

nicht ganz triviales Problem

Ausgangssituation:

	X	
0	X	

← Q-Spieler bekommt  
Reward für

State: 

0	X	

 Action | 1 | Value = 100

nächstes Spiel:

	X	

will diesen Wert und  
maximale Value erreichen

dafür muss next State generiert werden können:

next State könnte aber unterschiedlich aussehen:

0				
X				

0	X			

	X	0		

welchen nimmt man/oder alle?

⇒ Dann Generierung von 3x3 Actions und  
den maximalen Wert bestimmen also in diesem Fall 100.