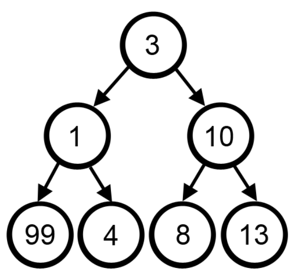
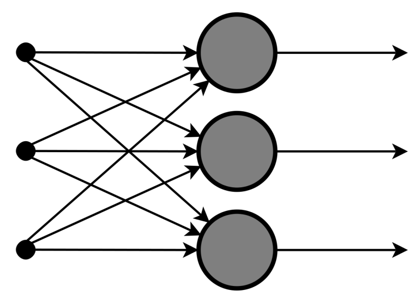
CSE 40171:
Artificial Intelligence





Probabilistic Read-Out Layers for Artificial Neural Networks:
Bayes' Theorem and Knowledge Representation

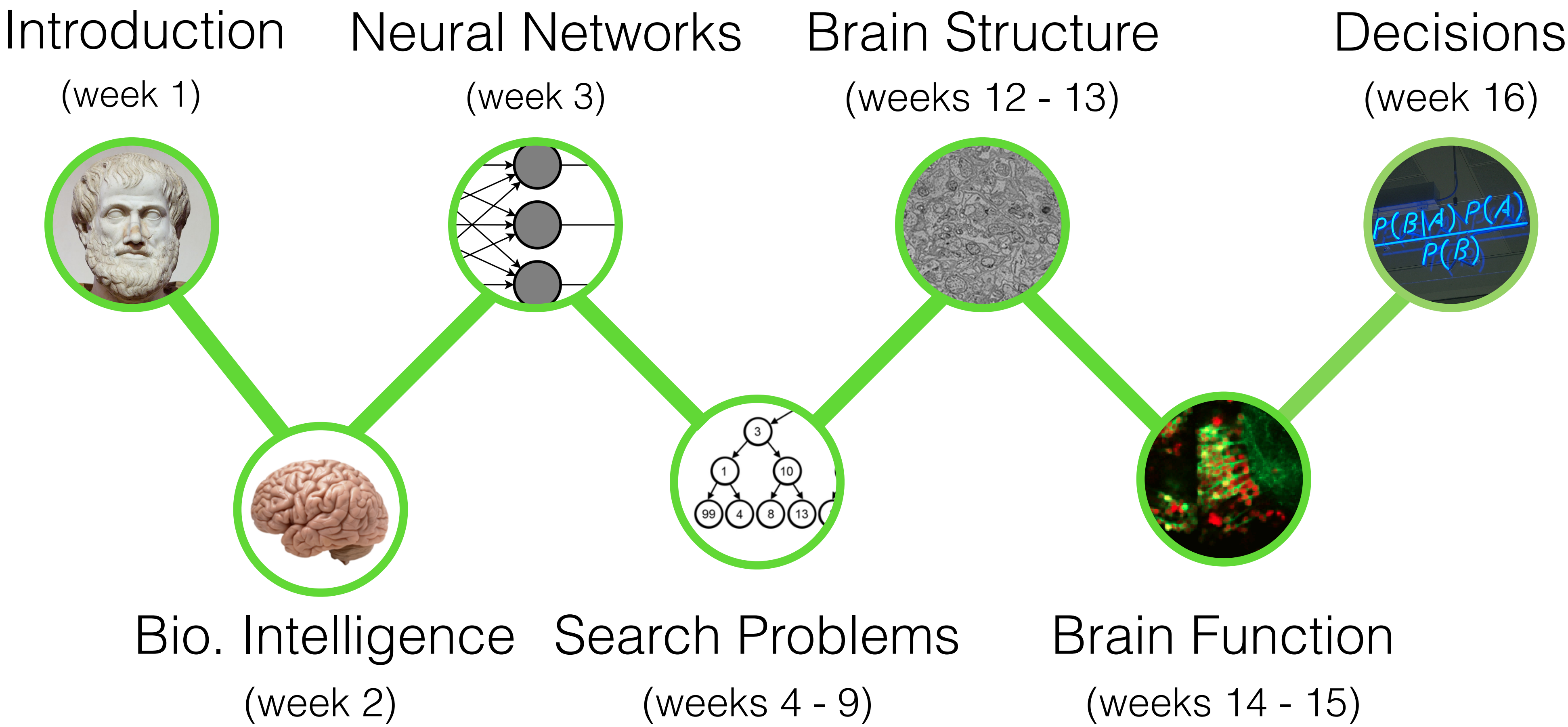
Homework #7 is due **tonight** at 11:59PM

Final Project Deliverable are Due
12/18 at 11:59PM

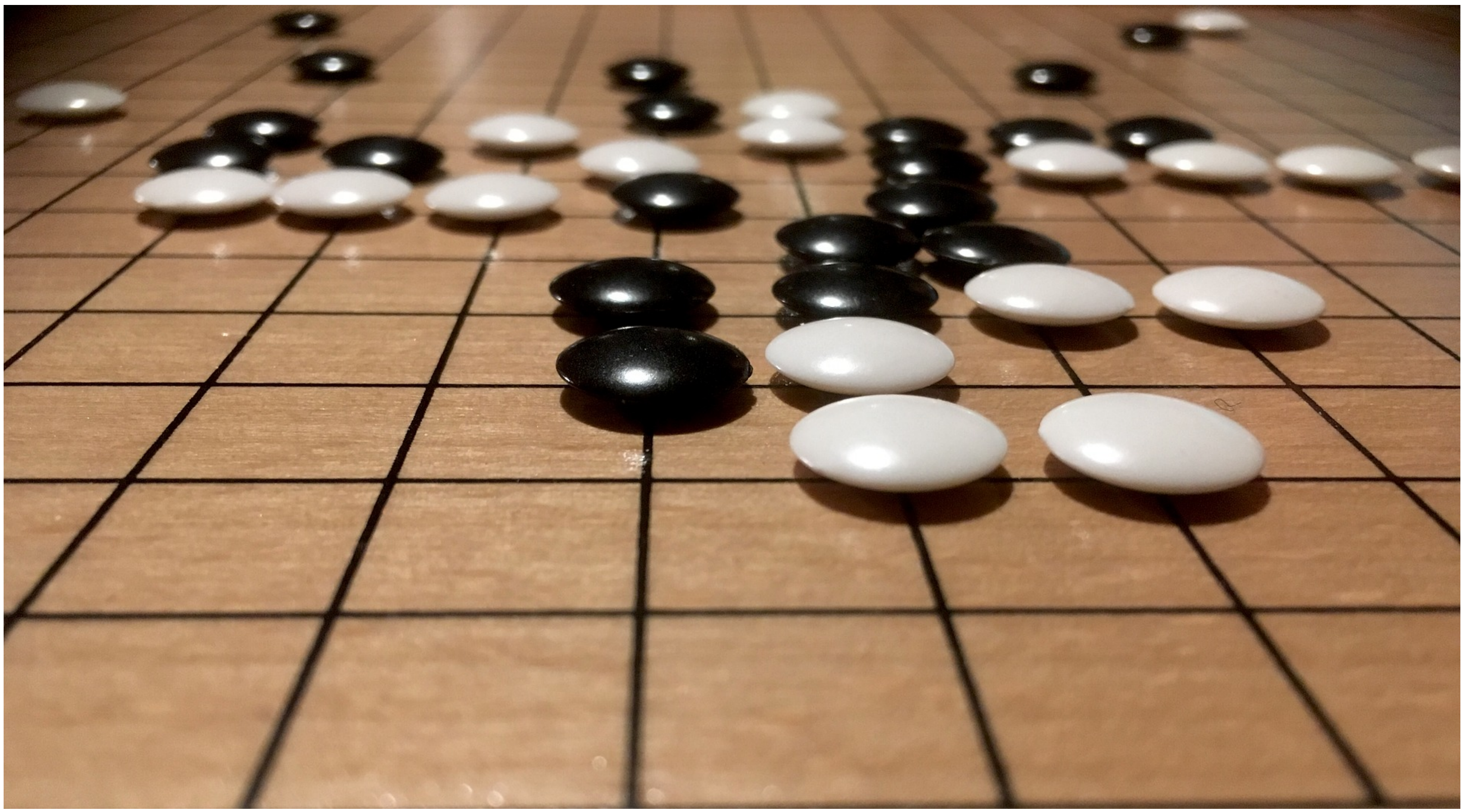
(See Course Website for Instructions)

Quiz #2 will take place on 12/11 in class. See review checklist on course website.

Course Roadmap



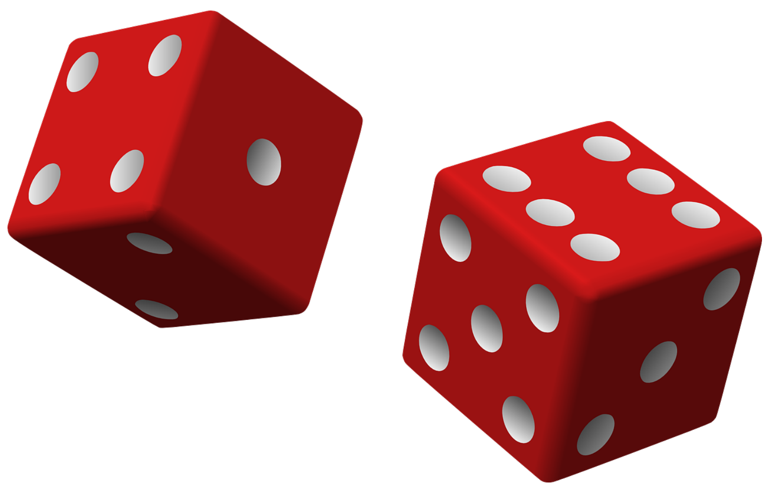
Games with no element of chance



Games with some element of chance

Seven Wonders Game

Games of Chance



Structured Text

“The Matrix has its roots in primitive arcade games,' said the voice-over, 'in early graphics programs and military experimentation with cranial jacks.' On the Sony, a two-dimensional space war faded behind a forest of mathematically generated ferns, demonstrating the spatial possibilities of logarithmic spirals; cold blue military footage burned through, lab animals wired into test systems, helmets feeding into fire control circuits of tanks and war planes. 'Cyberspace. A consensual hallucination experienced daily by billions of legitimate operators, in every nation, by children being taught mathematical concepts... A graphic representation of data abstracted from the banks of every computer in the human system. Unthinkable complexity. Lines of light ranged in the nonspace of the mind, clusters and constellations of data. Like city lights, receding...”

Unstructured Environment

Green Park, London - April

Acting Under Uncertainty

Agents need to handle uncertainty

- Due to partial observability
- Due to non-determinism
- Due to a combination of the two

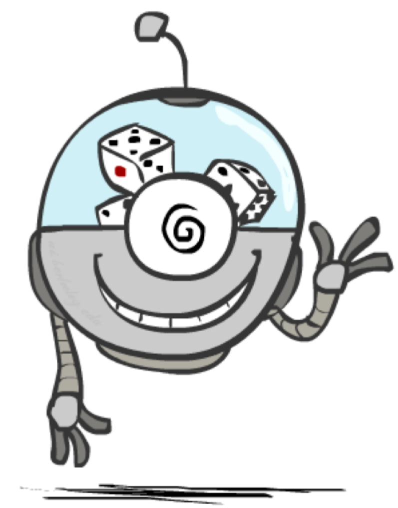


Image credit: Dan Klein and Pieter Abbeel, UC Berkeley CS 188

Problems with Belief States

- Partial sensor information: the agent has to consider **every logically possible** explanation for the available observations.
- Contingency plans: as the state space grows, so does the space for contingency planning
- What if we don't have a plan? We still need to choose an action

Rationality

Russell and Norvig tell us: The right thing to do — the *rational decision* — therefore depends on both the relative importance of various goals and the likelihood that, and the degree to which, they will be achieved.

The Frequentist Philosophy

Numbers can only come from experiments

Example: If we test 100 people and find that 10 of them have a cavity, then we can say the probability of a cavity is approximately 0.1.

From any finite sample, we can estimate the true fraction and also calculate how accurate our estimate is likely to be

Summarizing Uncertainty

Let's try to diagnose a patient's toothache:

$$\cancel{Toothache \Rightarrow Cavity.}$$

$$\cancel{Toothache \Rightarrow Cavity \lor GumProblem \lor Abscess \ldots}$$

$$\cancel{Cavity \Rightarrow Toothache.}$$

Why does logic fail for medical diagnosis?

- **Laziness:** It is too much work to list the complete set of antecedents or consequences needed to ensure an exceptionalness rule and too hard to use such rules
- **Theoretical ignorance:** Medical science has no complete theory for the domain
- **Practical ignorance:** Even if we know all the rules, we might be uncertain about a particular patient because not all the necessary tests have been or can be run

Probability Theory as an Alternative

Probability provides a way of summarizing the uncertainty that comes from our laziness and ignorance

We might not know what for sure afflicts a particular patient, but maybe there is an 80% chance that a patient with a toothache has a cavity

This belief could come from statistical data

Or it could come from general dental knowledge

Or it could come from some combo of evidence

Reminder: Probabilities

Example: Traffic on freeway

- Random variable: T = whether there's traffic
- Outcomes: T in $\{none, light, heavy\}$
- Distribution: $P(T = \text{none}) = 0.25$, $P(T = light) = 0.50$, $P(T = heavy) = 0.25$

0.25

0.50

0.25

Slide credit: Dan Klein and Pieter Abbeel, UC Berkeley CS 188

Reminder: Probabilities

Some laws of probability (more later):

- Probabilities are always non-negative
- Probabilities over all possible outcomes sum to one

As we get more evidence, probabilities may change:

- $P(T = heavy) = 0.25,\ P(T = heavy \mid Hour = 8am) = 0.60$
- We'll talk about methods for reasoning and updating probabilities later

0.60

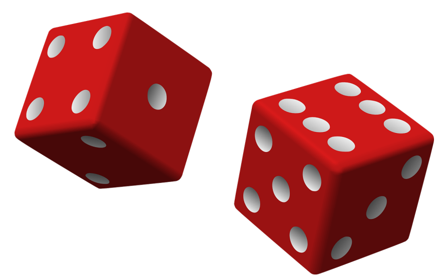
Slide credit: Dan Klein and Pieter Abbeel, UC Berkeley CS 188

Sample Space

The set of all possible worlds is called the *sample space*

The possible worlds are mutually exclusive and exhaustive

Example: rolling two dice



36 possible worlds: (1,1), (1,2), …, (6,6)

Probability Model

Ω = sample space ω = elements of the space

$$0 \leq P(\omega) \leq 1 \text{ for every } \omega \text{ and } \sum_{\omega \in \Omega} P(\omega) = 1$$

Probability of each possible world

Need to sum to 1

Propositions

Probabilistic assertions and queries are not usually about particular possible worlds, but about sets of them

We will call these sets **events**

The events are always described by propositions in a formal language

The probability associated with a proposition is defined to be the sum of the probabilities of the worlds in which it holds:

$$\text{For any proposition } \phi, P(\phi) = \sum_{\omega \in \phi} P(\omega)$$

Unconditional or Prior Probabilities

Probabilities such as $P(Total = 11)$  and $P(doubles)$  are called unconditional or prior probabilities

Such probabilities refer to degrees of belief in propositions in the **absence** of any other information

Image credit: https://freeonlinedice.com/

Conditional or Posterior Probabilities

Most of the time, we have some evidence that has already been revealed

Example: we have rolled two dice. The first die shows a 5, and we are waiting for a result from the second.  **?**

What is the probability of rolling doubles given the first die is a 5?

$$P(doubles \mid Die_1 = 5)$$

Conditional or Posterior Probabilities

Conditional probabilities are defined in terms of unconditional probabilities; for any propositions a and b, we have:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)}$$

Which holds true whenever $P(b) > 0$

Simpler form: product rule

$$P(a \wedge b) = P(a|b)P(b)$$

Bayes' Theorem

Bayes' Theorem

Two ways to factor a joint distribution over two variables:

$$P(x,y) = P(x|y)P(y) = P(y|x)P(x)$$

Dividing, we get:

$$P(x|y) = \frac{P(y|x)}{P(y)}P(x)$$

Slide credit: Dan Klein and Pieter Abbeel, UC Berkeley CS 188

Bayes' Theorem

Posterior **Likelihood** **Prior**

$$P(x|y) = \frac{P(y|x)}{P(y)} P(x)$$

Marginal Likelihood

Inference with Bayes' Rule

Example: Diagnostic probability from causal probability:

$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause})P(\text{cause})}{P(\text{effect})}$$

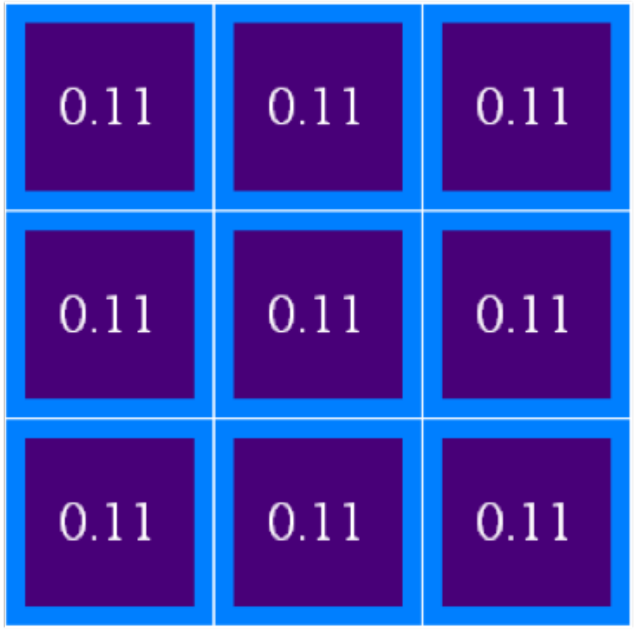
m: meningitis, s: stiff neck

$$\left.\begin{aligned} P(+m) &= 0.0001 \\ P(+s|+m) &= 0.8 \\ P(+s|-m) &= 0.01 \end{aligned}\right\} \text{Example givens}$$

Pacman as Ghostbuster

Let's say we have two distributions:

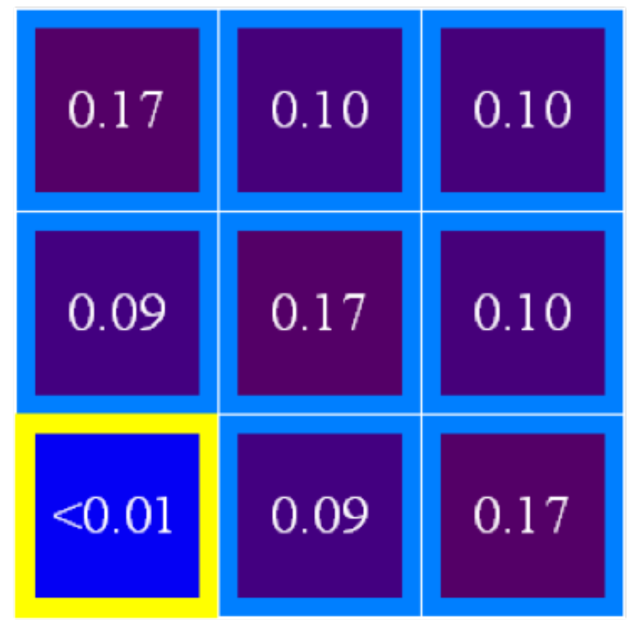
- Prior distribution over ghost location: $P(G)$
  - Let's say this is uniform
- Sensor reading model: $P(R \mid G)$
  - Given: we know what our sensors do
  - R = reading color measured at $(1,1)$
  - e.g., $P(R = \text{yellow} \mid G = (1,1)) = 0.1$



Pacman as Ghostbuster

We can calculate the posterior distribution $P(G|r)$ over ghost locations given a reading using Bayes' rule:

$$P(g|r) \propto P(r|g)P(g)$$



Conditional Independence

Conditional independence is our most basic and robust form of knowledge about uncertain environments.

X is conditionally independent of Y given Z $\quad X \perp\!\!\!\perp Y | Z$

If and only if:

$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

Or, equivalently, if and only if:

$$\forall x, y, z : P(x|z, y) = P(x|z)$$

Conditional Independence and the Chain Rule

Chain rule: $P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$

Trivial decomposition: $P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$

$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}, \text{Traffic})$$

With assumption of conditional independence:

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$

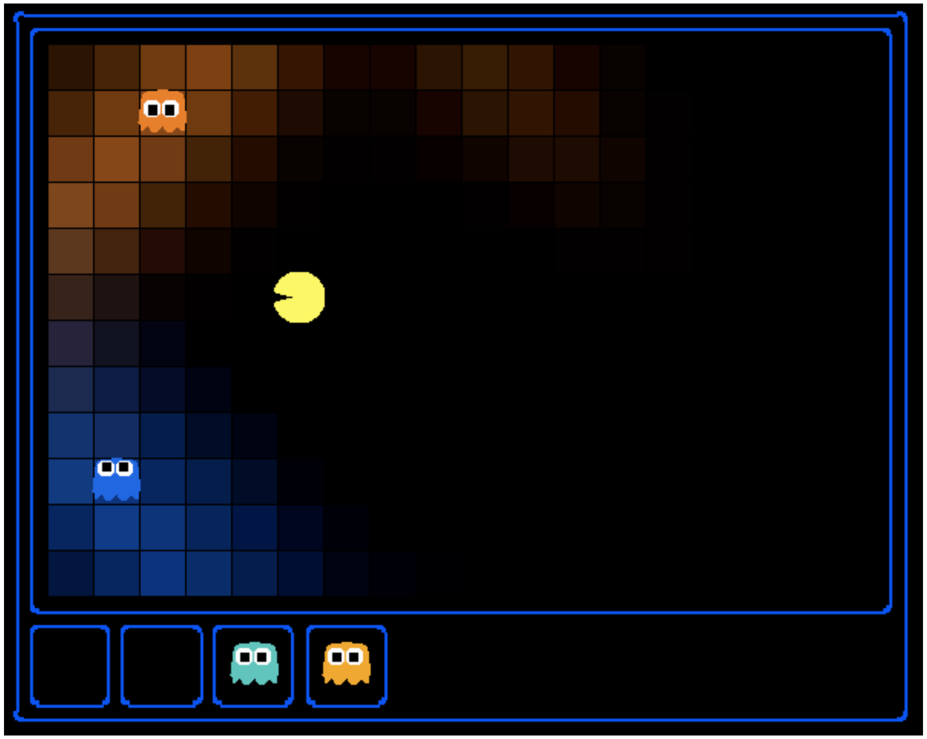
Bayes' nets / graphical models help us express conditional independence assumptions

Slide credit: Dan Klein and Pieter Abbeel, UC Berkeley CS 188

Pacman as Ghostbuster: The Chain Rule

Each sensor depends only on where the ghost is

That means, the two sensors are conditionally independent, given the ghost position



Slide credit: Dan Klein and Pieter Abbeel, UC Berkeley CS 188

Pacman as Ghostbuster: The Chain Rule

T: Top square is red

B: Bottom square is red

G: Ghost is in the top

$$P(\mathrm{T}, \mathrm{B}, \mathrm{G}) = P(\mathrm{G})\, P(\mathrm{T} \mid \mathrm{G})\, P(\mathrm{B} \mid \mathrm{G})$$

Givens:

$P(+g) = 0.5$

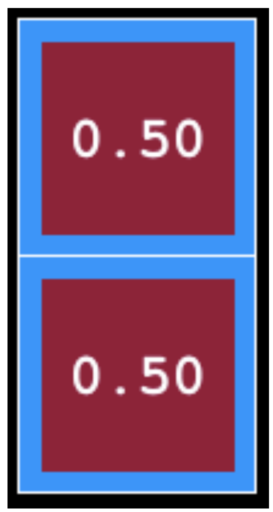
$P(-g) = 0.5$

$P(+t \mid +g) = 0.8$

$P(+t \mid -g) = 0.4$

$P(+b \mid +g) = 0.4$

$P(+b \mid -g) = 0.8$



T	B	G	P(T,B,G)
+t	+b	+g	0.16
+t	+b	-g	0.16
+t	-b	+g	0.24
+t	-b	-g	0.04
-t	+b	+g	0.04
-t	+b	-g	0.24
-t	-b	+g	0.06
-t	-b	-g	0.06

Slide credit: Dan Klein and Pieter Abbeel, UC Berkeley CS 188