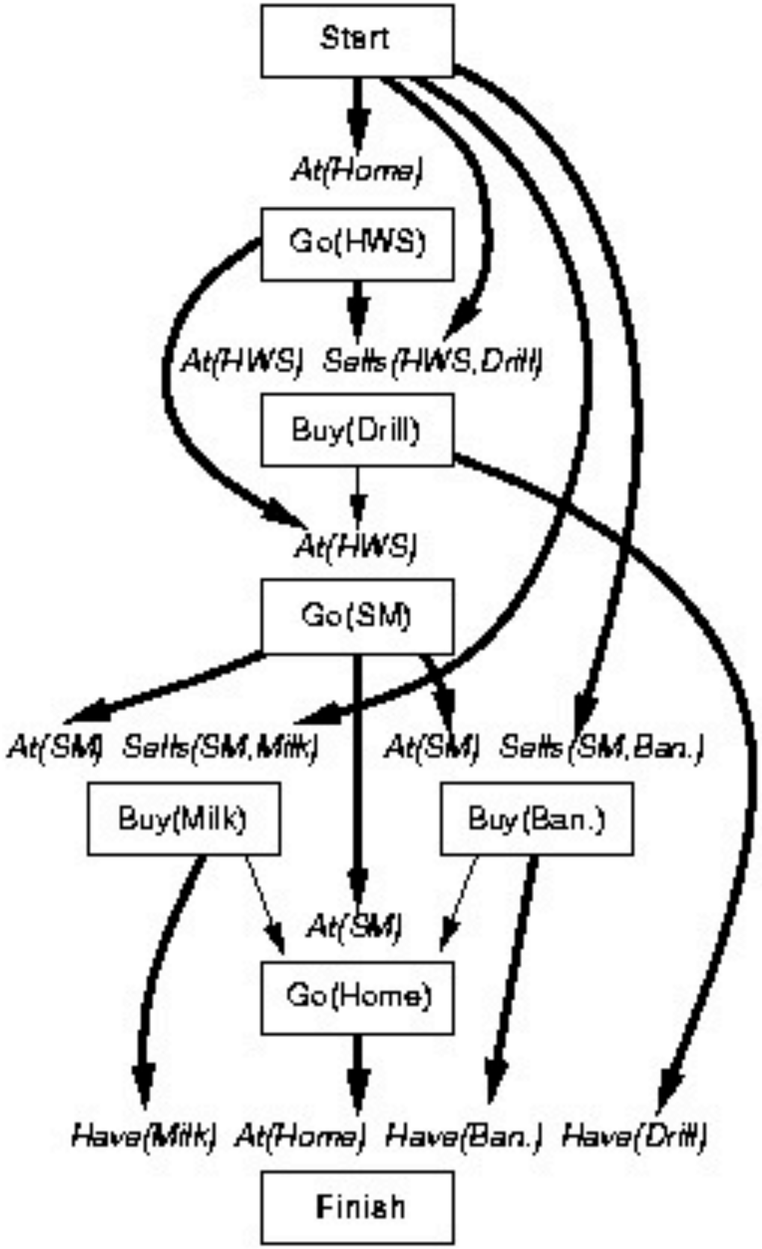

Start
At(x)
Go(HWS)
At(x)
Go(SM)
At(HWS), Sells(HWS,Drill)
At(SM), Sells(SM,Milk)
At(SM), Sells(SM,Bananas)
Buy(Drill)
Buy(Milk)
Buy(Bananas)
Have(Drill) , Have(Milk) , Have(Bananas) , At(Home)
Finish


Start
At(Home)
Go(HWS)
At(Home)
Go(SM)
At(HWS). Sells(HWS.Drill)
At(SM). Sells(SM.Milk)
At(SM). Sells(SM.Bananas)
Buy(Drill)
Buy(Milk)
Buy(Bananas)
Have(Drill) . Have(Milk) . Have(Bananas) . At(Home)
Finish

Resolving Threats




Start
At(Home)
At(HWS)
Go(HWS)
Go(SM)
At(HWS), Sells(HWS,Drill)
At(SM), Sells(SM,Milk)
At(SM), Sells(SM,Bananas)
At(SM)
Buy(Drill)
Buy(Milk)
Buy(Bananas)
Go(Home)
Have(Drill) , Have(Milk) , Have(Bananas) , At(Home)
Finish



Start
At(Home)
Go(HWS)
At(HWS) Sells(HWS,Drill)
Buy(Drill)
At(HWS)
Go(SM)
At(SM) Sells(SM,Milk)
At(SM) Sells(SM,Ban.)
Buy(Milk)
Buy(Ban.)
At(SM)
Go(Home)
Have(Milk) At(Home) Have(Ban.) Have(Drill)
Finish

Bayesian Reasoning

Chapter 13

Sources of Uncertainty

- Uncertain **inputs**
  - Missing data
  - Noisy data
- Uncertain **knowledge**
  - Multiple causes lead to multiple effects
  - Incomplete enumeration of conditions or effects
  - Incomplete knowledge of causality in the domain
  - Probabilistic/stochastic effects
- Uncertain **outputs**
  - Abduction and induction are inherently uncertain
  - Default reasoning, even in deductive fashion, is uncertain
  - Incomplete deductive inference may be uncertain

▸ Probabilistic reasoning only gives probabilistic results (summarizes uncertainty from various sources)

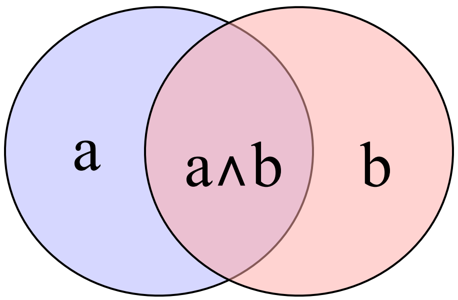
Decision Making with Uncertainty

- **Rational** behavior:
  - For each possible action, identify the possible outcomes
  - Compute the **probability** of each outcome
  - Compute the **utility** of each outcome
  - Compute the probability-weighted **(expected) utility** over possible outcomes for each action
  - Select the action with the highest expected utility (principle of **Maximum Expected Utility**)

Why Probabilities Anyway?

- Kolmogorov showed that three simple axioms lead to the rules of probability theory
  - De Finetti, Cox, and Carnap have also provided compelling arguments for these axioms

1. All probabilities are between 0 and 1:
   - $0 \leq P(a) \leq 1$
2. Valid propositions (tautologies) have probability 1, and unsatisfiable propositions have probability 0:
   - $P(true) = 1$; $P(false) = 0$
3. The probability of a disjunction is given by:
   - $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



Probability Theory

- **Random variables**
  - Domain
- Alarm, Burglary, Earthquake
  - Boolean (like these), discrete, continuous
- **Atomic event**: complete specification of state
- Alarm=True ∧ Burglary=True ∧ Earthquake=False
  alarm ∧ burglary ∧ ¬earthquake
- **Prior probability**: degree of belief without any other evidence
- P(Burglary) = .1
- **Joint probability**: matrix of combined probabilities of a set of variables
- P(Alarm, Burglary) =

	alarm	¬alarm
burglary	.09	.01
¬burglary	.1	.8

Probability Theory (cont.)

- **Conditional probability**: probability of effect given causes
- **Computing conditional probs**:
  - $P(a \mid b) = P(a \wedge b) / P(b)$
  - $P(b)$: **normalizing** constant
- **Product rule**:
  - $P(a \wedge b) = P(a \mid b) P(b)$
- **Marginalizing**:
  - $P(B) = \Sigma_a P(B, a)$
  - $P(B) = \Sigma_a P(B \mid a) P(a)$ (**conditioning**)

- P(burglary | alarm) = .47
  P(alarm | burglary) = .9
- P(burglary | alarm) =
  P(burglary ∧ alarm) / P(alarm)
  = .09 / .19 = .47
- P(burglary ∧ alarm) =
  P(burglary | alarm) P(alarm) =
  .47 * .19 = .09
- P(alarm) =
  P(alarm ∧ burglary) +
  P(alarm ∧ ¬burglary) =
  .09+.1 = .19

Example: Inference from the Joint

	alarm		¬alarm	
	earthquake	¬earthquake	earthquake	¬earthquake
burglary	.01	.08	.001	.009
¬burglary	.01	.09	.01	.79

Example: Inference from the Joint

	alarm		¬alarm	
	earthquake	¬earthquake	earthquake	¬earthquake
burglary	.01	.08	.001	.009
¬burglary	.01	.09	.01	.79

P(burglary | alarm)

Example: Inference from the Joint

	alarm		¬alarm	
	earthquake	¬earthquake	earthquake	¬earthquake
burglary	.01	.08	.001	.009
¬burglary	.01	.09	.01	.79

P(burglary | alarm)

P(burglary | alarm) = P(burglary ^ alarm) / P(alarm)

Example: Inference from the Joint

	alarm		¬alarm	
	earthquake	¬earthquake	earthquake	¬earthquake
burglary	.01	.08	.001	.009
¬burglary	.01	.09	.01	.79

P(burglary | alarm)

P(burglary | alarm) = P(burglary ^ alarm) / P(alarm)

P(alarm) = P(alarm ^ burglary ^ earthquake) + P(alarm ^ burglary ^ ¬earthquake) + P(alarm ^ ¬burglary ^ earthquake) + P(alarm ^ ¬burglary ^ ¬earthquake)

Example: Inference from the Joint

	alarm		¬alarm	
	earthquake	¬earthquake	earthquake	¬earthquake
burglary	.01	.08	.001	.009
¬burglary	.01	.09	.01	.79

P(burglary | alarm)

P(burglary | alarm) = P(burglary ^ alarm) / P(alarm)

P(alarm) = P(alarm ^ burglary ^ earthquake) + P(alarm ^ burglary ^ ¬earthquake) + P(alarm ^ ¬burglary ^ earthquake) + P(alarm ^ ¬burglary ^ ¬earthquake)

P(alarm) = 0.19

Example: Inference from the Joint

	alarm		¬alarm	
	earthquake	¬earthquake	earthquake	¬earthquake
burglary	.01	.08	.001	.009
¬burglary	.01	.09	.01	.79

P(burglary | alarm)

P(burglary | alarm) = P(burglary ^ alarm) / P(alarm)

P(alarm) = P(alarm ^ burglary ^ earthquake) + P(alarm ^ burglary ^ ¬earthquake) + P(alarm ^ ¬burglary ^ earthquake) + P(alarm ^ ¬burglary ^ ¬earthquake)

P(alarm) = 0.19

P(burglary | alarm) = 0.09 / 0.19 = 0.474

Example: Inference from the Joint

	alarm		¬alarm	
	earthquake	¬earthquake	earthquake	¬earthquake
burglary	.01	.08	.001	.009
¬burglary	.01	.09	.01	.79

P(burglary | alarm)

P(burglary | alarm) = P(burglary ^ alarm) / P(alarm)

P(alarm) = P(alarm ^ burglary ^ earthquake) + P(alarm ^ burglary ^ ¬earthquake) + P(alarm ^ ¬burglary ^ earthquake) + P(alarm ^ ¬burglary ^ ¬earthquake)

P(alarm) = 0.19

P(burglary | alarm) = 0.09 / 0.19 = 0.474

P(¬burglary | alarm)?

Exercise: Inference from the Joint

p(smart ∧ study ∧ prep)	smart		¬smart	
	study	¬study	study	¬study
prepared	.432	.16	.084	.008
¬prepared	.048	.16	.036	.072

- **Queries:**
  - **What is the prior probability of *smart*?**
  - **What is the prior probability of *study*?**
  - **What is the conditional probability of *prepared*, given *study* and *smart*?**
- **Save these answers for next time! ☺**

Independence

- When two sets of propositions do not affect each others' probabilities, we call them **independent**, and can easily compute their joint and conditional probability:
  - Independent (A, B) $\Leftrightarrow$ $P(A \wedge B) = P(A)\,P(B)$, $P(A \mid B) = P(A)$
- For example, {moon-phase, light-level} might be independent of {burglary, alarm, earthquake}
  - Then again, it might not: Burglars might be more likely to burglarize houses when there's a new moon (and hence little light)
  - But if we know the light level, the moon phase doesn't affect whether we are burglarized
  - Once we're burglarized, light level doesn't affect whether the alarm goes off
- We need a more complex notion of independence, and methods for reasoning about these kinds of relationships

Exercise: Independence

p(smart ∧ study ∧ prep)	smart		¬smart	
	study	**¬study**	**study**	**¬study**
prepared	.432	.16	.084	.008
¬prepared	.048	.16	.036	.072

- **Queries:**
  - **Is *smart* independent of *study*?**
  - **Is *prepared* independent of *study*?**

Conditional Independence

- Absolute independence:
  - A and B are **independent** if $P(A \wedge B) = P(A)\,P(B)$; equivalently, $P(A) = P(A \mid B)$ and $P(B) = P(B \mid A)$
- A and B are **conditionally independent** given C if
  - $P(A \wedge B \mid C) = P(A \mid C)\,P(B \mid C)$
- This lets us decompose the joint distribution:
  - $P(A \wedge B \wedge C) = P(A \mid C)\,P(B \mid C)\,P(C)$
- Moon-Phase and Burglary are ***conditionally independent given*** Light-Level
- Conditional independence is weaker than absolute independence, but still useful in decomposing the full joint probability distribution

Exercise: Conditional Independence

p(smart ∧ study ∧ prep)	smart		¬smart	
	study	**¬study**	**study**	**¬study**
prepared	.432	.16	.084	.008
¬prepared	.048	.16	.036	.072

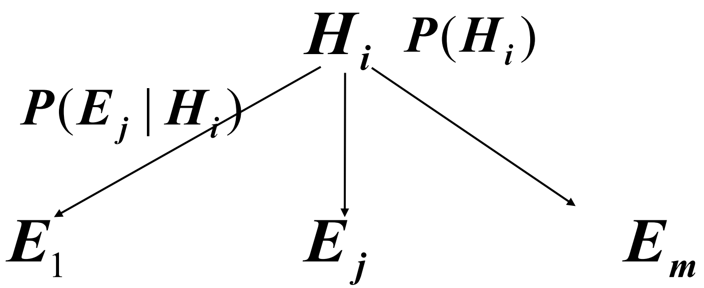
- **Queries:**
  - **Is *smart* conditionally independent of *prepared*, given *study*?**
  - **Is *study* conditionally independent of *prepared*, given *smart*?**

Bayes's Rule

- Bayes's rule is derived from the product rule:
  - P(Y | X) = P(X | Y) P(Y) / P(X)
- Often useful for diagnosis:
  - If X are (observed) effects and Y are (hidden) causes,
  - We may have a model for how causes lead to effects (P(X | Y))
  - We may also have prior beliefs (based on experience) about the frequency of occurrence of effects (P(Y))
  - Which allows us to reason abductively from effects to causes (P(Y | X)).

Bayesian Inference

- In the setting of diagnostic/evidential reasoning



hypotheses

evidence/manifestations

  - Know prior probability of hypothesis $P(H_i)$

    conditional probability $P(E_j \mid H_i)$

  - Want to compute the *posterior probability* $P(H_i \mid E_j)$

- Bayes's theorem (formula 1):

$$P(H_i \mid E_j) = P(H_i)P(E_j \mid H_i) / P(E_j)$$

Simple Bayesian Diagnostic Reasoning

- Knowledge base:
  - Evidence / manifestations: $E_1, \ldots E_m$
  - Hypotheses / disorders: $H_1, \ldots H_n$
    - E_j and H_i are **binary**; hypotheses are **mutually exclusive** (non-overlapping) and **exhaustive** (cover all possible cases)
  - Conditional probabilities: $P(E_j \mid H_i)$, $i = 1, \ldots n$; $j = 1, \ldots m$
- Cases (evidence for a particular instance): $E_1, \ldots, E_l$
- Goal: Find the hypothesis H_i with the highest posterior
  - $\text{Max}_i \; P(H_i \mid E_1, \ldots, E_l)$

Bayesian Diagnostic Reasoning II

- Bayes' rule says that
  - $P(H_i \mid E_1, \ldots, E_l) = P(E_1, \ldots, E_l \mid H_i)\ P(H_i) / P(E_1, \ldots, E_l)$
- Assume each piece of evidence E_i is conditionally independent of the others, *given* a hypothesis H_i, then:
  - $P(E_1, \ldots, E_l \mid H_i) = \prod^{l}_{j=1} P(E_j \mid H_i)$
- If we only care about relative probabilities for the H_i, then we have:
  - $P(H_i \mid E_1, \ldots, E_l) = \alpha\ P(H_i) \prod^{l}_{j=1} P(E_j \mid H_i)$

Limitations of Simple Bayesian Inference

- Cannot easily handle multi-fault situations, nor cases where intermediate (hidden) causes exist:
  - Disease D causes syndrome S, which causes correlated manifestations M_1 and M_2
- Consider a composite hypothesis $H_1 \wedge H_2$, where H_1 and H_2 are independent. What is the relative posterior?
  - $P(H_1 \wedge H_2 \mid E_1, \ldots, E_l) = \alpha\, P(E_1, \ldots, E_l \mid H_1 \wedge H_2)\, P(H_1 \wedge H_2)$
    $= \alpha\, P(E_1, \ldots, E_l \mid H_1 \wedge H_2)\, P(H_1)\, P(H_2)$
    $= \alpha \prod^{l}_{j=1} P(E_j \mid H_1 \wedge H_2)\, P(H_1)\, P(H_2)$
- How do we compute $P(E_j \mid H_1 \wedge H_2)$??

Limitations of Simple Bayesian Inference II

- Assume H1 and H2 are independent, given E1, …, El?
  - $P(H_1 \wedge H_2 \mid E_1, \ldots, E_l) = P(H_1 \mid E_1, \ldots, E_l)\, P(H_2 \mid E_1, \ldots, E_l)$
- This is a very unreasonable assumption
  - Earthquake and Burglar are independent, but *not* given Alarm:
    - P(burglar | alarm, earthquake) << P(burglar | alarm)
- Another limitation is that simple application of Bayes's rule doesn't allow us to handle causal chaining:
  - A: this year's weather; B: cotton production; C: next year's cotton price
  - A influences C indirectly: $A \rightarrow B \rightarrow C$
  - $P(C \mid B, A) = P(C \mid B)$
- Need a richer representation to model interacting hypotheses, conditional independence, and causal chaining
- Next time: conditional independence and Bayesian networks!