Open Access Library Journal
2025, Volume 12, e14330
ISSN Online: 2333-9721
ISSN Print: 2333-9705

The Reward and Punishment System (Reinforcement Learning System) for Strong Artificial Intelligence and Human

Hongwen Cheng

Department of Medicine, The Third People's Hospital of Zhongxiang City, Zhongxiang, China
Email: chenghwn@aliyun.com

How to cite this paper: Cheng, H.W. (2025) The Reward and Punishment System (Reinforcement Learning System) for Strong Artificial Intelligence and Human. *Open Access Library Journal*, **12**: e14330.
https://doi.org/10.4236/oalib.1114330

Received: September 22, 2025
Accepted: November 1, 2025
Published: November 4, 2025

Abstract

The reward and punishment system is one of the fundamental reasons why humans possess intelligence. In order for strong artificial intelligence software to have true intelligence, it must have an appropriate reward and punishment system. This article discusses the structure and functions of the reward and punishment system for strong artificial intelligence from the perspectives of programming and personification, as well as some important reward and punishment phenomena. On this basis, this paper discussed the reward and punishment system that obtained through programming based on my theory how processes stimulus information under the purpose, how to adapt the thinking and behavior of strong artificial intelligence software to the environment through reward and punishment learning, and how to accelerate its learning speed and ability to adapt to the environment.

Subject Areas

Artificial Intelligence, Neuroscience

Keywords

Strong Artificial Intelligence, State Centre, Expectation of Rewards and Punishments, Purpose Centre, Strength Centre, Mindset

1. 前言

奖惩系统(与强化学习系统对应)是人和强人工智能软件获得智能的关键。适应是人的奖惩系统进化形成的根本原因，也是我们设计强人工智能软件奖惩系统结构功能的根本标准。

强人工智能是当今科学界的热点。强人工智能的关键是智能，而学习与

适应是智能最重要的特点。奖惩中枢与学习[1]这篇文章证明了，智能实体要具有学习与适应环境的能力，它就必须有逃避惩罚与追求奖赏的能力，也就是说必须具有奖惩中枢及奖惩预期中枢。这里的奖惩学习与强化学习相对应，深度学习正因为引入了强化学习的概念，才使深度学习得到大的发展。但深度学习的强化理论相对于智能生物的奖惩学习就显得过于浅显了，还不足以实现真正的智能。这篇文章主要介绍的是如何模拟人脑的奖惩系统进行编程设计，并讨论了这样编程设计的奖惩系统的结构功能，及如何在这些结构功能的基础上实现智能的。

强人工智能的学习必须是奖惩学习[1]，奖惩系统就是与奖惩学习有关的一些“大脑”结构。强人工智能软件只有具有相对完善的奖惩系统，才能获得真正的智能。

本文是在《拟人智能的实现》[2]这本书的基础之上进行的理论探讨，亦为如何进一步编程获得完整强人工智能软件提供理论依据。本文关于强度中枢与奖惩预期的关系，奖惩预期并不需要“完全”客观的反应奖惩的认知及目的是如何涌现的讨论对我们理解智能是如何形成的至关重要。

《拟人智能的实现》这本书的第十一章(对环境的适应与产生智能的必要条件)从理论上证明了智能实体要实现真正的智能必须具有奖惩、奖惩预期、目的、注意力的分配这些功能及与这些功能相对应的相关结构。这些结构与功能就是本文讨论的奖惩系统的结构与功能。奖惩系统包含奖惩中枢、奖惩预期中枢、状态中枢、强度中枢等中枢。人脑的奖惩系统除了包含边缘系统中的大多数结构外，应该还包含大部分额叶皮质。

奖惩预期是思想行为的动力，人(强人工智能软件)的任何有意识的思想行为都存在奖惩预期。通过长期的奖惩学习，人(强人工智能软件)的思想行为模式最终是与环境相适应的思想行为模式。通过奖惩学习对“大脑”兴奋的选择、强化、抑制作用，最终，所有的有意识的思想行为都会与奖惩相关，本文只讨论一些我认为重要的奖惩问题。

2. 奖赏、惩罚

人与其它高等生物的奖惩系统使它们能够更好的适应环境，我们要编程获得的强人工智能软件的奖惩系统，可以部分模拟生物的奖惩系统。

通过奖惩实践学习，最终，所有的“后天”奖惩(比如夸奖的言语)都来源于“先天”的基础奖惩(比如食物)，所有的奖惩预期也是来源于基础奖惩预期。因为记忆、学习的作用，所有的奖惩预期都会是“混合”的奖惩预期(比如在获得食物时，可能会有各种言语的奖惩，或者是体罚等等奖惩体验，而这些体验通过记忆最终都会对这个食物产生的奖惩预期产生一定的影响)。基础奖惩类型基本可以分为：与基础生理相关的奖赏、惩罚，与伤害刺激相关的惩罚。(生物与强人工智能软件都必须有与基础“生理”相关的奖惩，但由于“生理”结构的不同，在具体的基础奖惩的设计上，强人工智能软件与生物的又必然有所不同)

2.1. 奖惩与内外环境的关系

人与其它生物的奖惩系统最重要的功能就是使生物获得有利于生物生存

的内外环境，而逃避不利于生物生存的内外环境。在《获得性遗传与细胞结构功能》[3]这篇文章的“什么情况下产生获得获得性遗传，奖惩、应急反应”这一节讨论了环境、神经内分泌、奖惩之间的关系。

人与其它生物的奖惩系统是适者生存，进化选择的产物，它使生物能够更好的适应环境，能够获得更好的生存环境。人与其它生物获得惩罚(包括伤害性刺激)时会有不舒服的感觉(相应的神经内分泌兴奋)，而当这个惩罚去除后，不舒服的感觉去除，会有相对舒服的感觉，获得奖赏。生物的奖惩系统、生存环境与神经内分泌是统一的，进化使奖惩系统、神经内分泌与环境相适应。在适宜的生存环境下，生物会有相应的神经内分泌，生物会感觉到“舒服”，生物会去追求这个环境状态，这种情况下，生物一般会处于奖赏状态。不适宜的生存环境，生物会有相应的神经内分泌，生物会处于惩罚状态，生物会逃避这一状态。

奖赏与惩罚是对立统一的：

1) 对生物或者强人工智能软件来说，奖赏刺激之所以能够成为奖赏刺激，是因为能够产生奖赏刺激的“对象”的缺乏。它们注意到这个“缺乏的对象”时，会有惩罚的感觉，要逃避这个惩罚，就需要相应的奖赏刺激。逃避了惩罚就相当于获得了奖赏，奖赏与惩罚是对立统一的。

2) 伤害刺激，当伤害刺激发生时，生物会应激，会有相应的神经内分泌，是惩罚。而当逃避了这个伤害刺激后，相当于逃避了惩罚，获得了奖赏。

3) 《特殊记忆柱群的奖惩学习》[4]这篇文章论述了人在无法获得奖赏的情况下，会带来惩罚的机理，这也适用于其它生物与强人工智能软件。

综上所述，生物的奖赏与惩罚是对立统一的，逃避惩罚会带来奖赏，无法获得奖赏会带来惩罚。预期能够逃避惩罚，会产生奖赏预期……。

那么，对于强人工智能软件来说，我们该如何来设计它的奖惩系统哪？设计奖惩系统的根本目的是为了让强人工智能软件适应环境。如何适应环境，可以借鉴人类的奖惩系统，但也不应该完全模仿人的奖惩系统。设计的奖惩系统应该使基本的生存学习功能得到实现，比如：设计基本需求(能量等等)、逃避伤害刺激的奖惩系统、中介奖惩系统。

智能机器人需要设置类似于生物的内分泌系统的功能吗？(可以用特殊的神经系统来模拟)。如果智能机器人需要“应急”或者其各个重要组成结构需要适宜的内外环境，我们就可以给它设计相应的神经“内分泌”系统，使它能够“应急”及追求适宜的“神经内分泌”，逃避不适宜的“神经内分泌”。

2.2. 奖惩系统的结构功能及编程设计

强人工智能软件奖惩系统的编程设计可以参考人脑的智能实现及人脑奖惩系统的功能。

《奖惩中枢与学习》《注意力问题的系统讨论》[6]这两篇文章讨论了奖惩、奖惩预期、注意力、目的与学习的大概关系，指出了高等生命(高级智能)必然会有奖惩、奖惩预期、目的下的注意力分配(状态兴奋)，并讨论了它们相互之间的功能、结构关系。

关于奖惩、奖惩预期、状态兴奋的具体编程设计，我的小程序中[2] [5]有简单的相关编程设计，这种设计反应了它们之间的大概关系。

思维的中心与动力是目的(目的对象被奖惩预期赋值)。由主注意对象及亚主注意对象根据奖惩预期情况形成相应目的及目的对象，然后由目的对象根据对应的奖惩预期值易化兴奋状态中枢对应的记忆柱，最后由状态中枢强烈易化兴奋其它中枢对应的记忆柱群。从而使状态中枢根据奖惩、奖惩预期强烈影响主注意对象的选择、亚主注意对象的奖惩预期等等所有的与思想行为相关的兴奋。最终强人工智能软件根据奖惩预期影响它的思想行为的走向。

奖惩是奖惩预期的基础，奖惩预期来源于奖惩。我在《拟人智能的实现》这本书中的相关章节，系统的讨论了这些结构功能实现的编程问题，并进行了简单的编程。

可以说，强人工智能软件如果没有与目的(目的中枢的功能可以由状态中枢的兴奋涌现)、奖惩(奖惩中枢)、奖惩预期(奖惩预期中枢)、注意力分配(状态中枢)相对应的“结构功能”，便不可能真正具有高级智能。

本文会根据既往实践经验进一步讨论奖惩中枢、奖惩预期中枢、状态中枢的结构功能设置问题，力求能够加深大家对这些功能、结构的认识，从而能够更好的编程。

强人工智能软件的奖惩除了“先天”的基本设置，都应该是通过长期的实践学习，获得的经验奖惩。任何奖惩及奖惩预期都是建立在基础的奖惩、奖惩预期之上(奖惩预期可以相互组合，相互叠加，而形成新的奖惩预期)。

(编程的时候，也可以设置专门的结构来控制状态中枢的奖记忆柱及惩记忆柱的兴奋易化，这个结构的兴奋的最强影响因素是奖惩中枢及奖惩预期中枢，但这个结构能够与其它中枢建立记忆，并能够被其它中枢兴奋，并能够被习惯性兴奋，最终可以撇开奖惩预期而影响对应奖记忆柱及惩记忆柱的兴奋易化，这个有利于习惯性思维活动顺利运行。)

奖惩系统主要包括目的中枢、奖惩中枢、奖惩预期中枢、状态中枢、强度中枢。

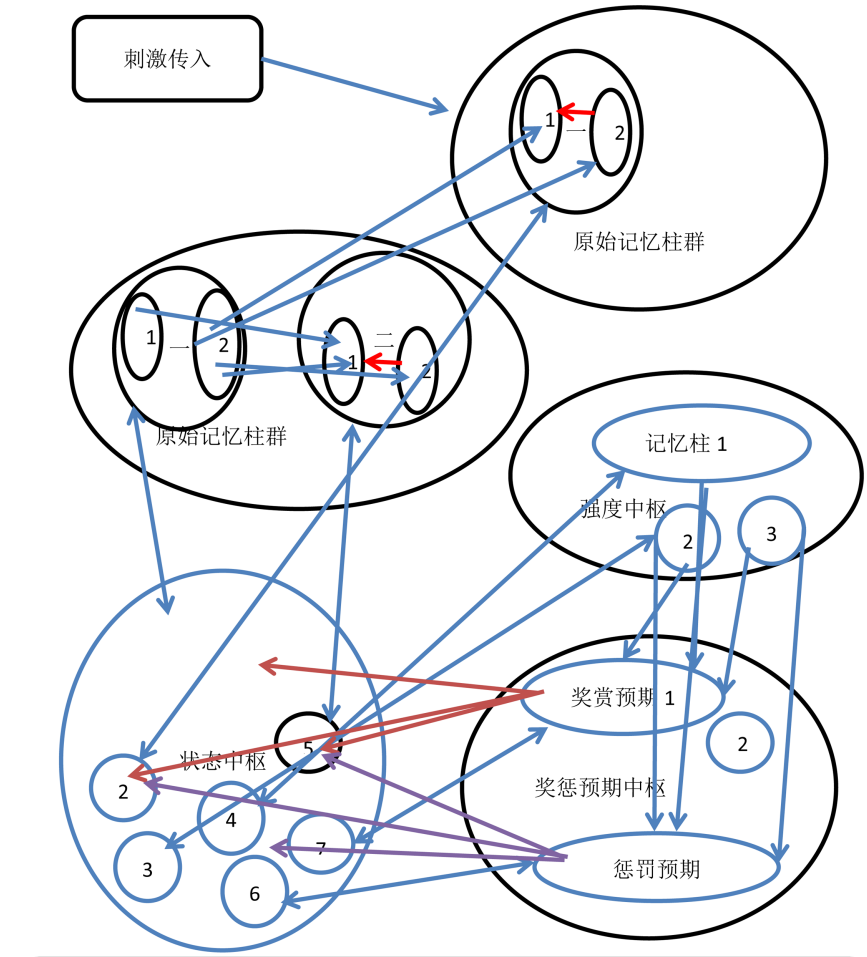
如图 1，介绍了，刺激传入后皮质中枢与奖惩系统各个结构之间的互动过程。

2.3. 奖惩中枢、奖惩预期中枢

我们在进行强人工智能编程时，该如何具体设置奖惩中枢那？对于奖惩中枢的设计，我部分参考了人的奖惩中枢。

奖惩中枢主要分为奖赏中枢与惩罚中枢两部分。奖赏中枢又由几个接受不同先天奖赏刺激传入的亚中枢组成，每个亚中枢只接受一种类型的奖赏刺激，每个亚中枢的可兴奋强度受“欲望”情况及先天奖赏刺激情况共同影响(比如饥饿感、甜感等等刺激对对应奖惩中枢的兴奋影响)。

每个亚中枢都有相应的联络区，其接受对应亚奖惩中枢的兴奋传入及其它中枢联络区的兴奋传入。亚奖惩中枢对其联络区的兴奋能力与它的兴奋强度密切相关，这个联络区的兴奋强度又会兴奋对应的强度中枢，从而与其它亚奖惩中枢的联络区共同参与奖惩预期的计算。



大写数字代表基本记忆柱群，原始记忆柱群中的小写数字 1 代表易兴奋的记忆柱群，小写数字 2 代表不易兴奋的记忆柱群。原始记忆柱群红色箭头代表抑制，蓝色箭头代表易化兴奋。橙色箭头表示对奖赏记忆柱的易化，绿色箭头表示对惩罚记忆柱的易化。原始记忆柱群由基本记忆柱群组成，基本记忆柱群由易兴奋和不易兴奋的记忆柱组成。原始记忆柱群分别与传入刺激及主注意对象对应

图 1. 刺激传入后与奖惩系统各个结构之间的互动过程

惩罚中枢的结构功能与奖赏中枢的类似。

如果外界刺激直接兴奋奖惩预期中枢相应的联络区，无疑会加快强人工智能软件对外界刺激的反应速度。如图 2，外界刺激 3 通过记忆的兴奋通路可以直接兴奋奖惩预期中枢的联络区，进而使奖惩预期中枢发生相应的兴奋，从而产生奖惩预期。我称它为奖惩预期通路 1。

而外界刺激 3 通过记忆兴奋各个亚奖惩中枢的联络区，再通过联络区的兴奋，对应兴奋相应强度中枢，再通过与对应的亚奖惩中枢的激活情况综合计算(比如通过食物产生的奖赏刺激及奖赏预期的大小与饥饿程度密切相关)后兴奋对应奖惩预期中枢(当然也可以设计成不需要强度中枢的计算中介，其本身就可计算)，产生奖惩预期，显然前一通路所需要的计算少很多。这一通路是奖惩预期通路 2。

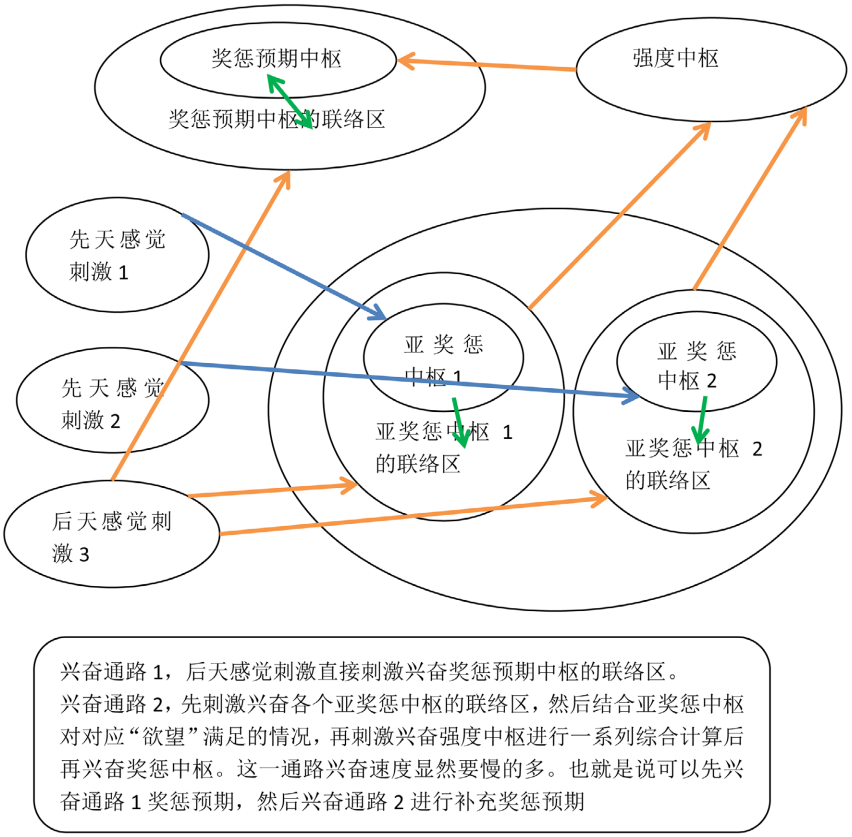


图 2. 奖惩预期的两种兴奋通路

刺激 3 的这两个奖惩预期通路最终在奖惩预期中枢产生综合兴奋，而获得最终的奖惩预期。并在联络区产生相应记忆。后面的兴奋通路是前面的兴奋通路的基础，前面的兴奋通路来源于后面兴奋通路对奖惩预期中枢兴奋后的记忆。

惩罚预期的计算机制相同。

环境及目的状态下，一个物体对象产生的一视觉刺激传入感觉中枢产生相应易化兴奋，然后被选择成为主注意对象并传出兴奋，同时状态中枢对应的记忆柱群兴奋，传出的兴奋会根据记忆兴奋奖惩预期中枢的联络区、强度中枢等等中枢并产生一系列的计算，这一刺激如果产生的奖惩预期值的变化速率大到一定程度，就会迅速诱发注意，这时的主注意对象及亚主注意对象就会成为主注意目的对象，而动力预期最弱的目的对象就会被“踢出”目的队列……。而如果这个刺激引起的奖惩预期值的变化速率不大，不会引起注意，这个环境对象会产生进一步的视觉刺激，会通过奖惩预期通路 2 进一步奖惩预期，如果产生的动力预期值大于目的“队列”中奖惩预期值最小的目的对象的奖惩预期值，这时的主注意目的对象及亚主注意目的对象就会成为新的目的对象……。

饥饿感、甜感是先天的奖赏刺激，而产生甜感的食物视觉刺激则可以成为后天奖赏刺激。

通过奖惩学习，一些后天奖惩刺激的刺激往往能够同时兴奋多个亚奖惩

中枢的联络区(比如如果一个食物能够带来甜感、香气刺激，这个食物的视觉刺激就能够兴奋……)，就如后天奖惩刺激 3。

2.4. 强度中枢

思维运行过程中会大量涉及到神经通路的兴奋强度，而且往往不是简单的涉及，这种情况下，强度中枢就会发挥其作用。

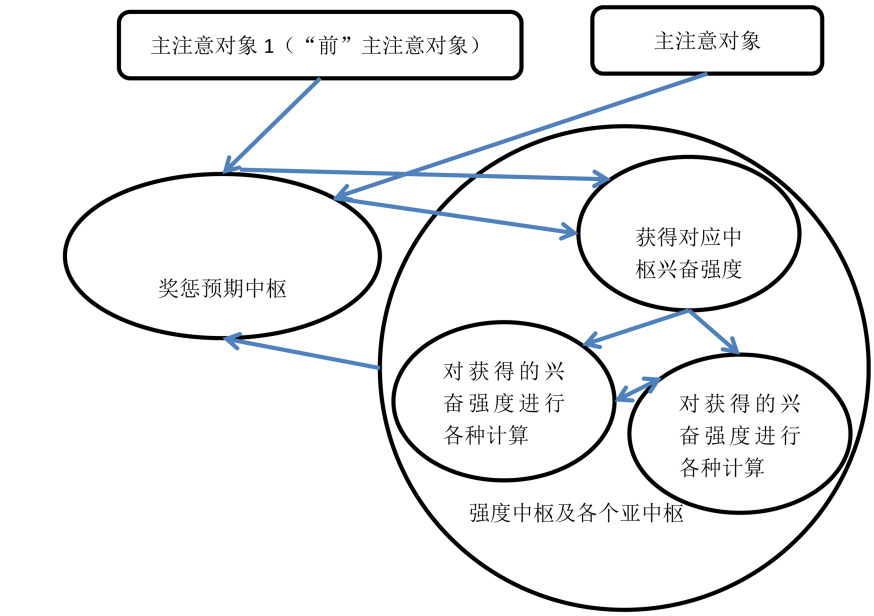
顾名思义，强度中枢的功能就是感知各个中枢及这些中枢的各个局部的兴奋强度分布(包括易兴奋的记忆柱、不易兴奋的记忆柱)，从而“感知”各个中枢及局部能够表征环境对象的各种属性的兴奋。并进行一系列的计算(可由专门的结构来参与)，提取与内外环境的“重要”属性密切相关的“数据”，从而获得大量对思维的“正确”运行至关重要的“数据”，以此来调控思想行为的运行，使之与环境相适应。

大脑的任何功能都是由其特定结构的兴奋实现的，而强度中枢通过感知大脑的兴奋状态来“了解”大脑能够反应大脑的功能状态，并能够根据“了解”影响奖惩预期，从而调控思想行为。这是强度中枢能够调控思维的基础。

对我的智能理论来说，强度中枢是最重要的中枢之一。

强度中枢、奖惩预期中枢、奖惩预期

一个新的对象产生刺激会使奖惩预期中枢的兴奋(奖惩预期值)在原有的基础上发生一定的变化，这种变化可以被强度中枢感知(见图 3)。



1、主注意对象直接兴奋奖惩预期中枢的联络区，产生奖惩预期是通过奖惩预期通路 1。
2、兴奋的奖惩预期中枢兴奋强度中枢相应的一系列亚中枢，对传入的兴奋强度进行记忆、计算。
3、计算获得的兴奋强度变化达到一定程度，会诱发相应亚中枢兴奋并进行计算。同时根据计算结果兴奋相应亚中枢。（兴奋强度变化速率快，说明对象产生的奖惩预期值可能比较大，迅速引起注意。在兴奋强度的变化平稳后，就可以得到对象产生的奖惩预期值）。

图 3. 主注意对象的奖惩预期值的计算与应用

一个对象的动力预期值的大小随着时间的流逝，会不断衰减，同时由于记忆柱之间的相互抑制，也会使动力预期值不断衰减。而主注意对象持续兴奋动力预期中枢，随动力预期中枢的兴奋不断增强，相同兴奋强度的传入刺激对它的兴奋能力会不断衰减，最后兴奋与衰减的速度会在一定范围内达到平衡。

奖惩预期中枢的兴奋强度被强度中枢感知后，通过“适当”计算，会获得奖惩预期的各种变化特点，然后通过这些特点来调控强人工智能软件的思想行为。

比如强人工智能软件的主注意对象产生前计算获得的动力预期值，与产生后获得的动力预期值的差值的计算。可以最先通过奖惩预期的通路 1 来计算实现(如图 2)，这样使强人工智能软件能够对重要的环境对象迅速进行注意(“重要”说明产生的奖惩预期值高，被它对应兴奋的奖惩预期中枢兴奋强，奖惩预期值变化的速率快)。这个动力值的差值可由专门的中枢计算获得(如果编程，可由专门的程序片段计算)，计算后兴奋专门的结构，并产生记忆，这些记忆包括随后正确的注意过程(正确的注意带来对环境对象正确的反应，从而正确逃避惩罚，获得奖赏，这个奖赏会影响今后的奖惩预期)，这些“正确”的记忆使动力差值超过适当范围(对应的环境刺激及动力预期值)时，能够迅速诱发“适当”的注意行为。当然，主注意对象要能够迅速引起注意，需要这个主注意对象本身或者曾经与能够带来强烈奖惩的对象存在联系，而使它能够产生强烈的奖惩预期。这个对象如果是特殊的新奇刺激也可以，新奇刺激本身能够带来强的中介奖惩预期，再加上其它的奖惩预期带来的综合奖惩预期。

又比如计算的动力预期值变化平稳后，先后两次动力预期值平稳后的值的差的计算。

动力预期达到平衡的时间不会太长，如果太长，会影响思维的速度。在完整动力预期的情况下，动力预期中枢的传入刺激会基本恒定，这时的动力预期值作为传入动力预期兴奋恒定时的动力预期值，来参与对象的动力预期值的计算，应该也能够获得与环境相适应的计算结果。

环境及目的状态下，一个物体对象产生的一视觉刺激传入感觉中枢产生相应易化兴奋，然后被选择成为主注意对象并传出兴奋，同时状态中枢对应的记忆柱群兴奋，传出的兴奋会根据记忆兴奋奖惩预期中枢的联络区、强度中枢等等中枢并产生一系列的计算，这一刺激如果产生的奖惩预期值的变化速率大到一定程度，就会迅速诱发注意，这时的主注意对象及亚主注意对象就会成为主注意目的对象，而动力预期最弱的目的对象就会被“踢出”目的队列……。而如果这个刺激的奖惩预期值的变化速率不大，不会引起迅速注意，这个环境对象会产生进一步的视觉刺激，会通过奖惩预期通路 2 进一步奖惩预期，如果产生的动力预期值大于目的“队列”中奖惩预期值最小的目的对象的奖惩预期值，这时的主注意目的对象及亚主注意目的对象就会成为新的目的对象……。

新奇刺激对应的主注意对象之间记忆联系相对较弱(较少建立记忆联系)，兴奋也较弱，易兴奋的记忆柱作为主注意对象的时间较长，兴奋强度较

强，不易兴奋的记忆柱一样，这样的兴奋属性应该能够被强度中枢感知。

对于选择出主注意对象那种雪崩式的计算模式，可能不需要强度中枢，但不同感觉中枢之间选择出主注意对象可能需要强度中枢的参与。强度中枢中枢只需要感知各个中枢局部的兴奋属性，并不需要感知单个记忆柱的兴奋属性。

2.5. 状态中枢

简单的说，状态中枢是奖惩影响皮质记忆柱群兴奋的中间转换结构，它负责将奖惩及奖惩预期转换为对皮质记忆柱群的强烈易化、抑制，从而使奖惩强烈影响思想行为的进程，而使强人工智能软件的思想行为与环境相适应。

状态中枢的结构功能的设计可以见《对编程实现拟人智能可行性的论证》[7]这篇文章的状态中枢这一节的讨论。

状态中枢的基本记忆柱群的兴奋强度是根据主记忆柱的基本兴奋强度，再结合其奖赏记忆柱、惩罚记忆柱的兴奋情况综合计算获得(计算模式，根据奖惩对思想行为的影响情况来设计)，最后通过主记忆柱传出兴奋(传出的兴奋会强烈易化兴奋状态中枢相关基本记忆柱群，并强烈影响其它中枢对应记忆柱的易化兴奋)。而奖赏预期强烈影响奖赏记忆柱的兴奋，惩罚预期强烈影响惩罚记忆柱的兴奋(见图 4)。

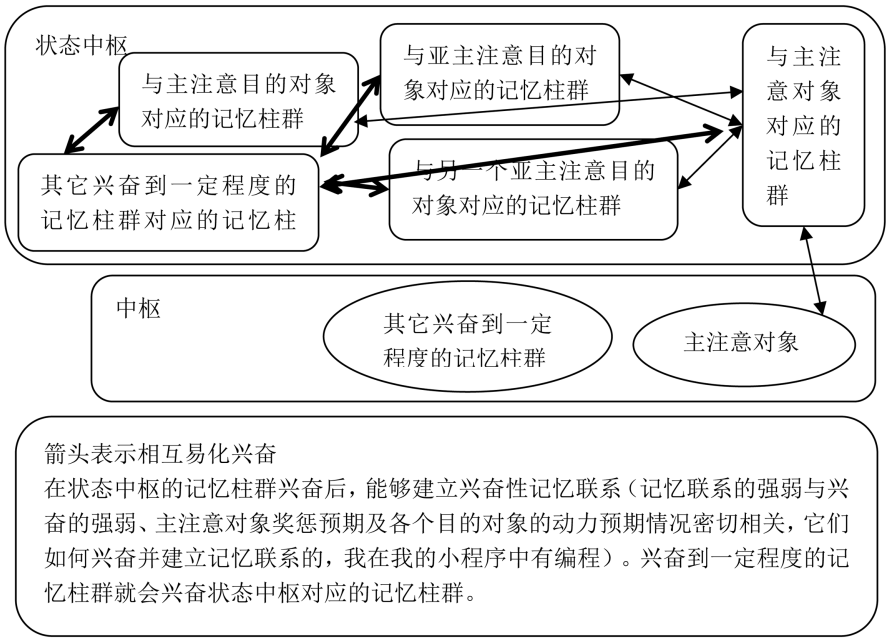


图 4. 主注意对象、兴奋到一定程度的记忆柱群与各个目的对象之间的关系图及相应说明

2.6. 目的中枢

这章的后面论述了，目的中枢的功能是如何在奖惩预期中枢及状态中枢的交互兴奋过程中涌现的(见图 5)。

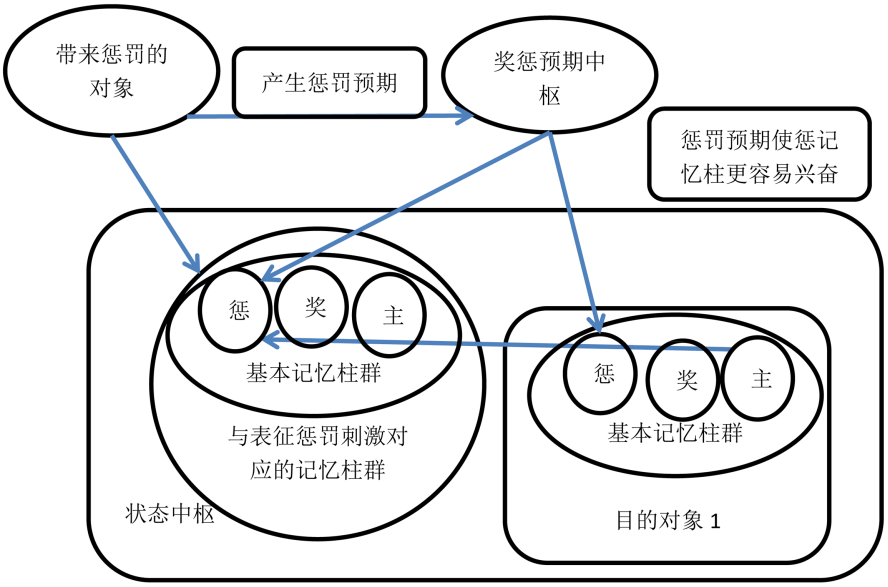


图 5. 状态中枢与奖惩预期的关系

目的中枢的概念及模型是我在思考、模拟人的思维过程中获得的。之所以强调目的，是因为如果没有目的这个概念，我们便无法对思维进行清晰的描述、讨论。清醒状态下，我们的思维一般是有目的地，而目的肯定是某一大脑结构的功能。

目的中枢它是一个相对宏观的概念、模型。目的中枢的功能是实现思维过程中目的地功能。目的中枢、奖惩预期中枢、状态中枢这三个根据功能编程设计的中枢是强人工智能软件(人)实现智能的核心中枢(思维的发生、发展、终止都是在它们的控制下实现的。至于它们是如何在实现智能的过程在发挥作用的，在《对编程实现拟人智能可行性的论证》这篇文章中的“思维模式的学习获得”[7]那一章节中有详细论述)。

目的是由一群目的对象组成的，而且目的的功能也是由目的对象来实现的。

对象成为目的就会被分配注意力(注意力的分配是通过对状态中枢对应记忆柱群的易化兴奋来实现的)。那些需要注意的有助于目的完成的对象，会被目的赋予动力，从而被分配注意力。注意一个对象有助于目的的完成，则注意，如果不利于目的的完成，则注意它的动力预期会下降，不注意。

思维过程中可以形成多个目的。任意两个目的之间可以是兼容的、不兼容的、或者一个目的来源于另一个目的。

2.6.1. 简单计算的目的中枢的设计

目的中枢的结构、功能，是我为了容易思考、编程，根据目的地“宏观特点”设计的。

目的中枢的结构功能的设计可以参考《对编程实现拟人智能可行性的论证》这篇文章中的目的中枢这一节的讨论，本节是在这些基础上的进一步讨论。

2.6.2. 亚主注意对象如何成为目的对象的组成

复杂一点的强人工智能软件由于一个主注意对象无法完成一次完整的奖

惩预期，而需要多个主注意对象(当时的主注意对象与前几个主注意对象)共同完成一次完整的奖惩预期(对复杂环境对象的区分、标志，往往需要多个主注意对象的共同参与。而对环境对象的“正确”奖惩预期需要能够正确区分这个对象的情况下才能够“正确”发生)，因而目的对象的确定与我在《对编程实现拟人智能可行性的论证》这篇文章中对目的对象的确定方法有所不同，(我在主注意对象、刺激传入、奖惩预期、目的对象、亚主注意对象这篇博文[8]中有论述)：可以设置一个虚拟的亚主注意对象中枢来记录亚主注意对象。亚主注意对象中枢记录的亚主注意对象数目只有几个，主注意对象及其它兴奋到一定程度的记忆柱群，在状态中枢对应的基本记忆柱群就会成为亚主注意对象，从而被亚主注意对象中枢记录，而最早成为亚主注意对象的对象，不会再被亚主注意对象中枢记录。如果主注意对象通过奖惩预期成为目的对象，这个目的对象也会包含亚主注意对象中枢所记录的亚主注意对象(见图 6)。

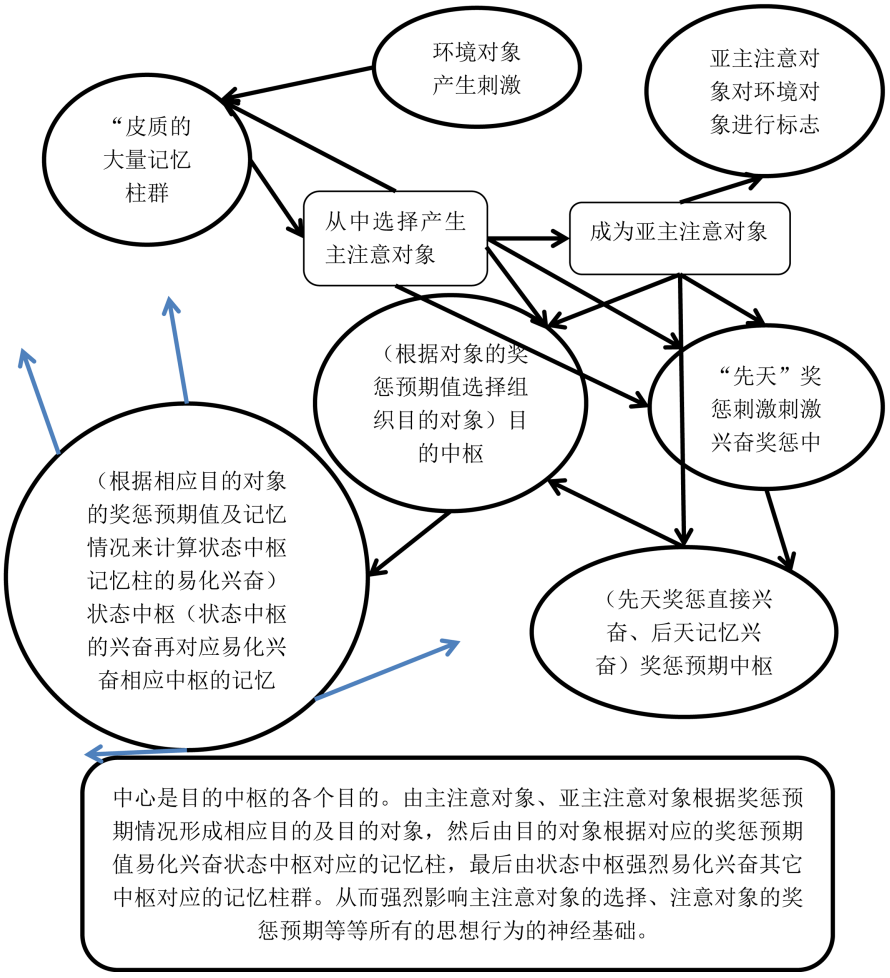


图 6. 目的在思维中的核心作用

新形成的主注意对象的动力预期的值大于动力预期值最低的一个目的对象(a)的动力预期值时，最近的几个亚主注意对象及主注意对象成为目的对象，而 a 不再成为目的对象。

在状态中枢，某一刻形成的目的对象其组成核心是由当时的主注意对象、亚主注意对象在状态中枢对应的记忆柱群所组成成分，一般是动力预期相对平稳后亚主注意对象成为新的目的对象的组成成分。

被赋值的目的对象的动力预期是短期记忆，会迅速被遗忘。而使目的对象的动力预期值下降。(我们也可以通过编程而强制在一定时间内不断降低目的对象的动力预期值，就如我编程的小程序)

当主注意目的对象的动力预期值“严重”下降，而不能成为主注意目的对象时，在这个目的下被“严重”关注注意的对象的动力预期值也会“严重”下降(思维模式下，在关注这些目的对象时，必然会影响目的的完成，从而影响它的奖惩预期值)。而不是被“严重”目的对象，其动力本来就不高，再迅速被遗忘，它们对新目的的影响较小。

2.6.3. 主注意对象、各个目的对象之间的兴奋记忆联系

在《对编程实现拟人智能可行性的论证》中的主注意对象与各个目的对象之间的兴奋记忆联系这一节有相对详细的论述。

编程设计使：强人工智能软件“皮质”中兴奋到一定程度的记忆柱群(包括主注意对象)会兴奋状态中枢对应的记忆柱群。这些兴奋的状态中枢的记忆柱群能够与主注意目的对象、亚主注意目的对象在状态中枢对应的记忆柱群(它们也兴奋，并传出兴奋)建立记忆联系，它们越是兴奋强，越是与主注意对象建立的记忆联系强。这样设置的好处是，主注意对象能够与短时间内发生的“重大”事件(高动力预期，能够成为目的)建立记忆联系(联系强弱与动力预期密切相关)。

(程序就是这样设计的，这可以看我的小程序中相关的计算)

这种目的参与的记忆联系除了使最近发生的“重要”事件之间能够建立记忆联系，还能够通过注意使发生时间间隔长的事件之间建立记忆联系。这种记忆联系的能力是各种高级功能的基础。比如：归因、奖惩预期、赋值等等。没有这种能力就不可能有高级智能。

以归因为例，这种目的记忆设计使，1) 短时间内发生的事件能够直接参与归因。2) 两个事件如果间隔时间比较长，可以通过回忆注意，使表征两个事件的兴奋在同一个主注意目的中成为目的，从而产生归因预期。

2.6.4. 通过适当设计状态中枢的功能来直接实现目的地功能

下面的讨论只是理论探讨，尚缺乏具体编程验证。

目的中枢的结构功能，是我根据宏观的思维现象设计的。关于它的实现问题，我在本章前面有论述。

我在这里向大家介绍一种更简单、“微观”的实现方法。本节主要讨论编程设计的状态中枢，是如何通过“微观”上的计算，“涌现”目的中枢的功能的。

下面通过状态中枢，设计一个能够适应复杂环境的有关目的计算模型。

相对于我前面介绍的简单目的模型，这种模型在编程上要复杂的多，虽然我描述的比较简单，但在实现上相对要麻烦的多(主要可能是编程参数的设

计上比较麻烦)。而且只有将前面的目的模型与这个模型结合起来思考，才能更好的理解这个模型。

相关规则设计

只要对状态中枢的“计算”进行一些基本的规则设计就能使它实现目的的功能。

这种设计不需要专门的设计目的中枢，只需要对状态中枢及主注意对象进行相应的设计就可以了。这种设计也能够使我们能够更好的推测人脑的结构功能。(我们人脑如何实现目的功能的？如果专门设计一个目的中枢，在人脑中什么中枢来执行目的功能的？，但如果是状态中枢来执行目的功能，目的的涌现就好理解了)

下面是设计的规则限制，及相关计算描述。

1) 状态中枢的记忆柱群的兴奋数量受到限制(可以通过扩布抑制及自身兴奋衰减来实现，扩布抑制就是一个记忆柱群兴奋的越强，它向周围传出的无差别抑制越强。我在简单编程中的“皮质”记忆柱的兴奋抑制部分有类似编程设计。扩布抑制在人脑皮质记忆柱有相关研究)。

2) 状态中枢的记忆柱群的兴奋强度，在每一个兴奋周期，都迅速衰减一个与状态中枢整体兴奋程度相关的一个数值(这种情况下，兴奋越弱的记忆柱群衰减的越快，这是因为由于扩布性抑制使每个记忆柱群兴奋衰减的值基本相同，在衰减值相同的情况下，兴奋越弱的记忆柱群，其衰减的比例越高)。

3) 状态中枢的记忆柱群兴奋的最主要影响因素是奖惩。因而，主注意对象的奖惩预期如果比较弱，它在状态中枢对应记忆柱群的兴奋强度也不会强，它(状态中枢)的兴奋就会迅速衰减。而如果它的奖惩预期强，对应的记忆柱群兴奋强，衰减的就相对较慢，它就会在更长时间内发挥目的对象的作用。(在这种模式下，皮质记忆柱群的兴奋应该有一个不应期，不包括状态中枢的记忆柱群，如果没有这个不应期，主注意目的对象可能使对应的皮质记忆柱群反复兴奋而影响思维的进行)。在这种模式下，目的及目的对象就是一个统计学意义上的目的与目的对象。而且这种兴奋必须是模糊兴奋。

4) “曾经”的主注意对象在状态中枢的兴奋会不断衰减，会与奖惩预期中枢不断建立记忆联系，可以设计使它在一个时间周期内衰减速度慢，而超过了这个时间会迅速衰减，也就是说，只有曾经成为主注意对象时间不长的对象，才能与奖惩预期中枢建立强的记忆联系。(这样设计的目的是使成为一个目的的主注意对象、亚主注意对象与动力预期中枢建立的记忆联系强度相似，从而使组成一个目的的目的对象的动力预期值差不多。而且也不需要专门设计一个亚主注意对象中枢)。(状态中枢记忆联系的强弱与兴奋的强弱，与是否成为主注意对象及不再成为主注意对象后的时间长短有关。)

5) 一个对象引起注意后，会被注意一段时间来进行奖惩预期，再加上记忆联系，多个对象记忆的奖惩预期“值”差不多(奖惩预期被赋值)。这些对象组成一个目的，它们是目的对象。按照衰减速度，这个目的的动力下降到一定程度的时候，就不能成为目的，组成它的各个对象不断被抑制而不能兴奋。在宏观上就表现为：动力预期值最弱的目的不能成为目的，而新的动力预期值大于这个目的地动力预期值的主注意对象及几个亚主注意对象成为新的目

的对象。

目的功能的实现

目的的功能我在前面有论述，本节论述状态中枢是如果在上面的规则下，实现目的的功能的(不是直接编程实现，而是计算涌现)。

目的对象这个概念只是对状态中枢某群具有相关兴奋特点的记忆柱群的定义(目的对象如何来的，前面有讨论)。

组成一个目的的一群目的对象，它们的动力应该差不多(根据目的对象动力预期记忆建立的过程可知)，因而它们在状态中枢的兴奋强度应该也差不多，它们在状态中枢被抑制不再兴奋的时间也应该差不多。这样，组成一个目的对象的一群记忆柱群，它们的兴奋程度及兴奋持续时间更应该差不多。当一个目的的一个目的对象不再兴奋后，这个目的的其它目的对象随后也会不再兴奋(因为它们的兴奋程度差不多，使它们兴奋的衰减速度也会差不多，既然组成这个目的的一个目的对象被“当时”的主注意对象及亚主注意对象代替，组成它的其它的目的对象也会被代替)……。

这个过程不断进行，其在宏观上表现为，兴奋越强的目的对象，对其它记忆柱群的兴奋易化越强，从而使兴奋能够“围绕”目的“展开”。同时，那些“旧的”动力预期值最小的目的被新的动力预期值大于它的目的“挤下”。

大家可能认为，既然奖惩预期中枢与状态中枢能够实现目的中枢的功能，那为什么我还要理论上设置一个目的中枢？首先，目的现象是思维的“宏观”现象，而奖惩预期中枢与状态中枢相互作用表现出来的特点是“微观”现象，设置目的中枢这个虚拟的功能中枢，有利于我们对思想行为的讨论。也有利于我们对状态中枢的功能的正确设置。

适当的编程设计下，目的中枢的功能可以是状态中枢在“计算”过程中涌现出来的(见图 7)。

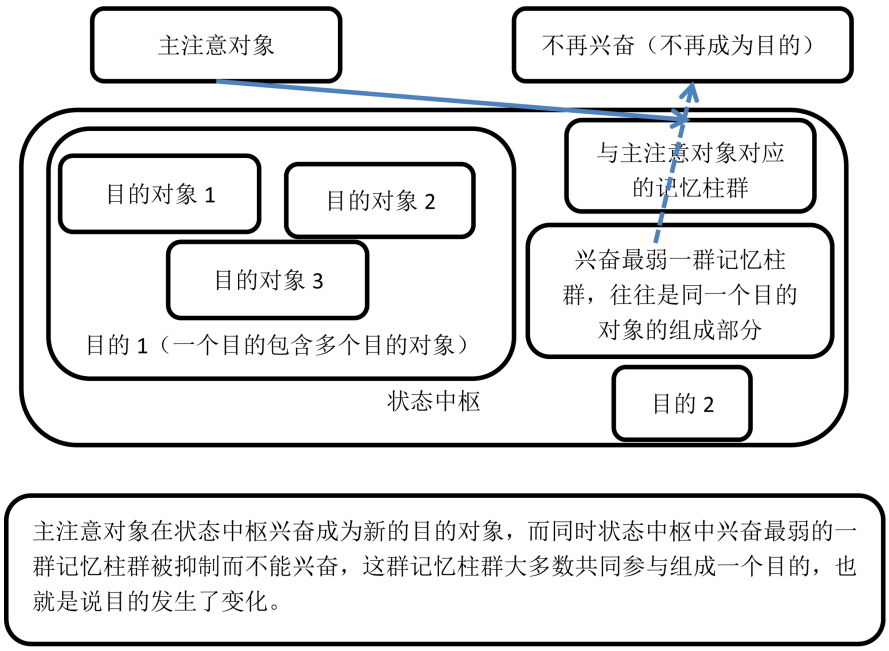


图 7. 目的模型与状态中枢

2.7. 一些重要的奖惩现象的意义及形成机理

奖惩现象，可以大概分为编程设计的基础奖惩现象，及在基础奖惩之上，通过奖惩学习获得的后天奖惩现象。重要的标准是它们对学习(包括促进不断探索学习、加快学习速度)、适应的重要程度。

2.7.1. 奖惩学习

奖惩学习的机理，我在奖惩中枢与学习[2]，对编程实现拟人智能可行性的论证[7]等等文章中都有讨论。下面只是结合编程简单的进行讨论。

对强人工智能软件来说，如果一个对象带来惩罚预期，状态中枢兴奋的记忆柱群的惩记忆柱处于更易兴奋的状态，不同主记忆柱与不同惩记忆柱，相互之间建立记忆联系(可以参考编程设计的状态中枢的基本记忆柱群、主记忆柱、惩记忆柱、赏记忆柱之间“纤维”联系，及相互之间的兴奋记忆关系)。也就是说，这时目的对象与带来惩罚预期的对象建立了惩罚(与惩记忆柱建立记忆联系，而惩记忆柱会抑制它所在记忆柱群的兴奋)记忆联系。带来惩罚预期的对象或者过程会在强人工智能软件追求目的的过程中被相对抑制(目的对象会通过它们的主记忆柱兴奋与带来惩罚相对应的基本记忆柱群的惩记忆柱，从而抑制这个基本记忆柱群)。如图 8 惩罚学习奖赏学习的机理相似。

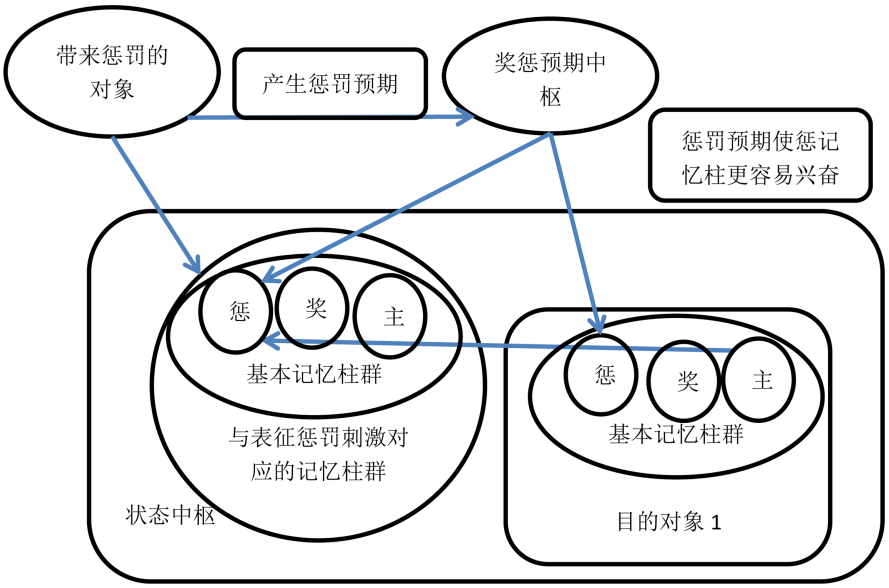


图 8. 奖惩学习

最终的后果是，那些有利于目的实现(获得奖赏)的兴奋会被强化，今后在追求目的地过程会更容易被兴奋，而不利于目的完成的兴奋会被抑制，……。

2.7.2. 奖惩预期

奖惩预期的计算基础与标准来源于奖惩，而为了更好的适应环境，“先天”的固有的奖惩，其值的大小是随内外环境的变化而变化的，比如饥饿与食物(饥饿感越强，所获得的惩罚越强，对应的惩罚预期越强。饥饿感越强，同一个食物所能够产生的奖惩预期越强，获得这个食物的动力越强)，编程设

计的奖惩的变化应该参考这种变化。

奖惩预期是记忆的奖惩预期，是条件记忆、遗忘下的奖惩预期，一个对象的奖惩预期值，会随遗忘而不断改变。奖惩预期是环境、目的状态下的奖惩预期，对象的奖惩预期值随环境、目的地变化会不断的变化，也就是说对象的奖惩预期值会随环境、目的、注意、条件等等的变化而不断的变化。

对智能实体来说，奖惩预期是一种主观预期，主观预期的根本是奖惩刺激产生的奖惩，是适应环境。主观预期并不需要完全客观的表征奖惩，它只需要对客观奖惩的表征能够带来所需要的奖赏就可以了。奖惩预期是情景预期(状态预期，状态预期的各个组成条件都是主观的。奖惩预期的过程是对环境状态模拟的过程，在这种模拟的过程中，模拟会受到智能实体所在环境刺激状态、“大脑”的兴奋状态等等多种因素的影响。因而模拟只能是部分客观的模拟)。因而，要完全客观的表征奖惩，是不可能的，也是没有必要的。为了完全客观的表征奖惩，我曾经设计了大量的奖惩预期模型，但始终无法完全客观的表征奖惩。(在这里对它强调的原因是，如果大家不明白这个道理，就必然会在奖惩预期上浪费大量的时间)。同时大家也会发现，在理论设计、思考时，我们要想完全客观的表征奖惩，许多与环境相适应的奖惩功能便不能实现。

奖惩预期是环境、目的状态下的奖惩预期，是对奖惩的模拟预期，是经验奖惩预期，是主观奖惩预期，是学习、适应下的不断发展变化的奖惩预期，是与环境相适应的奖惩预期，是思维状态下的奖惩预期。比如，我在特殊记忆柱群的奖惩学习这篇博文中讨论了预期奖赏没有获得的情况下，婴儿(强人工智能软件)是如何产生惩罚预期的。

认识到奖惩预期是不断变化的奖惩预期，并且不执着于奖惩预期“完全真实”的反应客观奖惩，对我们进行奖惩预期的理论讨论及奖惩预期的编程设计是至关重要的，可以大大简化我们对奖惩预期模型的设计。

2.8. 动力预期值计算的应用

动力预期值的计算模型遵循适应的原则，即什么样的计算模型有利于强人工智能软件适应内外环境(包括兴奋、思维等等)，就可以采取什么样的模型。

动力预期值的大小可以根据奖惩预期中枢的兴奋强度计算获得的(强度中枢可以参与)。

动力预期计算的主要意义是目的的确定及对象注意力的分配。注意力分配的关键是各个目的对象动力预期值的大小。

在环境、目的状态下，被计算的对象其动力预期值有几个重要的参数：基础动力预期值；对象兴奋时的动力预期值；对象兴奋前的动力预期值(可以参与计算动力预期改变速率)；对象兴奋时，动力预期值的改变速率平稳时的动力预期值；对象兴奋时，动力预期值的改变较小，可以忽略时的动力预期值。可以根据上面的几个参数来计算对象的动力预期值的各种属性，来调控强人工智能的思想行为，使之与环境相适应。

比较动力预期值的功能应该是亚强度中枢的作用。兴奋的强度(动力预期值)变化平稳后(可以是动力预期值，也可以是动力预期值的变化速率，它们会被记忆下来，为后面的计算做准备)，就会对对象兴奋前后的动力预期值(变化平稳)进行比较，(包括主注意对象动力预期的兴奋强度的比较，主注意目的对象动力预期兴奋强度的比较)。

兴奋强度改变，相关结构(特别强度中枢)产生相应强度的兴奋，从而会产生相关计算。主注意对象奖惩预期的兴奋强度(值)改变后会兴奋相应结构，兴奋平稳后会兴奋相应结构(普通强度中枢之上的结构)，兴奋平稳改变后会兴奋相应结构。兴奋了相应结构就会产生计算、记忆，然后按一定的规则计算兴奋。

对象的动力预期值就是强人工智能软件或者人注意对象时动力改变平稳后的动力预期值(总体的)减去注意对象前的动力预期值，所获得的值就是“当时”这个对象产生的动力预期值。这个计算获得的值是状态性的，它既是主观的，也是客观的。(可以看强度中枢、奖惩预期中枢、奖惩预期这一节)

计算对象动力预期值的方法多样，比较简单的方法是：在动力预期值迅速改变后，会逐渐平稳改变，当改变值在弱的范围内平稳时，可以根据前后奖惩预期值按一定规则来计算获得。

然后再根据计算获得的对象的动力预期值来进行注意力的分配(对状态中枢的兴奋)。这样能够迅速对环境刺激做出反应，有利于环境的适应。在一定时间内其动力分配情况会受到随后的主注意对象动力预期的动力预期值及注意力分配的影响(通过在奖惩预期中枢及状态中枢建立记忆联系来影响)

主注意对象引起的动力预期值的改变速率超过了一定值(动力预期值改变的速率越快，对象的动力预期值就越大，)，可以直接诱发注意，这样能够使强人工智能迅速注意“重要”的对象。(相对于这种注意模式，被注意对象在动力预期改变平稳后再引发注意力的分配显然要慢几个主注意对象兴奋的周期)

通过动力预期中枢计算获得的当时动力预期值，主注意对象兴奋前的动力预期值，这两个基础参数受到长期奖惩实践所获得的奖惩经验的综合影响(经验的奖惩预期)。

2.9. 动力预期的概率问题

环境下，一个目的对象被实现的概率的预期，不需要专门的计算(只需要计算与之相关中枢的兴奋强度并综合起来就可以了)。从上面的讨论可知：奖惩预期的计算是各个亚奖惩中枢的兴奋强度的综合计算。这种计算模式应该能够大概反应目的实现的概率，因为每次目的实现与不实现都会相应兴奋对应的亚奖惩中枢的联络区及奖惩预期中枢的联络区，并产生经验记忆，这种记忆的“强度”与记忆强化的次数密切相关。今后再进行类似奖惩预期时，以前实现目的与没有实现目的的奖惩经验都会发生一定的影响，从而部分反应目的实现的概率。只要这样的部分反应在实践中能够带来奖赏就可以了，而且这样的预期可以通过学习不断得到完善。

要使对对象动力预期的结果与环境相适应，需要通过奖惩学习不断调整动力预期的计算值使最后的计算结果与环境相适应。

强人工智能软件在某些情况下会预期目的无法完成，或者完成的可能性小，动力预期就会下降。“某些情况”是奖惩学习的结果，强人工智能软件会在一定时间及努力范围内预期一个目的无法完成，这个时间长短及努力范围是通过奖惩学习获得的。

比如在某一环境条件下，强人工智能软件在追求目的达到一定程度后(a)，会认为目的难于完成，其动力预期会急剧下降到一定程度。但如果这时有一个可信之人说你在坚持一下就能够完成目的，强人工智能软件坚持一会达到一个新的程度(b)后就真的完成了目的(程度 b 就会与奖赏建立记忆联系，而程度 a 如果放弃，就会与惩罚建立记忆联系，如果继续追求就会与奖赏建立记忆联系)。这样多进行几次后，当强人工智能软件在追求目的达到一定程度后(a)，不会认为目的难于完成，其动力预期值也不会急剧下降，它就会继续追求目的。(这也说明奖惩预期是一种主观预期)

2.10. 与环境相适应的奖惩预期

才设计出的还没有学习过的强人工智能软件，因为泛化及模糊兴奋，它的奖惩预期会有下面的特点。

泛化使兴奋记忆联系具有无限可能性，而奖惩学习选择(分化)使奖惩预期、思想行为与环境相适应。

为了论述的方便，我在下面以婴儿为例进行讨论。

当一个对象给婴儿带来奖赏后，婴儿可能看到或者想到这个对象就会预期奖赏，从而追求这个对象，但这种追求往往无法达到目的，通过长期奖惩学习，他就不会有这种奖惩预期及相应的思想行为了。

模拟奖惩预期能够使强人工智能软件对奖惩对象的追求作出正确的选择。因为它们都与目的对象、目的对象兴奋下的奖惩预期中枢建立了记忆联系，记忆联系的强弱与获得的奖赏成大概正比的关系，带来的结果是环境、目的下，什么对象产生的奖惩强，在相似环境目的下对它的奖惩预期也会强，目的追求的是奖惩预期强的对象，也是能够带来奖惩强的对象。强人工智能软件总是追求更能带来奖赏的对象，是正确的选择。

2.10.1. 丢失奖赏对象、逃避惩罚

奖惩是对立统一的，获得某些惩罚是因为缺乏对应奖赏对象的刺激，丢失了相关奖赏对象的刺激相当于获得了惩罚，而逃避了某些惩罚，相当于获得了对应奖赏对象的刺激，获得了奖赏，智能实体在奖惩实践过程中获得这些奖惩后，就会产生记忆，以这些记忆为基础，会产生一系列的奖惩现象。泛化与模糊兴奋是下面所述记忆联系现象通过奖惩学习建立的基础。

失去与获得的奖惩经验记忆，是失去与获得的奖惩预期的基础。人或者强人工智能软件丢失奖赏对象，通过经验奖惩预期计算(拥有这个对象会有相应的奖惩体验，并有相应的奖惩预期，通过记忆联系，在不拥有这个对象的情况下也有相应的体验与预期，两次的奖惩预期值比较会使相应的强度中枢

及奖惩预期中枢产生相应的兴奋。比如饥饿时不拥有食物，可以有惩罚体验，而这种惩罚体验可以与“计算”获得的奖惩预期差值建立记忆联系。今后只要计算产生这种差值便能够兴奋与这种惩罚体验对应的奖惩预期)，使相应的惩罚中枢兴奋，并产生相应的行为，而如果这种思想行为有利于获得奖赏，适应环境，就会得到强化。获得能够带来奖赏的对象，逃避或者获得能够带来惩罚的对象的机理相似。

简单的说：一些先天的奖惩之间的某些关系(这里指获得惩罚是因为缺乏奖赏刺激，获得了奖赏就相当于逃避了惩罚刺激)发生后会产生一系列记忆联系，其它不具有这种关系的奖惩，可以通过共同的神经通路(相同的某些属性)兴奋相关的记忆联系，而具有这种关系，而如果这种关系有利于目的的完成，就会得到强化。

这些经验的学习、记忆、强化是复杂的奖惩学习过程(可以从人类的婴儿时期开始讨论失去与获得的各种奖惩、情绪经验体验)。通过长期的奖惩学习，最终与丢失奖赏对象、逃避惩罚对应的思想行为、“神经内分泌”都是有利于获得奖赏，适应环境的。

强人工智能软件认为可以获得一个对象，但追求时又得不到这个对象，某种程度上，相当于拥有这个对象，又失去这个可能获得的对象(它们计算的奖惩预期值的差都相似，这种相似性可以使它们产生相似的兴奋过程，而如果相似的兴奋过程能够带来奖赏便能够被强化确认)，从而产生相应的奖惩预期，并会产生相应的思想行为、“神经内分泌”。

2.10.2. 目的完成与不能完成现象

特殊记忆柱群的奖惩学习这篇博文讨论了，强人工智能软件(或者人)如果无法完成目的会获得惩罚(可以通过经验学习获得)。有利于目的完成的对象、思想行为(它们在奖惩预期时会被标志并与奖惩预期建立记忆联系——赋值)，会被奖赏预期赋值(见动力预期的赋值这篇博文)，而不利于目的完成的对象、思想行为过程，会被惩罚预期赋值。奖赏会被强化，惩罚会被抑制(通过状态中枢来实现)。这样通过长期的奖惩学习我们及智能机器人的模式化的思想行为过程一般就会是有利于目的完成的模式。

目的不能完成相当于丢失了奖赏对象(既然有目的，就表示认为有可能完成目的获得奖赏，目的最终没有完成，相当于本来应该得到的奖赏，结果没有得到)，它们会产生相对应的奖惩预期。

强人工智能软件(婴儿)在想象性回忆的情况下，回忆起某一主注意对象，从而产生奖惩预期(比如饥饿时想到食物，如果编程设计使强人工智能软件像人一样有饥饿，在饥饿的情况下获得食物会获得奖赏。之所以这样讨论主要是为了便于大家的理解及我的论述，这样更贴近我们的经验)，其奖惩预期的大小是获得食物的动力大小(才开始学习时的情况)。(整个过程是理想化的、确定性论述)

强人工智能软件在饥饿(编程设计)状态下，饥饿到一定程度的饥饿刺激成为目的对象。它在饥饿刺激时可能会回忆起香蕉，产生奖惩预期(有获得并吃香蕉经验。在吃香蕉的情况下，强人工智能软件获得奖赏，并逃避了饥饿

这一惩罚，其奖惩预期中枢相应兴奋，香蕉就与其建立了兴奋性记忆联系，再回忆到香蕉时通过记忆就能相应兴奋奖惩预期中枢，从而产生奖惩预期)，获得香蕉以及标志想象性回忆的部分记忆柱群成为主注意目的对象，强人工智能软件会根据经验不断追求香蕉的获得。在追求香蕉的过程中，由于现实中不存在香蕉，主注意目的对象的动力不断下降(追求过程中不断会有各种惩罚影响动力预期，同时随时间的流逝，动力预期值也会自动的有一定的下降)，追求香蕉的过程中，饥饿刺激会时不时成为主注意对象，它会更强的兴奋惩罚中枢，并产生动力预期。在状态中枢，惩罚中枢的“强烈”兴奋使与获得香蕉及标志想象性回忆相对应的记忆柱群(主注意目的对象)的惩记忆柱强烈兴奋，并相互建立记忆联系。这样，今后，标志想象性回忆(回忆起的对象在当时现实环境中不存在)对应的记忆柱群兴奋后，就会通过记忆，兴奋那些状态中枢中，与获得香蕉对应的奖惩预期，相对应的记忆柱群的惩罚记忆柱，从而抑制对奖赏预期中枢相应记忆柱群的兴奋，而使动力预期下降。

也就是说，通过这样的奖惩学习：1) 想象性回忆的情况下，回忆到香蕉时产生的奖惩预期值的大小，会大幅下降。通过长期的奖惩学习，最终，想象性回忆回忆起某一能够带来奖赏的对象时，一般情况下，强人工智能软件并不会再产生追求这一对象的动力。2) 追求获得香蕉不得的过程中，会不断产生惩罚记忆。同时，预期能够获得一个对象从而获得奖赏，结果这个对象因为“丢失”或者其它原因，而不能获得这个对象，就会产生惩罚体验，并产生惩罚预期，并不断记忆下来。

通过长期奖惩学习，强人工智能软件会获得下面的能力：1) 想象性回忆会一般成为不可信的思维模式。2) 追求一个目的不成，会产生相应的惩罚预期及惩罚体验。3) 预期可以获得的某一奖赏，通过实践但最终却没有获得这一奖赏，会通过记忆的奖惩学习记忆，产生一定的惩罚预期。

强人工智能软件追求奖赏不得，而获得惩罚的功能。有两种获得方法，一是设计获得，二是学习获得。都有可能，但不管怎样，必然能够获得。设计获得这样的功能这里就不讨论了。

上面讨论的是不可信思维模式是如何对思想行为产生影响的。

2.11. 奖赏与惩罚现象，强化，中介奖惩，好奇心等等

中介奖惩对早期“基础思想行为”的自主学习非常重要，相关讨论见《对编程实现拟人智能可行性的论证》这篇文章中的中介奖惩学习及视觉系统这两节。它们主要讨论了，强人工智能软件的运动中枢通过编程设计的“神经兴奋模式”自主兴奋，组合，如果能够带来“正确”的运动，就会产生强的刺激传入，从而获得强的中介奖惩并产生记忆。适当条件下，强人工智能软件就会追求这种运动，从而不断强化运动模式，最终获得能够使各种运动模式正常产生的基本运动。

强人工智能软件对新奇刺激的注意可以获得更多更强的中介奖惩刺激，同时它对新奇刺激的注意有利于获得新的知识与技能，对获得奖赏有利，从而使它对新奇刺激的追求、注意不断得到强化。新奇刺激的特点是：在去除其它奖惩的情况下，这种刺激产生的中介奖惩强度高于其它普通刺激的产生

的中介奖惩刺激，这种刺激能够被标志，并强化。(相对普通刺激，新奇刺激对应的记忆柱群之间由于缺乏兴奋记忆联系，因而记忆柱群之间要发生兴奋转移需要更强的兴奋，这样就会产生更强的中介奖惩刺激)

获得奖赏的思想行为被强化，获得惩罚的思想行为被抑制。最终的结果是环境、目的下的思想行为过程基本与环境相适应。其特点是，见《特殊记忆柱群的奖惩学习》。

短期与长期记忆现象，赋值，更多的赋值是短期赋值，遗忘

思维过程中动力预期的赋值：

当主注意对象兴奋奖惩预期中枢后，其它目的对象就能够与兴奋的奖惩预期中枢建立一定程度的记忆联系，这种记忆联系主要发生在状态中枢，是短期记忆。这种联系就是我说的赋值。短期内，通过一个对象回忆到这里的某个目的对象(回忆过程是可信的)，就能够根据记忆兴奋奖惩预期中枢，产生动力预期，这个对象就会与奖惩预期中枢建立短期的记忆联系，从而对这个对象完成赋值。这种短期记忆能力是我们的思维能够与环境相适应的基础。比如：(我们可以通过编程设计使香蕉能够使强人工智能软件获得奖赏)强人工智能软件推开门，就能够获得香蕉，从而获得奖赏。回忆预期到推开门后会获得香蕉，产生奖惩预期，预期到推开门能够获得奖赏，这里获得香蕉是主注意目的，推开门是亚主注意目的，进屋到香蕉旁是亚主注意目的，获得香蕉完成奖惩预期(使奖惩预期中枢相应的记忆柱群及与这些记忆柱群对应的状态中枢的相应记忆柱群兴奋)。这个奖惩预期对应的记忆柱群在状态中枢与推开门对应的记忆柱群就能够建立短期的记忆联系。再由推开门进行回忆，就能够由推开门对应的状态中枢的记忆柱群，根据短期记忆，兴奋奖惩预期中枢相应的记忆柱群，从而产生奖惩预期(这个奖惩预期值就是获得香蕉产生的奖惩预期值)，赋值完成。

这里必须是短期记忆，能够被迅速遗忘，否则奖惩预期系统会崩溃的。比如推开门被“获得香蕉”产生的动力赋值，如果不能被迅速遗忘……。动力预期的赋值，是思维“高级化”、复杂化的必要条件。

强人工智能软件的思想行为需要动力，最初，它有获得香蕉的动力，没有推开门的动力(不会追求推开门)，当它预期到推开门能够获得香蕉时，便将获得香蕉的的动力部分赋值给推开门，而当推开门被赋值后，它便有了推开门的动力，便会追求如何推开门(目的)。这种赋值的能力对人或者强人工智能软件的思维是至关重要的。(长期奖惩学习后，赋值模式是一种可信的思维模式)

2.12. 短期记忆、长期记忆参与产生的一些重要的奖惩预期现象

奖惩预期是学习的奖惩预期。奖惩、奖惩预期的记忆首先是短期记忆，然后不断强化，才发展成长期记忆。

奖赏记忆消退后自发回复

短期与长期记忆、遗忘相互作用，会产生了一个重要的奖惩现象：一个对象如果往往能够带来奖赏，动物会追求这个对象的实现。但最近动物追求这个对象往往会带来惩罚，这一追求行为会被抑制，而不再发生，而过段时

候后，动物又会再追求这个对象的实现。其原因是短期惩罚记忆在短期内占据优势，使对对象的奖惩预期总体为负，行为被抑制。但动物在一定时间内不再追求这个对象后，短期的奖惩记忆会迅速遗忘，而长期记忆遗忘的较慢，短期的惩罚学习带来的长期惩罚记忆相对不强。所以这个追求事件在一段时间没有发生后，对象带来奖赏的动力预期会重新大于惩罚，从而使动物会再追求这个对象的实现(见图 9)。

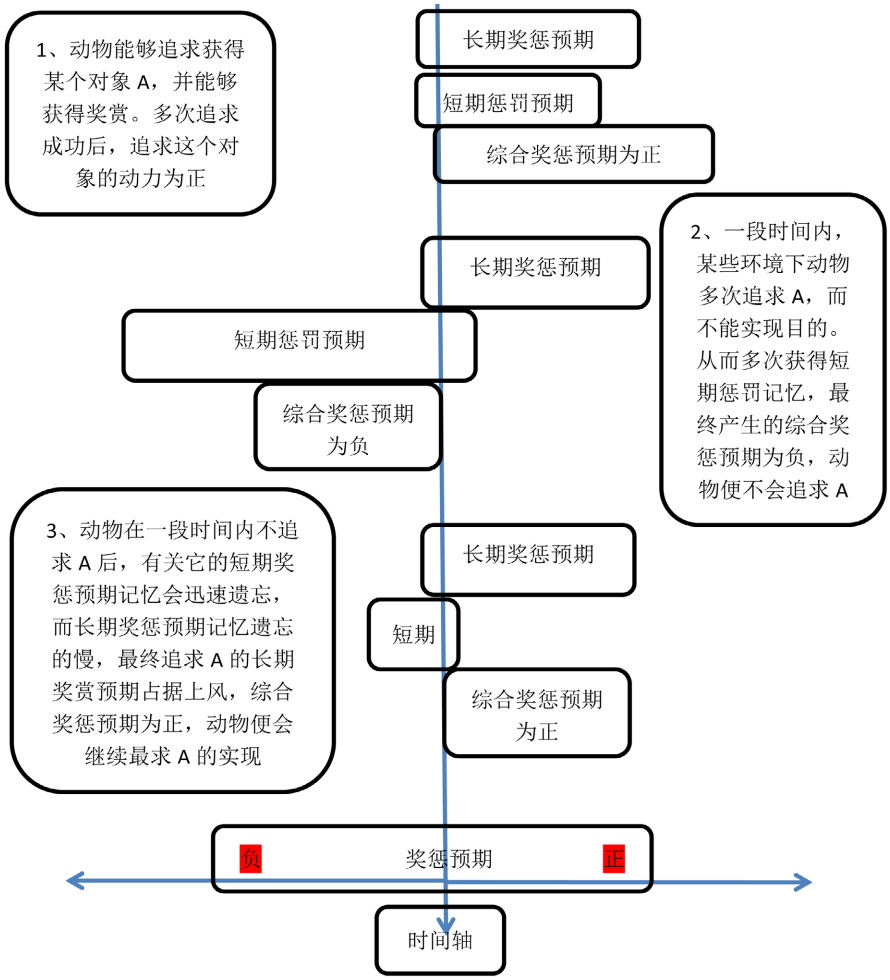


图 9. 短期、长期记忆与奖惩预期

这种现象是缺乏强化的结果。而且往往主要是第一种类型(奖惩中枢、奖惩预期中枢这一节)的奖惩预期，两种类型奖惩预期的记忆联系都下降，而第二种下降的更快、更多。

这种现象对学习非常重要。有了这种现象，动物就不会因为一个对象带来的短期的惩罚经验而不会再追求这个对象。也就是说动物不会因为某一次或者某一段时间的惩罚经验，而放弃这种奖惩实践。不断的奖惩实践是动物获得新知识、新技能适应环境的重要手段。

动物追求一个对象能否实现与环境条件密切相关，而动物在奖惩预期时模拟的环境条件并不一定与“真实”环境条件完全一致。因而，动物在一些

环境条件下，追求一个对象的目的能够达到，在另一些相似的环境条件下不能够达到，但动物可能并不能区分这些环境。也就是说，动物奖惩预期时，可能并不能正确的模拟重要的环境条件，而相对正确的模拟，往往需要长期的奖惩实践。也就是说早期学习阶段，对于复杂环境下产生的各种奖惩事件，动物经常无法“正确”奖惩预期及归因(有时认为能够达成目的，但不能达成，而不能达成目的原因，并不是人或者动物认为的原因)，在可能经常无法达到目的并且归因错误的情况下。在以上情况下，如果没有这种奖惩预期现象，动物会迅速丧失探索环境的欲望，这对动物的生存不利。

实际上这种奖惩预期现象产生的原因比我在这里论述的更复杂：首先，在各种环境下，不能带来奖赏的思想行为、对象，在动物的实践过程中，会更少的引起注意，即使注意，动力的绝对值也不高，也就是说，它们得到的强化机会少，即使强化，强化的也不强，相对来说更容易遗忘。而且有实验证明有专门的基因调控加快某些惩罚记忆的遗忘[9] (其意义可能使生物更愿意探索环境，不会因为无法经常获得奖赏达到目的，而丧失探索环境追求目的的欲望)。

2.13. 归因现象

归因就是使对象与奖惩对象建立特殊的可信的记忆联系。人或者强人工智能之所以能够归因，是因为他们能通过注意使任何多个对象及奖惩之间建立记忆联系。这种记忆联系如果有利于获得奖赏，就会被强化。从而使正确的归因方式被强化。归因有利于正确奖惩预期，有利于对环境的适应。

归因能力是随奖惩实践学习不断发展的。归因能力越强强人工智能软件正确奖惩预期的能力也会越强，它们对环境适应的能力也会越强。

一些重要的奖惩现象还包括探索及好奇心(本文及《对编程实现拟人智能可行性的论证》都有讨论)、模仿能力(《镜像功能的形成机理》[10]、《对编程实现拟人智能可行性的论证》有讨论)、可信对象及权威对象对思想行为的影响能力、及其它一些新的欲望等等的获得与发展(《对编程实现拟人智能可行性的论证》有讨论)……。

3. 奖惩系统与思维的关系

思维的目的就是为了获得奖赏逃避惩罚。奖惩塑造思维。奖惩预期控制所有思维的发生、发展、终结。

3.1. 学习获得的各种思维模式，有利于思维正确完成

(大家可以结合具体的思维过程)当强人工智能软件在一目的下时，产生一个思维过程后，产生目的无法完成的预期，就会产生相应的惩罚预期(来源于奖惩经验)，而使在这个思维过程中，相关的对象，在目的下，短期与惩罚预期中枢建立了记忆联系，强人工智能软件短期内再在这个目的下时，这个思维过程就会受到抑制。这样的后果是，通过长期的奖惩学习，强人工智能软件思维过程中产生的各种思维过程(包括模式)，一般都应该是有利于思维

正确完成的思维过程。

比如推开门获得香蕉那个例子中，强人工智能软件在思维的过程中，开始可能会思维：推开墙上的窗户能否获得香蕉，但通过思考预期到不能获得香蕉。一般情况下，在后面思维时，与推开墙上的窗户相关的某些思维内容便会在这一目的下受到抑制。

3.2. 学习

强人工智能软件学习的最终目的是适应环境，这需要其神经系统的“自主”兴奋能够适应内外环境，因而在学习的早期，其神经系统的兴奋能力应该受到限制(就如人类婴儿早期的神经系统)。这样就能够减少其自发的兴奋，而使其兴奋往往是在内外环境的刺激下产生，并接受奖惩系统的选择。(相关讨论可以看我的《拟人智能的实现》这本书的，关于为何要那样设置主注意对象的章节)

强人工智能软件的奖惩系统，应使强人工智能软件具有基本的奖惩学习能力，同时具有不断的自主探索、学习适应环境的能力。强人工智能软件要具有不断的自主探索、学习适应环境的能力，就需要设计中介奖惩学习(使强人工智能软件能够不断的追求新奇刺激)、短期长期奖惩学习(使强人工智能软件不会因为“暂时”的惩罚而失去追求的欲望，没有它中介奖惩学习的作用也会有限)。

在强人工智能软件追求奖赏的过程中，就会学习获得对思想行为至关重要的能力及与之相关的奖惩现象。比如思想行为过程中，兴奋的都是有利于思想行为完成的兴奋。目的完成与不能完成、丢失产生的奖惩。

强人工智能在不断的探索学习的过程中，其学习能力不断增强。归因能力(能迅速归因，无疑会较少为了正确归因而探索的时间)、模仿能力(能够通过模仿他人的思想行为，而迅速获得正确思想行为的能力)、交流能力，文化指导(使他人的实践经验成为自己的经验……)等等[7]，会不断加快正确学习的速度，从而不断提高强人工智能软件对环境的适应能力。

3.3. 编程设计的结构功能与人脑的结构功能的对应的假设

关于奖惩中枢，神经科学界研究的比较多，它们主要位于中脑的一些核团中，那里核团的功能一般与基础的生理相关，它们涉及到最基本、简单的奖惩学习适应。

奖惩预期中枢发挥着奖惩预期的功能，人在思维过程中奖惩预期起到统筹的作用，它与其它各个中枢存在密切的联系，它应该是最“忙碌”的中枢之一(这些功能联系对它的纤维联系及体积有要求)。符合这些特点的中枢，我感觉应该是额叶及其相关中枢。因而，我认为奖惩预期的功能可能主要由额叶及其相关中枢来实现。奖惩预期中枢涉及到高级复杂的学习。

对于强度中枢，当今神经科学界没有相关的研究(是指感知兴奋强度的功能)，根据其功能可知，它与各个中枢联系密切(这对它的体积有要求)，特别是与奖惩预期中枢的兴奋联系，在思维过程中，它应该也是最为忙碌的中枢

之一。根据描述的强度中枢的纤维联系及兴奋特点，我认为强度中枢的一些功能可能主要由扣带回来实现。之所以对思维无比重要的一个中枢，神经科学界却没有什么研究，我觉得是因为思维理论研究的限制。

至于状态中枢，根据其兴奋特点及它在短期及长期记忆中的作用，我认为，状态中枢的功能应该主要是由海马中枢来实现[6]。

4. 小结

我模拟人脑的功能设置的奖惩系统主要包括：奖惩中枢、奖惩预期中枢、状态中枢、强度中枢。奖惩中枢、奖惩预期中枢、状态中枢。这几个中枢的结构功能、编程在《拟人智能的实现》这本书中有详细的讨论，而对强度中枢这一关键中枢的讨论较少，并缺少相关的编程设计。本文除了更进一步讨论了这些中枢的结构功能的设置外，还详细的讨论了强度中枢的结构、功能，及它在思维过程中重要的作用(也为强度中枢的编程提供了理论依据)。

强度中枢通过“感知计算”各个中枢局部及整体的各种结构的兴奋特点来参与奖惩预期的计算，从而直接参与了注意、奖惩预期、注意力的分配、目的的形成、思维的形成、持续、转换等等重要功能的实现。

在以上结构功能的基础上，强人工智能软件在奖惩实践过程中会形成一些重要的奖惩现象。这些奖惩现象对学习适应环境具有重要的作用。

对于才编程获得的强人工智能软件，“大脑”各个结构之间虽然由于泛化的原因存在“无限”联系的可能，但它们之间并不存在兴奋记忆联系，经验奖惩学习使它们之间建立起特殊的兴奋记忆联系，并使强人工智能软件的思想行为与环境相适应。

兴奋记忆要与环境相适应，就需要不断感知环境，并根据感知不断调节兴奋以适应环境。智能软件需要有不断探索环境刺激的欲望，以学习适应环境。

中介奖惩、好奇心、奖赏记忆消退后自发回复、……，这些奖惩现象使奖惩智能软件能够不断的自主探索新奇(未知)刺激，同时不会因为各种偶然的惩罚而迅速丧失探索环境的欲望。

在不断的探索各种环境刺激的过程中，智能软件不断的塑造自己的思想行为，使它们与环境相适应。最终，通过长期奖惩学习，使思想行为都是按适应环境的模式进行。

各种与环境相适应的兴奋模式有：思维过程的思维模式使思维能够“正确”的进行下去、注意模式、兴奋记忆模式、回忆模式等等，它们都是通过奖惩学习获得的。

理论上，随着各种思想行为模式的不断形成与发展优化，强人工智能软件的各种能力会不断发展增强，使它的学习能力与速度，及对环境的迅速适应能力不断增强。比如：注意能力、兴奋记忆能力、回忆能力、模仿能力、交互能力等等。

最终，强人工智能软件会从神经通路的兴奋到思想行为的发生都是与环境刺激相适应的。

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Cheng, H.W. (2007) Relationship between Reward and Punishment Nerve Centers and Learning. *Mind and Computation*, **1**, 208-212.

[2] 程洪文. 拟人智能的实现[M]. 武汉: 汉斯出版社, 2023. https://www.hanspub.org/books/bookmanage?BookID=241

[3] 程洪文. 获得性遗传与细胞的结构功能[J]. 生物医学, 2024, 14(1): 37-47. https://doi.org/10.12677/HJBM.2024.141004

[4] 程洪文. 特殊记忆柱群的奖惩学习[EB/OL]. https://blog.csdn.net/chenghwn/article/details/139699577?spm=1001.2014.3001.5502, 2024-06-15.

[5] 程洪文. 简化的证明小程序[EB/OL]. http://bbs.bioguider.com/home-space-uid-68-do-blog-id-8200.html, 2020-12-22.

[6] Cheng, H.W. (2010) System Discussion of Attention Problems. *Mind and Computation*, **4**, 216-227.

[7] 程洪文. 对编程实现拟人智能可行性的论证[EB/OL]. 汉斯预印本, 2022, 7(1): 1-39. https://www.hanspub.org/journal/paperinformation?paperid=58850

[8] 程洪文. 主注意对象、刺激传入、奖惩预期、目的对象、亚主注意对象[EB/OL]. https://blog.csdn.net/chenghwn/article/details/141438635?spm=1001.2014.3001.5502, 2024-08-22.

[9] Yang, Q., *et al.* (2023) Spontaneous Recovery of Reward Memory through Active Forgetting of Extinction Memory. *Current Biology*, **33**, 838-848.E3. https://doi.org/10.1016/j.cub.2023.01.022

[10] 程洪文. 镜像功能的形成机理[J]. 心智与计算, 2009, 3(1): 46-50.

Appendix (Abstract and Keywords in Chinese)

强人工智能软件、人的奖惩系统(强化学习系统)

摘要：奖惩系统是人具有智能的根本原因之一，强人工智能软件要具有真正的智能，必须具有适当的奖惩系统。本文从编程与拟人角度讨论了强人工智能的奖惩系统的结构功能及一些重要的奖惩现象。并在此基础上，讨论了，根据我的理论编程获得的奖惩系统是如何在目的下处理刺激信息，如何通过奖惩学习使强人工智能软件的思想行为与环境相适应，并加快提高它学习适应环境的速度、能力。

关键词：强人工智能，状态中枢，奖惩预期，目的中枢，强度中枢，思维模式