心灵测试和最小能力集

Li Yaguang

email: arkliyg@gmail.com

毫无疑问，意识是以物质为基础的，在已经完成的观察和实验中，可以确定意识是以神经元之间的放电为物理基础的信息活动[1]，我们已经无需再向它处寻找意识。对意识本身，我们能够观察到的只是神经元之间放电时看似杂乱无章的节奏，但实际上却传导着结构化的有序的信息，且随之化学性地以神经元之间的连接存储为记忆。这是我们观察到的一切，物质上的现象再无其他。我们不必再向实验观察奢求什么，实验最多只会再添加一些在机缘巧合的条件下对脑内物质运动细节的理解，这些终究只局限于对简单问题的解答[2]。对于简单问题的探究，我们已经做得足够多，虽然也许还有更多实现上的细节需要考量，但我们已经处在了合适的研究范畴。

而对于困难问题，也无需纠结太多。对它我们已无处可躲，巨大的事实已经向我们扑来，承认答案的确定存在只是我们接下来必须要做出的抉择。基于心理分析，我们已经有了一个已经基本被广泛认同的理念：自我意识实际上是对“自我”概念的循环引用。我们能感受到自身的思考活动的存在，进而产生了“自我”的概念。与此同时，我们之所以能感受到“自我”，正是因为意识本身，“自我”这个概念是存续于思考之中的。在

脑内的信息活动中，这种对“自我”循环引用[3]，标志着自我意识的确立。在看到了简单问题的答案的基本轮廓后，我们只能把这种动态神经元在信息活动中对自己的循环引用称为自我意识。我们不得不承认，这是自我意识在物质运动中唯一的表象，动态的一簇神经元[1]的持续运动性的激活，自我的循环引用，所有的对于感觉的困难问题的根源也就生成于此。而循环引用必然是排外的，私密性和独特性也就蕴藏于此。

（示意图：自我的循环引用）

自我的循环引用的意义在于自我迭代，更新，扩展。设$\boldsymbol{Fe}$为可以实现自我循环引用的算法，$\boldsymbol{\Omega}$为自我意识，$\boldsymbol{\Omega}'$为更新后的自我意识，则有：

$$\boldsymbol{\Omega}' = \boldsymbol{Fe}(\boldsymbol{\Omega})$$

且：

$$\boldsymbol{\Omega}' \supset \boldsymbol{\Omega}$$

我们都已经熟知图灵测试，但关于图灵测试的原理及有效性却并没有被根本的论述过。或许由于困难问题的答案难以描述，我们一直是以它不能证伪而进行图灵测试的有效性的论述的，但这不是直接的从正面进行的证明[4]。图灵测试之所以在现阶段被广泛认可，是因为它正是我们可以对互相的同类进行的一种有效的测试。当然每一个心智健全的人都承认自己拥有意识，但无论如何，在某种情况下，如果我们需要向其他人证明我们是有意识的，那么此时其他人可以通过对我们进行图

灵测试加以判别，也就是我们通常所说的谈话。显而易见的是，我们对他人施以图灵测试可以有效地证明他们意识的存在与否，但这看起来简单的过程的背后却隐蔽着一个根本性的条件，那就是我们都相信彼此大脑的生理构造相同。既然自己在正常对话时是清醒并有心智的，那么推己及人，对方在正常对话时一样会有正常的心智。也就是这个根本性的条件，导致了图灵测试出现了旷日持久的争议。

机器毕竟不是人，我们无法相信彼此的神经架构是相同的，在推己及“机器”这一步骤上出现了障碍。其中僵尸体就是一个无法回避的挑战，即图灵测试无法区分真正有意识的受试体和一个看起来有意识的僵尸体。对此，图灵以数个反问做了不能证伪的论述，并认为没有必要去区分真正有意识的受试体或是伪装的看起来具有智能的僵尸体，因为这对旁观者来说是没有区别和影响的。或许图灵是对的，但可能通过图灵测试的机器是真的具有意识的，它要在内部有一个自我的信息，否则可能无法充分表现出像个具有正常心智的人。当然如果我们有种方法能深入受试者内心，那么在僵尸体内是探测不到意识的，这是我们对僵尸体的定义。

（示意图：对人的图灵测试，对机器的图灵测试）

我们可以看到，图灵测试之所以在机器上不是完全有效的，是因为我们无法相信机器与我们一样拥有合理的心理活动。实际上，对是否具有意识来说，对话的能力和其他任何表达或沟

通的能力并不是必须的[5]，而且我们的意识活动也是很难通过任何方式来精准表达的。当然，图灵基于所处的历史时期和科技发展水平做出了在当时最好的论断。但有幸近些年的脑科学研究以及信息技术和计算机科技的发展，使得我们可以观察到意识的活动。一个大脑是否有意识，现在可以通过探测脑电活动来做出很精准的判断，我们已经可以证实，一些特定的脑电活动反映了我们有意识的和下意识的意识活动。我们现在可以知道大脑在思考，但具体的思考内容还不能准确的得知。如果我们有能力解码这些脑电活动，以进一步读取心理信息，则是判断是否具有意识的最准确的方式。在此方面，脑神经研究已经有了一些初步的进展，但仍然很粗浅，无法得出较准确的信息。

而对于机器，我们是可以拥有这种“读心”能力的。如前所述，“自我”意识是一种循环引用，如果我们发现了机器对“自我”概念产生了循环引用的某种数据结构信息，就可认定它产生了自我意识。我们测试它是否有意识，不只观察它外在的表现，也不只观察它内在的“心理活动”，主要和根本的一点，是要观察这种对“自我”概念的循环引用。并且这种“自我”概念不要是刻意生成的，而是要在输入各种客观世界的信息后，进行各种逻辑运算自然而然得出的信息。就如同我们设计一个程序来计算π值，如果那只是一个输出已经存储了π值字符串的程序，虽然运行起来和真正实时计算并输出π值数字的程序

无异，但那却是拙劣且无效的。

（示意图：带有日志及内省过程的机器示意）

在机器内部对于这种所谓的“心理活动”，以及对自我概念的循环引用，看起来只是一种模仿。它模仿着我们人类对自身的自我概念。但模仿并不意味着是假的，我们可以类比飞机和鸟类，它们的理论却都基于空气动力学，并且在天空的飞翔都是实在的事实。如果内在的循环引用仍然是看上去很像是故意模仿，那么可以反身思考我们自己，我们的意识的背后是在遵循客观定律的条件下运行的某种信息活动，我们的意识本身也是大脑内部产生出来的一种信息现象，这与机器内部发生的关于“自我”的信息活动别无二致，如果说机器在模仿，那么我们自己何尝不也是在模仿一个拥有意识的信息活动体。

可将此种深入信息活动体内部检验是否生成自我概念的方法视为图灵测试的进阶版本，我们亦可以称之为“心灵测试”。即不只测试可用语言或任何沟通方式所表达出来的信息，还要测试那些信息活动体内部的无法用语言等方式表达出的信息活动。如果产生了对“自我”概念的循环引用，则标志着真正意识的存在。

产生自我意识产生所需要的条件中，可以确定的必要条件有逻辑运算能力，抽象能力，联想能力，保存和访问信息系统自身的所有历史信息活动的能力。一个具有自我意识的系统，应该具备包含或不限于以上的信息处理能力，且配以解决简单问

题的适当算法和实现自我概念循环引用的机制。设$\boldsymbol{W}$为输入的信息[1]，$\boldsymbol{T}$为类人的记忆和回忆、联想等能力，$\boldsymbol{L}$为类人的逻辑思维以及其他产生意识所需的能力，$\boldsymbol{M}$为系统中保存的信息，$\boldsymbol{Fs}$为解决简单问题的算法和可以实现自我循环引用的机制，则有:

$$\{\boldsymbol{T}, \boldsymbol{L}', \boldsymbol{M}', \boldsymbol{\Omega}'\} = \boldsymbol{Fs}(\{\boldsymbol{T}, \boldsymbol{L}, \boldsymbol{M}, \boldsymbol{\Omega}, \boldsymbol{W}\})$$

式0

可得:

$$\boldsymbol{\Omega}' \subseteq \boldsymbol{Fs}(\{\boldsymbol{T}, \boldsymbol{L}, \boldsymbol{M}, \boldsymbol{\Omega}, \boldsymbol{W}\})$$

式1

设:

$$\boldsymbol{t} \subseteq \boldsymbol{T}, \quad \boldsymbol{l} \subseteq \boldsymbol{L}, \quad \boldsymbol{m} \subseteq \boldsymbol{M}, \quad \boldsymbol{w} \subseteq \boldsymbol{W}$$

且满足:

$$\boldsymbol{\Omega}' \subseteq \boldsymbol{Fs}(\{\boldsymbol{t}, \boldsymbol{l}, \boldsymbol{m}, \boldsymbol{\Omega}, \boldsymbol{w}\})$$

式2

同时，若有:

$$\boldsymbol{t}' \subseteq \boldsymbol{t}, \quad \boldsymbol{l}' \subseteq \boldsymbol{l}, \quad \boldsymbol{m}' \subseteq \boldsymbol{m}, \quad \boldsymbol{w}' \subseteq \boldsymbol{w}$$

且满足:

$$\boldsymbol{\Omega}' \subseteq \boldsymbol{Fs}(\{\boldsymbol{t}', \boldsymbol{l}', \boldsymbol{m}', \boldsymbol{\Omega}, \boldsymbol{w}'\})$$

则必有:

$$\boldsymbol{t}' = \boldsymbol{t}, \quad \boldsymbol{l}' = \boldsymbol{l}, \quad \boldsymbol{m}' = \boldsymbol{m}, \quad \boldsymbol{w}' = \boldsymbol{w}$$

[1] 不只是外部信息，还包含系统中所发生的信息活动的记录

即$\{t, l, m, w\}$是可以满足式1的最小的集合。

同时我们知道在初始时刻，信息系统内没有保存的信息，自我意识也是从无到有产生的，则初始时式2可为：

$$\Omega = Fs(\{t, l, w\})$$

式3

构造Fs和$\{t, l, w\}$，进行心灵测试，找到满足式3的Fs和$\{t, l, w\}$，即可得到初始的自我意识Ω。此时，我们可以称满足条件的$\{t, l\}$为产生意识的最小能力集。根据式0，Ω在Fs的作用下，输入W后可以将L中的各种能力迭代、发展和扩充，据此，我们可以逐渐得到一个类人或超人的L。

References

[1] Gerald M. Edelman, Giulio Tononi (2001), "A Universe of Consciousness: How Matter Becomes Imagination"

[2] Chalmers, D. J. (2007), "Phenomenal Concepts and the Explanatory Gap"

[3] Douglas Hofstadter R. (2007), "I Am a Strange Loop"

[4] Alan M. Turing (1950), "Computing Machinery and Intelligence". Mind, LIX (236): 433-460

[5] Nordgren RE, Markesbery WR, Fukuda K, Reeves AG (1971), "Seven cases of cerebromedullospinal disconnection: the 'locked-in' syndrome". Neurology. 21 (11): 1140-8.