A) 2D problems

Problem 1
There are several solutions but they all have in common the same policy in part of the states:
*The “o“ sign mean there are several choices for an optimal policy.*

Problem 2
The *generic* policy is slightly different but there are still several optimal policies:

Problem 3
The optimal policies can be quite different from each other:
Problem 4
The following policy can be mirrored or rotated because of the symmetries of the map.

The “o” sign means there are several possibilities.

Is the policy different if we limit the maximum number of steps?

- If we limit for instance the number of steps to 10, some of the “o”s will have to be replaced by an arrow in direction of the external ring or the internal ring depending on where they stand: for instance if an agent starts in 1 (see figures below) and follows the previous policy plus a down arrow on the starting state, it will collect 7 rewards in the 10 steps allowed (on the left). However, if the agent follows the same policy with a right arrow on the starting state, it will collect 8 rewards (on the right).

Problem 5
If all states have a reward, this policy will allow cumulating all rewards in the smallest amount of steps, whatever the initial state is.