



第9章 展望

第26讲 智能哲学



第26讲 智能哲学

面对逻辑计算的局限性，人类的心智活动到底能不能归结为计算过程呢？显然对这一问题的回答，不仅在于我们对人类心脑机制本身的了解程度，也在于我们是如何理解计算这一尚无严格定义的事物的。

但在这一节中，我们仅讨论基于逻辑计算（即把计算定格在丘奇-图灵论题的意义上）人类的心智能不能归结为计算的回答。而在下一节中，我们再突破这一局限性，讨论更广泛意义的心智计算问题。



1、心智能否被计算

基于逻辑计算的心智实现，有一种观点的回答是肯定的，强调我们的**心智活动**不过是一种**信息加工过程**，因此从根本上讲，随着技术的发展，完全是**能够用逻辑计算方法来实现的**。

特别是如果计算便意味着表象和算法，因此只要**心智内容**可以用**形式化语言来描写**，而**心智过程**可以用**形式化算法来描述**，那么就可以将**心智活动**归结为计算的。



1、心智能否被计算

也许问题要比想象的要稍微复杂一点，对心智内容需要用一种以上的形式化语言或类似符号系统来描写，而心智活动的不同部分也需要以不用媒介进行计算，但这不是实质性。根据这样的观点，包括意识在内的心智是可以看作是一种计算模型，有一套程序或一组规则，类似于控制机器的规则来支配其活动。



1、心智能否被计算

特别是从微观上讲，你可以将心与脑的关系类比到软硬件关系之上，而更为序列式的操作系统运算，可以像意识活动一样，对所有“感觉”、“认知”及“运动”进程进行全局控制。这样如果机器的硬件具有神经物质一样的运转机制的话，那么就没有理由说，机器的软件就一定不能具有心智和意识的功能。



1、心智能否被计算

特别是，鉴于神经细胞种数、个数以及连接方式、电脉冲通信方式等均为有限，其原理就类似于组合有限的**电位脉冲反应**。这就意味着，可以用**形式化符号系统**来对大脑神经网络进行编码刻画，**不仅**给出**大脑状态**的形式描写，**也**给出**大脑状态变化**的形式描述。

于是，假若我们能够造出在**量级**上与神经系统一样复杂的机器，那么凭什么说，大脑能够具备的功能，机器就不能够具备呢？



1、心智能否被计算

实际上，从某种角度上讲，我们所有关于描述或形容人的心理状态的言辞都不过是在区分脑活动中出现的不同神经网络模式状态。而 10^{12} 个神经元及其联接构成的神经回路的稳定状态数，也就远非是我们的语言所能描述殆尽，所以才会有“不可言说”的情况，才会有那么多难以言表的情感体验、才会有含糊不清的意义涌现、才会有似乎是顿悟的创造性思维以及才会意识到我们似乎具有意识的自我体验等。但这一切都毫无例外地源于规模无比的神经元集群活动结果。



1、心智能否被计算

因此，这意味着，只要机器的集成电路中基本元件与连接规模（目前只有 10^8 左右）**超过**人脑的基本元件与连接规模（ 10^{16} ），那么无疑就可以指望机器能够像人脑一样自涌现出**高级心智现象**。

应该说，正是机器量级规模上的**局限性**，制约着人工智能实现这种高级心智的尝试。结果，我们就因此可以期待，实现心智活动采用什么基本元件是不重要的，重要的是**这样的元件之间复杂的相互作用行为必须具有某种协同性、必须达到某个临界规模**。



1、心智能否被计算

至于所谓那些机器局限性的困境并不能成为机器不能拥有心智的理由（并因此否定自明性意识的存在），因为，目前机器不可计算的那些形式问题同样人类也不可计算，比如停机问题对于机器是半可解问题，但对于人类不是同样也是半可解问题吗？你无法判定你自己是否已经死了，甚至有人认为，即使哥德尔定理也不能成为否定心智计算化的论点，因为他们认为人类心智能力同样也是由哥德尔和图灵的理论所界限的。



1、心智能否被计算

总之，这种肯定观点的要点就是认为，从根本上来讲心智能力的表现范围并没有超出形式化计算的范围。因此形式化计算的局限性不会对心智计算化实现构成本质上的障碍。

就此而言，将心智活动归结为计算问题并无不妥之处。或许，在不远将来的某一天，一种具有人类心智能力的机器就会展现在我们眼前。



1、心智能否被计算

不过，问题恐怕没有这么简单，持**否定态度**的观点却坚持认为，即使人类能够造出一台能做人们明确告诉机器去做的任意事情，也无法造出一台具有情感、意识、幽默感等以及做出人们意料之外的事情。这种观点认为机器不可能具备思想和情感，更不用说是意识了（并因此认为自明性意识是存在的）。



1、心智能否被计算

诚然，如果从人类心智现象的种种独特性出发，可以发现机器所无法企及的心智活动能力，如整体局势判断和边缘意识感悟等问题。

由于机器是以一步步机械运算为基础的，对于人类一打眼就能辨识事物的心智行为，如果代之以机器，那就会陷于无穷无尽的细节辨别之中。

也许人们可以给出整个脑机制的形式化描述，但由于问题本身的复杂性，给出的形式化描述系统必定是逻辑不一致的，因此，即使能够给出形式化描述，也是徒劳无益的。



1、心智能否被计算

很明显用算法来刻画事物的手段是非常受局限的，在任何一个形式系统中总存在不能由公理和步骤法则证明或证伪的正确命题。

简要言之，世界万花筒般的复杂性不可能用可列的算法步骤来穷尽。

的确，与大多数现代计算机不一样的是，人脑不是一种通用机。在完全发育好以后，人脑的每部分都是特异化的，并在相互作用中完成整体的心智活动。因而，不可能将大脑还原为一组特殊的规则或公理。特别是意识和语言之中所不可避免的自指性已经远非是任意逻辑计算系统所能包容的。



1、心智能否被计算

从这种公理算法步骤不可穷尽性底下，我们还会发现机器难以逾越的根本障碍所在。思维过程和它的形式化模型之间存在着根本的区别。形式化的整个思想体系并没有超出抽象集合概念的范围，也没有超出对思维过程所做的纯集合性解释的范围。反之，意识和心智的固有属性则来源于它的独一无二的整体性质，这是任何形式化方法显然无法解释的。



1、心智能否被计算

那种认为人类心智能力同样受到哥德尔和图灵理论的限制无疑是忽略了人类心智能力中更为重要的心脑所具有的自反映能力，如语言和意识的自指性能力。这种自指性必定不可能有逻辑计算的方法所实现。从这个角度看，逻辑计算的方法既不能解释意识，也不能解释意识的表达内容，因而也就不可能解释作为标准设想中的心智。



1、心智能否被计算

一句话，机器运算基于的基础是因果性公式，是一种机器的、分析的、低级的、最简单的、最原始的联系形式；而心脑活动基于的基础则是非力相关性原理，是一种内在的、依存性的、整体自涌性的联系形式。两者之间有着根本的区别，绝不能同日而语的。

因此心智不可能被归结为计算问题。指望有朝一日我们就可以面对具有心智能力的机器，无疑是白日做梦。



2、来一场图灵测验

你看，强调机器能够拥有心智的观点似乎道理很充足；而强调机器不能拥有心智的观点也并非没有道理。我们似乎真正陷入了一个二律背反的境地。其实早在20世纪五十五年代初，伟大的英国数学家图灵就已经提出了这一问题。他在《心智》杂志上发表了“计算机器与心智”的文章，就首次明确提出了“机器能不能思维”这一重要命题，并给出了一种测验机器心智是否达到人类水平的测验，即著名的图灵测验。



2、来一场图灵测验

所谓图灵测验，指的是在两间隔离的房间里分别关有一个人与一台机器，然后通过向人或机器提问并根据他们的回答来判断谁是人，谁又是机器。

图灵认为，如果通过你的巧问，最终能够正确地将人与机器区分开来，那么

说明机器不同于人，否则就说明机器与人在心智上没有差别，起码在语言行为能力上是这样的。





2、来一场图灵测验

为了使读者有个直观的理解，我们不妨就来进行这么一次测验。测验的提问就是一个十分简单的问题：你多大年纪了？其中人们不难发现，用什么话语回答并不重要，重要的是，人的感情冲动导致那种出人意料的行为反应，是不可穷尽的事。机器则顶多在已知情况下来编定程序作出意料之中的回答，除此之外，别无选择。这就是**人与机器**之间一条跨越不过去的鸿沟，使得当你用巧妙式的提问后，**程式化的机器一定不会对机器的身份暴露无遗**，除非机器放弃预先编程方式。



2、来一场图灵测验

为了看清楚人类是如何做出“出人意料”反应的，我们举一个人类对话的例子，从而引出更加复杂的意识问题。德国生物学家福尔克·阿尔茨特和依曼努尔·比尔梅林在他们合著的《动物有意识吗》一书中，有如下一段精彩生动的对话（发生在一对男女恋人之间）。（对话是这么开始的）那位男士在变着法子逗弄他的女伴，半开玩笑半嗔怪地说她只是一台自动机器：“您是一台自动机器，夫人。”刚开始她的反应泰然自若：“你什么时候见过一台能出汗的自动机器？”



2、来一场图灵测验

他：“怎么没见过？每台机器都会变热。”

她：“可是你见过想喝饮料的机器吗？”

他：“机器需要加油是非常正常的。”

她：“我说的不是加油，而是喝的欲望。你难道见过对汽油怀有欲望的汽车吗？”

他：“当然！完全可以用计算机的语音提示代替油量表的指针。计算机也可以说：“我想要无铅汽油。””

她——现在已经不那么镇定了——：“我说的不是那个意思，见鬼了，再说一遍，我说的是我想喝东西的欲望。”

他：“没错呵，车载计算机在油料耗尽之前，也可以表达这种欲望呵！——您就是一台机器，夫人。”



2、来一场图灵测验

当我（指《动物有意识吗》的作者）正考虑，是不是去给这两位拿些饮料来，以排解这道不可解的难题时，谈话出现了令人吃惊的转折。

她：“按理我根本用不着那么认真的对待你。”

他：“为什么？”

她：“嗯，因为你很可能根本就不存在。”

他：“你怎么会这么想呢？你是看得见我的呀！”

她：“是的。可是我也许只是在做梦或者这只是一种错觉。要么你就得向我证明，确实有你这个人存在。可是你能证明这一点吗？”



2、来一场图灵测验

他——看得出来，对方以其人之道还之其人之身已经让他陷入困境——：“你可以掐我一下，如果我“啊呀”一声，那就……不行，那你肯定要说，不论你掐我那一下还是我“啊呀”那一声都是你梦见的事情。”

她：“没错！”

他——作为一种战略撤退——：“我确实无法向你证明我的存在。可那又怎么样呢？”

她——带着显而易见的胜利的喜悦——：“如果你根本不存在，那我是不是一台自动机器对你来说又有什么所谓呢？”接着她又加了一句，声音非常迷人：“你是不是也来一点儿无铅汽油？”



2、来一场图灵测验

可见，要逻辑上证明一个人是不是机器或者是不是存在是很困难的，因为这涉及到意识的**主观性体验**问题，他人是无法知道的（所谓他心知问题）。反过来，要逻辑上证明一台机器是不是一个人一定也是很困难的，这需要证明一台机器是不是跟你一样有**主观意识**，同样他人也是无法知道的。或者可以这样说，要证明一台机器是否拥有像人类一样的智能（意识），根本就**不是一个逻辑证明的问题**。



2、来一场图灵测验

不过，“尽管无法进行合乎逻辑的证明，但对他人之意识的揣测依然可以做到准确无误：只要涉及我们周围的人的行为，这种揣测基本上可以做到恰如其分、明白无误，而且可以加以解释和预测，平常我们谁也不会把它当成是“胡乱猜测”而置之不理。” 也就是，我们却可以通过机器的行为反应，来揣测机器是否拥有与我们一样的智能（相同境遇下有类似的行为反应）。从这个意义上讲，图灵测验在判断机器是否拥有与人类一样的智能上，有着不可替代的作用。



2、来一场图灵测验

而通过前面有关“你多大年纪了”的一再诘问中机器所暴露出来的那种缺陷，恰恰说明机器在遇到此类问题时，与人类的行为反应不能一致，缺少的正是出人意料的行为反应，这是预先编程不可克服的局限性。

由此可见，靠预先设计的程式化算法执行的机器是不可能仿造出同人类心智媲美的机器心智来的。其实，在《无心的机器》一书中就已经给出了这样的结论：“基于逻辑的机器，只能是无心的机器。”

End 2



3、钵中之脑的启示

在上面的讨论中我们涉及到一个人虚幻的存在性问题，对于理解机器能否达到跟人类一样的心智起着关键作用。为了一探究究竟，我再来做更加深入的哲学分析。为此，我们从一个称为“**钵中之脑**”的思想实验说起。“钵中之脑”的思想实验是美国哲学家希拉里·普特南在《理性、真理与历史》一书中提出来的，其原文如下：



3、钵中之脑的启示

这里有一个哲学家们所讨论的科学幻想中的可能事件：设想一个人（你可以设想这正是阁下本人）被一位邪恶的科学家作了一次手术。此人的大脑（阁下的的大脑）被从身体上截下并放入一个营养钵，以使之存活。神经末梢同一台超科学的计算机相连接，这台计算机使这个大脑的主人具有一切如常的幻觉。人群，物体，天空，等等，似乎都存在着，但实际上此人（即阁下）所经验到的一切都是从那架计算机传输到神经末梢的电子脉冲的结果。



3、钵中之脑的启示

这台计算机十分聪明，此人若要抬起手来，计算机发出的反馈就会使他“看到”并“感到”手正被抬起。不仅如此，那位邪恶的科学家还可以通过变换程序使得受害者“经验到”（即幻觉到）这个邪恶科学家所希望的任何情境或环境。他还可以消除脑手术的痕迹，从而该受害者将觉得自己一直是处于这种环境的。这位受害者甚至还会以为他正坐着读书，读的就是这样一个有趣但荒唐之极的假定：.....。



3、钵中之脑的启示

“钵中之脑”如图9.3所示，或许仅仅是一种理论假设。但由于不能排除其纯理论上的可能性，因此起码在哲学上给我们提出了这样两个隐喻性问题：（1）我们是否都是钵中之脑？

（2）操纵钵中之脑的超科学计算机真的存在吗？



钵中之脑思想实验



3、钵中之脑的启示

很明显，对于这两个问题的回答，有助于我们弄清人脑与机器在心智上到底存在不存在本质的差异这一问题。或许正好也顺带跳出了“他心知”命题的困境。那么“我们都是钵中之脑吗”？也就是说我，你还有你周围那么多的人难道都是像普特南描绘的那样，并非是完整自主的人，而仅仅是些养在营养钵中的、受超科学计算机控制的脑。其实，这一问题是一个典型的怀疑论质疑。



3、钵中之脑的启示

我国古代思想家庄子曾经提出过类似这样命题的寓言（《庄子·齐物论》）：“昔者庄周梦为胡蝶，栩栩然胡蝶也。自喻适志与！不知周也。俄然觉，则蘧蘧然周也。不知周之梦为胡蝶与？胡蝶之梦为周与？”

这个寓言换个角度讲同样也隐含地给出了这样一个哲学命题：“我们都在梦中？”应该说，从本质上讲，这一命题同“我们都是钵中之脑”是等价的，指的都是对现实生活的一种**虚幻假设**。



3、钵中之脑的启示

直觉上，大概我们谁也不会承认我们都在梦中且又不为我们所知道。但我们又怎能证明“**我们确实不在梦中**”呢？即使在你自认为是清醒的时候，要想通过实证手段给出有说服力的证明也是十分困难的。特别是由于你所能采取的一切实证手段，原则上也都会在梦中“实现”，于是你的这一切努力，都会被认为是“梦中所为”之事而使你为证明“不在梦中”的努力付注东流。



3、钵中之脑的启示

稍微深入分析我们不难看出，像“我们都在梦中”之类的命题，均有一个共同点，那就是命题的**自毁性或自我反驳性**。对于“我们都在梦中”也一样。如果该命题为真，那么我们就知道“我们都在梦中”，而“我们都在梦中”却意味着我们并不意识到，从而也就不知道“我们都在梦中”。这显然是一个矛盾，也就是说，“我们都在梦中”是不成立的。于是依此类推，“我们都是钵中之脑”同样也不能成立。



3、钵中之脑的启示

因此只要机器建立在逻辑运算之上，受到逻辑一致性约束，那么任何超越逻辑运算的努力，机器终将注定是难以胜任的。

像人类心智这样复杂的事物，那怕是局部实现，只要其含有超越逻辑运转的成份，就远非是逻辑机器所能胜任的，更不用说是心智中的意识活动了。这就是“钵中之脑”带给我们的启示。



3、钵中之脑的启示

或许在心智再现研究中，出路不可能通过逻辑拆解重建的方法来实现，如果存在什么方法的话，那也只有以复杂性对付复杂性的方法来寻找出路，通过特异化的“机器”本身来拥有某种心智能力，而非要靠逻辑编程的步骤来实现。



3、钵中之脑的启示

看来为了真正能够达到完全仿造心智的目的，我们必须另谋出路，那就是我们必须突破经典计算的局限性，寄希望于非逻辑运算（更非预先设定程序化的）机器。在这里，结合自然机制的“生物化”、“量子化”、“集群化”的非经典计算可能才是一条出路。

End 3