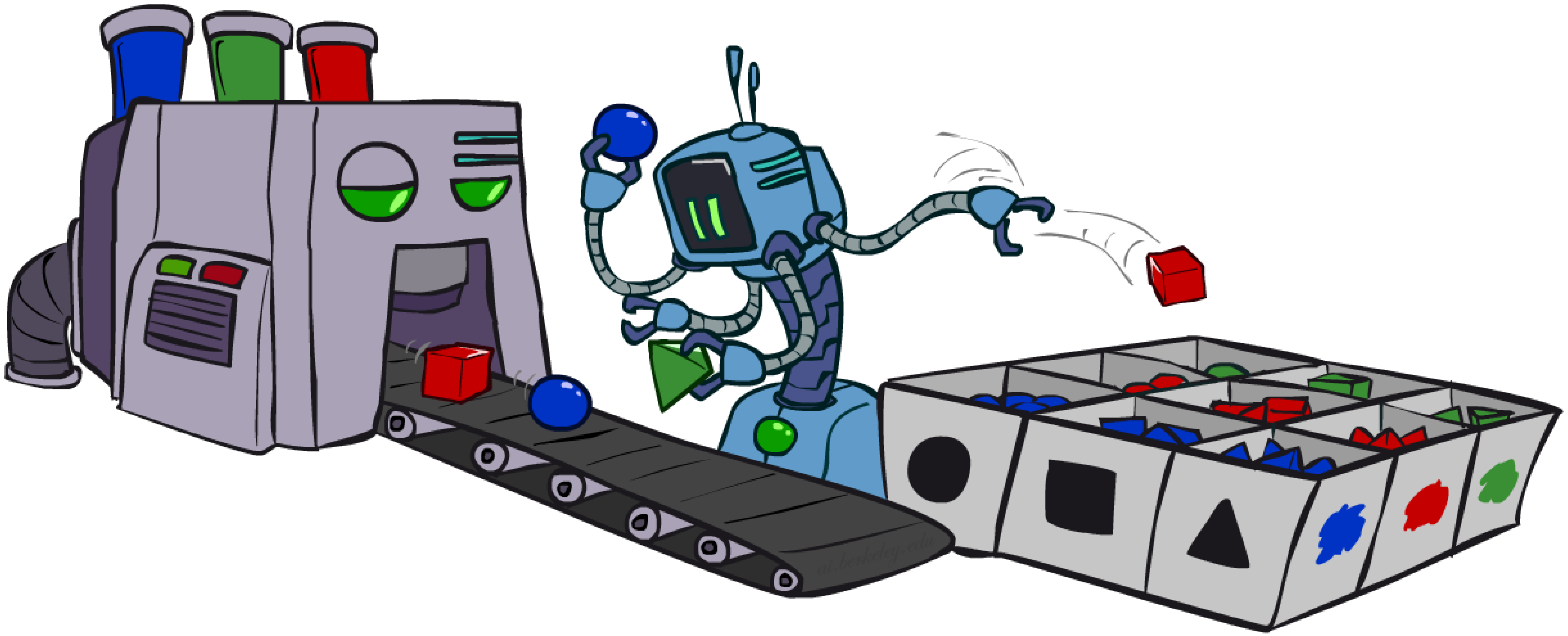
Bayes Nets: Approximate Inference



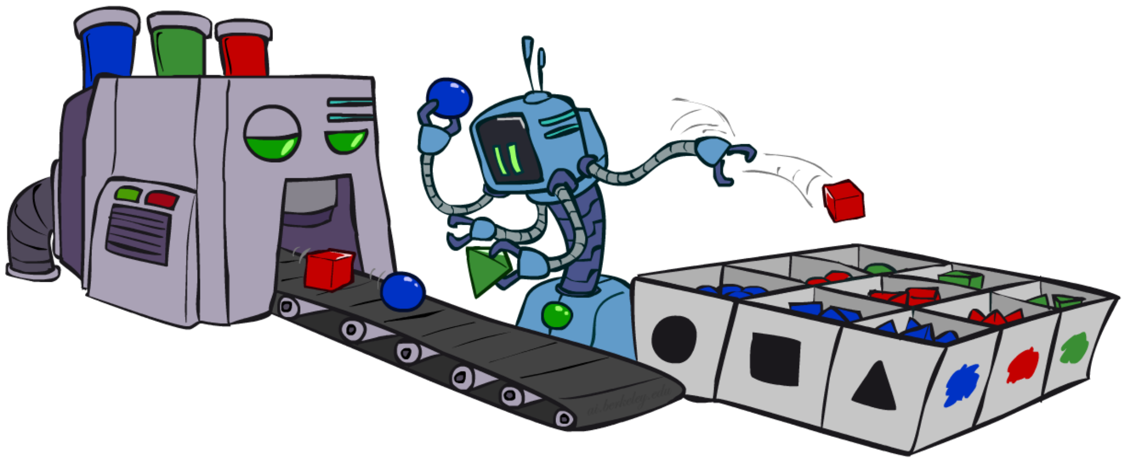
AIMA Chapter 14.5, PRML Chapter 11

[Adapted from slides by Dan Klein and Pieter Abbeel at UC Berkeley]

Sampling

- Goal: probability *P*
- Basic idea
  - Draw *N* samples from a sampling distribution *S*
  - Compute some quantity from the samples
  - Show this converges to the true probability *P*

- Why sample?
  - Often very fast to get a decent approximate answer
  - The algorithms are very simple and general (easy to apply to fancy models)
  - They require very little memory (O(n))

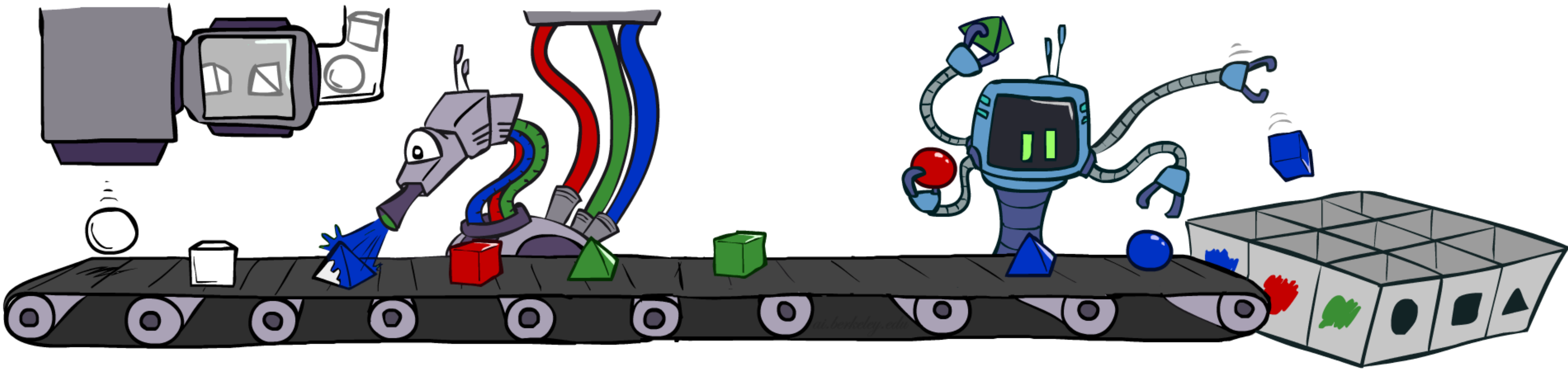


Sampling in Bayes Nets

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

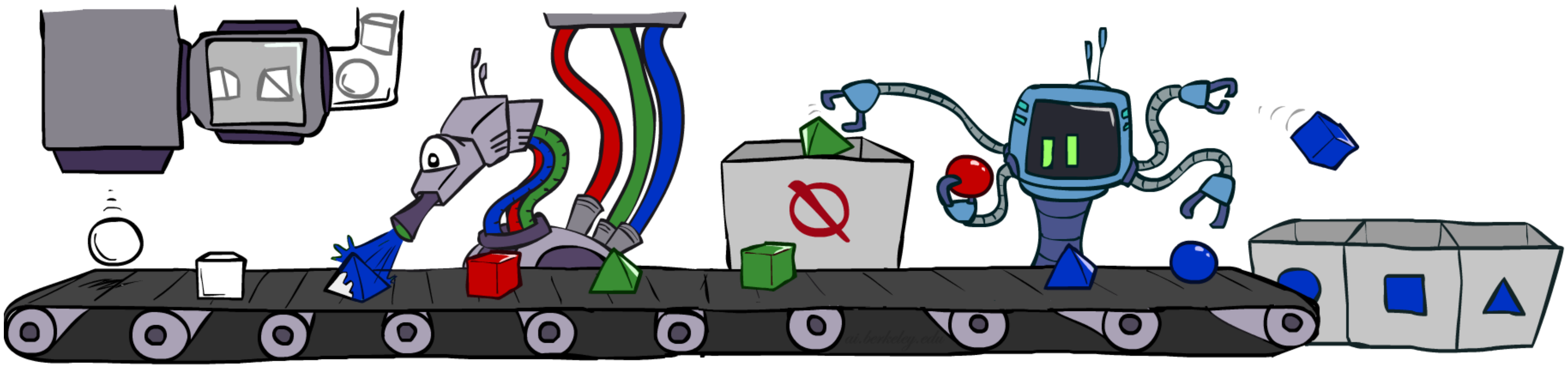
Prior Sampling

- For i=1, 2, …, n (in topological order)
  - Sample X_i from $P(X_i \mid parents(X_i))$
- Return $(x_1, x_2, \ldots, x_n)$

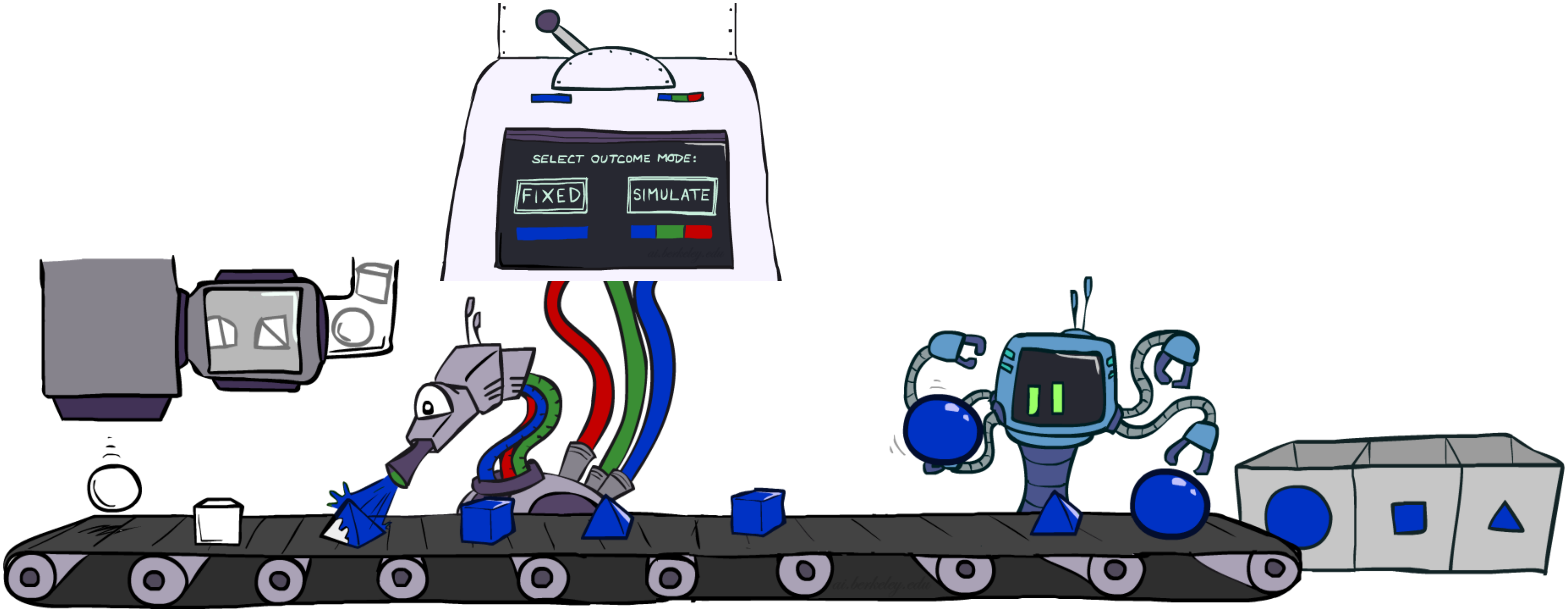


Rejection Sampling

- Input: evidence $e_1,..,e_k$
- 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 $(x_1, x_2, \ldots, x_n)$

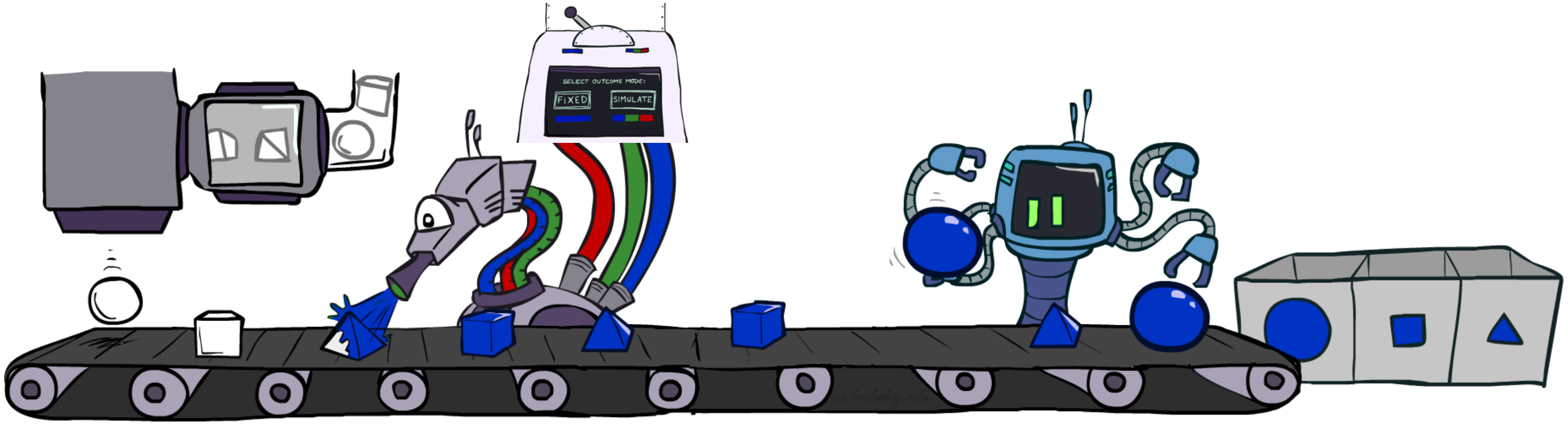


Likelihood Weighting



Likelihood Weighting

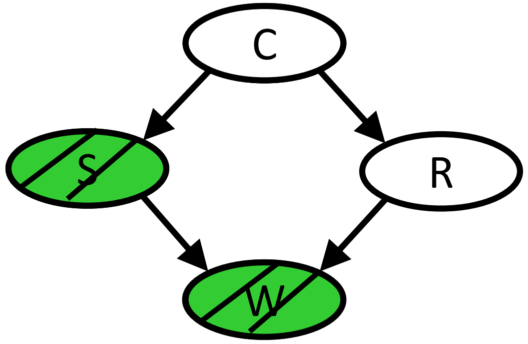
- Input: evidence $e_1,..,e_k$
- w = 1.0
- for i=1, 2, …, n
  - if X_i is an evidence variable
    - x_i = observed value$_i$ for X_i
    - Set $w = w * P(x_i \mid Parents(X_i))$
  - else
    - Sample x_i from $P(X_i \mid Parents(X_i))$
- return $(x_1, x_2, \ldots, x_n)$, w



Likelihood Weighting

- Sampling distribution (z is sampled and e is 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

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

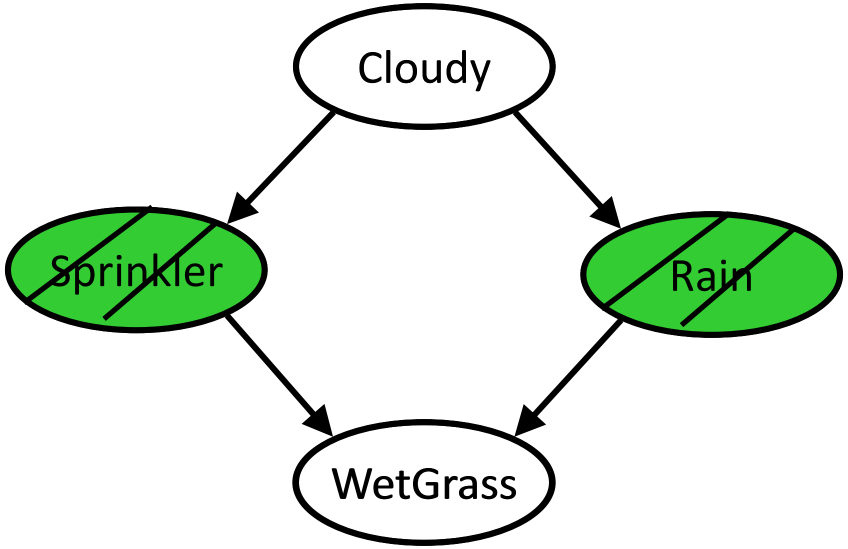
$$= P(\mathbf{z}, \mathbf{e})$$

Importance Sampling

- Likelihood weighting is an instance of importance sampling
  - Suppose it is difficult to sample from p(x)
  - Generate samples from a proposal distribution q(x)
    - q(x) is easy to draw samples from
  - Weight each sample by p(x)/q(x)
- The choice of q(x) would greatly influence the speed of convergence
  - If you want to estimate the expectation of f(x)
  - Then q(x) should be close to being proportional to |f(x)|p(x)

Likelihood Weighting

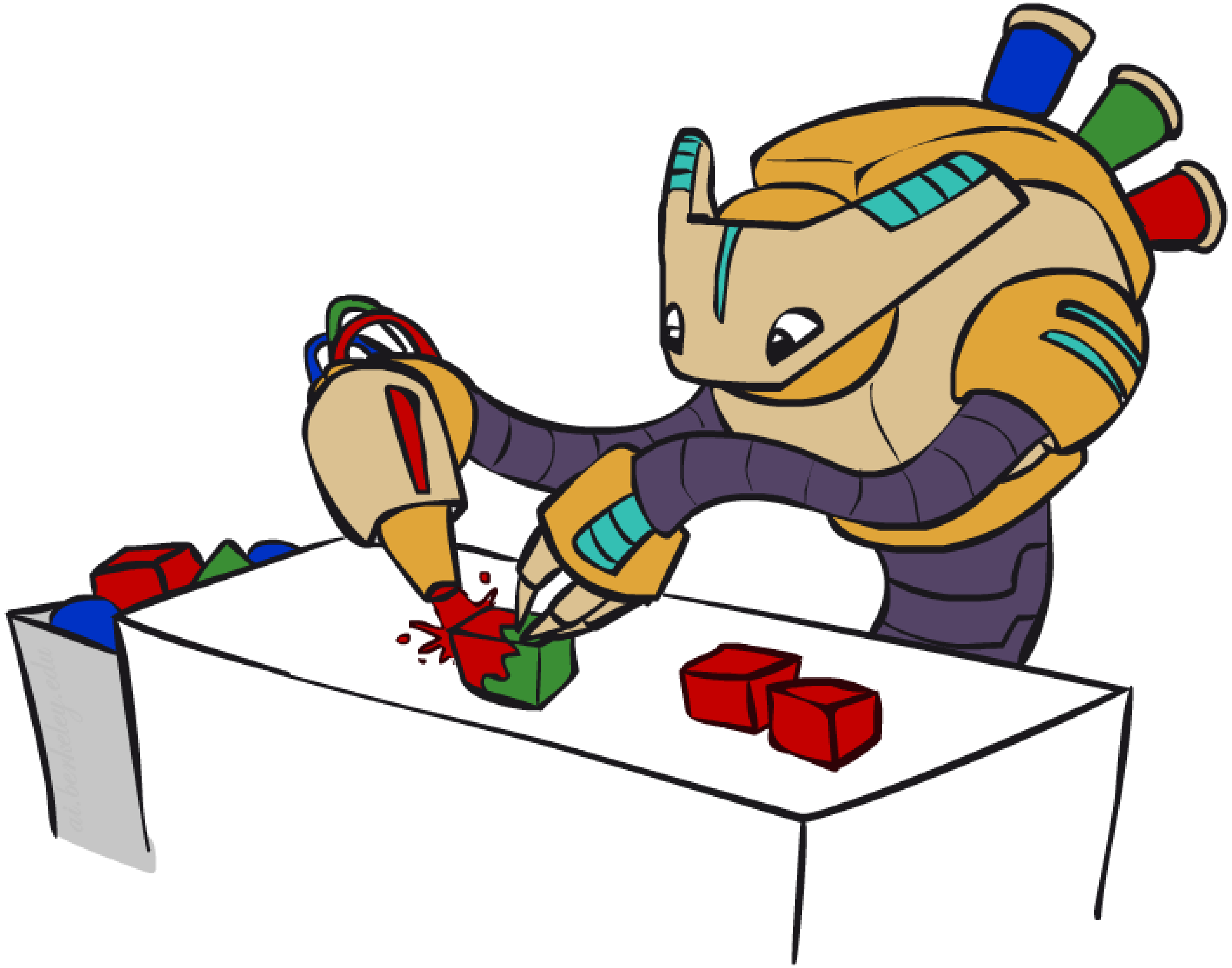
- Likelihood weighting is good
  - All samples are used
  - The values of ***downstream*** variables are influenced by ***upstream*** evidence



- Likelihood weighting still has weaknesses
  - The values of ***upstream*** variables are unaffected by ***downstream*** evidence
  - With many downstream evidence, we may
    - mostly get samples that are inconsistent with the evidence and thus have very small weights
    - get a few lucky samples with very large weights, which dominate the result

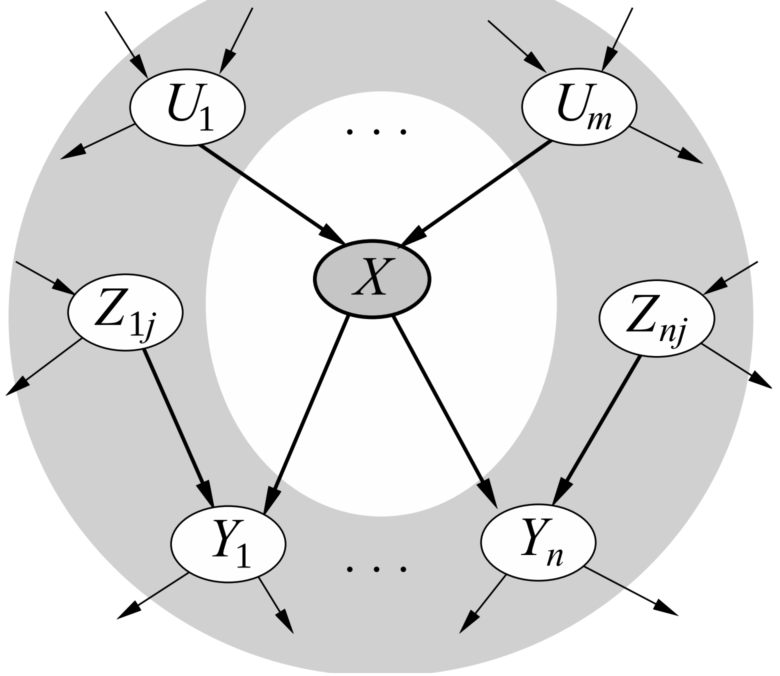
- We would like each variable to "see" ***all*** the evidence!

Gibbs Sampling



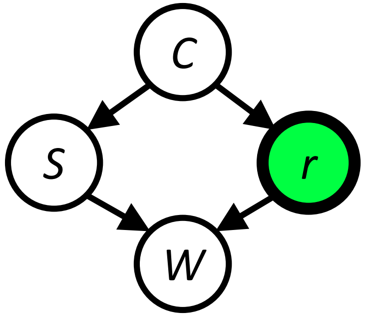
Gibbs Sampling

- Generate each sample by making a random change to the preceding sample
  - Evidence variables remain fixed. For each of the non-evidence variable, sample its value conditioned on all the other variables
  - $X_i' \sim P(X_i \mid x_1,..,x_{i-1},x_{i+1},..,x_n)$
  - In a Bayes net
    $P(X_i \mid x_1,..,x_{i-1},x_{i+1},..,x_n)$
    $= P(X_i \mid markov\_blanket(X_i))$
    $= \alpha\ P(X_i \mid u_1,..,u_m)\ \prod_j P(y_j \mid parents(Y_j))$

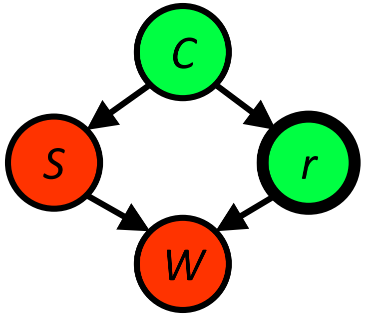


Gibbs Sampling Example: $P(S \mid r)$

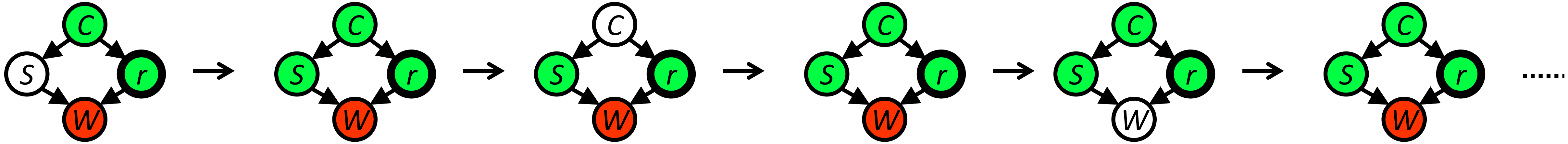
- Step 1: Fix evidence
  - R = true



- Step 2: Initialize other variables
  - Randomly



- Step 3: Repeat
  - Choose an arbitrary non-evidence variable X
  - Resample X from $P(X \mid markov\_blanket(X))$

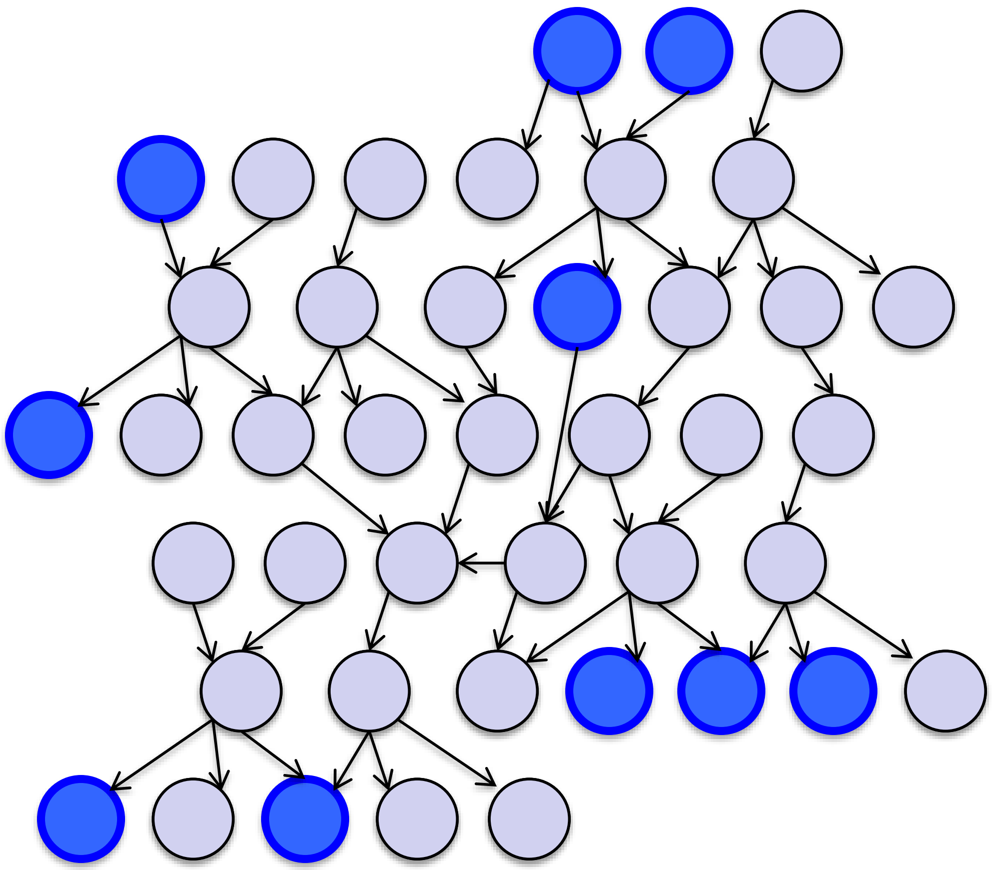


Sample $S \sim P(S \mid c, r, \neg w)$

Sample $C \sim P(C \mid s, r)$

Sample $W \sim P(W \mid s, r)$

Why doing this?



- Samples soon begin to reflect all the evidence in the network
- Eventually they are being drawn from the true posterior!
- Theorem: Gibbs sampling is consistent

Why does it work? (see AIMA 14.5.2 for details)

- Suppose we run it for a long time and predict the probability of reaching any given state at time t: $\pi_t(x_1,\ldots,x_n)$ or $\pi_t(\underline{\mathbf{x}})$
- Each Gibbs sampling step (pick a variable, resample its value) applied to a state $\underline{\mathbf{x}}$ has a probability $q(\underline{\mathbf{x}}' \mid \underline{\mathbf{x}})$ of reaching a next state $\underline{\mathbf{x}}'$
- So $\pi_{t+1}(\underline{\mathbf{x}}') = \sum_{\underline{\mathbf{x}}} q(\underline{\mathbf{x}}' \mid \underline{\mathbf{x}})\, \pi_t(\underline{\mathbf{x}})$ or, in matrix/vector form $\pi_{t+1} = \mathrm{Q}\pi_t$
- When the process is in equilibrium $\pi_{t+1} = \pi_t$ so $\mathrm{Q}\pi_t = \pi_t$
- This has a unique solution $\pi_t = \mathrm{P}(x_1,\ldots,x_n \mid e_1,\ldots,e_k)$
- So for large enough t the next sample will be drawn from the true posterior

- MC[...]ing som[...]
  - M[...]walk"), w[...]
  - M[...]ino

Markov Chain Monte Carlo (MCMC)

- MCMC is a family of randomized algorithms for approximating some quantity of interest over a very large state space
  - Markov chain = a sequence of randomly chosen states (“random walk”), where each state is chosen conditioned on the previous state
  - ~~Monte Carlo = a very expensive city in Monaco with a famous casino~~
  - Monte Carlo = an algorithm (usually based on sampling) that is likely to find a correct answer
- MCMC = sampling by constructing a Markov chain
- Gibbs, Metropolis-Hastings, Hamiltonian, Slice, etc.

Metropolis-Hastings

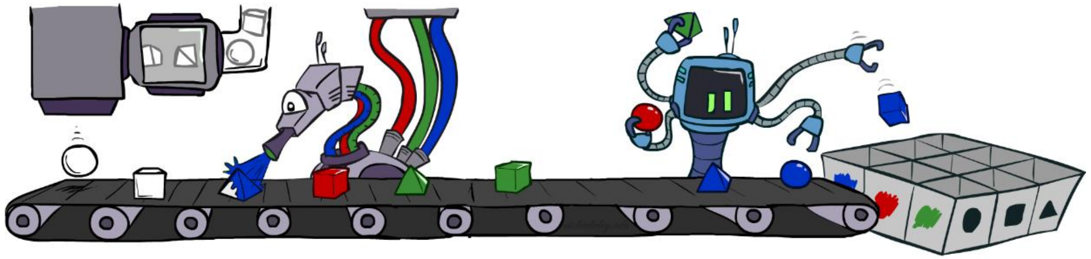
- Repeat
  1. Draw a sample from a proposal distribution $g(x'|x)$
     - $g(x'|x)$ is typically easy to sample from
  2. Accept this sample with probability

$$\min\left(1, \frac{P(x')g(x|x')}{P(x)g(x'|x)}\right)$$

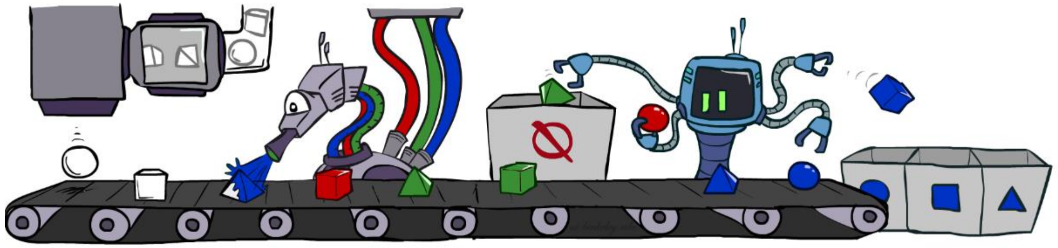
- Gibbs is a special case of Metropolis-Hastings in which the acceptance rate is always 1

Summary

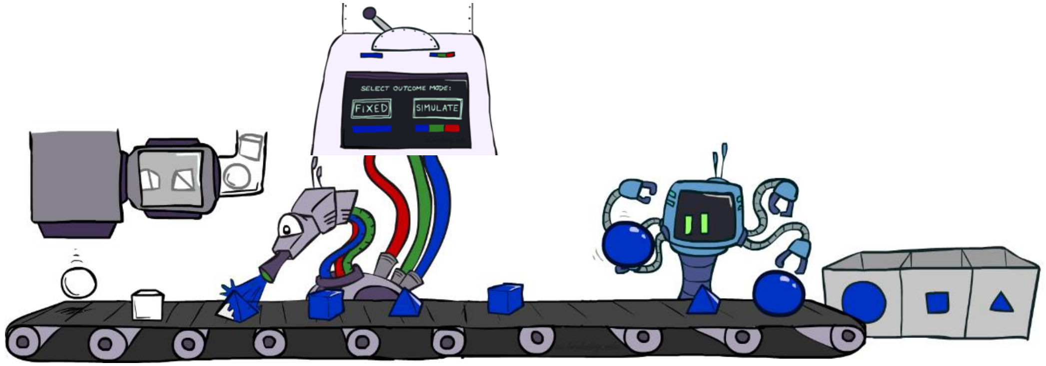
- Prior Sampling P



- Rejection Sampling P(Q | e)



- Likelihood Weighting P(Q | e)



- Gibbs Sampling P(Q | e)