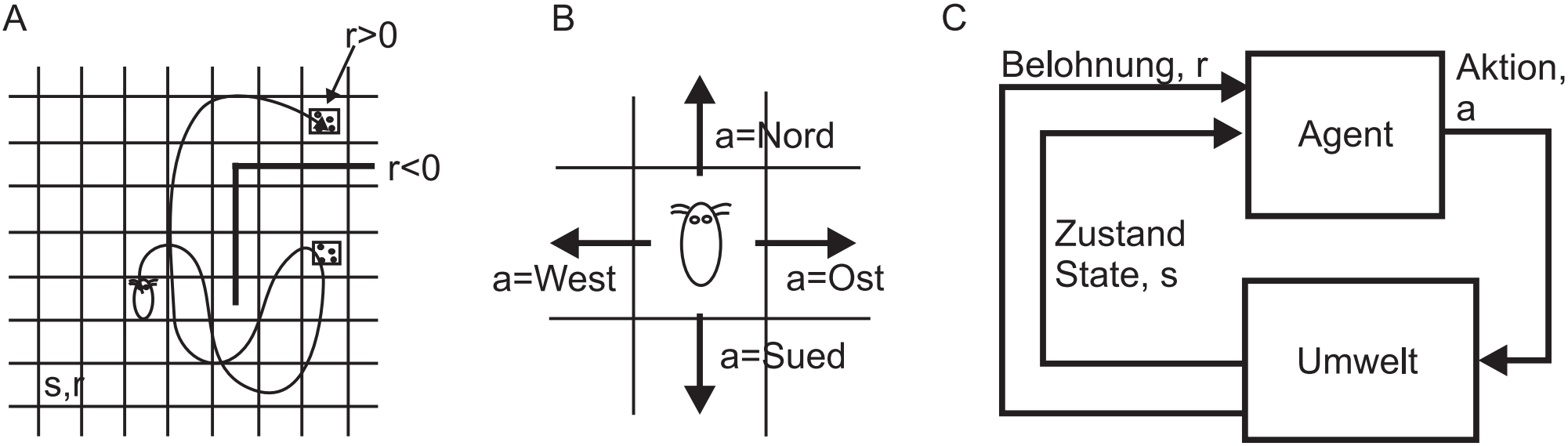
$\alpha = 0.5$

$Q(s, \mathrm{Ost})$

δ


A
r>0
r<0
s,r
B
a=Nord
a=West
a=Ost
a=Sued
C
Belohnung, r
Aktion,
a
Agent
Zustand
State, s
Umwelt

S

s_t

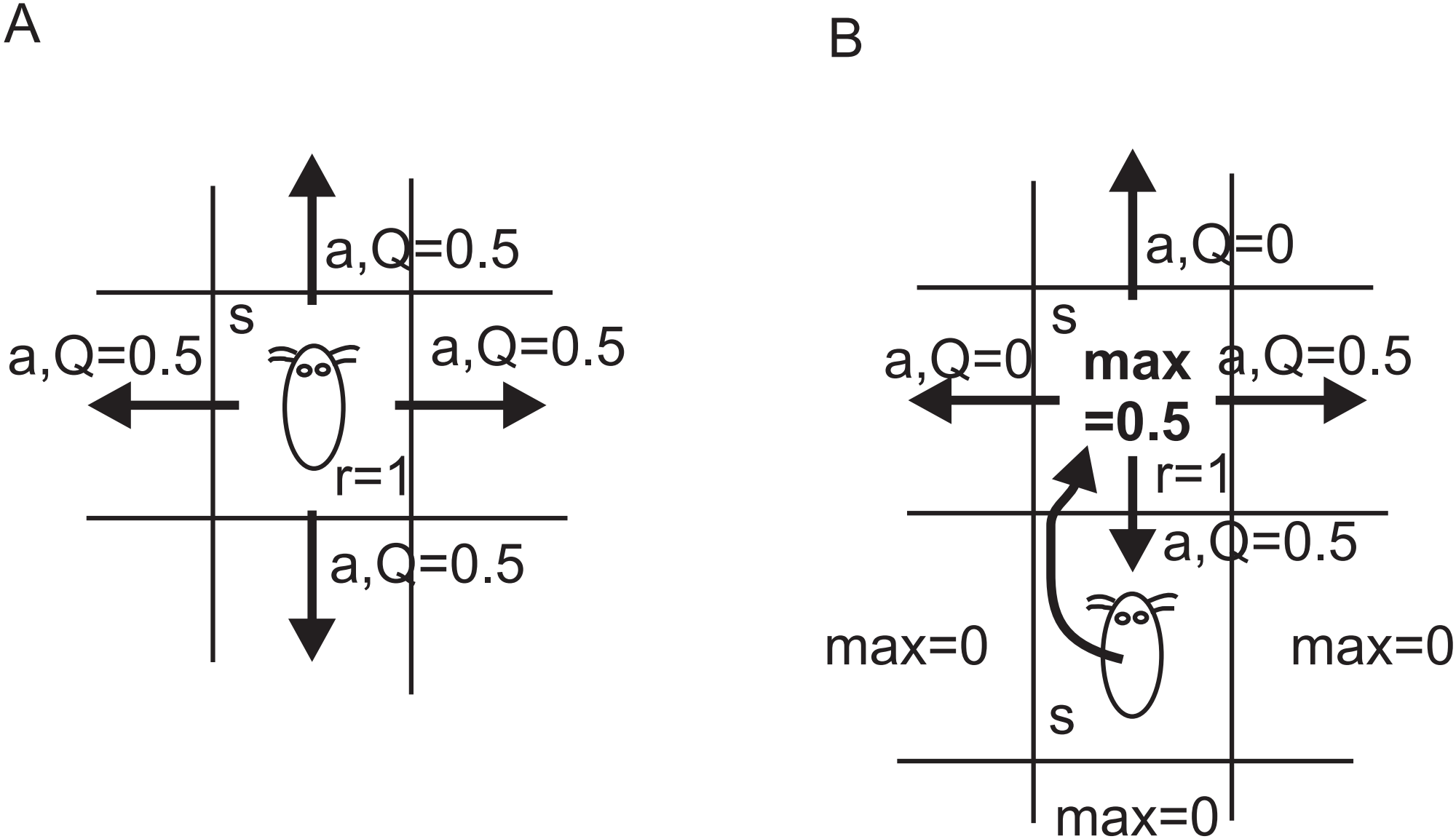
γ

$r > 0$

$r < 0$

$r = 0$

a


A
a,Q=0.5
s
a,Q=0.5
a,Q=0.5
r=1
a,Q=0.5
B
a,Q=0
s
a,Q=0
max
=0.5
a,Q=0.5
r=1
a,Q=0.5
max=0
max=0
s
max=0

$$Q(s,a) \leftarrow Q(s,a) + \alpha \underbrace{\left[r(s) + \gamma \max_{a'} Q(s',a') - Q(s,a) \right]}_{\delta(s,a)}$$

$0 < \gamma < 1$

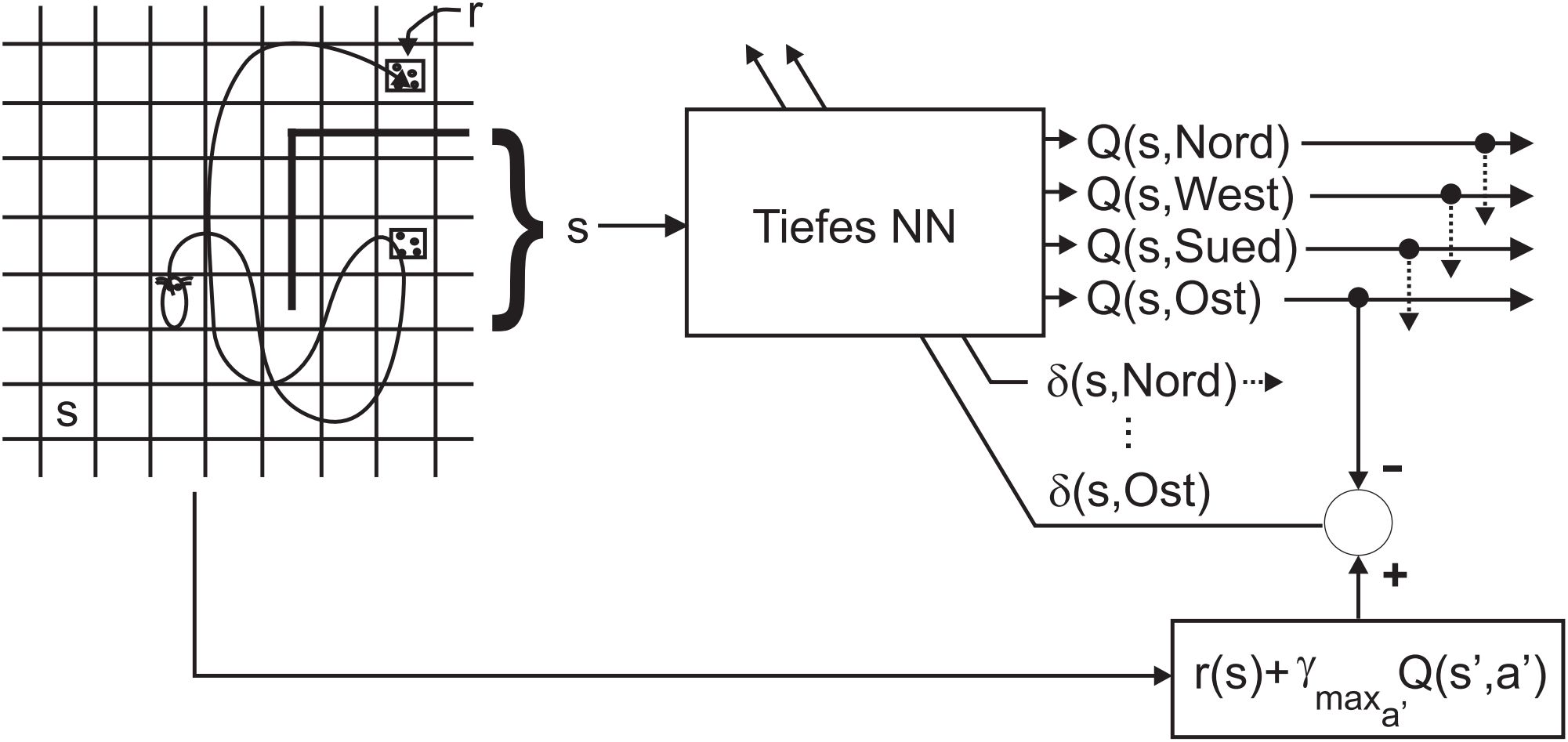
$\gamma = 1$

$\max_{a'} Q(s', a')$

$\delta(s, a)$

T

$$Q(s,a) = \sum_{t=0}^{T} \gamma^t r_t$$


r
s
s
Tiefes NN
Q(s,Nord)
Q(s,West)
Q(s,Sued)
Q(s,Ost)
δ(s,Nord)
δ(s,Ost)
-
+
r(s)+γ max a' Q(s',a')

$Q(s, \text{Nord}), Q(s, \text{Sued}), Q(s, \text{West}), Q(s, \text{Ost})$