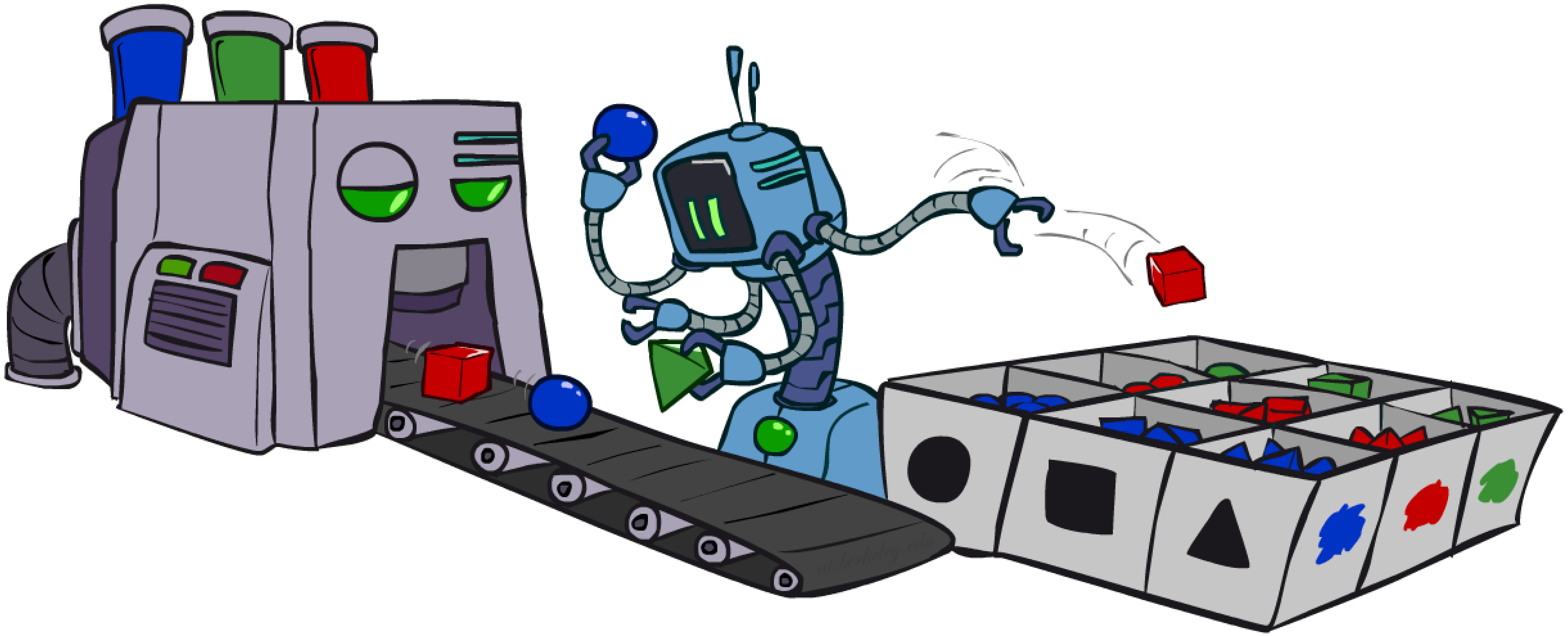
CS 188: Artificial Intelligence

Bayes' Nets: Sampling



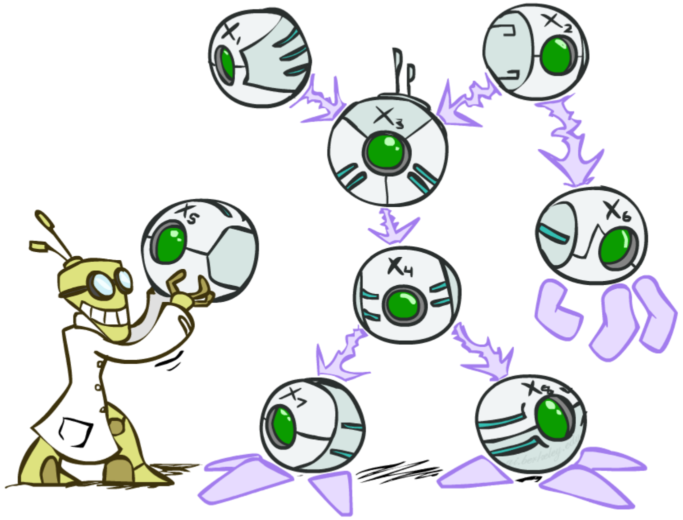
Instructors: Adam Janin, Josh Hug --- University of California, Berkeley

[These slides were created by Dan Klein, Pieter Abbeel, Anca Dragan, and Josh Hug. http://ai.berkeley.edu.]

Bayes' Net Representation

- A directed, acyclic graph, one node per random variable
- A conditional probability table (CPT) for each node
  - A collection of distributions over X, one for each combination of parents' values

$$P(X|a_1 \ldots a_n)$$



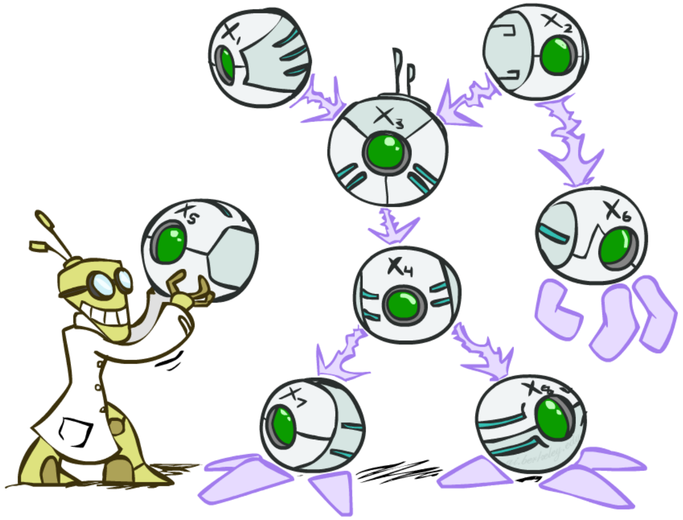
Bayes' Net Representation

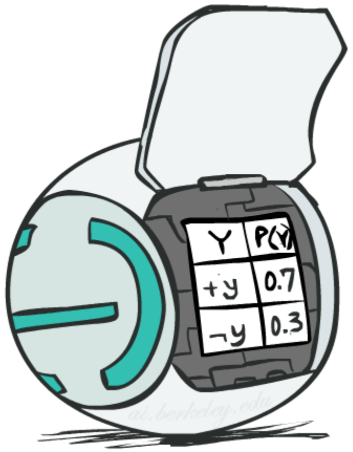
- Bayes' nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

  - Less complex than chain rule (valid for all distributions):

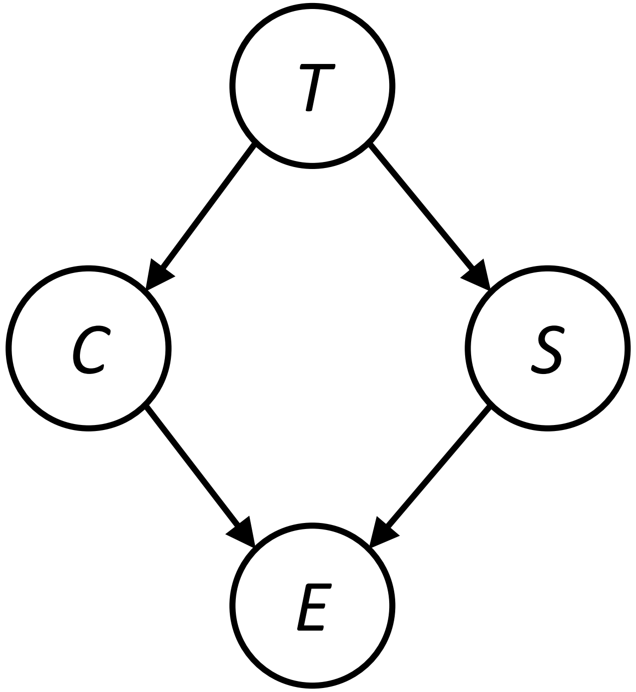
$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | x_1 \ldots x_{i-1})$$



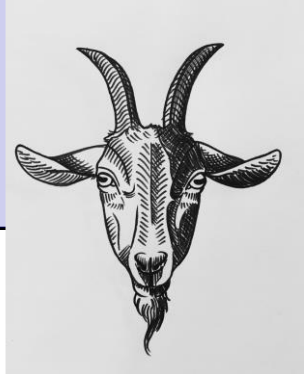


Example Model: Escaping a Dungeon

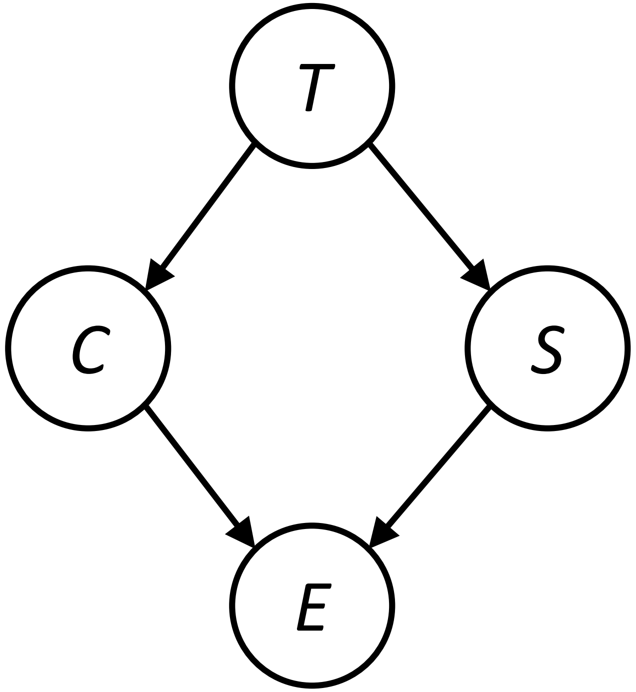
- If we take a treasure:
  - Cage trap may fall on us.
  - Snakes may be released.
  - These two events are determined independently.
- Our chance of escaping depends on presence of snakes and cages.



Example Model: Escaping a Dungeon



- If we take a treasure:
  - Cage trap may fall on us.
  - Snakes may be released.
  - These two events are determined independently.
- Our chance of escaping depends on presence of snakes and cages.

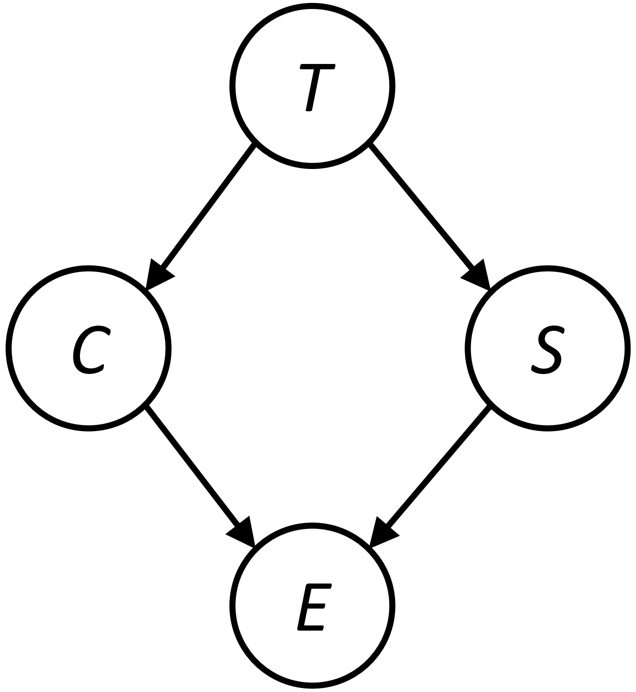


- **Question**: What four factors does our Bayes Net model provide?
  - Bonus question: How many rows do they have?

Example Model: Escaping a Dungeon

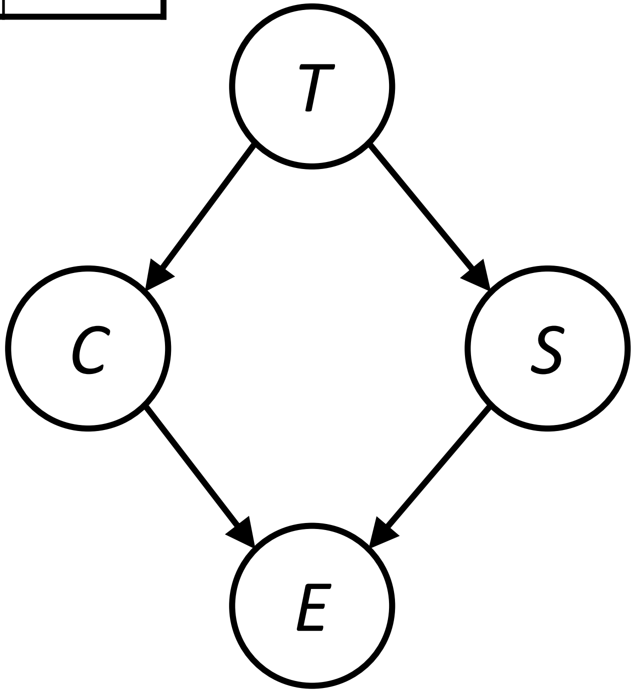
- If we take a treasure:
  - Cage trap may fall on us.
  - Snakes may be released.
  - These two events are determined independently.
- Our chance of escaping depends on presence of snakes and cages.

- **Question**: What four factors does our Bayes Net model provide?
  - P(T): Probability we **T**ake treasure.
  - P(C | T): Probability of **C**age falling on us given whether we take.
  - P(S | T): Probability of **S**nake release given whether we take.
  - P(E | C, S): Probability of **E**scaping given status of cage and snakes.



Example Model: Escaping a Dungeon

T	P(T)
+t	0.75
-t	0.25



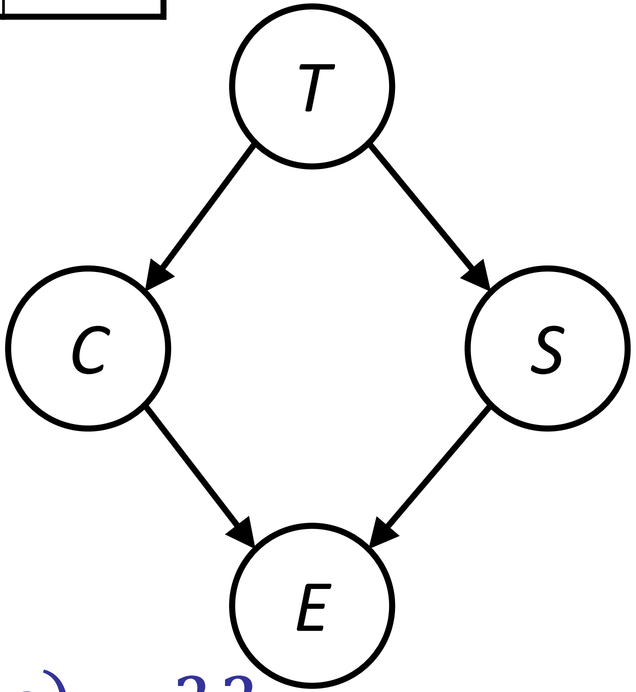
T	C	P(C|T)
+t	+c	0.95
+t	-c	0.05
-t	+c	0.0
-t	-c	1.0

T	S	P(S|T)
+t	+s	0.1
+t	-s	0.9
-t	+s	0.01
-t	-s	0.99

C	S	E	P(E|C,S)
+c	+s	+e	0.1
+c	+s	-e	0.9
+c	-s	+e	0.2
+c	-s	-e	0.8
-c	+s	+e	0.3
-c	+s	-e	0.7
-c	-s	+e	0.8
-c	-s	-e	0.2

Computing the Joint PDF

T	P(T)
+t	0.75
-t	0.25



T	C	P(C|T)
+t	+c	0.95
+t	-c	0.05
-t	+c	0.0
-t	-c	1.0

T	S	P(S|T)
+t	+s	0.1
+t	-s	0.9
-t	+s	0.01
-t	-s	0.99

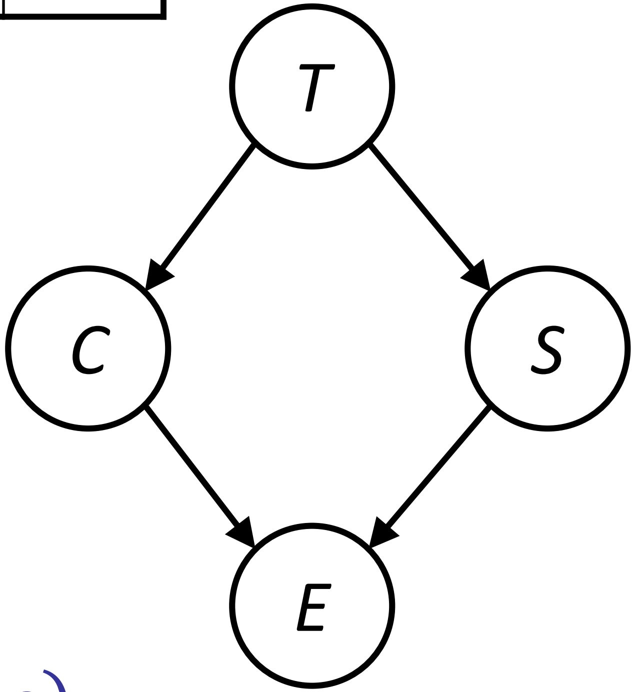
C	S	E	P(E|C,S)
+c	+s	+e	0.1
+c	+s	-e	0.9
+c	-s	+e	0.2
+c	-s	-e	0.8
-c	+s	+e	0.3
-c	+s	-e	0.7
-c	-s	+e	0.8
-c	-s	-e	0.2

- $P(+t, +c, -s, +e) = ??$

Computing the Joint PDF

T	P(T)
+t	0.75
-t	0.25

T	C	P(C|T)
+t	+c	0.95
+t	-c	0.05
-t	+c	0.0
-t	-c	1.0

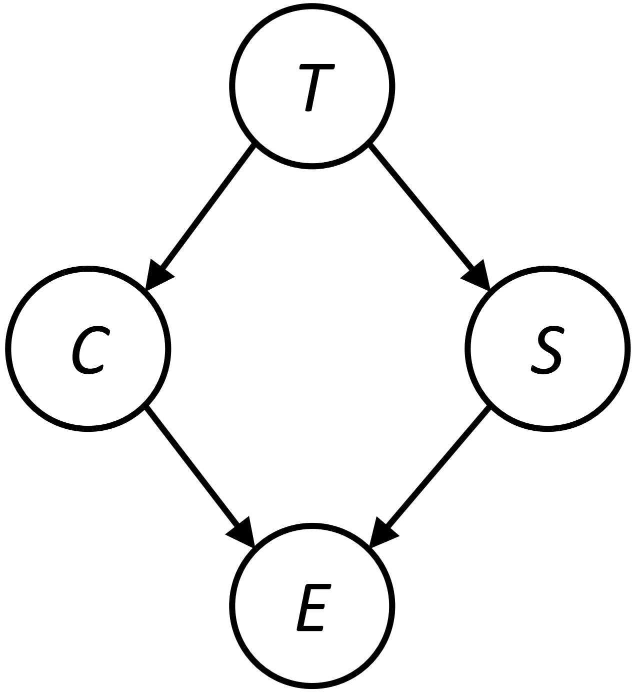


T	S	P(S|T)
+t	+s	0.1
+t	-s	0.9
-t	+s	0.01
-t	-s	0.99

C	S	E	P(E|C,S)
+c	+s	+e	0.1
+c	+s	-e	0.9
+c	-s	+e	0.2
+c	-s	-e	0.8
-c	+s	+e	0.3
-c	+s	-e	0.7
-c	-s	+e	0.8
-c	-s	-e	0.2

- $P(+t, +c, -s, +e)$
  $= P(+t)P(+c|+t)P(-s|+t)P(+e|+c, -s)$
  $= 0.75 \cdot 0.95 \cdot 0.9 \cdot 0.2 = 0.12825$

The Joint PDF

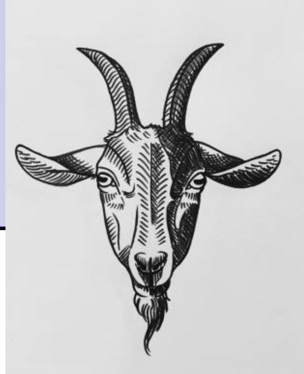


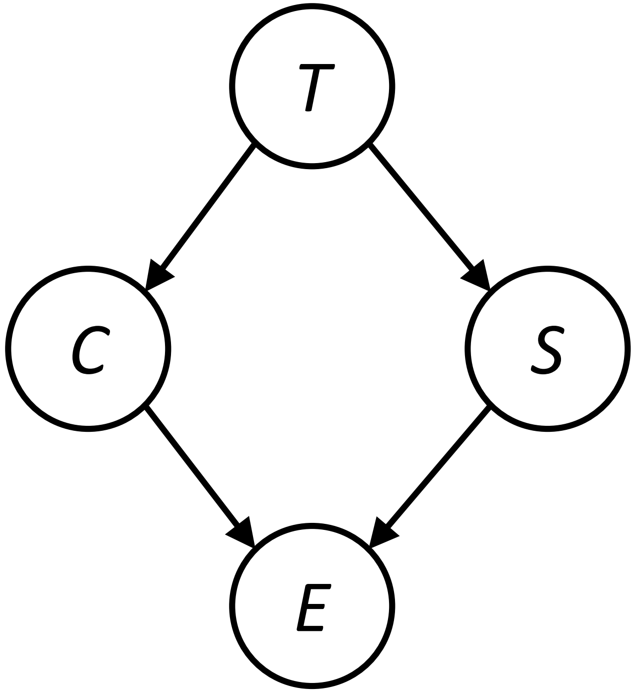
T	C	S	E	P(T, C, S, E)	T	C	S	E	P(T, C, S, E)
+t	+c	+s	+e	0.007125	-t	+c	+s	+e	0.0
+t	+c	+s	-e	0.064125	-t	+c	+s	-e	0.0
+t	+c	-s	+e	0.12825	-t	+c	-s	+e	0.0
+t	+c	-s	-e	0.513	-t	+c	-s	-e	0.0
+t	-c	+s	+e	0.001125	-t	-c	+s	+e	0.00075
+t	-c	+s	-e	0.002625	-t	-c	+s	-e	0.00175
+t	-c	-s	+e	0.0027	-t	-c	-s	+e	0.198
+t	-c	-s	-e	0.00675	-t	-c	-s	-e	0.0495

P(T, C, S, E): 16 row PDF

- $P(+t, +c, -s, +e)$
$$= P(+t)P(+c|+t)P(-s|+t)P(+e|+c, -s)$$
$$= 0.75 \cdot 0.95 \cdot 0.9 \cdot 0.2 = 0.12825$$

Inference by Enumeration



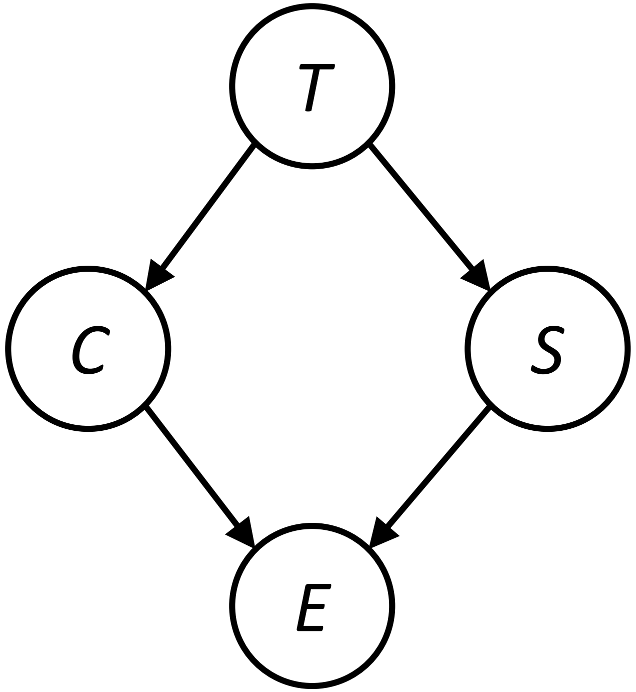


T	C	S	E	P(T, C, S, E)	T	C	S	E	P(T, C, S, E)
+t	+c	+s	+e	0.007125	-t	+c	+s	+e	0.0
+t	+c	+s	-e	0.064125	-t	+c	+s	-e	0.0
+t	+c	-s	+e	0.12825	-t	+c	-s	+e	0.0
+t	+c	-s	-e	0.513	-t	+c	-s	-e	0.0
+t	-c	+s	+e	0.001125	-t	-c	+s	+e	0.00075
+t	-c	+s	-e	0.002625	-t	-c	+s	-e	0.00175
+t	-c	-s	+e	0.0027	-t	-c	-s	+e	0.198
+t	-c	-s	-e	0.00675	-t	-c	-s	-e	0.0495

P(T, C, S, E): 16 row PDF

- Suppose we want to calculate P(T | +e)?
  - What is this? A number? A table? If a table, how big?
  - How would we compute it? (If webcast viewing, compute it!)

Inference by Enumeration



T	C	S	E	P(T, C, S, E)	T	C	S	E	P(T, C, S, E)
+t	+c	+s	+e	0.007125	-t	+c	+s	+e	0.0
+t	+c	+s	-e	0.064125	-t	+c	+s	-e	0.0
+t	+c	-s	+e	0.12825	-t	+c	-s	+e	0.0
+t	+c	-s	-e	0.513	-t	+c	-s	-e	0.0
+t	-c	+s	+e	0.001125	-t	-c	+s	+e	0.00075
+t	-c	+s	-e	0.002625	-t	-c	+s	-e	0.00175
+t	-c	-s	+e	0.027	-t	-c	-s	+e	0.198
+t	-c	-s	-e	0.00675	-t	-c	-s	-e	0.0495

P(T, C, S, E): 16 row PDF

- Suppose we want to calculate P(T | +e)?

Inference by Enumeration

T	C	S	E	P(T, C, S, E)	T	C	S	E	P(T, C, S, E)
+t	+c	+s	+e	0.007125	-t	+c	+s	+e	0.0
~~+t~~	~~+c~~	~~+s~~	~~-e~~	~~0.064125~~	~~-t~~	~~+c~~	~~+s~~	~~-e~~	~~0.0~~
+t	+c	-s	+e	0.12825	-t	+c	-s	+e	0.0
~~+t~~	~~+c~~	~~-s~~	~~-e~~	~~0.513~~	~~-t~~	~~+c~~	~~-s~~	~~-e~~	~~0.0~~
+t	-c	+s	+e	0.001125	-t	-c	+s	+e	0.00075
~~+t~~	~~-c~~	~~+s~~	~~-e~~	~~0.002625~~	~~-t~~	~~-c~~	~~+s~~	~~-e~~	~~0.00175~~
+t	-c	-s	+e	0.027	-t	-c	-s	+e	0.198
~~+t~~	~~-c~~	~~-s~~	~~-e~~	~~0.00675~~	~~-t~~	~~-c~~	~~-s~~	~~-e~~	~~0.0495~~

P(T, C, S, E): 16 row PDF

$$P(T\,|+e) = \frac{P(T,+e)}{P(+t,+e)+P(-t,+e)}$$

$$P(+t|+e) = \frac{0.007125\ +\ 0.12825\ +\ 0.001125\ +\ 0.027}{0.007125\ +\ 0.12825\ +\ 0.001125\ +\ 0.027\ +\ 0.00075\ +\ 0.198}$$

- Suppose we want to calculate P(T | +e)?
  - Keep only the rows consistent with the “evidence”: +e.
  - Sum out hidden variables S and C, then normalize.

T	P(T| +e)
+t	0.45135
-t	0.54865

Inference by Enumeration (what we just did)

- General case:
  - Evidence variables: $E_1 \ldots E_k = e_1 \ldots e_k$
  - Query* variable: Q
  - Hidden variables: $H_1 \ldots H_r$

$$\left.\begin{array}{l} E_1 \ldots E_k = e_1 \ldots e_k \\ Q \\ H_1 \ldots H_r \end{array}\right\} X_1, X_2, \ldots X_n$$

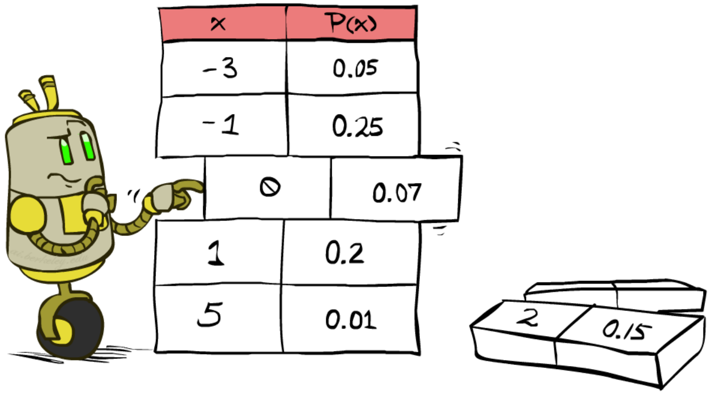
All variables

- We want:

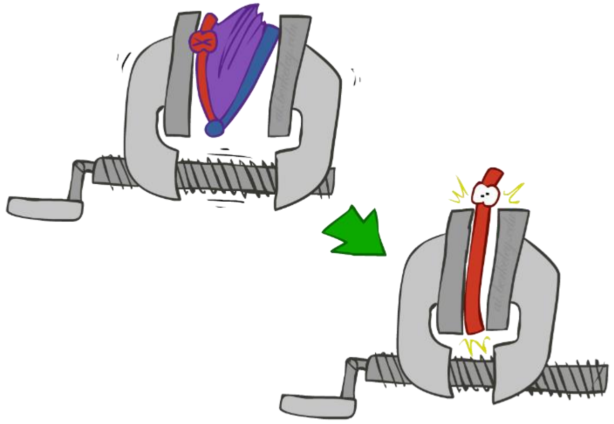
** Works fine with multiple query variables, too*

$$P(Q|e_1 \ldots e_k)$$

- Step 1: Select the entries consistent with the evidence



- Step 2: Sum out H to get joint of Query and evidence



$$P(Q, e_1 \ldots e_k) = \sum_{h_1 \ldots h_r} \underbrace{P(Q, h_1 \ldots h_r, e_1 \ldots e_k)}_{X_1, X_2, \ldots X_n}$$

- Step 3: Normalize

$$\times \frac{1}{Z}$$

$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

$$P(Q|e_1 \cdots e_k) = \frac{1}{Z} P(Q, e_1 \cdots e_k)$$

Inference by Enumeration and the Joint PDF

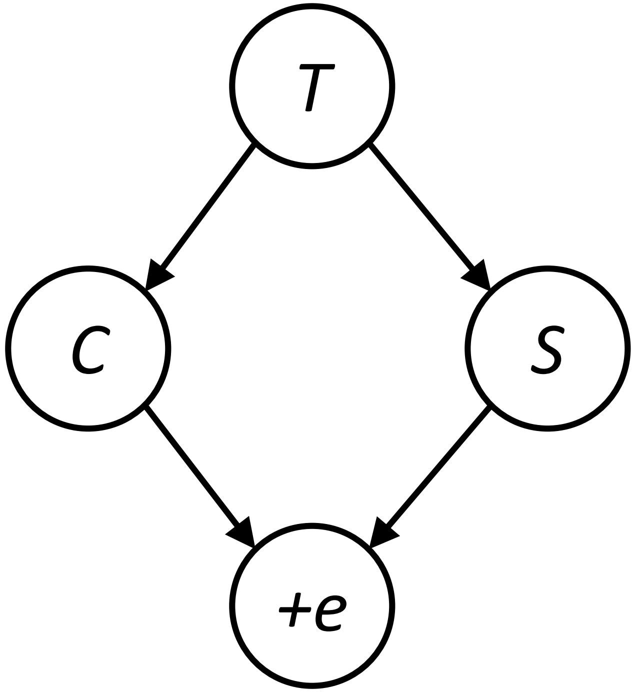
- The joint PDF lets us answer any probabilistic inference question trivially using "inference by enumeration" or IBE.
  - IBE is dumb, but powerful.
  - IBE is computationally too expensive for practical use (exponentially large joint PDF).
- Variable Elimination is an alternate approach (from last time).
  - Avoids needs to compute joint PDF.

"They who control the joint PDF, control the universe."

Inference on the Joint PDF

- Variable Elimination is an alternate approach (from last time).
  - Avoids needs to compute joint PDF.
  - Basic idea: Interleave “joins” with “sums”.
    - Ideally: Smaller maximum table size.
    - In the worst case, no savings.

Computing P(T | +e) using Variable Elimination



$\boxed{P(T)}$ $\boxed{P(S|T)}$ $\boxed{P(C|T)}$ $\boxed{P(+e|C,S)}$

Eliminate C (two step process)

multiply

$\boxed{P(+e,C|T,S)}$

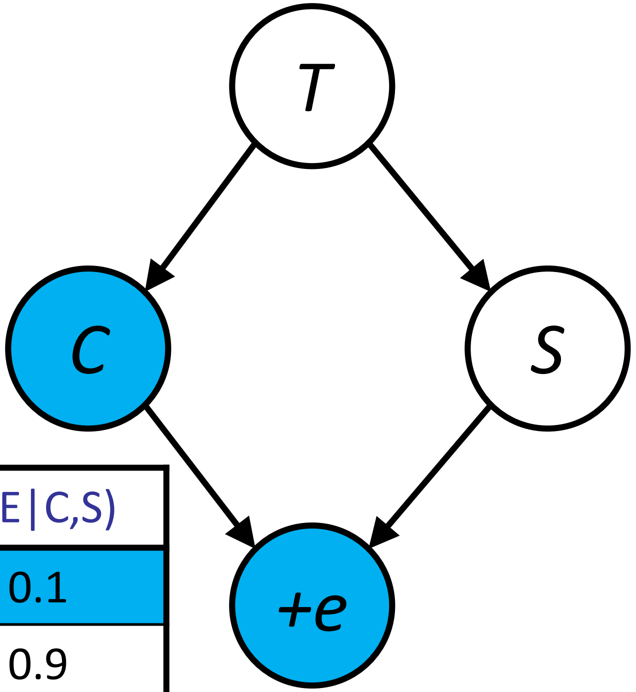
sum over c

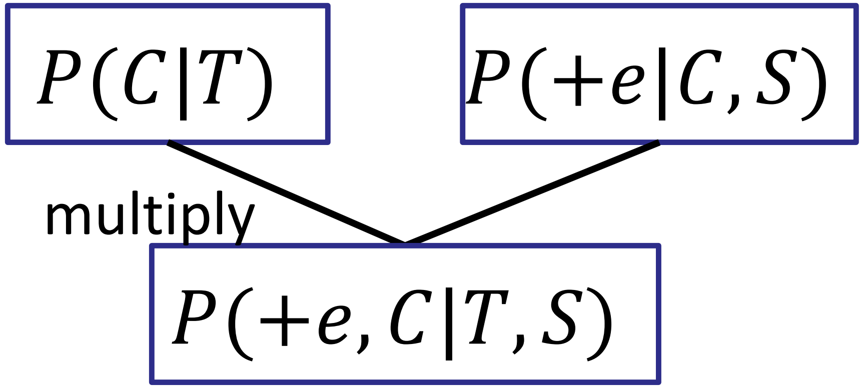
$\boxed{P(+e|T,S)}$

$\boxed{P(T)}$ $\boxed{P(S|T)}$ $\overset{f_1}{\boxed{P(+e|T,S)}}$

Computing P(T | +e) using Variable Elimination

T	C	P(C|T)
+t	+c	0.95
+t	-c	0.05
-t	+c	0.0
-t	-c	1.0



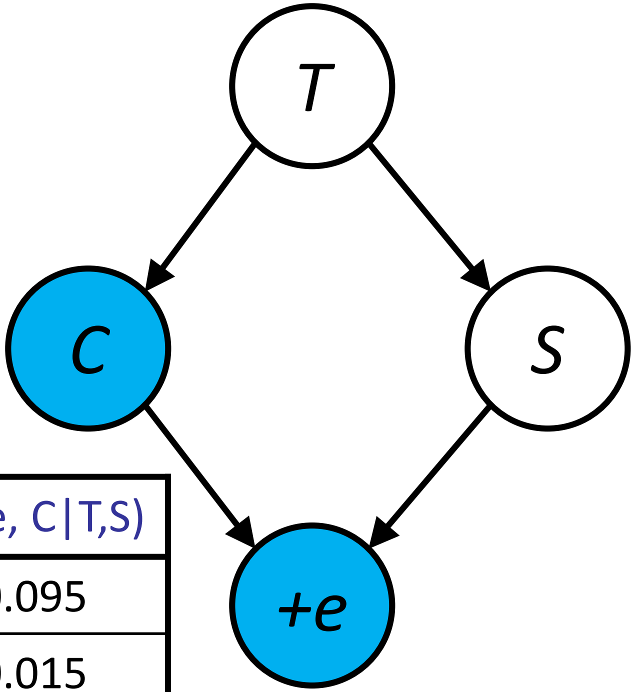


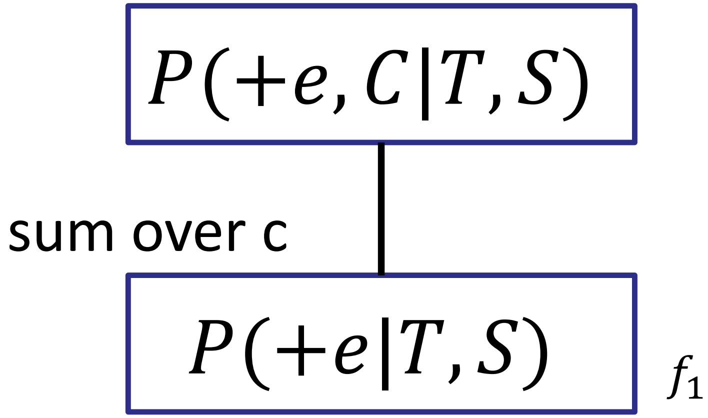
C	S	E	P(E|C,S)
+c	+s	+e	0.1
+c	+s	-e	0.9
+c	-s	+e	0.2
+c	-s	-e	0.8
-c	+s	+e	0.3
-c	+s	-e	0.7
-c	-s	+e	0.8
-c	-s	-e	0.2

P(+e, C|T,S) does not sum to anything in particular.
This is OK, it's just a "Type 3" factor.

T	S	C	P(+e, C|T,S)
+t	+s	+c	0.095
+t	+s	-c	0.015
+t	-s	+c	0.19
+t	-s	-c	0.04
-t	+s	+c	0.0
-t	+s	-c	0.3
-t	-s	+c	0.0
-t	-s	-c	0.8

Computing P(T | +e) using Variable Elimination





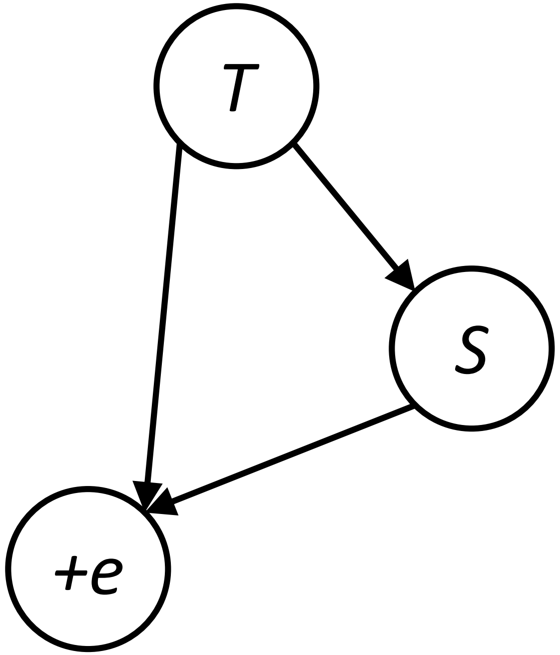
T	S	C	P(+e, C|T,S)
+t	+s	+c	0.095
+t	+s	-c	0.015
+t	-s	+c	0.19
+t	-s	-c	0.04
-t	+s	+c	0.0
-t	+s	-c	0.3
-t	-s	+c	0.0
-t	-s	-c	0.8

sum over c →

f_1

T	S	P(+e|T,S)
+t	+s	0.11
+t	-s	0.23
-t	+s	0.3
-t	-s	0.8

Computing P(T | +e) using Variable Elimination



$P(T)$ $P(S|T)$ $P(+e|T,S)$

f_1

Eliminate S (two step process)

multiply

$P(+e,S|T)$

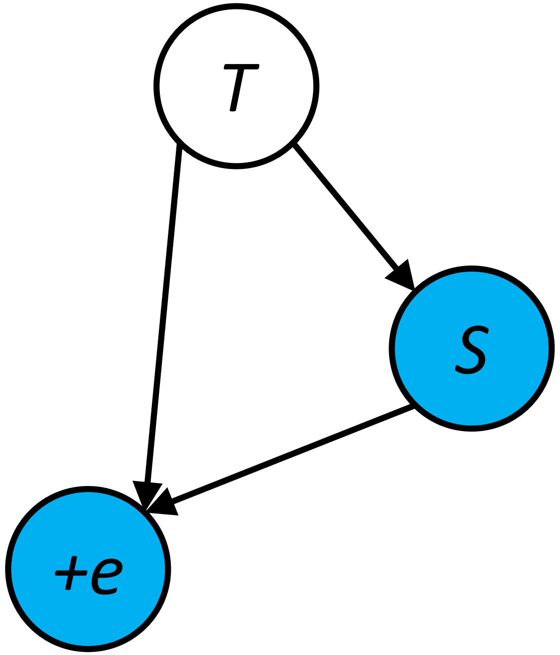
sum over s

$P(+e|T)$

f_2

$P(T)$ $P(+e|T)$

Computing P(T | +e) using Variable Elimination



$P(T)$

$P(S|T)$ $\quad$ $P(+e|T,S)$ $\quad f_1$

multiply

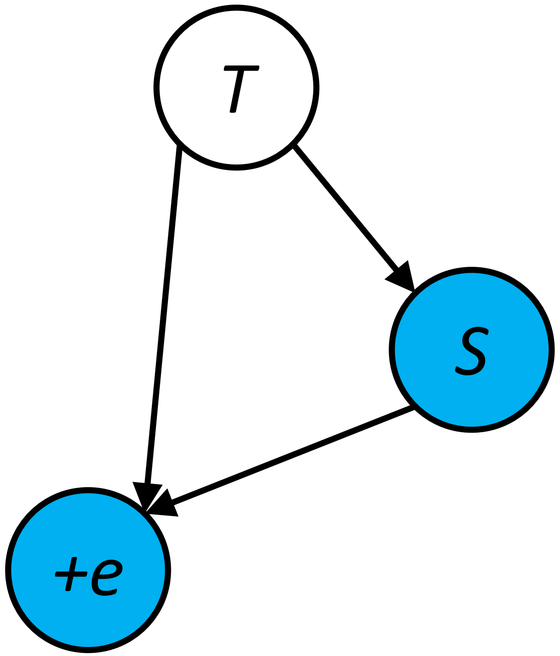
$P(+e,S|T)$

T	S	P(S|T)
+t	+s	0.1
+t	-s	0.9
-t	+s	0.01
-t	-s	0.99

T	S	P(+e|T,S)
+t	+s	0.11
+t	-s	0.23
-t	+s	0.3
-t	-s	0.8

T	S	P(+e, S|T)
+t	+s	0.011
+t	-s	0.207
-t	+s	0.003
-t	-s	0.792

Computing P(T | +e) using Variable Elimination



$\boxed{P(T)}$

$\boxed{P(+e, S|T)}$

sum over s

$\boxed{P(+e|T)}$ f_2

T	S	P(S|T)
+t	+s	0.1
+t	-s	0.9
-t	+s	0.01
-t	-s	0.99

T	S	P(+e|T,S)
+t	+s	0.11
+t	-s	0.23
-t	+s	0.3
-t	-s	0.8

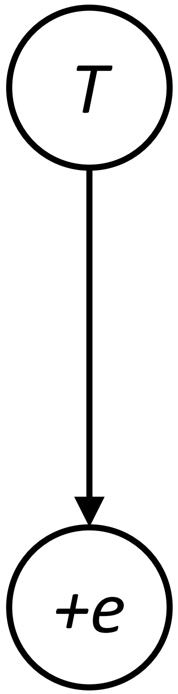
T	S	P(+e, S|T)
+t	+s	0.011
+t	-s	0.207
-t	+s	0.003
-t	-s	0.792

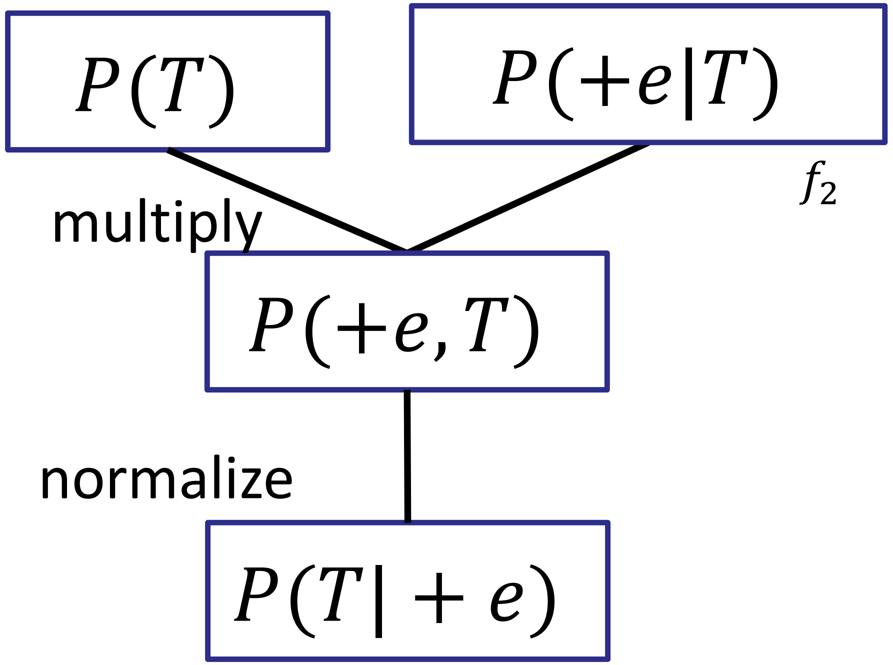
sum over s

f_2

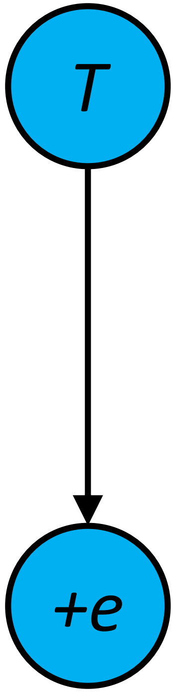
T	E	P(+e|T)
+t	+e	0.218
-t	+e	0.795

Computing P(T | +e) using Variable Elimination



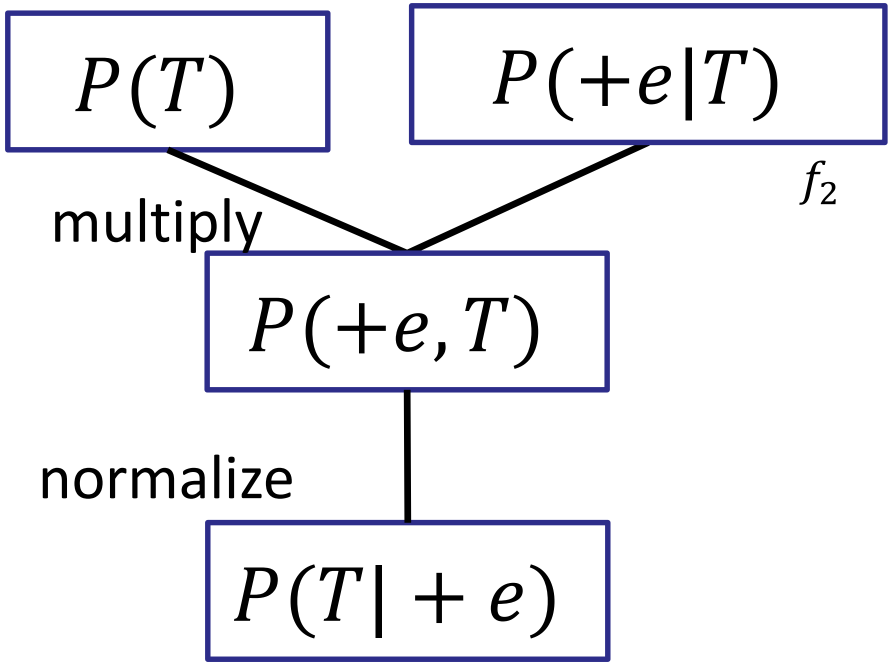


Computing P(T | +e) using Variable Elimination



T	P(T)
+t	0.75
-t	0.25

T	E	P(+e|T)
+t	+e	0.218
-t	+e	0.795



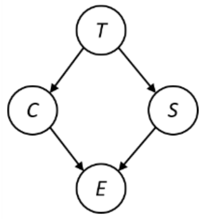
T	P(+e, T)
+t	0.1635
-t	0.19875

normalize →

T	P(T | +e)
+t	0.45134
-t	0.54865

Handy for debugging lectures!

IBE vs. Variable Elimination



calculate joint pdf

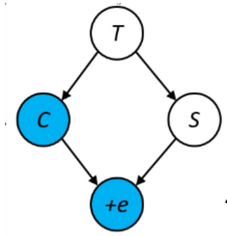
T	C	S	E	P(T, C, S, E)	T	C	S	E	P(T, C, S, E)
+t	+c	+s	+e	0.007125	-t	+c	+s	+e	0.0
+t	+c	+s	-e	0.064125	-t	+c	+s	-e	0.0
+t	+c	-s	+e	0.12825	-t	+c	-s	+e	0.0
+t	+c	-s	-e	0.513	-t	+c	-s	-e	0.0
+t	-c	+s	+e	0.001125	-t	-c	+s	+e	0.00075
+t	-c	+s	-e	0.002625	-t	-c	+s	-e	0.00175
+t	-c	-s	+e	0.027	-t	-c	-s	+e	0.198
+t	-c	-s	-e	0.00675	-t	-c	-s	-e	0.0495

Select consistent entries

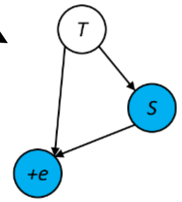
T	C	S	E	P(T, C, S, E)	T	C	S	E	P(T, C, S, E)
+t	+c	+s	+e	0.007125	-t	+c	+s	+e	0.0
~~+t~~	~~+c~~	~~+s~~	~~-e~~	~~0.064125~~	~~-t~~	~~+c~~	~~+s~~	~~-e~~	~~0.0~~
+t	+c	-s	+e	0.12825	-t	+c	-s	+e	0.0
~~+t~~	~~+c~~	~~-s~~	~~-e~~	~~0.513~~	~~-t~~	~~+c~~	~~-s~~	~~-e~~	~~0.0~~
+t	-c	+s	+e	0.001125	-t	-c	+s	+e	0.00075
~~+t~~	~~-c~~	~~+s~~	~~-e~~	~~0.002625~~	~~-t~~	~~-c~~	~~+s~~	~~-e~~	~~0.00175~~
+t	-c	-s	+e	0.027	-t	-c	-s	+e	0.198
~~+t~~	~~-c~~	~~-s~~	~~-e~~	~~0.00675~~	~~-t~~	~~-c~~	~~-s~~	~~-e~~	~~0.0495~~

Sum out hidden variables
and normalize

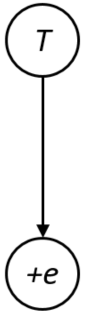
T	P(T | +e)
+t	0.45134
-t	0.54865



Eliminate C



Eliminate S



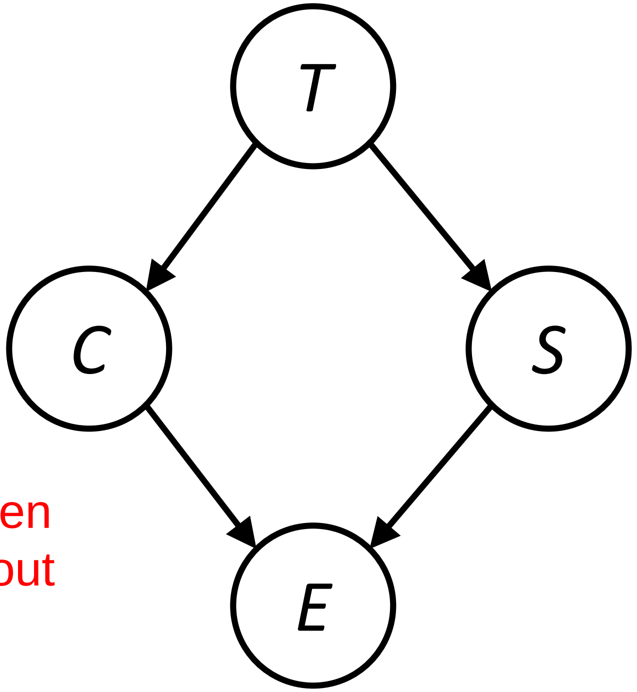
T	P(T | +e)
+t	0.45134
-t	0.54865

Inference by Enumeration (alternate view)

Joint PDF with hidden variables summed out

$$P(T \mid +e) = \frac{P(T,+e)}{\sum_t P(t,+e)}$$

Normalization factor Z



$$P(T,+e) = \sum_{c,s} P(T,c,s,+e)$$

Joint PDF with hidden variables summed out

$$= \sum_{s} \sum_{c} P(T)P(s|T)P(c|T)P(+e|c,s)$$

Variable Elimination(alternate view)

$$P(T \mid +e) = \frac{P(T,+e)}{\sum_t P(t,+e)}$$

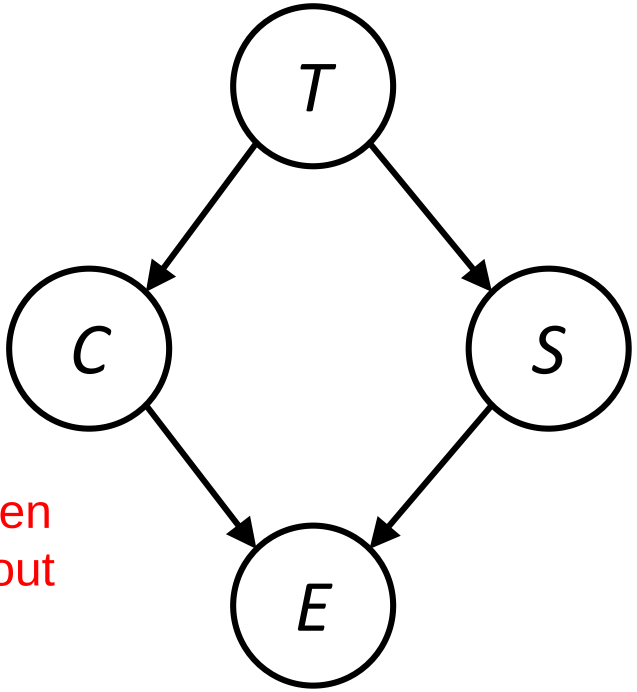
Joint PDF with hidden variables summed out

Normalization factor Z

$$P(T,+e) = \sum_{c,s} P(T,c,s,+e)$$

Joint PDF with hidden variables summed out

$$= P(T) \sum_{s} P(s|T) \overbrace{\sum_{c} P(c|T) P(+e|c,s)}^{f_1}$$



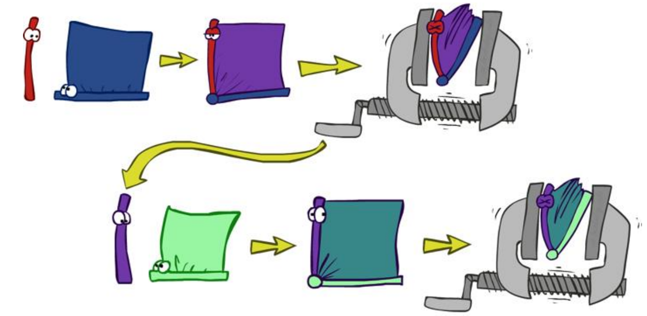
Variable Elimination

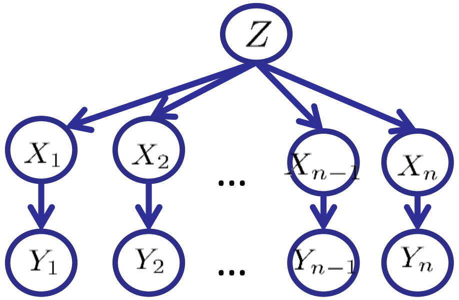
- Interleave joining and marginalizing

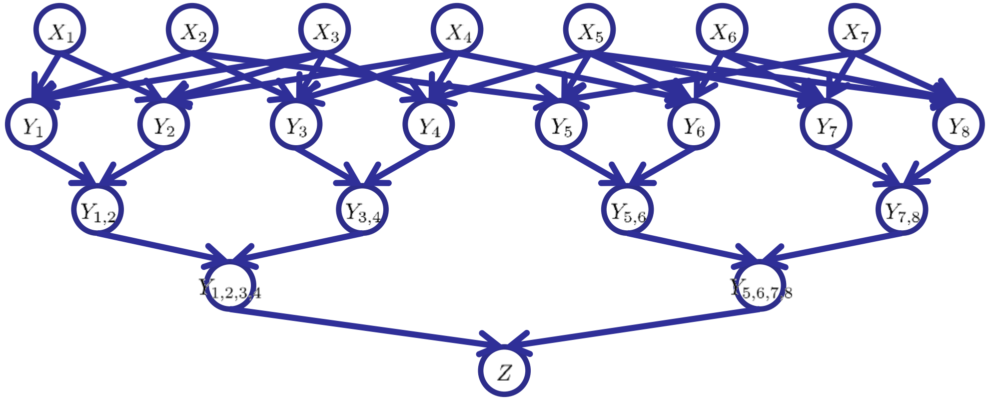
- d^k entries computed for a factor over k variables with domain sizes d

- Ordering of elimination of hidden variables can affect size of factors generated

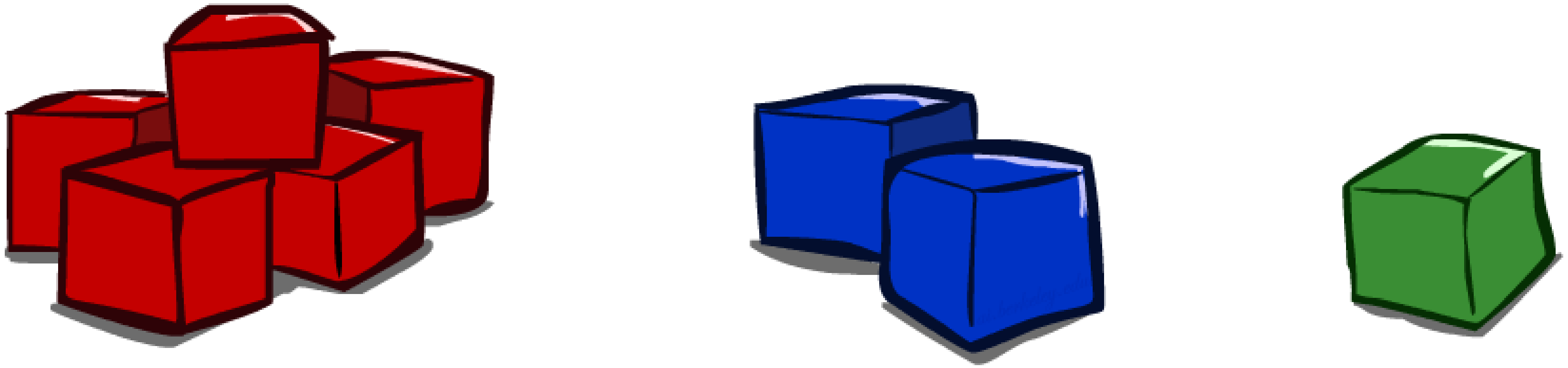
- Worst case: running time exponential in the size of the Bayes' net







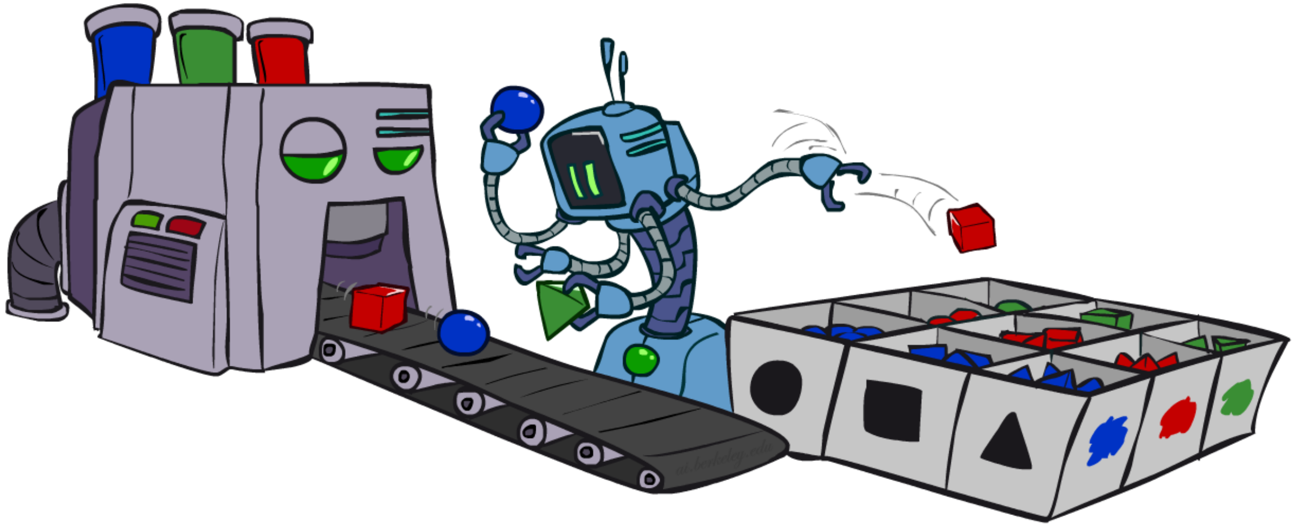
Approximate Inference: Sampling



Sampling

- Basic idea
  - Draw N samples from a sampling distribution S
  - Compute an approximate posterior probability
  - Show this converges to the true probability P

- Why sample?
  - Inference: getting a sample is faster than computing the right answer (e.g. with variable elimination)



Sampling Basics

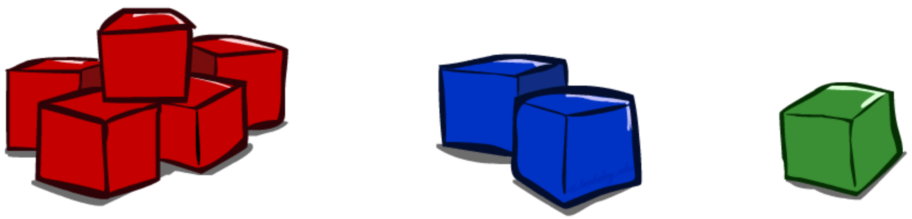
- Sampling from given distribution
  - Step 1: Get sample u from uniform distribution over [0, 1)
    - E.g. random() in python
  - Step 2: Convert this sample u into an outcome for the given distribution by having each outcome associated with a sub-interval of [0,1) with sub-interval size equal to probability of the outcome

- Example

C	P(C)
red	0.6
green	0.1
blue	0.3

$$0 \leq u < 0.6, \rightarrow C = red$$
$$0.6 \leq u < 0.7, \rightarrow C = green$$
$$0.7 \leq u < 1, \rightarrow C = blue$$

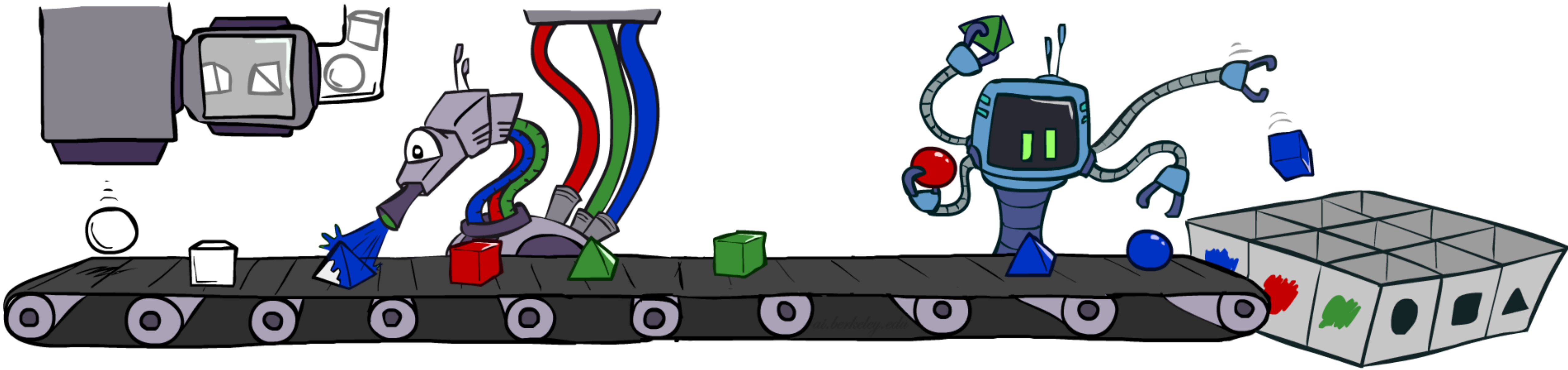
- If random() returns u = 0.83, then our sample is C = blue
- E.g, after sampling 8 times:



Sampling in Bayes’ Nets

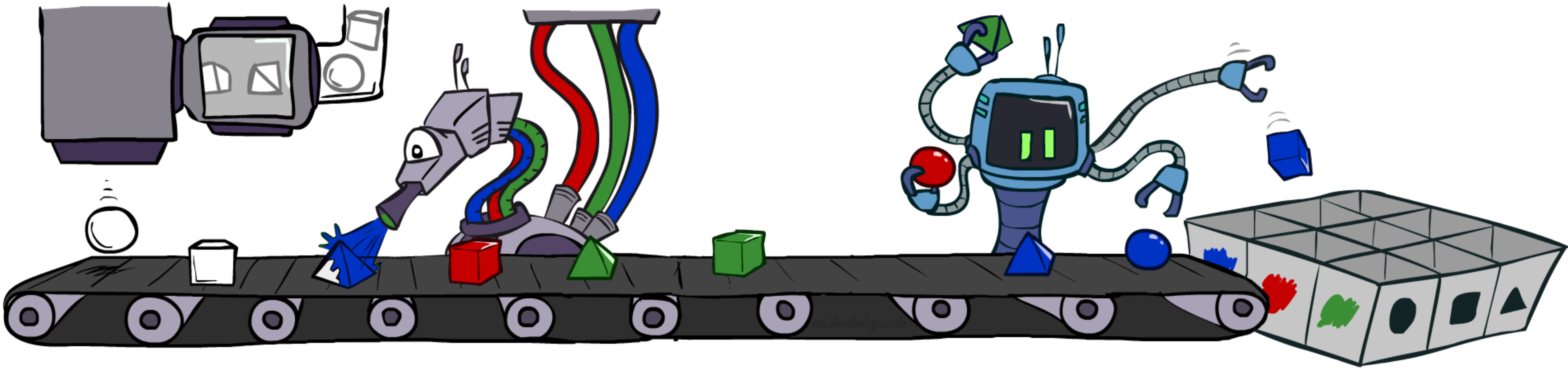
- Prior Sampling
- Rejection Sampling
- Likelihood Weighting
- Gibbs Sampling

Prior Sampling



Prior Sampling

- Ignore evidence. Sample from the joint probability.
- Do inference by counting the right samples.



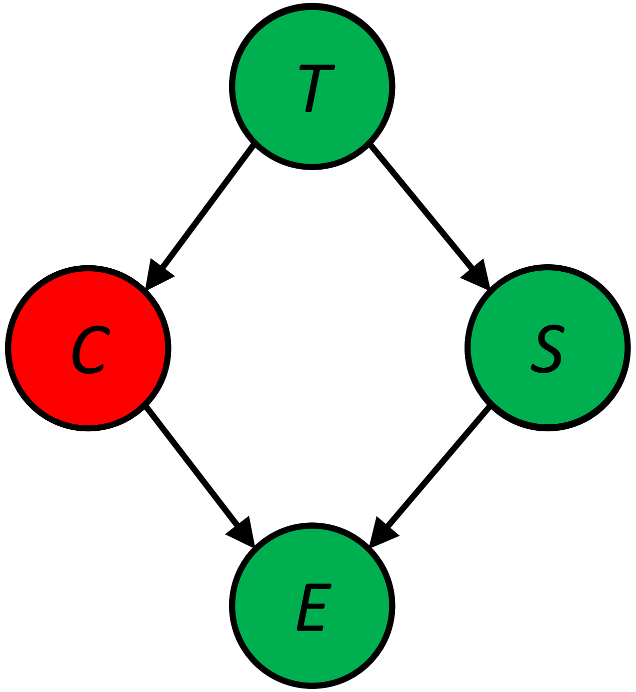
Example Model: Escaping a Dungeon

$P(T)$

+t	0.75
-t	0.25

$P(C|T)$

+t	+c	0.95
+t	-c	0.05
-t	+c	0.0
-t	-c	1.0



$P(S|T)$

+t	+s	0.1
+t	-s	0.9
-t	+s	0.01
-t	-s	0.99

$P(E|C,S)$

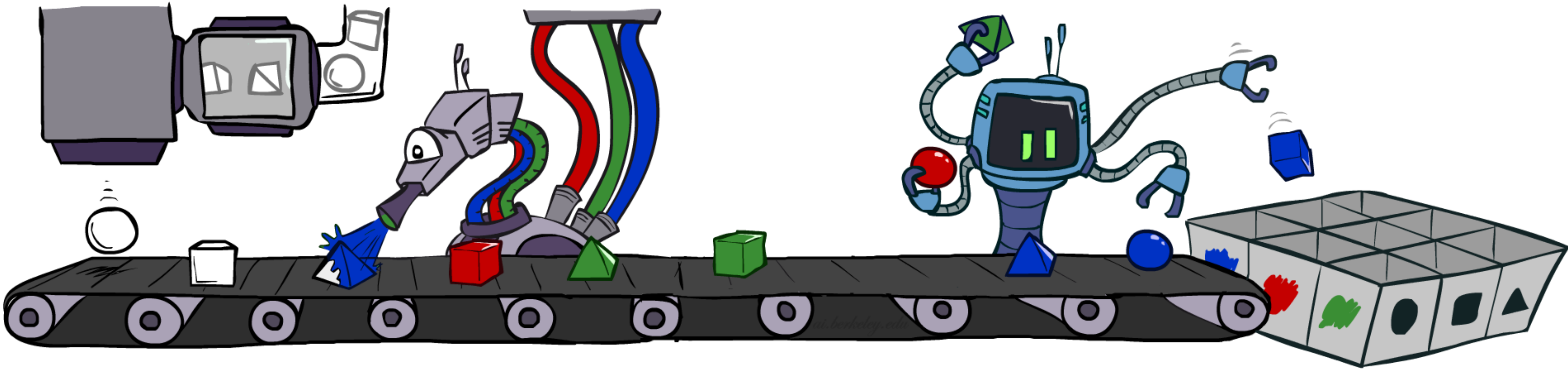
+c	+s	+e	0.1
+c	+s	-e	0.9
+c	-s	+e	0.2
+c	-s	-e	0.8
-c	+s	+e	0.3
-c	+s	-e	0.7
-c	-s	+e	0.8
-c	-s	-e	0.2

Samples:

```
+t, -c, +s, +e
-t, -c, -s, -e,
...
```

Prior Sampling

- For i=1, 2, …, n
  - Sample x_i from $P(X_i \mid Parents(X_i))$
- Return $(x_1, x_2, \ldots, x_n)$



Example

- We' ll get a bunch of samples from the BN:

  +t, -c, +s, +e

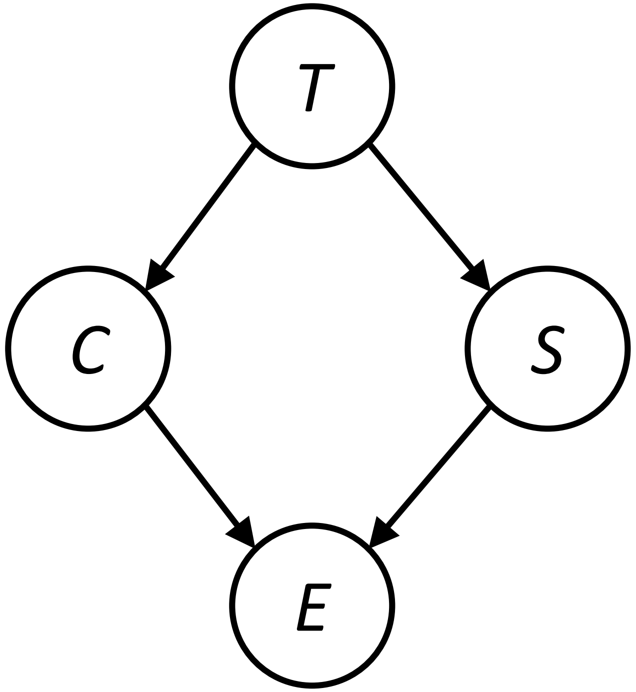
  +t, +c, +s, +e

  -t, -c, +s, -e

  +t, -c, +s, +e

  -t, -c, -s, +e

- If we want to know P(E)
  - We have counts <+e:4, -e:1>
  - Normalize to get P(E) = <+e:0.8, -e:0.2>
  - This will get closer to the true distribution with more samples
  - Can estimate anything else, too
  - What about P(T| +e)? P(T| +s, +e)? P(S| -t, +c)?
  - Fast: can use fewer samples to save time (what's the drawback?)



Prior Sampling Analysis

- Two things we'd like to prove:
  - Proof 1: The samples drawn from the right distribution.
  - Proof 2: A normalized count of samples from a distribution provides a good estimate for an event.

Prior Sampling Analysis

Proof 1

- This process generates samples with probability:

$$S_{PS}(x_1 \ldots x_n) = \prod_{i=1}^{n} P(x_i|\text{Parents}(X_i)) = P(x_1 \ldots x_n)$$

…i.e. the BN's joint probability

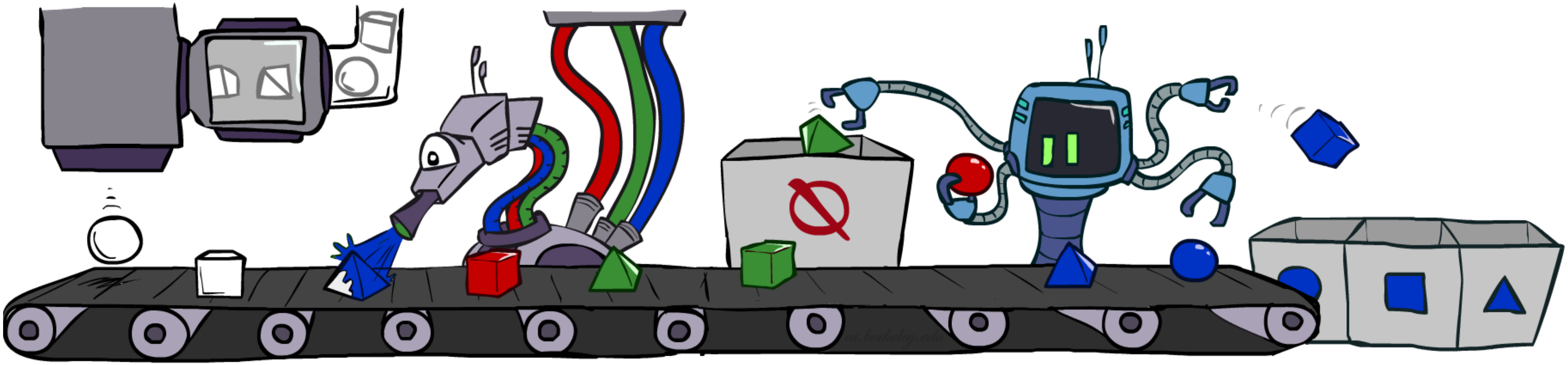
Proof 2

- Let the number of samples of an event be $N_{PS}(x_1 \ldots x_n)$

- Then
$$\begin{aligned} \lim_{N\to\infty} \widehat{P}(x_1, \ldots, x_n) &= \lim_{N\to\infty} N_{PS}(x_1, \ldots, x_n)/N \\ &= S_{PS}(x_1, \ldots, x_n) \\ &= P(x_1 \ldots x_n) \end{aligned}$$

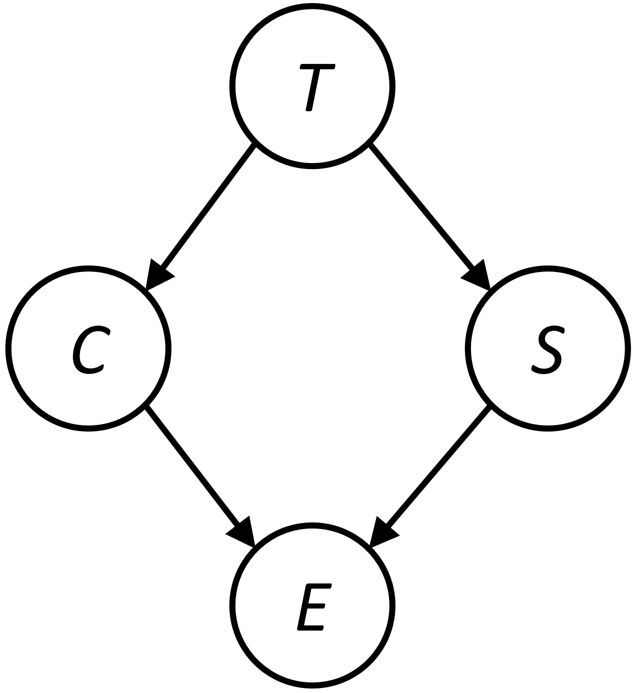
- I.e., together these proofs tell us the sampling procedure is consistent

Rejection Sampling



Rejection Sampling

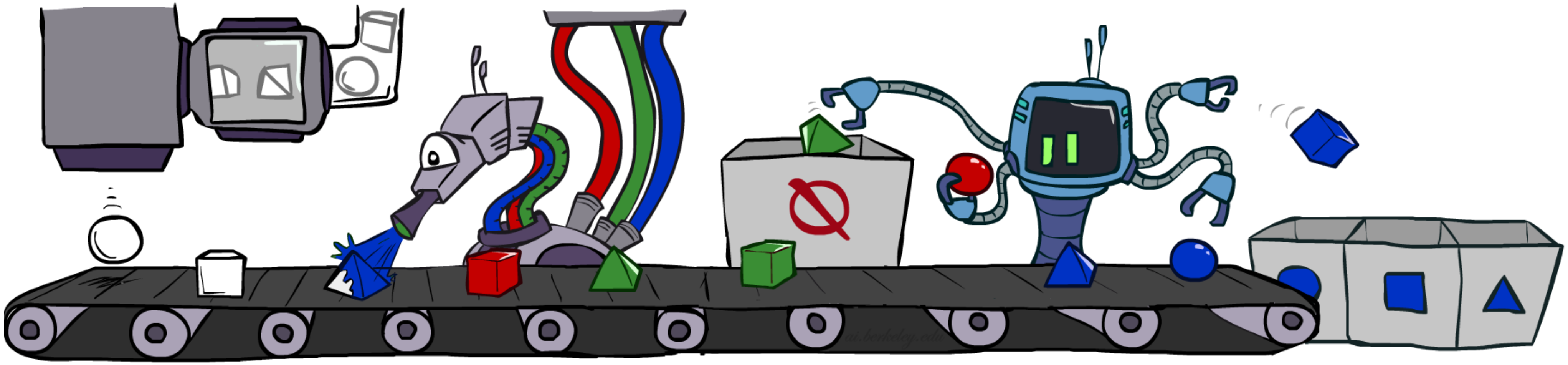
- Let’s say we want P(T| +c)
  - Tally T outcomes, but ignore (reject) samples which don’t have C=+c
  - This is called rejection sampling
  - It is also consistent for conditional probabilities (i.e., correct in the limit)



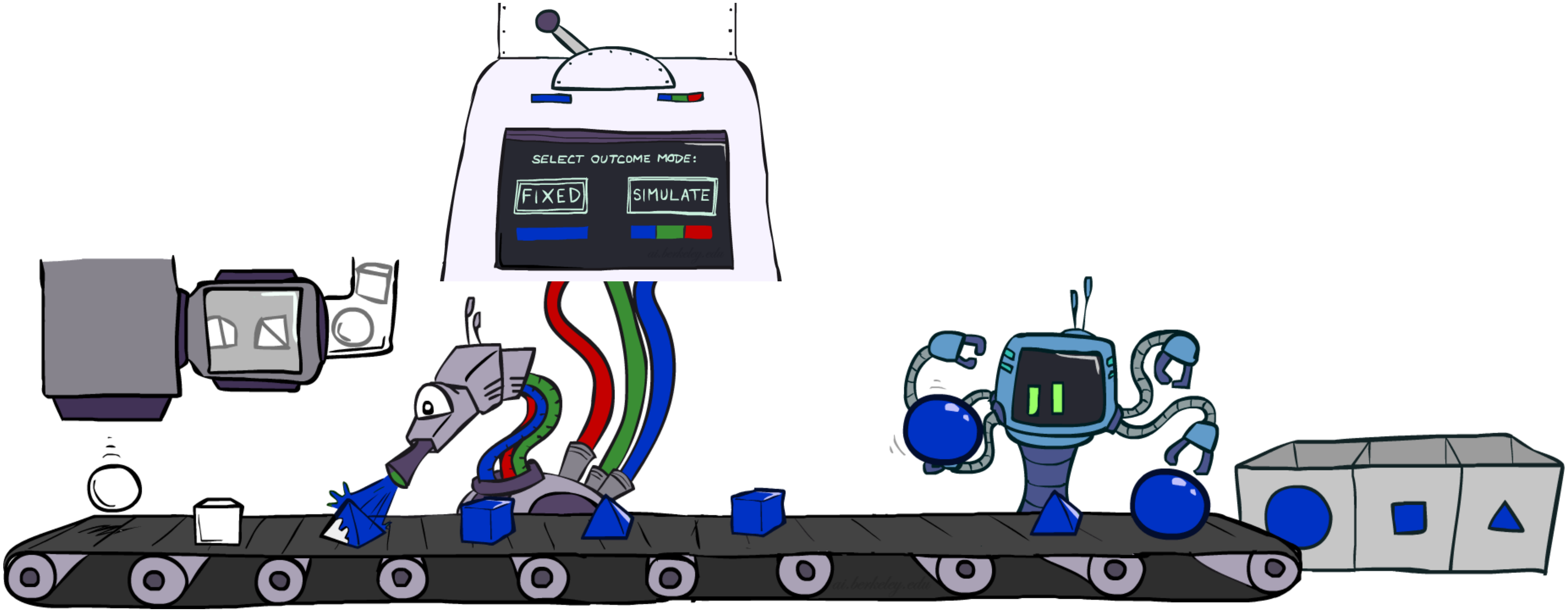
~~+t, -c~~
+t, +c, +s, +e
~~-t, -c~~
+t, +c, -s, +e
~~+t, -c~~

Rejection Sampling

- IN: evidence instantiation
- For i=1, 2, …, n
  - Sample x_i from $P(X_i \mid Parents(X_i))$
  - If x_i not consistent with evidence
    - Reject: Return, and no sample is generated in this cycle
- Return $(x_1, x_2, …, x_n)$

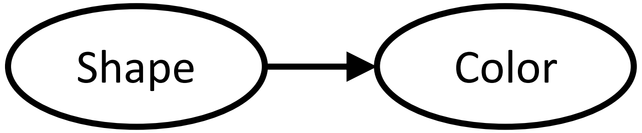


Likelihood Weighting

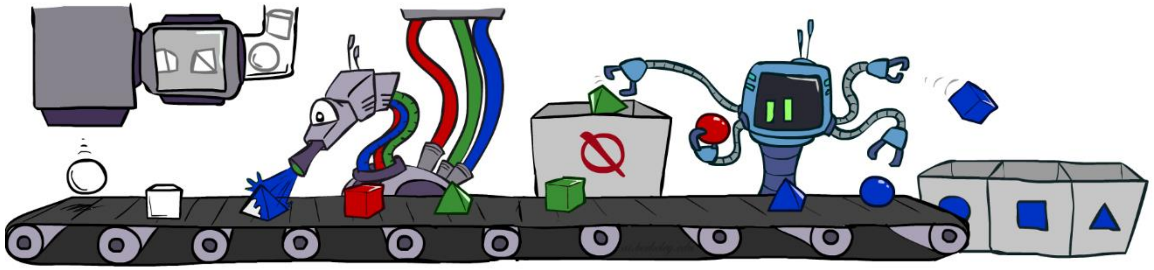


Likelihood Weighting

- Problem with rejection sampling:
  - If evidence is unlikely, rejects lots of samples
  - Evidence not exploited as you sample
  - Consider P(Shape|blue)

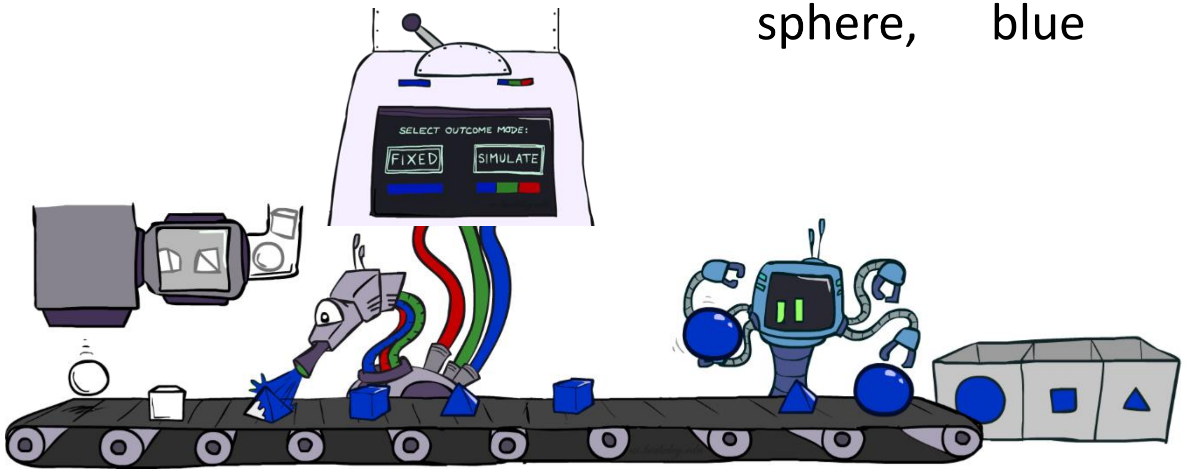


~~pyramid, green~~
~~pyramid, red~~
sphere, blue
~~cube, red~~
~~sphere, green~~



- Idea: fix evidence variables and sample the rest
  - Problem: sample distribution not consistent!
  - Solution: weight by probability of evidence given parents

pyramid, blue
pyramid, blue
sphere, blue
cube, blue
sphere, blue



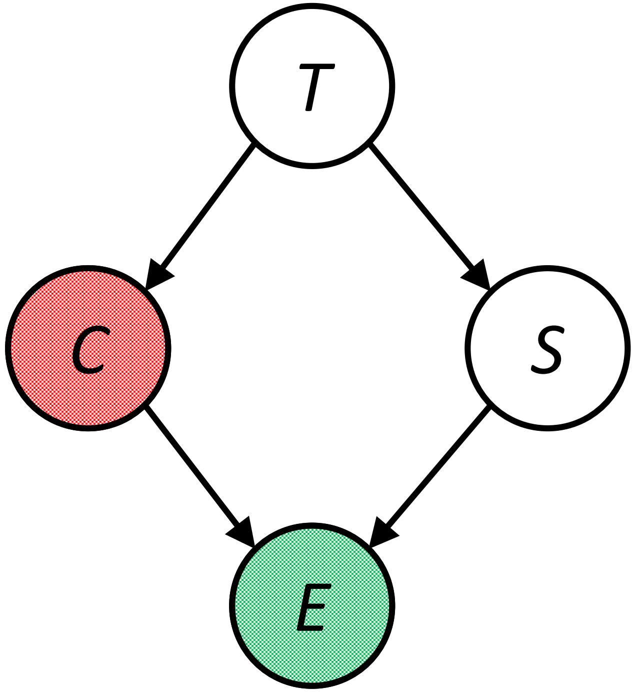
Example: Calculating P(T | -c, +e)

$P(T)$

+t	0.75
-t	0.25

$P(C|T)$

+t	+c	0.95
+t	-c	0.05
-t	+c	0.0
-t	-c	1.0



$P(S|T)$

+t	+s	0.1
+t	-s	0.9
-t	+s	0.01
-t	-s	0.99

$P(E|C,S)$

+c	+s	+e	0.1
+c	+s	-e	0.9
+c	-s	+e	0.2
+c	-s	-e	0.8
-c	+s	+e	0.3
-c	+s	-e	0.7
-c	-s	+e	0.8
-c	-s	-e	0.2

Samples:

```
    ...
6: ??, -c, ??, +e
    ...
```

$$w_6 = 1.0$$

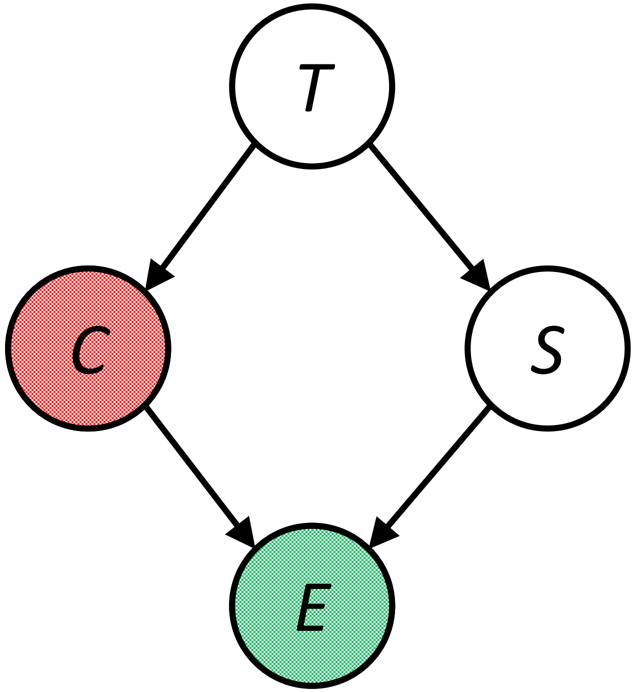
Example: Calculating P(T | -c, +e)

$P(T)$

+t	0.75
-t	0.25

$P(C|T)$

+t	+c	0.95
+t	-c	0.05
-t	+c	0.0
-t	-c	1.0



$P(S|T)$

+t	+s	0.1
+t	-s	0.9
-t	+s	0.01
-t	-s	0.99

$P(E|C,S)$

+c	+s	+e	0.1
+c	+s	-e	0.9
+c	-s	+e	0.2
+c	-s	-e	0.8
-c	+s	+e	0.3
-c	+s	-e	0.7
-c	-s	+e	0.8
-c	-s	-e	0.2

Samples:

```
    ...
6: ??, -c, ??, +e
    ...
```

$$w_6 = 1.0$$

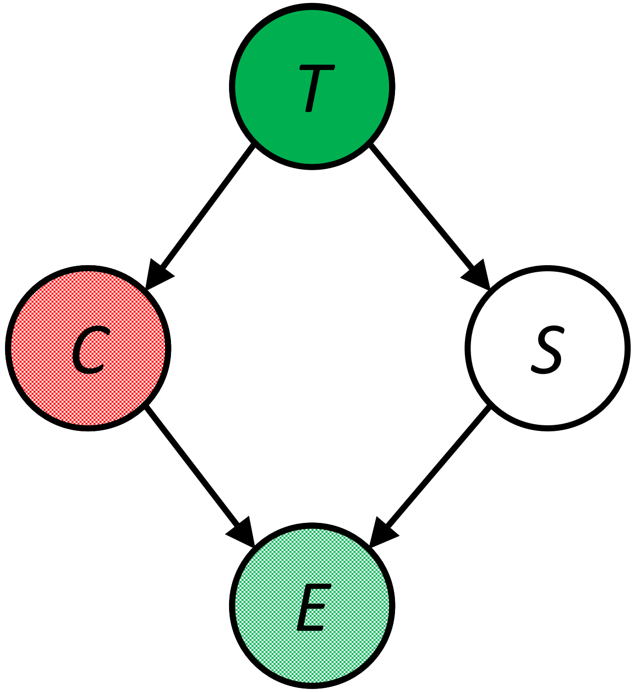
Example: Calculating P(T | -c, +e)

$P(T)$

+t	0.75
-t	0.25

$P(C|T)$

+t	+c	0.95
+t	-c	0.05
-t	+c	0.0
-t	-c	1.0



$P(S|T)$

+t	+s	0.1
+t	-s	0.9
-t	+s	0.01
-t	-s	0.99

$P(E|C,S)$

+c	+s	+e	0.1
+c	+s	-e	0.9
+c	-s	+e	0.2
+c	-s	-e	0.8
-c	+s	+e	0.3
-c	+s	-e	0.7
-c	-s	+e	0.8
-c	-s	-e	0.2

Samples:

```
    ...
6: +t, -c, ??, +e
    ...
```

$$w_6 = 1.0$$

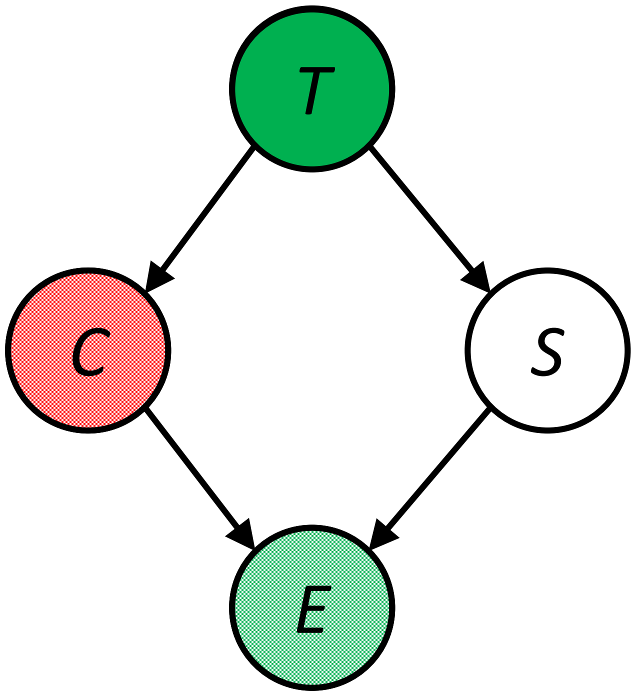
Example: Calculating P(T | -c, +e)

$P(T)$

+t	0.75
-t	0.25

$P(C|T)$

+t	+c	0.95
+t	-c	0.05
-t	+c	0.0
-t	-c	1.0



$P(S|T)$

+t	+s	0.1
+t	-s	0.9
-t	+s	0.01
-t	-s	0.99

$P(E|C,S)$

+c	+s	+e	0.1
+c	+s	-e	0.9
+c	-s	+e	0.2
+c	-s	-e	0.8
-c	+s	+e	0.3
-c	+s	-e	0.7
-c	-s	+e	0.8
-c	-s	-e	0.2

Samples:

```
    ...
6: +t, -c, ??, +e
    ...
```

$$w_6 = 1.0 \times 0.05$$

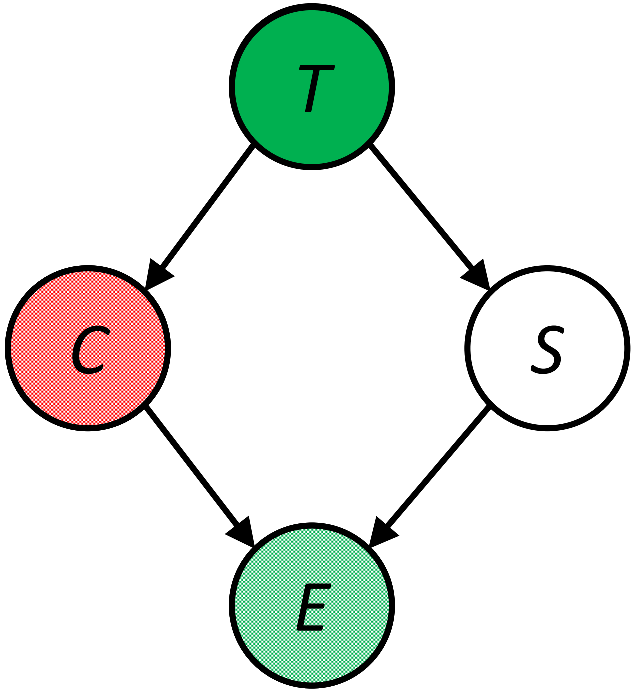
Example: Calculating P(T | -c, +e)

$P(T)$

+t	0.75
-t	0.25

$P(C|T)$

+t	+c	0.95
+t	-c	0.05
-t	+c	0.0
-t	-c	1.0



$P(S|T)$

+t	+s	0.1
+t	-s	0.9
-t	+s	0.01
-t	-s	0.99

$P(E|C,S)$

+c	+s	+e	0.1
+c	+s	-e	0.9
+c	-s	+e	0.2
+c	-s	-e	0.8
-c	+s	+e	0.3
-c	+s	-e	0.7
-c	-s	+e	0.8
-c	-s	-e	0.2

Samples:

```
    ...
6: +t, -c, ??, +e
    ...
```

$$w_6 = 1.0 \times 0.05$$

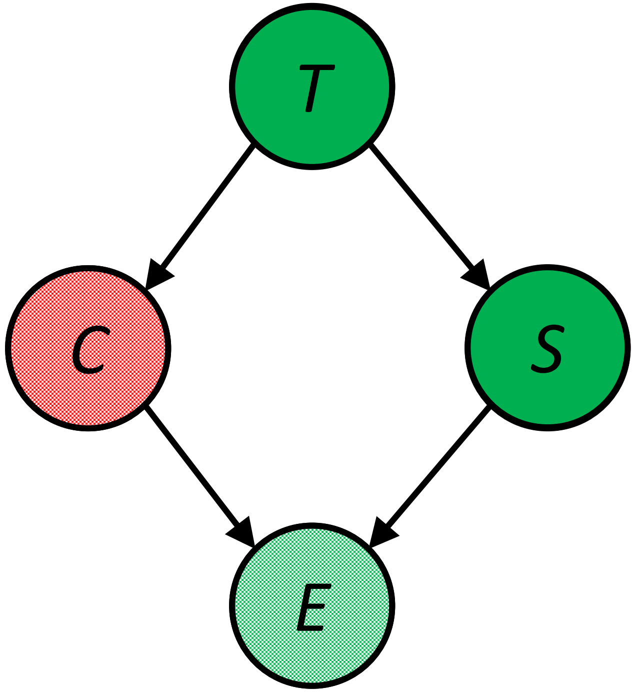
Example: Calculating P(T | -c, +e)

$P(T)$

+t	0.75
-t	0.25

$P(C|T)$

+t	+c	0.95
+t	-c	0.05
-t	+c	0.0
-t	-c	1.0



$P(S|T)$

+t	+s	0.1
+t	-s	0.9
-t	+s	0.01
-t	-s	0.99

$P(E|C,S)$

+c	+s	+e	0.1
+c	+s	-e	0.9
+c	-s	+e	0.2
+c	-s	-e	0.8
-c	+s	+e	0.3
-c	+s	-e	0.7
-c	-s	+e	0.8
-c	-s	-e	0.2

Samples:

```
    ...
6: +t, -c, +s, +e
    ...
```

$$w_6 = 1.0 \times 0.05$$

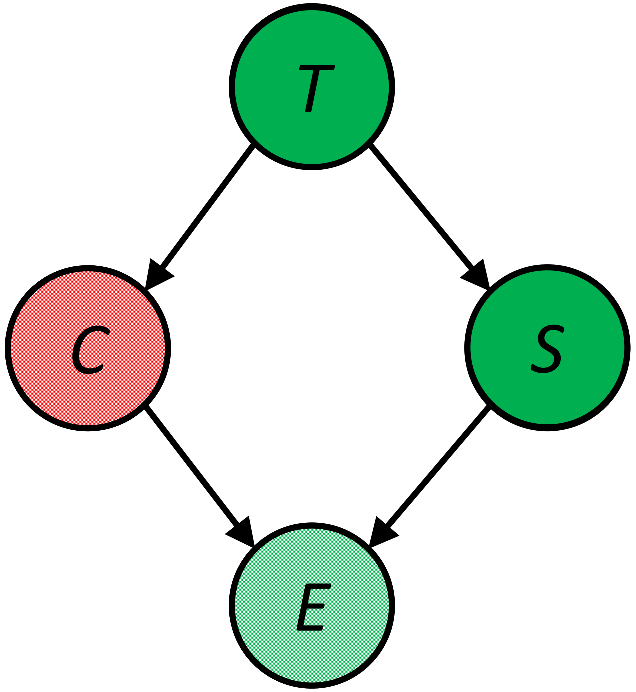
Example: Calculating P(T | -c, +e)

$P(T)$

+t	0.75
-t	0.25

$P(C|T)$

+t	+c	0.95
+t	-c	0.05
-t	+c	0.0
-t	-c	1.0



$P(S|T)$

+t	+s	0.1
+t	-s	0.9
-t	+s	0.01
-t	-s	0.99

$P(E|C,S)$

+c	+s	+e	0.1
+c	+s	-e	0.9
+c	-s	+e	0.2
+c	-s	-e	0.8
-c	+s	+e	0.3
-c	+s	-e	0.7
-c	-s	+e	0.8
-c	-s	-e	0.2

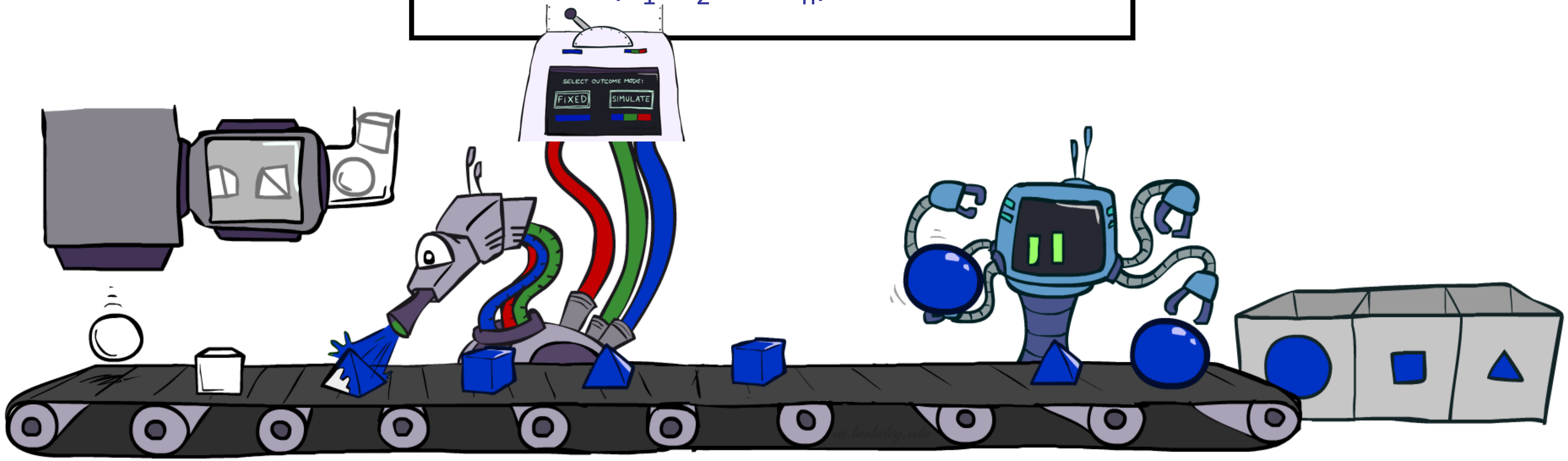
Samples:

```
    ...
6: +t, -c, +s, +e
    ...
```

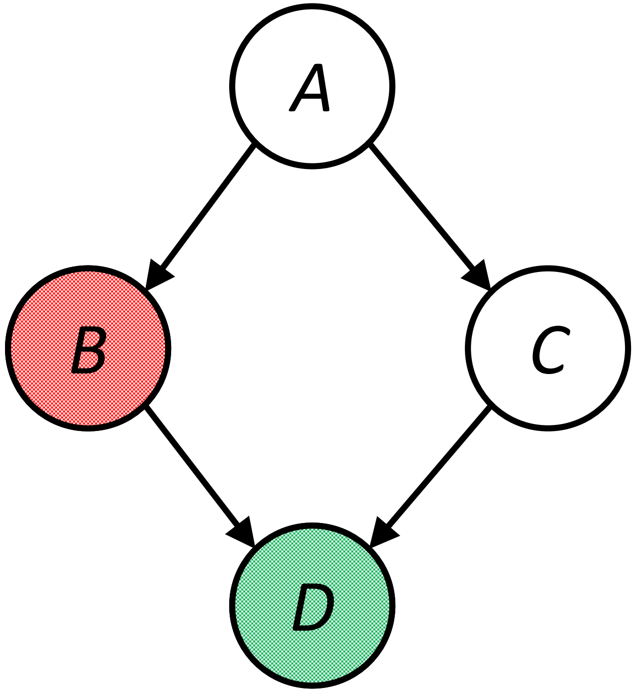
$$w_6 = 1.0 \times 0.05 \times 0.3 = 0.015$$

Likelihood Weighting

- IN: evidence instantiation
- w = 1.0
- for i=1, 2, …, n
  - if X_i is an evidence variable
    - X_i = observation x_i for X_i
    - Set w = w * $P(x_i \mid \text{Parents}(X_i))$
  - else
    - Sample x_i from $P(X_i \mid \text{Parents}(X_i))$
- return $(x_1, x_2, \ldots, x_n)$, w



Likelihood Weighting (Proof 1)



- Sampling distribution if z sampled and e fixed evidence

$$S_{WS}(\mathbf{z},\mathbf{e}) = \prod_{i=1}^{l} P(z_i|\text{Parents}(Z_i))$$

- Now, samples have weights

$$w(\mathbf{z},\mathbf{e}) = \prod_{i=1}^{m} P(e_i|\text{Parents}(E_i))$$

- Together, weighted sampling distribution is consistent

$$S_{\text{WS}}(z,e) \cdot w(z,e) = \prod_{i=1}^{l} P(z_i|\text{Parents}(z_i)) \prod_{i=1}^{m} P(e_i|\text{Parents}(e_i))$$

$$= P(\mathbf{z},\mathbf{e})$$

Likelihood Weighting (Proof 1)

- Sampling distribution if z sampled and e fixed evidence

$$S_{WS}(z_1, \ldots, z_p, e_1, \ldots, e_m) = \prod_i^p P(z_i | Parents(z_i))$$

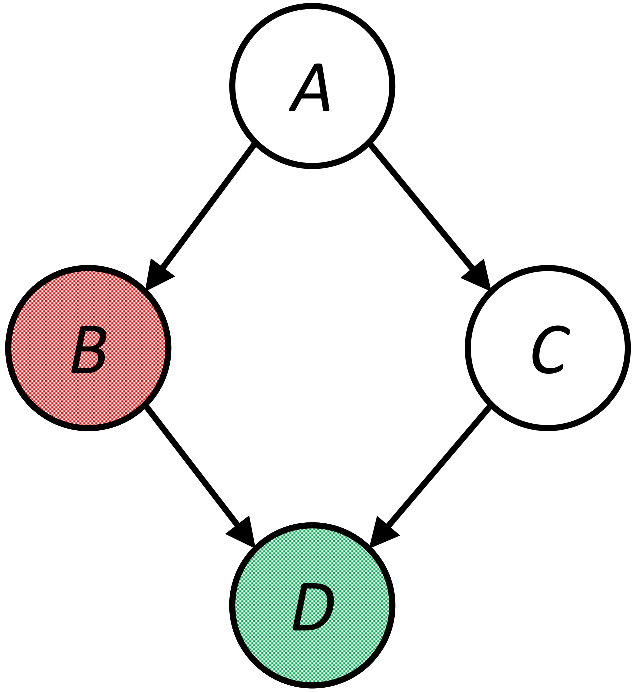
- Now, samples have weights

$$w(z_1, \ldots, z_p, e_1, \ldots, e_m) = \prod_i^m P(e_i | Parents(e_i))$$

- Together, weighted sampling distribution is consistent

$$S_{WS}(z_1, \ldots, z_p, e_1, \ldots, e_m) \cdot w(z_1, \ldots, z_p, e_1, \ldots, e_m) = \prod_i^p P(z_i | Parents(z_i)) \prod_i^m P(e_i | Parents(e_i))$$

$$= P(z_1, \ldots, z_p, e_1, \ldots e_m)$$



Upstream Variable Choices Can Be Bad

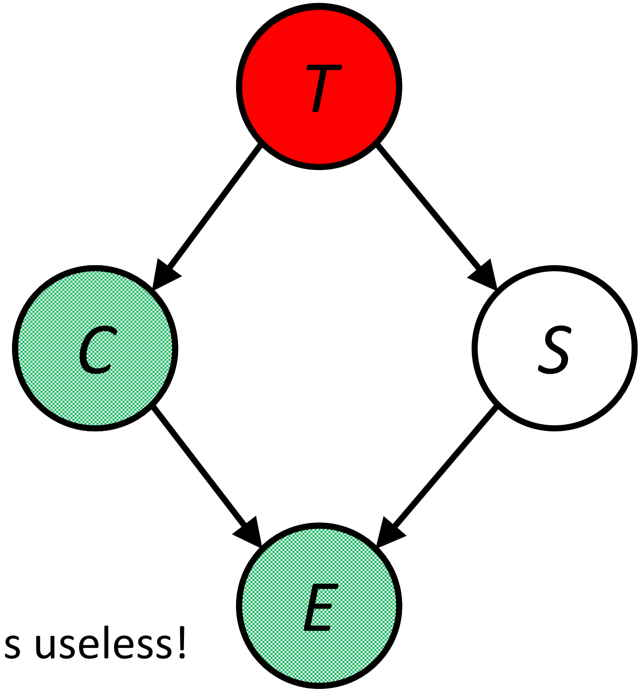
$P(T)$

+t	0.75
-t	0.25

Suppose we're calculating P(T, S | **+c, +e**)

$P(C|T)$

+t	+c	0.95
+t	-c	0.05
-t	+c	0.0
-t	-c	1.0



$P(S|T)$

+t	+s	0.1
+t	-s	0.9
-t	+s	0.01
-t	-s	0.99

$P(E|C,S)$

+c	+s	+e	0.1
+c	+s	-e	0.9
+c	-s	+e	0.2
+c	-s	-e	0.8
-c	+s	+e	0.3
-c	+s	-e	0.7
-c	-s	+e	0.8
-c	-s	-e	0.2

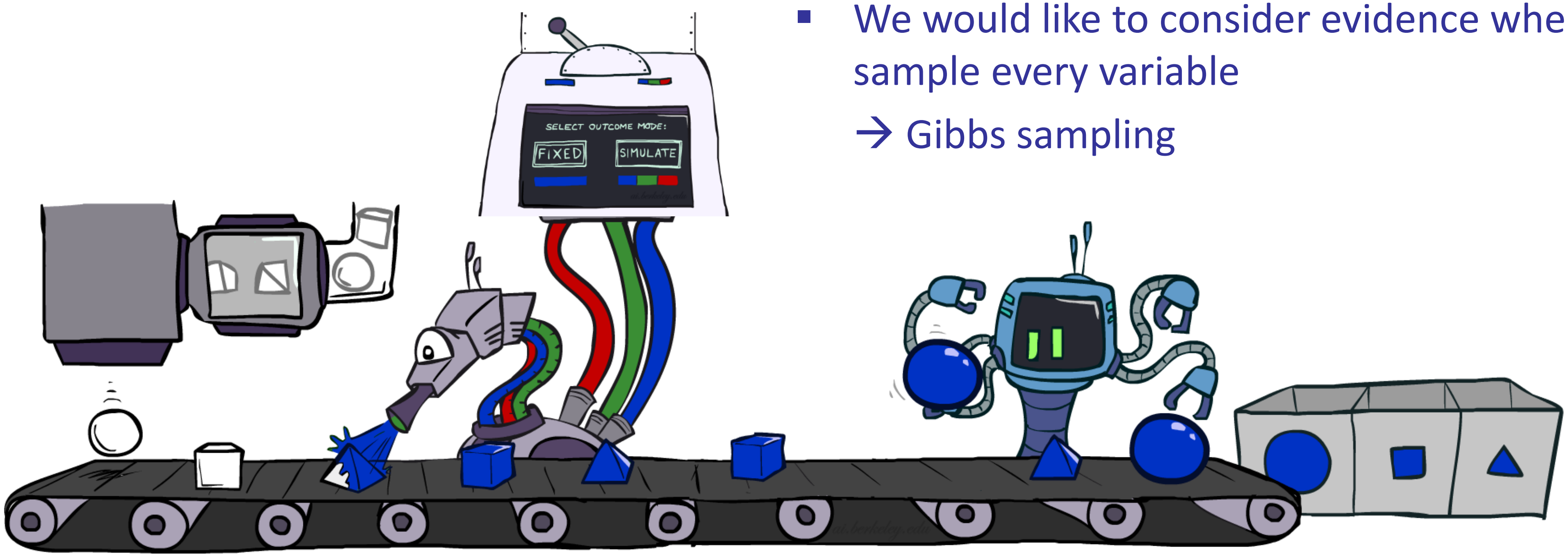
Any sample that starts with +t is useless!

- Will be given weight 0.

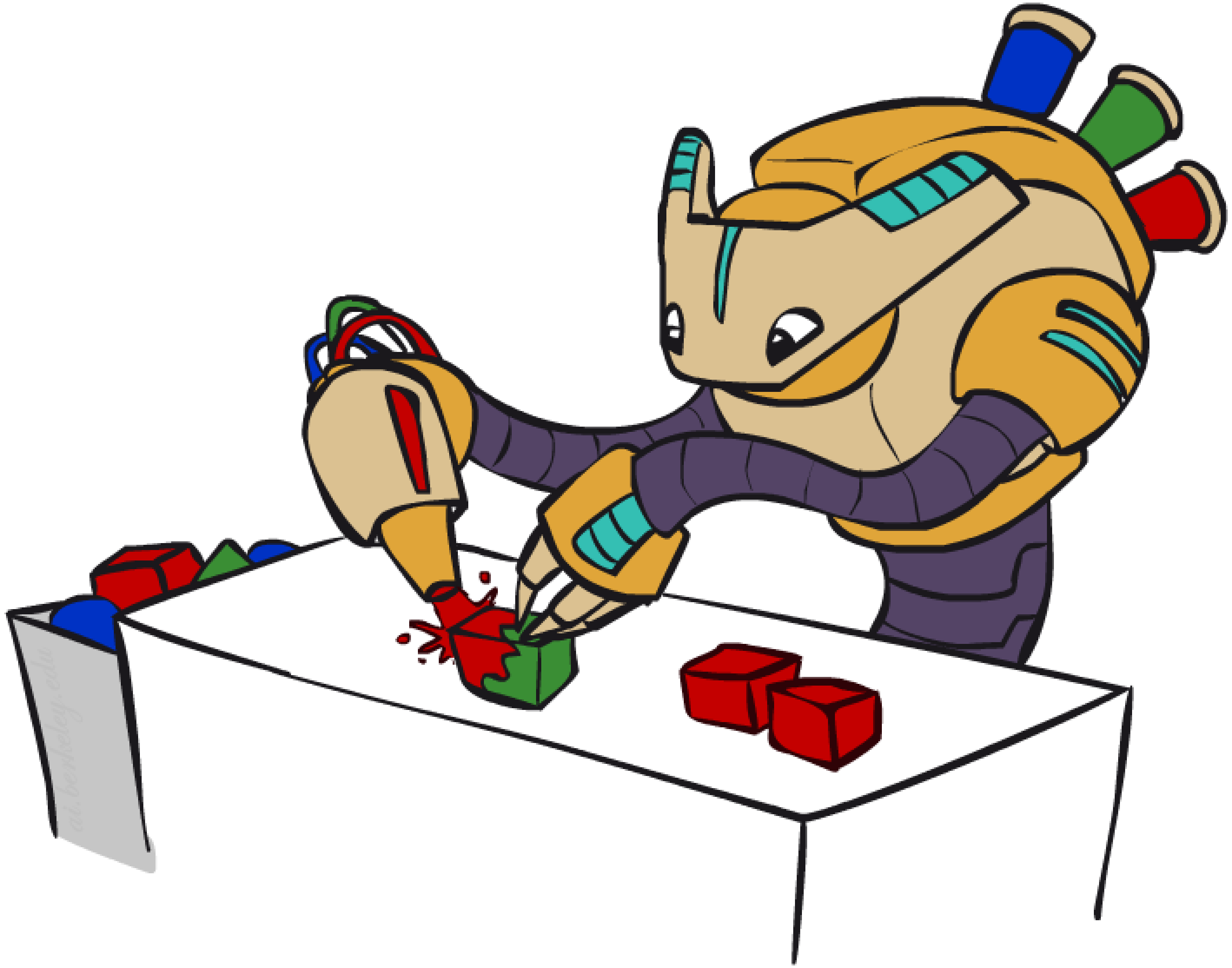
Likelihood Weighting

- Likelihood weighting is an improvement.
  - Samples never rejected for disagreeing with evidence
  - Thus, never need to abort a sample
- Likelihood weighting doesn't solve all our problems
  - Evidence influences the choice of downstream variables, but not upstream ones (T isn't more likely to get a value matching the evidence)
  - Still possible to generate useless or very low weight samples
- We would like to consider evidence when we sample every variable

  → Gibbs sampling



Gibbs Sampling

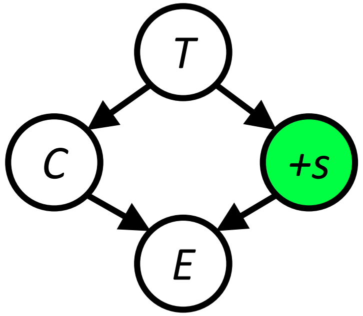


Gibbs Sampling

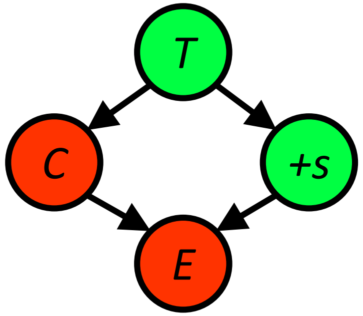
- *Procedure:* keep track of a full instantiation $x_1, x_2, \ldots, x_n$. Start with an arbitrary instantiation consistent with the evidence. Sample one variable at a time, conditioned on all the rest, but keep evidence fixed. Keep repeating this for a long time.
- *Property:* in the limit of repeating this infinitely many times the resulting sample is coming from the correct distribution
- *Rationale*: both upstream and downstream variables condition on evidence.
- In contrast: likelihood weighting only conditions on upstream evidence, and hence weights obtained in likelihood weighting can sometimes be very small. Sum of weights over all samples is indicative of how many "effective" samples were obtained, so want high weight.

Gibbs Sampling Example: P(T | +s)

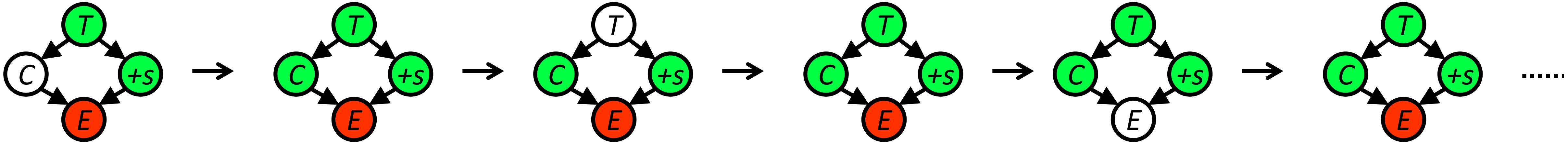
- Step 1: Fix evidence
  - S = +s



- Step 2: Initialize other variables
  - Randomly



- Steps 3: Repeat
  - Choose a non-evidence variable X
  - Resample X from P(X | all other variables)



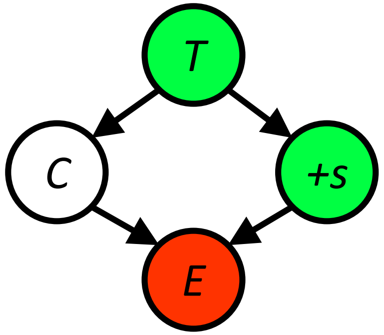
Sample from $P(C| +t, +s, -e)$

Sample from $P(T| +c, +s, -e)$

Sample from $P(E| +t, +s, +e)$

Efficient Resampling of One Variable

- Sample from P(C | +t, +s, -e)



$$P(C|+t,+s,-e) = \frac{P(C,+t,+s,-e)}{P(+t,+s,-e)}$$

$$= \frac{P(C,+t,+s,-e)}{\sum_c P(c,+t,+s,-e)}$$

$$= \frac{P(+t)P(C|+t)P(+s|+t)P(-e|C,+s)}{\sum_c P(+t)P(c|+t)P(+s|+t)P(-e|c,+s)}$$

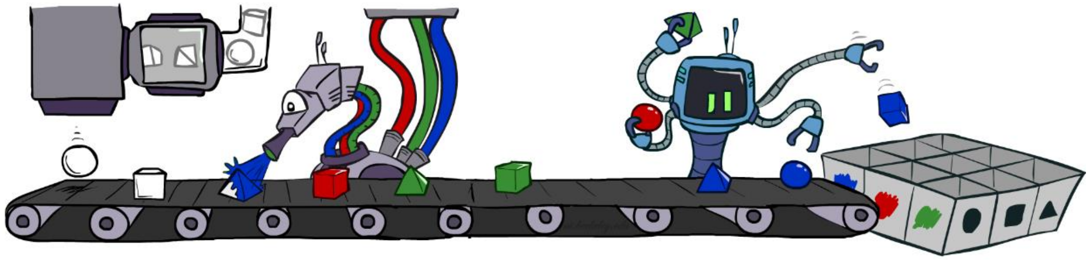
$$= \frac{P(+t)P(C|+t)P(+s|+t)P(-e|C,+s)}{P(+t)P(+s|+t)\sum_c P(c|+t)P(-e|c,+s)}$$

$$= \frac{P(\mathrm{C}|+t)P(-e|C,+s)}{\sum_c P(c|+t)P(-e|c,+s)}$$

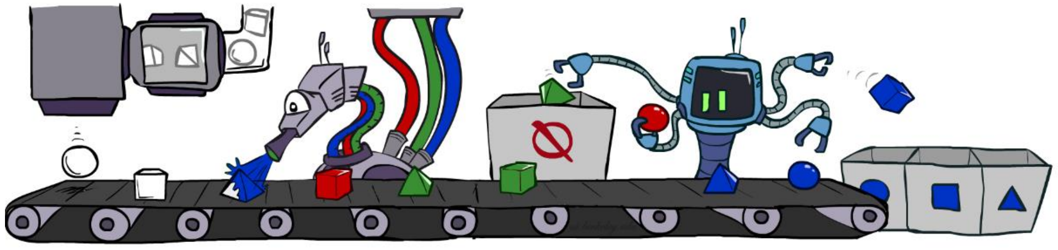
- Many things cancel out – only CPTs with S remain!
- Cool fact (we won't prove): For all BNs, only CPTs that have the resampled variable remain.

Bayes' Net Sampling Summary

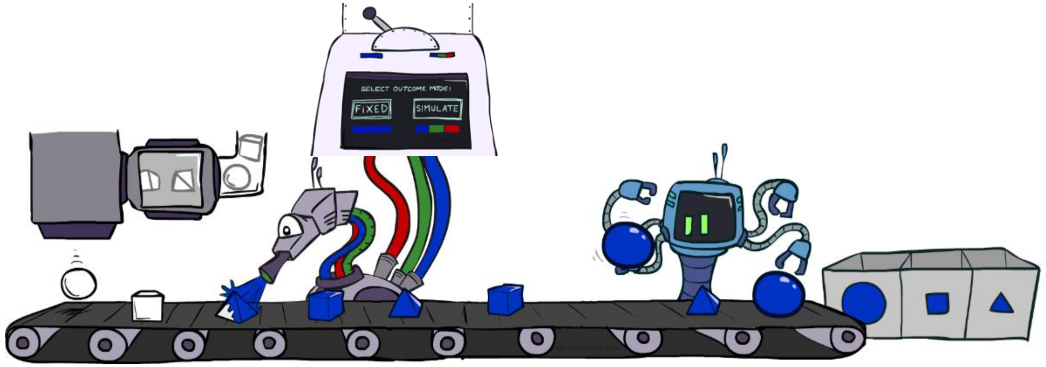
- Prior Sampling P



- Rejection Sampling P(Q | e)



- Likelihood Weighting P(Q | e)



- Gibbs Sampling P(Q | e)

Further Reading on Gibbs Sampling*

- Gibbs sampling produces sample from the query distribution P(Q | e) in limit of re-sampling infinitely often

- Gibbs sampling is a special case of more general methods called Markov chain Monte Carlo (MCMC) methods

  - Metropolis-Hastings is one of the more famous MCMC methods (in fact, Gibbs sampling is a special case of Metropolis-Hastings)

- You may hear about Monte Carlo methods – they're just sampling