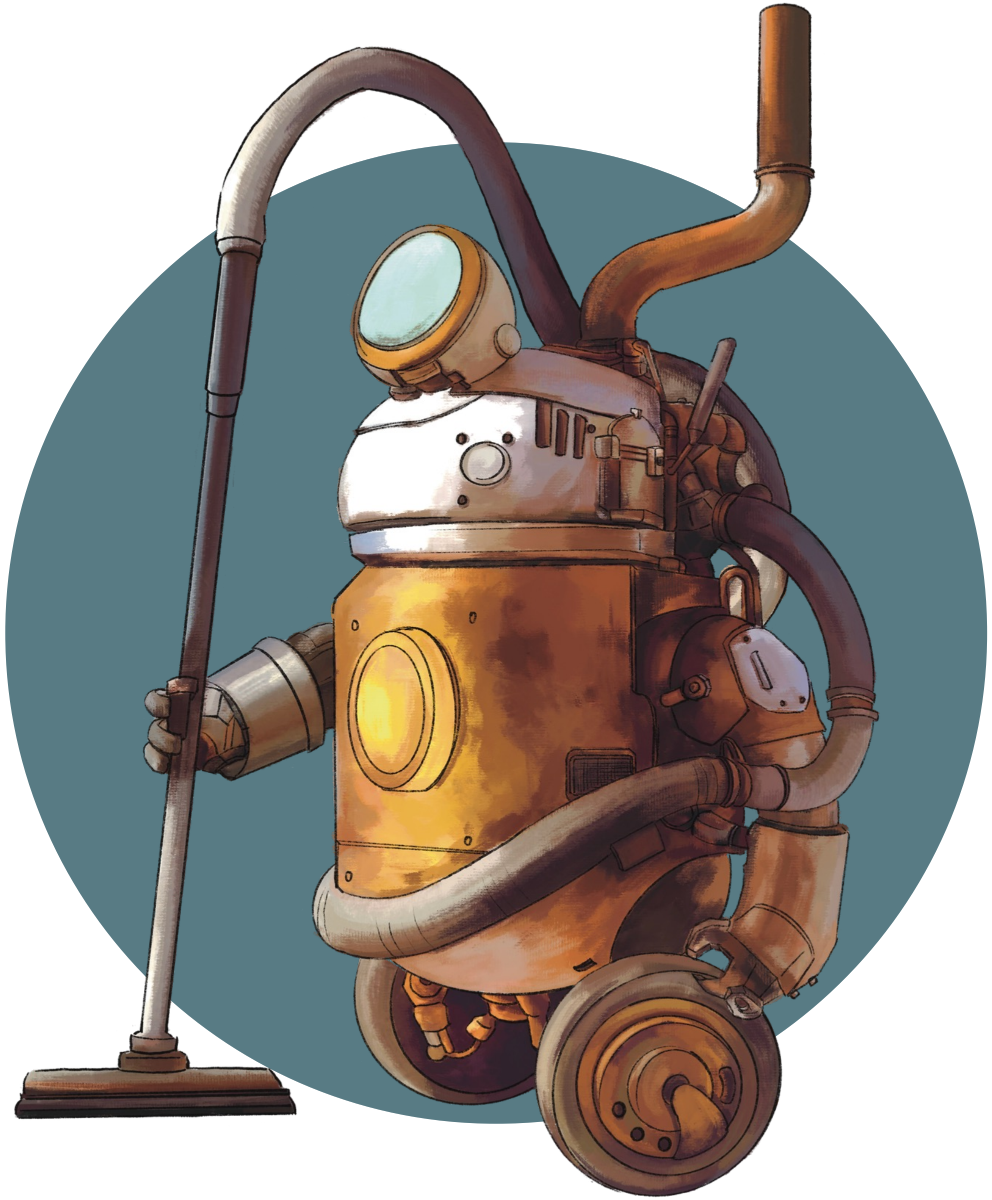
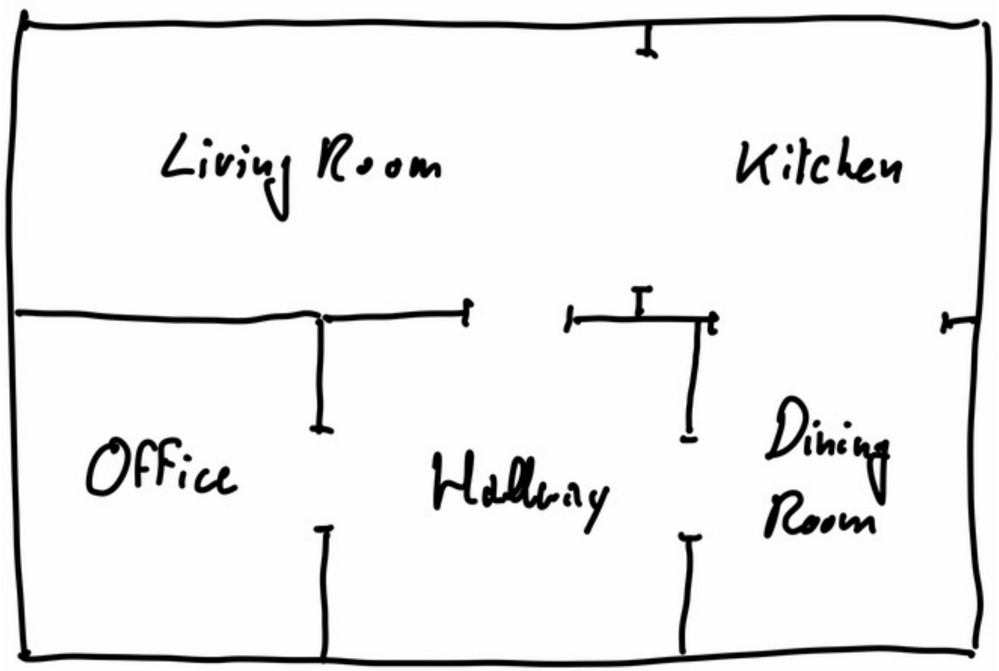
Lecture 9:
Sensing & perception



Frank Dellaert • Seth Hutchinson

Five rooms produce just three possible light readings.

SOUTH SIDE: usually brighter



bright	Living room · Kitchen
medium	Office · Dining room
dark	Hallway

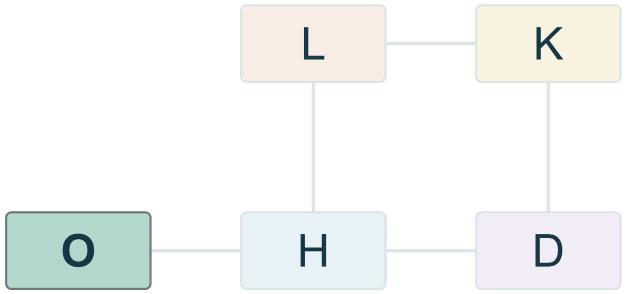
Typical **light levels, sometimes we get different readings.**

A sensor model predicts a reading given the room—not the other way around.

$$P(Z_t \mid X_t)$$

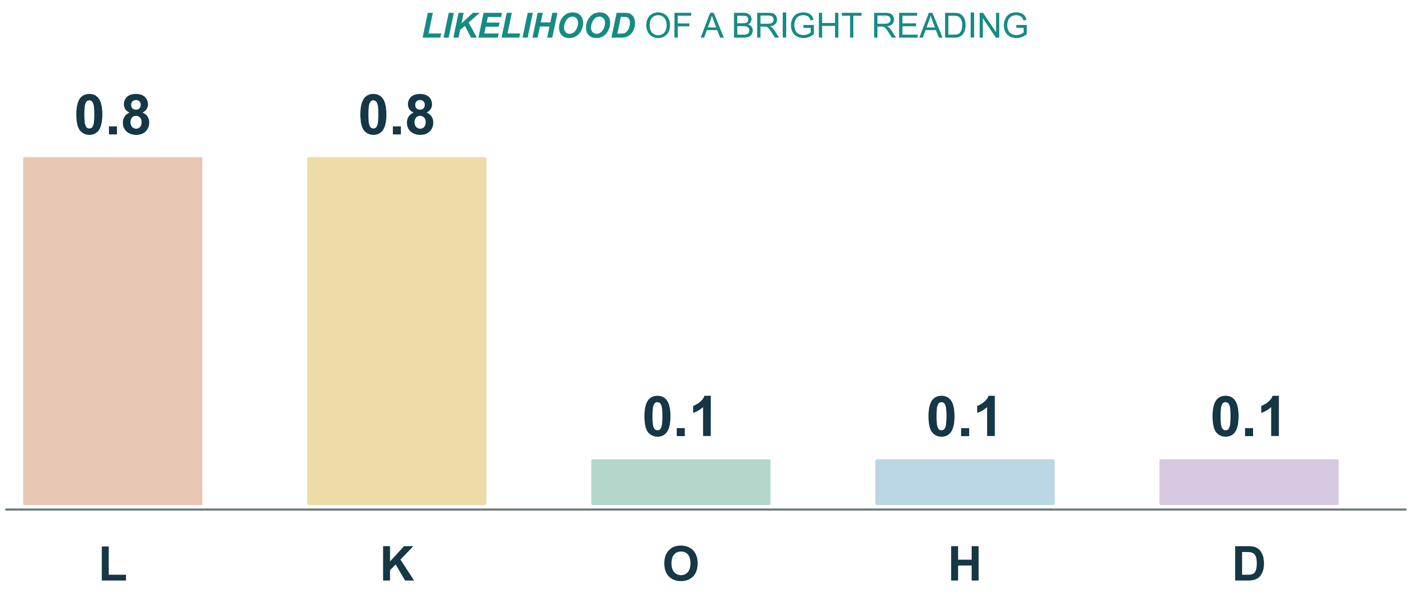
Current room	dark	medium	bright
Living room	0.1	0.1	**0.8**
Kitchen	0.1	0.1	**0.8**
Office	0.2	**0.7**	0.1
Hallway	**0.8**	0.1	0.1
Dining room	0.1	**0.8**	0.1

CURRENT ROOM: OFFICE

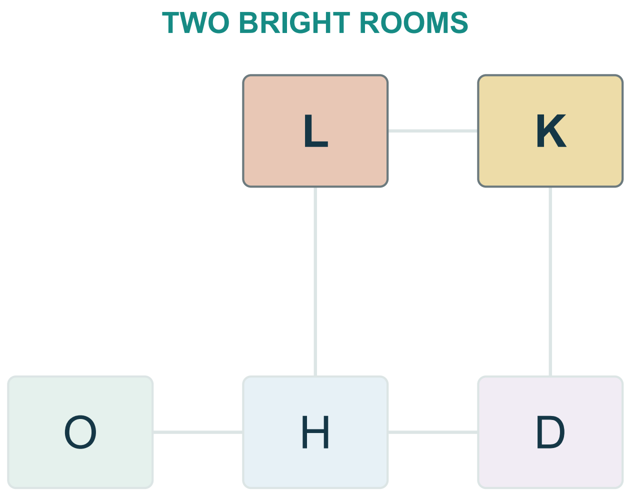


Choose one room; read its distribution across the three possible measurements.

The same reading can come from more than one room.



$$P(\text{bright} \mid L) = P(\text{bright} \mid K) = 0.8$$



Several rooms are *likely* given a single reading.

We can predict where we might be.
Now we want evidence of where we are.

Current room	dark	medium	bright
Living room	0.1	0.1	**0.8**
Kitchen	0.1	0.1	**0.8**
Office	0.2	**0.7**	0.1
Hallway	**0.8**	0.1	0.1
Dining room	0.1	**0.8**	0.1

RIGHT

Office →

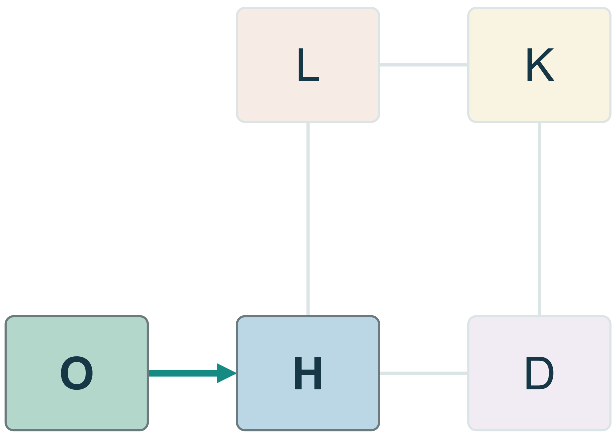
Our prediction

0	0	0.2	0.8	0
L	K	O	H	D

$$b_1 = b_0 M_R$$

dark

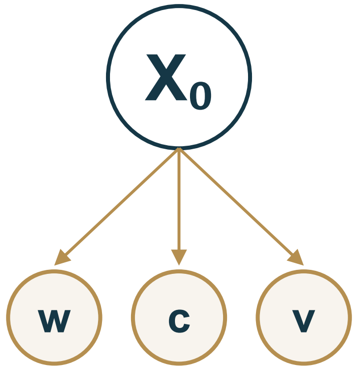
MAP REMINDER



Exercise: What is the posterior probability?

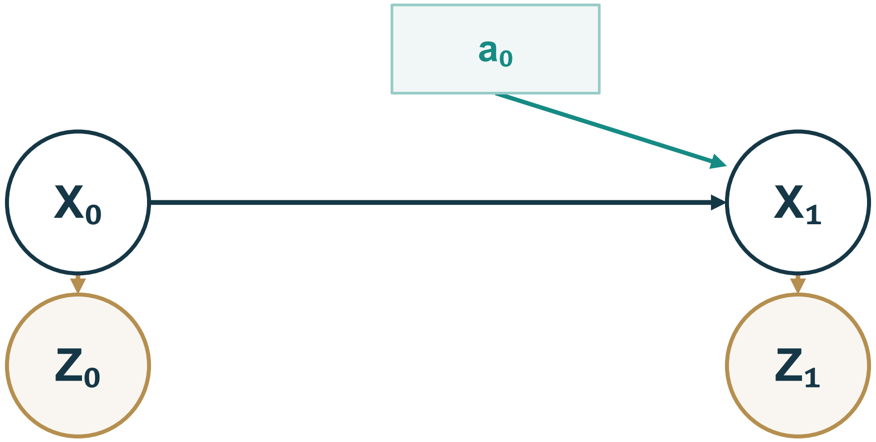
Unlike separate pieces of trash, one robot links measurements across time.

TRASH SORTING: separate items



No temporal link between the two items.

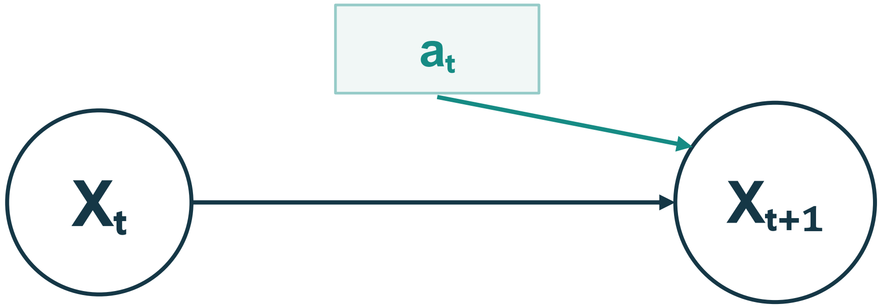
VACUUMING: one moving robot



A motion model links consecutive states.

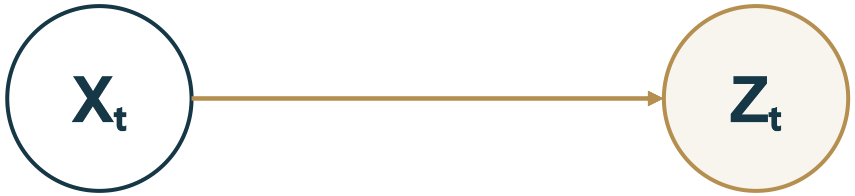
A node's parents tell us which conditional probability to use.

MOTION MODEL



$$P(X_{t+1} \mid X_t, a_t)$$

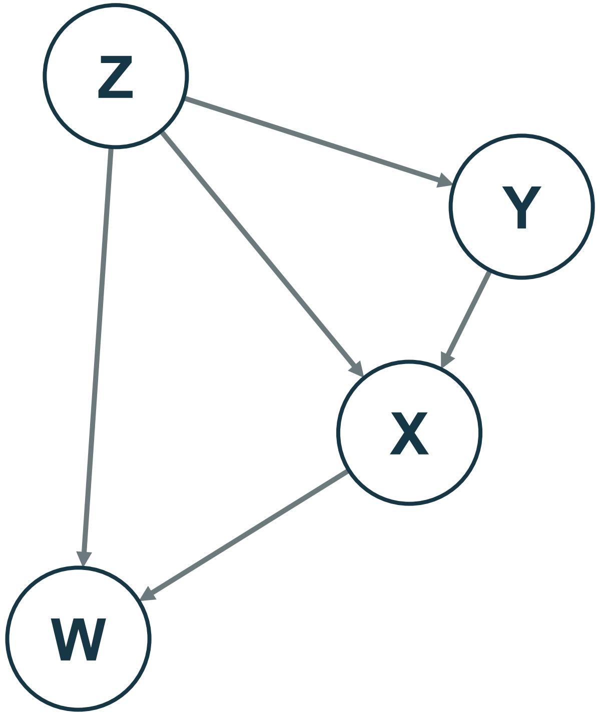
SENSOR MODEL



$$P(Z_t \mid X_t)$$

Know quantities are indicated with a square node

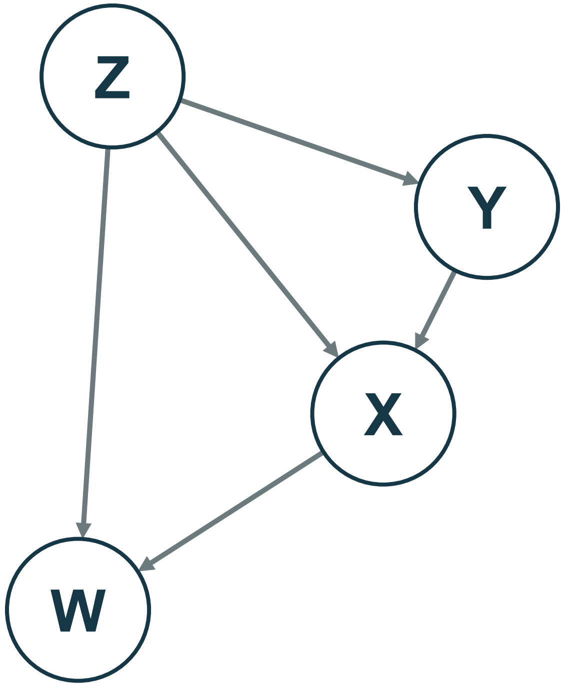
A *Bayes net* is a directed graph with no directed cycles.



Z	No parents
Y	Parent: Z
X	Parents: Y, Z
W	Parents: X, Z

Directed acyclic graph = DAG. Each node stores a local probability distribution.

Multiply one conditional per node to obtain the joint distribution.



$P(Z)$

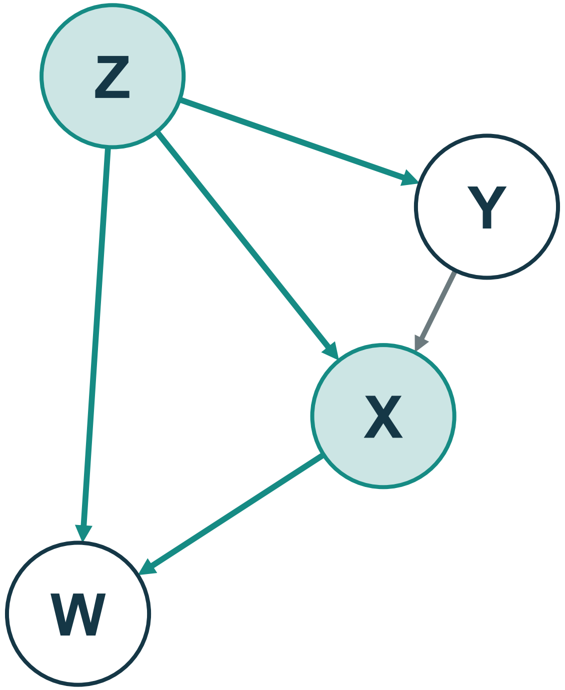
$P(Y \mid Z)$

$P(X \mid Y, Z)$

$P(W \mid X, Z)$

$$P(W, X, Y, Z) = P(W \mid X, Z)P(X \mid Y, Z)P(Y \mid Z)P(Z)$$

Given all of W's parents, Y adds no information about W.



CONDITION ON BOTH PARENTS

X

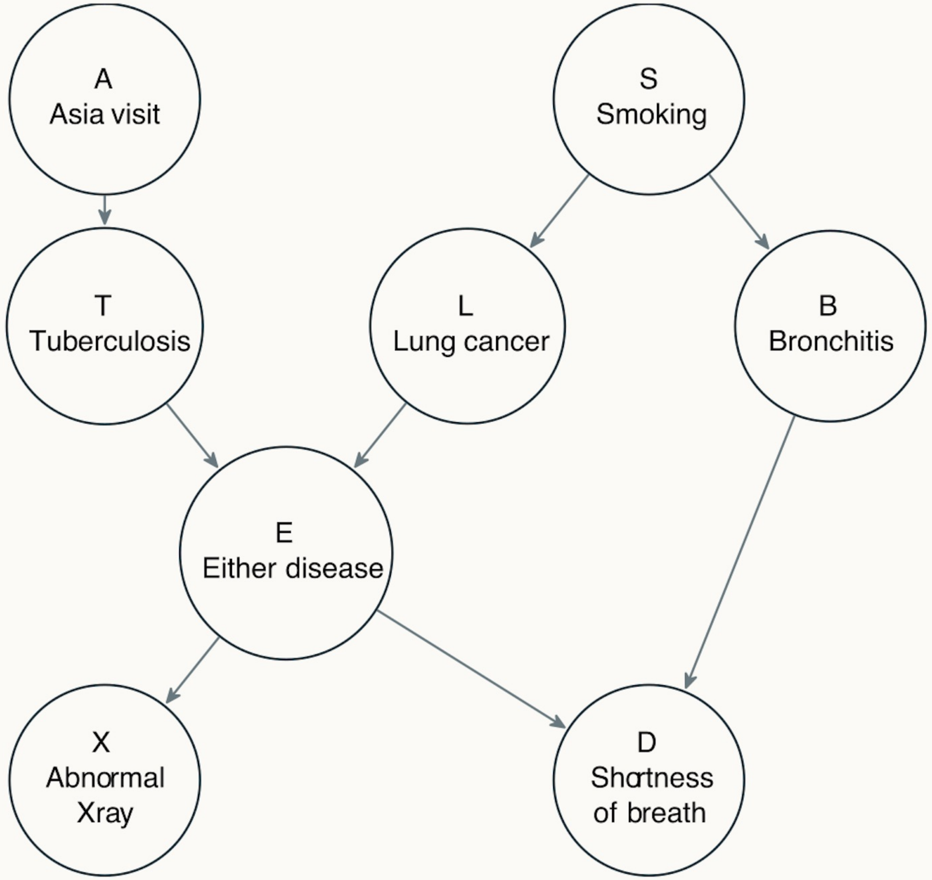
Z

$$P(W \mid X, Y, Z) = P(W \mid X, Z)$$

The conditioning set is {X, Z}, not X alone.

Asia: a Simplified Medical Diagnosis Example.

Asia network



Conditional probability table

$P(T \mid A)$

Asia visit	T = false	T = true
No	97%	**3%**
Yes	68%	**32%**

Equivalent factor

$f_T(A, T) = P(T \mid A)$

Zero-shot CPTs with Large Language Model (Jev)

A Asia visit

root: no parents

population	4%

S Smoking

root: no parents

population	20%

T Tuberculosis

condition on A

A=0	3%
A=1	32%

L Lung cancer

condition on S

S=0	1%
S=1	8%

B Bronchitis

condition on S

S=0	5%
S=1	69%

E Either disease

condition on T, L

T=0, L=0	1%
T=0, L=1	100%
T=1, L=0	100%
T=1, L=1	100%

X Abnormal X-ray

condition on E

E=0	11%
E=1	99%

D Shortness of breath

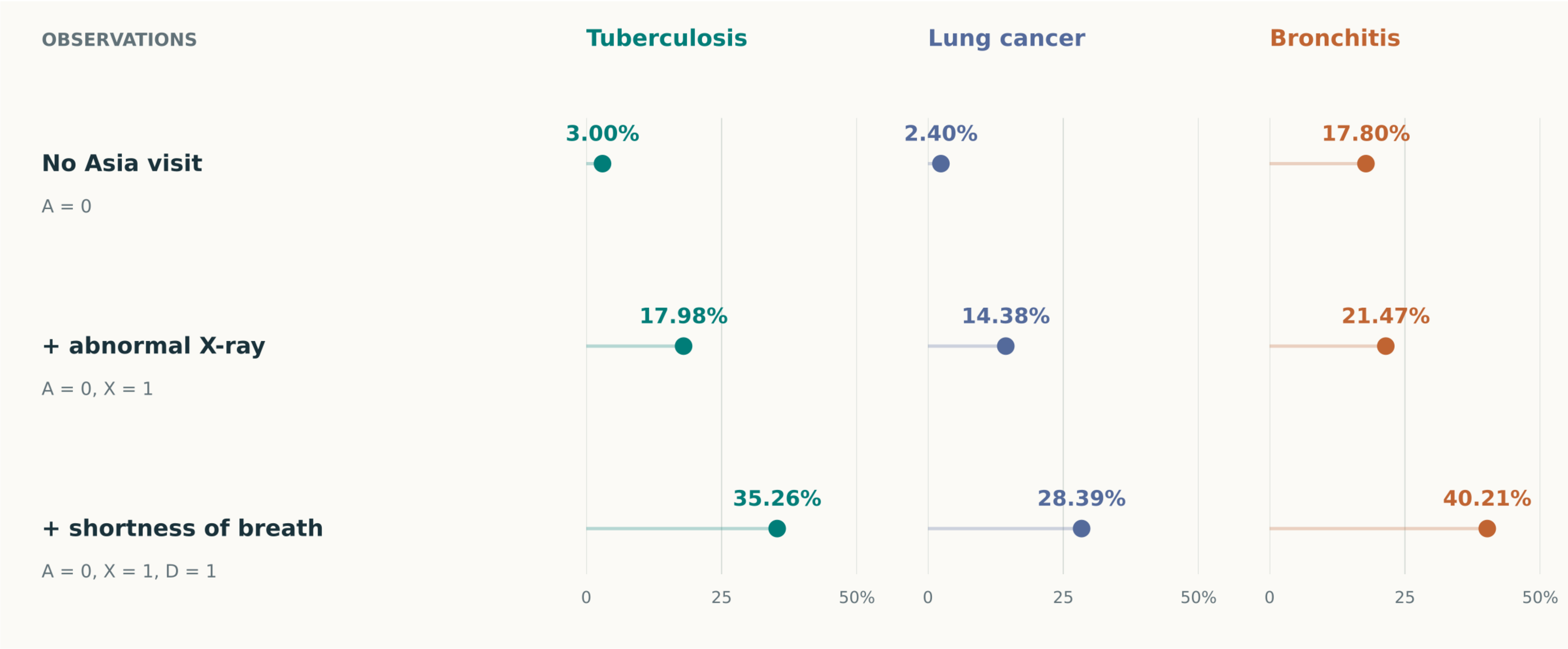
condition on E, B

E=0, B=0	7%
E=0, B=1	85%
E=1, B=0	95%
E=1, B=1	97%

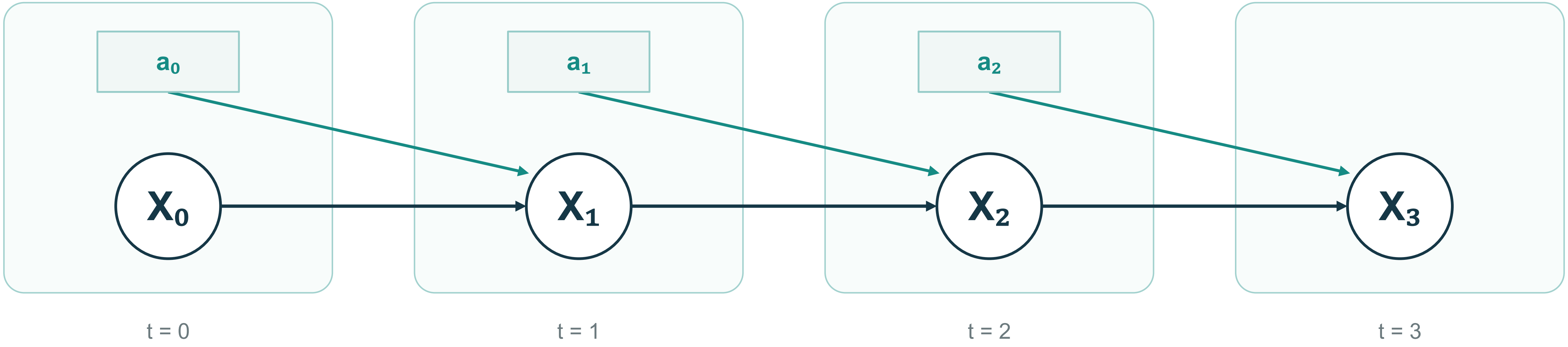
Teal = P(true); remainder = P(false). All bars run from 0 to 100%.

A Asia visit · S smoking · T tuberculosis · L lung cancer · B bronchitis · E either disease

Once CPTs are known, we can perform inference



Dynamic Bayes nets replicate a Bayes net fragment over time.



Exercise: How many possible trajectories?

The room can repeat. The time index does not run backward.

THE STORAGE ADVANTAGE

Local tables avoid storing every possible state history.

ENUMERATE AN 11-STATE JOINT

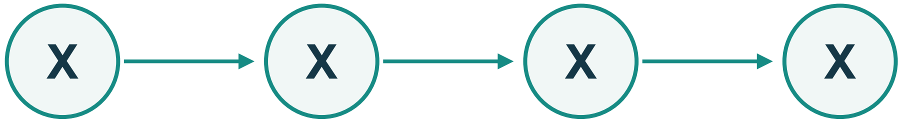
48,828,125

5^{11} entries

REUSE ONE LOCAL MOTION TABLE

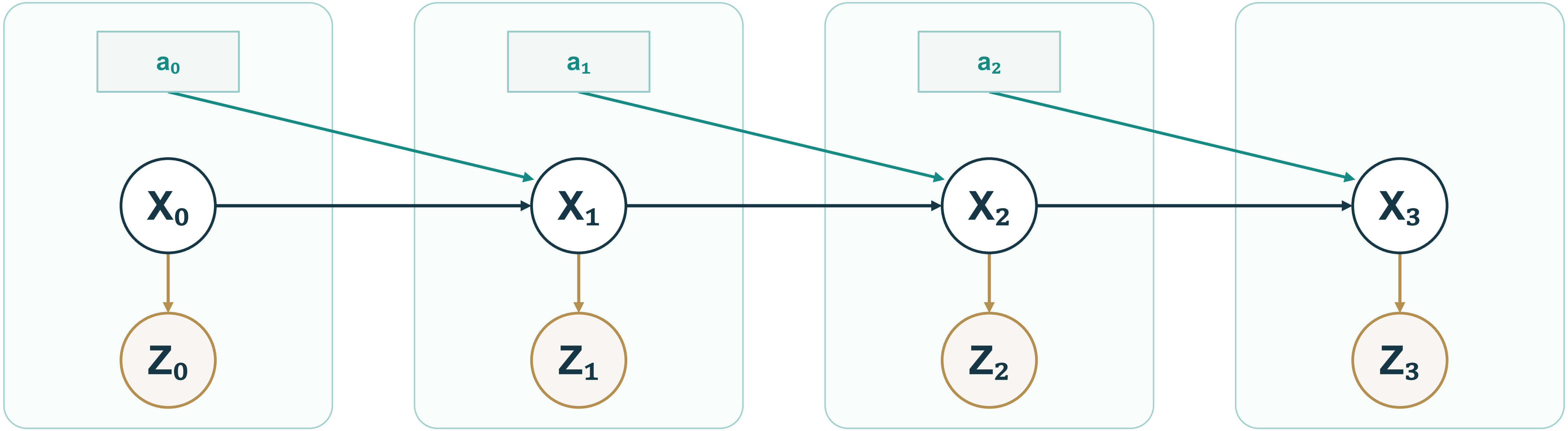
100 + 5

motion CPT + initial prior



Same model at every step.

Each new state generates one new sensor reading.



$$P(Z_t \mid X_t)$$

Motion links one state to the next; sensing links each state to its own measurement.

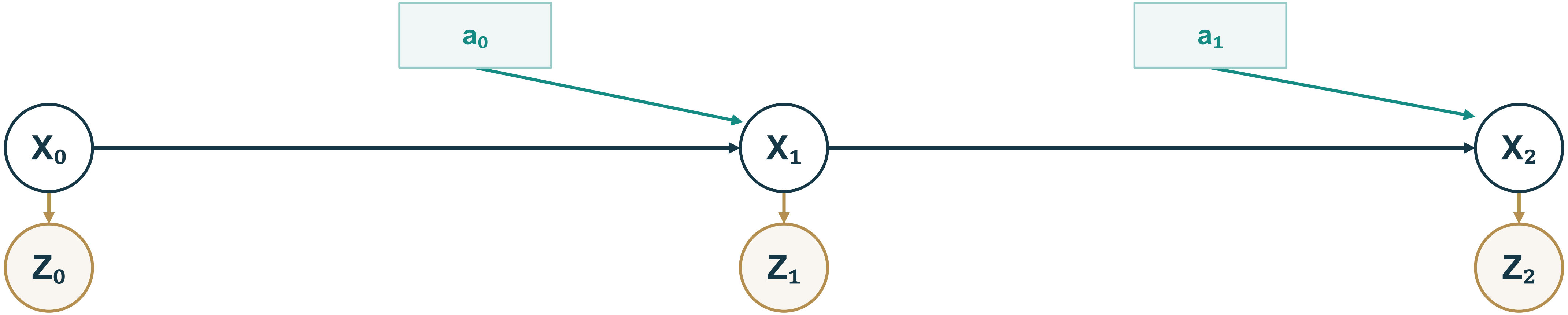
The sensor ignores the past once the current room is known.

Condition on this room.

Use only its sensor distribution.

$$P(Z_t \mid X_{0:t}, Z_{0:t-1}, a_{0:t-1}) = P(Z_t \mid X_t)$$

The joint is a prior *times* motion factors *times* sensor factors.



prior

$$P(X_{0:2}, Z_{0:2} \mid a_{0:1}) = P(X_0)\; P(X_1 \mid X_0, a_0)\, P(X_2 \mid X_1, a_1)$$
$$\times\, P(Z_0 \mid X_0)\, P(Z_1 \mid X_1)\, P(Z_2 \mid X_2)$$

Remember: Know quantities (actions in this case!) are indicated with a square node

How do we *simulate*?
First sample its parents, only then sample a node.

1 · INITIAL STATE

X_0

$$x_0 \sim P(X_0)$$

2 · MEASUREMENT

Z_t

$$z_t \sim P(Z_t \mid x_t)$$

3 · NEXT STATE

X_{t+1}

$$x_{t+1} \sim P(X_{t+1} \mid x_t, a_t)$$

Start from the prior.

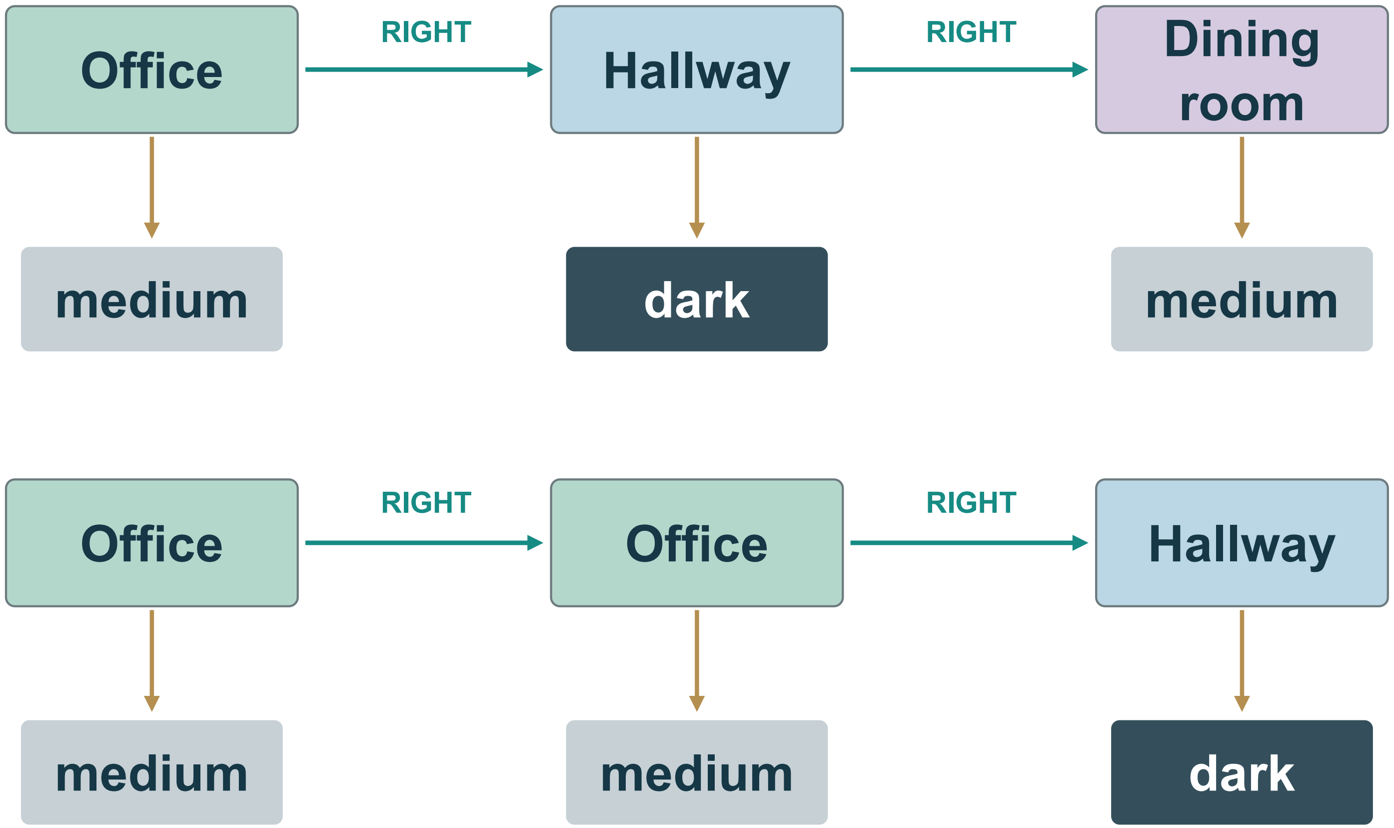
Sample this room's light.

Use the given action.

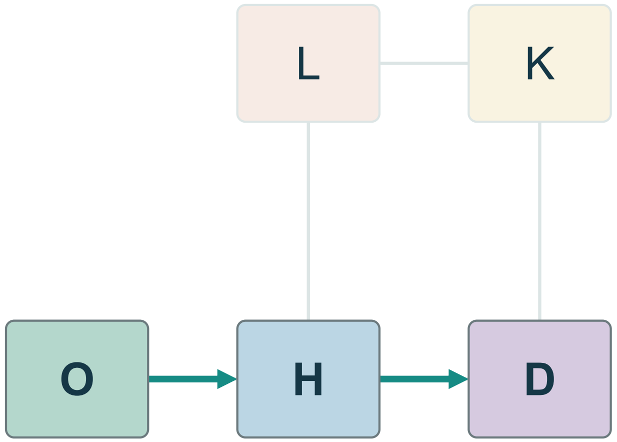
After advancing time, repeat measurement → next state. A *realization* is one possible world.

The same actions can produce different journeys and readings.

SAME START: OFFICE · SAME COMMANDS: RIGHT, RIGHT

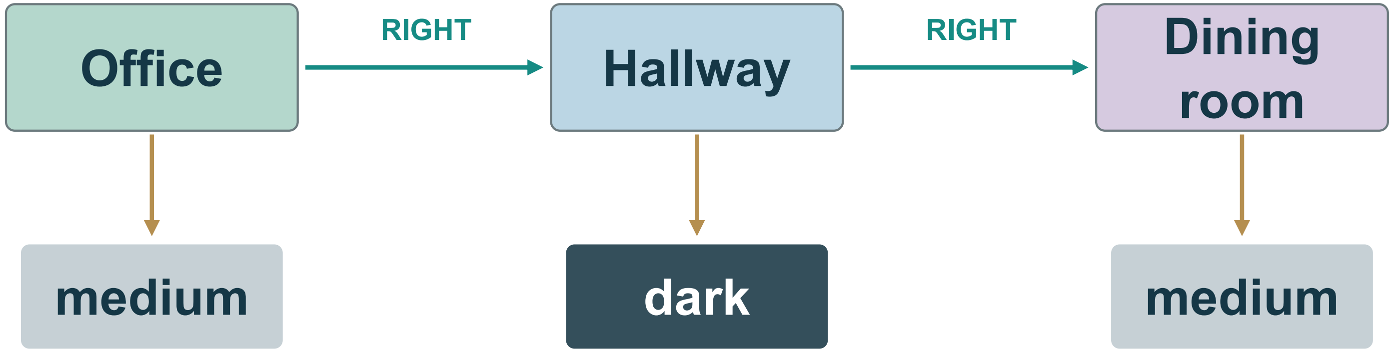


TWO POSSIBLE REALIZATIONS

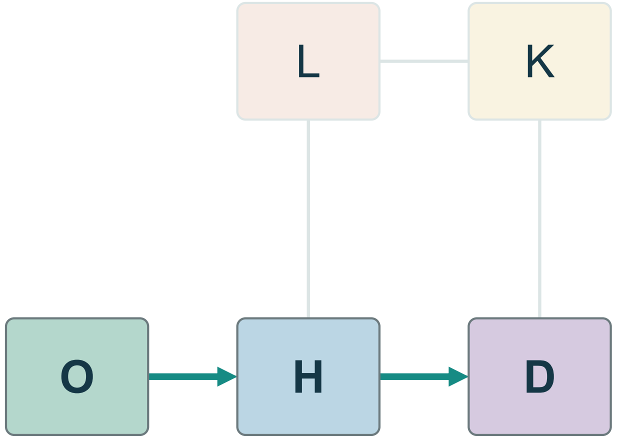


The same actions can produce different journeys and readings.

SAME START: OFFICE · SAME COMMANDS: RIGHT, RIGHT

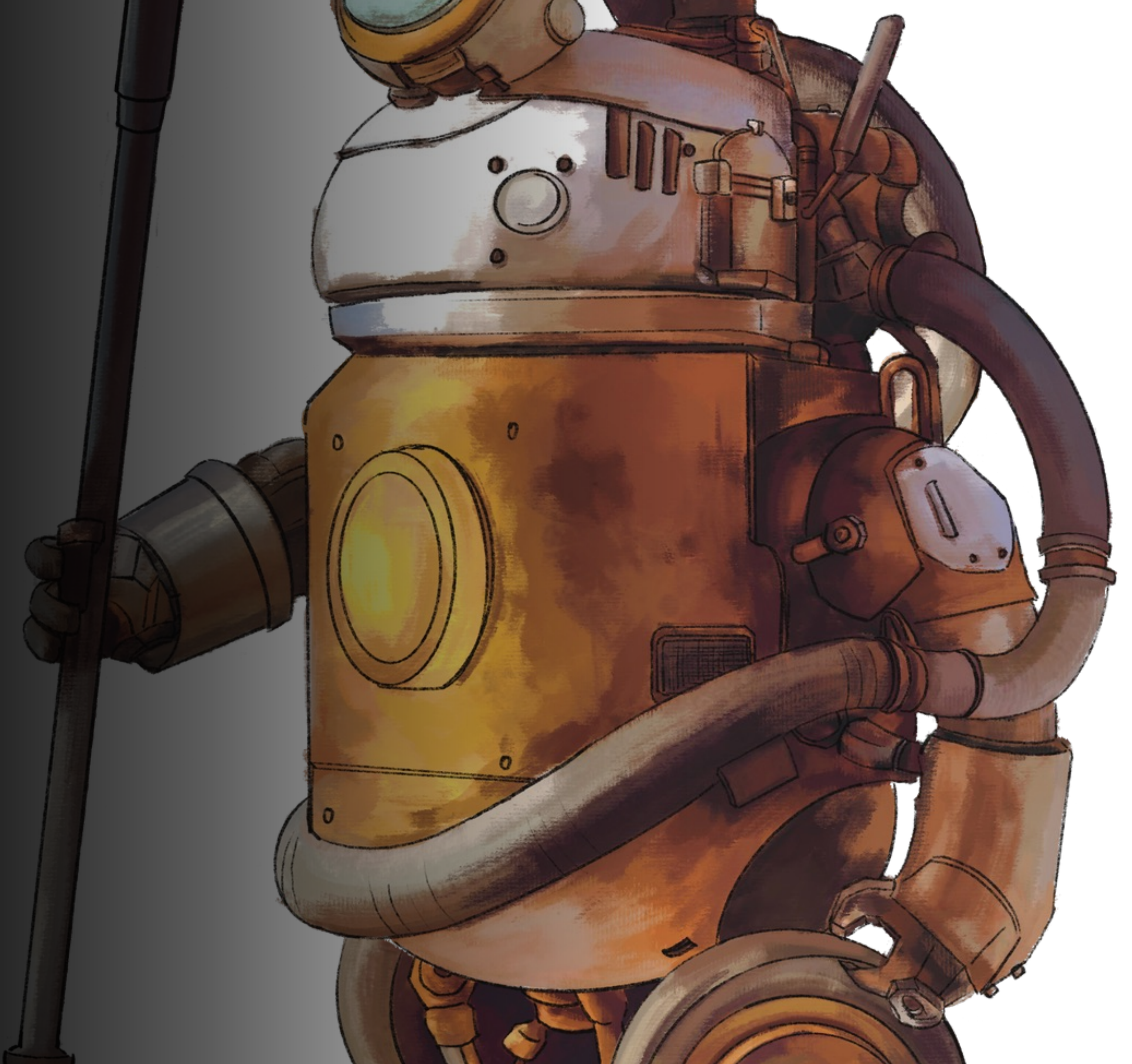


TWO POSSIBLE REALIZATIONS



Exercise: How many possible realizations?

Perception



Simulation reveals the states. *Perception* must infer them.

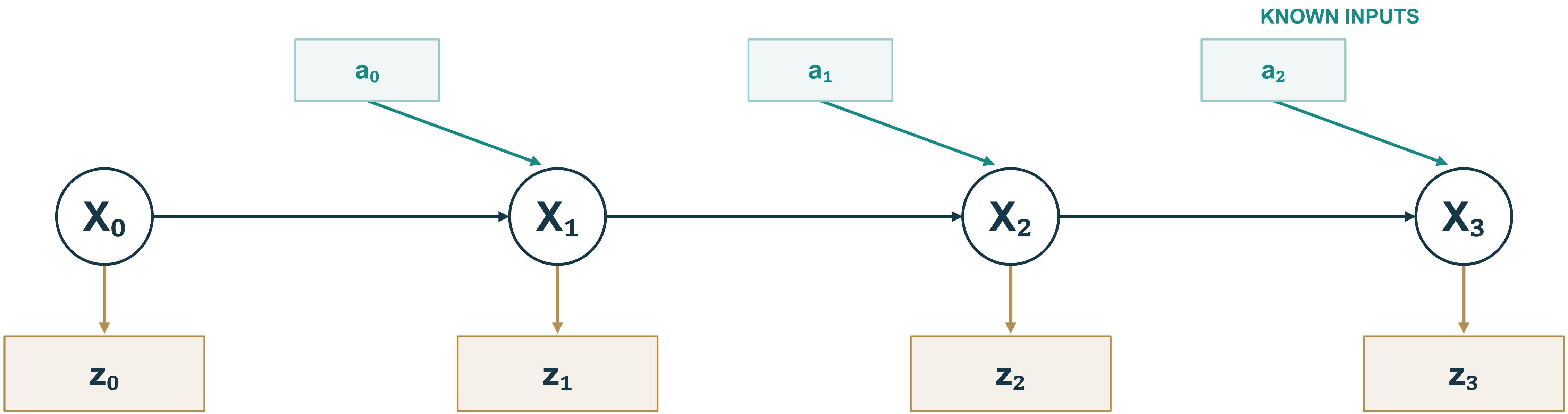
SIMULATION: GENERATE A POSSIBLE WORLD

$$P(X, Z \mid a)$$

PERCEPTION: EXPLAIN THE KNOWN READINGS

$$P(X \mid z, a)$$

HMMs provide a general framework for perception over time.



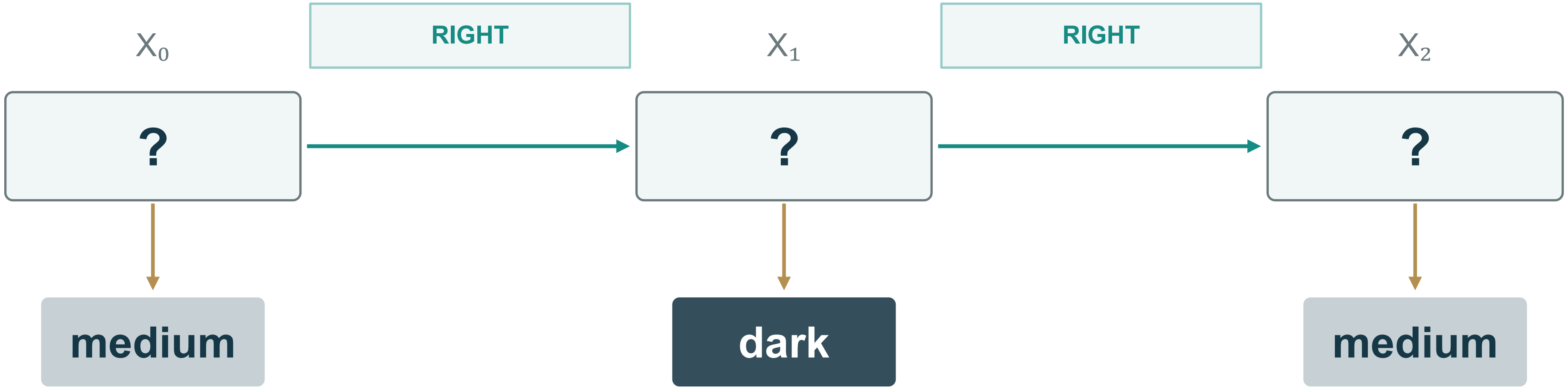
Hidden states: circles

Observed values: boxes

The state sequence is hidden; the actions and measurement values are given.

We infer a state sequence, not just the current room.

Unknown: the whole room history

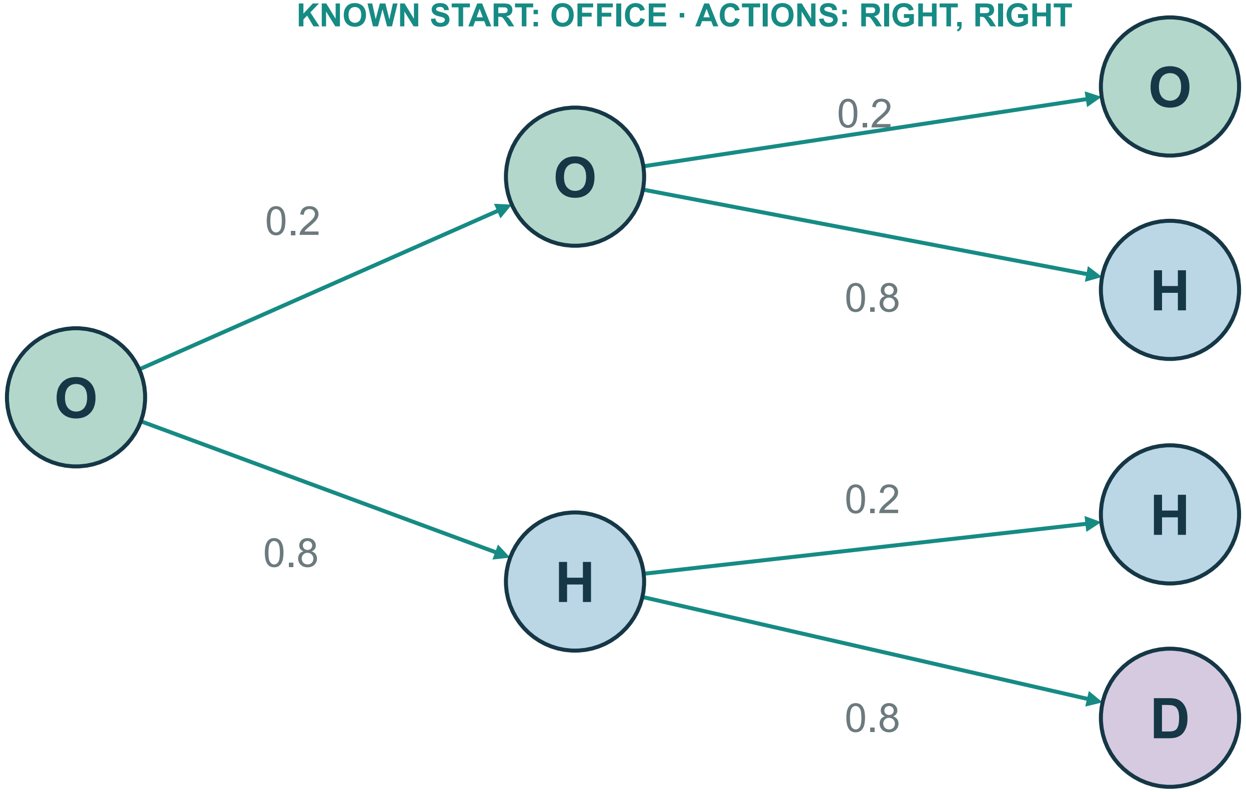


$$P(X_0, X_1, X_2 \mid z_0, z_1, z_2, a_0, a_1)$$

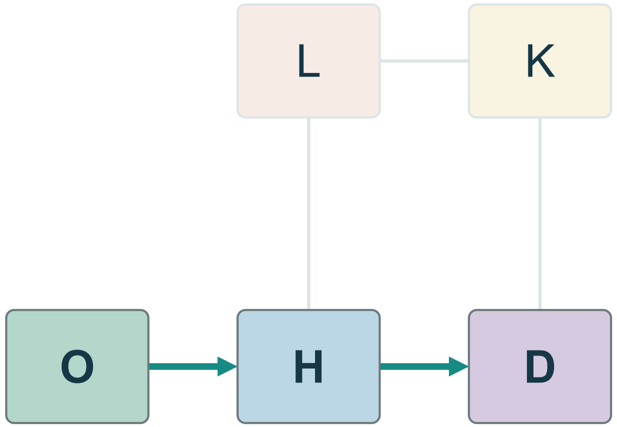
Exercise: How many possible hypotheses?

From the known Office start, only **four paths survive** two right commands.

KNOWN START: OFFICE · ACTIONS: RIGHT, RIGHT



MAP REMINDER



21 other assignments have zero motion probability.

Bayes Law:

$P(X|z;a)$ $posterior$

$\propto P(z|X;a)P(X;a)$ $\propto likelihood$

X prior

$\propto P(z,X;a)$ ∝ joint

using $P(A,B) = P(A|B)P(B)$

Bayes Law:

$$P(X_0, X_1, X_2 | z_0, z_1, z_2; a_0, a_1)$$ *posterior*

$$\propto P(z_0, z_1, z_2 | X_0, X_1, X_2; a_0, a_1)$$ $\propto$*likelihood*

$$* \, P(X_0, X_1, X_2; a_0, a_1)$$ X prior

$$\propto P(z_0, z_1, z_2, X_0, X_1, X_2; a_0, a_1)$$ $\propto$ joint

using $P(A, B) = P(A|B)P(B)$

Fix the evidence, then compare the joint scores of candidate paths.

O	O	O	→ score
O	H	H	→ score
O	H	D	→ score

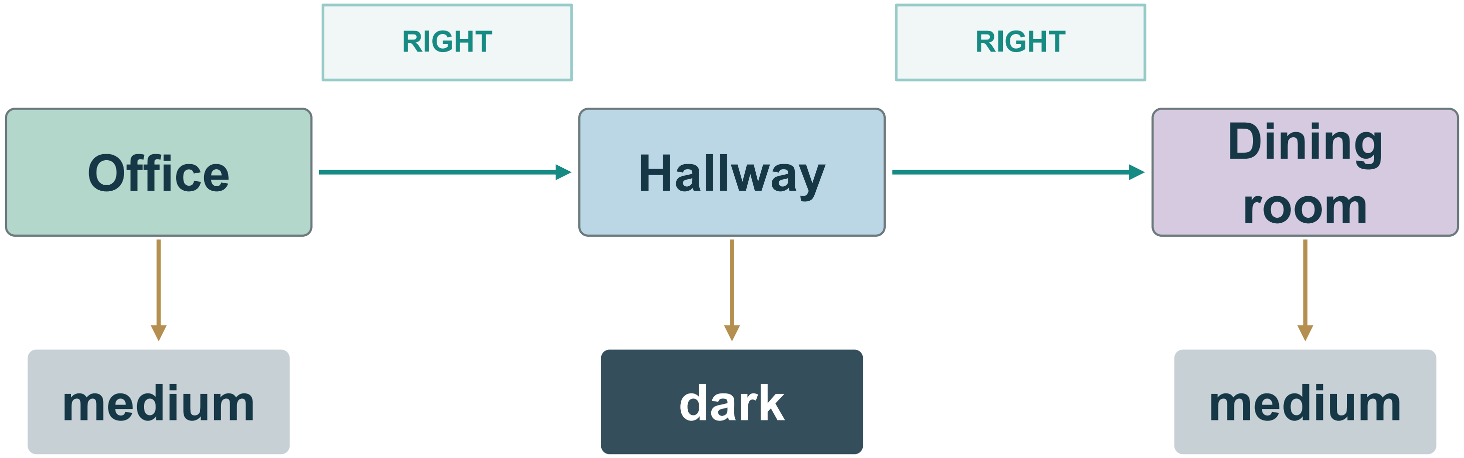
Same evidence
Same actions
Same normalizer

$$q(x) = P(X = x, Z = z \mid a)$$

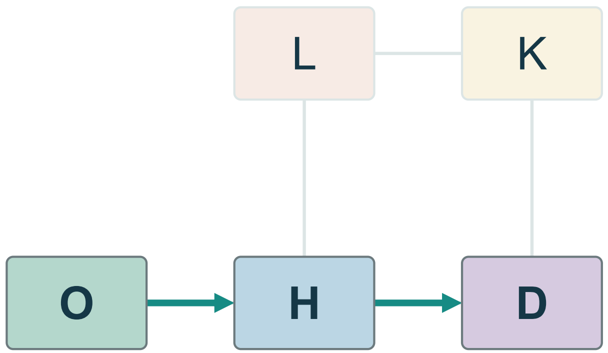
$$P(X = x \mid z, a) = \frac{q(x)}{P(Z = z \mid a)}$$

A path’s score multiplies its prior, motion, and sensor probabilities.

Current room	dark	medium	bright
Living room	0.1	0.1	**0.8**
Kitchen	0.1	0.1	**0.8**
Office	0.2	**0.7**	0.1
Hallway	**0.8**	0.1	0.1
Dining room	0.1	**0.8**	0.1



KNOWN START: OFFICE



1	×	0.7	×	0.8	×	0.8	×	0.8	×	0.8	=	**0.28672**
prior		sensor		motion		sensor		motion		sensor		

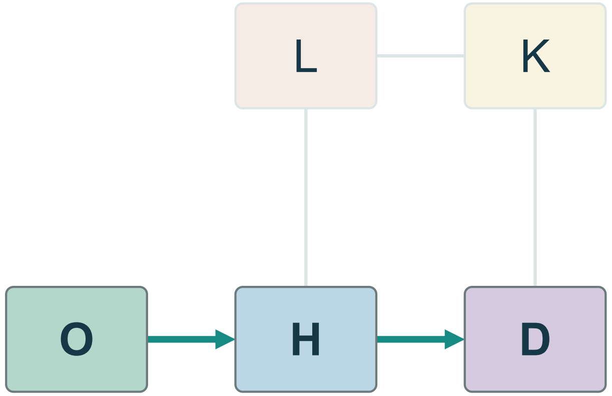
Joint score—not a posterior percentage.

Office → Hallway → Dining room has the highest joint score.

KNOWN START: OFFICE · READINGS: MEDIUM, DARK, MEDIUM

Candidate path	Joint score q(x)
O → O → O	0.00392
O → O → H	0.00224
O → H → H	0.00896
O → H → D	**0.28672**

MOST PROBABLE EXPLANATION



Same prior, actions, and measurements
for every candidate.

The most probable explanation is the highest-scoring complete path.

Illustrative ranking—not numerical scores yet.

$$x^* = \arg\max_{x} \; P(X = x \mid z, a) = \arg\max_{x} \; q(x)$$

The best explanation is the most probable, not certain.

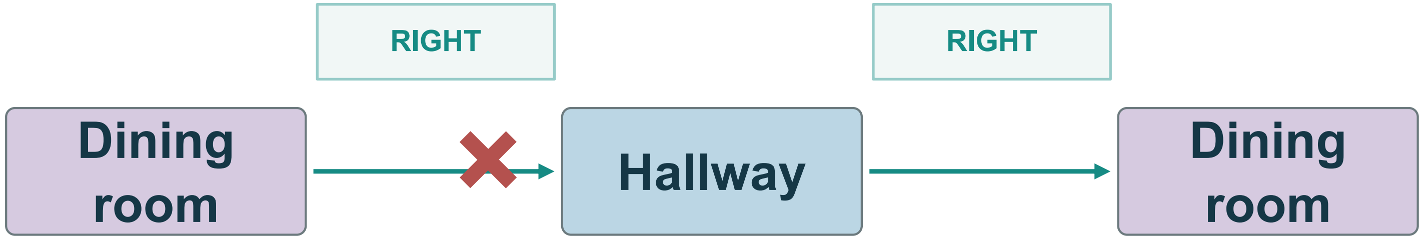
$$P(x \mid z, a) = \frac{q(x)}{0.30184}$$

The remaining paths still carry 5.01%.

The best room for each reading need not form a possible journey.

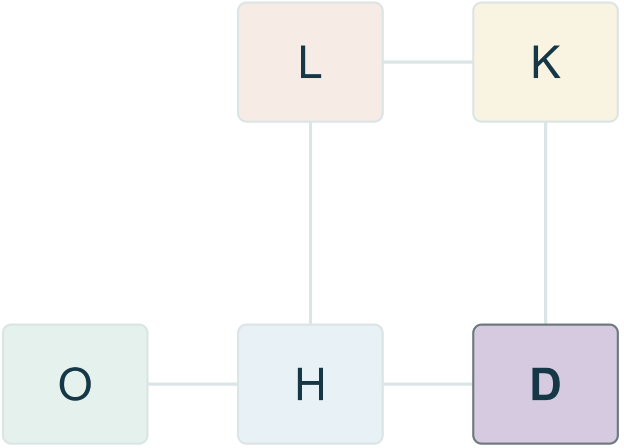
Current room	dark	medium	bright
Living room	0.1	0.1	**0.8**
Kitchen	0.1	0.1	**0.8**
Office	0.2	**0.7**	0.1
Hallway	**0.8**	0.1	0.1
Dining room	0.1	**0.8**	0.1

SENSOR-ONLY PICKS: MEDIUM → DARK → MEDIUM



$$P(H \mid D, \text{right}) = 0$$

RIGHT CANNOT LEAVE DINING



Compare complete paths—not three independent guesses.

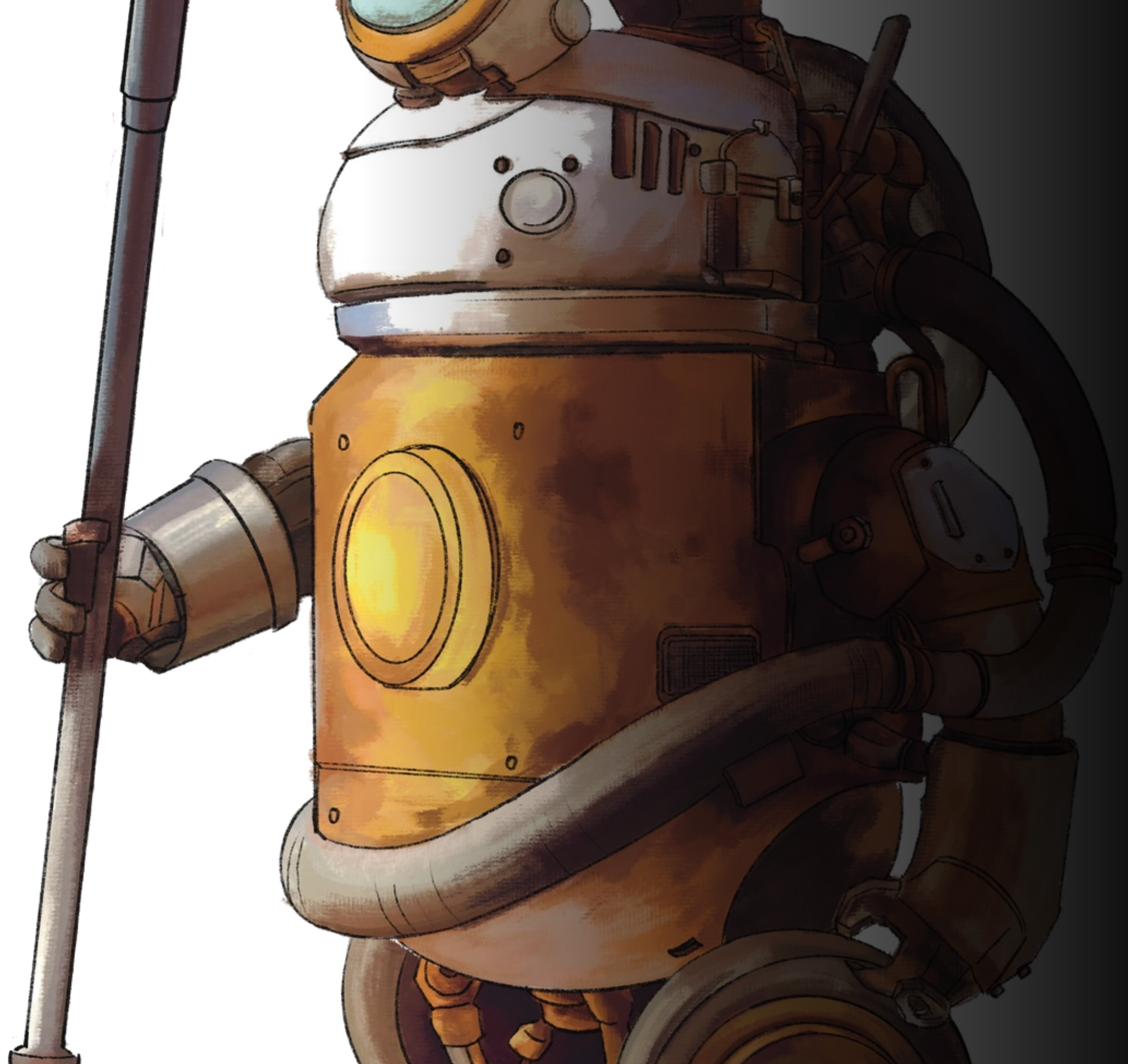
Scoring one path is easy; enumerating all paths is intractable in general !

two actions
25

ten actions
9.8 million

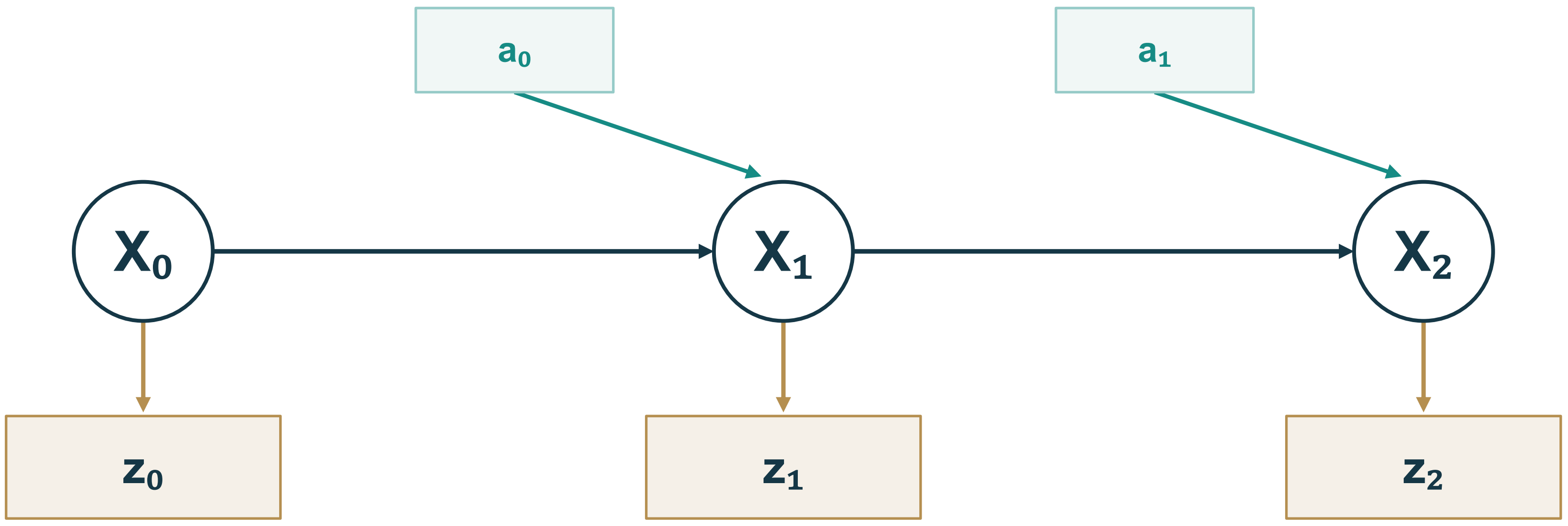
twenty actions
95 trillion

5^T candidate assignments (X_0 fixed)



Takeaways

A motion model links states; a sensor model links states to evidence.



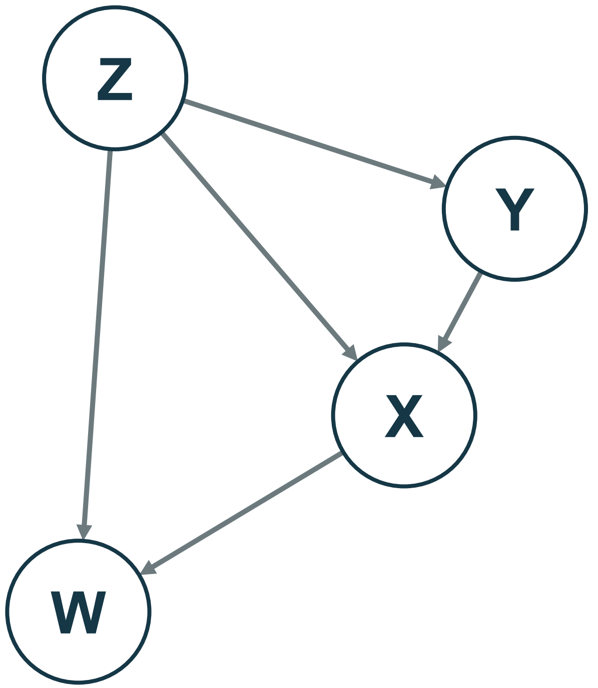
MODEL

SIMULATE

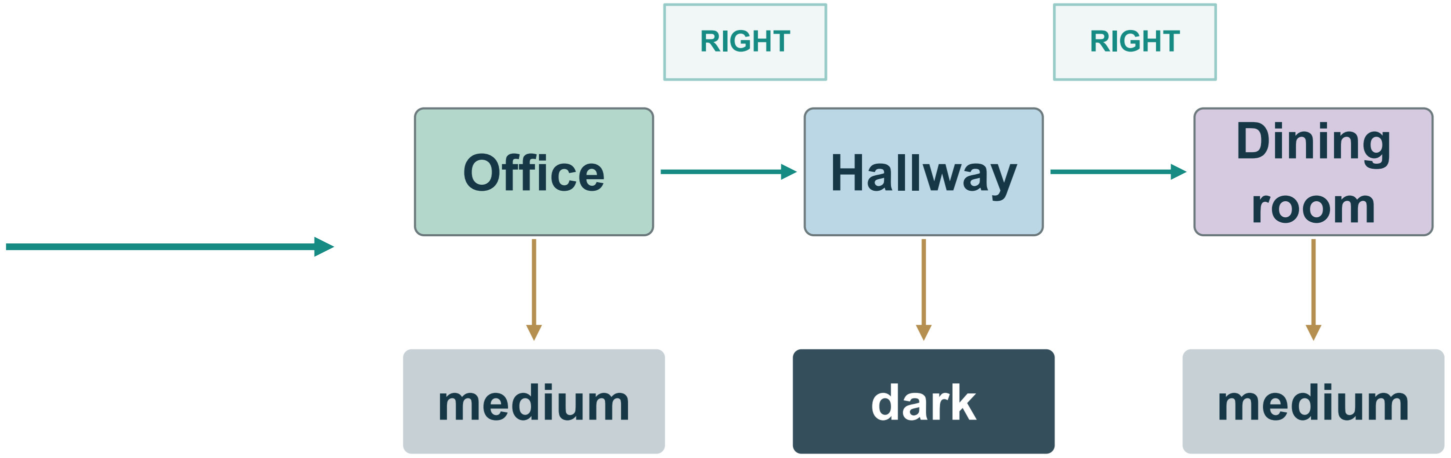
INFER

Prior × motion × sensing → joint model → an explanation of the observed journey.

Bayes nets model the journey.
Inference explains the measurements.



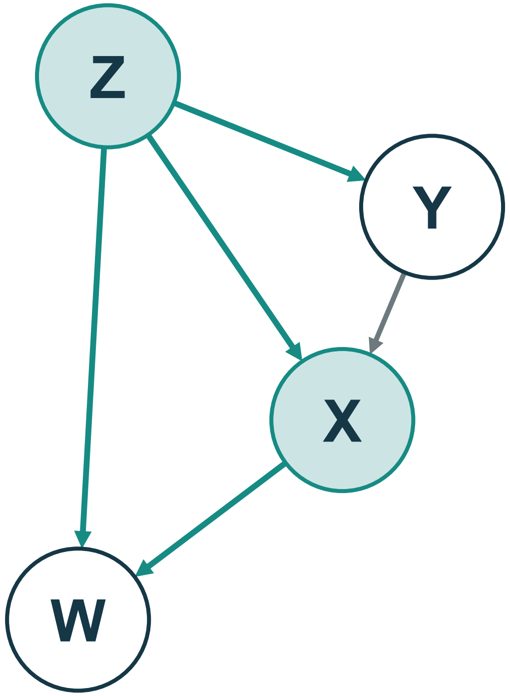
A factored generative model



A likely hidden-state sequence

Next question: how can we avoid enumerating every path?

The chain rule explains why the Bayes-net product works.



CHAIN RULE

$$P(W, X, Y, Z) = P(W \mid X, Y, Z)P(X \mid Y, Z)P(Y \mid Z)P(Z)$$

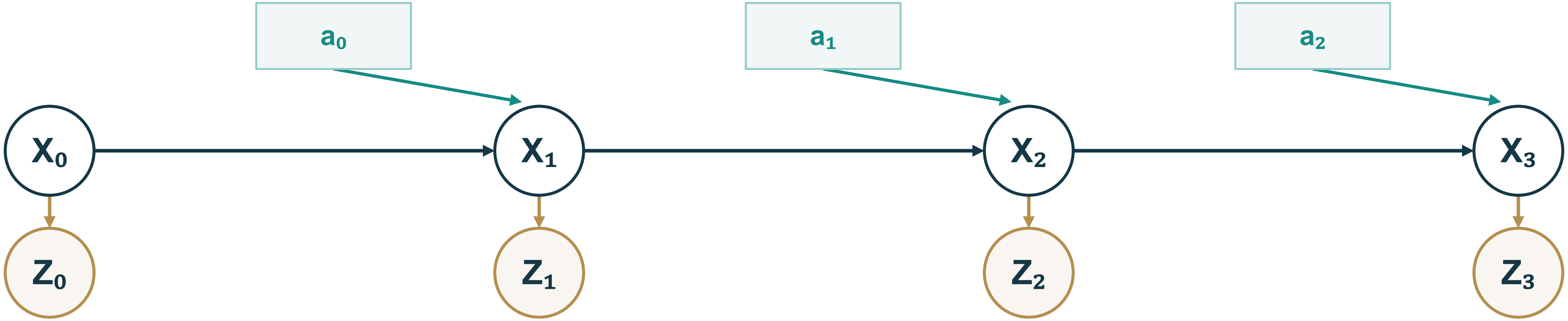
USE THE GRAPH

$$P(W \mid X, Y, Z) = P(W \mid X, Z)$$

SUBSTITUTE

$$P(W, X, Y, Z) = P(W \mid X, Z)P(X \mid Y, Z)P(Y \mid Z)P(Z)$$

The full sequence uses T actions and T+1 states and measurements.



$$P(X_{0:T}, Z_{0:T} \mid a_{0:T-1}) =$$

$$P(X_0) \prod_{t=0}^{T-1} P(X_{t+1} \mid X_t, a_t) \prod_{t=0}^{T} P(Z_t \mid X_t)$$