强化学习：第二次作业

阅读《RL》前两章的内容，写出自己不理解的三个知识点或概念

问题一

在1.3节：强化学习的组成要素，是不是强化学习会受到策略空间的约束？因为策略空间的限制，强化学习无法解决“从无到有”的问题？比如写句子，强化学习只能根据策略空间中定义好的语法句式来生成句子，而无法产生独立于策略空间中的新型句式？

问题二

2.3节：10-摇臂测试工具中的图2.2描绘了ε-贪心动作值方法在10-摇臂测试工具上的平均表现。书中描述了：“ε=0.01 的方法提升得更为缓慢, 但最终会在图中的两种测度上比ε=0.1的方法表现得更好. 也可以随时间逐步减小”。ε=0时的平均奖赏明显小于上述两种情况，这也就说明并不是ε越小越好的，那到底应该怎么选择ε的值？

问题三

想知道图2.3和图2.4中出现尖峰的原因。