Sample Solution Outline

Question Q1

(a)

```
froSearch([[Node|T]|_], Path):- goal(Node), reverse([Node|T],Path).
froSearch([[Node|T]|More],Path):-
    findall([Next,Node|T],arc(Node,Next),Children),
    add2frontier(Children,More,New),
    froSearch(New,Path).
```

(b) (i)

```
cost([H],0).
cost([A,B|T],C) :- arcCost(A,B,C0), cost([B|T], C1), C is C0+C1.
```

(ii) True.

(iii) cost(arc) = -1 and h = 0 OR *cost*(arc) = 0 and h(node)= 1/(1+depth(node))

(iv) False if arc costs vary. True, if arcs all have same positive cost (and finite branching).

(v) False: sufficient with 2 more conditions but not necessary

(c) (i) Only (A) — arc2([b,c],[c],KB) and arc3([b,c],[c],KB) but not arc([b,c],[c],KB) for KB described in question

(ii) False: loop on KB = [[a,b],[b,a]]

(iii) True: can only prove heads, and no need to prove them twice

(d) The explosion in the search space is offset by a search tree with depth = n, number of variables, and branching factor = max(s_i), allowing variable instantiation to be interleaved with constraint satisfaction and backtracking to avoid multiple fruitless searches.

Question Q2

(a) What to do at a state — i.e. to define a policy $\pi : S \to A$ in order to maximize reward

(b) Yes, if $p(s, a, s') > 0$. Since $\sum_{s'} p(s, a, s') = 1$, there must be some s' for which $p(s, a, s') > 0$. Reward specified by r may vary if it is construed as expected value (over a variety of transitions).

(c) choose a to maximize expected immediate reward

$$\arg\max_a \sum_{s'} p(s, a, s') r(s, a, s')$$

(d) (i) $q_0(s, a) = \sum_{s'} p(s, a, s') r(s, a, s')$ which is 0 for $s' \neq (3, 3), (2, 2)$

(ii) Recall that for $s_n = \sum_{i=0}^{n} b^i$,

$$s_{n+1} = 1 + bs_n = s_n + b^{n+1} \implies s_n = \frac{1 - b^{n+1}}{1 - b}$$

so

$$\begin{aligned} k \sum_{i=0}^{n} b^i &= k \frac{1 - b^{n+1}}{1 - b} \\ &= 10k(1 - .9^{n+1}) \quad \text{for } b = .9 \end{aligned}$$

turning ± 1 into ± 10 with $\gamma = .9$

$$\begin{aligned} q_0((2, 2), a) &= -1 \quad \text{immediate reward} \\ q_{n+1}((2, 2), a) &= -1 + .9Q_n((2, 2), a) \quad [= -1.9 \text{ for } n = 0] \\ &= -10(1 - .9^{n+2}) \quad [= -10(.19) = -1.9 \text{ for } n = 0] \quad k = -1 \\ &\to -10 \text{ as } n \to \infty \\ q_0((3, 3), a) &= 1 \quad \text{immediate reward} \\ q_{n+1}((3, 3), a) &= 1 + .9Q_n((3, 3), a) \quad [= 1.9 \text{ for } n = 0] \\ &= 10(1 - .9^{n+2}) \quad [= 10(.19) = 1.9 \text{ for } n = 0] \quad k = 1 \\ &\to 10 \text{ as } n \to \infty \end{aligned}$$

For limit $n \to \infty$, use eqns

$$q = -1 + .9q \text{ and } q' = 1 + .9q'$$

for

$$q = -10 \text{ and } q' = 10$$

(iii) $q_n((2,3)$,left) and Q_n((2,3),down) for every integer $n \geq 0$

$$q_n((2,3), \text{left}) = q_n((2,2), a) \rightarrow -10 \quad (-1 \text{ for } n = 0)$$
$$q_n((2,3), \text{down}) = q_n((3,3), a) \rightarrow 10 \quad (1 \text{ for } n = 0)$$

(e) What they have in common are outputs in the unit interval [0,1]. They differ in their inputs: subset S of sample space Ω for μ; proposition α for P, and value x of a variable X for the probabily distribution P_X. Note that

$$P_X(x) = P(X = x) = \mu(\{\omega \in \Omega \mid \omega \models X = x\})$$