T-79.5102 **Autumn 2006**
Special Course in Computational Logic
Tutorial 9

1. Sometimes MDPs are formulated with a reward function $R(s, a)$ that depends on the action taken or a reward function $R(s, a, s')$ that also depends on the outcome state.

(a) Write the Bellman equations for these formulations.

(b) Show how an MDP with reward function $R(s, a, s')$ can be transformed into a different MDP with reward function $R(s, a)$ such that optimal policies in the new MDP correspond exactly to optimal policies in the original MDP.

(c) Now do the same to convert MDPs with $R(s, a)$ into MDP with $R(s)$.

(R&N, Exercise 17.5)

2. There is a cat sleeping in the kitchen. There are two doors on the way out from the kitchen: one leading from the kitchen to the hall and then the front door. The states of the doors at time t is modelled in terms of two Boolean variables:

$$\begin{aligned} F_t &= \text{"the front door is open at time step } t\text{" and} \\ K_t &= \text{"the kitchen door is open at time step } t\text{".} \end{aligned}$$

The owner of the cat cannot directly observe the doors. However, the kitchen door is a fire door and has an alarm bell attached to it. On the other hand, it gets easily windy inside if the front door is open. These pieces of evidence are modelled in terms of two further variables.

$$\begin{aligned} A_t &= \text{"the alarm bell is ringing at time step } t\text{" and} \\ W_t &= \text{"it is windy inside at time step } t\text{".} \end{aligned}$$

The behaviour of the system is governed by the following parameters:

Transition model:	Sensor model:
$P(f_{t+1} \mid f_t) = 0.6$	$P(w_t \mid f_t) = 0.9$
$P(f_{t+1} \mid \neg f_t) = 0.1$	$P(w_t \mid \neg f_t) = 0.1$
$P(k_{t+1} \mid k_t) = 0.6$	$P(a_t \mid k_t) = 0.7$
$P(k_{t+1} \mid \neg k_t) = 0.3$	$P(a_t \mid \neg k_t) = 0.1$

The owner observes both wind inside and the alarm bell ringing at time step $t = 1$ but not at time step $t = 2$ when the cat wakes up and mews. With what probability the cat is able to escape?

The initial beliefs of the owner are based on the prior probabilities

$$P(f_0) = 0.1 \text{ and } P(k_0) = 0.3.$$