Inference in Bayesian Networks

Joseph E. Gonzalez

10701-Recitation

Convenient Probabilistic Model

- Compact Representation
  - Resistant to Over-fitting
  - Captures underlying structure
- Creates a “language” for describing probability relationships
- Permits the construction of generic inference techniques
  - Variable Elimination, Variational Methods, Belief Propagation

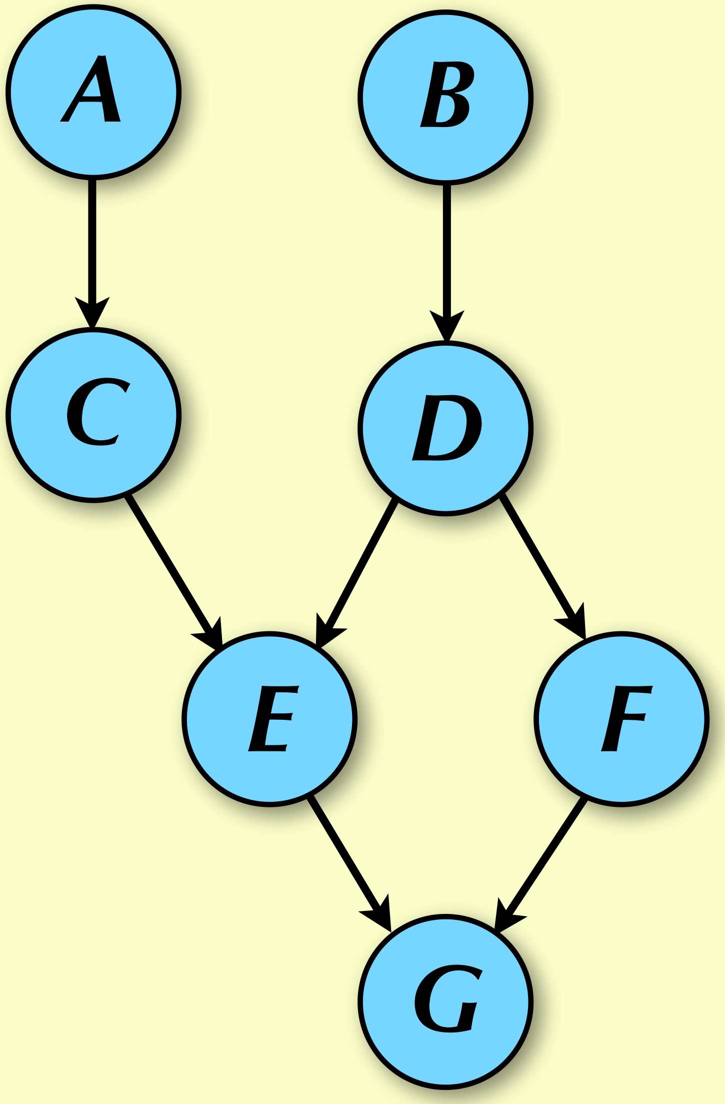
Joint Probability

- What can you compute with $P(X_1,X_2,\ldots,X_n)$?
  - Joint probability
  - Marginal Probability
  - Conditional distributions: $P(X_1 \mid X_2, X_3)$ (Predictions)
  - Most likely explanations: max $P(X_1, X_4 \mid X_2, X_3)$
  - Samples from the distribution
  - Information gain …

What is a Bayesian Network?

$P(A) = f_1(A)$
$P(B) = f_2(B)$
$P(C|A) = f_3(C,A)$
$P(D|B) = f_4(D,B)$
$P(E|C,D) = f_5(E,C,D)$
$P(F|D) = f_6(F,D)$
$P(G|E,F) = f_7(G,E,F)$

Conditional Probability Distributions / Tables / Functions



Directed Acyclic Graph

Conditional Probability … CPDs / CPTs

- Conditional Probability
  - CPD: Distributions
  - CPT: Tables
- Can be:
  - Discrete
  - Continuous

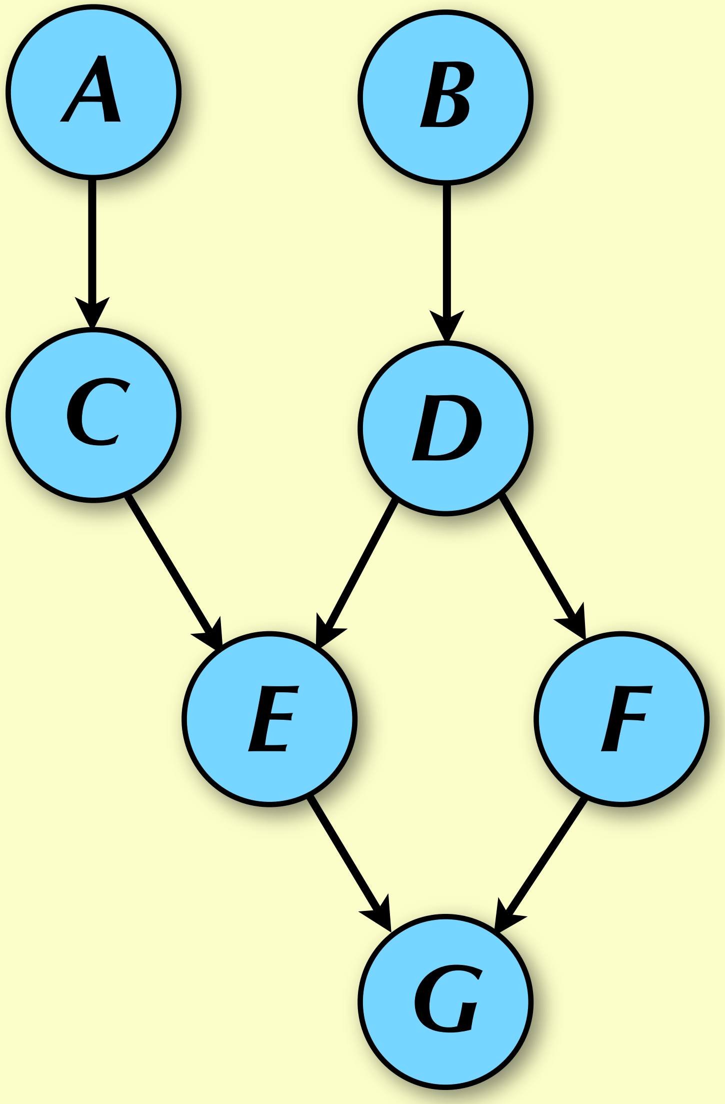
P(C|A)	A=T	A=F
C=T	0.3	0.6
C=F	0.7	0.4

OR

$$P(C|A) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{(C-A)^2}{2\sigma^2}\right)}$$

Graphs Describe Probability Distributions

$$P(A,B,C,D,E,F,G) = P(A)\,P(B)\,P(C|A)\,P(D|B)\,P(E|C,D)\,P(F|D)P(G|E,F)$$



Writing the Joint Probability

Bayesian Network

For a directed acyclic graph G=(V,E)

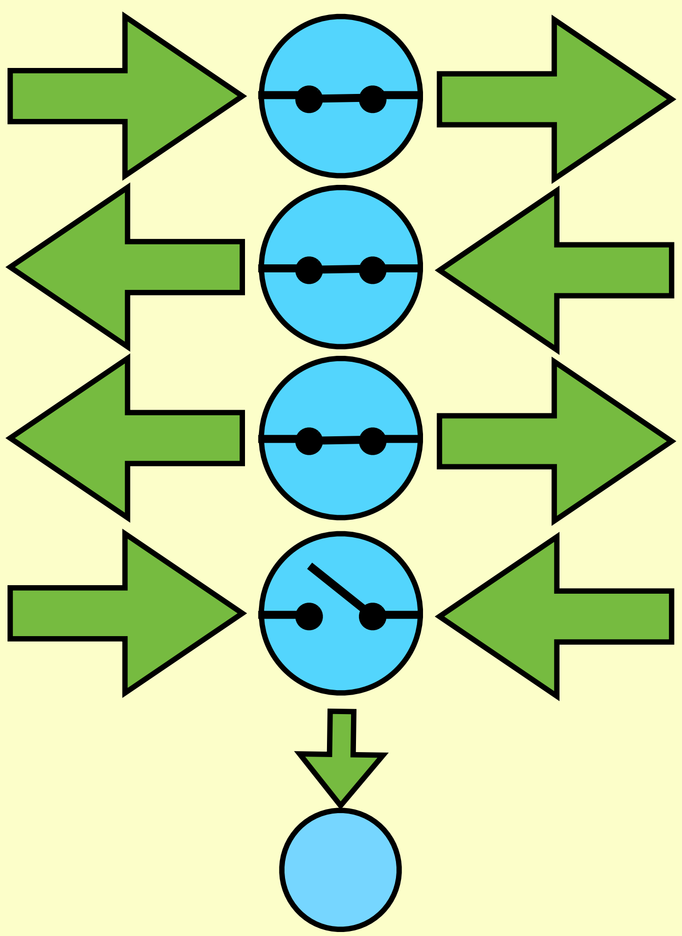
$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | Pa_G[X_i])$$

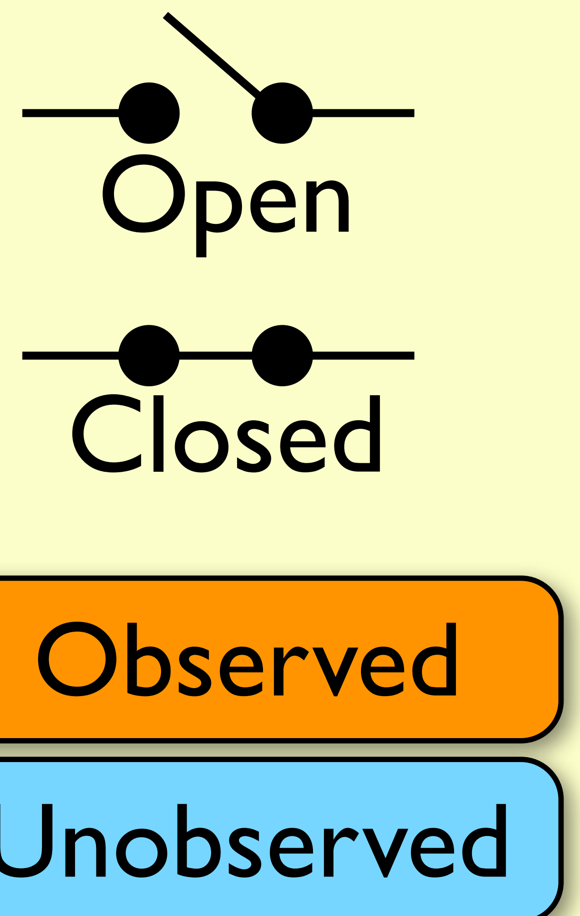
Where $Pa_G[X_i]$ are the parents of X_i

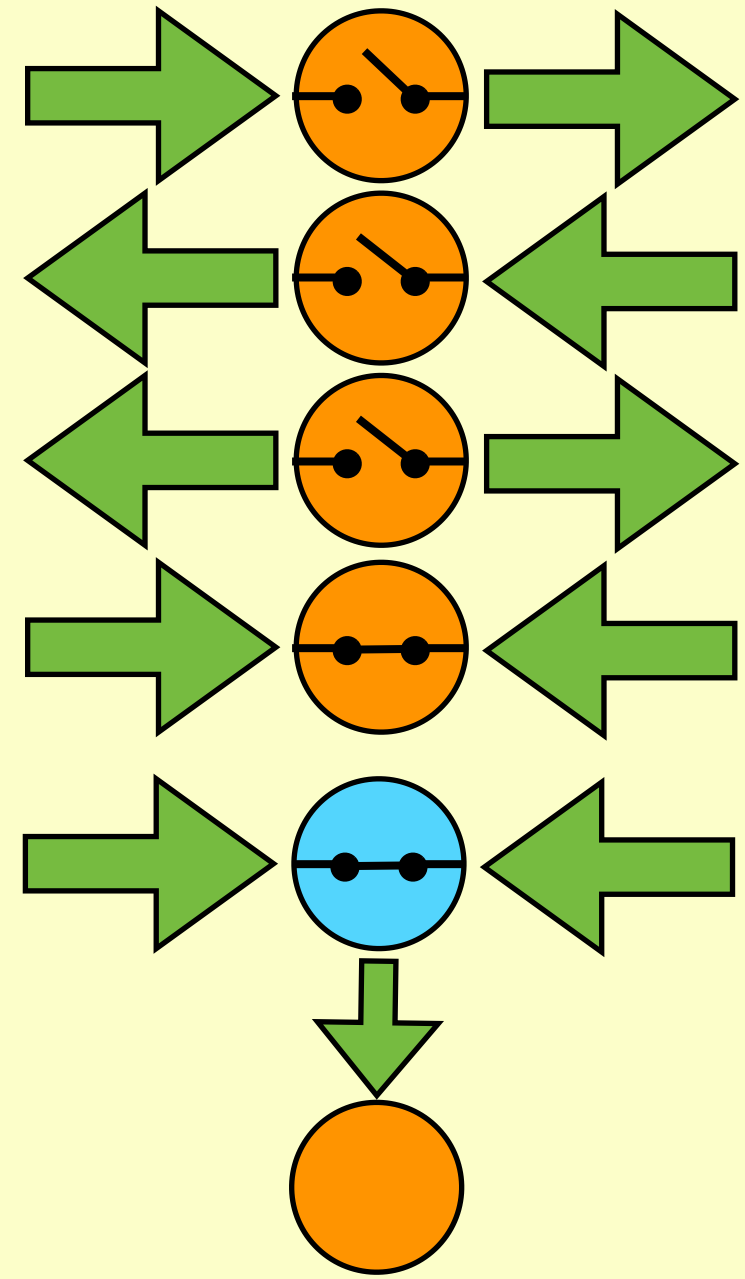
- We say that $P(X_1,\ldots,X_n)$ factors with respect to the graph G

Conditional Independence

If influence can flow from node A to node B (given observations) then nodes A and B are **not** (conditionally) independent.

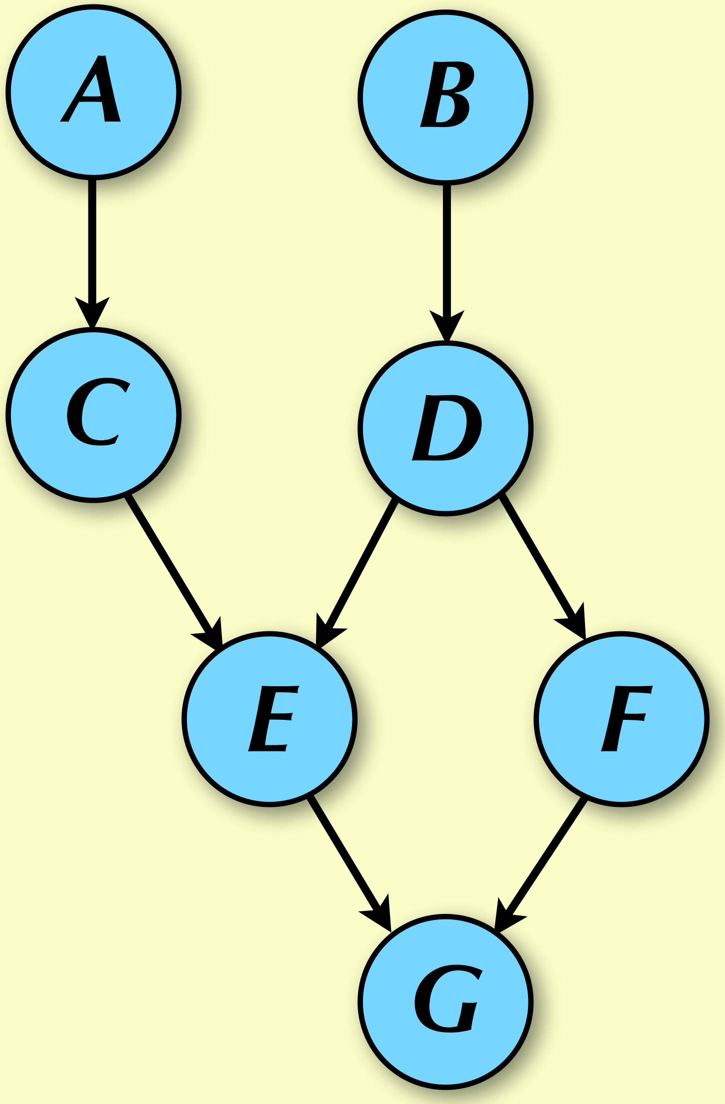






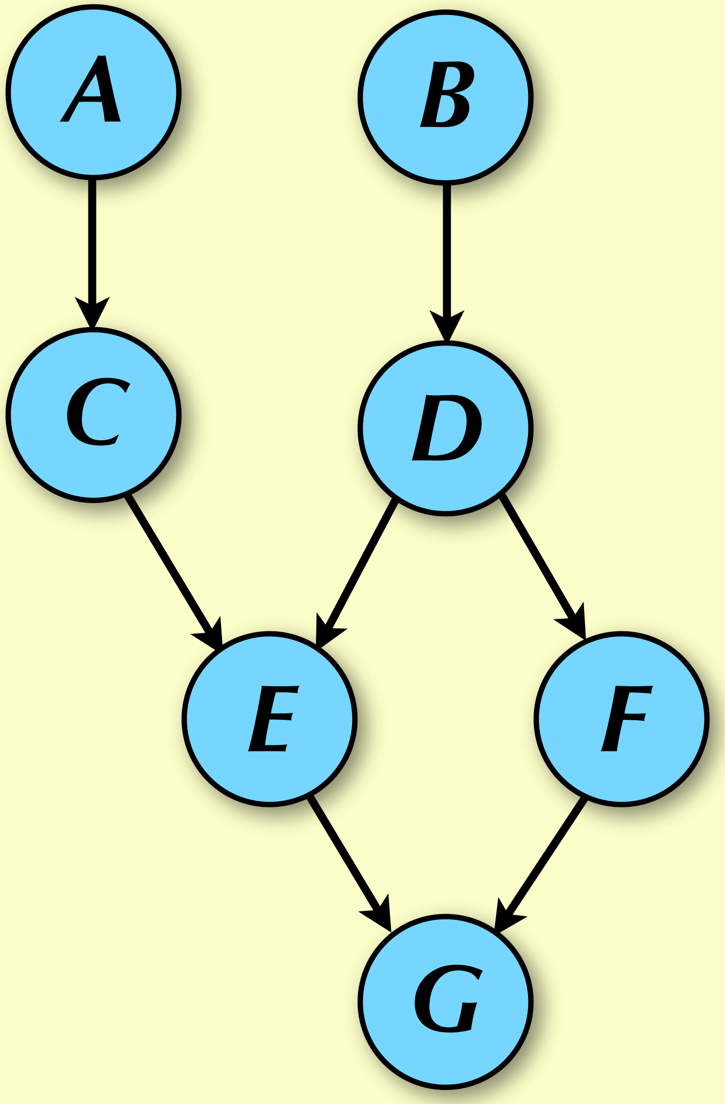
Independence

- A Indep B
  - Yes
- A Indep G
  - No
- A Indep B given G
  - No



Inference

- Computing Joint Assignment P(*A*=*a*, *B*=*b*,…, *B*=*b*)
  - Easy
- Computing Marginal
  - Tricky to Hard
- Computing Conditional
  - Requires Marginal



Why is it hard to Marginalize

Marginalize Unobserved Variables

You want $P(X_1,X_2)$ but you have $P(X_1,\ldots,X_n)$

$$P(X_1, X_2) = \sum_{x_3} \sum_{x_4} \ldots \sum_{x_n} P(X_1, X_2, x_3, x_4, \ldots, x_n)$$

- for X3 = 1:b
  - for X4 = 1:b
    - for X5 = 1:b
      - for X6 = 1:b ...

How can we save some work

Simple Algebra Example

Want to compute

$$\sum_a \sum_b \sum_c \sum_d f_1(a) f_2(b) f_3(c) f_4(d)$$

Rearrange the sums

Much Easier

$$\left(\sum_a f_1(a)\right)\left(\sum_b f_2(b)\right)\left(\sum_c f_3(c)\right)\left(\sum_d f_4(d)\right)$$

Return functions

Coupled Variables

Suppose we want to compute

$$\sum_a \sum_b \sum_c \sum_d f_1(a) f_2(b,a) f_3(c,b) f_4(d,c)$$

Functions rather than constants

$$\sum_a \sum_b \sum_c f_1(a) f_2(b,a) f_3(c,b) \sum_d f_4(d,c)$$

$g_1(c)$

Cache Computation

What is a Function

Assume variables are discrete

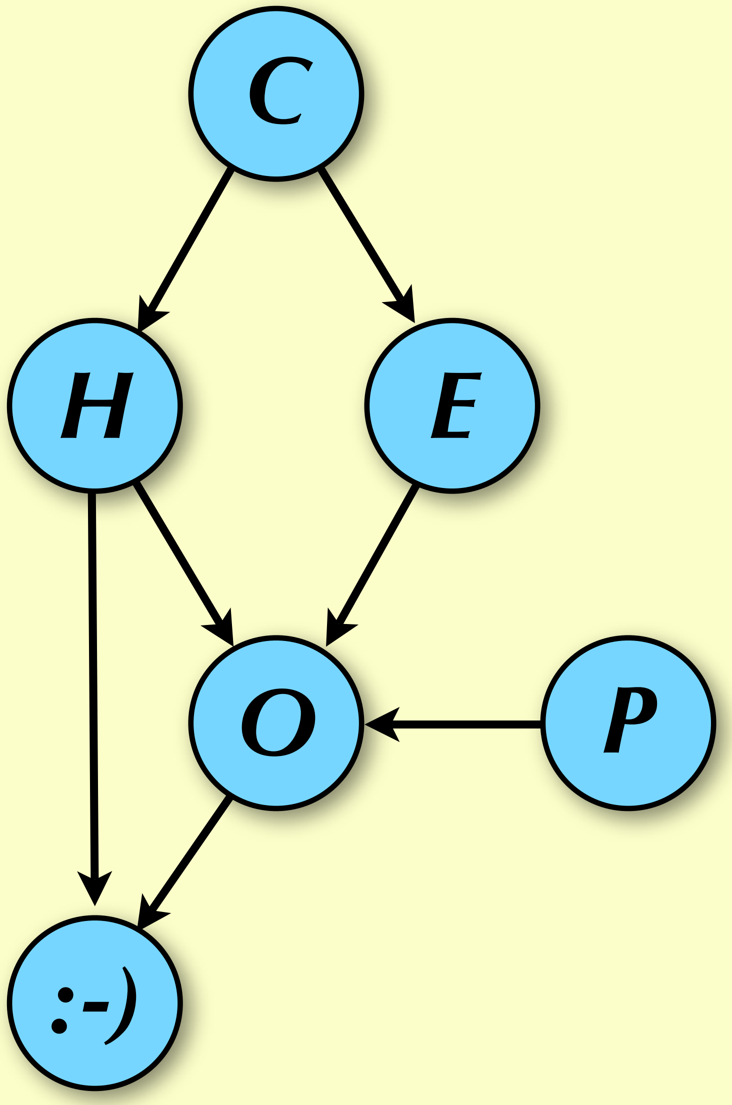
$$g(x_1, x_2, \ldots, x_n) = table[x_1, x_2, \ldots, x_n]$$

- Assume variables are discrete
- Each entry in the table is a cached “marginalization”

Example :: Graph

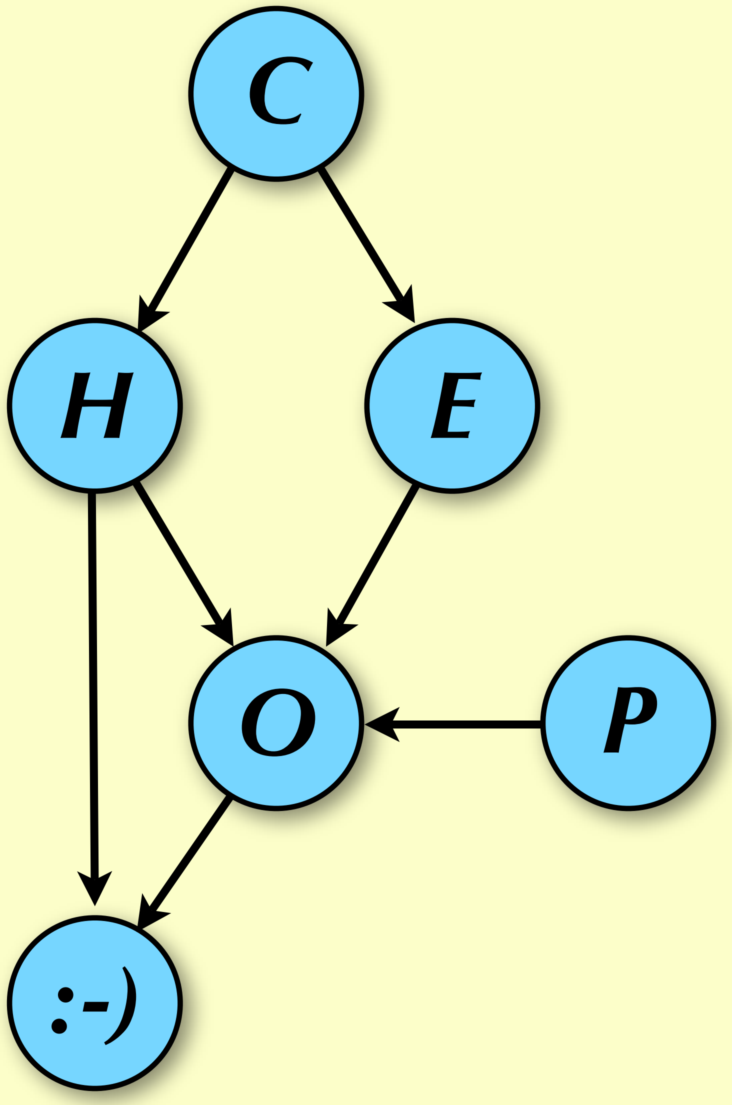
Define Some Variables

- (H) HW Grade = {P, F}
- (E) Exam Grade = {P,F}
- (P) Project Grade = {P,F}
- (O) Overall Grade = {P,F}
- (C) Class Attend = {T,F}
- (:-)) Happy = {T,F}



Example :: Joint Probability

$$\mathbf{P}(C, H, E, O, P, \text{:-)}) = \mathbf{P}(C)\mathbf{P}(P)\mathbf{P}(H \mid C)\mathbf{P}(E \mid C) \mathbf{P}(O \mid H, E, P)\mathbf{P}(\text{:-)} \mid H, O)$$



Example :: Tables

P(O|H,E,P)			O=p	O=f
H=p	E=p	P=p	0.95	0.05
H=p	E=p	P=f	0.6	0.4
H=p	E=f	P=p	0.6	0.4
H=p	E=f	P=f	0.3	0.7
H=f	E=p	P=p	0.3	0.7
H=f	E=p	P=f	0.2	0.8
H=f	E=f	P=p	0.1	0.9
H=f	E=f	P=f	0.01	0.99

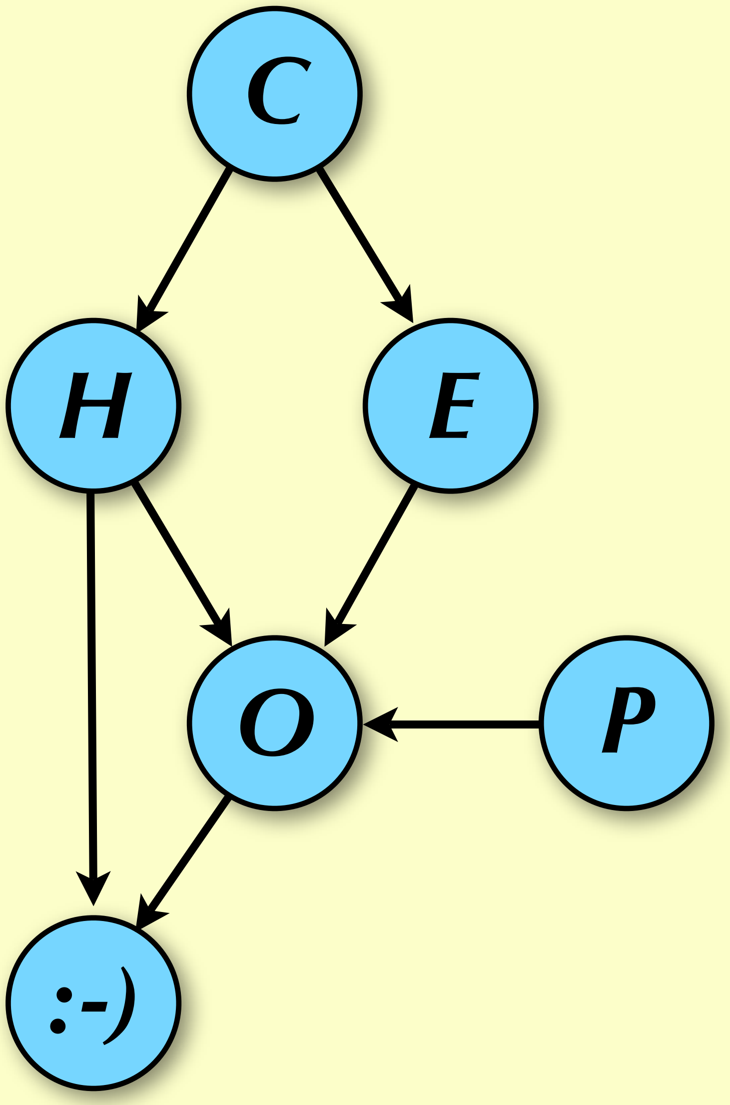
P(E|C)	E=p	E=f
C=t	0.7	0.3
C=f	0.4	0.6

P(H|C)	H=p	H=f
C=t	0.7	0.3
C=f	0.4	0.6

P(:-)|O,H)		:-)=t	:-)=f
O=p	H=p	0.7	0.3
O=p	H=f	0.8	0.2
O=f	H=p	0.2	0.8
O=f	H=f	0.1	0.9

P(P)	P=p	P=f
	0.7	0.3

P(C)	C=t	C=f
	0.8	0.2



Happiness and Class Attendance

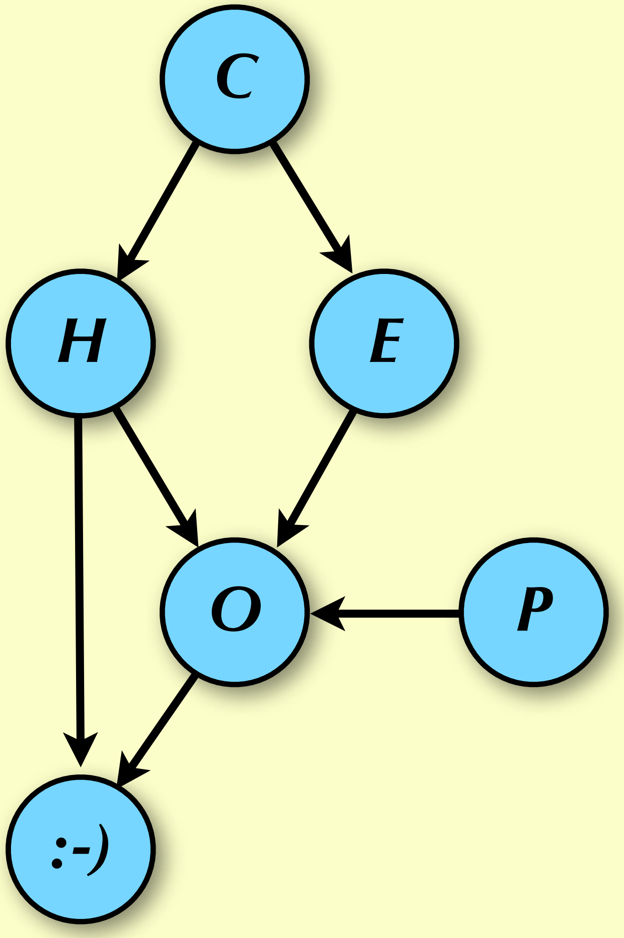
$$\mathbf{P}(:\text{-}), C) = \sum_H \sum_O \sum_P \sum_E \mathbf{P}(C, H, E, O, P, :\text{-}))$$

$$= \sum_{H,O,P,E} Pr(C)\mathbf{P}(P)\mathbf{P}(H \mid C)\mathbf{P}(E \mid C)$$

$$\mathbf{P}(O \mid H, E, P)\mathbf{P}(:\text{-}) \mid H, O)$$

$$= \sum_{H,O,E} Pr(C)\mathbf{P}(H \mid C)\mathbf{P}(E \mid C)\mathbf{P}(:\text{-}) \mid H, O)$$

$$\left(\sum_P \mathbf{P}(P)\mathbf{P}(O \mid H, E, P) \right)$$



This becomes a "function" table $g_1(O,H,E)$.

The Function g_1

$$g_1(O, H, E) = \sum_P \mathbf{P}(P)\mathbf{P}(O \mid H, E, P)$$

P(O|H,E,P)			O=p	O=f
H=p	E=p	P=p	0.95	0.05
H=p	E=p	P=f	0.6	0.4
H=p	E=f	P=p	0.6	0.4
H=p	E=f	P=f	0.3	0.7
H=f	E=p	P=p	0.3	0.7
H=f	E=p	P=f	0.2	0.8
H=f	E=f	P=p	0.1	0.9
H=f	E=f	P=f	0.01	0.99

P(P)	P=p	P=f
	0.7	0.3

g_1(O,H,E)			
O=p	H=p	E=p	0.7(0.95)+0.3(0.6)=0.845
O=p	H=p	E=f	0.7(0.6)+0.3(0.3)=0.510
O=p	H=f	E=p	0.7(0.3)+0.3(0.2)=0.270
O=p	H=f	E=f	0.7(0.1)+0.3(0.01)=0.073
O=f	H=p	E=p	0.7(0.05)+0.3(0.4)=0.155
O=f	H=p	E=f	0.7(0.4)+0.3(0.7)=0.490
O=f	H=f	E=p	0.7(0.7)+0.3(0.8)=0.730
O=f	H=f	E=f	0.7(0.9)+0.3(0.99)=0.927

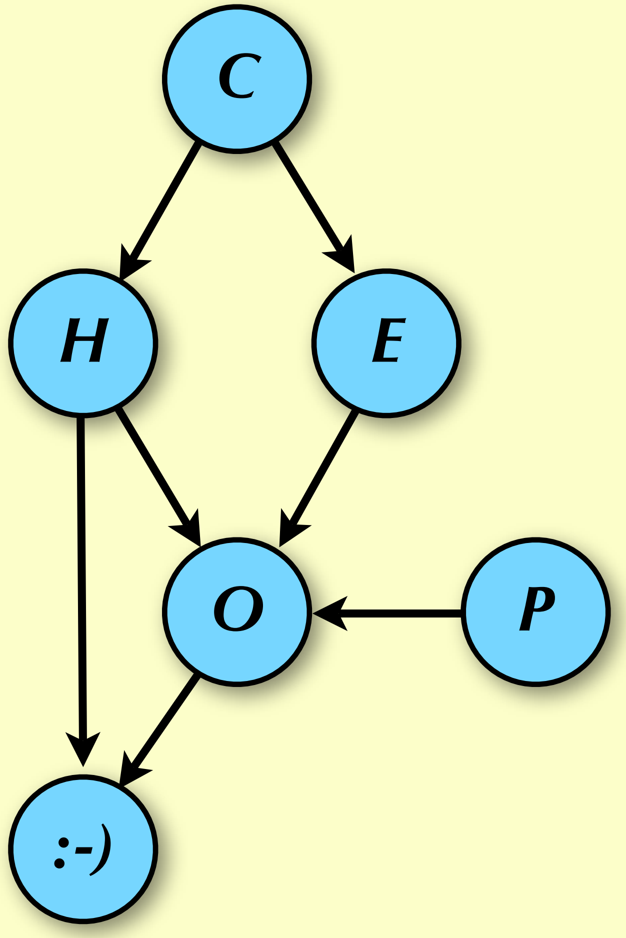
Happiness and Class Attendance

$$\mathbf{P}(:\text{-}), C) = \sum_{H,O,E} Pr(C)\mathbf{P}(H \mid C)\mathbf{P}(E \mid C)\mathbf{P}(:\text{-}) \mid H, O) g_1(O, H, E)$$

$$= \sum_{H,O} Pr(C)\mathbf{P}(H \mid C)\mathbf{P}(:\text{-}) \mid H, O) \left(\sum_{E} \mathbf{P}(E \mid C) g_1(O, H, E) \right)$$

$$g_2(O, H, C) = \sum_{E} \mathbf{P}(E \mid C) g_1(O, H, E)$$

Another table to compute



The Function g_2

$$g_2(O,H,C) = \sum_E \mathbf{P}(E \mid C) g_1(O,H,E)$$

g_1(O,H,E)			
O=p	H=p	E=p	0.845
O=p	H=p	E=f	0.510
O=p	H=f	E=p	0.270
O=p	H=f	E=f	0.073
O=f	H=p	E=p	0.155
O=f	H=p	E=f	0.490
O=f	H=f	E=p	0.730
O=f	H=f	E=f	0.927

P(E|C)	E=p	E=f
C=t	0.7	0.3
C=f	0.4	0.6

g_2(O,H,C)			
O=p	H=p	C=t	0.7(0.845)+0.3(0.510)=0.7445
O=p	H=p	C=f	0.4(0.845)+0.6(0.510)=0.644
O=p	H=f	C=t	0.7(0.270)+0.3(0.073)=0.2109
O=p	H=f	C=f	0.4(0.270)+0.6(0.073)=0.1518
O=f	H=p	C=t	0.7(0.155)+0.3(0.490)=0.2555
O=f	H=p	C=f	0.4(0.155)+0.6(0.490)=0.356
O=f	H=f	C=t	0.7(0.730)+0.3(0.927)=0.7891
O=f	H=f	C=f	0.4(0.730)+0.6(0.927)=0.8482

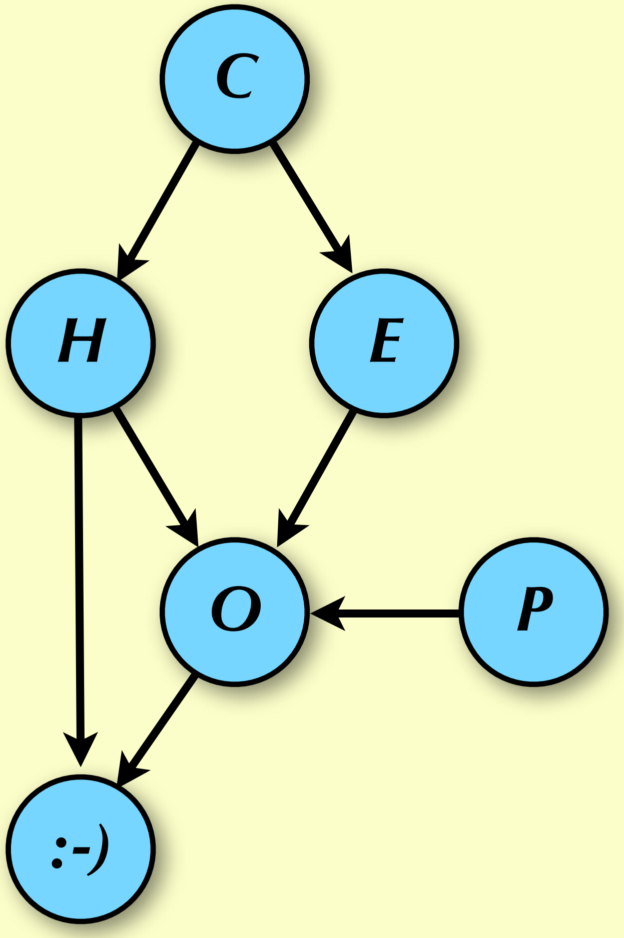
Happiness and Class Attendance

$$\mathbf{P}(:\text{-}), C) = \sum_{H,O} Pr(C)\mathbf{P}(H \mid C)\mathbf{P}(:\text{-}) \mid H, O) g_2(O, H, C)$$

$$= \sum_{H} Pr(C)\mathbf{P}(H \mid C) \left(\sum_{O} \mathbf{P}(:\text{-}) \mid H, O) g_2(O, H, C) \right)$$

$$g_3(:\text{-}), H, C) = \sum_{O} \mathbf{P}(:\text{-}) \mid H, O) g_2(O, H, C)$$

Another table to compute



The Function g_3

P(:-)|O,H)		:-)=t	:-)=f
O=p	H=p	0.7	0.3
O=p	H=f	0.8	0.2
O=f	H=p	0.2	0.8
O=f	H=f	0.1	0.9

$$g_3(:\text{-}), H, C) = \sum_O \mathbf{P}(:\text{-}) \mid H, O) g_2(O, H, C)$$

g_2(O,H,C)			
O=p	H=p	C=t	0.7445
O=p	H=p	C=f	0.644
O=p	H=f	C=t	0.2109
O=p	H=f	C=f	0.1518
O=f	H=p	C=t	0.2555
O=f	H=p	C=f	0.356
O=f	H=f	C=t	0.7891
O=f	H=f	C=f	0.8482

g_3(:-),H,C)			
:-)=t	H=p	C=t	0.7(0.7445)+0.2(0.2555)=0.5723
:-)=t	H=p	C=f	0.7(0.644)+0.2(0.356)=0.522
:-)=t	H=f	C=t	0.8(0.2109)+0.1(0.7891)=0.2468
:-)=t	H=f	C=f	0.8(0.1518)+0.1(0.8482)=0.2063
:-)=f	H=p	C=t	0.3(0.7445)+0.8(0.2555)=0.4278
:-)=f	H=p	C=f	0.3(0.644)+0.8(0.356)=0.478
:-)=f	H=f	C=t	0.2(0.2109)+0.9(0.7891)=0.752
:-)=f	H=f	C=f	0.2(0.1518)+0.9(0.8482)=0.793

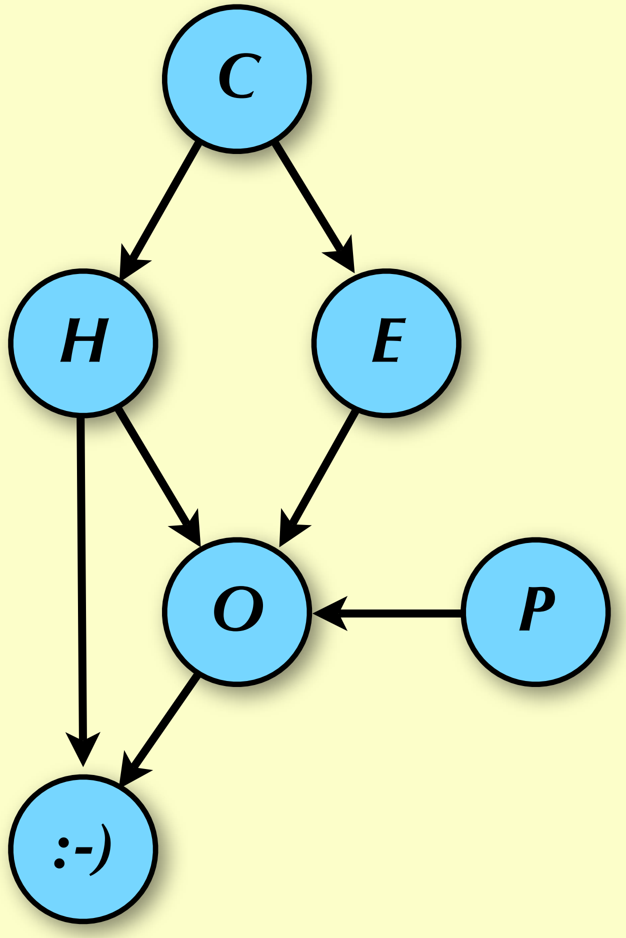
Happiness and Class Attendance

$$\mathbf{P}(:\text{-}), C) = \sum_H Pr(C)\mathbf{P}(H \,|\, C) g_3(:\text{-}), H, C)$$

$$= Pr(C) \sum_H \mathbf{P}(H \,|\, C) g_3(:\text{-}), H, C)$$

$$g_4(:\text{-}), C) = \sum_H \mathbf{P}(H \,|\, C) g_3(:\text{-}), H, C)$$

Last table to compute



The Function g_4

$$g_4(:\text{-}), C) = \sum_H \mathbf{P}(H \mid C) g_3(:\text{-}), H, C)$$

g3(:-),H,C)			
:-)=t	H=p	C=t	0.5723
:-)=t	H=p	C=f	0.522
:-)=t	H=f	C=t	0.2468
:-)=t	H=f	C=f	0.2063
:-)=f	H=p	C=t	0.4278
:-)=f	H=p	C=f	0.478
:-)=f	H=f	C=t	0.752
:-)=f	H=f	C=f	0.793

g4(:-),C)		
:-)=t	C=t	0.7(0.5723)+0.3(0.2468)=0.4747
:-)=t	C=f	0.4(0.522)+0.6(0.2063)=0.3326
:-)=f	C=t	0.7(0.4278)+0.3(0.752)=0.5251
:-)=f	C=f	0.4(0.478)+0.6(0.793)=0.667

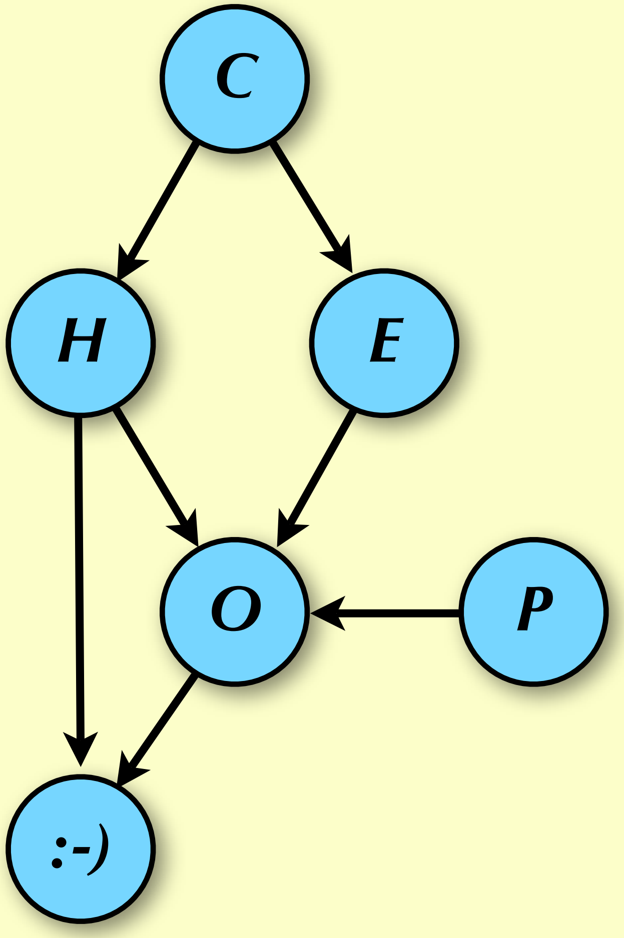
P(H|C)	H=p	H=f
C=t	0.7	0.3
C=f	0.4	0.6

Happiness and Class Attendance

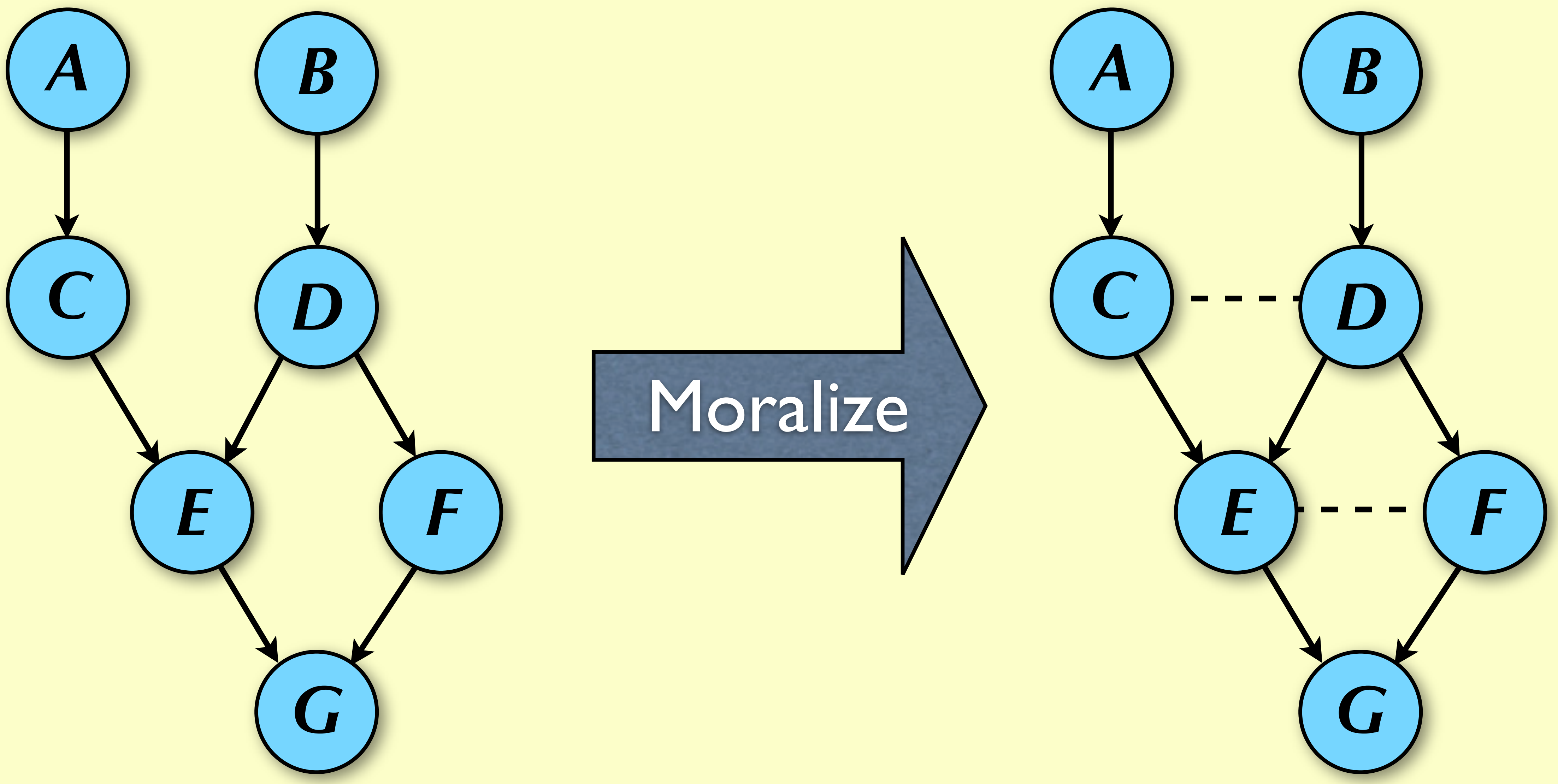
$$\mathbf{P}(\text{:-)}, C) = Pr(C) g_4(\text{:-)}, C)$$

$$\begin{aligned}\mathbf{P}(\text{:-)} \mid C) &= \frac{Pr(C) g_4(\text{:-)}, C)}{\sum_{\text{:-)}} Pr(C) g_4(\text{:-)}, C)} \\ &= \frac{Pr(C) g_4(\text{:-)}, C)}{Pr(C) \sum_{\text{:-)}} g_4(\text{:-)}, C)} \\ &= \frac{Pr(C) g_4(\text{:-)}, C)}{Pr(C)} \\ &= g_4(\text{:-)}, C)\end{aligned}$$

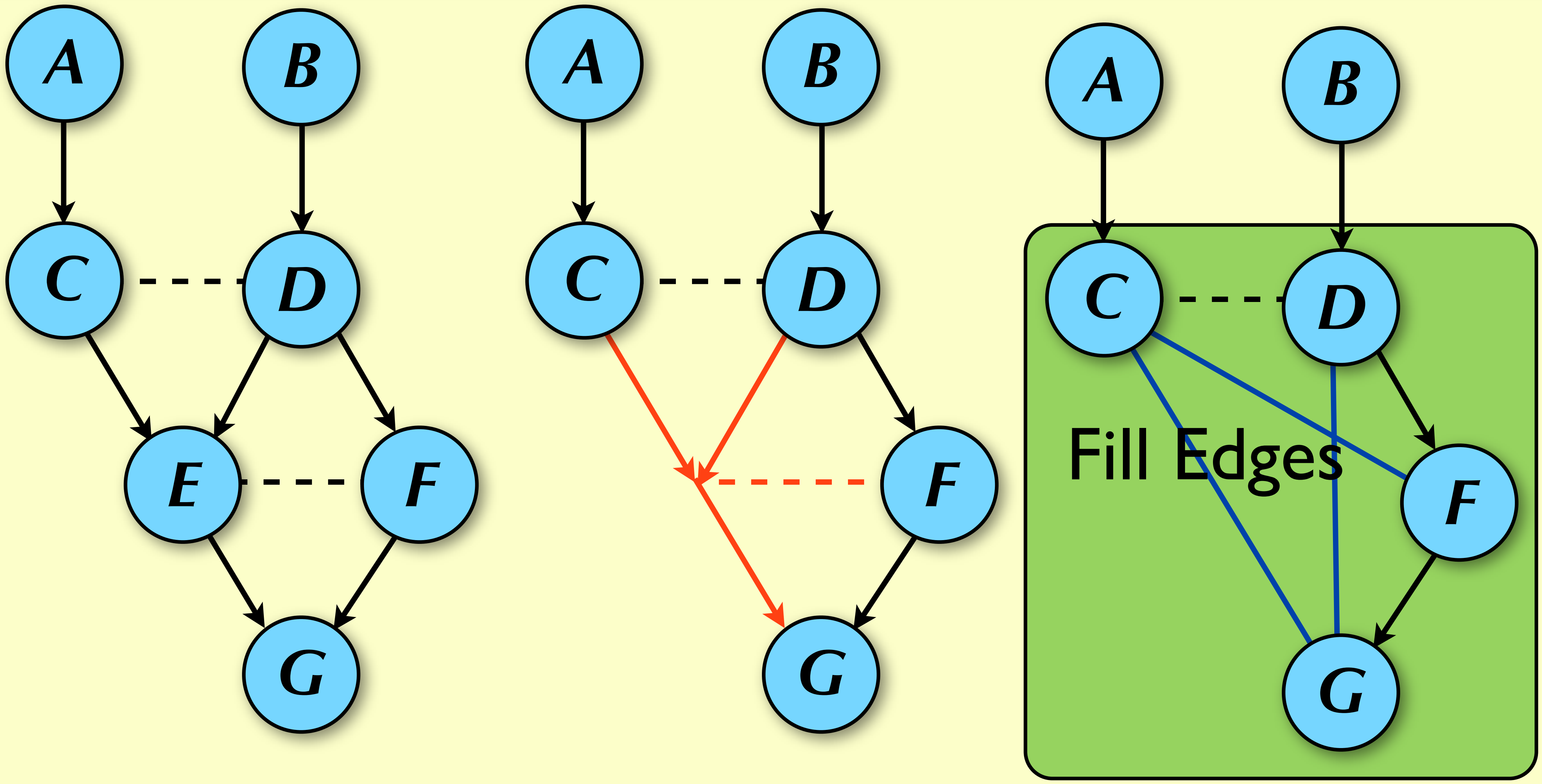
g_4(:-),C)		
:-)=t	C=t	0.4747
:-)=t	C=f	0.3326
:-)=f	C=t	0.5251
:-)=f	C=f	0.667



Variable Elimination as Pictures



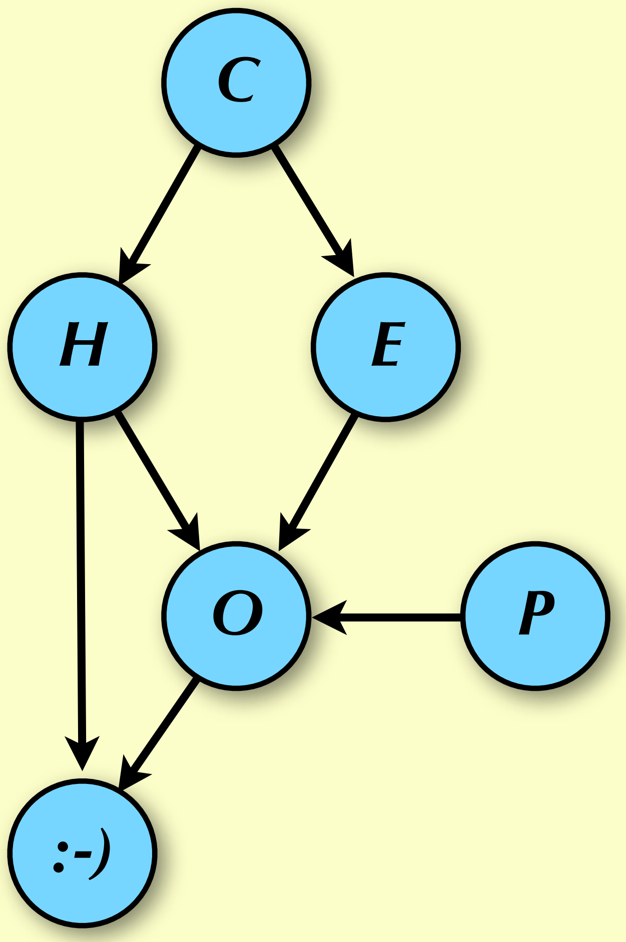
Variable Elimination and Fill Edges



$$g(C, D, G, F) = \sum_e P(E = e|C, D)P(G|E = e, F)$$

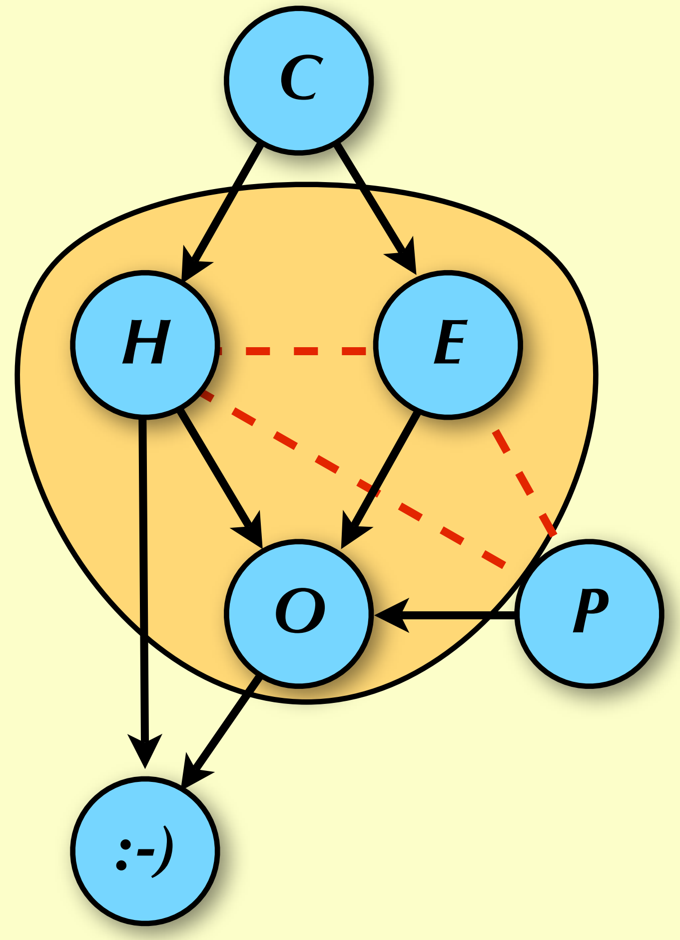
Recall The Happiness Problem

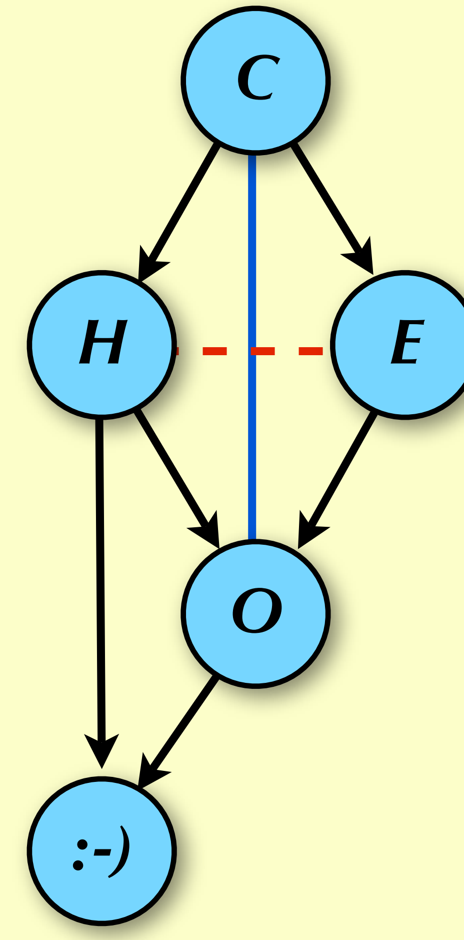
- We eliminated:
  - P,E,O,H
- Producing Functions
  - $g_1(O,H,E)$
  - $g_2(O,H,C)$
  - $g_3(:\text{-}),H,C)$
  - $g_4(:\text{-}),C)$

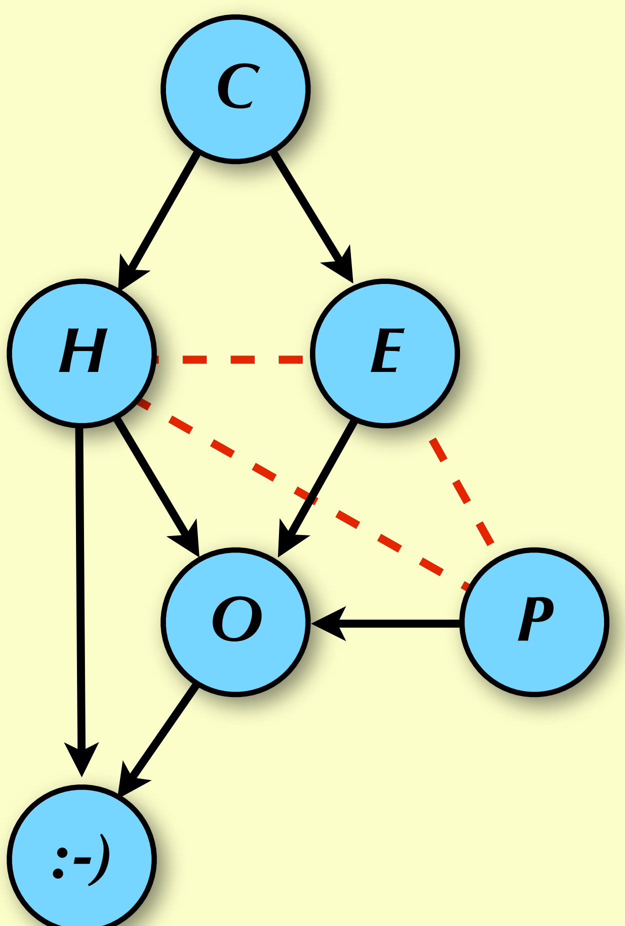


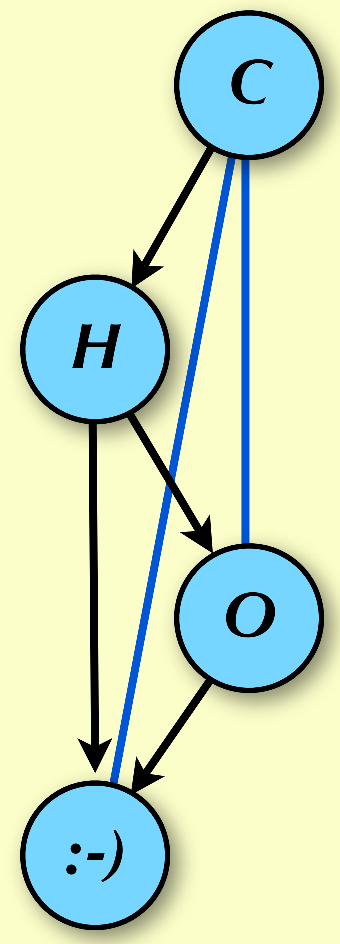
Recall The Happiness Problem

- P : $g_1(O,H,E)$
- E : $g_2(O,H,C)$
- O : $g_3(:\text{-}),H,C)$
- H : $g_4(:\text{-}),C)$









Summary

- Covered basic concepts of Variable Elimination
  - Worked through an example
  - Observed correspondence to graph operations
- Remember
  - There are other inference algorithms
  - Graphical models are a generic framework for modeling probability
    - Exploit structure (conditional independence) to manage representation and computational complexity
- Don't forget Midway milestone due Monday