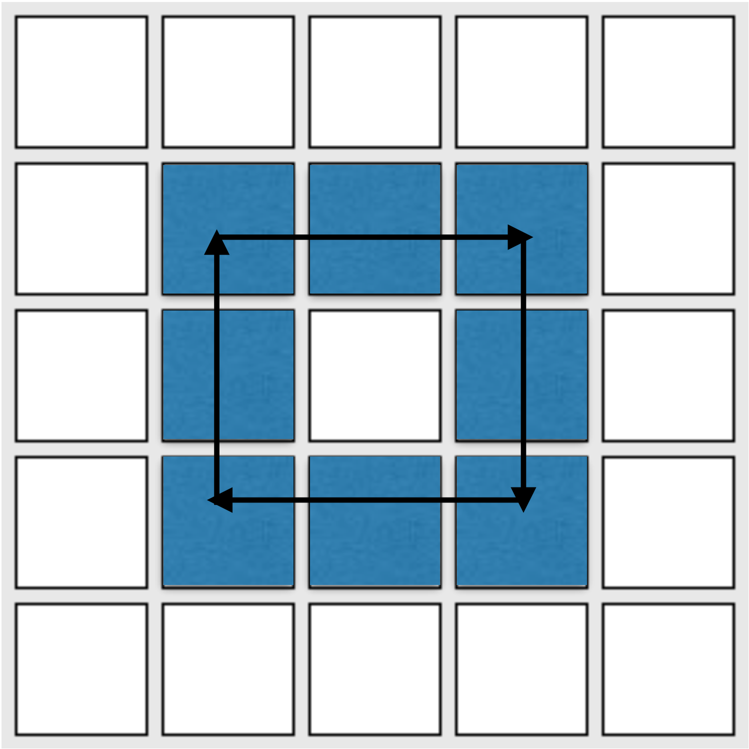
Machine Learning for Data Science (CS4786)
Lecture 21

Approximate Inference Via Sampling, particle filter

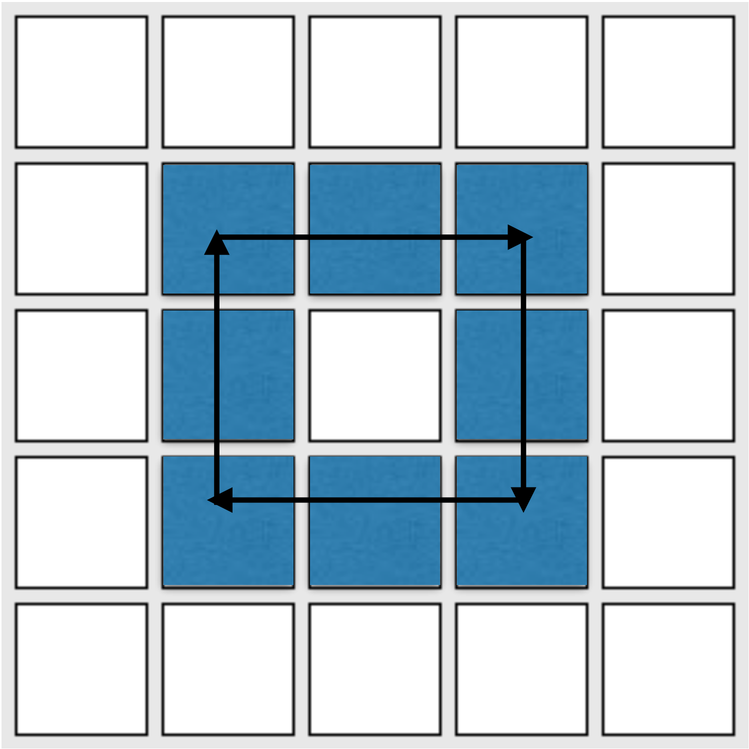
Hidden Markov Model (HMM)

Hidden Markov Model (HMM)



HIDDEN MARKOV MODEL (HMM)

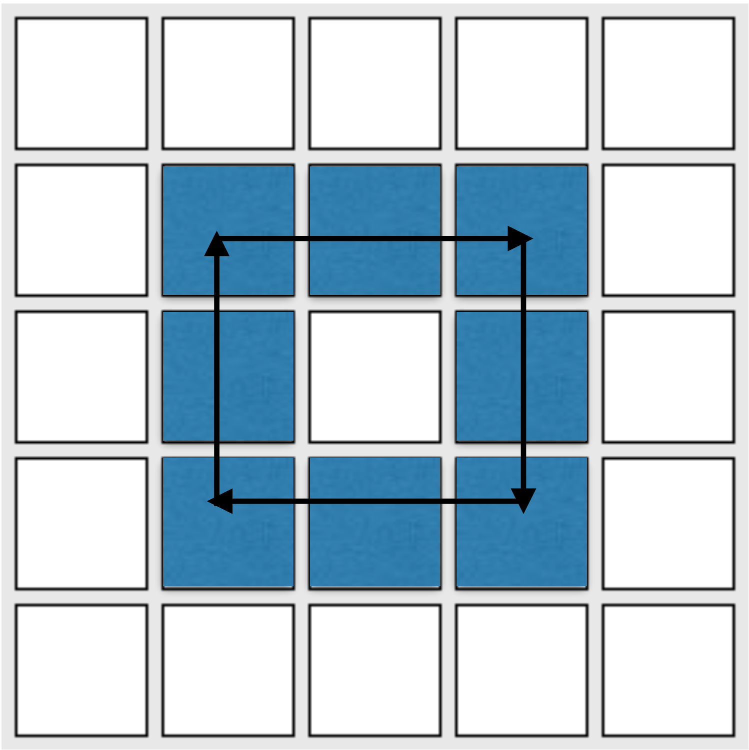
Example:



Hidden Markov Model (HMM)

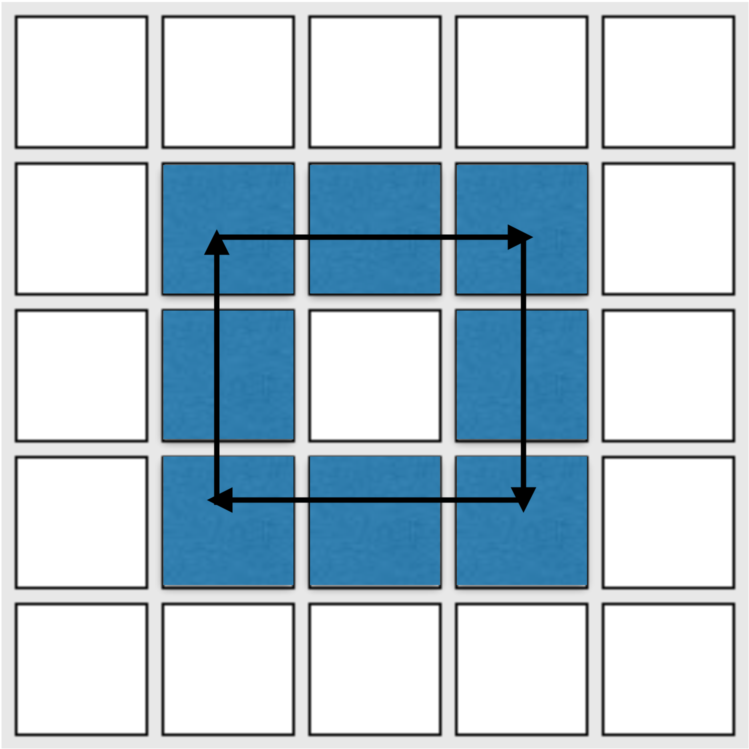
Example:

But you don't observe location (dark room)



Hidden Markov Model (HMM)

Example:

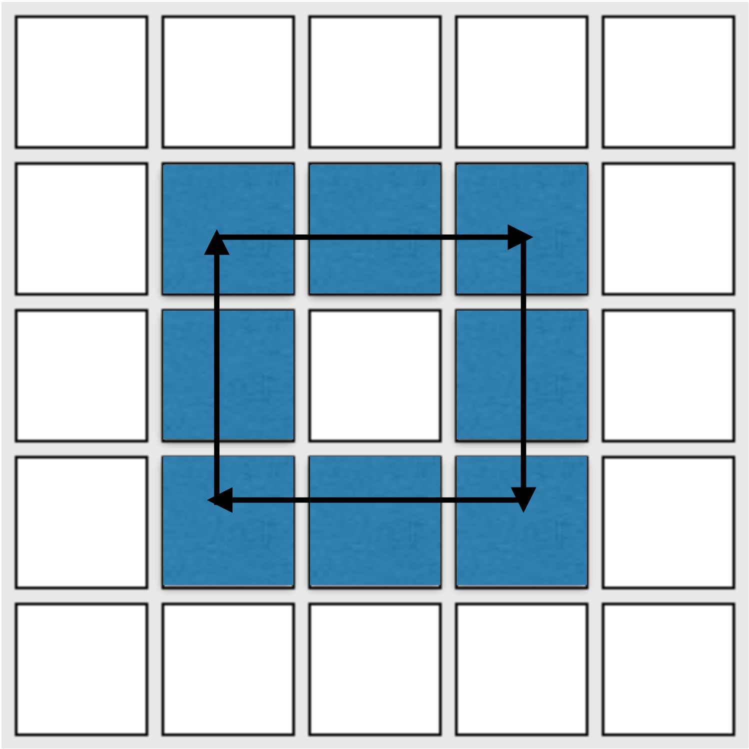


But you don’t observe location (dark room)

You hear how close the bot is!

Hidden Markov Model (HMM)

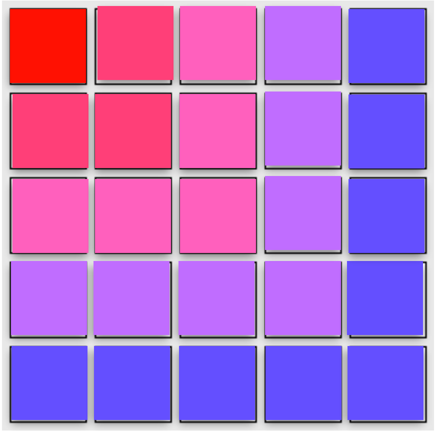
Example:



But you don't observe location (dark room)

You hear how close the bot is!

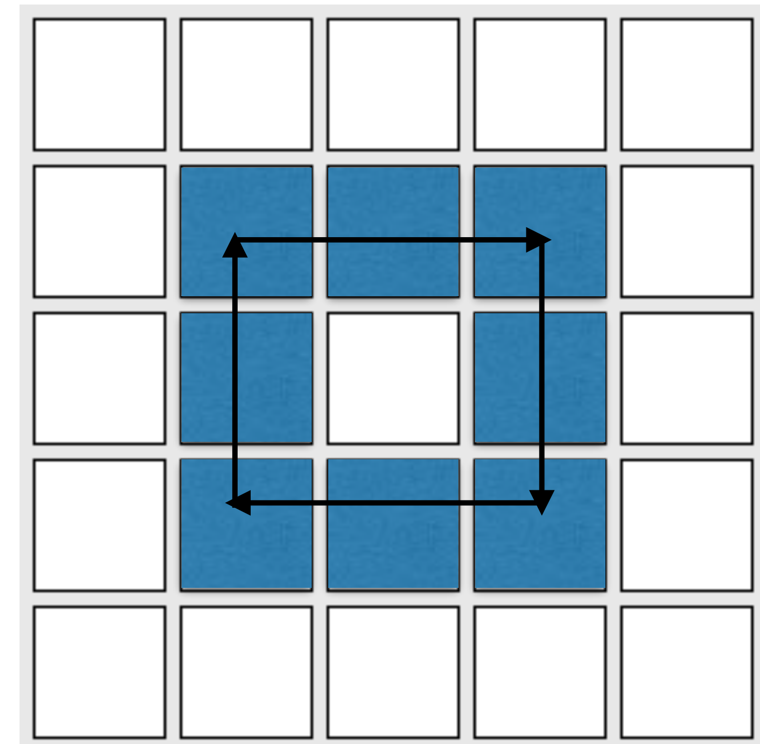
What you hear:



+ noise

HIDDEN MARKOV MODEL (HMM)

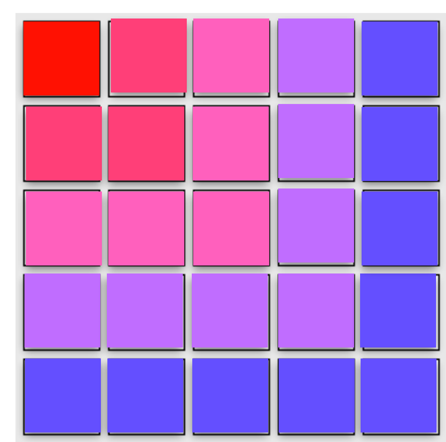
Example:



But you don't observe location (dark room)

You hear how close the bot is!

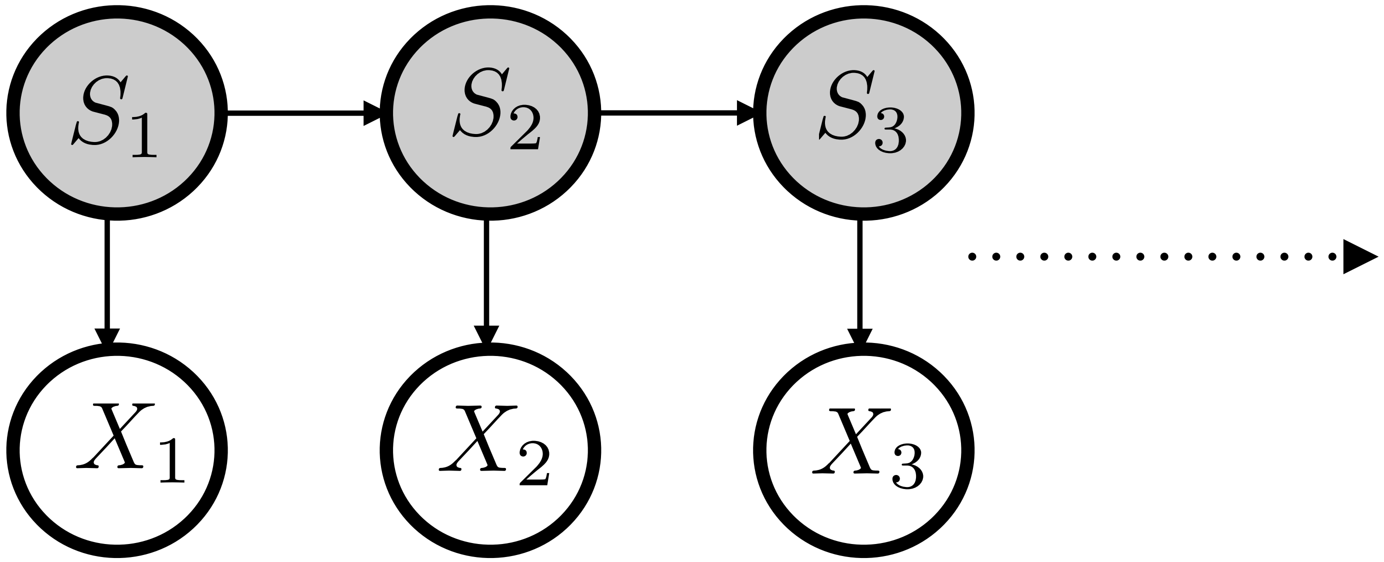
What you hear:



+ noise

Can you catch the Bot? In time?

Hidden Markov Model (HMM)



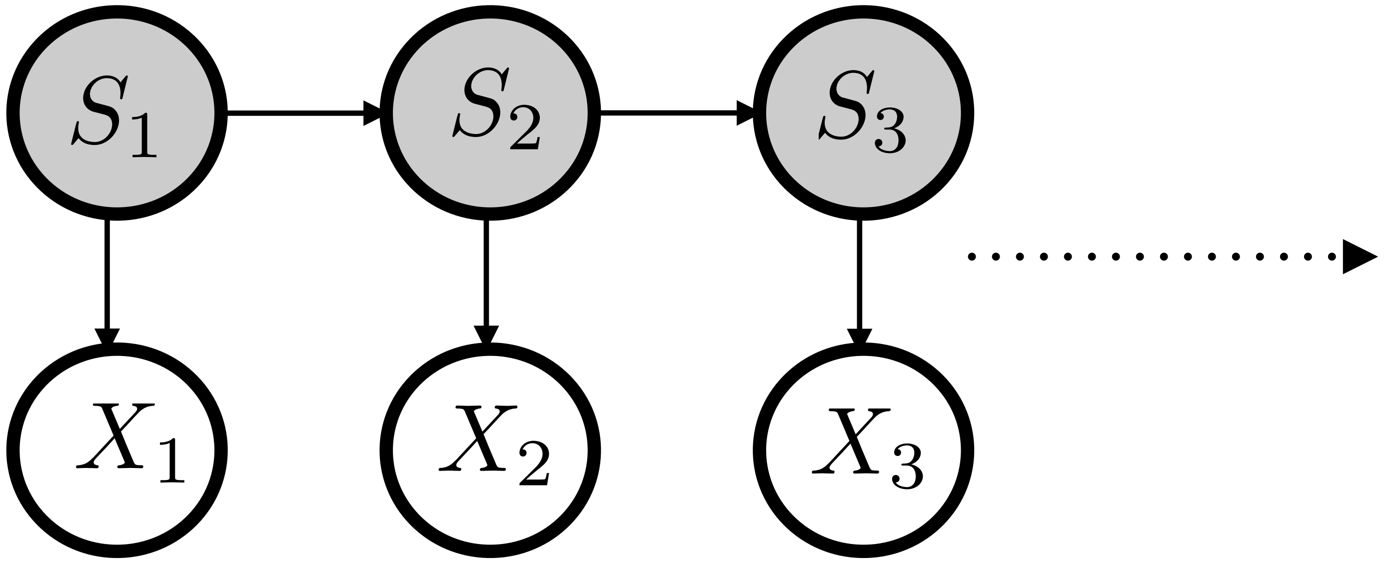
X_t's are what you hear (observation)

S_t's are the unseen locations (states)

Eg: for m x m grid we have, $K = m^2$ states

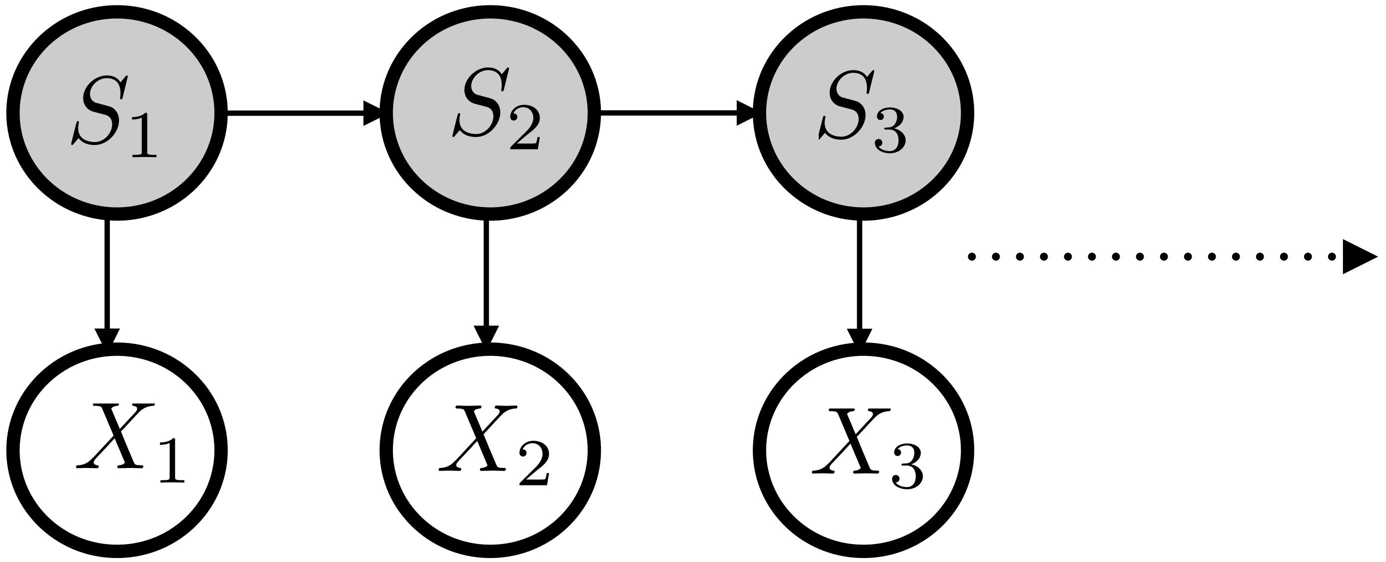
Number of alphabets = # colors you can observe

Hidden Markov Model (HMM)



Eg: for m x m grid we have, K = m^2 states

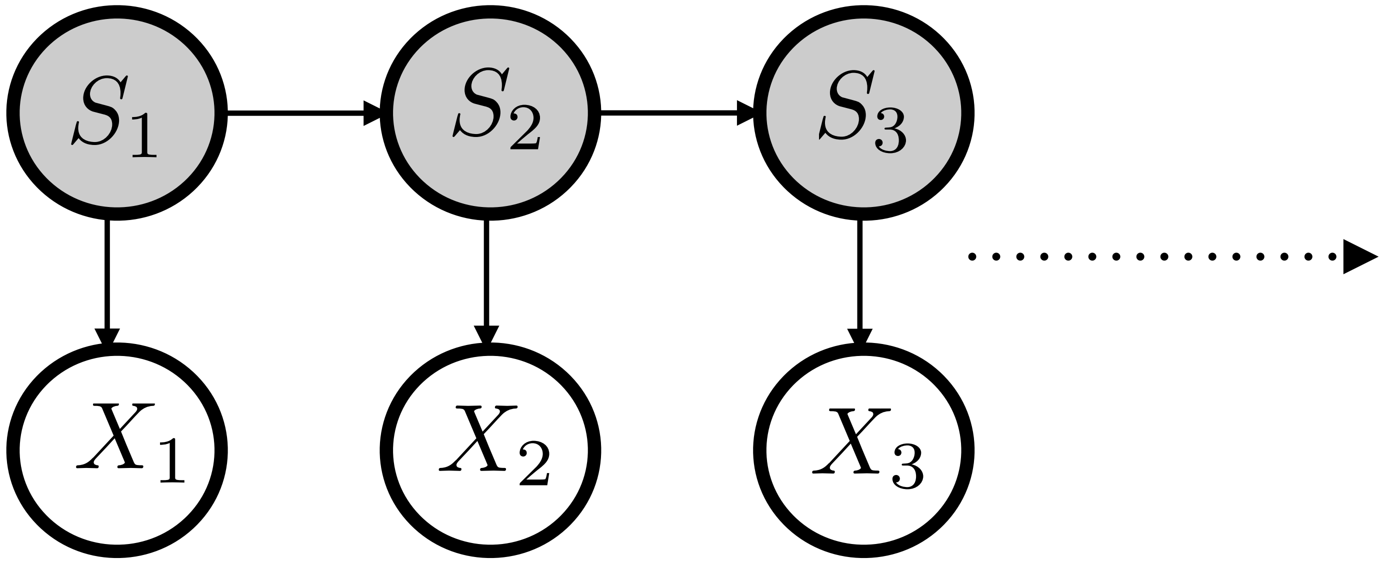
Hidden Markov Model (HMM)



Eg: for m x m grid we have, K = m^2 states

Transition matrix is K x K (too large)

Hidden Markov Model (HMM)



Eg: for m x m grid we have, $K = m^2$ states

Transition matrix is K x K (too large)

Use sampling to do approximate inference

Number of samples $n << m^4$

Inference Question

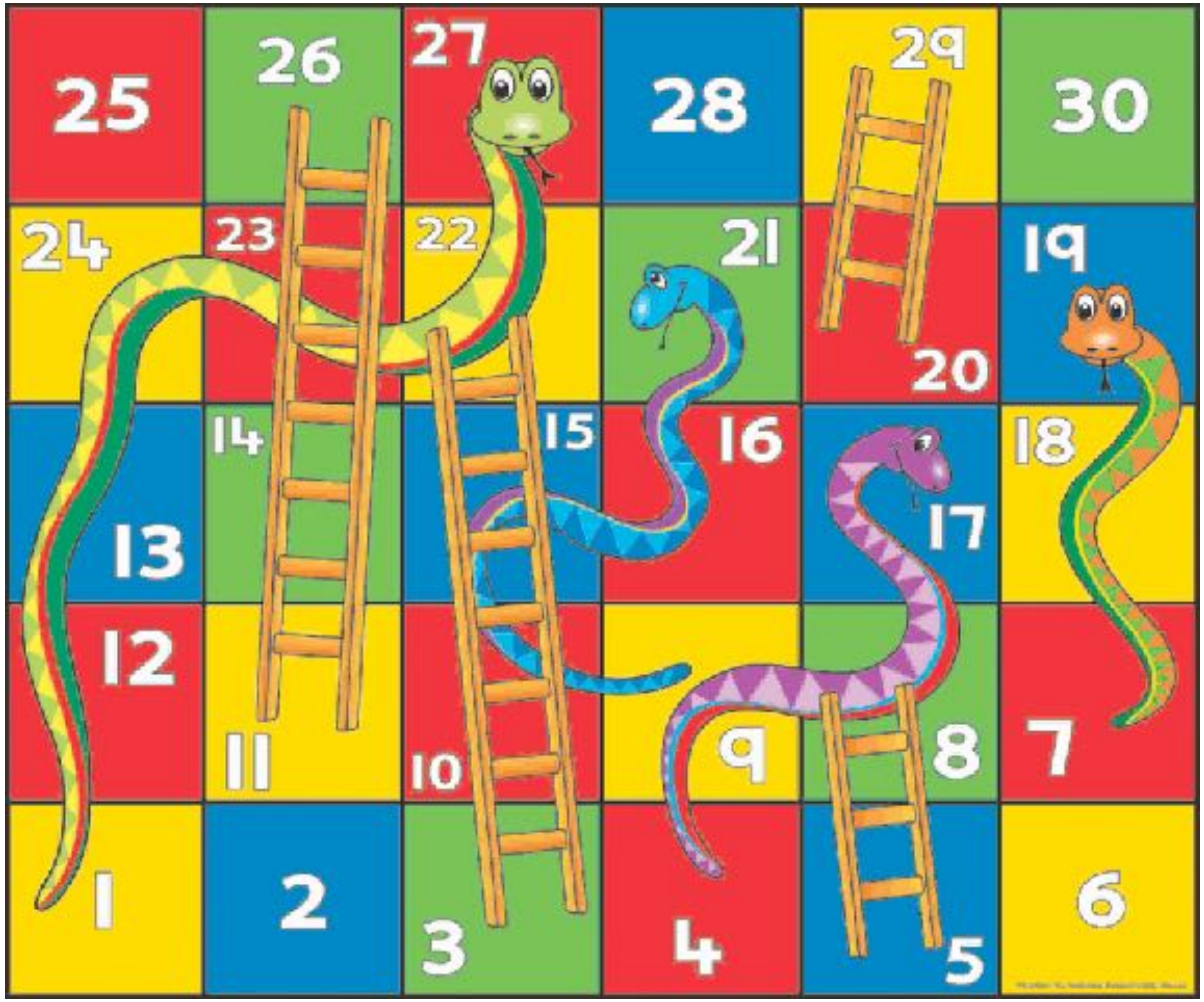
- Can we compute (efficiently and approximately)

$$P(S_t|x_1, \ldots, x_{t-1})$$

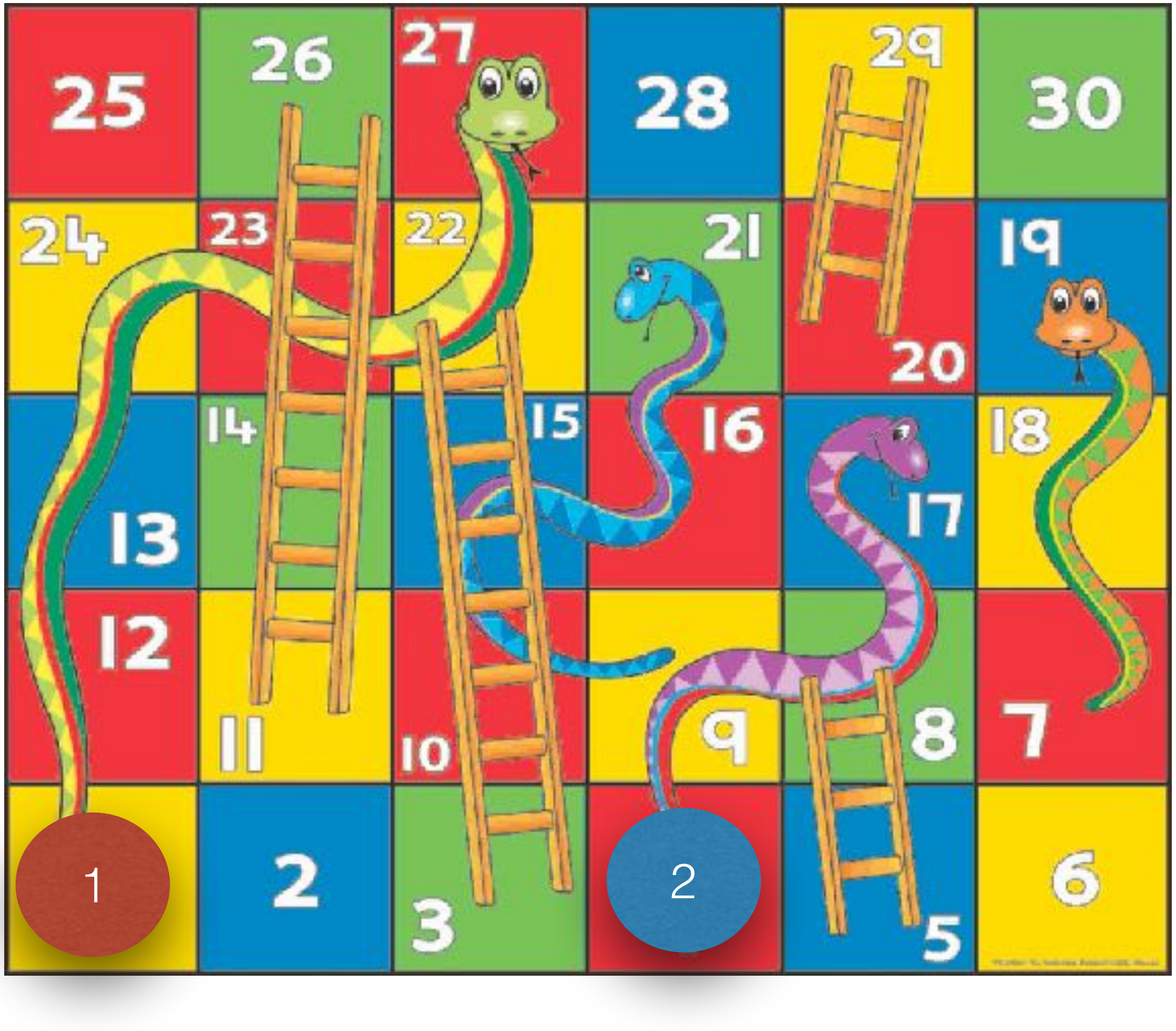
- We cant afford too much time to compute since we need to move the bot in time

- We can perform inference via sampling

Inference

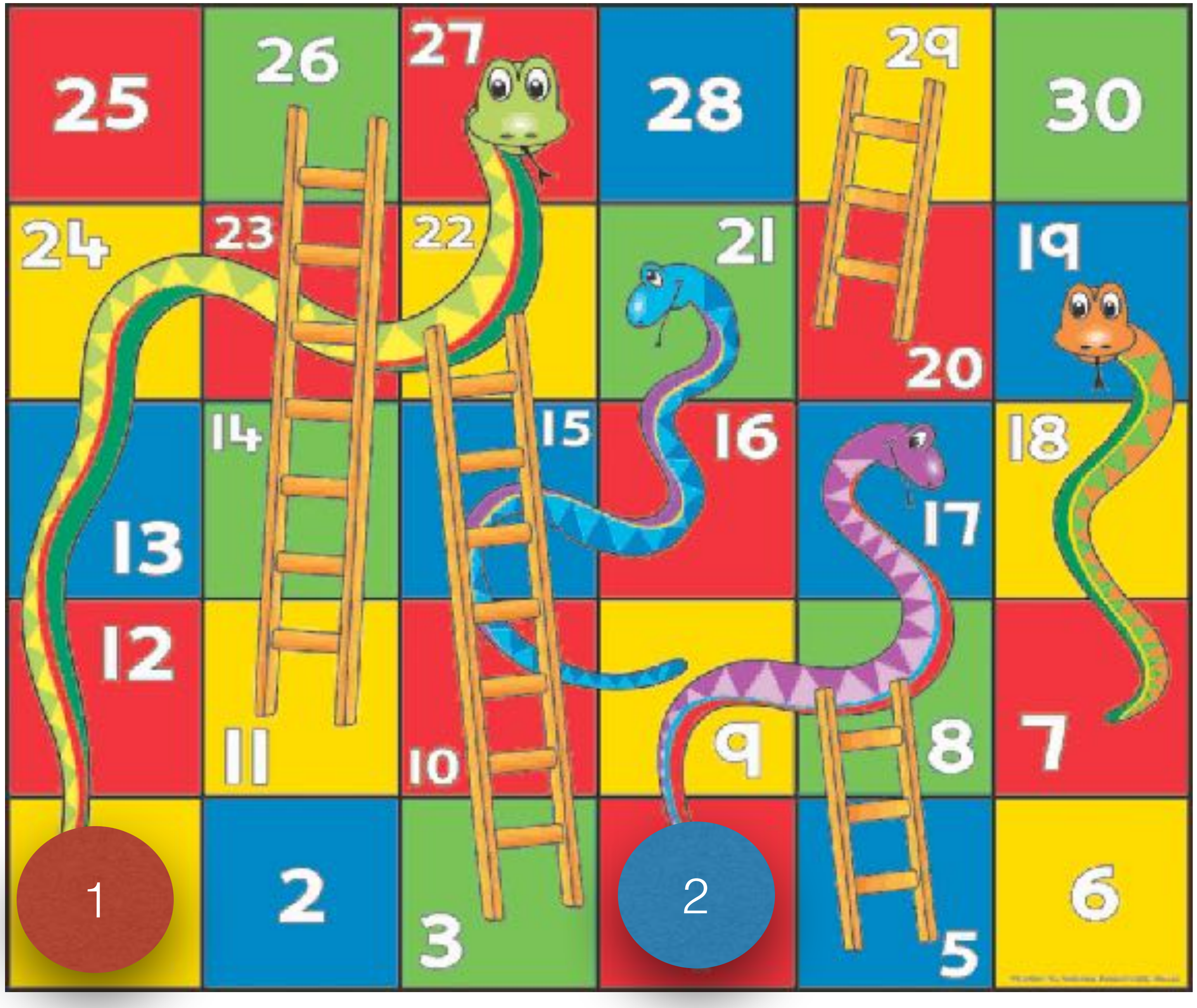

25
26
27
28
29
30
24
23
22
21
20
19
13
14
15
16
17
18
12
11
10
9
8
7
1
2
3
4
5
6

INFERENCE



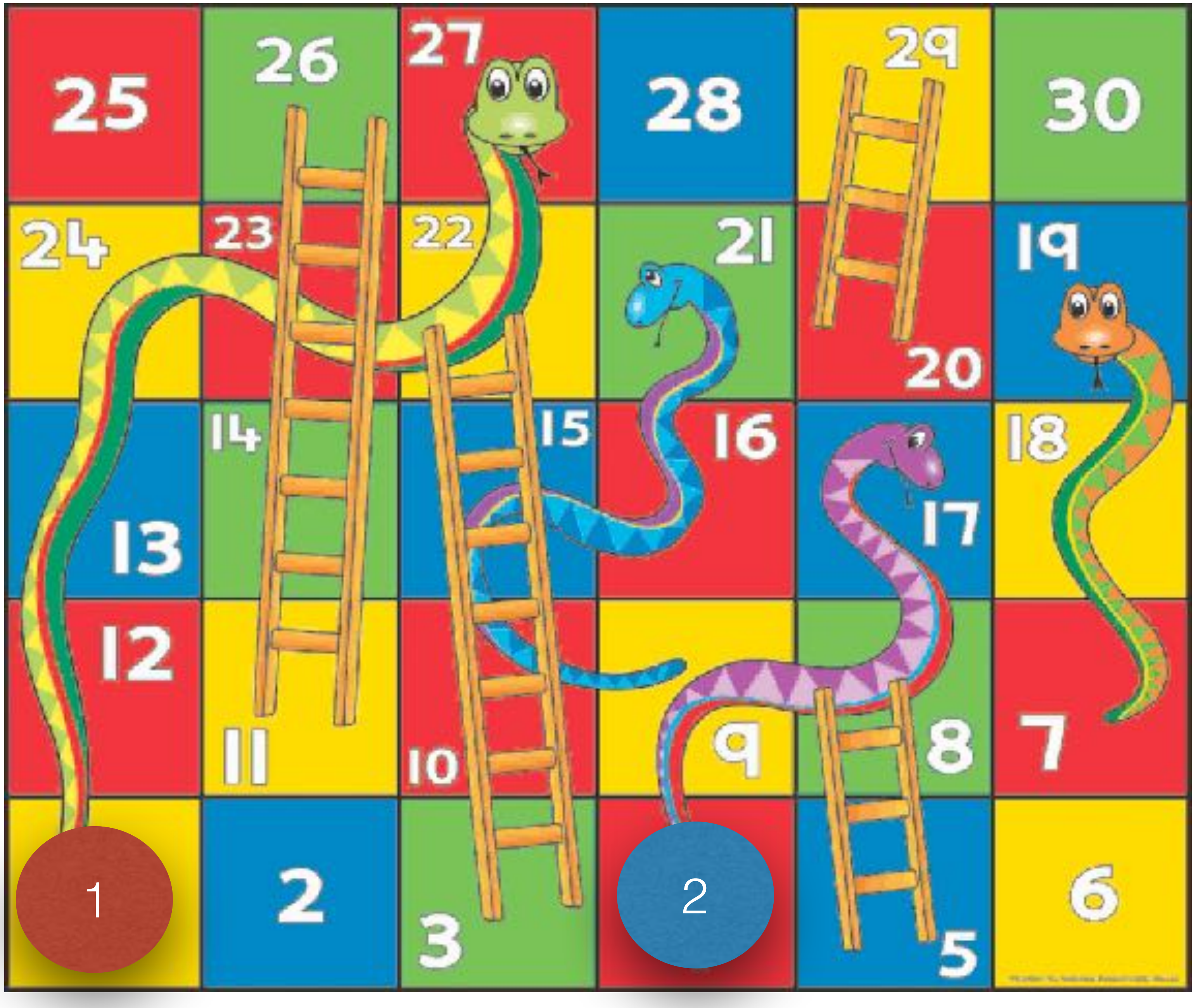
25
26
27
28
29
30
24
23
22
21
20
19
13
14
15
16
17
18
12
11
10
9
8
7
1
2
3
4
5
6

Inference



Who is more likely to win the game?

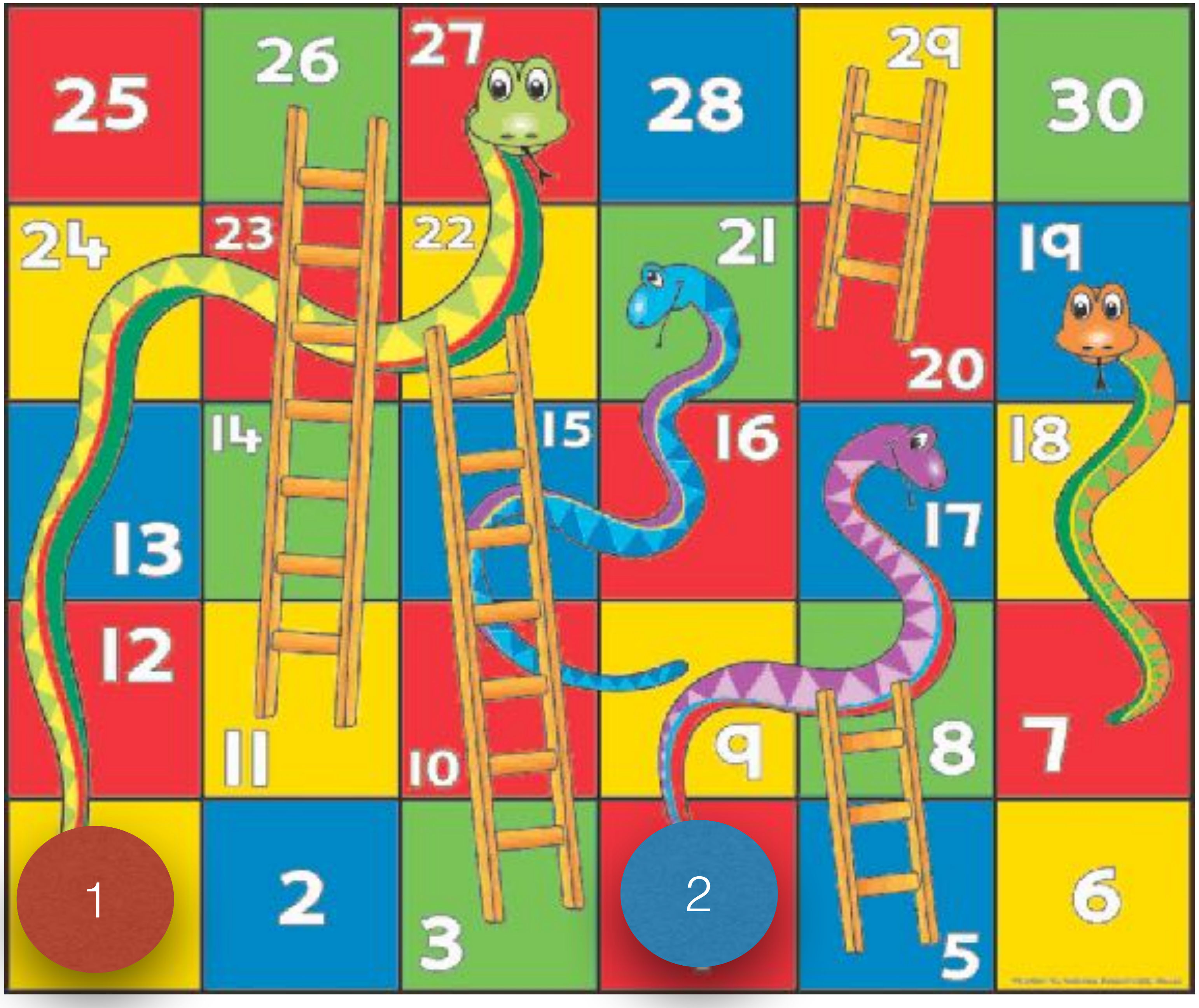
INFERENCE



Who is more likely to win the game?

Compute sum of exact probabilities of all possible sequence of moves leading to Player 1's victory

INFERENCE



Who is more likely to win the game?

Throw dice and simulate multiple games, see who wins more often

Inference Via Sampling

Inference Via Sampling

- Draw n samples from the sampling distribution

INFERENCE VIA SAMPLING

- Draw n samples from the sampling distribution
- Compute approximate probabilities by computing empirical frequencies

Inference Via Sampling

- Draw n samples from the sampling distribution
- Compute approximate probabilities by computing empirical frequencies
- Why sampling?

INFERENCE VIA SAMPLING

- Draw n samples from the sampling distribution
- Compute approximate probabilities by computing empirical frequencies
- Why sampling?
  - Getting multiple samples often faster than computing exact probabilities

Inference Via Sampling

- Draw n samples from the sampling distribution
- Compute approximate probabilities by computing empirical frequencies
- Why sampling?
    - Getting multiple samples often faster than computing exact probabilities
    - Inference is key step in learning

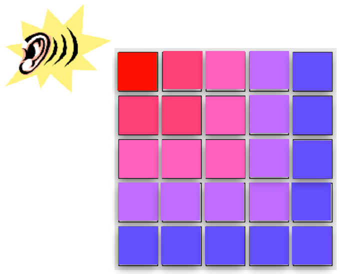
INFERENCE VIA SAMPLING

- Law of large numbers: empirical distribution using large samples approximates the true distribution
- Some approaches:
  - Rejection sampling: sample all the variables, retain only ones that match evidence
  - Importance sampling: Sample from a different distribution but then apply correction while computing empirical marginals
  - Gibbs sampling: iteratively sample from distributions closer and closer to the true one

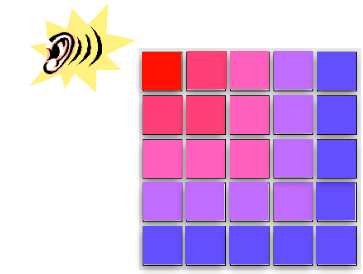
Hidden Markov Model (HMM)

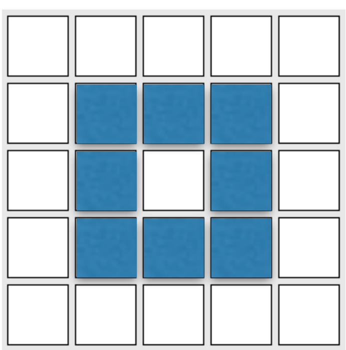
- Getting a sample from HMM given parameters is easy!
- Its accounting for observations (conditionally sampling given observations) that is hard!
- Can we use the fact that sampling from HMM is easy inference?

Hidden Markov Model (HMM)

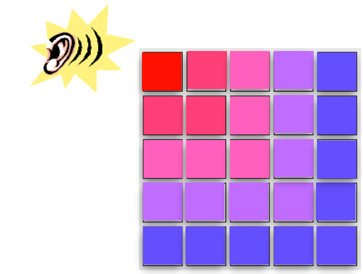


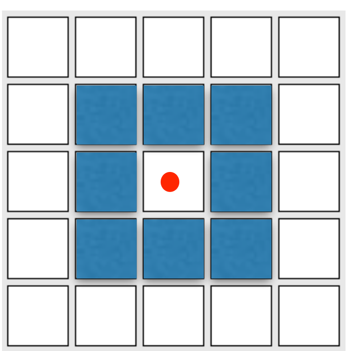
Hidden Markov Model (HMM)



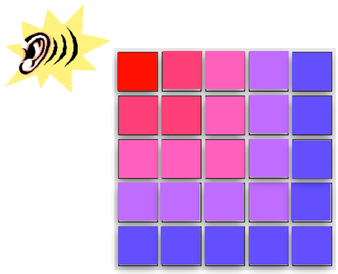


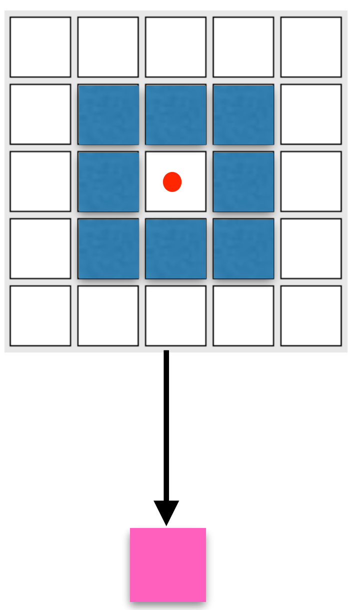
Hidden Markov Model (HMM)



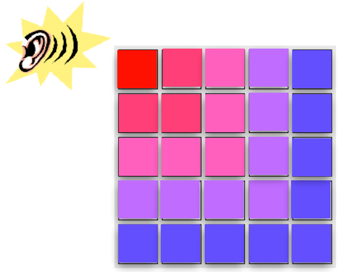


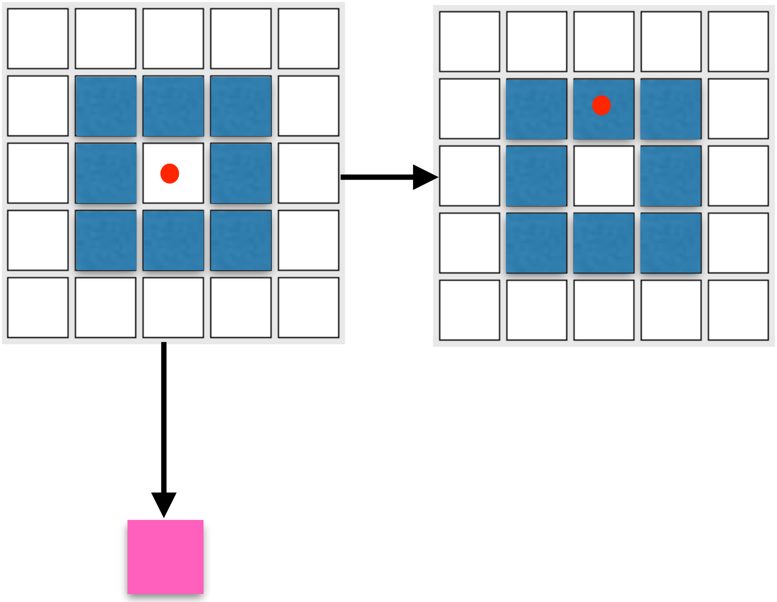
HIDDEN MARKOV MODEL (HMM)



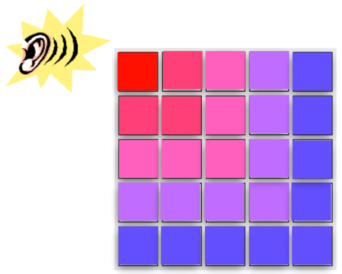


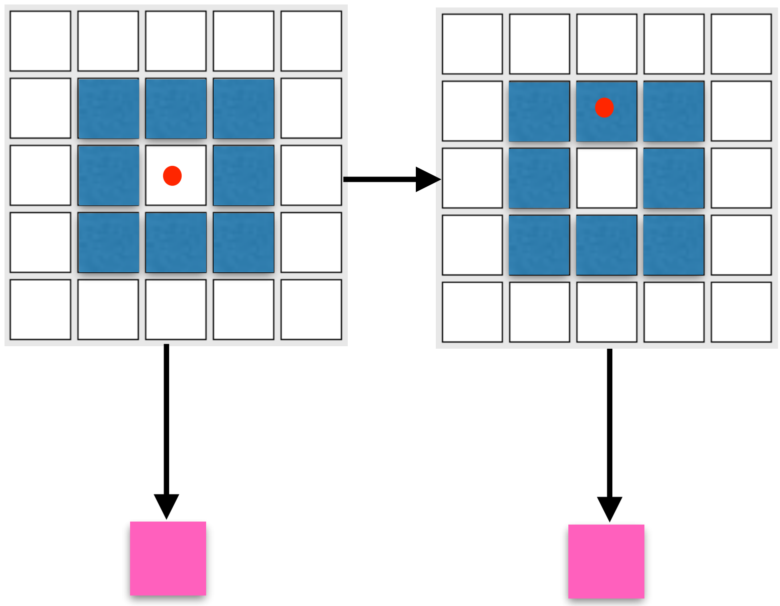
Hidden Markov Model (HMM)



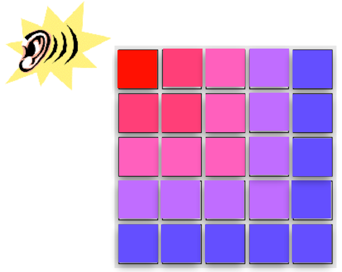


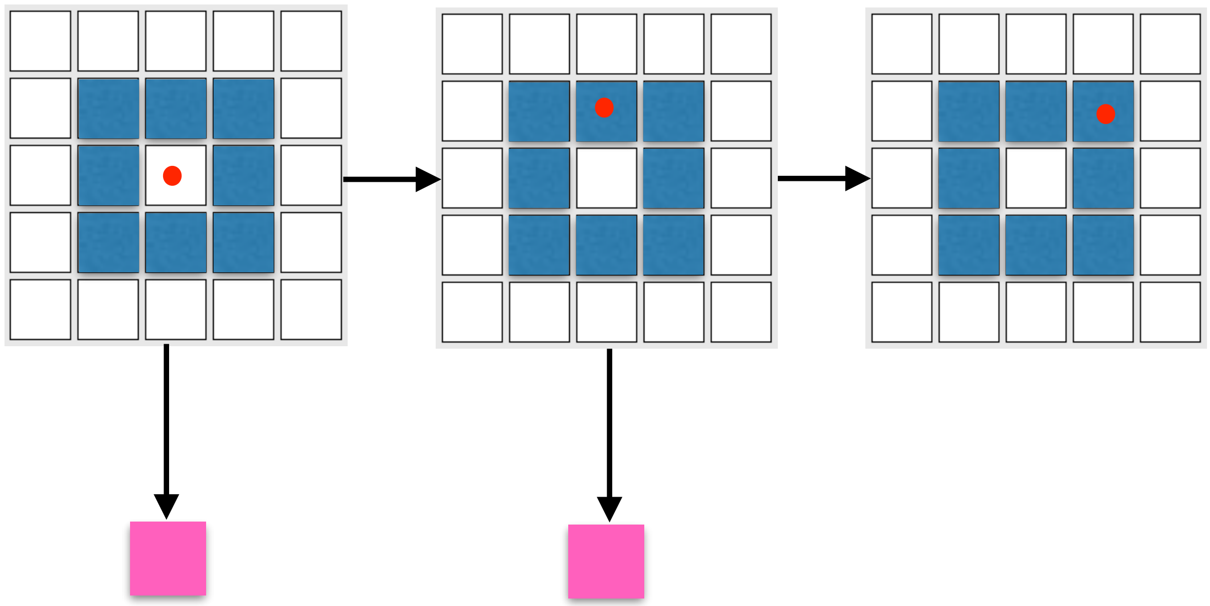
Hidden Markov Model (HMM)



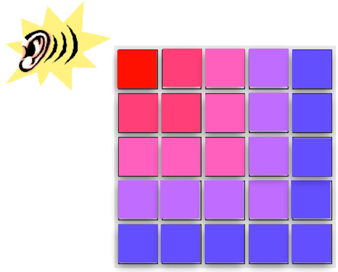


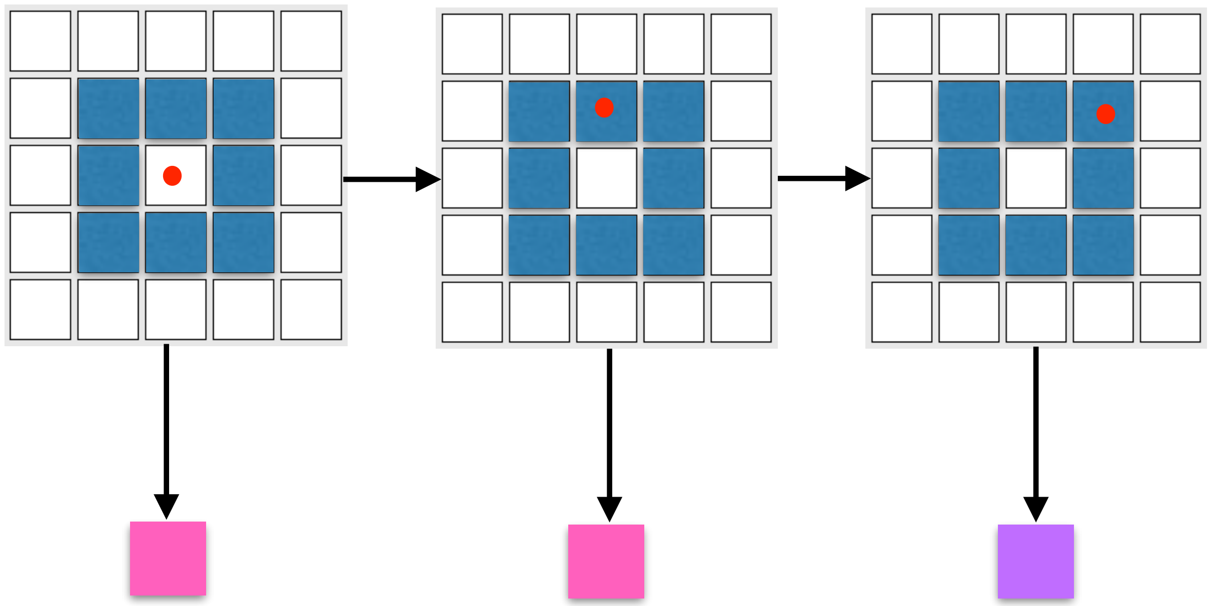
Hidden Markov Model (HMM)



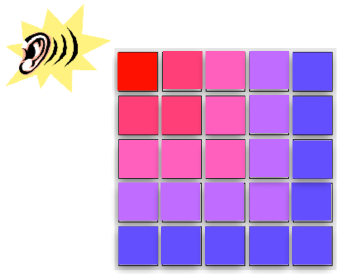


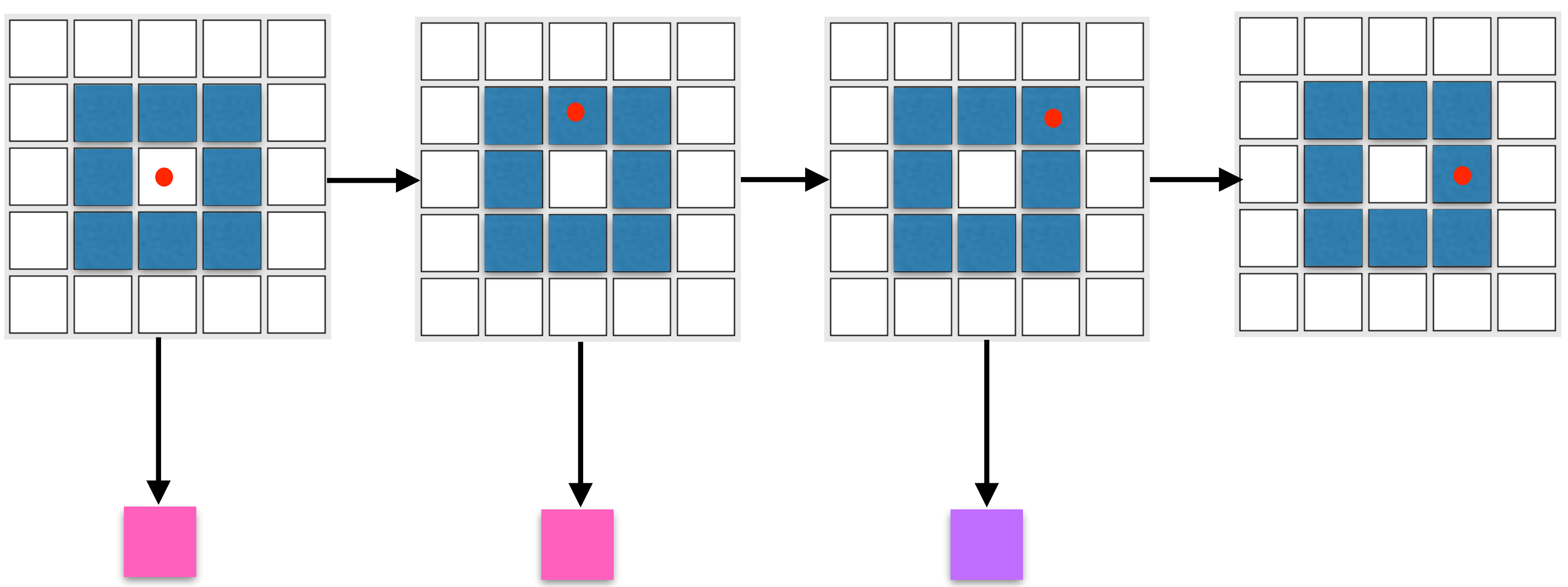
HIDDEN MARKOV MODEL (HMM)



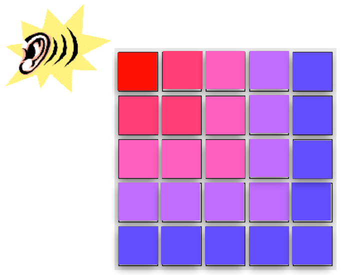


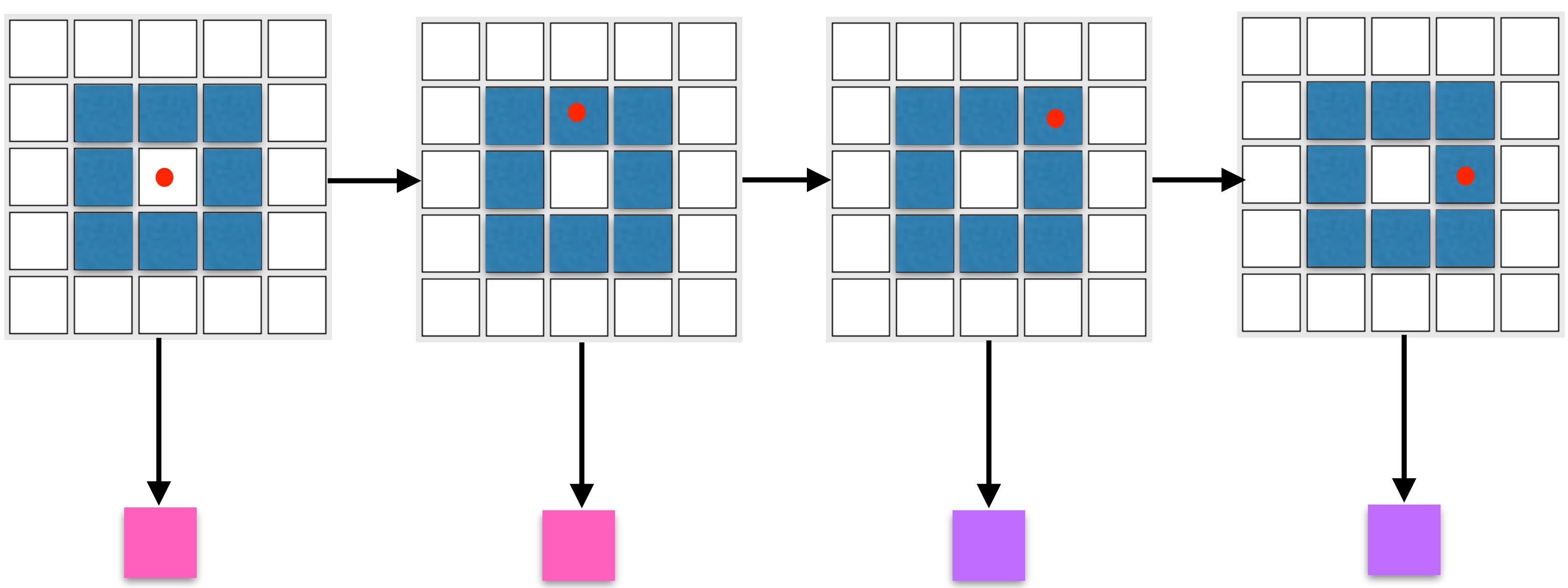
HIDDEN MARKOV MODEL (HMM)



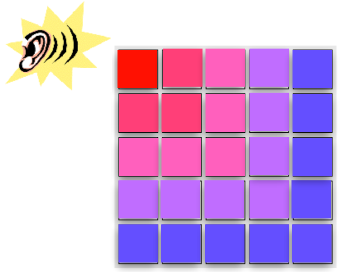


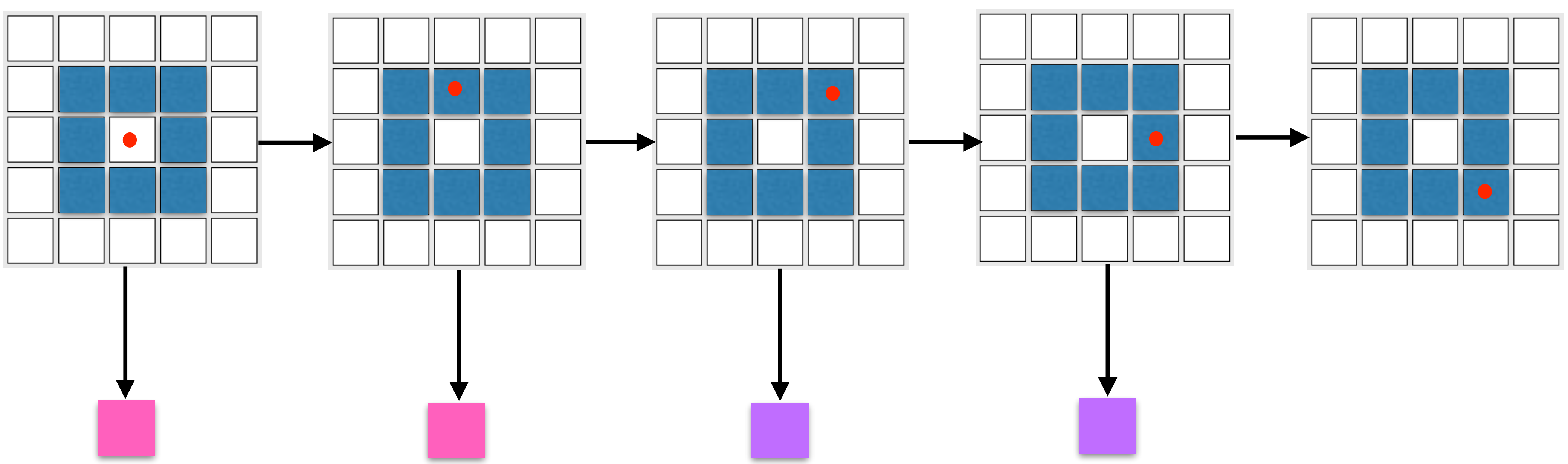
Hidden Markov Model (HMM)



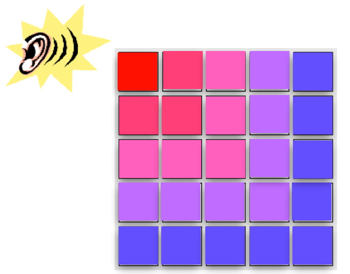


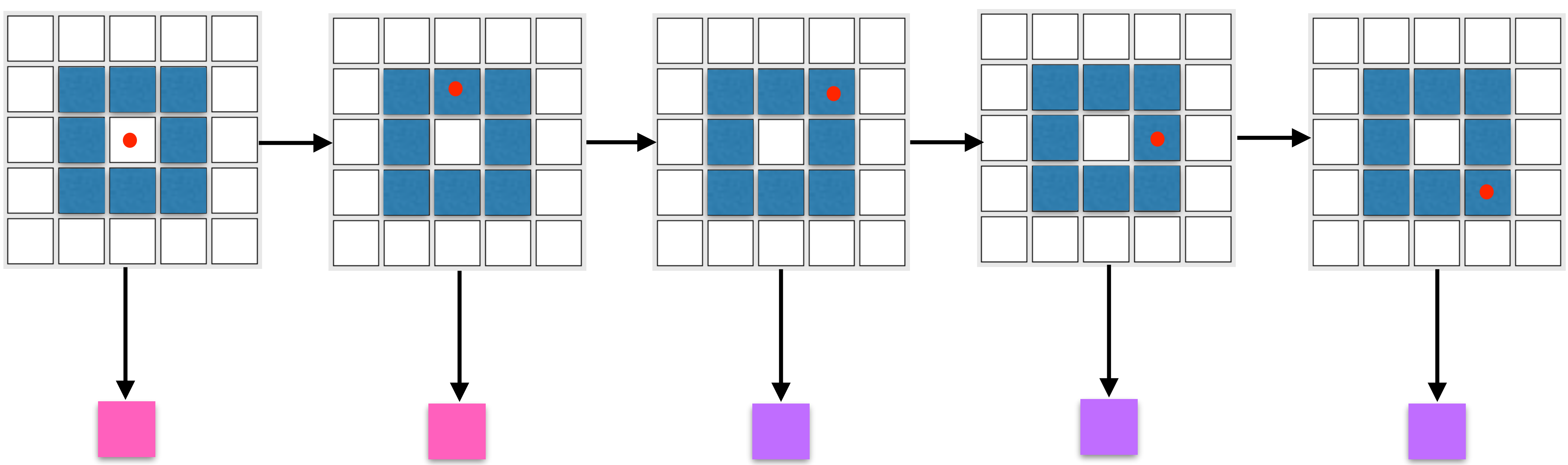
Hidden Markov Model (HMM)



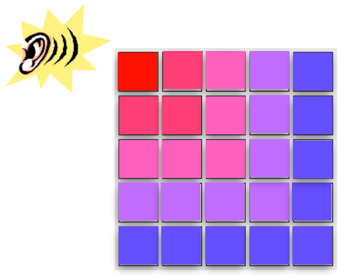


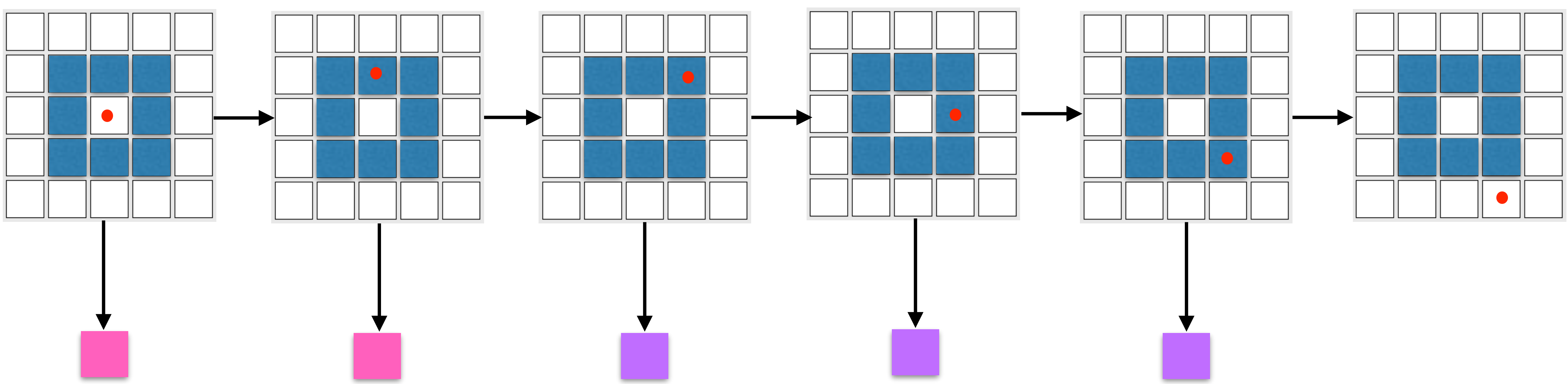
Hidden Markov Model (HMM)



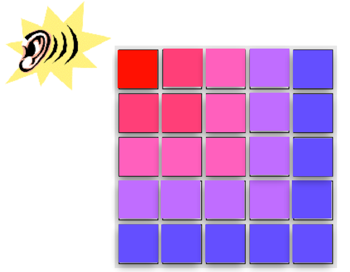


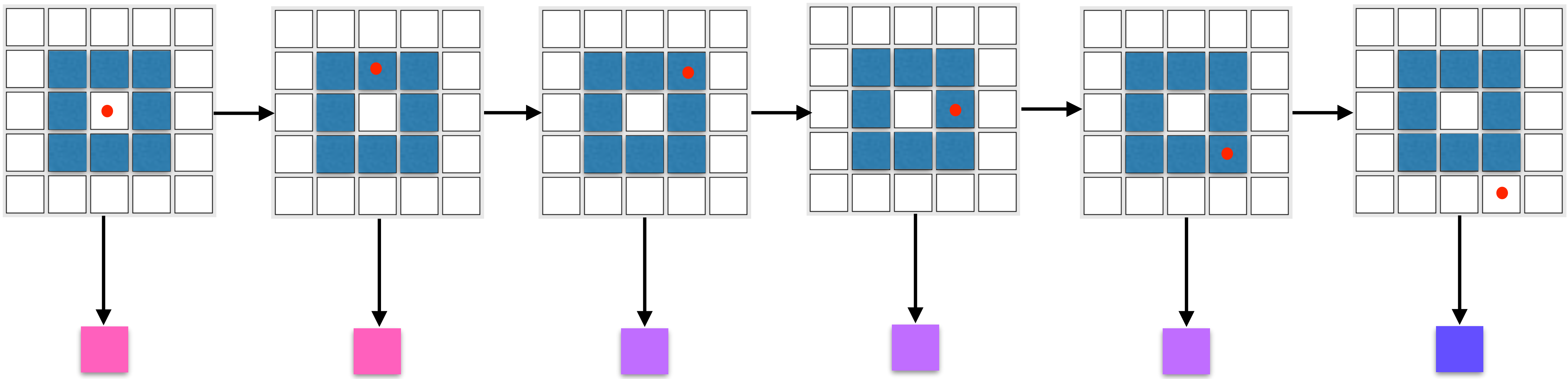
Hidden Markov Model (HMM)



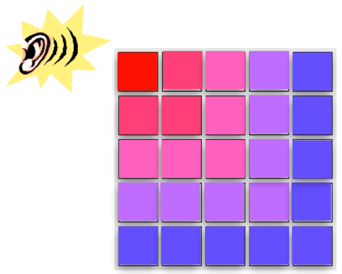


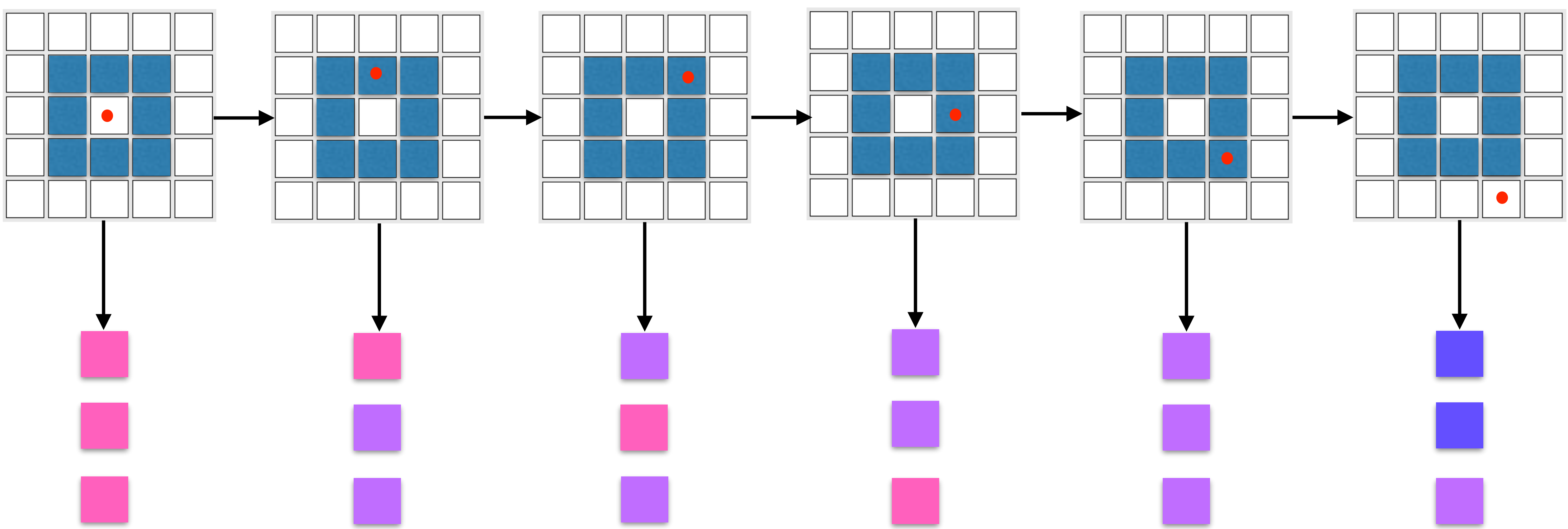
Hidden Markov Model (HMM)



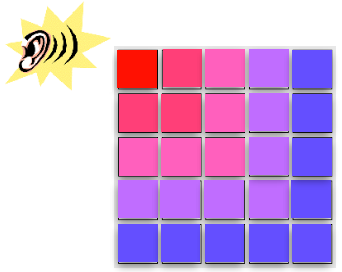


HIDDEN MARKOV MODEL (HMM)

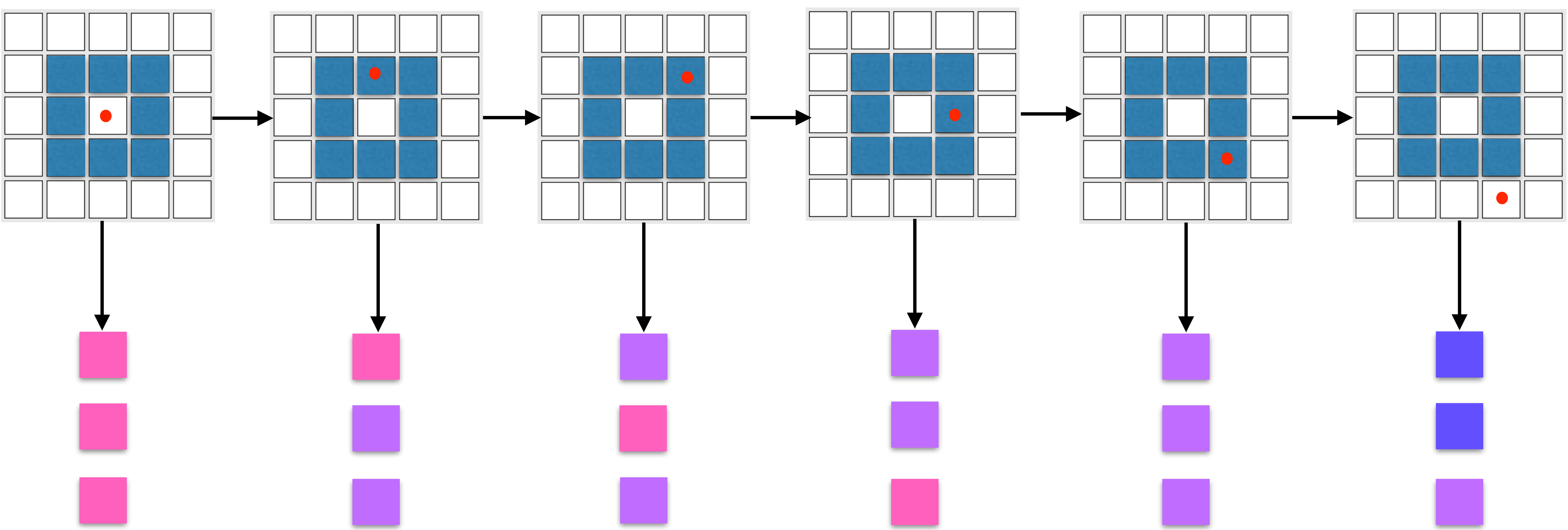




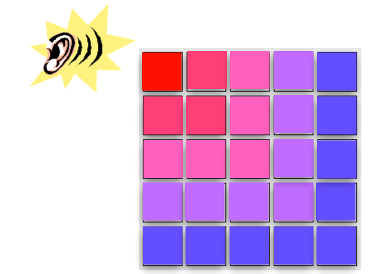
Hidden Markov Model (HMM)



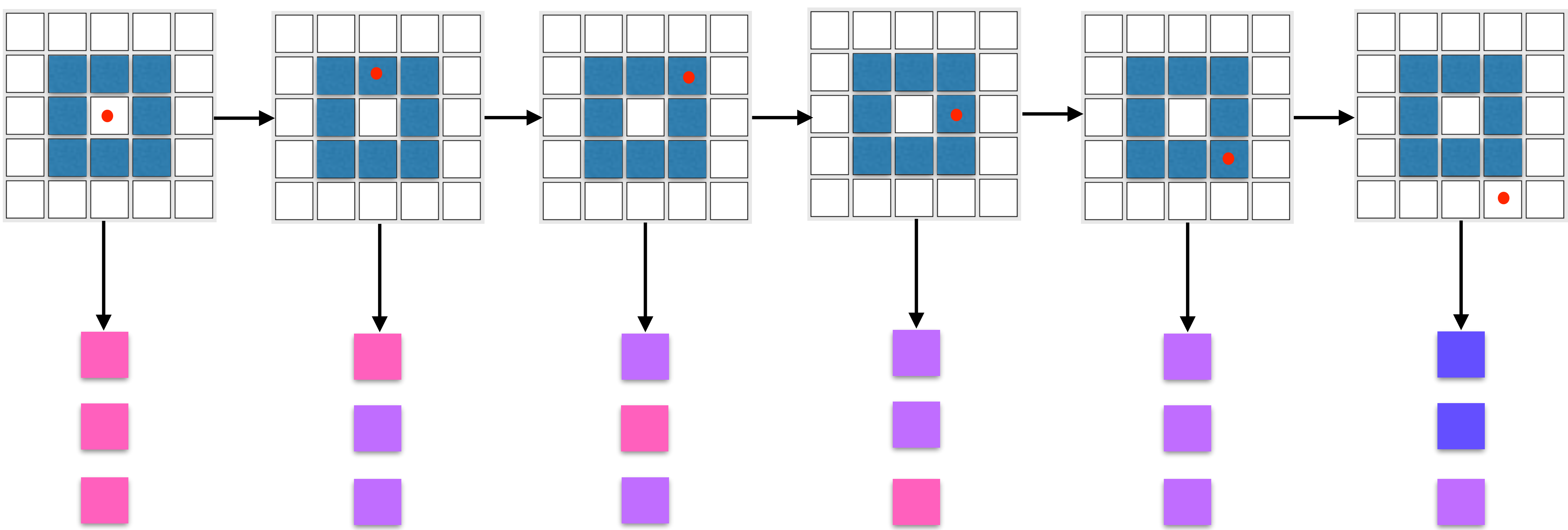
Eg: say observations were



HIDDEN MARKOV MODEL (HMM)

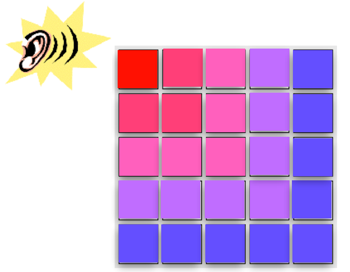


Eg: say observations were

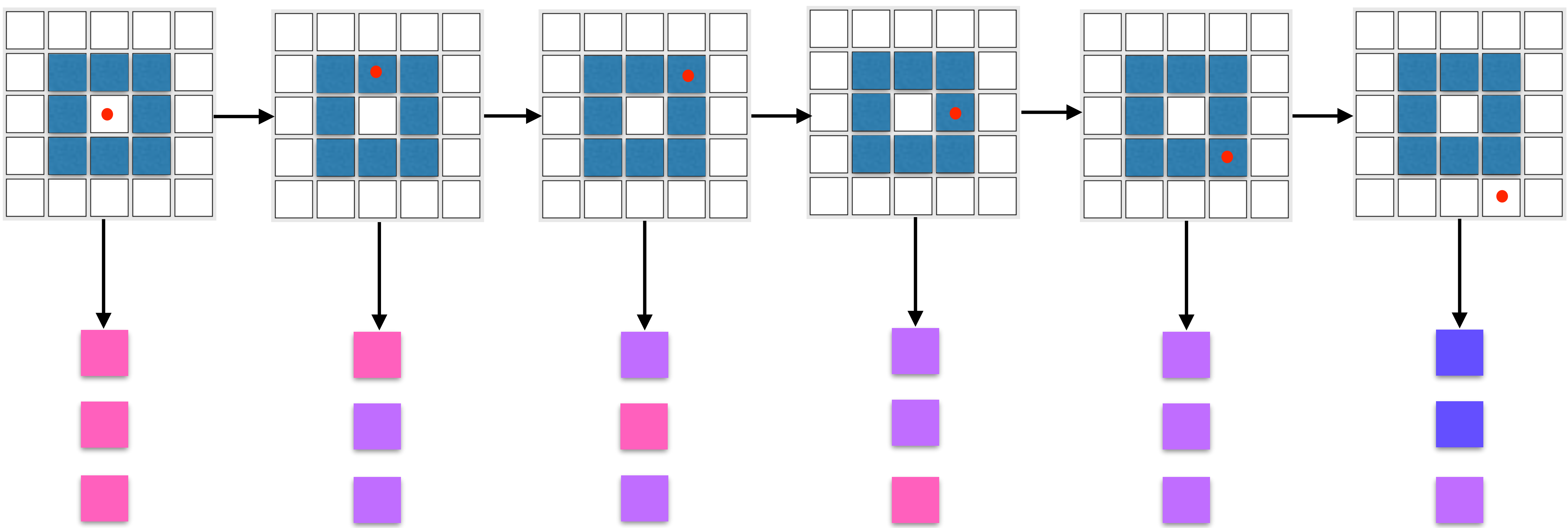


Rejection sampling: Reject samples that don’t match observations

HIDDEN MARKOV MODEL (HMM)



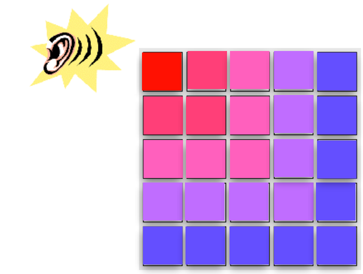
Eg: say observations were



Rejection sampling: Reject samples that don't match observations

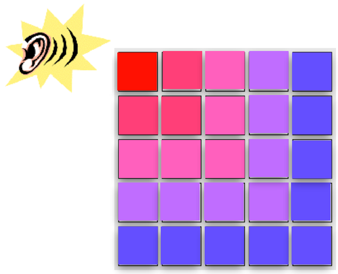
We can do this sequentially!

Hidden Markov Model (HMM)

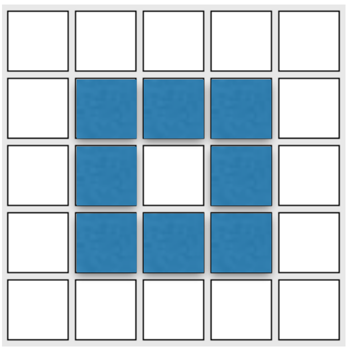


Eg: say observations were

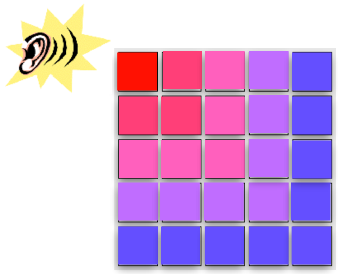
Hidden Markov Model (HMM)



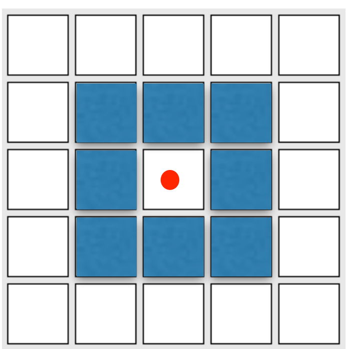
Eg: say observations were



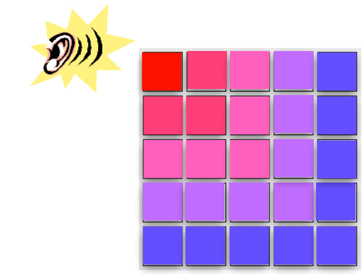
Hidden Markov Model (HMM)



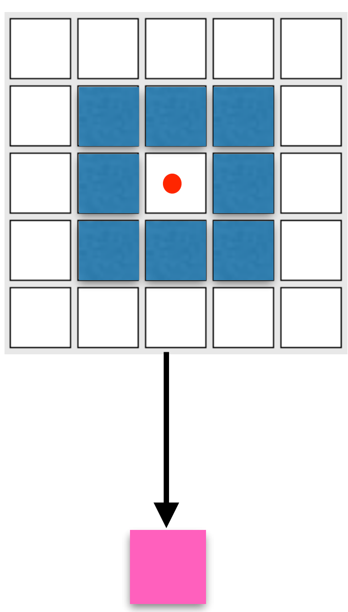
Eg: say observations were



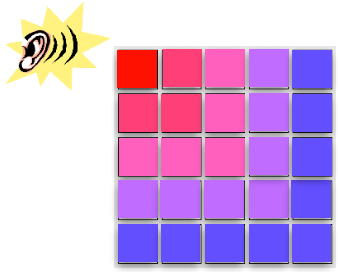
Hidden Markov Model (HMM)



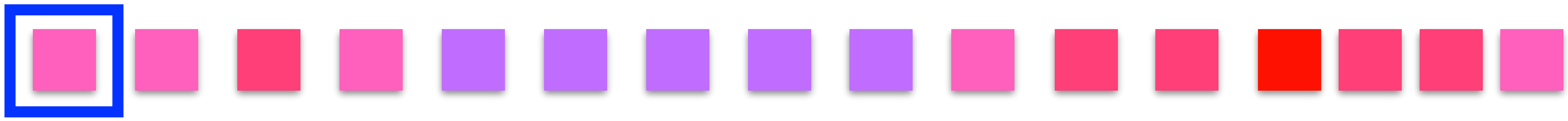
Eg: say observations were

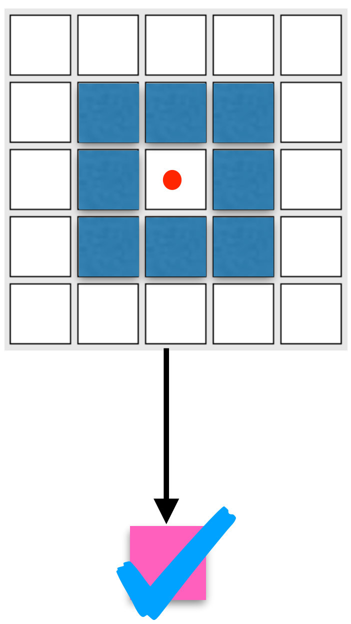


Hidden Markov Model (HMM)

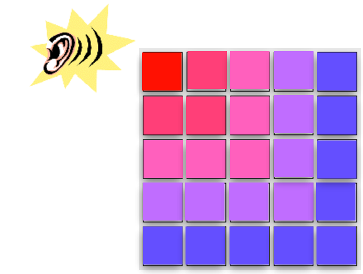


Eg: say observations were

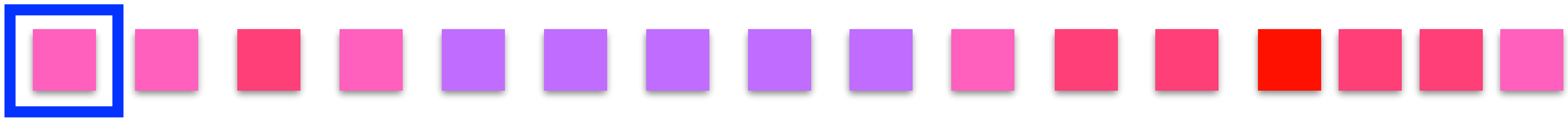


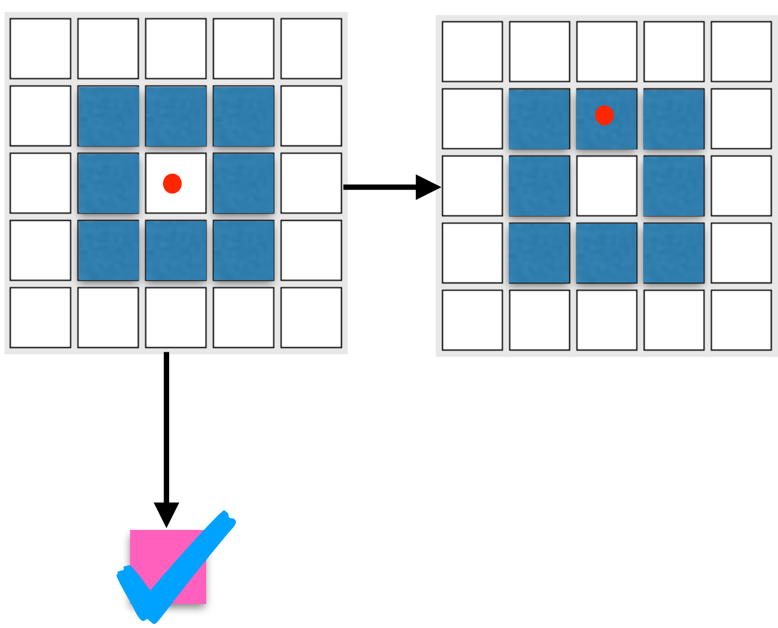


HIDDEN MARKOV MODEL (HMM)

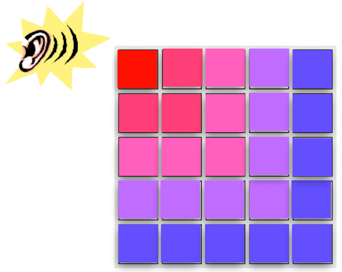


Eg: say observations were

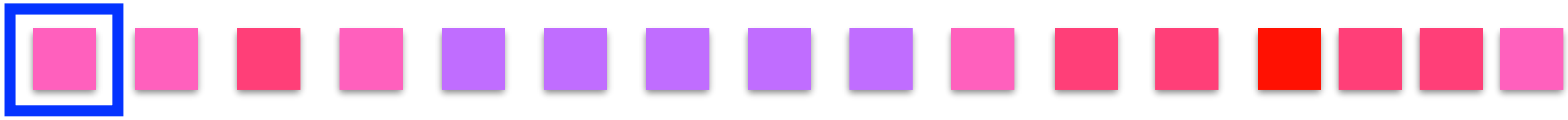


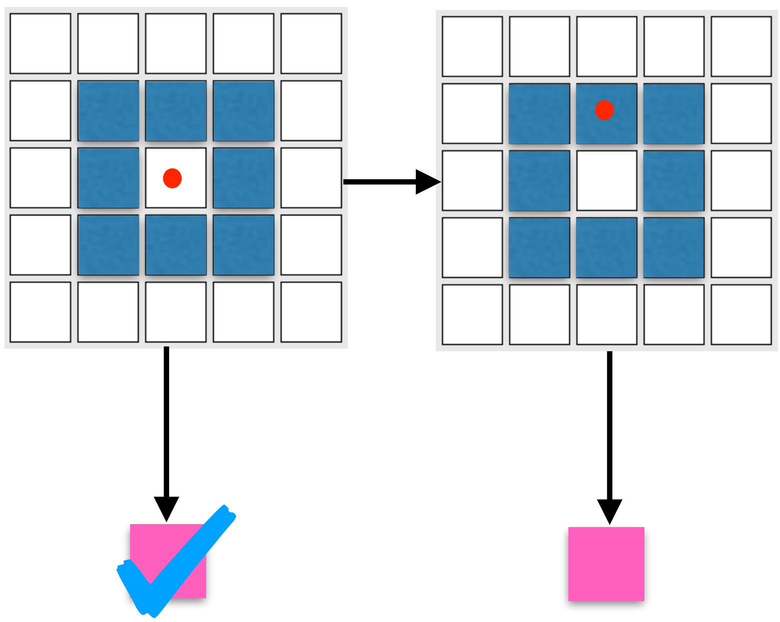


Hidden Markov Model (HMM)

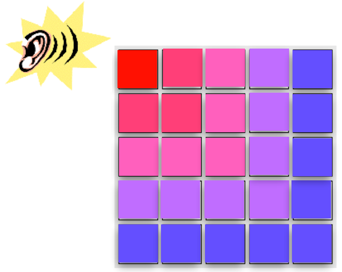


Eg: say observations were

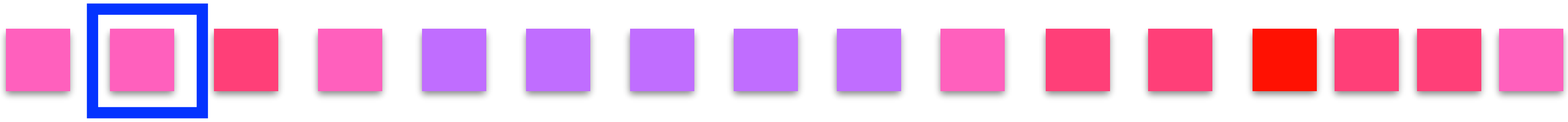


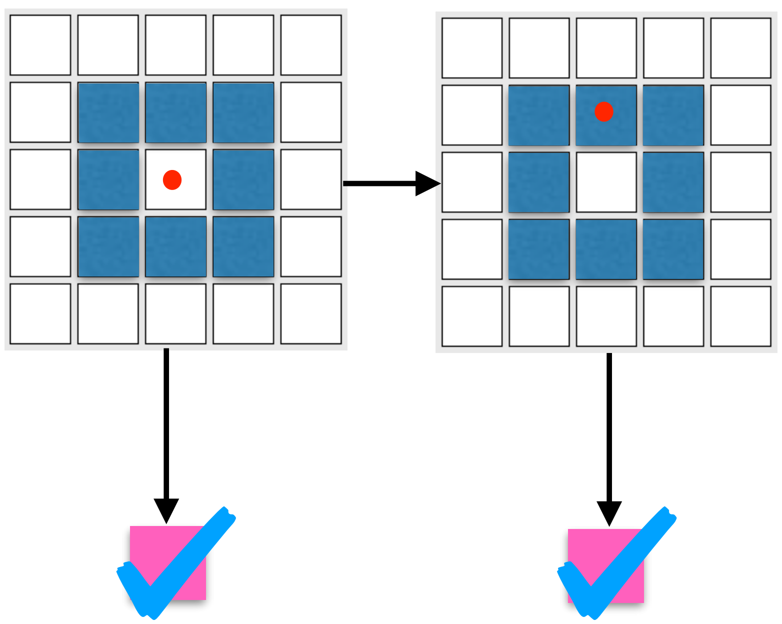


Hidden Markov Model (HMM)

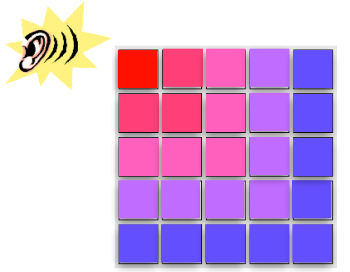


Eg: say observations were

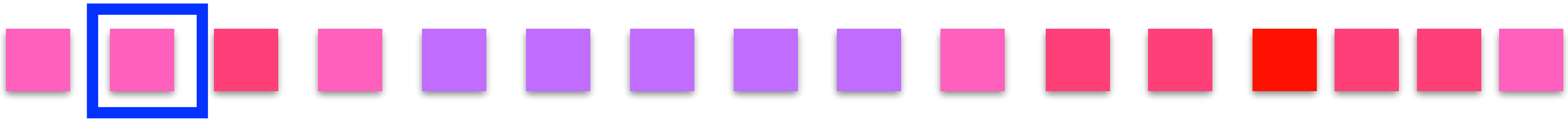


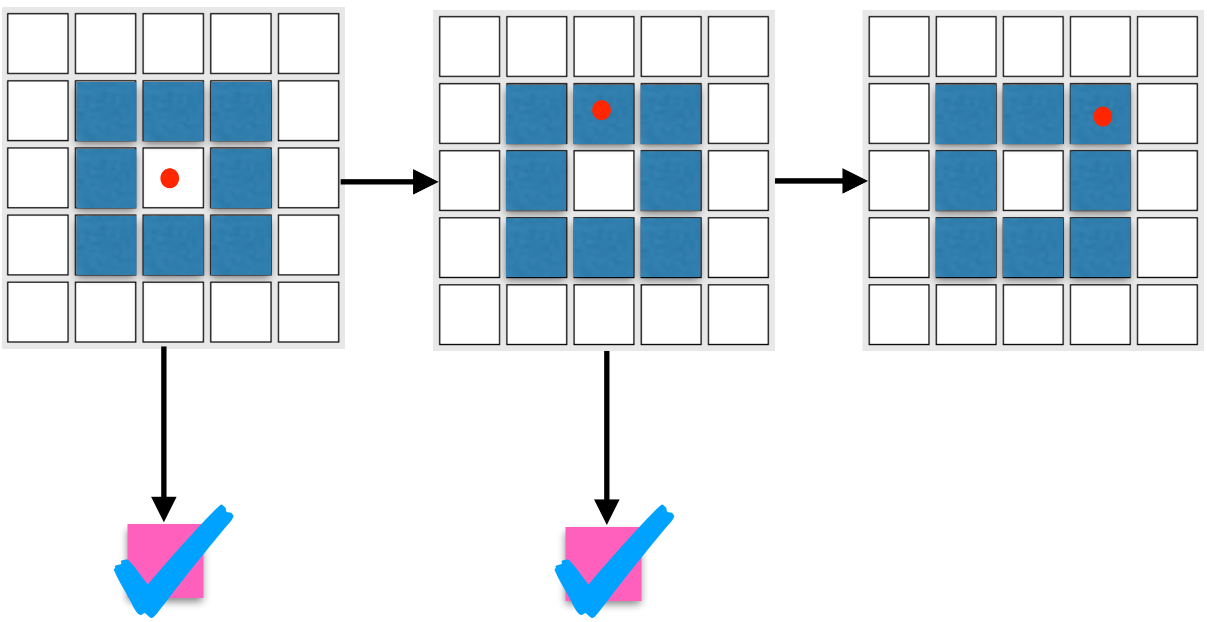


Hidden Markov Model (HMM)

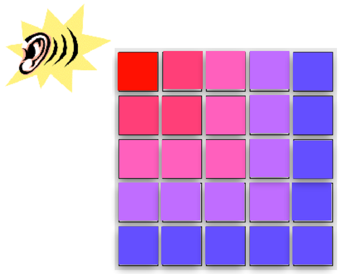


Eg: say observations were

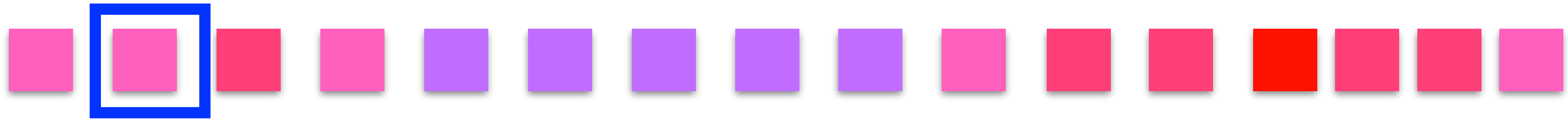


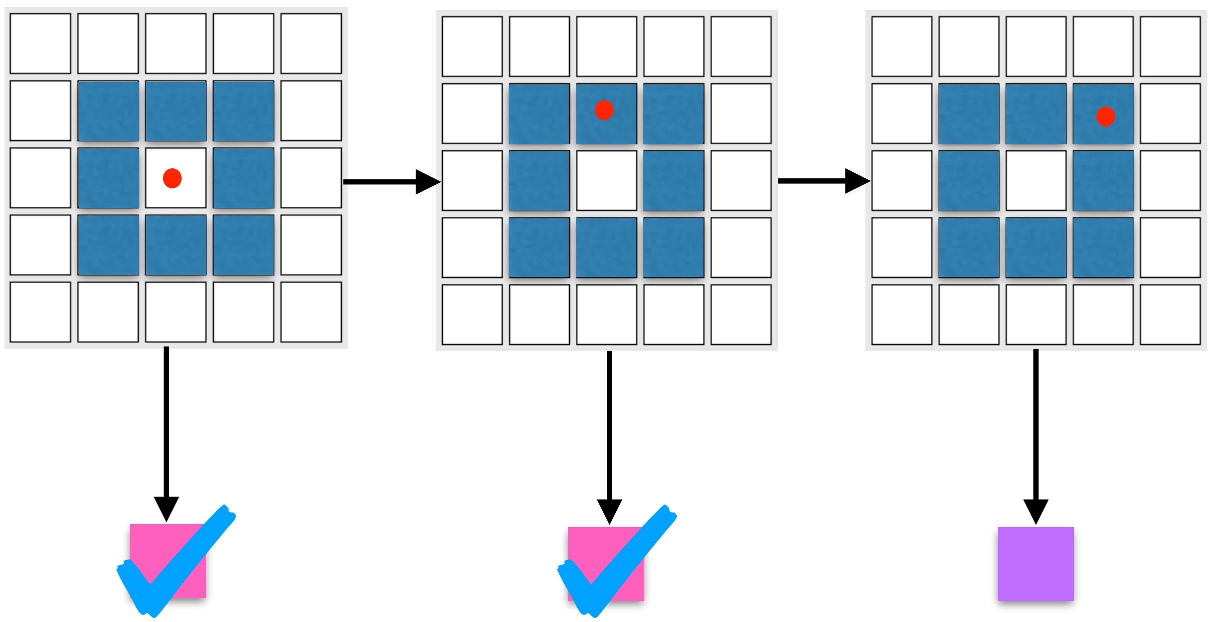


HIDDEN MARKOV MODEL (HMM)

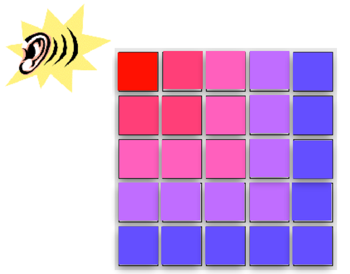


Eg: say observations were

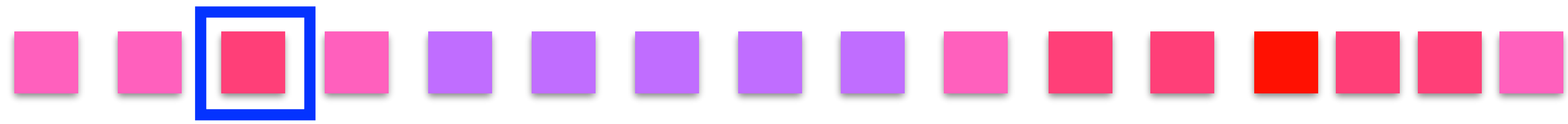


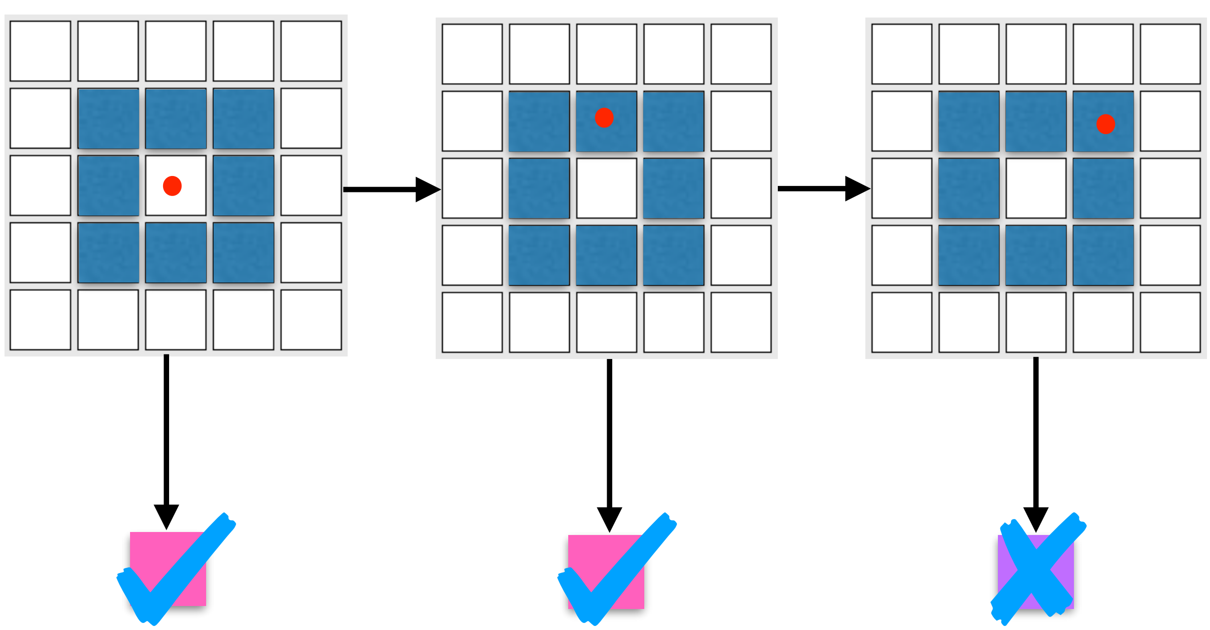


Hidden Markov Model (HMM)

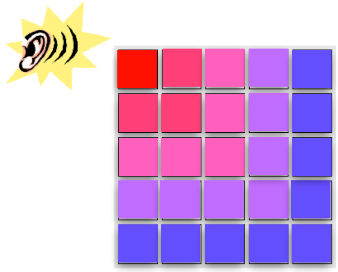


Eg: say observations were

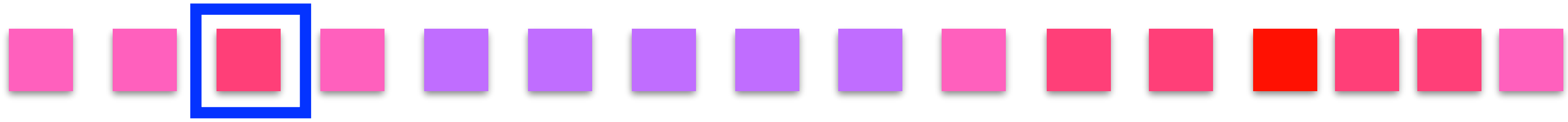


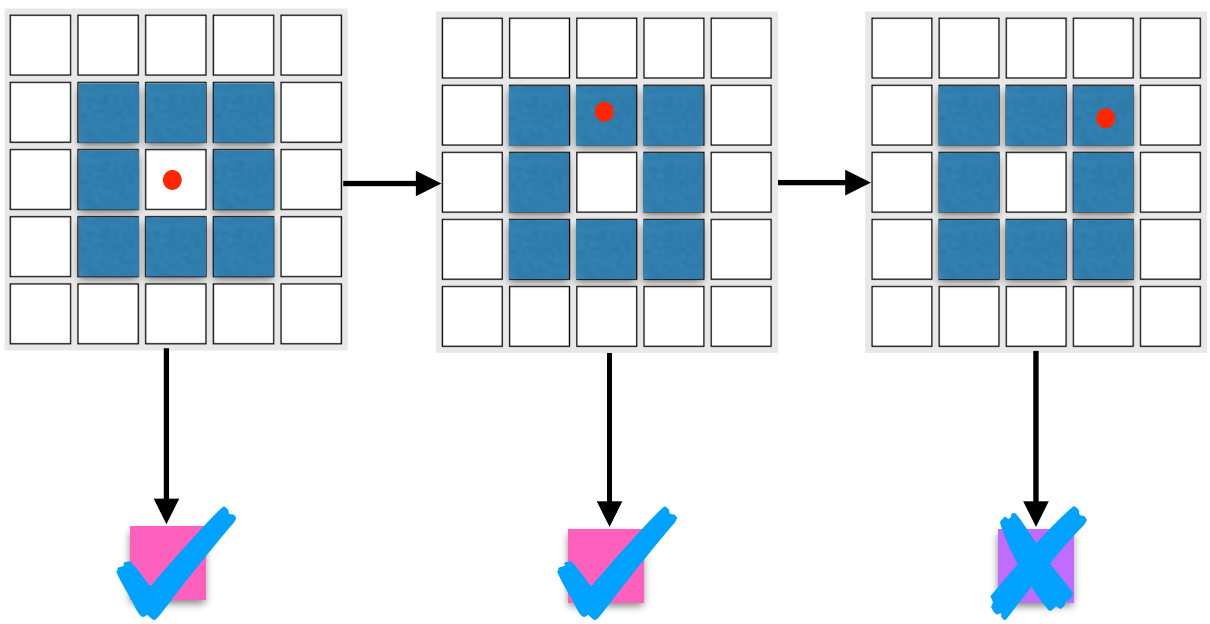


Hidden Markov Model (HMM)



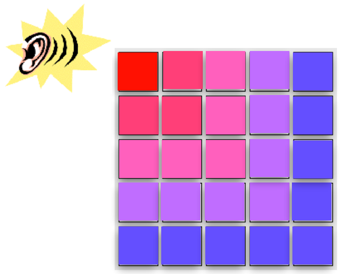
Eg: say observations were



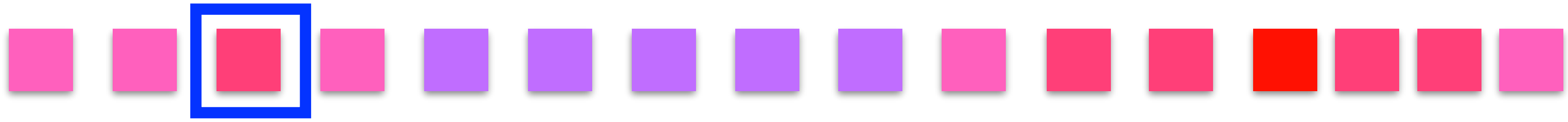


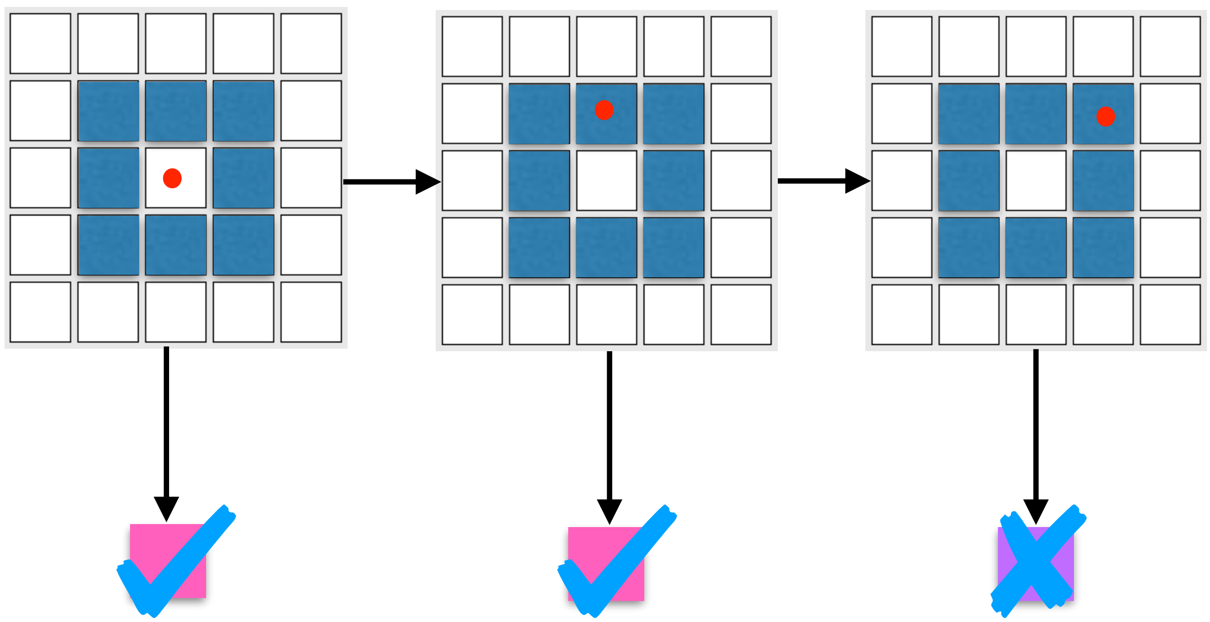
Multiple samples simultaneously.

Hidden Markov Model (HMM)



Eg: say observations were





Multiple samples simultaneously.

Problem: Most samples rejected

Importance Sampling

- We really want to draw from distribution P.
- But we can only draw from distribution Q easily
- Trick:
    - Draw $x_1, \ldots, x_n \sim Q$
    - Re-weight each sample x_t by $P(X = x_t)/Q(X = x_t)$

Importance Sampling

- Why does it work?

$$\mathbb{E}_{X \sim P}[f(X)] = \sum_x P(X = x) f(x)$$

Importance Sampling

- Why does it work?

$$\mathbb{E}_{X \sim P}[f(X)] = \sum_x P(X = x) f(x)$$
$$= \sum_x Q(X = x) \left(\frac{P(X = x)}{Q(X = x)} f(x) \right)$$

Importance Sampling

- Why does it work?

$$\begin{aligned}\mathbb{E}_{X\sim P}[f(X)] &= \sum_x P(X=x)f(x) \\ &= \sum_x Q(X=x)\left(\frac{P(X=x)}{Q(X=x)}f(x)\right) \\ &= \mathbb{E}_{X\sim Q}\left[\frac{P(X)}{Q(X)}f(X)\right]\end{aligned}$$

Importance Sampling

- Why does it work?

$$\begin{aligned}
\mathbb{E}_{X\sim P}[f(X)] &= \sum_x P(X=x)f(x) \\
&= \sum_x Q(X=x)\left(\frac{P(X=x)}{Q(X=x)}f(x)\right) \\
&= \mathbb{E}_{X\sim Q}\left[\frac{P(X)}{Q(X)}f(X)\right] \\
&\approx \frac{1}{n}\sum_{t=1}^{n}\frac{P(X=x_t)}{Q(X=x_t)}f(x_t)
\end{aligned}$$

Importance Sampling

- Why does it work?

$$
\begin{aligned}
\mathbb{E}_{X\sim P}[f(X)] &= \sum_x P(X=x) f(x) \\
&= \sum_x Q(X=x)\left(\frac{P(X=x)}{Q(X=x)} f(x)\right) \\
&= \mathbb{E}_{X\sim Q}\left[\frac{P(X)}{Q(X)} f(X)\right] \\
&\approx \frac{1}{n}\sum_{t=1}^{n} \frac{P(X=x_t)}{Q(X=x_t)} f(x_t)
\end{aligned}
$$

- Example: $f(X) = \mathbf{1}\{X \in \text{Set}\}$, then $\mathbb{E}_{X\sim P}[f(X)] = P(X \in \text{Set})$

Importance Sampling

- Why does it work?

$$\mathbb{E}_{X\sim P}[f(X)] = \sum_x P(X=x)f(x)$$
$$= \sum_x Q(X=x)\left(\frac{P(X=x)}{Q(X=x)}f(x)\right)$$
$$= \mathbb{E}_{X\sim Q}\left[\frac{P(X)}{Q(X)}f(X)\right]$$
$$\approx \frac{1}{n}\sum_{t=1}^{n}\frac{P(X=x_t)}{Q(X=x_t)}f(x_t)$$

- Example: $f(X) = \mathbf{1}\{X \in \mathsf{Set}\}$, then $\mathbb{E}_{X\sim P}[f(X)] = P(X \in \mathsf{Set})$
- Hence, using importance weighted sampling,

$$P(X \in \mathsf{Set}) \approx \frac{1}{n}\sum_{t=1}^{n}\mathbf{1}\{x_t \in \mathsf{Set}\}\frac{P(X=x_t)}{Q(X=x_t)}$$

Importance Sampling

IMPORTANCE SAMPLING

Importance Sampling

$$P(1) = 0.9, \quad \forall j \neq 1 \; P(j) = 0.1/5$$

Importance Sampling

$P(1) = 0.9, \quad \forall j \neq 1 \; P(j) = 0.1/5$

Importance Sampling

$P(1) = 0.9, \quad \forall j \neq 1 \; P(j) = 0.1/5$

$\forall j \quad Q(j) = 1/6$

Importance Sampling

$P(1) = 0.9, \quad \forall j \neq 1 \; P(j) = 0.1/5$

$\forall j \quad Q(j) = 1/6$

$$\text{Set} = \{2, 4, 6\}$$

Importance Sampling

$$P(1) = 0.9, \quad \forall j \neq 1 \ P(j) = 0.1/5$$

$$\forall j \quad Q(j) = 1/6$$

$$\text{Set} = \{2, 4, 6\}$$

What is P(Set)?

Importance Sampling

$P(1) = 0.9, \quad \forall j \neq 1 \ P(j) = 0.1/5$ $\quad \forall j \quad Q(j) = 1/6$

$$\text{Set} = \{2, 4, 6\}$$

What is P(Set)?

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}\{x_t \in \{2,4,6\}\}\frac{P(x_t)}{Q(x_t)} = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}\{x_t \in \{2,4,6\}\}\frac{0.1/5}{1/6}$$

Importance Sampling

$$P(1) = 0.9, \quad \forall j \neq 1 \ P(j) = 0.1/5$$

$$\forall j \quad Q(j) = 1/6$$

$$\text{Set} = \{2, 4, 6\}$$

What is P(Set)?

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}\{x_t \in \{2,4,6\}\} \frac{P(x_t)}{Q(x_t)} = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}\{x_t \in \{2,4,6\}\} \frac{0.1/5}{1/6}$$

$$= 0.12 \times \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}\{x_t \in \{2,4,6\}\} \approx 0.12 \times 0.5 = 0.06$$

Importance Sampling

- We had the problem of too many rejections because probability of getting our sample to match exactly the observation is very low!

Importance Sampling

- We had the problem of too many rejections because probability of getting our sample to match exactly the observation is very low!
- How do we fix this?

Importance Sampling

- We had the problem of too many rejections because probability of getting our sample to match exactly the observation is very low!
- How do we fix this?
- Fix observations and sample only states from the markov chain!

Importance Sampling

- We had the problem of too many rejections because probability of getting our sample to match exactly the observation is very low!
- How do we fix this?
- Fix observations and sample only states from the markov chain!
- Desired distribution P: $P(S_1, \ldots, S_N | X_1 = x_1, \ldots, X_N = x_N)$

Importance Sampling

- We had the problem of too many rejections because probability of getting our sample to match exactly the observation is very low!
- How do we fix this?
- Fix observations and sample only states from the markov chain!
- Desired distribution P: $P(S_1, \ldots, S_N | X_1 = x_1, \ldots, X_N = x_N)$
- Sampling distribution Q: $P(S_1, \ldots, S_N)$

Importance Sampling

For a given sample $s_1, \ldots, s_N$, importance weight given by:

$$\frac{P(S_1 = s_1, \ldots, S_N = s_n | X_1 = x_1, \ldots, X_N = x_N)}{P(S_1 = s_1, \ldots, S_N = s_N)}$$

Importance Sampling

For a given sample $s_1, \ldots, s_N$, importance weight given by:

$$\frac{P(S_1 = s_1, \ldots, S_N = s_n | X_1 = x_1, \ldots, X_N = x_N)}{P(S_1 = s_1, \ldots, S_N = s_N)}$$

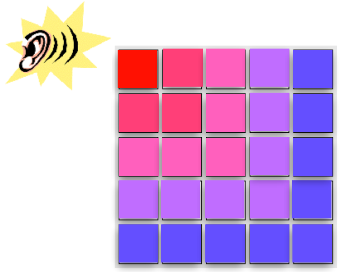
$$= \frac{P(X_1 = x_1, \ldots, X_N = x_N | S_1 = s_1, \ldots, S_N = s_n)}{P(X_1 = x_1, \ldots, X_N = x_N)}$$

Importance Sampling

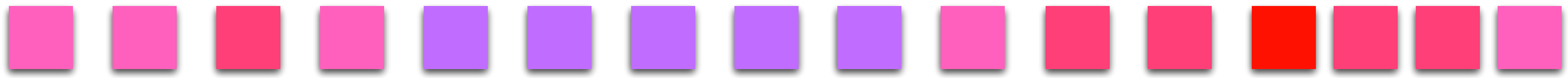
For a given sample $s_1, \ldots, s_N$, importance weight given by:

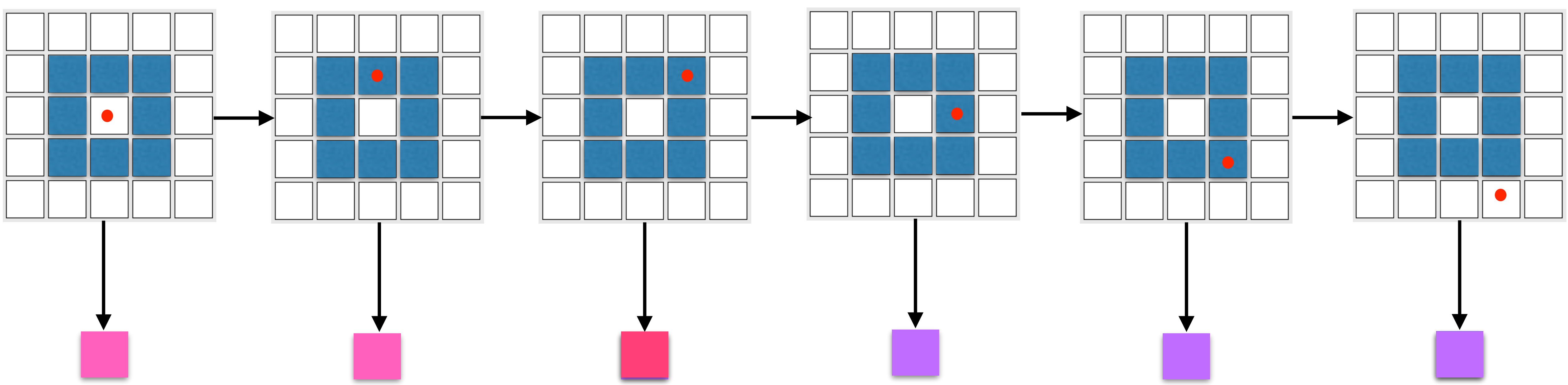
$$\frac{P(S_1 = s_1, \ldots, S_N = s_n | X_1 = x_1, \ldots, X_N = x_N)}{P(S_1 = s_1, \ldots, S_N = s_N)}$$
$$= \frac{P(X_1 = x_1, \ldots, X_N = x_N | S_1 = s_1, \ldots, S_N = s_n)}{P(X_1 = x_1, \ldots, X_N = x_N)}$$
$$= \frac{\prod_{t=1}^{N} P(X_t = x_t | S_t = s_t)}{P(X_1 = x_1, \ldots, X_N = x_N)} \propto \prod_{t=1}^{N} P(X_t = x_t | S_t = s_t)$$

Hidden Markov Model (HMM)

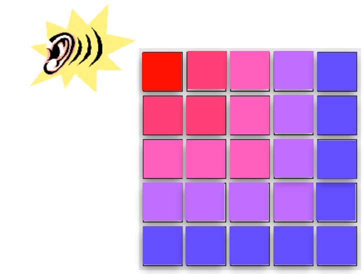


Eg: say observations were

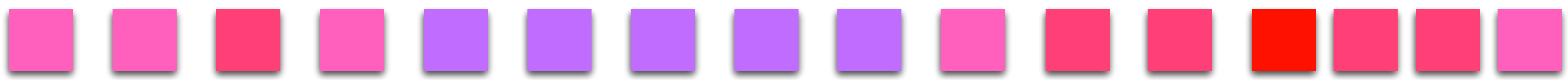


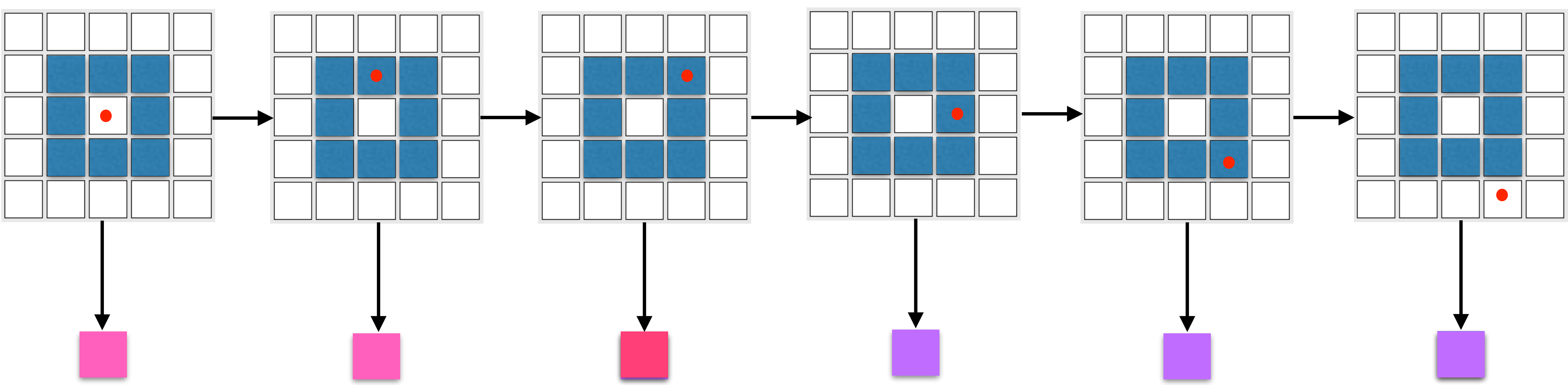


HIDDEN MARKOV MODEL (HMM)



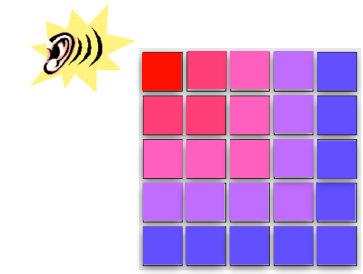
Eg: say observations were



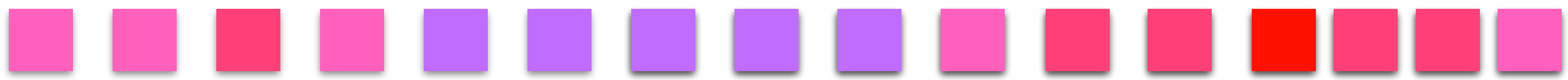


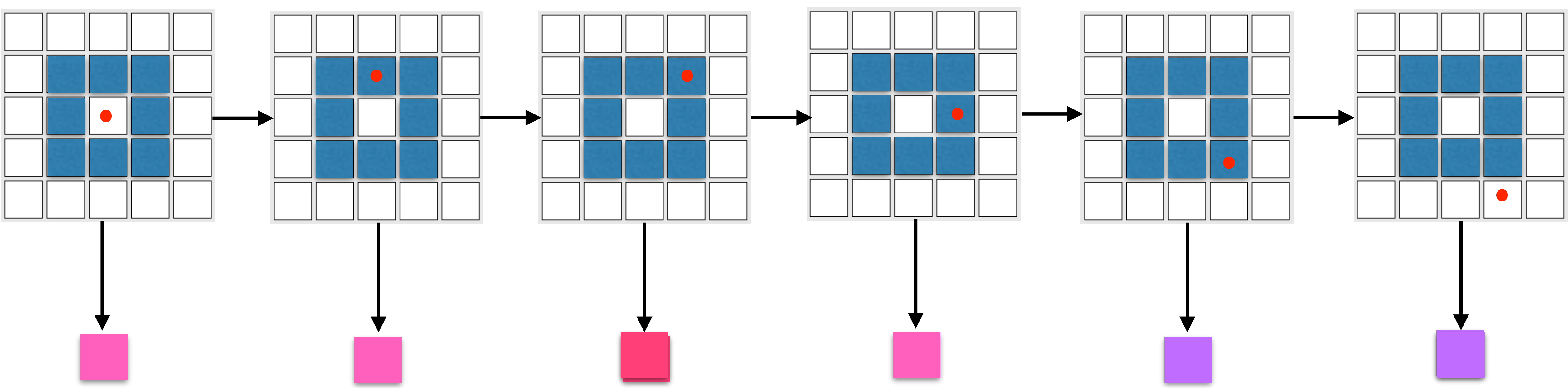
Importance weighting: weight samples

Hidden Markov Model (HMM)



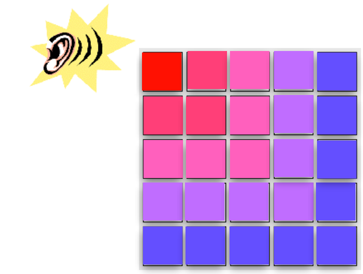
Eg: say observations were



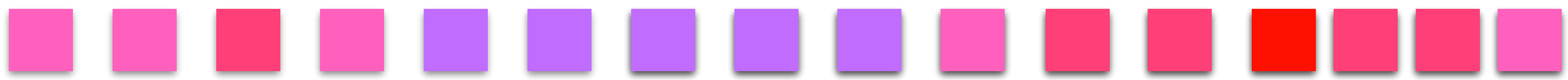


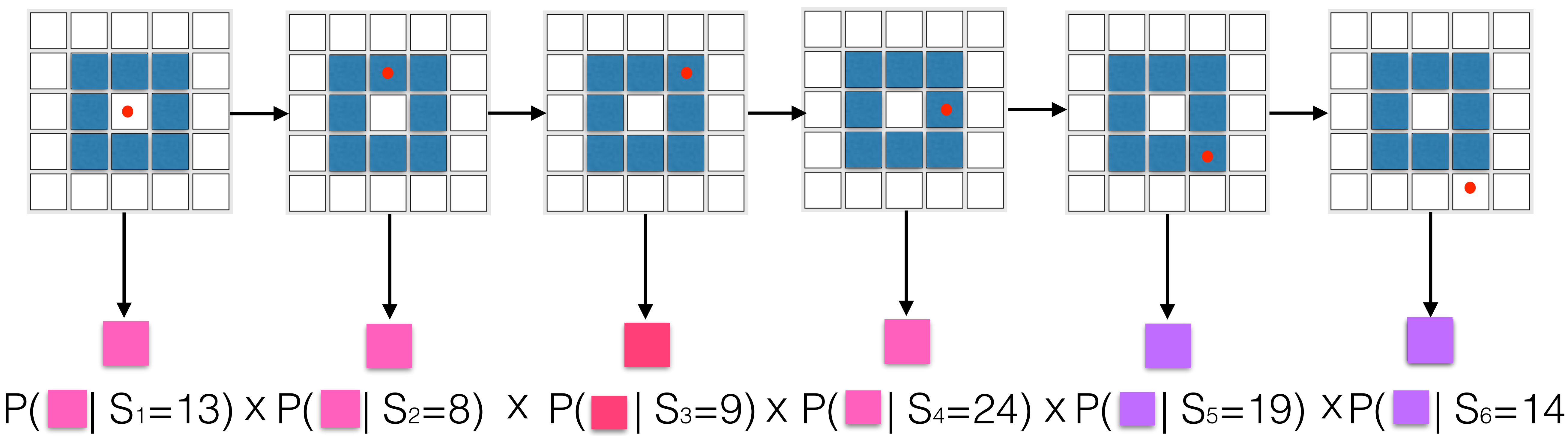
Importance weighting: weight samples

Hidden Markov Model (HMM)



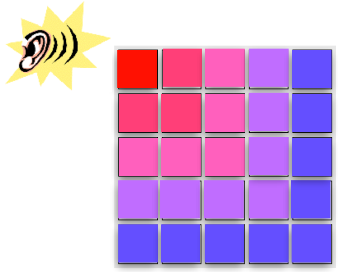
Eg: say observations were





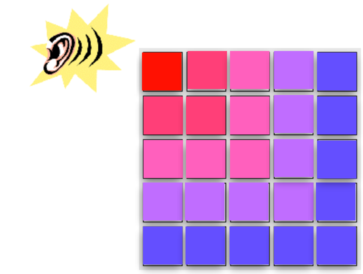
Importance weighting: weight samples

HMM Particle Filter

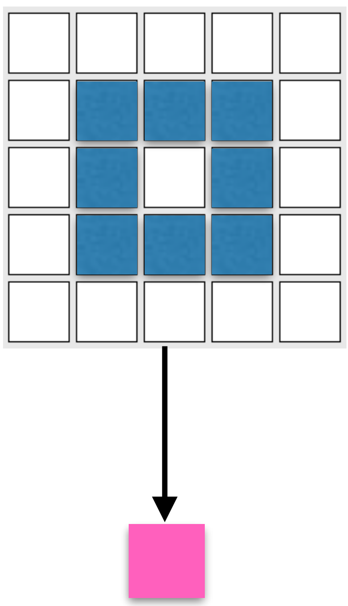


- Use multiple samples and track each ones weights.

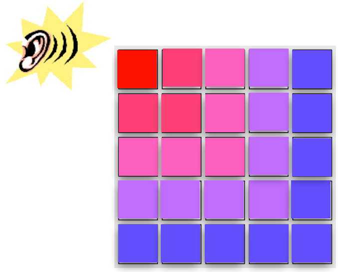
HMM Particle Filter



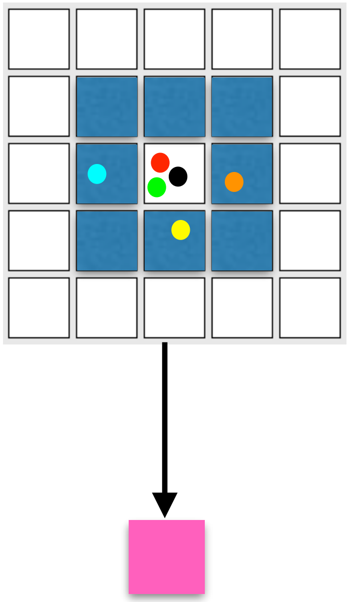
- Use multiple samples and track each ones weights.



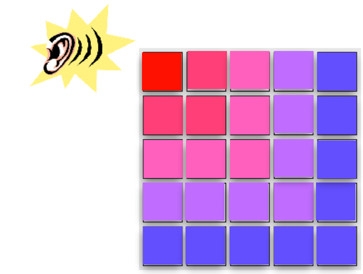
HMM Particle Filter



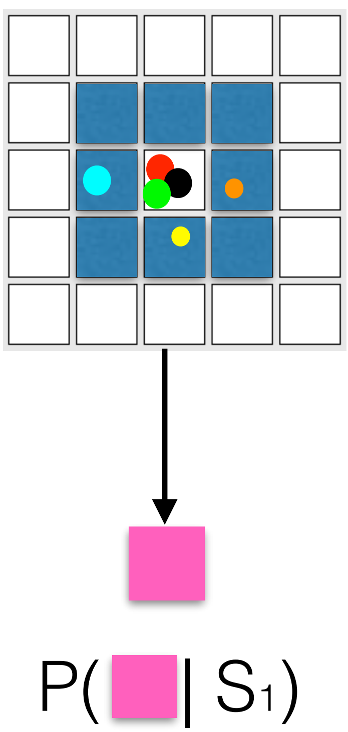
- Use multiple samples and track each ones weights.



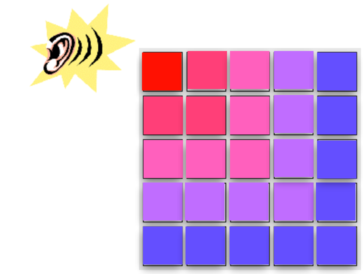
HMM Particle Filter



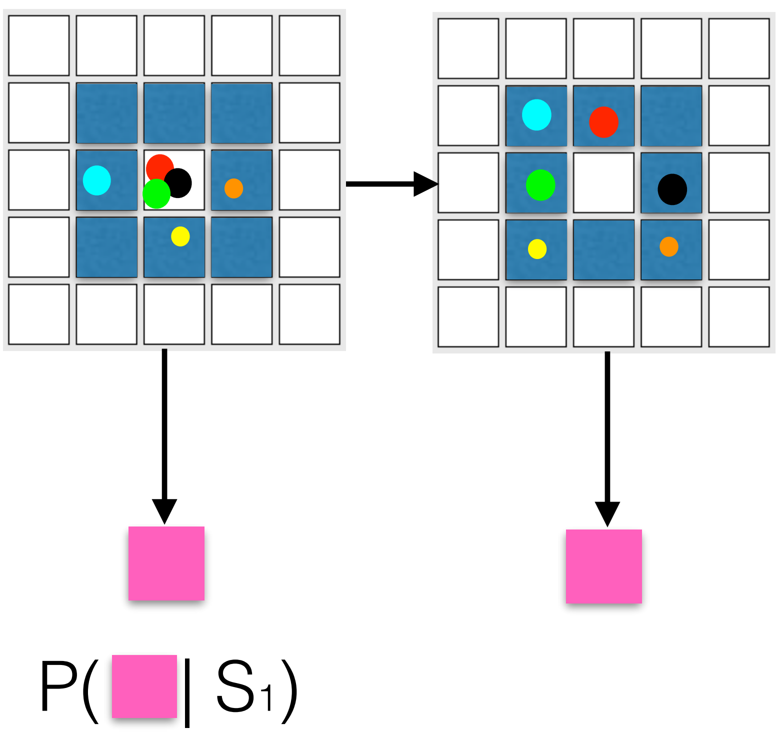
- Use multiple samples and track each ones weights.



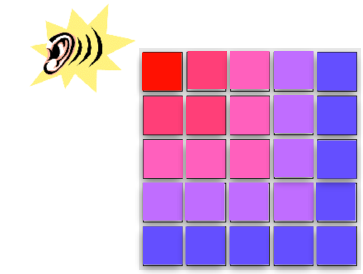
HMM Particle Filter



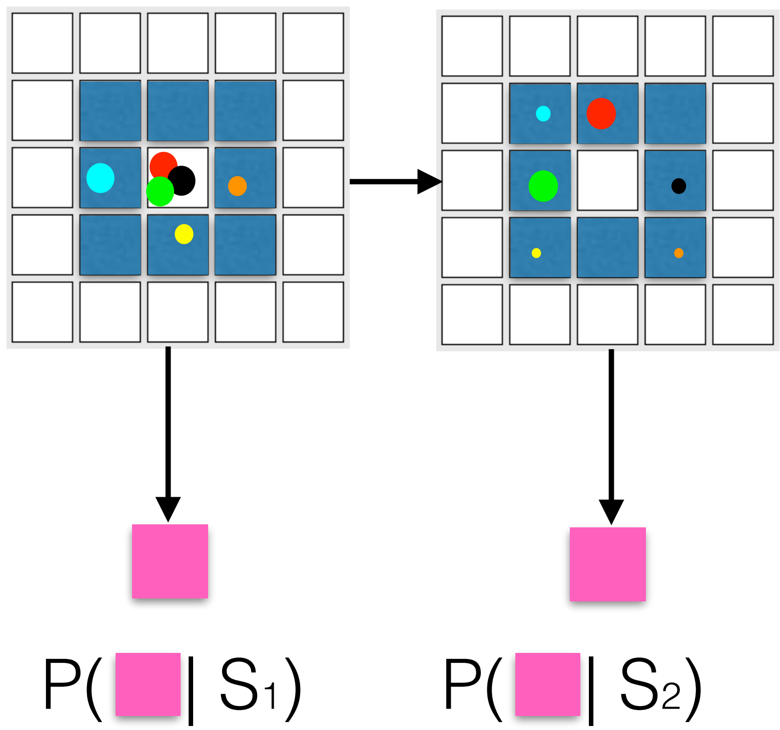
- Use multiple samples and track each ones weights.



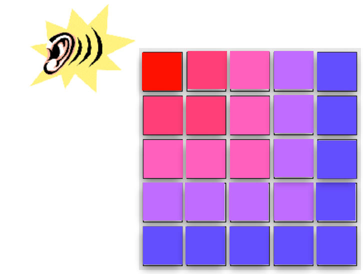
HMM Particle Filter



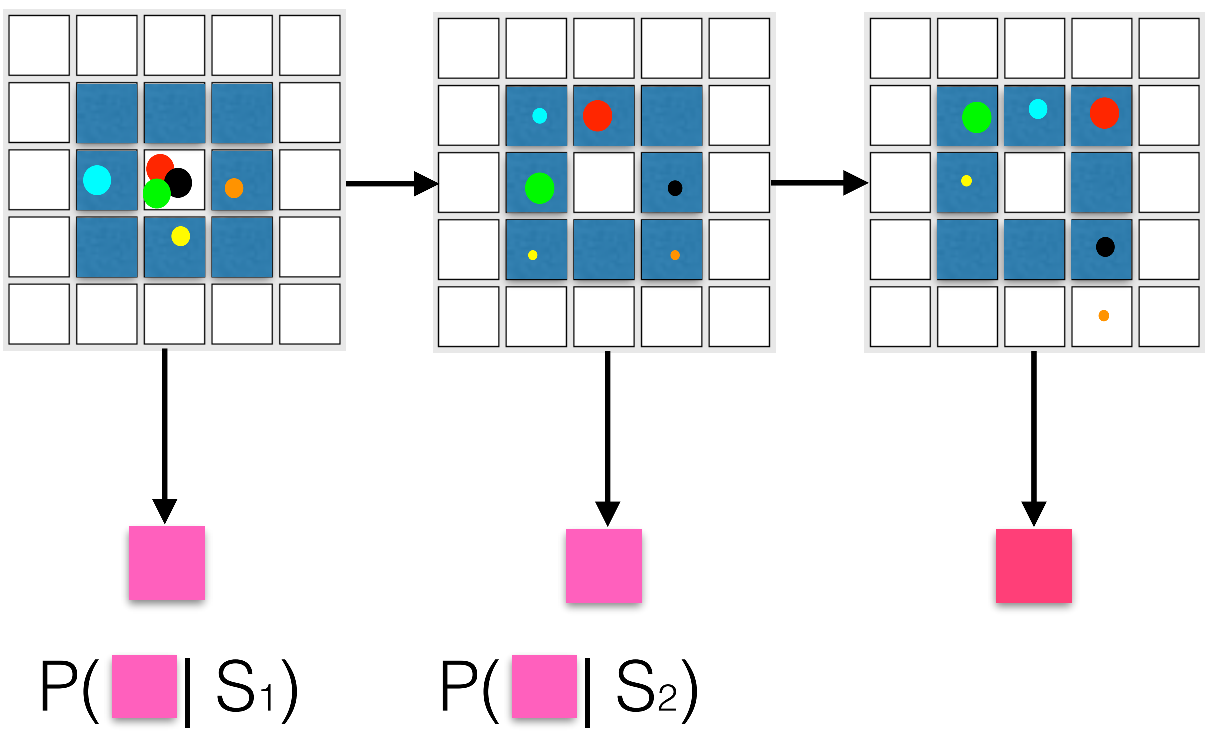
- Use multiple samples and track each ones weights.



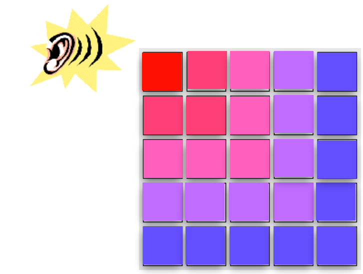
HMM Particle Filter



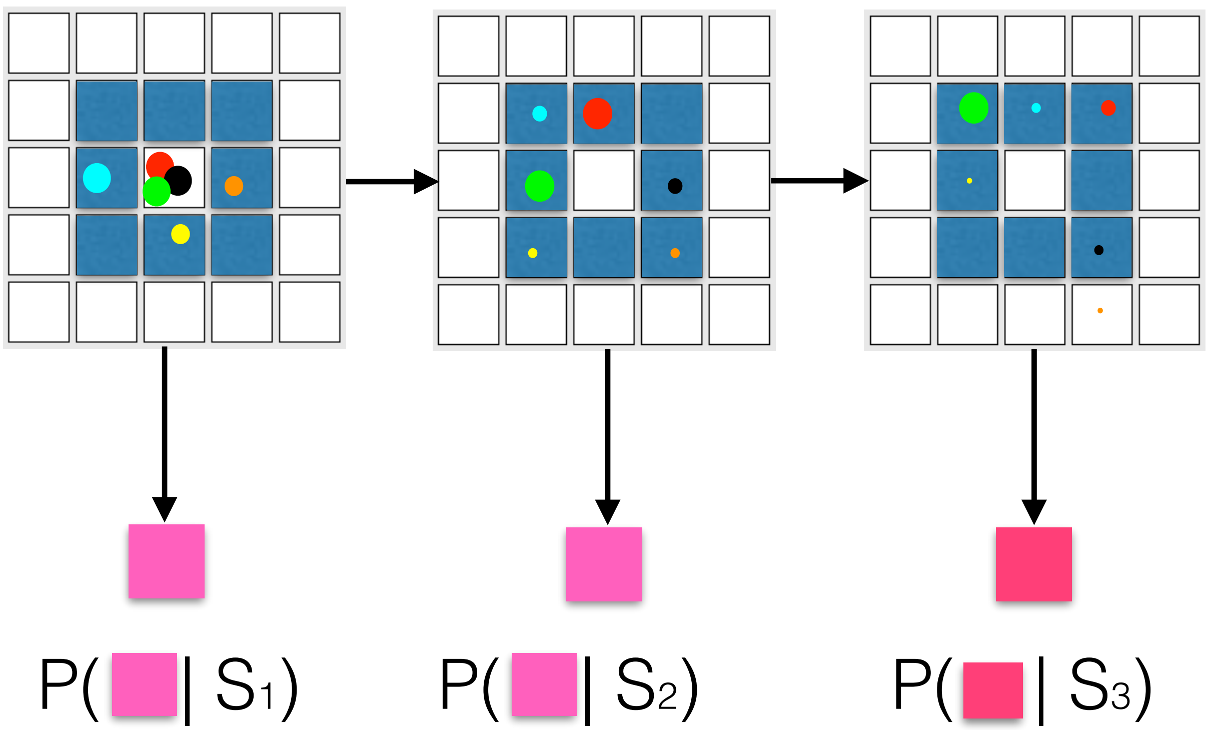
- Use multiple samples and track each ones weights.



HMM Particle Filter

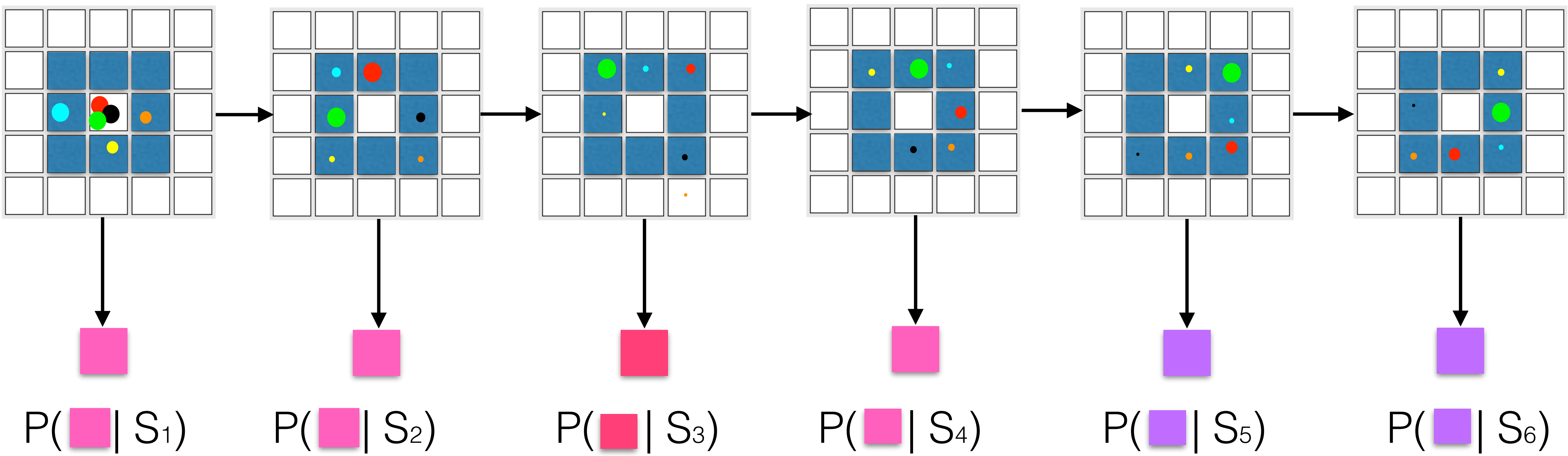


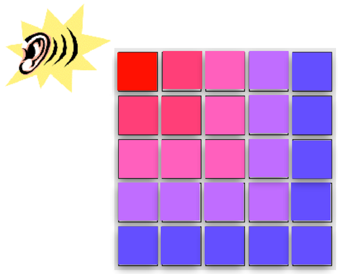
- Use multiple samples and track each ones weights.



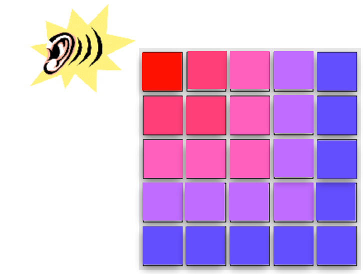
HMM Particle Filter

- Use multiple samples and track each ones weights.

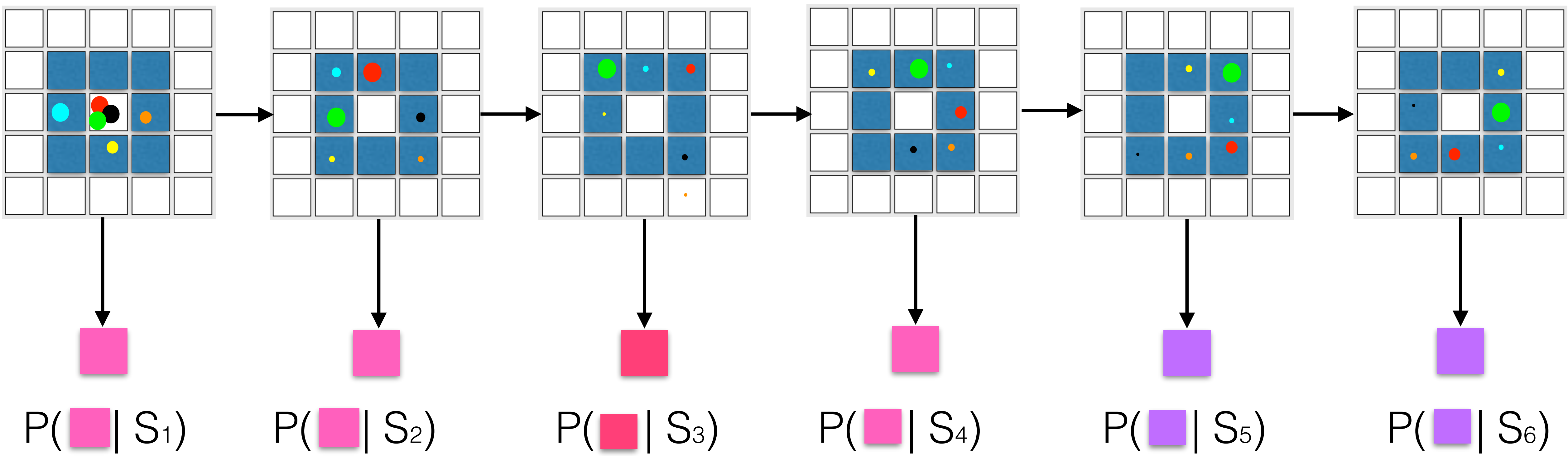




HMM Particle Filter

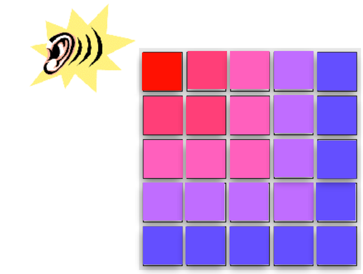


- Use multiple samples and track each ones weights.

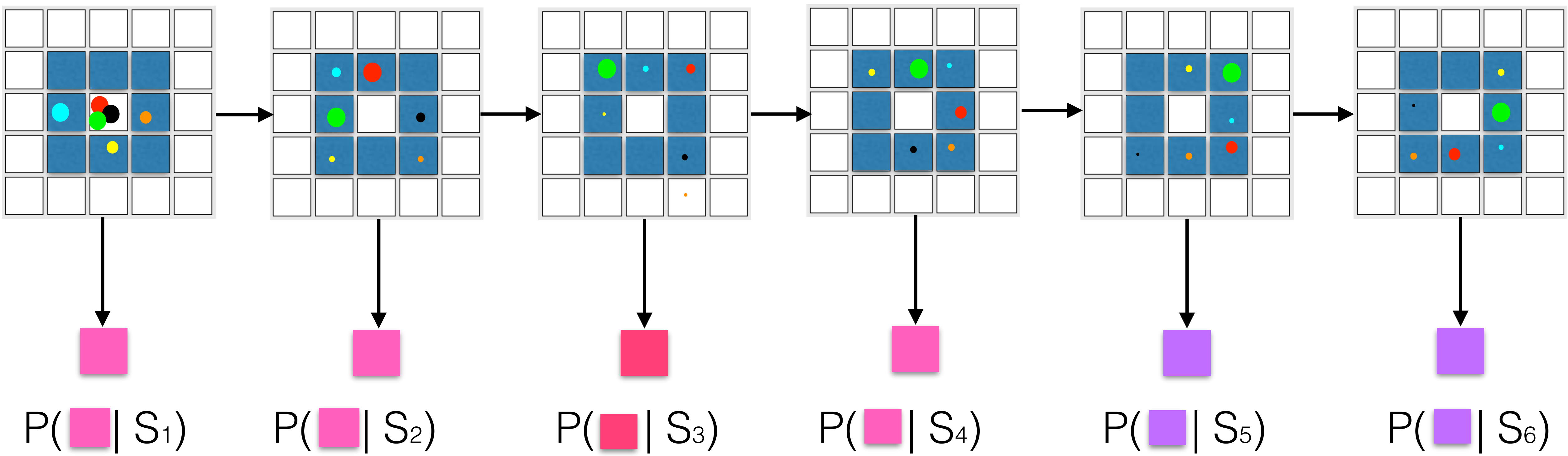


- This is same as 6 separate samples

HMM PARTICLE FILTER

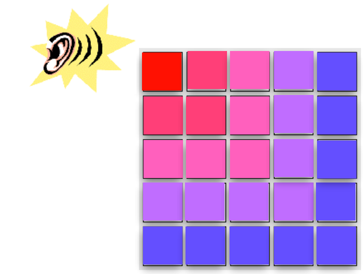


- Use multiple samples and track each ones weights.

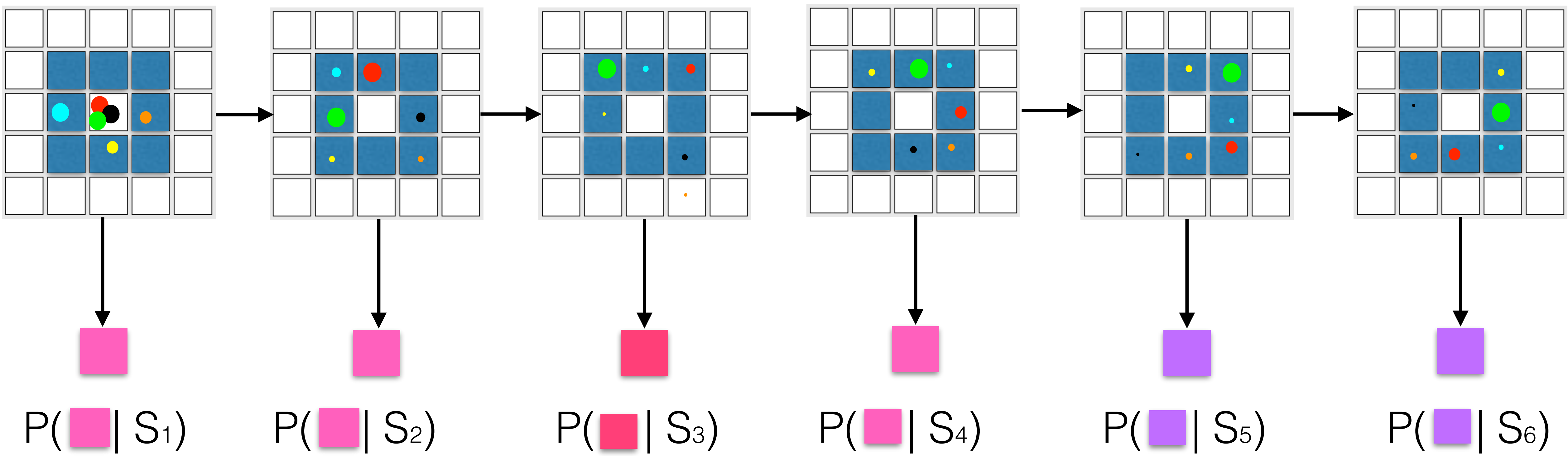


- This is same as 6 separate samples
- Instead of tracking each sample's weight, resample according to weights

HMM Particle Filter

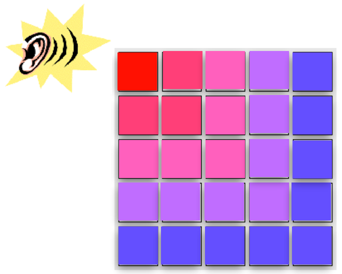


- Use multiple samples and track each ones weights.



- This is same as 6 separate samples
- Instead of tracking each sample's weight, resample according to weights
- Problem: Too many samples have negligible weight!

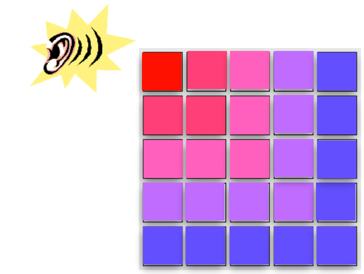
HMM Particle Filter

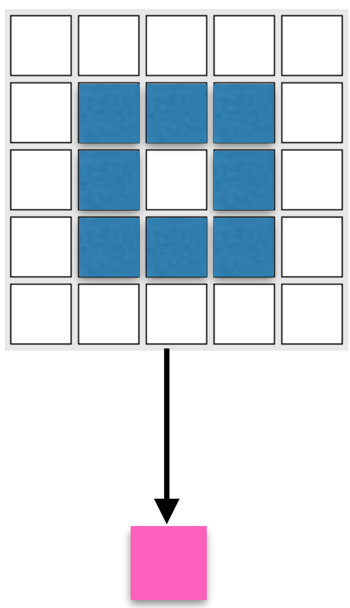


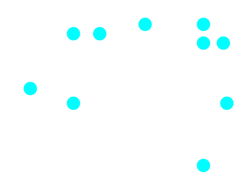
Instead of tracking each one, resample!

HMM Particle Filter

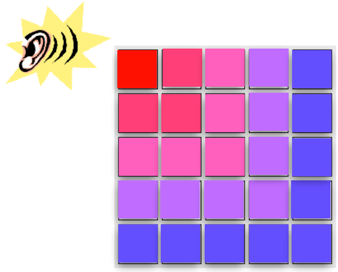
Instead of tracking each one, resample!



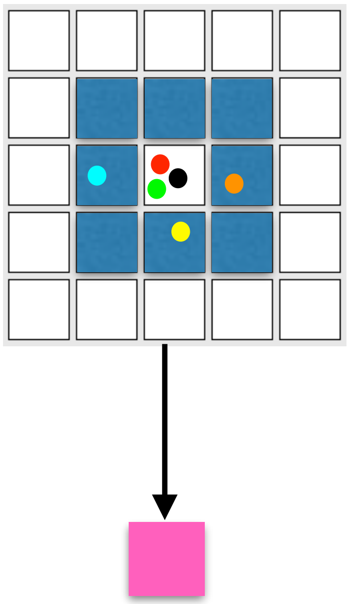




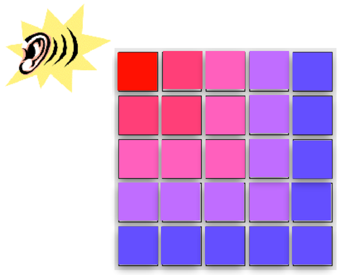
HMM Particle Filter



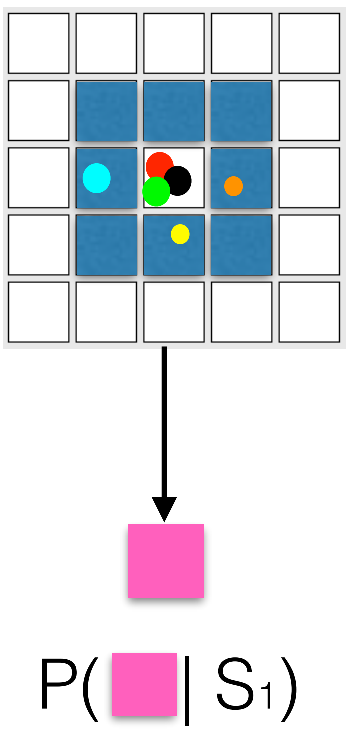
Instead of tracking each one, resample!



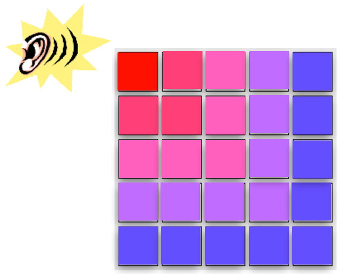
HMM Particle Filter



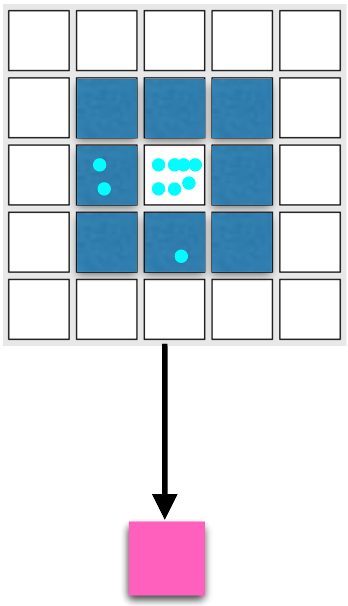
Instead of tracking each one, resample!



HMM Particle Filter

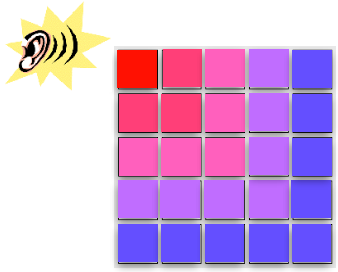


Instead of tracking each one, resample!

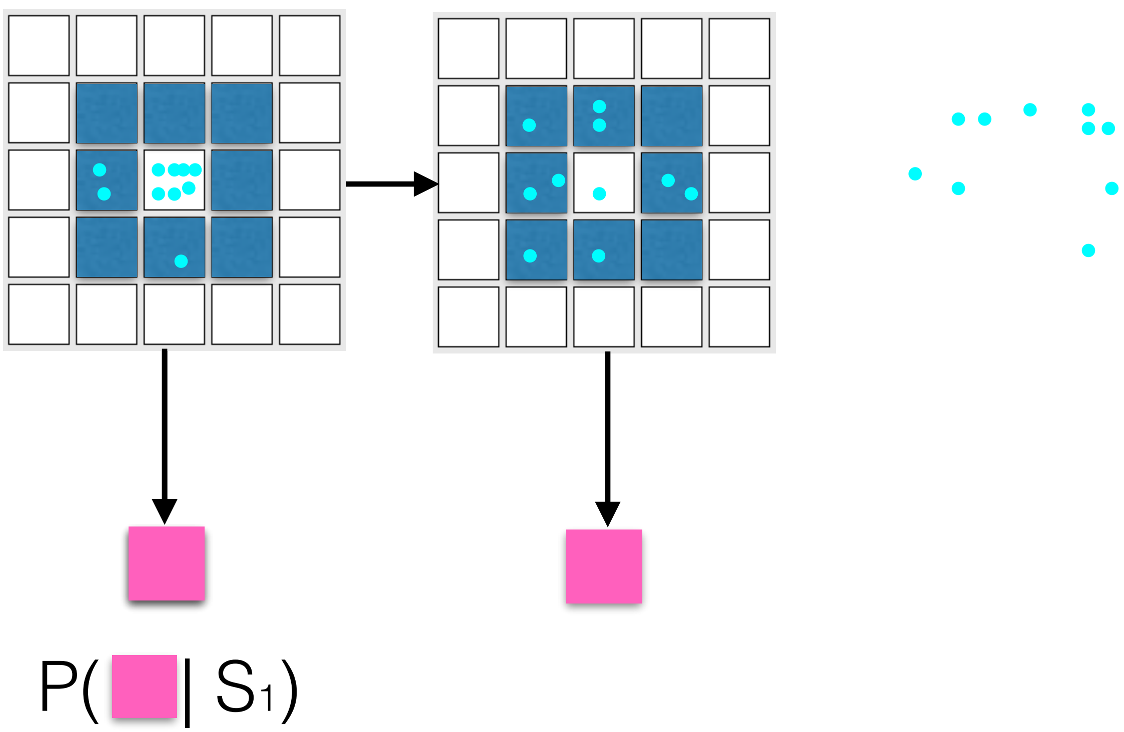


$P(\quad \mid S_1)$

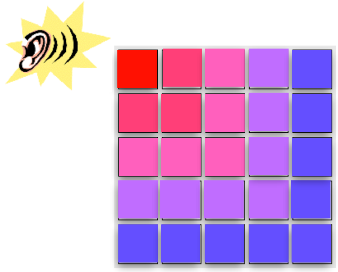
HMM Particle Filter



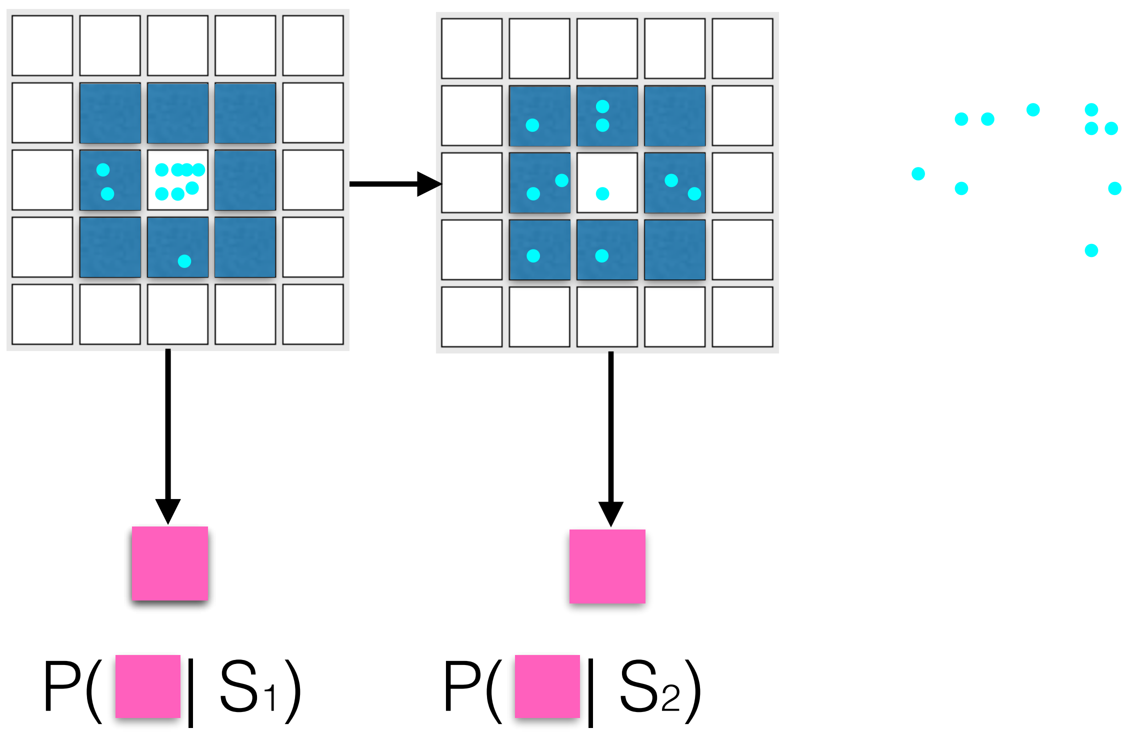
Instead of tracking each one, resample!



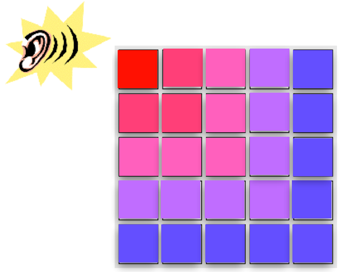
HMM Particle Filter



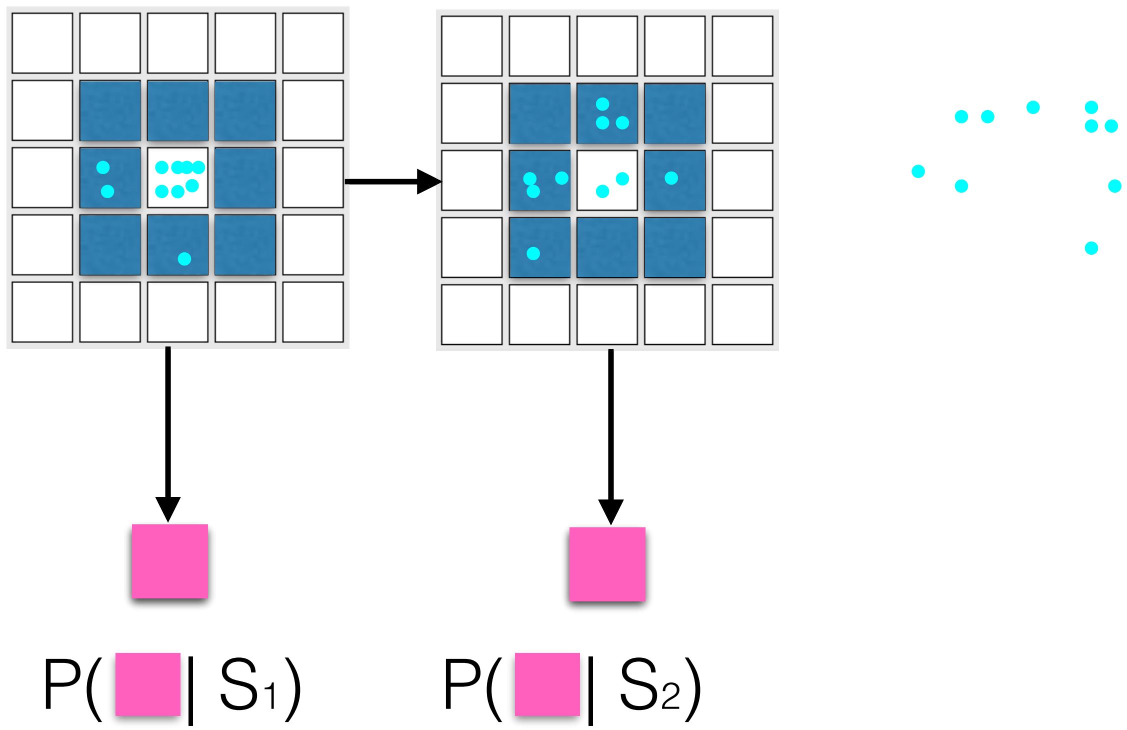
Instead of tracking each one, resample!



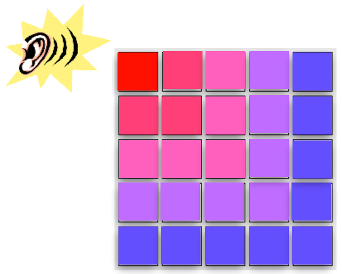
HMM Particle Filter



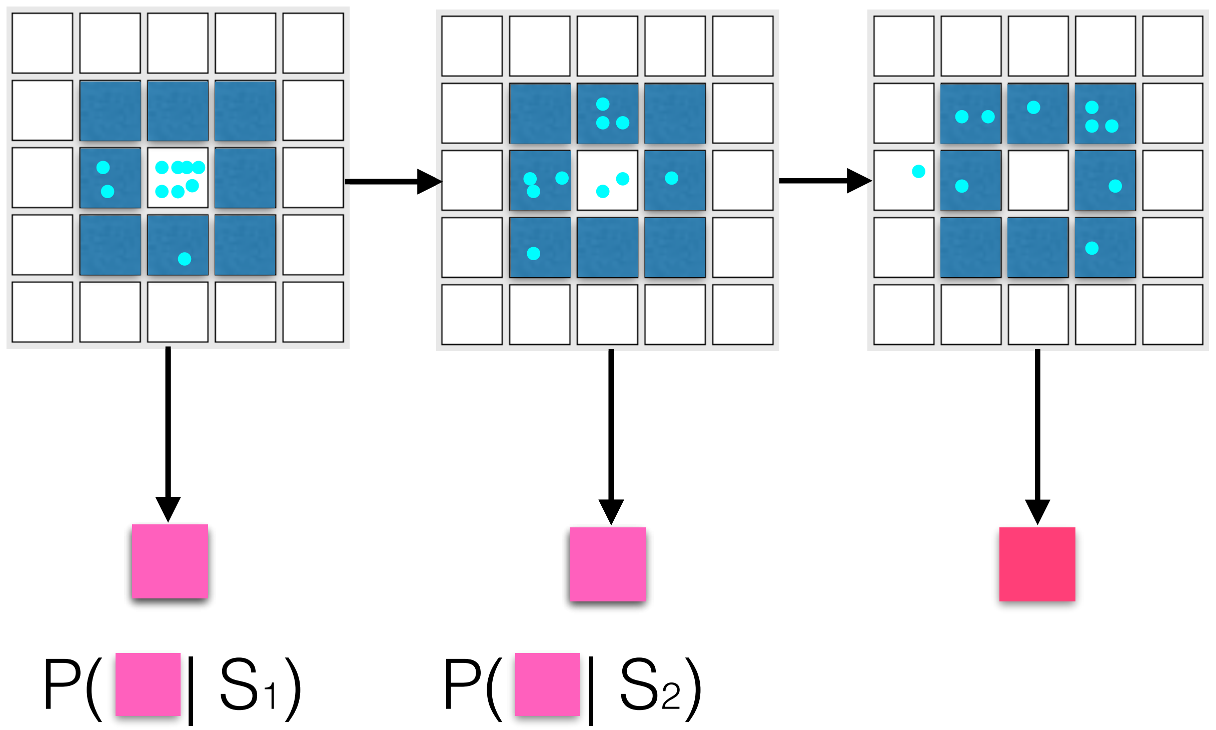
Instead of tracking each one, resample!



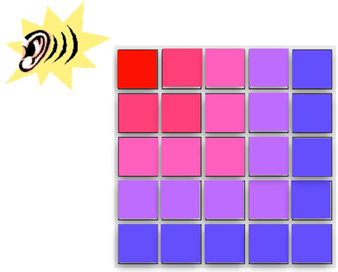
HMM Particle Filter



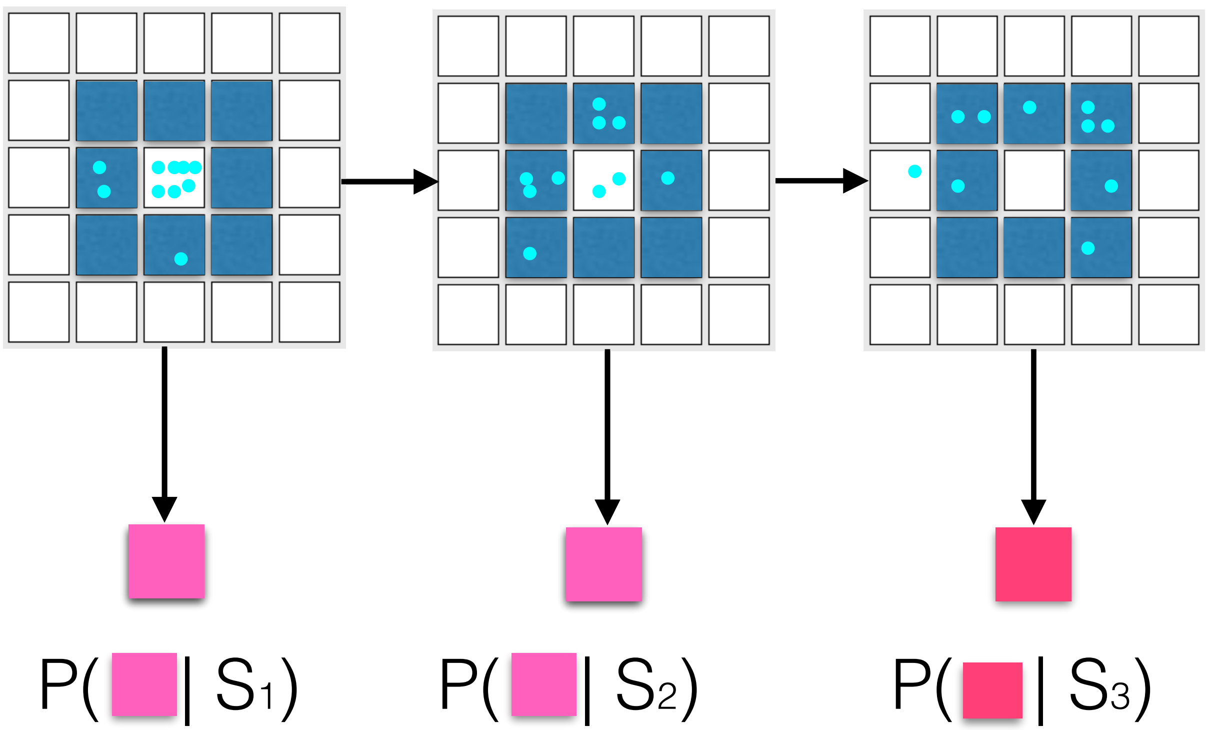
Instead of tracking each one, resample!



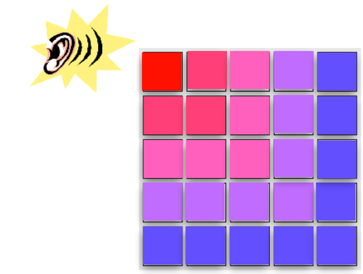
HMM Particle Filter



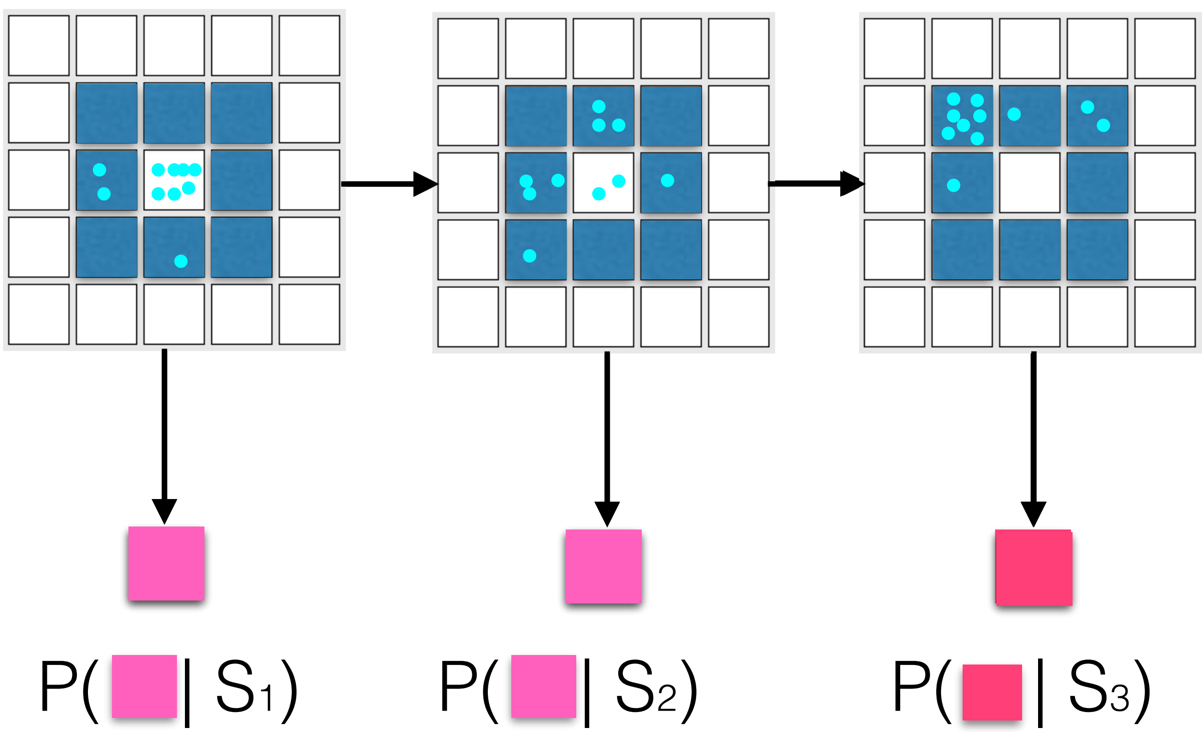
Instead of tracking each one, resample!



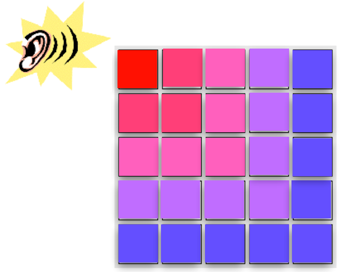
HMM Particle Filter



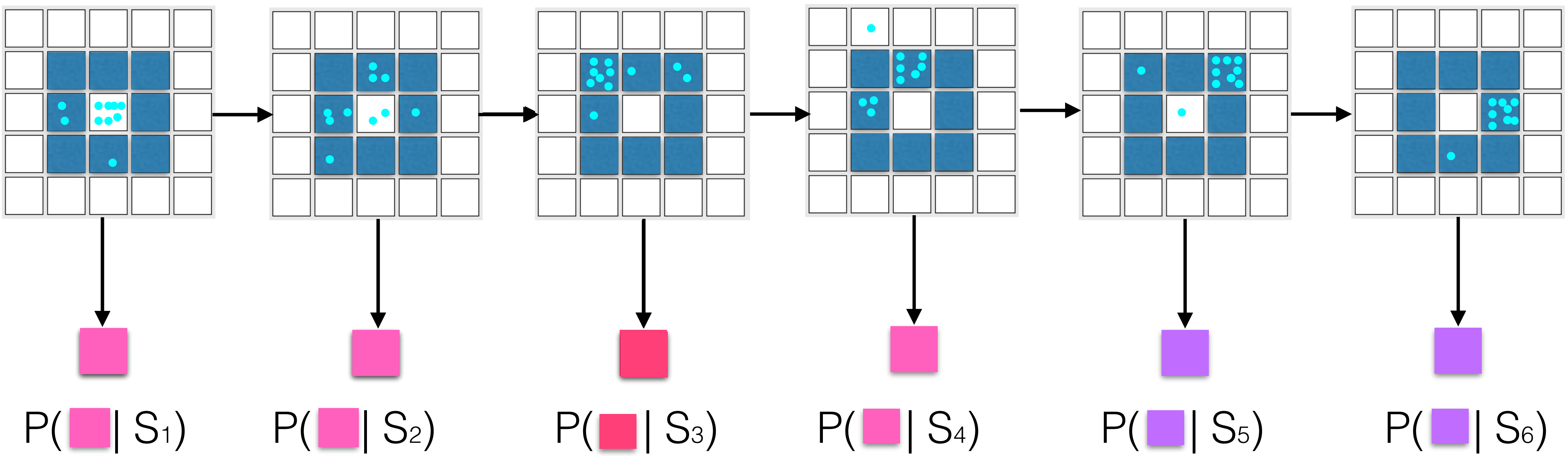
Instead of tracking each one, resample!



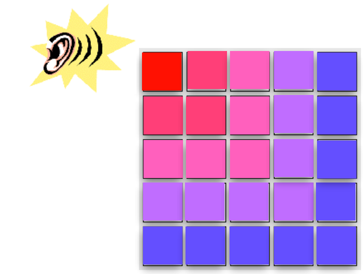
HMM Particle Filter



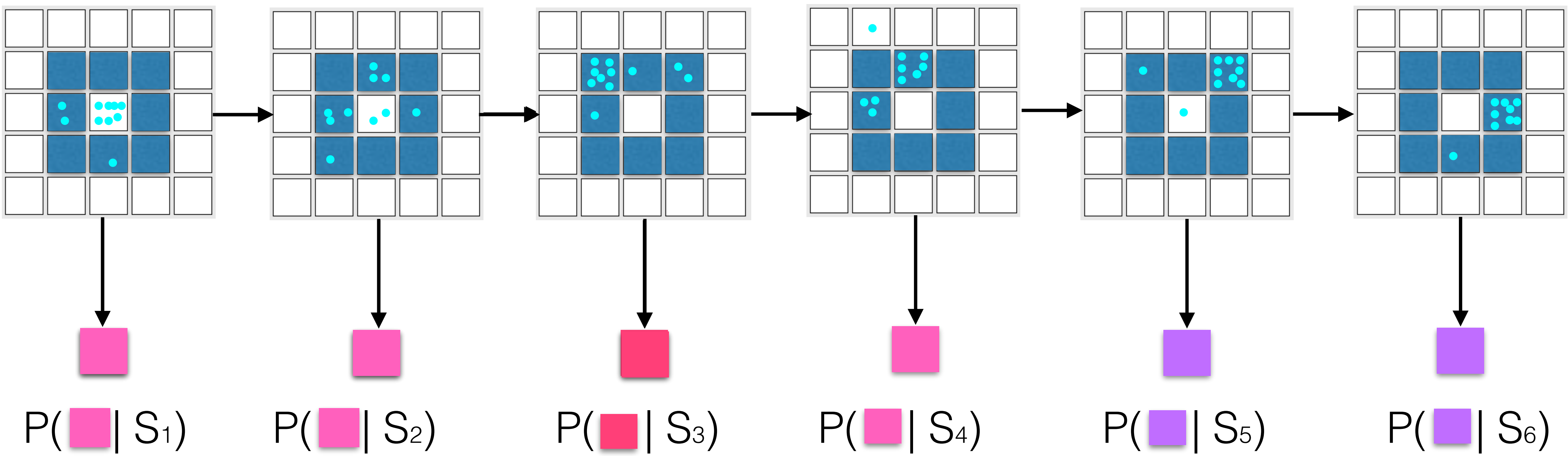
Instead of tracking each one, resample!



HMM Particle Filter

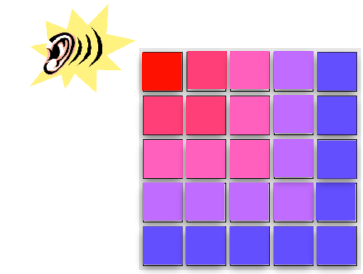


Instead of tracking each one, resample!

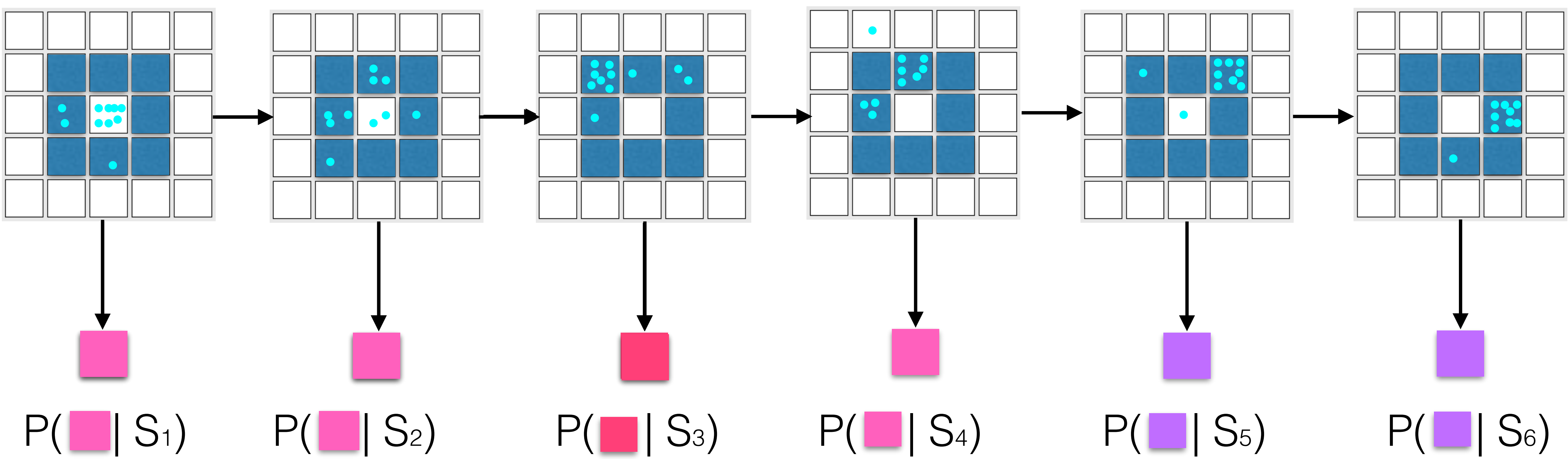


- On every round, transfer particles from previous states according to transition probability

HMM Particle Filter

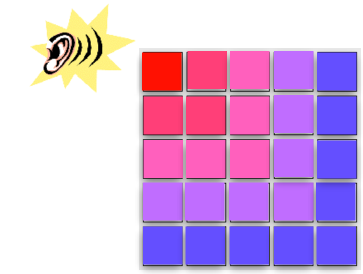


Instead of tracking each one, resample!

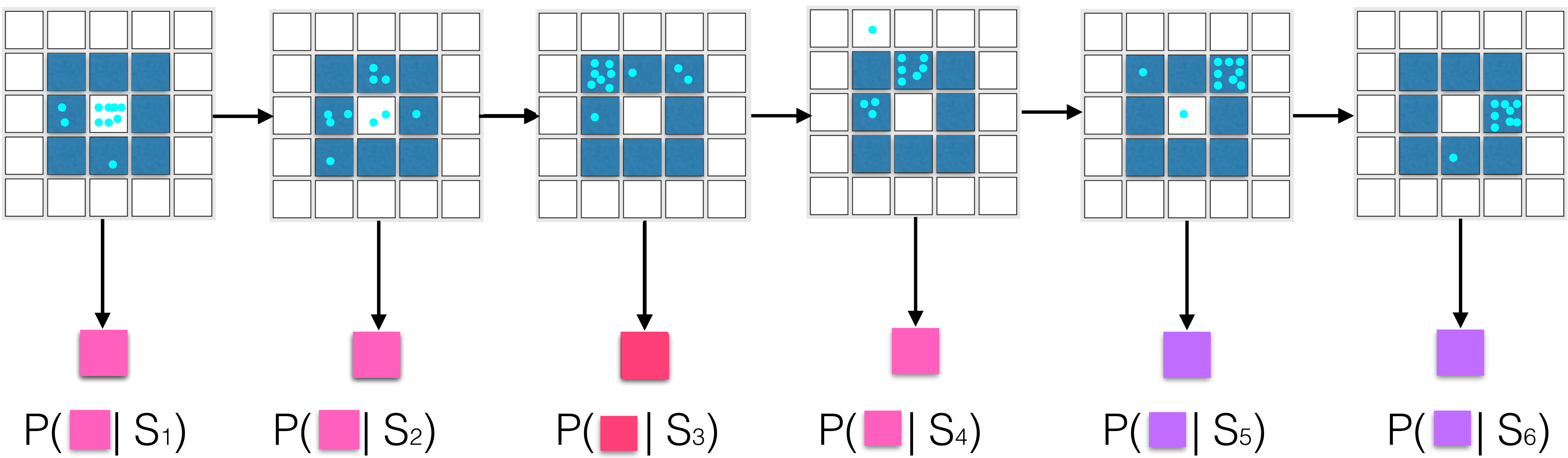


- On every round, transfer particles from previous states according to transition probability
- Resample particles according to P(observation|state)

HMM Particle Filter



Instead of tracking each one, resample!



- On every round, transfer particles from previous states according to transition probability
- Resample particles according to P(observation|state)
- Use new particles to proceed

Particle Filtering

Particle Filtering

- Without resampling, we carry many particles with very small probabilities

Particle Filtering

- Without resampling, we carry many particles with very small probabilities
  - too many samples needed for a good estimate

Particle Filtering

- Without resampling, we carry many particles with very small probabilities
  - too many samples needed for a good estimate
- By resampling, we got rid of samples with very small probabilities

Particle Filtering

- Without resampling, we carry many particles with very small probabilities
    - too many samples needed for a good estimate
- By resampling, we got rid of samples with very small probabilities
    - Hence fewer samples suffice

HMM Particle Filter

- Inference time only depends on number of samples
- Of course more the samples the better accuracy
- Often we don't need too many samples. Why ?

Gibbs Sampling

Gibbs Sampling

- Repeat n times for, n samples,

Gibbs Sampling

- Repeat n times for, n samples,
  - Start with arbitrary value for (latent) variables

Gibbs Sampling

- Repeat n times for, n samples,
  - Start with arbitrary value for (latent) variables
  - Replace each variable by new sample from P(Variable| all other variables)

Gibbs Sampling

- Repeat n times for, n samples,
  - Start with arbitrary value for (latent) variables
  - Replace each variable by new sample from P(Variable| all other variables)
  - Go over all (latent) variables multiple times

Gibbs Sampling

- Repeat n times for, n samples,
    - Start with arbitrary value for (latent) variables
    - Replace each variable by new sample from P(Variable| all other variables)
    - Go over all (latent) variables multiple times
    - Return final sample of the N variables

Gibbs Sampling

$$
\begin{aligned}
&P(S_t = k | S_1 = s_1, \ldots, S_{t-1} = s_{t-1}, S_{t+1} = s_{t+1}, \ldots, S_N = s_N, X_1 = x_1, \ldots, X_N = x_N) \\
&\propto P(S_1 = s_1, \ldots, S_{t-1} = s_{t-1}, S_t = k, S_{t+1} = s_{t+1}, \ldots, S_N, X_1 = x_1, \ldots, X_N = x_N) \\
&\propto \prod_{i=1}^{t-1} P(S_i = s_i | S_{i-1} = s_{i-1}) P(X_i = x_i | S_i = s_i) \times P(S_t = k | S_{t-1} = s_{t-1}) P(X_t = x_t | S_t = k) \\
&\qquad\qquad \times \prod_{j=t+1}^{N} P(S_j = s_j | S_{j-1} = s_{j-1}) P(X_j = x_j | S_j = s_j) \\
&\propto P(S_t = k | S_{t-1} = s_{t-1}) P(X_t = x_t | S_t = k) P(S_{t+1} = s_{t+1} | S_t = k) \\
&= \frac{P(S_t = k | S_{t-1} = s_{t-1}) P(X_t = x_t | S_t = k) P(S_{t+1} = s_{t+1} | S_t = k)}{\sum_{j=1}^{K} P(S_t = j | S_{t-1} = s_{t-1}) P(X_t = x_t | S_t = j) P(S_{t+1} = s_{t+1} | S_t = j)}
\end{aligned}
$$