[1-123]

From **Reinforcement Learning and Episodic Memory in Humans and Animals- An Integrative Framework.pdf**

1. A David, R.J.S., *Addiction as a computational process gone awry.* 2004. **306**(5703): p. 1944-1947.

2. A Ross, O., et al., *The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive.* 2013. **24**(5): p. 751-761.

这些结果表明，通过调节认知资源的可用性，可以逐个试验地控制多个学习系统之间的竞争。

3. A Ross, O., et al., *Working-memory capacity protects model-based learning from stress.* 2013. **110**(52): p. 20941-20946.

这些结果丰富了对急性应激，工作记忆和前额叶功能之间相互作用的现有描述，并表明执行功能可能对急性应激的有害影响具有保护作用。

4. Adams, C.D.J.T.Q.J.o.E.P.S.B., *Variations in the sensitivity of instrumental responding to reinforcer devaluation.* 1982. **34**(2b): p. 77-98.

5. Alec, S. and M.M. Botvinick, %J Proceedings of the National Academy of Sciences of the United States of America, *Evidence integration in model-based tree search.* 2015. **112**(37): p. 11708.

6. Alexander, G.E. and M.D.J.T.i.n. Crutcher, *c.* 1990. **13**(7): p. 266-271.

7. Barron, G. and I.J.J.o.b.d.m. Erev, *Small feedback‐based decisions and their limited correspondence to description‐based decisions.* 2003. **16**(3): p. 215-233.

8. Bayer, H.M. and P.W.J.N. Glimcher, *Midbrain dopamine neurons encode a quantitative reward prediction error signal.* 2005. **47**(1): p. 129-141.

因此，中脑多巴胺神经元的激发率通过理论描述在强化学习模型中使用的奖励预测误差信号的定量预测，其中该信号具有正值。我们还发现多巴胺系统继续计算奖励预测误差，即使动物的行为政策仅受此计算的微弱影响。

9. Bellman, R.J.v.P.B.P., *Dynamic Programming, Princeton, NJ: Princeton Univ.* 1957.

10. Ben, S., et al., *Serotonin selectively modulates reward value in human decision-making.* 2012. **32**(17): p. 5833-42.

在人类决策中建立神经调节剂5-羟色胺的功能已经证明非常困难，因为它在奖惩处理中起着复杂的作用。我们的研究结果为血清素在奖励处理中的惊人作用提供了证据，同时说明了其复杂和多种多样的影响。

11. Bertsekas, D.P. and J.N. Tsitsiklis, *Neuro-dynamic programming*. Vol. 5. 1996: Athena Scientific Belmont, MA.

12. Bettman, J.R., *Information processing theory of consumer choice*. 1979: Addison-Wesley Pub. Co.

13. Biele, G., I. Erev, and E.J.J.o.M.P. Ert, *Learning, risk attitude and hot stoves in restless bandit problems.* 2009. **53**(3): p. 155-167.

14. Bornstein, A.M., et al., *Reminders of past choices bias decisions for reward in humans.* 2017. **8**: p. 15958.

我们提供的证据表明，通过为过去的个人经历提供记忆来做出决策，并且这个过程可能偏向于使用附带提醒来支持过去的选择。 首先，在标准的奖励选择任务中，我们表明，使用过去结果的个体样本在决策时间估计价值的模型比规范增量学习模型更适合选择和决策相关的神经活动。 在第二个实验中，我们通过偶然提醒参与者个人过去的决定来偏向这个抽样过程。 提醒后的下一个决定显示所采取行动的强烈影响以及收到的提醒试验价值。 这些结果为决策架构提供了新的实证支持，该架构依赖于过去个体选择事件的样本而不是评估选项中的渐进平均奖励，并且对潜在的认知和神经机制具有暗示性意义。

15. Botvinick, M.M., Y. Niv, and A.C.J.C. Barto, *Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective.* 2009. **113**(3): p. 262-280.

对人类和动物行为的研究长期以来一直强调其层次结构 - 将持续行为分解为离散任务的可分性，这些任务由子任务序列组成，而子任务序列又由简单行为构成。行为的等级结构在神经科学中也具有持久的兴趣，其被广泛认为反映了前额皮质功能。在本文中，我们从计算强化学习的最新发展的角度重新审视行为层次及其神经基础。具体而言，我们将一组方法统称为分层强化学习，通过允许学习代理将动作聚合为可重用的子例程或技能来扩展强化学习范例。仔细研究分层强化学习的组成部分，可以看出它们如何映射到神经结构，特别是背外侧和眼眶前额叶皮层内的区域。它还提出了分层强化学习可以为现有的分层结构行为心理模型提供补充的具体方法。分层强化学习带来的一个特别重要的问题是学习如何识别可能为解决各种未来问题提供有用构建块的新行动惯例。在这里和许多其他方面，分层强化学习提供了一个吸引人的框架，用于研究分层结构行为的计算和神经基础。

16. Braver, T.S., J.D.J.C.o.c.p.A. Cohen, and p. XVIII, *On the control of control: The role of dopamine in regulating prefrontal function and working memory.* 2000: p. 713-737.

我们提出了一个模拟，建立了这些假定的神经生物学机制的计算可行性，用于保存控制功能

17. Brown, T.I., et al., *Cooperative interactions between hippocampal and striatal systems support flexible navigation.* 2012. **60**(2): p. 1316-1330.

对动物和人类的研究表明，海马体对于检索重叠信息序列的不同表示至关重要。最近有证据表明，尾状核和眶额皮质也参与了重叠空间表征的消歧。海马和尾状核是功能上不同的区域，但两者都与眶额皮质有解剖学联系。本研究使用基于fMRI的人体功能连接分析来检查海马，尾状和眶额皮质之间的功能关系，当参与者使用上下文信息来导航共享共同元素的良好学习的空间路线时。参与者在扫描仪外训练，从第一人称角度导航虚拟迷宫。重叠状态的迷宫开始并在不同的位置结束，但在中间聚合以与另一个迷宫共享一些走廊。非重叠状态的迷宫不与任何其他迷宫共用任何走廊。通过重叠的走廊成功导航需要识别当前导航路线的上下文信息，以指导针对给定试验的适当响应。结果显示，与非重叠的迷宫相比，重叠迷宫的海马，尾状和眶额皮质之间的功能连接性更强。目前的研究结果表明，海马和尾状核与前额结构相互作用，成功地依赖于上下文导航。

18. Carr, M.F., S.P. Jadhav, and L.M.J.N.n. Frank, *Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval.* 2011. **14**(2): p. 147.

海马是编码，合并和检索事件记忆所必需的。虽然这些过程背后的神经机制只是部分被理解，但最近的一系列论文指出唤醒内存重放是整合和检索的潜在贡献者。重播是表示先前经历的行为轨迹的海马位置细胞的顺序再激活，并且经常在清醒状态下发生，特别是在相对不动的时期。唤醒重放可以反映通过当前环境或先前访问的空间远程环境的轨迹。在压缩时间尺度上重复学习的序列非常适合于促进海马体外的分布式电路中的记忆巩固，这表明在清醒和睡眠动物中都发生了巩固。此外，感官信息可以影响唤醒重放的内容，表明唤醒重放在记忆检索中的作用。

19. Cassandra, A.R., L.P. Kaelbling, and M.L. Littman. *Acting optimally in partially observable stochastic domains*. in *Twelfth National Conference on Artificial Intelligence*. 1994.

20. Cohen, J.D., T.S. Braver, and J.W.J.C.o.i.n. Brown, *Computational perspectives on dopamine function in prefrontal cortex.* 2002. **12**(2): p. 223-229.

多巴胺和前额皮质对思想和行为至关重要。 最近，计算模型试图阐明多巴胺在前额叶皮层，神经生理学，系统和行为水平中的特定和复杂作用，并取得了不同程度的成功。

21. Cohen, J.Y., et al., *Neuron-type-specific signals for reward and punishment in the ventral tegmental area.* 2012. **482**(7383): p. 85.

多巴胺在动机和奖励方面发挥着核心作用。我们基于对气味和结果的反应发现了三种类型的神经元：大约一半的神经元（I型，52％）在以与奖励预测错误编码一致的方式奖励预测气味和奖励后显示相位激发;另一半神经元在气味和结果之间的延迟期间表现出持续的活动，其被结果的值调节为阳性（II型，31％）或阴性（III型，18％）。尽管I型神经元的活动对实际结果敏感（即，当奖励按预期递送时与意外省略时相比），II型和III型神经元的活动主要通过奖励预测气味来确定。我们用光敏蛋白通道视紫红质-2标记'多巴胺能和GABA能神经元，并根据它们在记录时对光刺激的反应进行鉴定。所有鉴定的多巴胺能神经元均为I型，所有GABA能神经元均为II型。这些结果表明，VTA GABAergic神经元发出预期奖励信号，这是多巴胺能神经元计算奖励预测误差的关键变量。

22. Collins, A.G., et al., *Working memory contributions to reinforcement learning impairments in schizophrenia.* 2014. **34**(41): p. 13747-13756.

这些结果表明，工作记忆系统对精神分裂症的学习障碍有很大帮助。

23. Collins, A.G. and M.J.J.E.J.o.N. Frank, *How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis.* 2012. **35**(7): p. 1024-1035.

因此，这项研究使我们能够指定高水平和低水平认知功能对工具性学习的不同影响，超越简单RL模型提供的可能性。

24. Cushman, F. and A.J.P.o.t.N.A.o.S. Morris, *Habitual control of goal selection in humans.* 2015. **112**(45): p. 13817-13822.

人类根据习惯和计划选择行动。 习惯控制在计算上是节俭的，但在新的情况下缓慢适应，而规划在计算上却很昂贵，但可以迅速适应。 目前的研究强调习惯和行为控制计划之间的竞争，但许多复杂的任务反而有利于他们的整合。 我们考虑一种层次结构，它利用习惯控制的计算效率来选择目标，同时保留计划实现这些目标的灵活性。 我们在强化学习环境中将此机制正式化，说明其成本和收益，并通过实验证明其在顺序决策任务中的自发应用。

25. Daphna, S. and A.D. Wagner, %J Neuron, *Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events.* 2008. **60**(2): p. 378-389.

这些发现为基于综合编码的泛化提供了证据，其中重叠的过去事件被整合到链接的助记表示中。海马 - 中脑相互作用支持经验的动态整合，提供了一种强大的机制，可以构建丰富的联想历史，超越个别事件。

26. Daw, N.D., *Advanced reinforcement learning*, in *Neuroeconomics*. 2014, Elsevier. p. 299-320.

本章回顾了当前强化学习理论及其神经基础研究的问题。我们考虑这些理论描述的状态，行动和奖励的正式结构如何被理解为映射到现实世界中生物有机体学习所经历的对应物。 在每种情况下，这种对应都涉及很大的困难。 然而，计算机科学中详细阐述的理论说明在每种情况下都阐明了如何将这些理论扩展到更现实的环境，同时仍然保留了神经经济学中突出的核心预测错误驱动的学习机制。

27. Daw, N.D., A.C. Courville, and D.S.J.N.c. Tourtezky, *Erratum: Representation and timing in theories of the dopamine system (Neural Computation (July 2006) 7,(1637)).* 2006. **18**(10): p. 2582.

28. Daw, N.D. and P.J.P.T.o.t.R.S.B.B.S. Dayan, *The algorithmic anatomy of model-based evaluation.* 2014. **369**(1655): p. 20130478.

尽管在二十世纪上半叶进行了许多辩论，但现在人类和其他动物建立环境模型并将其用于预测和控制，这在很大程度上是不言而喻的。 但是，基于模型（MB）的推理会带来严峻的计算挑战。 在强化学习文献中已经提出了替代的，计算上更简单的，无模型（MF）方案，并且已经提供了对行为和神经数据的有影响的描述。 在这里，我们研究了MB计算的实现，以及与MF值和评估方法结合在一起的方法。

29. Daw, N.D., et al., *Model-based influences on humans' choices and striatal prediction errors.* 2011. **69**(6): p. 1204-1215.

这些结果挑战了一个单独的无模型学习者的概念，并为高层次的人类决策提出了更加集成的计算架构。

30. Daw, N.D., Y. Niv, and P.J.N.n. Dayan, *Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control.* 2005. **8**(12): p. 1704.

广泛的神经和行为数据表明，大脑包含多个行为选择系统，包括一个与前额皮质相关的系统和另一个与背外侧纹状体相关的系统。 然而，这种过度控制会引发另一个选择问题：如何在系统不同意时对系统进行仲裁。 在这里，我们从规范的角度考虑使用强化学习的计算理论的双动作选择系统。 我们确定了一个关键的权衡，即计算简单性与灵活和统计上有效的经验使用。 在背外侧纹状体和前额叶系统之间的竞争中实现了权衡。 我们根据不确定性建议它们之间采用贝叶斯仲裁原则，因此每个控制器都应该在最准确的情况下进行部署。 这提供了关于有利于任一系统支配的因素的大量实验证据的统一说明。

31. Daw, N.D. and J.P. O’Doherty, *Multiple systems for value learning*, in *Neuroeconomics*. 2014, Elsevier. p. 393-410.

尽管选择通常是理论上的统一，但有许多经验证据表明决策是由多种，合作或竞争的神经和心理机制产生的。 我们回顾了人类和其他动物的决策受三种价值学习系统影响的证据：巴甫洛夫，习惯和目标导向。 这些系统在行为上是可解离的，由至少部分可区分的脑系统介导，并且体现了不同的计算原理。 我们讨论了这些系统之间用于行为控制的相互作用如何产生错误，效率低下和涉及强迫的障碍，以及这些系统如何与神经经济学中的其他双系统或多系统模型相关联。

32. Daw, N.D. and P.N. Tobler, *Value learning through reinforcement: the basics of dopamine and reinforcement learning*, in *Neuroeconomics*. 2014, Elsevier. p. 283-298.

我们还提供证据表明这种计算和神经生理机制影响决策和调节任务中的人和动物行为。

33. Dayan, P.J.N.C., *Improving generalization for temporal difference learning: The successor representation.* 1993. **5**(4): p. 613-624.

本文展示了TD如何用于学习这种表示，并使用导航任务说明了结果的适当分布性。

34. Dezfouli, A. and B.W.J.P.c.b. Balleine, *Actions, action sequences and habits: evidence that goal-directed and habitual action control are hierarchically organized.* 2013. **9**(12): p. e1003364.

虽然这些研究结果并不排除所有可能的无模型工具调节说明，但他们确实表明这些说明不是解释习惯性行为所必需的，并为理解目标导向和习惯性行为控制如何相互作用提供了新的基础。

35. Dickinson, A., B.J.S.s.h.o.e.p.L. Balleine, motivation, and emotion, *The role of learning in the operation of motivational systems.* 2002. **3**: p. 497-534.

36. Diuk, C., et al., *Hierarchical learning induces two simultaneous, but separable, prediction errors in human basal ganglia.* 2013. **33**(13): p. 5797-5805.

该结果表明，适当设计的任务可能揭示多巴胺能神经元中更复杂的发射模式。 此外，对这些信号的下游分离的需要意味着可能限制我们可以同时学习的不同任务级别的数量。

37. Dolan, R.J. and P.J.N. Dayan, *Goals and habits in the brain.* 2013. **80**(2): p. 312-325.

38. Doll, B.B., et al., *Model-based choices involve prospective neural activity.* 2015. **18**(5): p. 767.

决策可能通过“无模型”重复先前强化的行动或“基于模型”的评估而产生，人们普遍认为这是通过使用学习的地图或模型对行动后果的预期预期来实现的。虽然决策变量的选择和神经关联有时反映了其后果的知识，但仍不清楚这是否实际上来自于前瞻性评估。使用功能磁共振成像和顺序奖励学习任务，其中路径包含可解码的对象类别，我们发现人类基于模型的选择与在决策时观察到的未来路径的神经签名相关联，这表明了一种前瞻性的选择机制。勘探也与基于模型的决策变量的神经相关性的影响程度相关，并且与被认为是无模型学习基础的预测误差信号成反比。这些结果分离了基于模型和无模型评估的单独机制，并支持这样的假设：基于模型的选择和神经决策变量的影响来自于前景。

39. Eichenbaum, H. and N.J. Cohen, *From conditioning to conscious recollection: Memory systems of the brain*. 2004: Oxford University Press on Demand.

40. Engel, Y., S. Mannor, and R. Meir. *Reinforcement learning with Gaussian processes*. in *Proceedings of the 22nd international conference on Machine learning*. 2005. ACM.

41. Erev, I., E. Ert, and E.J.J.o.B.D.M. Yechiam, *Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions.* 2008. **21**(5): p. 575-597.

提出了三个实验，探讨了损失厌恶和敏感度降低驱使经验对选择行为的影响的断言。 实验侧重于重复选择任务，决策者在每个选择之间反复选择并获得反馈。 实验1a和1b表明，先前被解释为经验决策中损失规避的迹象的行为倾向被更好地描述为对绝对收益敏感性降低的产物。 实验2突出了名义上的幅度效应：名义收益幅度的减小消除了灵敏度降低的证据。 这些和相关的先前结果可以用一个模型来捕获，该模型假设依赖于主观体验的小样本，并且随着支付变化而增加灵敏度的降低。

42. Everitt, B.J. and T.W.J.N.n. Robbins, *Neural systems of reinforcement for drug addiction: from actions to habits to compulsion.* 2005. **8**(11): p. 1481.

药物成瘾越来越多地被视为从最初的药物使用开始的一系列转变的终点 - 当药物被自愿服用时，因为它具有增强的，通常是享乐的效果 - 通过失去对这种行为的控制，使得它成为习惯性的并且最终是强迫性的。 在这里，我们讨论证据表明这些转变取决于pavlovian和器乐学习过程之间的相互作用。 我们假设从自愿药物使用到更多习惯性和强制性药物使用的变化代表了从前额皮质到纹状体控制的药物寻求和吸毒行为的神经水平的转变以及从纹状体的腹侧到更多背侧区域的进展。 ，涉及其多巴胺能神经支配。 这些神经转变本身可能依赖于由药物的慢性自我施用诱导的皮质和纹状体结构中的神经可塑性。

43. Ezzyat, Y. and L.J.P.S. Davachi, *What constitutes an episode in episodic memory?* 2011. **22**(2): p. 243-252.

情节记忆的概念意味着存在一个将经验分成剧集的过程，以便它们可以存储在记忆中。因此令人惊讶的是，事件分割与将经验组织成记忆中的事件之间的联系尚未得到解决。我们发现，在参与者阅读叙述中不同位置包含时间事件边界的叙述后，他们对事件边界信息的长期联想记忆低于他们对事件内信息的记忆。这表明在编码期间的事件分段导致在存储器中对那些相同事件进行分段。此外，功能成像数据显示，在参与者之间，大脑活动与事件内信息的持续整合相一致，与这种助记符分割模式相关。这些数据是第一个解决支持将经验组织成长期记忆事件的机制。

44. Foster, D., R. Morris, and P.J.H. Dayan, *A model of hippocampally dependent navigation, using the temporal difference learning rule.* 2000. **10**(1): p. 1-16.

45. Frank, M.J., L.C. Seeberger, and R.C.J.S. O'reilly, *By carrot or by stick: cognitive reinforcement learning in parkinsonism.* 2004. **306**(5703): p. 1940-1943.

46. Fu, W.-T. and J.R.J.P.r. Anderson, *Solving the credit assignment problem: explicit and implicit learning of action sequences with probabilistic outcomes.* 2008. **72**(3): p. 321-330.

47. G Elliott, W. and S.J.S. Daphna, *Preference by association: how memory mechanisms in the hippocampus bias decisions.* 2012. **338**(6104): p. 270-273.

人们每天都在他们从未直接体验过的替代品之间做出新的选择。然而，这些决定往往是迅速和自信的。在这里，我们表明海马体传统上以其在构建长期陈述性记忆中的作用而闻名，它能够在记忆中传播价值，从而指导新选择之间的决策。在人类中使用功能性脑成像，我们发现给予人们金钱奖励导致激活预先建立的记忆网络，将奖励的正值传播到存储在记忆中的非奖励物品。后来，人们偏向于选择这些未经奖励的物品。这种决策偏倚是通过海马中的活动，相关记忆的重新激活以及大脑中记忆和奖励区域之间的连接来预测的。这些研究结果解释了新替代方案中的选择是如何从大脑建立记忆的联想机制中自动产生的。此外，我们的研究结果证明了海马体在基于价值的决策中的前所未有的作用。

48. Gabrieli, J.D.J.A.r.o.p., *Cognitive neuroscience of human memory.* 1998. **49**(1): p. 87-115.

目前的知识总结了人类大脑的长期记忆系统，记忆系统被定义为支持特定记忆过程的特定神经网络。该摘要整合了来自脑损伤患者的神经心理学研究和使用正电子发射断层扫描（PET）或功能磁共振成像（fMRI）的功能性神经影像学研究的趋同证据。关于海马和海马旁区域，杏仁核，基底神经节和各种新皮质区域在陈述性记忆中的具体作用的证据进行了回顾。还回顾了关于哪些大脑区域介导特定类型的程序记忆的证据，包括感觉运动，感知和认知技能学习;感性和概念重复启动;和几种形式的调理。根据正常记忆的功能性神经结构，记忆表现的年龄相关变化以及影响记忆的神经系统条件（如阿氏病，阿尔茨海默病，帕金森病和亨廷顿氏病）来讨论研究结果。

49. Gärtner, T., J.W. Lloyd, and P.A.J.M.L. Flach, *Kernels and distances for structured data.* 2004. **57**(3): p. 205-232.

本文将越来越重要的两条机器学习链结合在一起：核心方法和高度结构化的数据。我们提出了一种按照数据的句法结构构造内核的通用方法，该方法由数据的类型签名在高阶逻辑中定义。我们的主要理论结果是这样定义的任何核的肯定性。我们报告了一系列真实数据集的令人鼓舞的实验结果。通过将核转换为距离伪度量，我们可以将二萜数据集的最佳精度提高10%以上。

50. Geman, S., E. Bienenstock, and R.J.N.c. Doursat, *Neural networks and the bias/variance dilemma.* 1992. **4**(1): p. 1-58.

51. Gershman, S.J., D.M. Blei, and Y.J.P.r. Niv, *Context, learning, and extinction.* 2010. **117**(1): p. 197.

52. Gershman, S.J., A.B. Markman, and A.R.J.J.o.E.P.G. Otto, *Retrospective revaluation in sequential decision making: A tale of two systems.* 2014. **143**(1): p. 182.

在4个行为实验中，使用回顾性重估设计和认知负荷操纵，我们表明人类决策更符合协作体系结构，其中无模型系统控制行为，而基于模型的系统训练无模型系统通过重播和模拟经验。

53. Gershman, S.J., K.A. Norman, and Y.J.C.O.i.B.S. Niv, *Discovering latent causes in reinforcement learning.* 2015. **5**: p. 43-50.

54. Gilboa, I. and D. Schmeidler, *A theory of case-based decisions*. 2001: Cambridge University Press.

55. Gillan, C.M., et al., *Characterizing a psychiatric symptom dimension related to deficits in goal-directed control.* 2016. **5**: p. e11305.

56. Gläscher, J., et al., *States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning.* 2010. **66**(4): p. 585-595.

57. Glimcher, P.W. and E. Fehr, *Neuroeconomics: Decision making and the brain*. 2013: Academic Press.

58. Gonzalez, C. and V.J.P.r. Dutt, *Instance-based learning: Integrating sampling and repeated decisions from experience.* 2011. **118**(4): p. 523.

基于IBLT的单一认知模型（在抽样范式中添加了一个停止点规则）捕获了人类的选择，并预测了两个范式中的选择顺序。我们通过定量模型比较来整合范例，其中IBLT优于为每个范例分别创建的最佳模型。我们讨论了决策心理的影响。

59. Griffiths, T.L., et al., *Probabilistic models of cognition: exploring representations and inductive biases.* 2010. **14**(8): p. 357-364.

60. Gustafson, N.J. and N.D.J.P.C.B. Daw, *Grid cells, place cells, and geodesic generalization for spatial reinforcement learning.* 2011. **7**(10): p. e1002235.

强化学习（RL）提供了大脑学习做出有利选择的机制的一个有影响的特征。然而，一个重要的问题是，如何用一种能够有效学习的方式来表示复杂的任务。我们通过空间导航的透镜来考虑这个问题，研究大脑中的两个位置代表海马位置细胞和内生网格细胞如何适应作为RL在空间上近似值的基本功能。

61. Hare, T.A., et al., *Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors.* 2008. **28**(22): p. 5623-5630.

62. Hart, A.S., et al., *Phasic Dopamine Release in the Rat Nucleus Accumbens Symmetrically Encodes a Reward Prediction Error Term.* 2014. **34**(3): p. 698-704.

63. Hassabis, D. and E.A.J.P.T.o.t.R.S.o.L. Maguire, *The construction system of the brain.* 2009. **364**(1521): p. 1263-1271.

64. Hau, R., et al., *The description–experience gap in risky choice: the role of sample size and experienced probabilities.* 2010. **21**(5): p. 493-518.

这个模型分析表明，人们在两种类型的决策中对概率的处理是不同的，这是对大样本中剩余描述经验差距的一种解释。

65. Hertwig, R. and I.J.T.i.C.S. Erev, *The description-experience gap in risky choice.* 2009. **13**(12): p. 517-523.

汇集的研究结果表明，当人们根据经验做出决定时，根据客观概率，罕见事件的影响往往小于他们应得的影响。讨论了基于人类和动物经验的选择的显著相似性、这些选择的建模方法及其对风险和预防行为的影响。

66. Houk, J.C., J.L. Davis, and D.G.B.J.M.o.I.P.i.t.B. Ganglia, *A Model of How the Basal Ganglia Generate and Use Neural Signals That Predict Reinforcement.* 1995: p. 249-270.

67. Huys, Q.J.M., et al., *Interplay of approximate planning strategies.* 2015. **112**(10): p. 3098-103.

我们发现受试者利用域结构来建立子目标，其方法是在计算选择值的成本上达到几乎最大的降低，然后结合部分搜索和贪婪的局部步骤来解决子任务，并在遇到子任务时以一种自反的方式对子任务的决策树进行不适应性的修剪。客户损失。受试者特别喜欢特定的动作序列来实现子目标，创造新的复杂动作或“选项”。

68. Jäkel, F., B. Schölkopf, and F.A.J.T.i.C.S. Wichmann, *Does Cognitive Science Need Kernels?* 2009. **13**(9): p. 381-388.

69. Johnson, A. and A.D.J.J.o.N.t.O.J.o.t.S.f.N. Redish, *Neural Ensembles in CA3 Transiently Encode Paths Forward of the Animal at a Decision Point.* 2007. **27**(45): p. 12176.

70. Kahneman, D. and D.T.J.P.R. Miller, *Norm Theory: Comparing Reality to Its Alternatives.* 1986. **93**(2): p. 136-153.

71. Kahneman, D. and A.J.E. Tversky, *Prospect Theory: An Analysis of Decision under Risk.* 1979. **47**(2): p. 263-291.

72. Kearns, M. and S. Singh. *"Bias-Variance" Error Bounds for Temporal Difference Updates*. in *Thirteenth Conference on Computational Learning Theory*. 2000.

73. Keramati, M., A. Dezfouli, and P.J.P.C.B. Piray, *Speed/accuracy trade-off between the habitual and the goal-directed processes.* 2011. **7**(5): p. e1002055.

74. Knowlton, B.J., J.A. Mangels, and L.R. Squire, %J Science, *A neostriatal habit learning system in humans.* 1996. **273**(5280): p. 1399-1402.

75. Kruschke, J.K., %J Psychological Review, *ALCOVE: an exemplar-based connectionist model of category learning.* 1992. **99**(1): p. 22-44.

76. Kurth-Nelson, Z., et al., *Temporal structure in associative retrieval.* 2014. **4**(4).

77. Lansink, C.S., et al., *Hippocampus leads ventral striatum in replay of place-reward information.* 2009. **7**(8): p. e1000173.

将空间位置与奖励联系起来是自然环境中生存的基础，需要海马体和腹侧纹状体的完整性。在来自这些区域的联合多神经元记录中，海马纹状体在睡眠时一起重新激活。这一过程在海马细胞处理空间信息和腹侧纹状体放电与奖赏相关的配对中尤为强烈。重放主要由海马“位置”细胞在纹状体奖赏相关神经元前优先激活的细胞对所控制。我们的研究结果提示了巩固位置奖励关联的一个合理机制，并与巩固理论的核心原则相一致，表明海马体在投射区引导着再活化。

78. Lee, S.W., S. Shimojo, and J.P.J.N. O'Doherty, *Neural computations underlying arbitration between model-based and model-free learning.* 2014. **81**(3): p. 687-699.

79. Lieder, F., M. Hsu, and T.L. Griffiths, *The high availability of extreme events serves resource-rational decision-making.* 2017.

80. Love, B.C., D.L. Medin, and T.M. Gureckis, %J Psychological Review, *SUSTAIN: a network model of category learning.* 2004. **111**(2): p. 309-332.

81. Ludvig, E.A., C.R. Madan, and M.L.J.J.E.P.G. Spetch, *Priming Memories of Past Wins Induces Risk Seeking.* 2015. **144**(1): p. 24-29.

人们在不确定的情况下做决定时往往会规避风险。当这些决定是基于过去的经验时，人们必然依赖于他们的记忆。因此，在选择时记住的内容应该影响风险选择。我们在一个简单的风险选择任务中通过激发记忆来检验这一假设。在这项任务中，人们反复在一个安全的选择和一个风险的选择之间做出选择，这个选择以50/50的机会获得更大或更小的回报。一些试验之前有一个启动提示，之前与一个结果配对。我们发现，与胜利相关的启动线索导致人们成为风险寻求者，而与相对损失相关的启动线索几乎没有影响。这些结果表明，通过微妙地提醒人们以往的成功经验，人们可以被诱导成为更多的风险寻求者。

82. Ludvig, E.A., R.S. Sutton, and E.J.J.N.C. Kehoe, *Stimulus representation and the timing of reward-prediction errors in models of the dopamine system.* 2008. **20**(12): p. 3034-3054.

这种改进的拟合主要源于新模型中没有大的负误差，这表明在这种情况下，仅多巴胺就可以编码全范围的TD误差。

83. Lynch, J.G. and T.K.J.J.o.C.R. Srull, *Memory and Attentional Factors in Consumer Choice: Concepts and Research Methods | Journal of Consumer Research | Oxford Academic.* 1982. **9**(1): p. 18-37.

虽然消费者研究人员对决策过程中的认知过程表现出相当大的兴趣，但工作主要集中在对产品信息的有意识的心理操纵上。 本文讨论了可能在意识水平之下发生的记忆和注意过程。 提出了认知心理学领域开发的方法，以补充消费者研究人员常用的标准过程追踪方法。

84. Madan, C.R., et al., *Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions.* 2014. **21**(3): p. 629-636.

在根据过去的经验做出决定时，人们必须依靠他们的记忆。人类记忆有许多众所周知的偏见，包括更好地记住高度突出事件的倾向。我们提出了一个极端结果规则，即这种记忆偏差导致人们增加最大收益和最大损失，导致相对收益​​的风险更大，而不是相对损失。为了测试这个规则，在两个实验中，人们在固定期权和风险期权之间反复选择，其中风险期权可能比固定期权更多或更少。正如预测的那样，人们寻求相对收益的风险要高于相对损失。在随后的记忆测试中，人们倾向于首先回想起极端结果，并且还认为极端结果更频繁地发生。在个人中，风险选择任务中的风险偏好与这些记忆偏差相关。这种极端结果规则提供了一种记忆影响决策的新机制。

85. Mahadevan, S. and M. Maggioni, *Proto-value Functions: A Laplacian Framework for Learning Representation and Control in Markov Decision Processes*. 2007. 2169-2231.

86. Mathias, P., et al., *Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans.* 2006. **442**(7106): p. 1042-1045.

工具性学习的理论集中在理解如何使用成功和失败来改善未来的决策1。这些理论强调了奖励预测错误在更新与可用动作相关的值时的核心作用2。在动物中，大量证据表明神经递质多巴胺可能通过其调节皮质 - 纹状体突触功效的能力在这种类型的学习中具有关键功能3。然而，没有直接证据表明人类的多巴胺，纹状体活动和行为选择。在这里，我们表明，在仪器学习期间，纹状体中表达的奖赏预测误差的大小通过施用增强（3,4-二羟基-1-苯丙氨酸; 1-DOPA）或还原（氟哌啶醇）多巴胺能功能的药物来调节。因此，相对于用氟哌啶醇治疗的受试者，用I-DOPA治疗的受试者更倾向于选择最有益的行动。此外，将预测误差的大小结合到标准动作值学习算法中，准确地再现了在不同药物条件下的受试者的行为选择。我们得出结论，多巴胺依赖的纹状体活动调节可以解释人类大脑如何使用奖励预测错误来改善未来的决策。

87. Montague, P.R. and T.J.J.J.o.N. Dayan PSejnowski, *A framework for mesencephalic dopamine systems based on predictive Hebbian learning.* 1996. **16**(5): p. 1936-1947.

我们开发了一个理论框架，显示中脑多巴胺系统如何向目标分发代表未来预期信息的信号。特别地，我们展示了大脑皮质中的活动如何能够预测未来获得的奖励以及在基线水平之上和之下的弥漫性多巴胺系统中神经元活动水平的波动将如何表示传递给皮层和皮质下的这些预测中的错误。目标。我们提出了一个模型，用于如何在真实大脑中构建这样的错误，这与位于腹侧被盖区域和周围多巴胺能神经元的多巴胺能神经元子集的生理结果一致。该理论还在一个简单的决策任务上对人类选择行为进行了可测试的预测。此外，我们表明，通过对突触可塑性的简单影响，多巴胺释放的波动可以以适当的方式改变预测。

88. Murty, V.P., et al., *Episodic Memories Predict Adaptive Value-Based Decision-Making.* 2016. **145**(5): p. 548.

之前的研究表明，记忆可以指导基于价值的决策。例如，先前的工作暗示了工作记忆和程序记忆（即强化学习）在指导选择中的作用。然而，其他类型的记忆，例如情景记忆，也可能影响决策。在这里，我们测试情景记忆特定项目与联想记忆的作用，支持基于价值的选择。参与者完成了一项任务，他们首先了解了与试用独特彩票相关的价值。在短暂的延迟之后，他们完成了一项决策任务，他们可以选择重新接触之前遇到的彩票，或者从未见过的新彩票。最后，参与者完成了对彩票及其相关值的惊喜记忆测试。结果表明，参与者选择更频繁地重新使用彩票，从而导致高奖励和低奖励。重要的是，参与者不仅形成了与个人彩票相结合的奖励价值的详细联想记忆，而且只有当他们具有完整的联想记忆时才表现出适应性决策。我们进一步发现，自适应选择和联想记忆之间的关系推广到更复杂，生态有效的选择行为，例如社会决策。然而，个人更强烈地编码社会违规的经历 - 例如被不公平对待，暗示个人如何在社会环境中形成联想记忆的偏见。总之，这些发现提供了情节记忆和决策文献的重要整合，以更好地理解支持自适应行为的关键机制。

89. Neil, S., G.D.A. Brown, and C.J.P.R. Nick, *Absolute identification by relative judgment.* 2005. **112**(4): p. 881-911.

90. Niv, Y.J.J.o.M.P., *Reinforcement learning in the brain.* 2009. **53**(3): p. 139-154.

91. Nosofsky, R.M., %J J Exp Psychol Gen, *Attention, similarity, and the identification-categorization relationship.* 1986. **115**(1): p. 39-61.

92. O'Reilly, R.C. and M.J.J.N.C. Frank, *Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia.* 2006. **18**(2): p. 283-328.

长期以来，人们一直认为前额叶皮层可以提供工作记忆（在线处理信息以进行处理）和执行功能（决定如何操作工作记忆和执行处理）。尽管已经开发了许多工作记忆的计算模型，但执行功能的机制基础仍然是难以捉摸的，通常相当于一个小人。本文介绍了通过强大的学习机制解构这个小说的尝试，该机制允许前额叶皮层的计算模型以战略性，任务适当的方式控制自身和其他大脑区域。这些学习机制基于中脑，基底神经节和杏仁核中的皮质下结构，它们共同构成了演员 - 评论家的建筑。评论系统学习哪些前额表示与任务相关并训练演员，这反过来又提供了用于控制工作记忆更新的动态选通机制。在计算上，学习机制旨在同时解决时间和结构信用分配问题。该模型的性能优于基于标准反向传播的时间学习机制，适用于具有挑战性的1-2-AX工作记忆任务和其他基准工作记忆任务。

93. Ormoneit, D. and Ś.S.J.M. Learning, *Kernel-Based Reinforcement Learning.* 2002. **49**(2-3): p. 161-178.

94. Packard, M.G., J.L. Mcgaugh, %J Neurobiology of Learning, and Memory, *Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning.* 1996. **65**(1): p. 65-72.

研究结果表明：（1）海马和尾状核分别选择性地介导了位置和反应学习的表达（2），在视觉提示的外部环境中，海马依赖的地方学习获得比尾状依赖性反应学习更快，（ 3）当动物通过延长训练转变为尾状依赖性反应学习时，基于海马的位置表示保持完整。

95. Parker, N.F., et al., *Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target.* 2016. **19**(6): p. 845-854.

中脑中的多巴胺能（DA）神经元提供丰富的纹状体地形神经支配，并且是学习和产生动作的核心。尽管这种DA神经支配的重要性，但仍然不清楚DA神经元是否以及如何根据其纹状体靶标的位置进行专门化。因此，我们试图在仪器逆转学习任务的背景下比较靶向不同纹状体亚区的DA神经元亚群的功能。我们确定了背侧与腹侧纹状体中多巴胺末端的奖赏和选择编码的关键差异：腹侧纹状体的DA末端对奖励消耗和奖励预测线索的反应更强烈，而背内侧纹状体的DA末端对对侧选择的反应更强烈。在两种情况下，终端编码奖励预测错误。我们的研究结果表明，纹状体的DA调节是空间组织的，以支持目标亚区的特殊功能

96. Passingham, R.E.J.N., *The hippocampus as a cognitive map.* 1979. **4**(6): p. 863.

97. Pfeiffer, B.E. and D.J. Foster, %J Nature, *Hippocampal place-cell sequences depict future paths to remembered goals.* 2013. **497**(7447): p. 74-79.

有效导航需要规划扩展路线以记住目标位置。 海马位置细胞已被提议在导航规划中发挥作用，但缺乏直接证据。 在这里，我们显示在开放场地中的目标导向导航之前，大鼠海马体产生编码空间轨迹的简短序列，强烈偏向于从受试者的当前位置进展到已知的目标位置。 即使在开始和目标位置的特定组合是新颖的情况下，这些序列也可以预测即将来临的行为。 这些结果表明，先前在线性约束环境中表征为“重放”的海马序列事件也能够支持目标导向的轨迹发现机制，该机制在需要记忆检索的特定时间识别重要位置和相关行为路径， 并以可用于控制后续导航行为的方式。

98. Poldrack, R.A., et al., *Interactive memory systems in the human brain.* 2001. **414**(6863): p. 546-50.

人类的学习和记忆依赖于几种记忆系统，这些记忆系统似乎具有可分离的大脑底物1,2。一个基本问题涉及这些内存系统是否以及如何相互作用。在这里，我们展示了使用功能磁共振成像（FMRI），这些记忆系统可能在人类的分类学习期间彼此竞争。在分类学习期间，内侧颞叶和基底神经节在受试者之间的不同参与取决于该任务是否强调陈述性或非陈述性记忆，即使待学习材料和表现水平没有差异。与动物研究[3,4]和神经影像学5所表明的记忆系统之间的竞争相一致，这些区域的活动在个体之间呈负相关。使用事件相关FMRI对分类学习的进一步研究表明，在学习开始时这些区域的活动快速调节，表明受试者在学习早期依赖于内侧颞叶。

99. Reid, R.J.Q.J.o.E.P., *A test of sensory pre-conditioning in pigeons.* 1952. **4**(2): p. 49-56.

100. Riesbeck, C.K. and R.C.J.L.E.A. Schank, *Inside case-based reasoning.* 1989.

101. Ross, R.S., K.R. Sherrill, and C.E.J.B.R. Stern, *The hippocampus is functionally connected to the striatum and orbitofrontal cortex during context dependent decision making.* 2011. **1423**(2): p. 53-66.

我们的许多日常行为仅适用于某些情况，选择适当的行为需要我们使用当前背景和以往的经验来指导我们的决策。目前的研究在一项任务期间检查了与前额叶和纹状体区域的海马功能连接，该任务要求参与者基于重叠顺序表示的上下文检索做出决策。参与者学习了四个序列，每个序列由六个面组成。通过使具有两个相同面的两个序列作为中间图像来创建重叠条件。非重叠条件包含两个不共享它们之间任何面的序列。在呈现期间和关键选择期间评估海马功能连接，其中参与者必须做出依赖于上下文的决定。在参与者知道序列后，与重叠状态相比，左侧海马体在重叠的呈现期间显示出与背侧和腹侧纹状体和前扣带皮层的功能连接性显着增加。在重叠状态的关键选择点，左侧海马与眶额皮质显示出更强的功能连接性。这些功能连通性结果表明海马体可能通过预测接下来可能出现的可能性在决策过程中发挥作用，允许眶额和纹状体区域评估预期的选择选项，以便在选择点做出正确的动作。

102. Sadacca, B.F., J.L. Jones, and G.J.e. Schoenbaum, 5,, *Midbrain dopamine neurons compute inferred and cached value prediction errors in a common framework.* 2016. **5**.

103. Schacter, D.L., et al., *The future of memory: remembering, imagining, and the brain.* 2012. **76**(4): p. 677-694.

在过去几年中，研究人员在想象力和未来思维中研究记忆的作用急剧增加。这项工作揭示了记忆过去与想象或模拟未来之间惊人的相似之处，包括发现共同的大脑网络既是记忆又是想象的基础。在这里，我们讨论了近年来出现的一些关键点，特别关注在记忆和想象力分析中区分时间和非时间因素的重要性，记忆过去和想象未来之间差异的本质，识别包含支持基于存储器的模拟的默认网络的组件过程，以及该网络可以灵活地与其他网络耦合以支持复杂的目标导向模拟的发现。这一不断增长的研究领域通过强调记忆支持适应性功能的多种方式，拓宽了我们对记忆的概念。

104. Scholkopf, B. and A.J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. 2002.

105. Schonberg, T., et al., *Selective impairment of prediction error signaling in human dorsolateral but not ventral striatum in Parkinson's disease patients: evidence from a model-based fMRI study.* 2010. **49**(1): p. 772-781.

106. Schultz, W., ., P. Dayan, ., and P.R. Montague, %J Science, *A neural substrate of prediction and reward.* 1997. **275**(5306): p. 1593-1599.

107. Shohamy, D. and N.D.J.C.O.i.B.S. Daw, *Integrating memories to guide decisions.* 2015. **5**: p. 85-90.

适应性决策以过去的经验为指导。 然而，决策通常是在以前没有直接经历过的替代方案之间做出的，需要将记忆整合到多个过去的事件中。 我们回顾新出现的研究结果，支持至少两种看似截然不同的机制，即大脑如何利用选择服务中的记忆。 当面临新的决定时，会触发预期整合，允许在现场进行推论。 在实际面临决策之前，补充的回顾机制将现有记忆整合到相关经验的网络中。 我们讨论支持这些机制的证据以及理解记忆在决策中的作用的意义。

108. Simonsohn, U. and G.J.E.J. Loewenstein, *Mistake #37: The Effect of Previously Encountered Prices on Current Housing Demand\*.* 2010. **116**(508): p. 175-199.

109. Simonson, I. and A.J.J.o.M.R. Tversky, *Choice in Context: Tradeoff Contrast and Extremeness Aversion.* 1992. **29**(3): p. 281-295.

消费者的选择往往受到背景的影响，由所考虑的一系列替代品定义。 提出了关于语境对选择的影响的两个假设。 第一个假设，即权衡对比，表明偏好替代方案的倾向会增强或受阻，这取决于所考虑的集合内的权衡是否有利于或不利于该选择。 第二个假设，即极端厌恶，表明如果期权是选择集中的中间期权，则期权的吸引力会增强，如果是极端期权则会减少。 这些假设可以解释以前的发现（例如，吸引和妥协效应）并预测一些新的效果，在一系列以消费品作为选择替代品的研究中得到证实。 讨论了这些发现的理论和实践意义。

110. Skaggs, W.E. and B.L. Mcnaughton, %J Science, *Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience.* 1996. **271**(5257): p. 1870-1873.

睡眠期间大鼠海马锥体细胞的相关活性反映了早期空间探索中这些细胞的活性。 现在，已经发现睡眠期间的活动模式反映了在空间探索期间细胞发射的顺序。 在行为会话之后，这种关系对于睡眠来说比以前更加强烈; 因此，睡眠期间的活动反映了经验产生的变化。 对于神经元放电的时间顺序的这种记忆可以通过长期增强的时间整合特性和尖峰活动相对于海马θ节律的相移之间的相互作用产生。

111. Squire, L.R., %J Psychological Review, *Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans.* 1992. **99**(2): p. 195-231.

陈述性记忆与不需要海马体的非陈述性（隐性）记忆能力的异质集合（技能和习惯，简单的条件反射和引发现象）形成对比。临时需要海马体将新皮层中的分布式部位结合在一起，这些部位一起代表整个记忆。

112. Steinberg, E.E., et al., *A causal link between prediction errors, dopamine neurons and learning.* 2013. **16**(7): p. 966-973.

113. Stewart, N., N. Chater, and G.D.J.C.P. Brown, *Decision by sampling.* 2006. **53**(1): p. 1-26.

114. Sutton, R.S., *Learning to Predict by the Methods of Temporal Differences*. 1988. 9-44.

115. Sutton, R.S.J.A.S.B., *Dyna, an integrated architecture for learning, planning, and reacting.* 1991. **2**(4): p. 160-163.

116. Tenenbaum, J.B., V. Silva, De, and J.C. Langford, %J Science, *A global geometric framework for nonlinear dimensionality reduction.* 2000. **290**(5500): p. 2319-2323.

117. Tulving, E.J.O.o.M., *Episodic and semantic memory.* 1972. **381**(79): p. 381-403.

118. Vaidya, A.R. and L.K. Fellows, %J Journal of Neuroscience the Official Journal of the Society for Neuroscience, *Ventromedial Frontal Cortex Is Critical for Guiding Attention to Reward-Predictive Visual Features in Humans.* 2015. **35**(37): p. 12813-23.

119. Volodymyr, M., et al., *Human-level control through deep reinforcement learning.* 2015. **518**(7540): p. 529.

120. Wasserman, L., *All of Nonparametric Statistics*. 2006.

121. Wimmer, G.E., et al., *Episodic memory encoding interferes with reward learning and decreases striatal prediction errors.* 2014. **34**(45): p. 14901.

学习对于适应性决策至关重要。已知纹状体及其多巴胺能输入支持基于奖励的增量学习，而已知海马体支持单个事件（情景记忆）的编码。虽然传统上分开研究，即使是简单的经历，这两种类型的学习也可能共同发生并可能相互作用。在这里，我们通过研究增量奖励学习如何与并发情景记忆编码相关来寻求理解这种互动的本质。在实验过程中，人类参与者在两个选项（彩色方块）之间做出选择，每个选项与漂移的奖励概率相关联，目标是赚取尽可能多的钱。每个选项都覆盖了与选择无关的附带的试验唯一对象图片。第二天，参与者对这些照片进行了惊喜记忆测试。我们发现，更好的情景记忆与最近的奖励经历对参与者内部和之间的选择的影响减少有关。 fMRI分析进一步揭示，在学习期间，当情景记忆更强时，规范的纹状体奖励预测误差信号显着减弱。纹状体中奖赏预测误差信号的这种减少与选择时海马和纹状体之间增强的功能连接性相关。我们的研究结果提出了一种机制，通过这种机制，记忆编码可以竞争纹状体处理，并提供洞察不同形式的学习之间的相互作用如何指导基于奖励的决策。

122. Yael, N., et al., *Reinforcement learning in multidimensional environments relies on attention mechanisms.* 2015. **35**(21): p. 8145-57.

123. Zilli, E.A. and M.E.J.H. Hasselmo, *Modeling the role of working memory and episodic memory in behavioral tasks.* 2010. **18**(2): p. 193-209.

已经使用强化学习理论研究了目标导向行为的机制，但是这些理论技术并未经常用于解决记忆系统在执行行为任务中的作用。 这项工作通过提供一种方式来解决这个缺点，即工作记忆（WM）和情景记忆可以包含在强化学习框架中，然后模拟六个行为任务的成功获取和表现，这些任务来自大鼠实验文献或受其启发， 需要WM或情景记忆以获得正确的表现。 在任务期间没有施加任何延迟，使用WM的模拟可以在机会级别之上解决所有任务。 当施加延迟时，具有情景记忆和WM的模拟可以解决除了气味序列任务的歧义消除之外的所有任务。