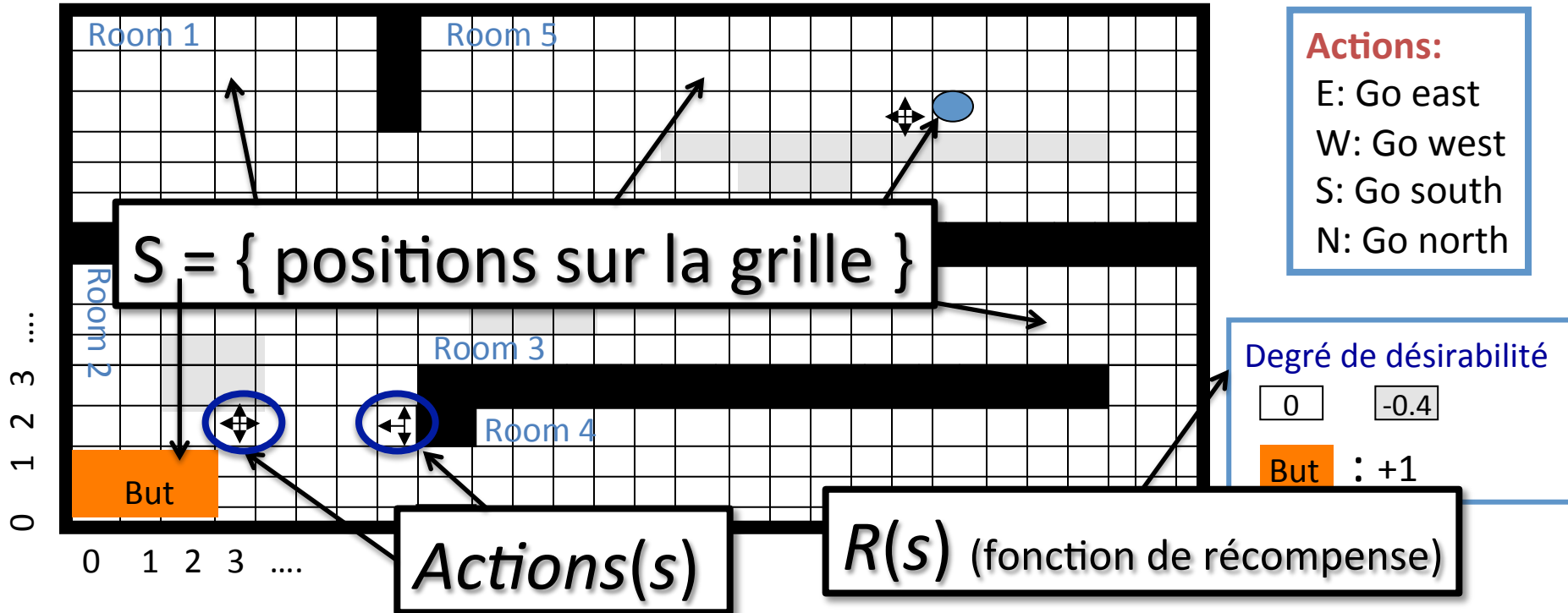


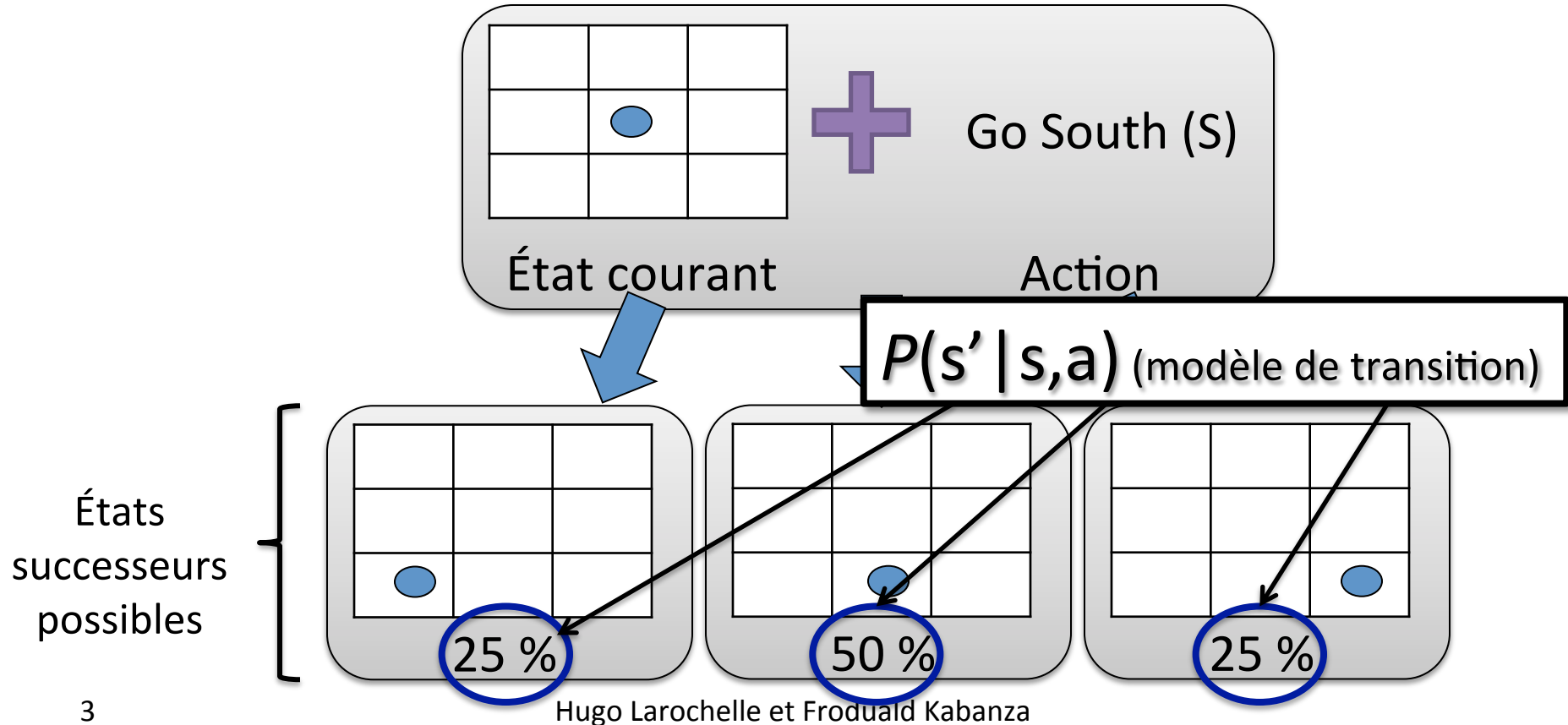
# Processus de décision markovien

- Un **processus de décision markovien** (*Markov decision process*, ou **MDP**) est défini par:
  - ◆ un **ensemble d'états**  $S$  (incluant un état initial  $s_0$ )
  - ◆ un **ensemble d'actions** possibles  $Actions(s)$  lorsque je me trouve à l'état  $s$
  - ◆ un **modèle de transition**  $P(s'|s, a)$ , où  $a \in A(s)$
  - ◆ une **fonction de récompense**  $R(s)$  (utilité d'être dans l'état  $s$ )
- Un MDP est un modèle général pour un environnement stochastique dans lequel un agent peut prendre des décisions et reçoit des récompenses
  - ◆ on y fait une supposition markovienne (de premier ordre) sur la distribution des états visités
  - ◆ requière qu'on décrive un objectif à atteindre à partir d'une fonction de récompense basée seulement sur l'état courant

# Grille (occupancy grid)

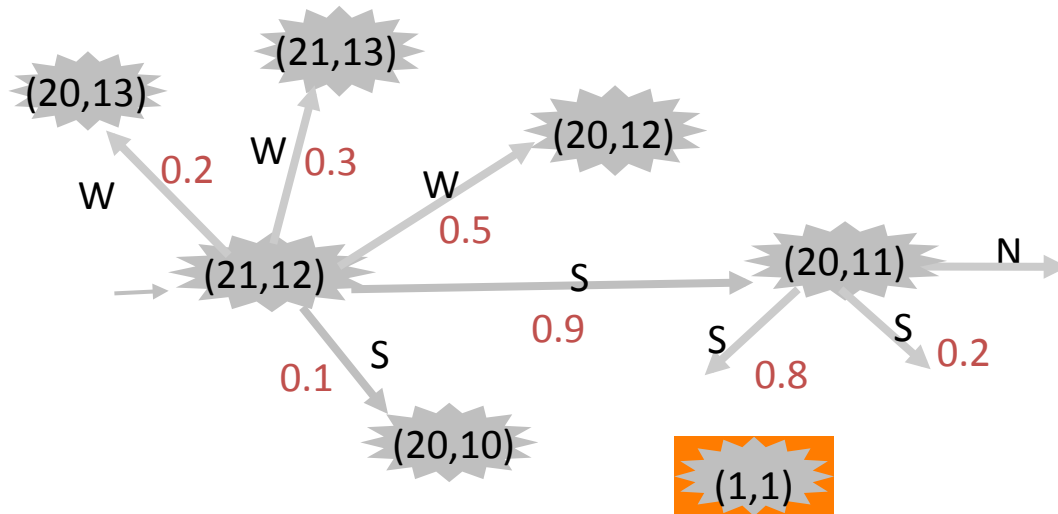


# Actions aux effets incertains



# Décision

- Une **décision** est un choix d'une action dans un état
  - ◆ c'est une règle « *if state then action* »



## Exemples:

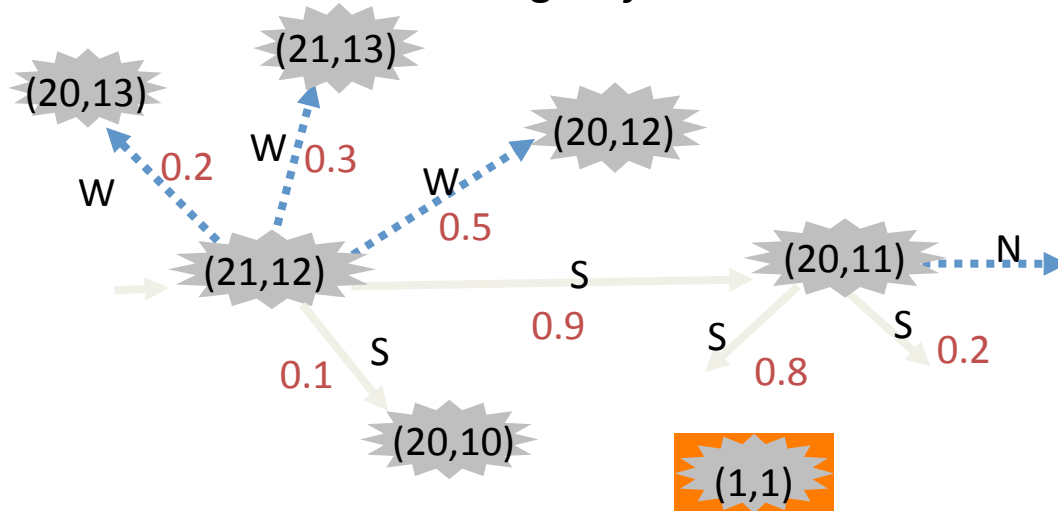
$(21,12) \rightarrow W$

ou

$(21,12) \rightarrow E$

# Plan (politique)

- Un **plan** est une stratégie: choix d'une action (décision) **pour chaque état**
  - ◆ un plan est également appelé une politique (*policy*)
  - ◆ c'est un **ensemble** de règles *if state then action*



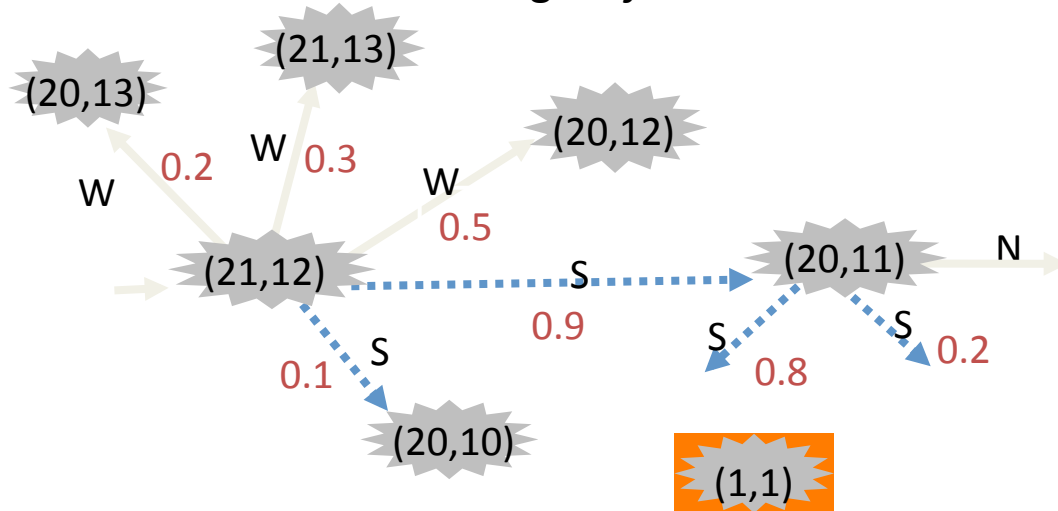
## Examples:

### Plan $\pi_1$

$\{ (21,12) \rightarrow W,$   
 $(20,13) \rightarrow S,$   
 $(21,13) \rightarrow S,$   
 $\dots$   
 $\}$

# Plan (politique)

- Un **plan** est une stratégie: choix d'une action (décision) **pour chaque état**
  - ◆ un plan est également appelé une politique (*policy*)
  - ◆ c'est un **ensemble** de règles *if state then action*



## Examples:

### Plan $\pi_1$

$\{ (21,12) \rightarrow W,$   
 $(20,13) \rightarrow S,$   
 $(21,13) \rightarrow S,$

...

}

### Plan $\pi_2$

$\{ (21,12) \rightarrow S,$   
 $(20,11) \rightarrow S,$   
 $(21,10) \rightarrow E,$   
.... }

# Exécution d'un plan (politique)

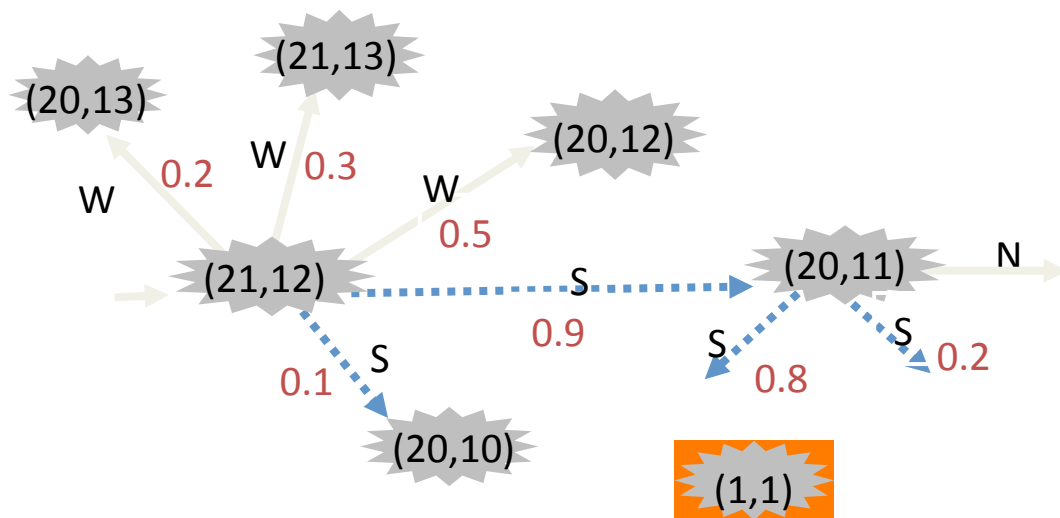
- Notons  $\pi(s)$  l'action désignée par le plan  $\pi$  dans l'état  $s$ 
  - ◆ voici un algorithme d'exécution ou d'application d'un plan

```
While (1)
{
  1.  $s$  = état courant de l'environnement;
  2.  $a = \pi(s)$ ;
  3. execute  $a$ ;
}
```

- L'étape 1 peut impliquer de la détection (*sensing*) et de la localisation
- L'état résultant de l'exécution de l'action à l'étape 3 est stochastique

# Interprétation/application d'un plan

- L'application d'un plan dans un MDP **résulte en une chaîne de Markov**
  - ◆ le modèle de transition est donné par  $P(s'|s, \pi(s))$



## Examples:

### Plan $\pi_1$

$\{ (21,12) \rightarrow W,$   
 $(20,13) \rightarrow S,$   
 $(21,13) \rightarrow S,$

...

}

### Plan $\pi_2$

$\{ (21,12) \rightarrow S,$   
 $(20,11) \rightarrow S,$   
 $(21,10) \rightarrow E,$   
.... }