

· 西文中译 ·

机器有意识吗?

- 作者 / 王小塞, 俞筱钥
- 单位 / 北京大学外国哲学研究所、北京大学哲学系(宗教学系), 北京, 100871

摘要: 基于生物工程的最新进展, 本文论证了标题中的问题“机器能拥有意识吗?”, 可以得到肯定的回答。

关键词: 人工智能; 机器意识; 生物工程; 人机神经界面

DOI: 10. 12184/wsprzkkWSP2515-528813. 20220602

艾伦·图灵(Alan Turing)认为有可能(至少在理论上)制造出能够喜欢吃英国夏日最受喜爱的食物——草莓和奶油的机器。由此我们可以推断, 他也认为有可能(同样, 至少在理论上)制造出有意识的机器。因为如果你没有意识, 你就不可能真的喜欢草莓和奶油, 不是吗? 无论如何, 图灵都非常明确地表示他认为机器可以有意识。然而, 他认为这样的机器不可能很快就被造出来。这并不是因为他认为这项任务特别困难, 而是因为他认为不值得为之付出努力。他在颇具影响力的《计算机器与智能》(Computing Machinery and Intelligence)中写道:“也许可以制造出一台能够享用这道佳肴的机器, 但任何这样的尝试都是愚蠢的。”他还补充道, 就连只是提到这种缺

作者简介 王小塞, 北京大学外国哲学研究所研究员、北京大学哲学系(宗教学系)助理教授, 北京大学博古睿研究中心研究员, 研究方向: 智能哲学、维特根斯坦、分析哲学等。

俞筱钥, 北京大学哲学系(宗教学系)本科生。

陷，或许都会让他的读者感到无聊又可笑。他解释道：

“这种缺陷的重要之处在于，它导致了其他一些缺陷，例如，人与机器之间难以像白人与白人之间或黑人与黑人之间一样发展出友谊”^①。

因此，图灵提醒我们（正如他总是习惯的那样），即使在人类物种内部，人类也总是很难接受某些其他的个体也具有与自己同等的基本价值或能力；他说，（在他看来）机器也许不能享用草莓和奶油这件事的重要之处在于，它体现着一种更广泛的缺陷即机器无法共享人类生活形式的某些要素。大体上看，图灵想得没错，就像他也准确地预测了人工神经网络、一般的机器学习、尤其是强化学习的成功一样。

然而，图灵一再表明，人们不会为了造一台无智能但带有明显人类特征的机器而大费力气，这终归还是预测错了。恰恰相反，对聊天机器人、虚拟助手以及护理、性爱等家用机器人日益增长的需求清楚地表明，事实正相反。因此，如果机器可以被制造成有意识的，我们大概迟早会设计它们，无论是有意还是无意的。此外，由于大多数人认为意识通常能对世界作出重大的改变（尤其从伦理的角度来看），机器能否拥有意识这个问题就显得越发重要，值得更多人思考该如何去发问。

当然，首先要做的是明确我们对“机器”一词的定义。当图灵考虑机器是否能思考的问题时，他将“机器”一词限制为指数字计算机，也就是与我们现代绝大多数计算设备（从智能手机到超级计算机）相同类型的机器。在他写作的那个时代，即1950年左右，他刚刚助力让这样的机器成为现实。顺便一提，他其实早在1936年就为这类机器的创制奠定了数学基础，今天我们称之为“通用图灵机（Universal Turing Machine）”^②。只是当时数字计算机毕竟是新兴产物，图灵仍有大量的解释工作要做。而今，大多数人至少已经在直观上了解了计算机基本的能力，那么关于计算机器的理论解释，我们这里就不再赘述。实际上，我们不需要把我们对“机器”的理解限制在数字计算机上。正如我们即将看到的，本文对机器能否拥有意识这个问题的探讨，只需要规定所用的相关工程方法不主要具有生物本质。

① Turing, A. M. 1950. Computing machinery and intelligence. *Mind* 59, 448.

② 参见 Turing, A. M. 1937. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society* 42, 230-265.

到目前为止, 机器是否会有意识的问题更为棘手的部分是要确定“意识”这个词应该被理解成什么含义。意识仍然是这个世界上最神秘的东西之一。可以肯定的是, 人类对意识非常熟悉, 因为一个人类个体的意识是完全由其私人主观经验构成的。在这个术语的一般意义上, 意识是这样一种特殊性质: 在一个特定时间的特定心灵状态 (mental state) 下的主观感觉 (what it is like)。因此许多人倾向于认为, 即使在他们所能想象的最精巧的机器人中也一定缺少这个特殊性质。人们在应对机器人也能像人类一样享受草莓和奶油 (或经历任何主观经验) 这个命题的时候会感到一种荒谬感, 这种荒谬感正是“机器能有意识吗?” 这个问题想要强调的。现在解答这个问题的主要困难是, 尽管我们对意识自然地熟悉, 我们对其本质仍然非常无知。关于意识是什么并没有一个广泛认同的理论, 我们当然也不知道如何从头开始构建它。毫无疑问, 这种无知在很大程度上造成了这个问题的模糊性。正如我们即将看到的, 诀窍就在于规避这种无知, 转而利用我们对意识自然熟悉的这个基础。

那么我们能想到一个有希望的方法来制造一个有意识的东西吗? 当然, 可以想到是生殖, 包括辅助生殖技术, 但这将违反我们的限制条件, 即工程方法不主要地具有生物本质。再生的方法也是一样, 例如神经干细胞治疗和将皮肤细胞变成功能性神经细胞的组织纳米转染。^③ 生殖和再生这两类方法都可以直接或间接地 (通过恢复相关功能) 用于制造有意识的事物, 特别是人类。然而, 由于这里采用的方法明显是生物性的, 我们自然会认为结果不是有意识的机器, 而是有意识的人类。

考虑一种不同但类似的方法。生物工程的最新进展已经做到了用双向连接的假肢实现人类部分神经系统的功能恢复, 使得运动控制和本体感觉 (对

^③ 例如, 各见 Jiao, Y., Liu, Y.-W., Chen, W.-G., Liu, J. 2021. Neuroregeneration and functional recovery after stroke: advancing neural stem cell therapy toward clinical application. *Neural Regeneration Research* 16, 80-92. 和 Gallego-Perez, D., Pal, D., Ghatak, S, Malkoc, V., Higuera-Castro, N., Gnyawali, S., Chang, L., Liao, W.-C., Shi, J., Sinha, M., Singh, K., Steen, E., Sunyecz, A., Stewart, R., Moore, J., Ziebro, T., Northcutt, R. G., Homsy, M., Bertani, P., Wu, L., Roy, S., Khanna, S., Rink, C., Sundaresan, V. B., Otero, J. J., Lee, L. J., and Sen, C. K. 2017. Topical tissue nanotransfection mediates non-viral stroma reprogramming and rescue. *Nature Nanotechnology* 12, 974-979.

肢体位置的直观认识)得到加强,并减少幻肢感知。^④该技术仍处于早期阶段,但它为人类神经系统的一部分可以使用硅等半导体材料进行修复的原则提供了经验证明。我们似乎有理由期待,这项技术以及正在进行的有关人机神经接口的相关研究将在未来取得进一步的进展,从而使神经系统的更多部分可以用硅这样的材料进行恢复或替换。例如,一个四肢截肢者可能能够恢复全部的手臂和腿部功能,包括触觉、本体感觉、温度感觉等等。

现在让我们做下面三点非常符合常识的假设:第一,人类的神经系统(包括大脑和位于其中心的脊髓)组成了人类意识,也即任何特定人类的意识状态的总和。第二,一个人类个体的意识状态通常包括基于肢体的感觉体验,因为神经系统延伸到肢体。第三,一些截肢者的意识状态包括基于假肢的感觉体验,因为他们神经系统的相关部分已经被恢复。

由此可见,这种截肢者的假肢也部分地构成了他们的个体意识,因为他们的神经系统延伸到了他们的假肢中。此外,由于设计这种截肢者的假肢的相关方法主要不是生物性质的,我们可以得出结论,机器部分地构成了他们的个人意识。然而,这并不意味着机器可以有意识。因为显然,人体剩下的某些部分可能最终仍是必要的,这似乎特别有说服力,考虑到迄今为止所考虑的机器部件与神经系统的整合仅仅影响到周围神经系统,而不涉及通常所说的中枢神经系统(即大脑和脊髓)。

然而,是中央还是非中央的部件被替换会造成什么与之相关的差别并不显而易见。可以肯定的是,中枢系统和周围系统之间有许多重要的区别,所以在周围神经系统里成功换上一个硅基部件,并不意味着实践层面上就有可能在中枢系统里也做这种替换,一只脚可不是一个大脑。然而,抽象到一定的层面上,二者却又确实属于同一类东西,因而很难否认神经系统的各个部件理论上都能被替换。再次强调,随着最近生物工程的进展,我们似乎有理

④ 特别地,参见 Srinivasan, S. S., Tuckute, G., Zou, J., Gutierrez-Arango, S., Song, H., Barry, R. L., and Herr, H. M. 2020. Agonist-antagonist myoneural interface amputation preserves proprioceptive sensorimotor neurophysiology in lower limbs. *Science Translational Medicine* 12 (573). 和 Srinivasan, S. S., Gutierrez-Arango, S., Teng, A. C.-E., Israel, E., Song, H., Bailey, Z. K., Carty, M. J., Freed, L. E., and Herr, H. M. 2021. Neural interfacing architecture enables enhanced motor control and residual limb functionality postamputation. *Proceedings of the National Academy of Sciences* 118 (9).

由预测, 当下和未来的关于人机神经接口以及有理由预测, 当下和未来的物理神经网络^⑤、记忆电阻^⑥和记忆系统^⑦的研究, 将使人类能够用像硅这样的材料来恢复或替换人类神经系统的越来越多的部分, 不妨想象下面这种可能场景。

一百年后, 技术有了稳定的进步。在这一时期的某个时刻, 小布在还年轻时就患上了一种新的神经系统的疾病。但好在小布随时都能做上她需要的植入手术, 同样幸运的是, 因为手术的间隔时间足够长, 得益于各种治疗手段和她自己神经系统持续重塑神经的能力, 每当她有一部分神经要被替换时, 上次植入的神经就已经很好地整合进了她的神经系统。

小布的神经系统究竟需要被植入物取代多少才能让人认为她是一台机器而非一个人, 这因人而异。有些人可能认为有必要替换小布的所有神经系统, 包括大脑和脊髓, 甚至是她的整个身体。无论是哪种方式, 改造的原则都是一样的, 因此, 在此基础上, 每个认可近期假肢方面的进展可以证明人类一部分神经系统可以用硅这样的材料来恢复的人, 都能同意小布最终会成为一台机器。此外, 他们将因此能够接受小布最终会成为一台像人类一样有意识的机器(或者, 无论如何, 就像小布曾经是的那样有意识)。

毫无疑问, 有些人会想反对在小布的逐渐转变过程中被假设的心理连续性。这种反对意见可能主要针对被假设的小布意识的连续性, 或者不那么直接地针对她的人格(personal identity)的连续性(因为缺乏人格同一性也会使人们对假设的意识连续性产生怀疑)。为了证明他们疑虑的有效性, 这种反对者不得不解释, 在他们看来, 事情可能会在哪里出错。想必, 他们相信有一个时刻, 在这个时刻转化的过程将变得明显更加激烈。当然, 如果顺着这个思路能发现任何这样的时刻, 都将会有很大的科学意义。但至少在那之前, 人们有理由相信只要时间足够多、步骤足够小, 人类的神经系统可以被

⑤ 参见 Wright, L. G., Onodera, T., Stein, M. M., Wang, T., Schachter, D. T., Hu, Z., and McMahon, P. L. 2021. Deep physical neural networks enabled by a backpropagation algorithm for arbitrary physical systems. arXiv: 2104.13386 (cs.LG) .

⑥ 参见 Strukov, D. B., Snider, G. S., Stewart, D. R., and Williams, R. S. 2008. The missing memristor found. *Nature* 453, 80-83.

⑦ 参见 Zidan, M. A., Strachan, J. P., and Lu, W. D. 2018. The future of electronics based on memristive systems. *Nature Electronics* 1, 22-29.

其他材料的替换件所替代，这样一个人类就可以慢慢地被转变机器并同时保留着意识。

参考文献

- [1] Gallego-Perez D, Pal D, Ghatak S, et al. Topical tissue nano-transfection mediates non-viral stroma reprogramming and rescue[J]. *Nature Nanotechnology*, 2017, 12, 974-979.
- [2] Jiao Y, Liu Y W, Chen W G, et al. Neuroregeneration and functional recovery after stroke: advancing neural stem cell therapy toward clinical application[J]. *Neural Regeneration Research*, 2021, 16, 80-92.
- [3] Srinivasan S S, Gutierrez-Arango S, Teng A C E, et al. Neural interfacing architecture enables enhanced motor control and residual limb functionality postamputation[J]. *Proceedings of the National Academy of Sciences*, 2021, 118(9).
- [4] Srinivasan S S, Tuckute G, Zou J, et al. Agonist-antagonist myoneural interface amputation preserves proprioceptive sensorimotor neurophysiology in lower limbs[J]. *Science Translational Medicine*, 2020, 12(573).
- [5] Strukov, D. B., Snider, G. S., Stewart, D. R., and Williams, R. S., The missing memristor found. *Nature*, 2008. 453, 80-83.
- [6] Turing A M. On computable numbers, with an application to the Entscheidungsproblem [J]. *Proceedings of the London Mathematical Society*, 1937, 42, 230-265.
- [7] Turing A M. Computing machinery and intelligence[J]. *Mind*, 1950, 59, 433-460.
- [8] Wright L G, Onodera T, Stein M M, et al. Deep physical neural networks enabled by a backpropagation algorithm for arbitrary physical systems[J/OL]. 2021. arXiv: 2104.13386 (cs.LG).
- [9] Zidan M A, Strachan J P, Lu W D. The future of electronics based on memristive systems [J]. *Nature Electronics*, 2018, 1, 22-29.