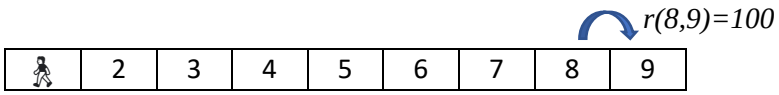
2.5 Mise en place du facteur d'amortissement

La formule précédente est correcte mais elle ne nous permettra pas de résoudre le problème du *Q*-Learning. Nous allons utiliser une extension de cette formule utilisant un facteur d'amortissement γ compris dans l'intervalle]0,1[:

$$Q_{e,a} = \sum_{e' \in E} p(e'|e,a).\left[\bar{r}(e,a,e') + \gamma.\max_{a'} Q_{e',a'}\right] \quad (1)$$

Son intérêt n'est pas évident au premier abord. Prenons un exemple d'un jeu avec transition déterministe. Le personnage se trouve dans un couloir, il peut se déplacer vers la droite ou vers la gauche. Le trésor se trouve à droite et il commence son aventure à gauche. Les valeurs des récompenses sont nulles, sauf lorsqu'il atteint la dernière case où se trouve le trésor :



Comme les transitions sont déterministes, on a $p(9|8, droite) = 1$. Comme la case 9 termine la partie, on obtient en appliquant la formule (1) :

$$Q_{8,d} = \bar{r}(8,d,9) = 100$$

On peut calculer les autres valeurs de Q. Pour cela, il suffit de remarquer que les autres récompenses sont nulles et que la politique $MaxQ$ nous fait sélectionner l'action nous amenant vers la droite. Ainsi :

$$Q_{7,d} = \gamma.Q_{8,d}$$

Et ensuite :

$$Q_{5,d} = \gamma.Q_{6,d} \qquad Q_{6,d} = \gamma.Q_{7,d}$$

Calculons les valeurs $Q_{i,d}$ associée à un déplacement vers la droite depuis la i-ème case du labyrinthe, avec γ = 0.9, nous obtenons :

$Q_{1,d}$	$Q_{2,d}$	$Q_{3,d}$	$Q_{4,d}$	$Q_{5,d}$	$Q_{6,d}$	$Q_{7,d}$	$Q_{8,d}$
47,8	53,1	59,1	65,6	72,9	81	90	100

La situation est identique à une stratégie utilisant une carte des distances. Pour guider notre personnage vers la récompense, il suffit de choisir à chaque case une nouvelle case augmentant la valeur Q courante. C'est simple ! Cependant, si l'on prend une valeur γ égale à 1, nous aurions obtenu :

$Q_{1,d}$	$Q_{2,d}$	$Q_{3,d}$	$Q_{4,d}$	$Q_{5,d}$	$Q_{6,d}$	$Q_{7,d}$	$Q_{8,d}$
100	100	100	100	100	100	100	100

Ces valeurs ne permettent pas de mettre une stratégie en place : où que l'on aille les valeurs Q sont identiques. Le personnage navigue dans le brouillard, il est impossible de l'orienter.