CSCE A405/A605 (Adv) Artificial Intelligence

Probabilistic Reasoning over Time

Ref: Artificial Intelligence: A Modern Approach, 4th ed by Stuart Russell and Peter Norvig, chapter 14

Instructor: Masoumeh Heidari (mheidari2@Alaska.edu)

UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

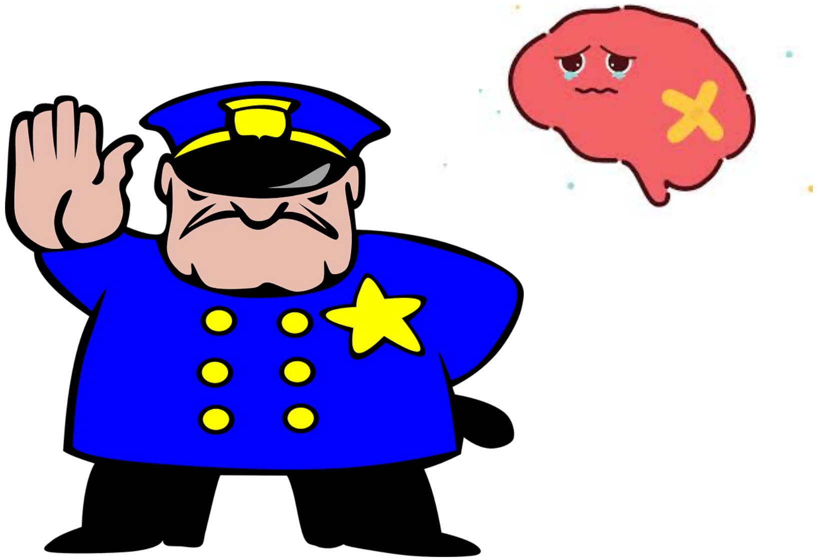
Example

Security guard stationed at a secret underground installation.

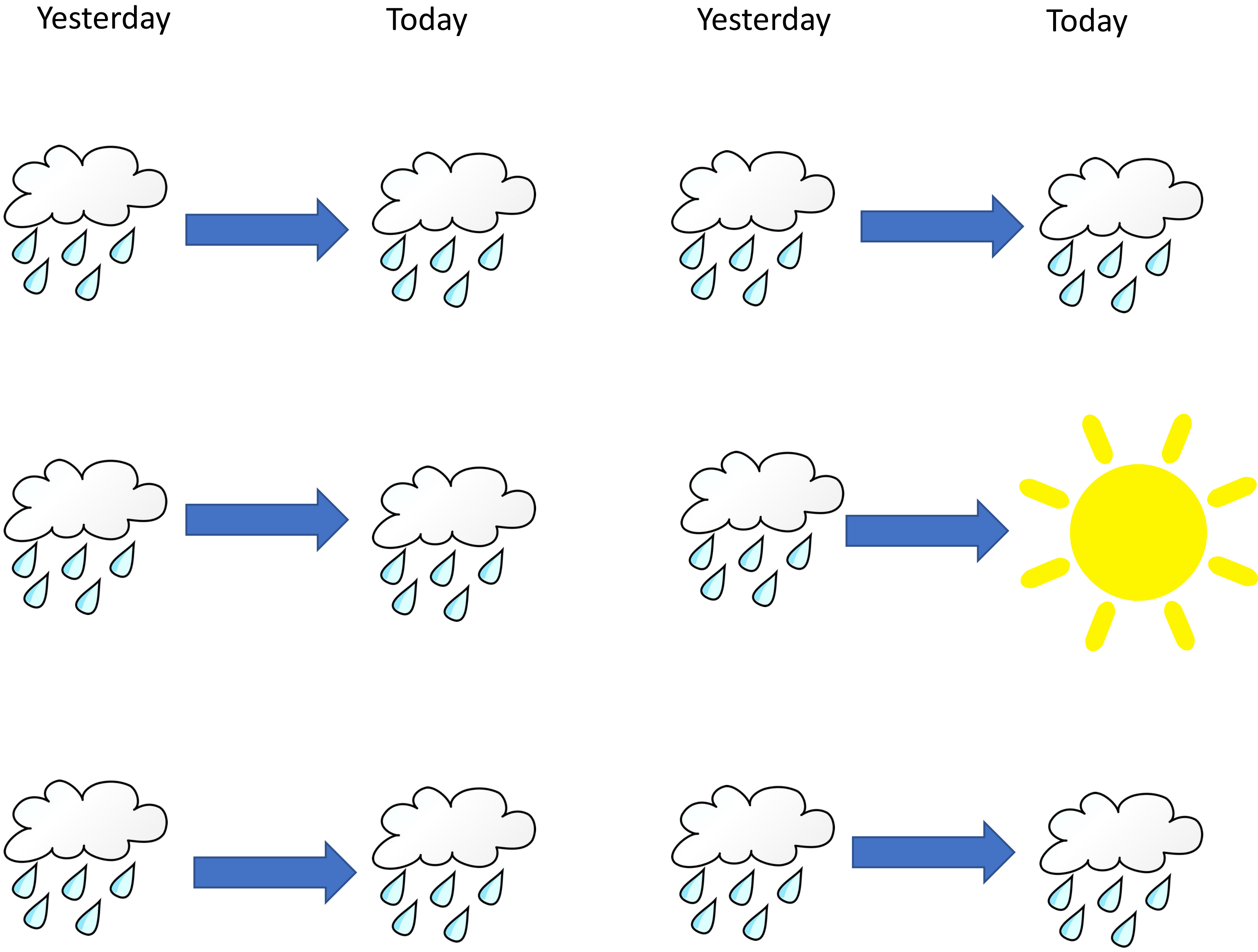
You want to know whether it’s raining today,
but your only access to the outside world occurs each morning when you see the director coming in with, or without, an umbrella.

Example

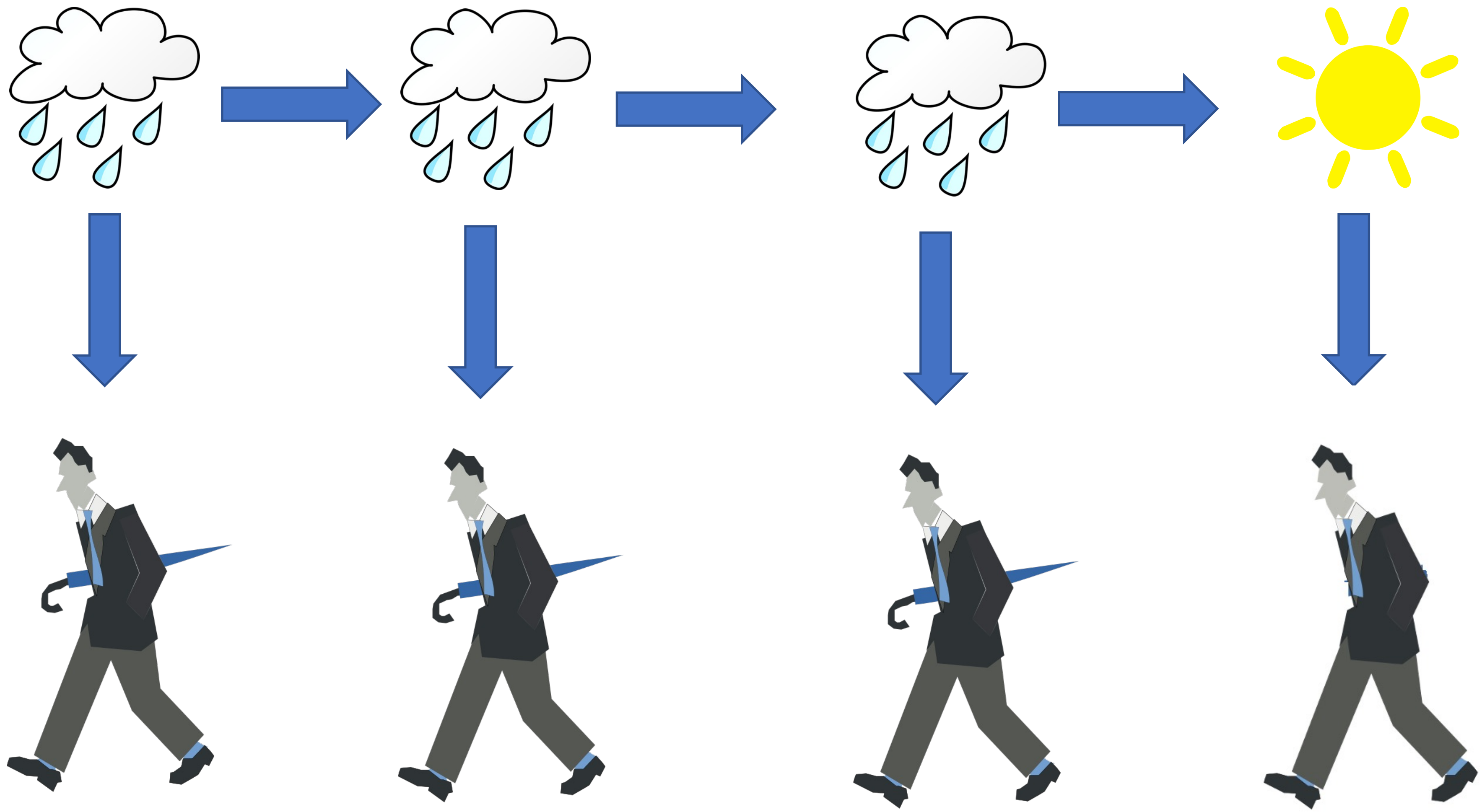
During your childhood you discovered that most people bring umbrellas because it is raining, but sometimes it's just because they think it may rain!



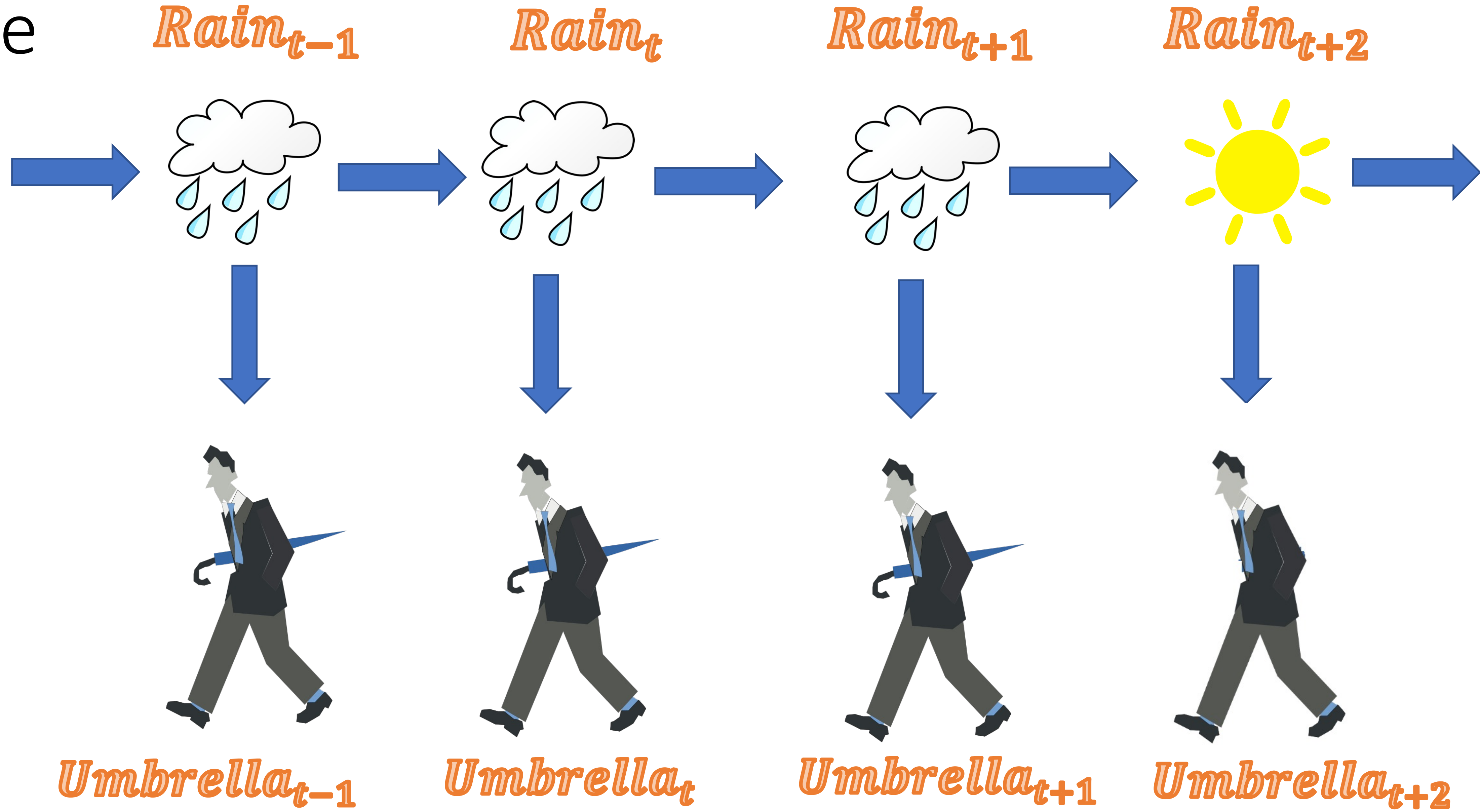
Example



Example



Example


$Rain_{t-1}$
$Rain_t$
$Rain_{t+1}$
$Rain_{t+2}$
$Umbrella_{t-1}$
$Umbrella_t$
$Umbrella_{t+1}$
$Umbrella_{t+2}$

Probabilistic reasoning over time

- We try to interpret the present, understand the past, and perhaps predict the future, even when very little is crystal clear.

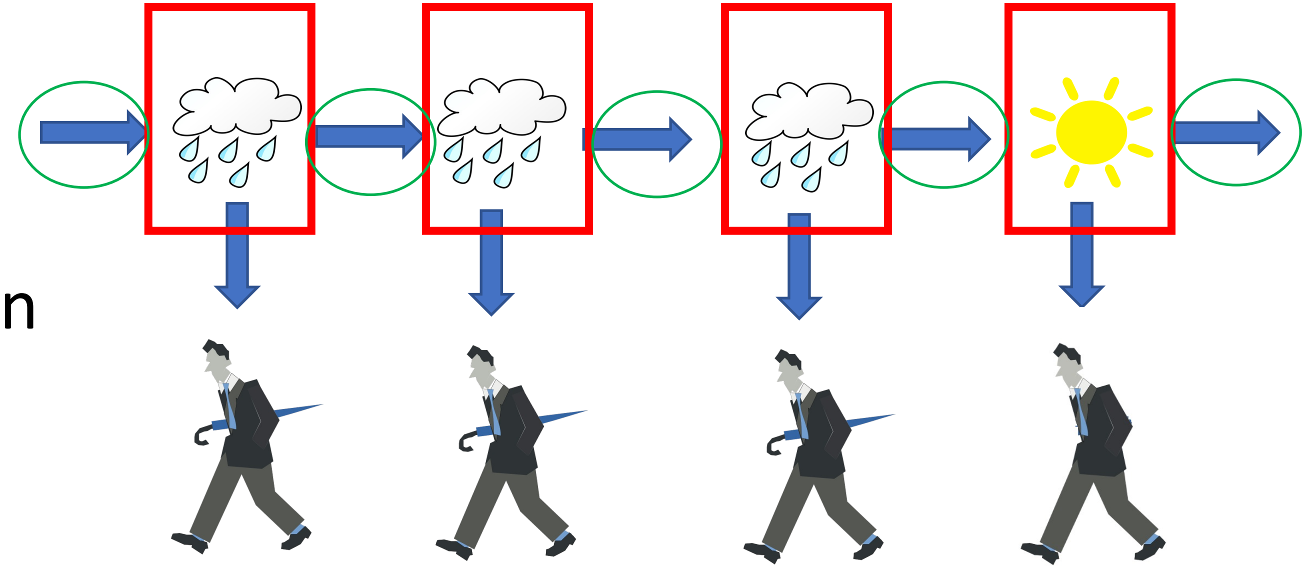
- Agents in partially observable environments must be able to keep track of the current state, to the extent that their sensors allow.

Probabilistic reasoning over time

- In Section 4.4 we showed a methodology for doing that:

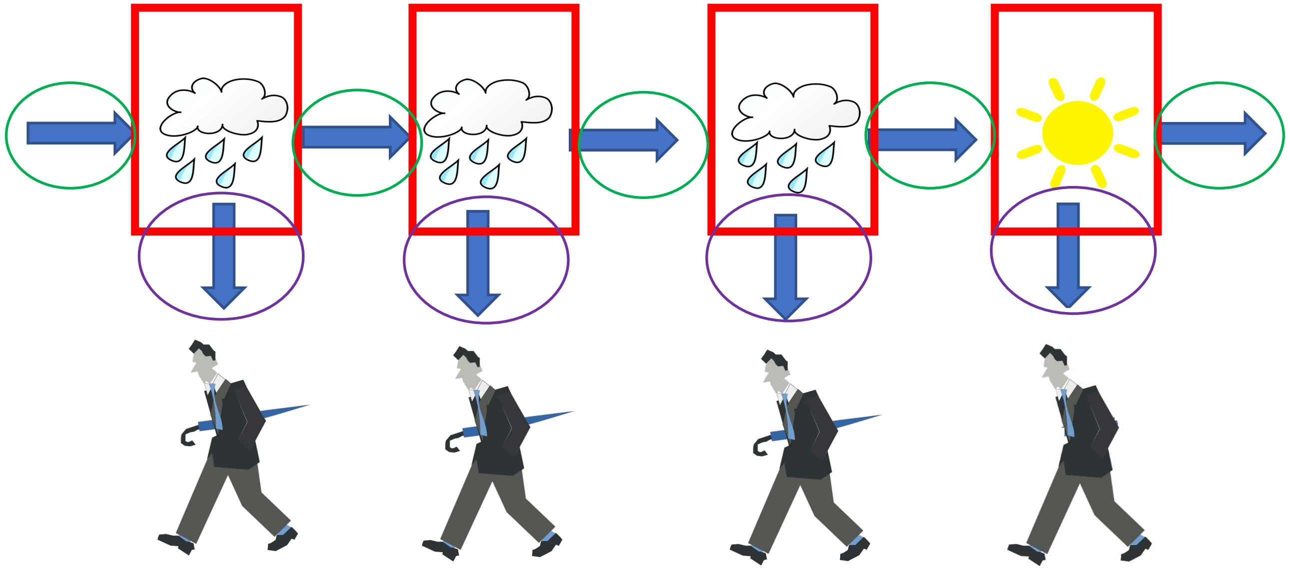
an agent maintains a belief state that represents which states of the world are currently possible.

From the belief state and a transition model, the agent can predict how the world might evolve in the next time step.



Probabilistic reasoning over time

an agent maintains a belief state that represents which states of the world are currently possible.



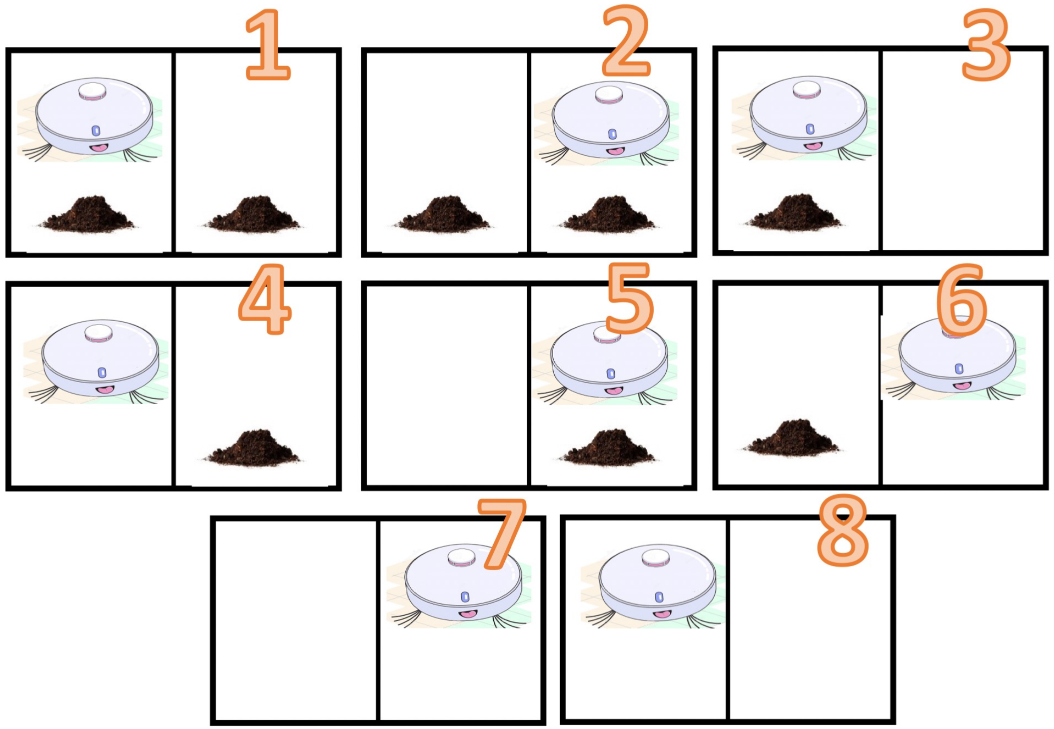
From the belief state and a transition model, the agent can predict how the world might evolve in the next time step.

From the percepts observed and a sensor model, the agent can update the belief state.

In Chapter 4 belief states were represented by explicitly enumerated sets of states.

Probabilistic reasoning over time

Enumerated sets of states



- Initial belief state?
  {1,2,3,4,5,6,7,8}

What if the agent moves right?

Updated belief state: {2,5,6,7}

The agent has gained information without perceiving anything!

The action of moving right aimed to reduce uncertainty about the current state.

[Right, Suck] -> {6,7}
[Right, Suck, Left, Suck] -> {8}

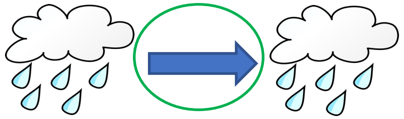
Probabilistic reasoning over time

- That approach defined belief states in terms of which world states were possible, but could say nothing about which states were likely or unlikely.
- In this chapter, we use probability theory to quantify the degree of belief in elements of the belief state.
- A changing world is modeled using a variable for each aspect of the world state at each point in time.

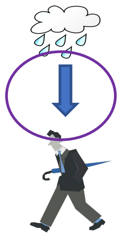
Probabilistic reasoning over time

- The transition and sensor models may be uncertain:

- **Transition model**: the transition model describes the probability distribution of the variables at time *t*, given the state of the world at past times.



- **Sensor model:** describes the probability of each percept at time *t*, given the current state of the world.



Time and uncertainty

- We have developed our techniques for probabilistic reasoning in the context of static worlds. In static worlds each random variable has a single fixed value.

- For example, when repairing a car, we assume that whatever is broken remains broken during the process of diagnosis;

- Our job is to infer the state of the car from observed evidence, which also remains fixed.

Time and uncertainty

- Now consider a slightly different problem: treating a diabetic patient.
- As in the case of car repair, we have evidence such as recent insulin doses, food intake, blood sugar measurements, and other physical signs.
- The task is to assess the current state of the patient, including the actual blood sugar level and insulin level.
- Given this information, we can make a decision about the patient’s food intake and insulin dose.

Time and uncertainty

- Unlike the case of car repair, here the dynamic aspects of the problem are essential.

- Blood sugar levels and measurements thereof can change rapidly over time, depending on recent food intake and insulin doses, metabolic activity, the time of day, and so on.

- To assess the current state from the history of evidence and to predict the outcomes of treatment actions, we must model these changes.

Time and uncertainty

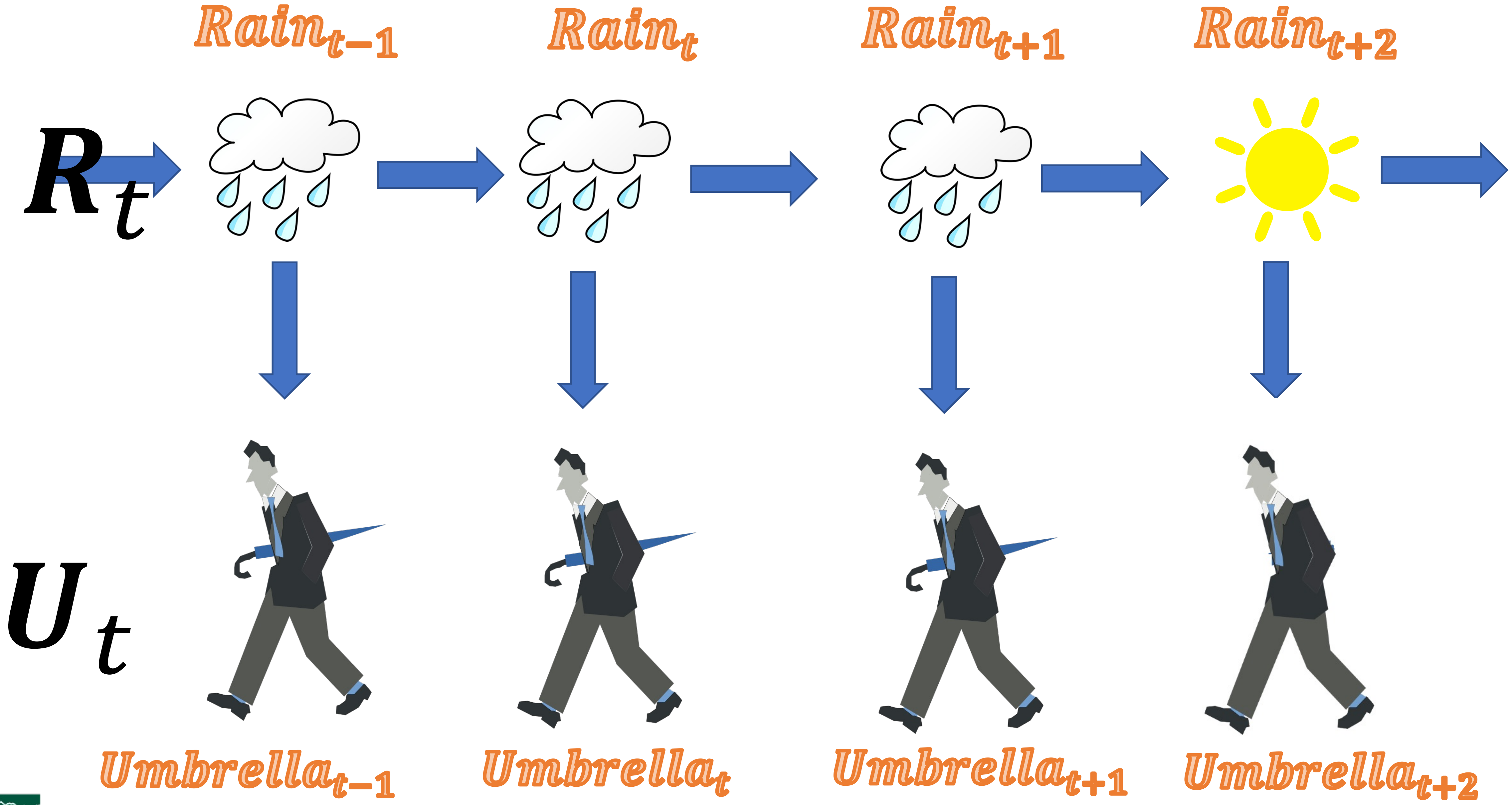
- The same considerations arise in many other contexts, such as:
  - Tracking the location of a robot.
  - Tracking the economic activity of a nation.
  - Making sense of a spoken or written sequence of words.
- We first consider discrete-time models, in which the world is viewed as a series of snapshots or time slices.

Time and uncertainty

- In the discrete-time models we'll just number the time slices 0, 1, 2, and so on, rather than assigning specific times to them.
- Typically, the time interval Δ between slices is assumed to be the same for every interval.
- For any particular application, a specific value of Δ has to be chosen:
  - Sometimes this is dictated by the sensor; for example, a video camera might supply images at intervals of $\frac{1}{30}$ of a second.
  - In other cases, the interval is dictated by the typical rates of change of the relevant variables; for example, in the case of blood glucose monitoring, things can change significantly in the course of ten minutes, so a one-minute interval might be appropriate.
  - On the other hand, in modeling continental drift over geological time, an interval of a million years might be fine.

Time and uncertainty

- Each time slice in a discrete-time probability model contains a set of random variables, some observable and some not.
- For simplicity, we will assume that the same subset of variables is observable in each time slice.
- We will use $\boldsymbol{X}_t$ to denote the set of state variables at time t, which are assumed to be unobservable, and $\boldsymbol{E}_t$ to denote the set of observable evidence variables. The observation at time t is $\boldsymbol{E}_t = \boldsymbol{e}_t$ for some set of value $\boldsymbol{e}_t$.


$Rain_{t-1}$
$Rain_t$
$Rain_{t+1}$
$Rain_{t+2}$
R_t
U_t
$Umbrella_{t-1}$
$Umbrella_t$
$Umbrella_{t+1}$
$Umbrella_{t+2}$

Time and uncertainty

- Other problems can involve larger sets of variables. In the diabetes example, the evidence variables might be $MeasuredBloodSuger_t$ and $PulseRate_t$ while the state variables might include $BloodSuger_t$ *and* $StomachContents_t$.

- We will assume that the state sequence starts at *t=0* and evidence start at *t=1.*

- Hence, our umbrella world is represented by state variables $R_0, R_1, R_2, \ldots$ and evidence variables $U_1, U_2, \ldots$ we will use the notation *a:b* to denote the sequence of integers from a to b and the notation $\boldsymbol{X_{a:b}}$ to denote the set of variables from $\boldsymbol{X_a}$ to $\boldsymbol{X_b}$ inclusive. For example, U_1, U_2, U_3.

Transition and sensor models

- With the set of state and evidence variables for a given problem decided on, the next step is to specify how the world evolves (the transition model) and how the evidence variables get their values (the sensor model).

- The transition model specifies the probability distribution over the latest state variables, given the previous values, that is, $P(X_t|X_{0:t-1})$.

- Problem? The set $X_{0:t-1}$ is unbounded in size as *t* increases.
- Solution: Markov assumption.

Transition and sensor models

- **Markov assumption:** the current state depends on only a finite fixed number of previous state.

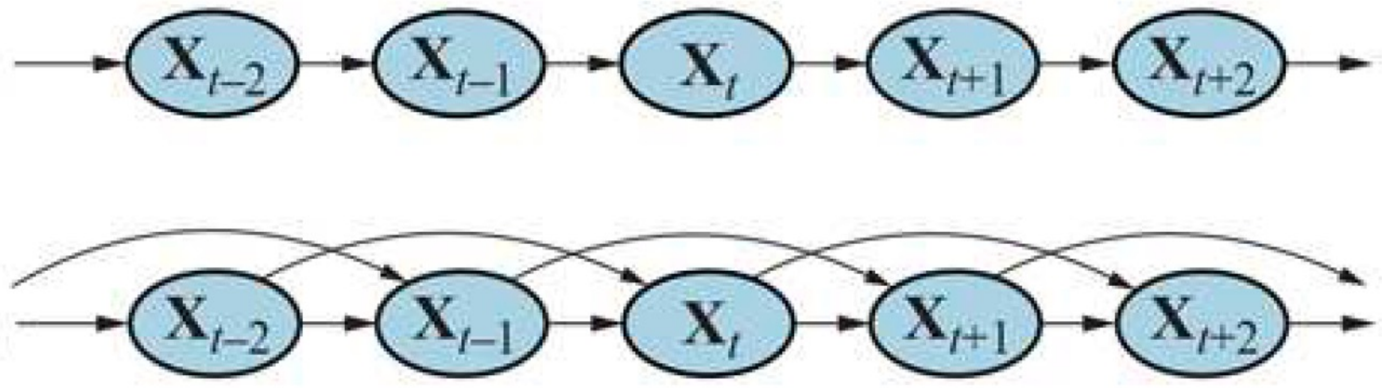
- The simplest is the first order Markov process, in which the current state depends only on the previous state and not on any earlier states.
- In other words, a state provides enough information to make the future conditionally independent of the past, and we have $\boldsymbol{P}(\boldsymbol{X}_t|\boldsymbol{X}_{0:t-1}) = \boldsymbol{P}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})$.

Transition and sensor models

- Hence, in a first-order Markov process, the transition model is the conditional distribution $\boldsymbol{P(X_t|X_{t-1})}$.

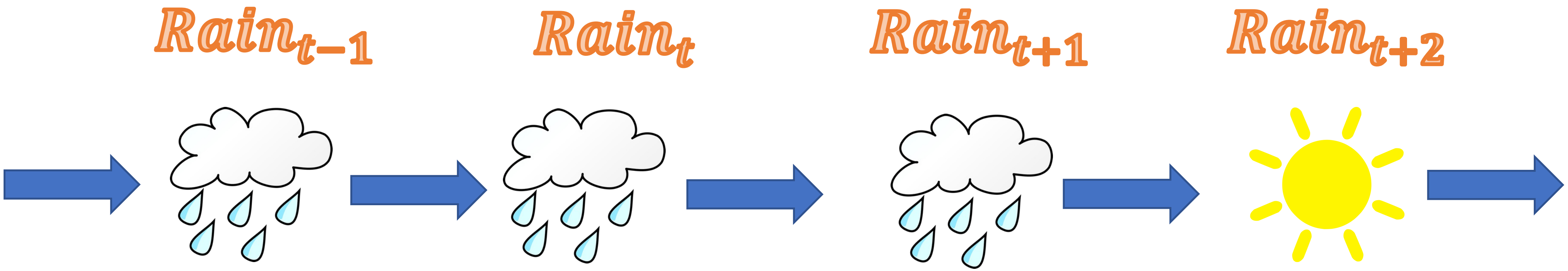
- The transition model for a second-order Markov process is the conditional distribution $\boldsymbol{P(X_t|X_{2-1}, X_{t-1})}$.

- The Bayesian network structures corresponding to first-order and second-order Markov processes:



Transition and sensor models

- Another Problem? there are infinitely many possible values of t. Do we need to specify a different distribution for each time step?
- Solution: We avoid this problem by assuming that changes in the world state are caused by a time-homogeneous process—that is, a process of change that is governed by laws that do not themselves change over time.



Transition and sensor models

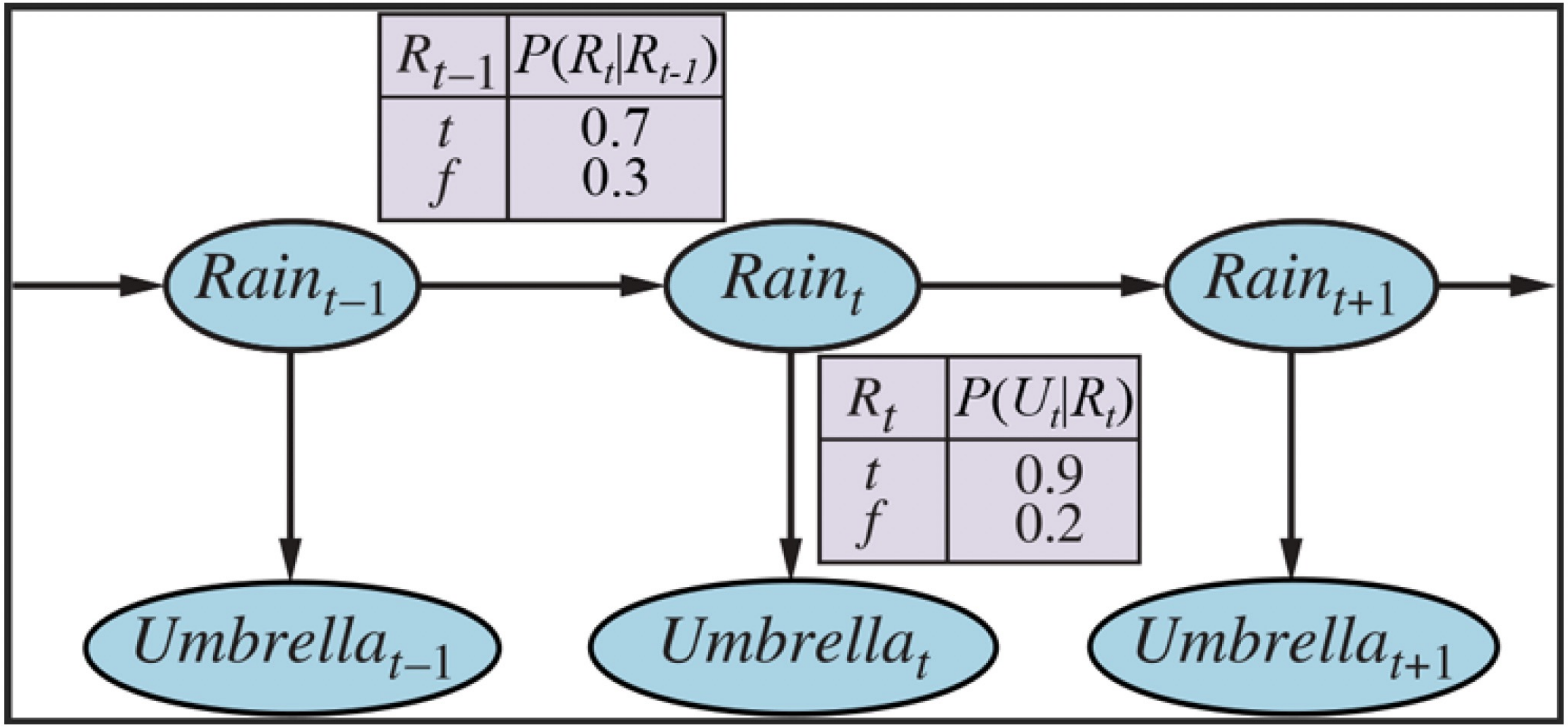
- In the umbrella world, then, the conditional probability of rain, $\boldsymbol{P(R_t|R_{t-1})}$, is the same for all *t*, and we need specify only one conditional probability table.

- The evidence variables could depend on previous variables as well as the current state variables, but any state that's worth its salt should suffice to generate the current sensor values.

Sensor model/observation model

- **Sensor Markov assumption:** $\boldsymbol{P(E_t|X_{0:t}, E_{1:t-1})} = \mathbf{P(E_t|X_t)}$.

Transition and sensor models



Transition and sensor model

- In addition to specifying the transition and sensor models, we need to say how everything gets started—the prior probability distribution at time 0, $\boldsymbol{P}(\boldsymbol{X}_{\mathbf{0}})$.

- Now we have a specification of the complete joint distribution over all the variables. For every time step t,

$$\mathbf{P}(\mathbf{X}_{0:t}, \mathbf{E}_{1:t}) = \mathbf{P}(\mathbf{X}_0) \prod_{i=1}^{t} \mathbf{P}(\mathbf{X}_i | \mathbf{X}_{i-1}) \mathbf{P}(\mathbf{E}_i | \mathbf{X}_i).$$

Transition and sensor model

$$\mathbf{P}(\mathbf{X}_{0:t}, \mathbf{E}_{1:t}) = \mathbf{P}(\mathbf{X}_0) \prod_{i=1}^{t} \mathbf{P}(\mathbf{X}_i | \mathbf{X}_{i-1}) \mathbf{P}(\mathbf{E}_i | \mathbf{X}_i).$$

The initial state model

The transition model

The sensor model

This equation defines the semantics of the family of temporal models represented by the three terms.

Inference in temporal models

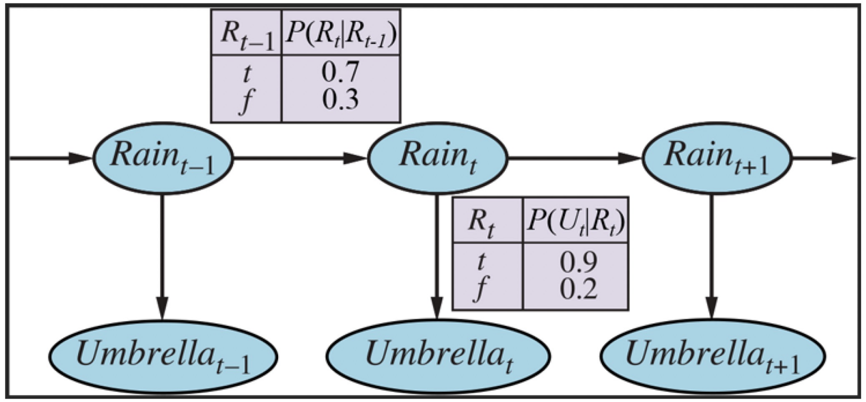
- Having set up the structure of a generic temporal model, we can formulate the basic inference tasks that must be solved.

- **Filtering** or **state estimation** is the task of computing the belief state $P(X_t|e_{1:t})$, the posterior distribution over the most recent state given all evidence to date.

- In the umbrella example, this would mean computing the probability of rain today, given all the umbrella observations made so far.

- Filtering is what a rational agent does to keep track of the current state so that rational decisions can be made.

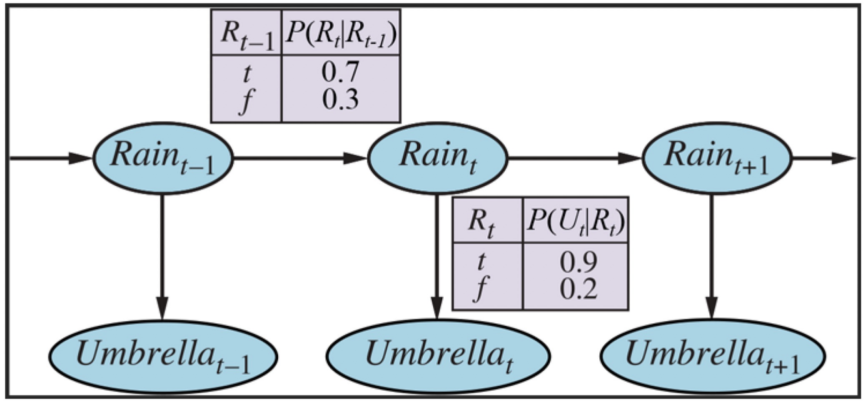
Inference in temporal models



$P(R_0)$

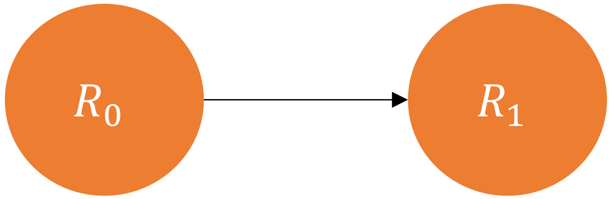
Rain	¬Rain
0.5	0.5

Inference in temporal models



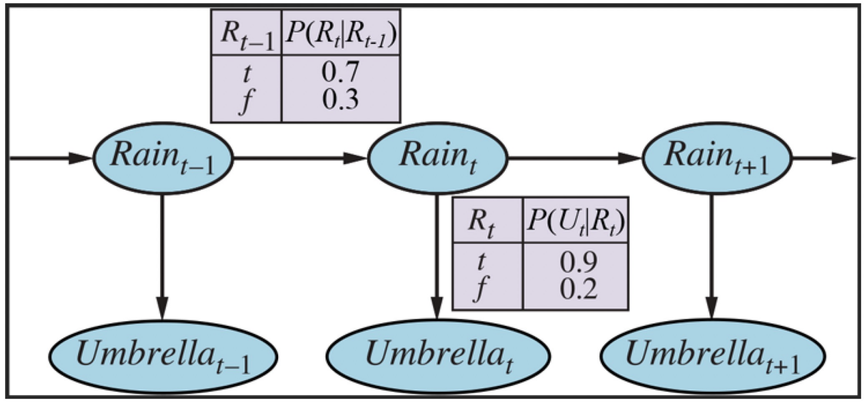
Rain	¬Rain
0.5	0.5

Day 1



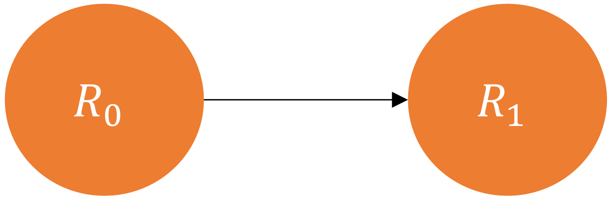
$P(R_1) = ?$

Inference in temporal models



Rain	¬Rain
0.5	0.5

Day 1



$P(R_1) =$

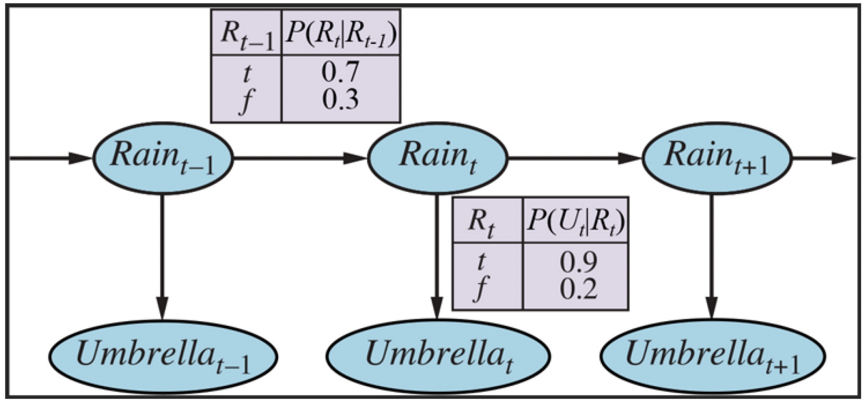
$$\sum_{R_0} P(R1|R_0)P(R_0) = P(R1|r_0)P(r_0) + P(R_1|\neg r_0)P(R_1|\neg r_0)$$

$$= < 0.7, 0.3 > \times 0.5 + < 0.3, 0.7 > \times 0.5$$

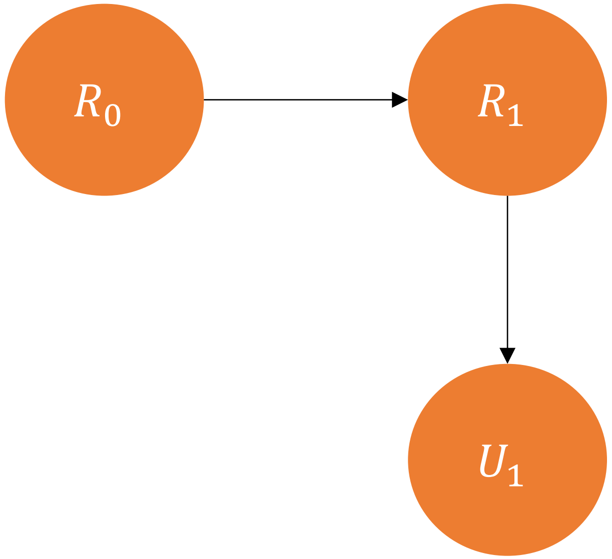
$$= < 0.5, 0.5 >$$

Inference in temporal models

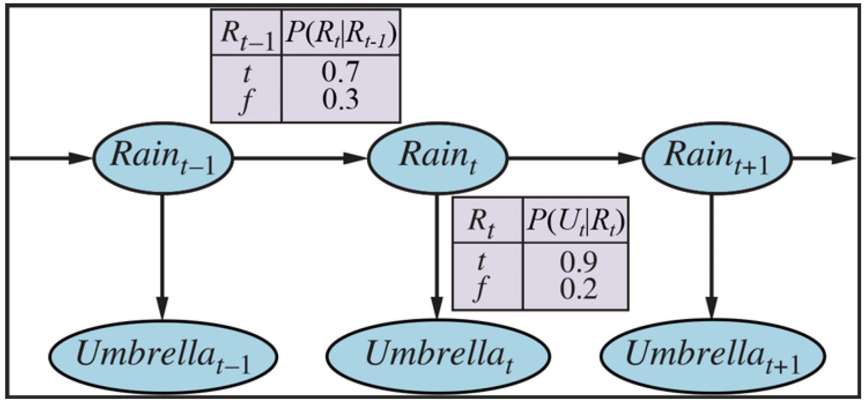
Day 1



Rain	¬Rain
0.5	0.5

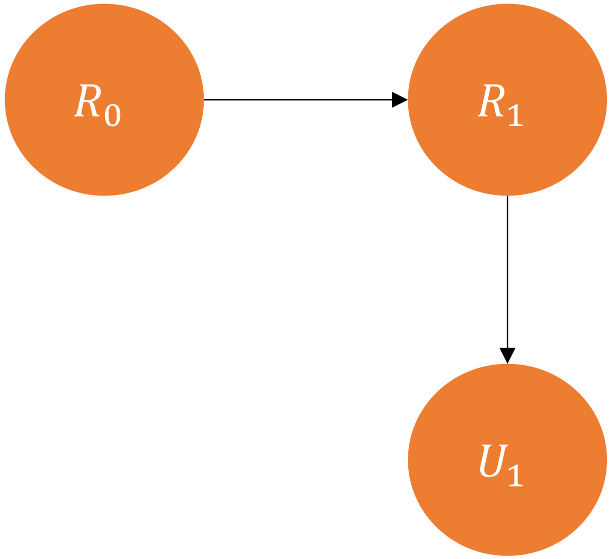


Inference in temporal models



Day 1

Rain	¬Rain
0.5	0.5

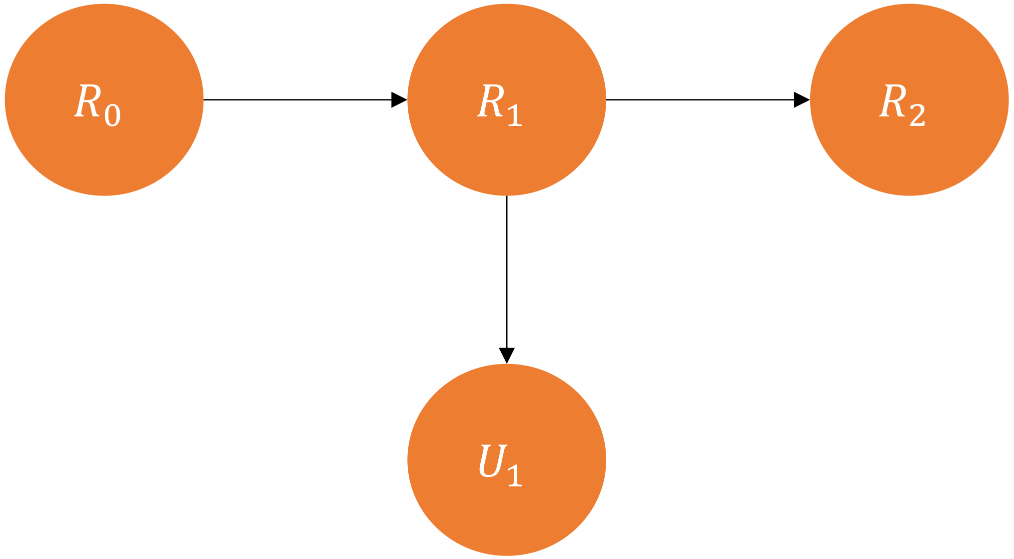


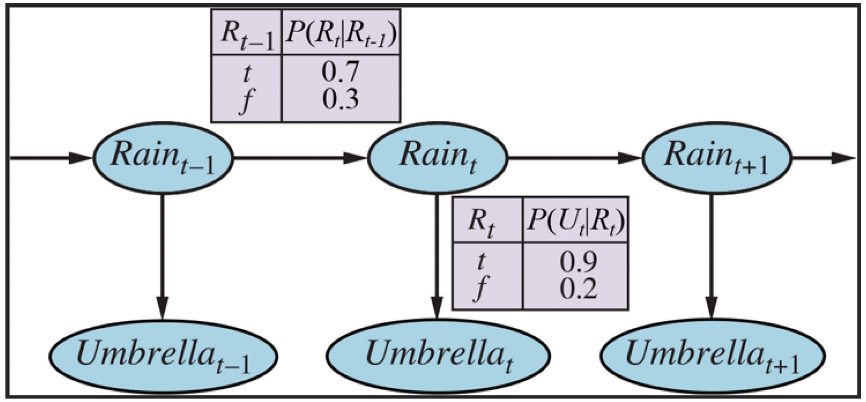
$$P(R_1|u_1) = \alpha\, P(u_1|R_1)P(R_1) = \alpha < 0.9,0.2 >< 0.5,0.5 >$$

$$\alpha < 0.9,0.2 >< 0.5,0.5 > = \alpha <0.45,0.1> \approx < 0.818,0.182 >$$

Inference in temporal models

Day 1



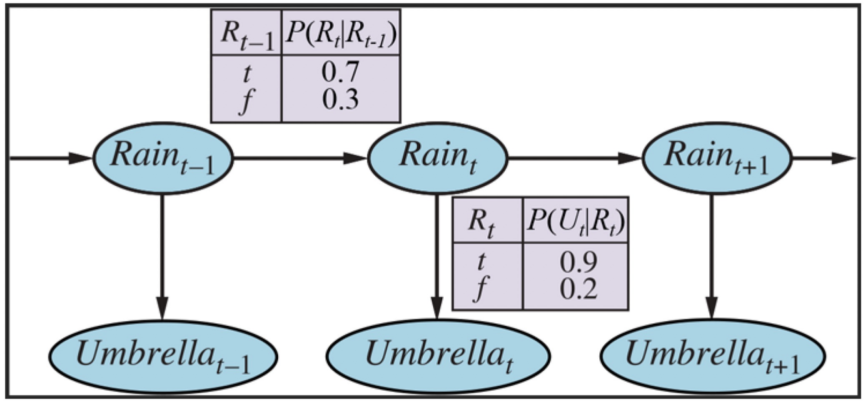


Rain	¬Rain
0.5	0.5
0.818	0.182

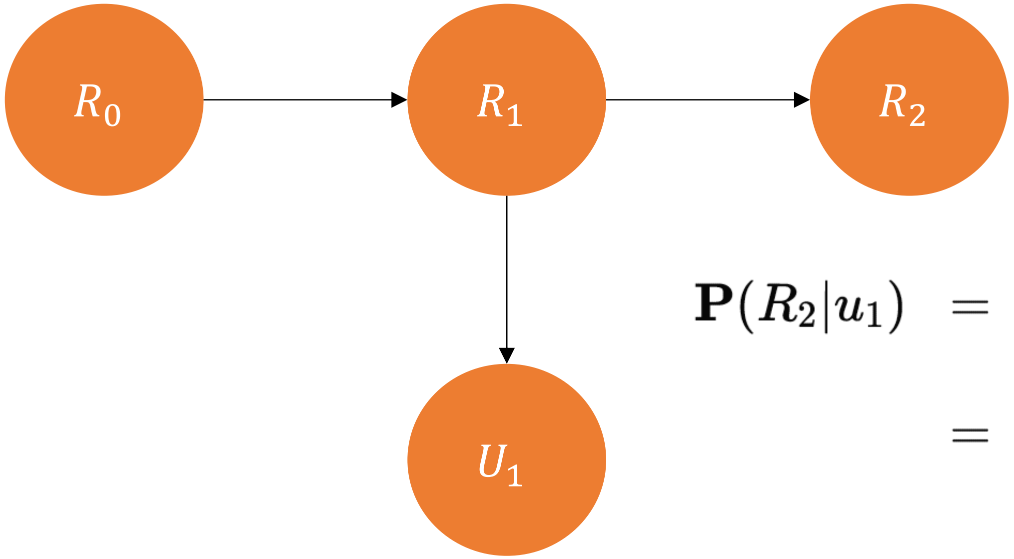
UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

Inference in temporal models

Day 1

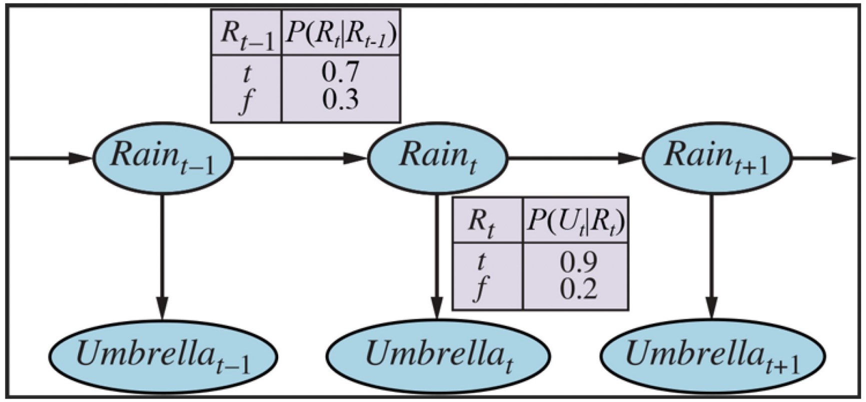


Rain	¬Rain
0.5	0.5
0.818	0.182



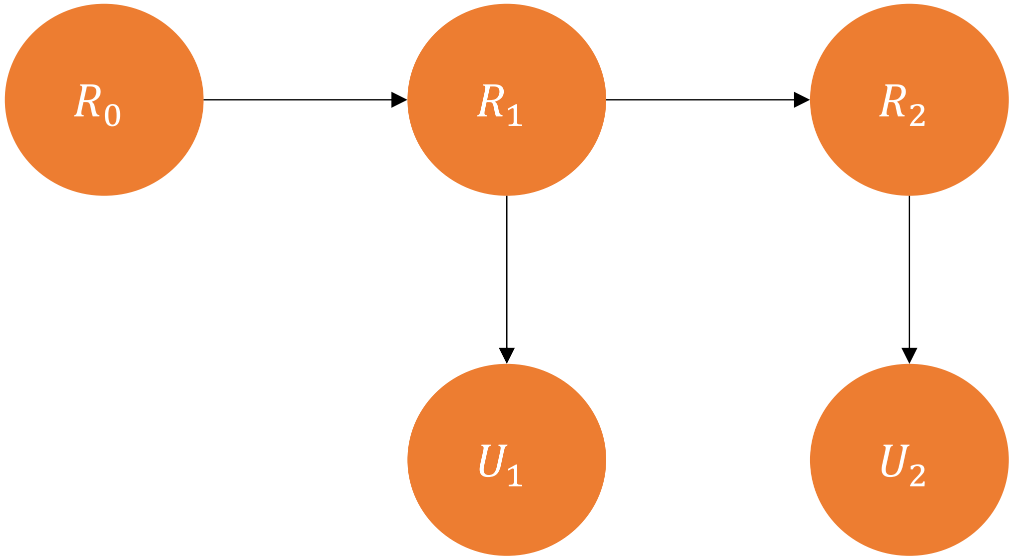
$$\begin{aligned}\mathbf{P}(R_2|u_1) &= \sum_{r_1} \mathbf{P}(R_2|r_1)P(r_1|u_1) \\ &= \langle 0.7, 0.3\rangle \times 0.818 + \langle 0.3, 0.7\rangle \times 0.182 \approx \langle 0.627, 0.373\rangle\end{aligned}$$

Inference in temporal models



Day 1

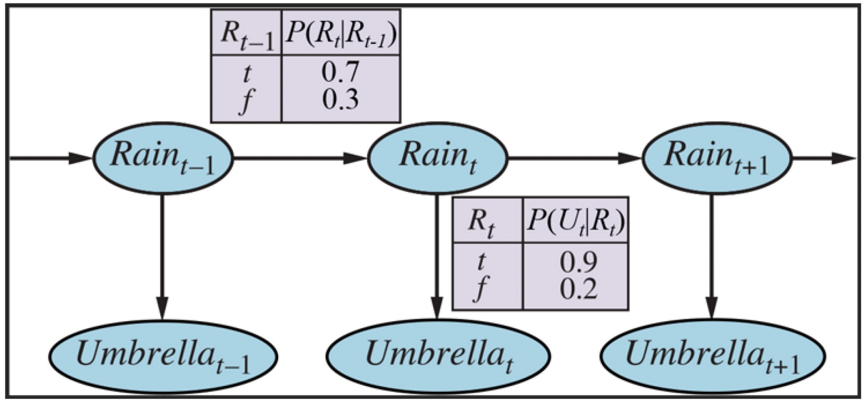
Day 2



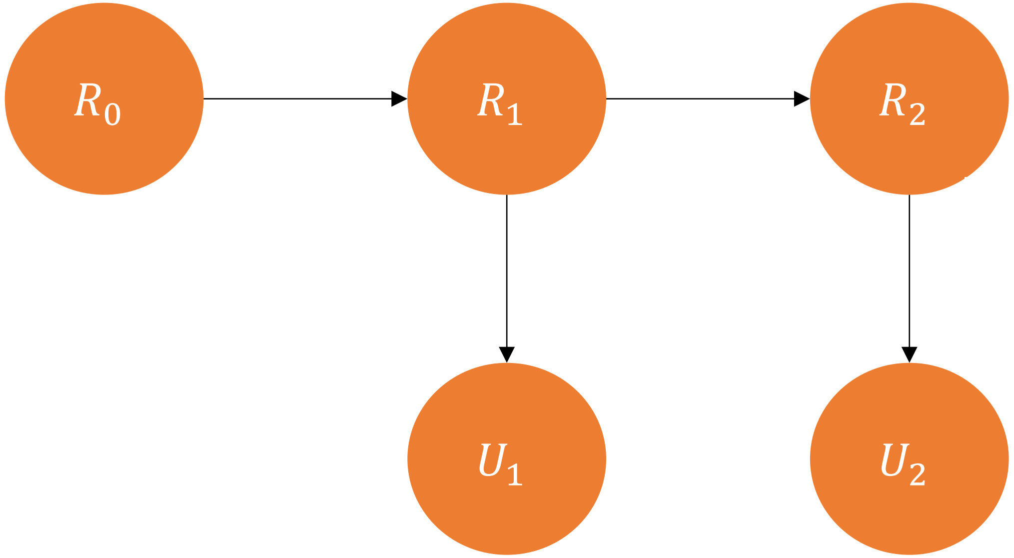
Rain	¬Rain
0.5	0.5
0.818	0.182

Inference in temporal models

Day 1 Day 2

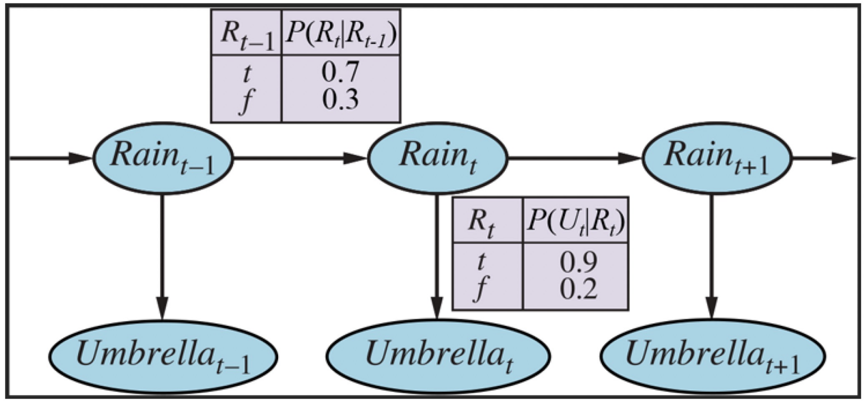


Rain	¬Rain
0.5	0.5
0.818	0.182



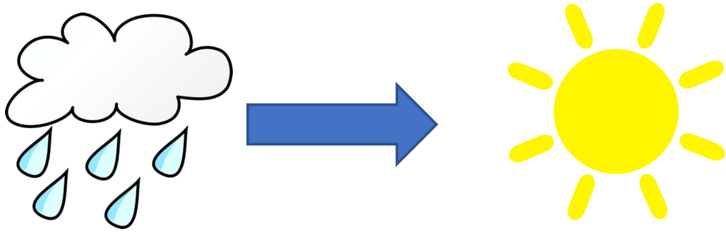
$$\begin{aligned}\mathbf{P}(R_2|u_1,u_2) &= \alpha\mathbf{P}(u_2|R_2)\mathbf{P}(R_2|u_1) = \alpha\,\langle 0.9, 0.2\rangle\,\langle 0.627, 0.373\rangle \\ &= \alpha\,\langle 0.565, 0.075\rangle \approx \langle 0.883, 0.117\rangle\,.\end{aligned}$$

Inference in temporal models



Transition model: $P(X_t|X_{t-1})$

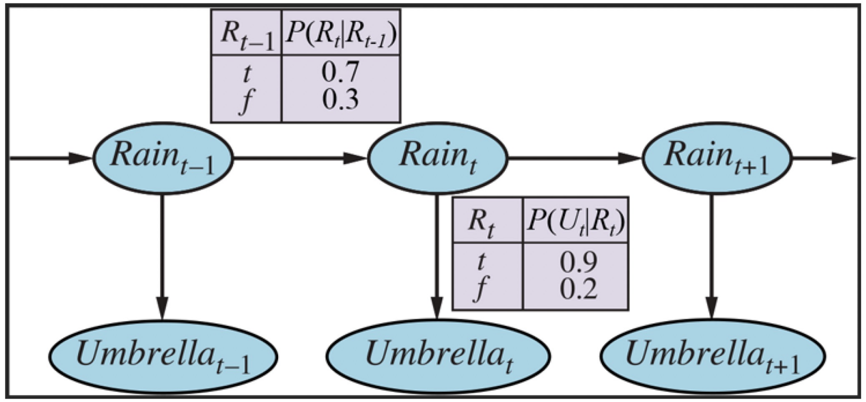
$$T_{ij} = P(X_t = j|X_{t-1} = i)$$



$$T = P(X_t|X_{t-1}) = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}$$

UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

Inference in temporal models

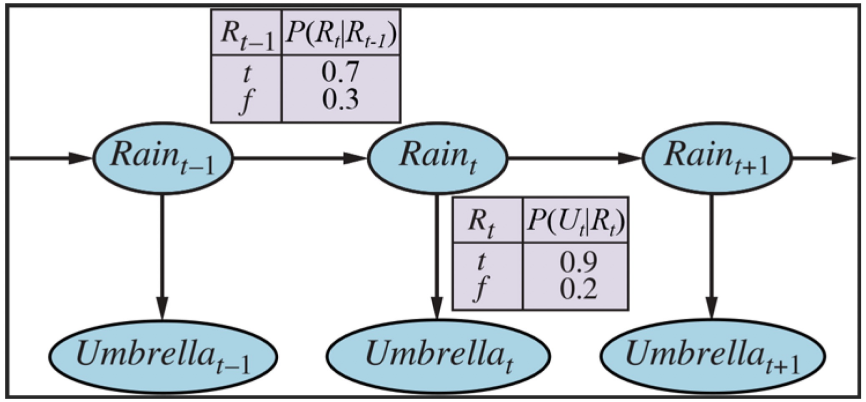


Sensor model: $P(e_t|X_t = i)$

For mathematical convenience, we place these values into an $S \times S$ diagonal observation matrix O_t, one for each time step.

The ith diagonal entry of O_t is $P(e_t|X_t = i)$ and the other entries are 0.

Inference in temporal models

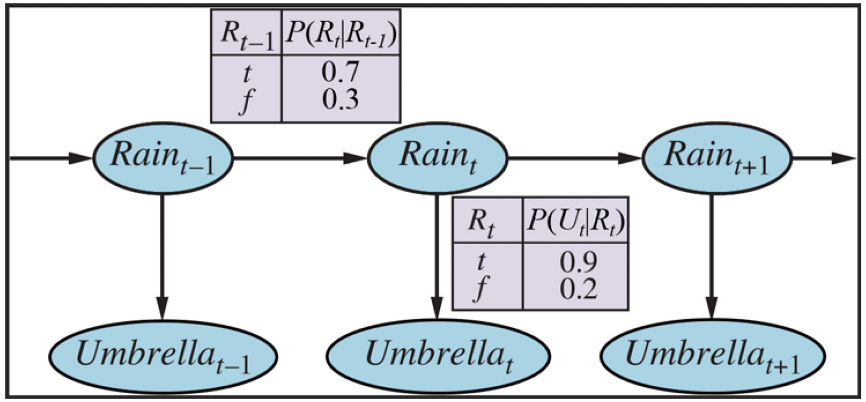


Sensor model: $P(e_t|X_t = i)$

The ith diagonal entry of O_t is $P(e_t|X_t = i)$ and the other entries are 0.

$$\mathbf{O}_1 = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.2 \end{pmatrix}; \qquad \mathbf{O}_3 = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.8 \end{pmatrix}.$$

Inference in temporal models

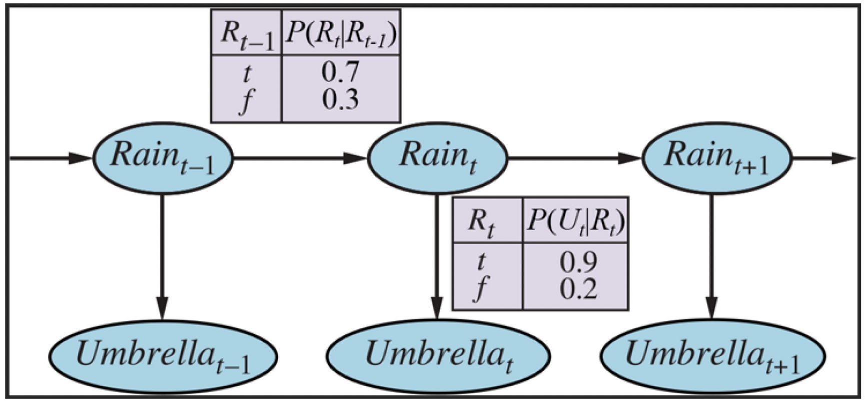


All the computations become simple matrix–vector operations.

$$\mathbf{f}_{1:t+1} = \alpha \mathbf{O}_{t+1} \mathbf{T}^{\top} \mathbf{f}_{1:t}$$

UAA College of Engineering
UNIVERSITY of ALASKA ANCHORAGE

Inference in temporal models



All the computations become simple matrix–vector operations.

$$T = P(X_t|X_{t-1}) = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix} \qquad \mathbf{O}_1 = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.2 \end{pmatrix};$$

$$\mathbf{f}_{1:t+1} = \alpha \mathbf{O}_{t+1} \mathbf{T}^\top \mathbf{f}_{1:t}$$

Prediction

- This is the task of computing the posterior distribution over the future state, given all evidence to date.

$P(X_{t+k}|e_{1:t})$, k>0

This might mean computing the probability of rain three days from now, given all the observations to date. Prediction is useful for evaluating possible courses of action based on their expected outcomes.

Smoothing

- This is the task of computing the posterior distribution over a past state, given all evidence up to the present.

  $P(X_k|e_{1:t}),\ 0 \leq k < t$

- In the umbrella example, it might mean computing the probability that it rained last Wednesday, given all the observations of the umbrella carrier made up to today.

- Smoothing provides a better estimate of the state at time *k* than was available at that time, because it incorporates more evidence.

Most likely explanation

- Given a sequence of observations, we might wish to find the sequence of states that is most likely to have generated those observations.

$$\underset{x_{1:t}}{\arg\max}\, P(x_{1:t}|e_{1:t})$$

- For example, if the umbrella appears on each of the first three days and is absent on the fourth, then the most likely explanation is that it rained on the first three days and did not rain on the fourth.

- Algorithms for this task are useful in many applications, including speech recognition—where the aim is to find the most likely sequence of words, given a series of sounds—and the reconstruction of bit strings transmitted over a noisy channel.

Learning

- In addition to these inference tasks, we also have learning

Recap

- The changing state of the world is handled by using a set of random variables to represent the state at each point in time.
- Representations can be designed to (roughly) satisfy the Markov property, so that the future is independent of the past given the present. Combined with the assumption that the process is time-homogeneous, this greatly simplifies the representation.
- A temporal probability model can be thought of as containing a transition model describing the state evolution and a sensor model describing the observation process.
- The principal inference tasks in temporal models are filtering (state estimation), prediction, smoothing, and computing the most likely explanation. Each of these tasks can be achieved using simple, recursive algorithms whose run time is linear in the length of the sequence.