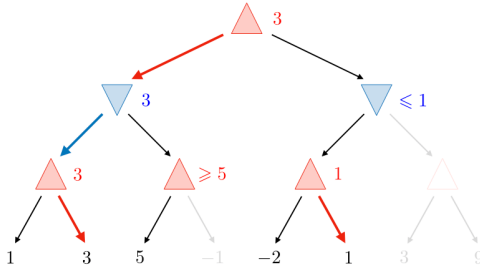


### 2.3.1 Accélération de minimax

□ **Fonction d'évaluation** – Une fonction d'évaluation estime de manière approximative la valeur  $V_{\text{minimax}}(s)$  selon les paramètres du problème. Elle est notée  $\text{Eval}(s)$ .

*Remarque : l'analogue de cette fonction utilisé dans les problèmes de recherche est  $\text{FutureCost}(s)$ .*

□ **Élagage alpha-bêta** – L'élagage alpha-bêta (en anglais *alpha-beta pruning*) est une méthode exacte d'optimisation employée sur l'algorithme de minimax et a pour but d'éviter l'exploration de parties inutiles de l'arbre de jeu. Pour ce faire, chaque joueur garde en mémoire la meilleure valeur qu'il puisse espérer (appelée  $\alpha$  chez le joueur maximisant et  $\beta$  chez le joueur minimisant). À une étape donnée, la condition  $\beta < \alpha$  signifie que le chemin optimal ne peut pas passer par la branche actuelle puisque le joueur qui précédait avait une meilleure option à sa disposition.



□ **TD learning** – L'apprentissage par différence de temps (en anglais *temporal difference learning* ou *TD learning*) est une méthode utilisée lorsque l'on ne connaît pas les transitions/récompenses. La valeur est alors basée sur la politique d'exploration. Pour pouvoir l'utiliser, on a besoin de connaître les règles du jeu  $\text{Succ}(s,a)$ . Pour chaque  $(s,a,r,s')$ , la mise à jour des coefficients est faite de la manière suivante :

$$w \leftarrow w - \eta [V(s,w) - (r + \gamma V(s',w))] \nabla_w V(s,w)$$

### 2.3.2 Jeux simultanés

Ce cas est opposé aux jeux joués tour à tour. Il n'y a pas d'ordre prédéterminé sur le mouvement du joueur.

□ **Jeu simultané à un mouvement** – Soient deux joueurs  $A$  et  $B$ , munis de possibles actions. On note  $V(a,b)$  l'utilité de  $A$  si  $A$  choisit l'action  $a$  et  $B$  l'action  $b$ .  $V$  est appelée la matrice de profit (en anglais *payoff matrix*).

□ **Stratégies** – Il y a principalement deux types de stratégies :

- Une stratégie pure est une seule action :

$$a \in \text{Actions}$$

- Une stratégie mixte est une loi de probabilité sur les actions :

$$\forall a \in \text{Actions}, \quad 0 \leq \pi(a) \leq 1$$

□ **Évaluation de jeu** – La valeur d'un jeu  $V(\pi_A, \pi_B)$  quand le joueur  $A$  suit  $\pi_A$  et le joueur  $B$  suit  $\pi_B$  est telle que :