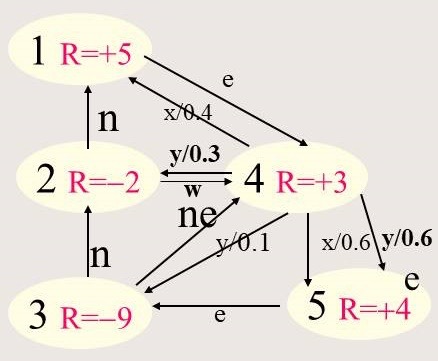
**3) Reinforcement Learning**

Consider the following World called ABC is given:



1. Give the Bellman equations for all 5 states! Assume γ=0.5! Propose a policy (using your own intelligence and not reinforcement learning) which maximizes an agent’s reward for the ABC world[[1]](#footnote-2)!

To be presented on Mo., March 1 by Group E

1. Now we apply temporal difference learning, assuming the agent starts in state 3 and applies the operator sequence **ne-y(ending up in state 3)-ne-y(ending up ins state 3)**; what are the final utilities of state 3 and 4—give each step of the computation? Assume the initial utilities are 0; also assume α=0.5 and γ=0.5!

To be presented on We., Match 3 by Group F

1. Write a program which solves the Bellman Equation System for the ABC world—the equations were presented by Group E earlier— assuming γ=0.5—and report your findings. Interpret the utilities for the 5 states you obtained. Compare you findings with the strategies suggested by group E!

To be presented on Mo., March 8 by Group G

Group E:

|  |  |
| --- | --- |
| David | Heller |
| Josue | Hernandez |
| Esai | Hernandez |
| Jake | Higdon |
| Brian | High |
| Sham | Hintolay |

Group F:

|  |  |
| --- | --- |
| Jonathan | Hirsch |
| David | Hoang |
| Bach | Hoang |
| Tyler | Hu |
| Henderson | Hua |
| Junren | Huang |

Group G:

|  |  |
| --- | --- |
| Emily | Humble |
| Nadiia | Hutcherson |
| Abdulkadir | Jaffar |
| Jarrod | Johnson |
| William | Lam |
| Karla | Lemus |

1. This might be challenging! [↑](#footnote-ref-2)