

# “图灵测试”、“反转图灵测试”与心智的意义

蒋 柯\*

**[摘 要]** “图灵测试”被设计出来是为了检验计算机是否拥有了人类的智能。但是“图灵测试”一旦确立,同时也成为了计算机智能(人工智能)发展的导航标,即人工智能的发展都指向了能够满足“图灵测试”的技术标准的方向。文章通过对“图灵测试”的历史性分析,并结合“反转图灵测试”的论证,进一步指出:人类心智是一种历史性建构,即是人类的社会性交往塑造了人类心智的计算逻辑;自从有了计算机以后,人类智能和计算机智能的互动也在进行着另一种心智建构活动,其结果就是“人—机智能”。“人—机智能”建构将是后现代人类智能和计算机智能的弥合进路。

**[关键词]** 人工智能;图灵测试;反转图灵测试;心智;建构

近年来,随着计算机技术和人工智能领域内不断涌现出来的新兴成果,社会公众与学术界都不由得产生了这样的担忧:机器有可能拥有和人类一样的智能吗?如果机器拥有了智能,会不会超越人类而成为这个世界的新的主宰呢?拥有了智能的机器还是机器吗,人类是否应该让拥有智能的机器享有和人类一样的社会和伦理权利呢?诸如此类的问题,有的表达了人们对未来的某种担忧,有的则是欣喜,但所有的追问实际上都表达了这样一个根本性的问题,即“如果机器拥有了智能将会怎样?”要能够言说这个问题,我们首先要思考另一个问题:“心智究竟是什么?”

关于心智的思考由来已久,其中的争论与分歧也非常多。本文并不敢斗言承前人之意而定今日之理,要为心智的意义作注之类。本文的目的是希望梳理这样一条线索:即自计算机技术发端以来,伴随着人工智能技术的演进,以及人类在机器智能化进路上的所取得的成绩,关于“心智”意义的解释策略也在悄然变化;进而论证:心智本身并不是一个固定不变的概念,心智是在人和世界、人和机器的交互作用过程中的一种历史性建构。

## 一、“图灵测试”与“唯一心智”问题

70多年前,图灵构想了一个衡量计算机是否“是有智能的”的标准,这就是“图灵测试”。标准的“图灵测试”是:“如果计算机能在5分钟内回答由人类测试者提出的一系列问题,并且有超过30%测

---

\*心理学博士,温州医科大学精神医学学院教授,325035。本文是国家社科基金一般项目“健康中国背景下社区心理学的智慧模式研究”(18BSH120)的阶段性成果。

试者认为这是人类的回答,那么这台计算机就算通过了测试。也就是说,计算机的表现与人类相当”(Turing, 1950)。2014年6月7日,一个模拟13岁乌克兰男孩的计算机程序被认为通过了图灵测试<sup>①</sup>。2016年3月9日至15日,由谷歌开发的围棋程序阿尔法狗(AlphaGo)与世界围棋冠军、韩国顶尖职业棋手李世石展开了五局对弈,结果“阿尔法狗”以4:1的成绩取得胜利。2017年5月23-27日,“阿尔法狗”再次以3比0的绝对优势战胜中国围棋冠军柯洁。2017年,一个完全不依赖人类经验的新一代围棋程序“阿尔法元”(AlphaGo Zero)以100比0的优势战胜“阿尔法狗”(见《自然》杂志2017年10月19日的报道)<sup>②</sup>。这一系列人工智能领域的成就再次引发人类的终结焦虑:人工智能如果完全达到并超越了人类的智能,世界将会变成什么样子?这样的焦虑其实从人类大规模使用机器时就已经暗流涌动了,每一次机器的技术革新都会引发人们重新理解人和机器的关系,并再一次为新型的人-机关系而感到焦虑(Minsky, 1985;熊哲宏, 2002;Wallach & Allen, 2009, 2009)。

当我们面对计算机技术的发展而感到担忧时,首先要明白,我们的担忧所指向的目标是什么。也就是,计算机技术所创造的“人工智能”是否是真正的“智能”,以及通过了图灵测试的计算机程序和战胜了人类棋手的阿尔法围棋程序所表现出来的“聪明才智”是否真的是“智能”?

塞尔曾经通过著名的“中文屋”论证阐述了人类的“心智”与机器的计算之间的区别。塞尔假想有一个不懂得中文的人被关在一间小屋子里,只能通过字条与外面的人进行中文符号的交流。这就像是在进行一场“图灵测试”。这个不懂中文的人按照一套详尽的操作规则对字条上的符号做出应答。比如,当他接收到一张写有中文“你好”的字条,尽管他不认识中文,但是他可以根据预制的规则将一张写有“你好,很高兴认识你”的字条递出去。塞尔假设,只要这个人拥有的规则足够完备,那么屋子外面的人就不会怀疑屋子里的人是懂中文的。塞尔希望通过这个论证表明,通过图灵测试的计算机以及会下围棋的程序,它们和中文屋里面不懂中文的人一样,只是对符号进行规则化加工而不论符号的意义;而人能够对“意义”做出应答(塞尔, 2006;王佳等, 2010)。然而,实际上塞尔的论证仅仅表明“通过图灵测试的计算机可以没有智能”,却没有提供足够的证据表明“计算机事实上没有智能”(杨小爱等, 2013)。所以,当阿尔法围棋战胜了围棋世界冠军李世石,我们并不能肯定它能否体会到在这种竞争性游戏中取胜的喜悦或自豪;或者对它来说,下棋仅仅是完成了一系列对符号的计算过程。

事实上,当一个人遭遇一台可能拥有智能的计算机时,与遇一个“人”并需要判断他(她)是否拥有正常的“心智”时,所面临的问题是一样的。笛卡尔通过著名的命题“我思故我在”论证了心智的存在(笛卡尔, 2000)。笛卡尔论证的预设是“凡有动作必有施动者”,所以,有“我思”必有“我”。笛卡尔的论证很强,但是,自此以后心智的存在却成为了哲学家们最苦恼的问题。根据笛卡尔的论证,心智的存在基于“我思”,即只有当我自己觉察到自己的思维活动时,才能确认我自己的心智的存在(周晓亮, 2005;高申春, 2005)。然而我是无法觉察或体验到其他任何人的思维活动的,那么,我怎么确认他人和我一样是具有心智的呢?这就是“唯一心智问题”(黎黑, 2013)。在逻辑上,“唯一心智问题”是无解的。因为“唯一心智问题”是在“我思故我在”的基础上产生的,所以,要解决“唯一心智问题”必须要首先放弃“我思故我在”的心智存在论证(刘晓力, 2014;陈晓平, 2015)。当我们放弃“我思故我在”的论证之后,又如何确认心智的存在呢?

假设有一人——我们姑且称他(她)为“象罔”<sup>③</sup>——初次来到一个陌生的星球,遇见一些长相和自己类似的生物,他们和“象罔”互相打量。“象罔”不知道面前的这些生物是否有“心智”,这些生

①《中国新闻网》2014年6月10日:《电脑冒充13岁男孩首次通过“图灵测试”》<http://www.chinanews.com/cul/2014/06-10/6263081.shtml>

②维基百科——AlphaGo Zero: [https://en.wikipedia.org/wiki/AlphaGo\\_Zero](https://en.wikipedia.org/wiki/AlphaGo_Zero)

③《庄子》中杜撰的人物,指心智未开化的人。

物也不知道“象罔”是否有“心智”。那么他们之间的互动会如何开始呢？他们都停下来，仔细地看对方，有一方先伸出手，张开手掌在空中摇晃几下，表明自己没有武器；另一方也做出同样的动作。于是，他们彼此的戒备稍解。然后一方开始向对方伸出一个手指，慢慢地靠近对方；对方也伸出一个手指。终于两个手指尖触碰在一起。这是两个互相陌生生命的第一次接触。在斯皮尔伯格的著名科幻电影《第三类接触》中，曾经想象了这种互动的开始。在这个过程中，无论是“象罔”还是陌生星球上的生命，他们相互都是在试探对方“是否有心智”。实际上，他们每一方都是根据对方与自己的互动来判断：对方是否有可能与自己产生“有意义的交流”。只有双方之间存在“意义”的交流，每一方才有可能认为对方是有“心智”的。而所谓“意义”却正是在双方的互动过程中建构出来的。在这种“意义”建构的基础上，我们有理由认为，“心智”——任何一方的，或者双方的——也是在这种互动过程中被建构出来的（蒋柯，2016）。因为如果这个世界上真的只有一个“心智”，那么这个“心智”本身的存在也失去了“意义”。每一个“心智”都只有在与其它“心智”的互动过程中才体现出作为“心智”的存在意义。

于是，当我们放弃了笛卡尔关于“心智”的“我思故我在”论证，就必须要建立一种新的论证，并且这个论证的预设前提是：“当且仅当意义存在，心智才是存在的。”因此，如果我们承认人际互动建构了“意义”，那么我们同样也必须承认人际互动建构了“心智”。

实际上，图灵本人在构想“图灵测试”时是参考了当时流行的一种“测谎游戏”。测试者和被测试者被一个屏风隔开，被测试者努力伪装自己的性别去迷惑对方，而测试者的任务就是通过一些问题来判断被测试者的真实性别是什么（Turing, 1950）。因此，我们可以把“象罔”与陌生生命的互动过程看作一个“图灵测试”。即“象罔”通过图灵测试判断对方是否具有“心智”；而对方也通过同样的“图灵测试”来判断“象罔”是否具有“心智”。正是在这样的相互的“图灵测试”过程中，“心智”本身被建构出来了。

当人类通过“图灵测试”来考察计算机是否具有“心智”的时候，是否计算机也在通过“图灵测试”考察人类的“心智”呢？人类和计算机的交流难道不正像是“象罔”与陌生生物的交流一样吗？如果“象罔”通过与陌生生物之间相互的“图灵测试”而建构了他们共同的“心智”，那么，人类和计算机之间是否也正在通过相互的“图灵测试”而建构了人—机共同“心智”呢？

## 二、“反转图灵测试”与人机“心智”建构

接下来，本文将通过“反转图灵测试”的思维实验来论证：人与计算机之间的互动正是双方共同建构共有“心智”的过程。

为了理解人的智能与计算机智能之间的差异，我们可以设计一个“反转的图灵测试”，即让一个人努力伪装成为计算机，让参与测试的人误以为自己是在与一个计算机程序对话（蒋柯，2018）。当我们努力想要识别一个伪装成为计算机的人类时，就会察觉到哪些任务是计算机可以轻松地完成而人类是无法完成的。标准的“图灵测试”是为了识别一个伪装成为人类的计算机，所以，测试者所关注的问题在于哪些任务是人类可以完成而计算机不能完成的。如果“图灵测试”的标准得到认可<sup>①</sup>，那么两者相比，“图灵测试”是关于“计算机拥有与人类一样智能”的充分性检验，而“反转图灵测试”则是关于这个判断的必要性检验。也就是说，我们可以认为通过了“图灵测试”的计算机可能拥有和人类一

---

<sup>①</sup>尽管有诸如塞尔的“中文屋”论证等质疑意见，但是计算机科学领域依然把“图灵测试”当作检验计算机程序的智能水平的重要标准。由于篇幅的限制，本文不希望在此讨论“图灵测试”的理论意义。有关这个问题的讨论见笔者另外一篇文章《计算机模拟大脑与功能性计算策略》（《南京师大学报（社会科学版）》，2017年第1期）。

样的智能水平,也可能拥有比人类更高级的智能水平。而通过了“反转图灵测试”的人类可以被认为是拥有和计算机一样的智能,也可能拥有比计算机更高级的智能。所以,只有当“图灵测试”和“反转图灵测试”都被计算机或人类通过了,我们才可能说“人类和计算机拥有同样的智能”,或者“人类的智能和计算机的智能是等价的”。

因此,“图灵测试”作为检验计算机是否拥有智能的标准,本身只是一个“充分性论证”,而“反转图灵测试”是“必要性论证”。正是在设想如何通过“反转图灵测试”,以及如何识别伪装成为计算机的人类的过程中,人们才能认识到了人与计算机的差异。本文以为,这个过程同时还揭示了人类正在努力使自己的“心智”特征向计算机程序化的转移。

根据“图灵测试”的标准,通过了测试的计算机即有可能拥有与人类一样、甚至比人类更高级的智能。这引起了人类关于人自身意义的怀疑和恐慌。所以,一方面,心灵哲学家努力证明人类心智的独特性,即一定存在某些特征是计算机无法模拟的,诸如所谓自我意识、感受质、情绪和情感体验、道德两难判断,等等<sup>①</sup>。但是另一方面,心灵哲学家每提出一个特征,就成为计算机科学家模拟的目标,即努力通过计算机程序来实现和人类一样的表达性特征。双方的努力好比一场竞赛,人类努力向前跑(奔跑的方向其实是个问题),而计算机努力在追赶。当心灵哲学家提出来一个又一个心灵问题被计算机一一实现的时候,比如,当计算机通过了“图灵测试”、当计算机在被认为是最复杂的计算游戏——围棋中战胜了世界顶尖棋手时……人类开始担忧,计算机的智能总有一天会超越人类,进而开始怀疑人类自身心智的意义。这种怀疑是一次“哥白尼式革命”。它引起的震撼可能要比“地球不再是宇宙的中心”(哥白尼)、“人的心智不是受理性的掌控”(弗洛伊德)、“个体不再是认识活动的中心”(皮亚杰)这些经典的“去自我中心化”造成的震撼更强烈。

如果说“图灵测试”反映了计算机追逐人类智能的努力,那么通过“反转图灵测试”我们就能够看出人类努力奔跑的方向。当一个人努力想要模仿计算机时,他所做出来的伪装反映了他对于计算机活动特征的理解;当测试者想要识别是否是人伪装成为计算机时,他所设计的问题也表达了对于人和计算机的不同的看法。针对这个问题,笔者提出,简单而大量的数字计算以及大量符号的快速复制和存储等,这类任务是区分人和计算机的关键。这些任务可以揭示出,人与计算机之间的差异在计算机拥有人类无法达到的进行符号计算的速度。计算速度是计算机相对于人类的显著优势,也是人类制造计算机时努力追求的目标。人类按照自己需求和愿望设计并制造了计算机,计算机所能够实现和应该实现的目标正是人类希望能够实现却不能够独立实现的目标。这好比汽车——可以相应地被称为“移动机”——被人类制造出来是为了实现超越人类身体极限的移动速度。反过来,汽车的出现也改变了人类的迁移方式,这是人类和汽车之间的共同建构。从计算机最初被建造出来到逐步升级,计算机的出现也正在改变着人类的思维方式,这个过程也体现了人与计算机之间的共同建构。首先,人制造机器是为了实现机器“像人一样工作”的目的,即机器总是对人的模仿,或者替代人来实现原本由人完成的工作。另一方面,人制造机器也表达了某种自身的期望,比如制造汽车是为了实现比仅仅依靠人力更快速地移动,而制造计算机则是为了实现比人的计算更快速的计算。即人按造自己的工作要求来制造机器,同时也通过机器表达自己的发展预期。当仅仅依靠人力无法达成预期目标时,人就制造某种机器来辅助以实现这个目标。在这个意义上,人类是在借助机器而朝着机器所限定的方向发展。也就是,人制造“像人一样的”机器的同时,也正在使得自己朝着“像机器一样”的方向发展。

在“反转图灵测试”中,如果伪装计算机的人作弊,他身边有一台便携式计算机,随时可以帮助

---

<sup>①</sup>以内格尔围绕心身关系所做的一系列讨论最著名,例如他(1965)对物理主义的论证,以及(1974)著名的“蝙蝠论证”等。

他快速地完成符号计算工作,那么测验者将很难识别出他是人还是计算机。这就是说,只要人稍微借助一点计算机的帮助,就可能成功地伪装成为计算机。同样地,如果计算机在“图灵测试”中作弊,也只要稍微借助一点人的帮助,就能够成功地骗过测试者。这说明,自计算机的诞生起,通过人和计算机的互动,随着计算机的不断升级以及人类自身认知体系的演变,人和计算机之间的差异正在逐步消解,即机器按照人类的设计逐步学习人类的“思维方式”,而人类自己的知识和心智建构也正在朝着能够适应与机器的合作的方向发展,也就是“人正在变得像机器一样”。在现实生活中,人们常常是和计算机一道协同工作的。当我们通过网络与另一个“人/计算机”发生互动时,我们怎么知道“他/它”究竟是人还是计算机,还是人和计算机的联合呢?这时,人和计算机之间的界限已经消解了。

这就是:人建造了计算机,并且和计算机一起建构了“人一机心智”的意义。

### 三、从生命逻辑向科学逻辑的趋同

我们通过“反转图灵测试”的论证揭示了人类自身的思维正在朝着与机器协同的方向演化。相反,通过与人类的互动,机器正在变得越来越“智能化”,所谓智能就是指机器在与人类的互动过程中表现得越来越良好的适应性;我们不能否认,在这个过程中,人类自己也在发生某种改变,使得人与机器的合作越来越容易。这种双方面的改变就是“人一机心智”的建构过程。接下来,本文将引证生命逻辑向科学逻辑趋同的发展趋势,进一步阐释“人一机心智建构”的过程。

逻辑作为一种支持公共交流的法则,其本身不等于“意义”,更不同于“心智”。但是逻辑的使用和逻辑特征的演变却能够反映出心智实现意义交流的需要。如前所述,是公共交流建构了人类的“心智”,而交流双方持相同或相似的逻辑体系正是支持公共交流的基础。所以可以说,如果没有交流双方共同认可的逻辑体系,就不可能实现“意义”的建构,更不可能有“心智”的建构。当然,交流双方共有的逻辑体系本身也是建构的结果。于是,我们应该说,“逻辑”、“意义”和“心智”三者在时间上或空间上的顺序区分是没有意义的,在某种意义上它们是“统一的”,它们都是交流过程的建构。

人们在日常生活中使用的逻辑体系可以有很多种,我们不妨称之为“生命逻辑”,但是被官方话语所认可的、被当作“逻辑学”的标准体系的只有一种,本文称之为“科学逻辑”。“生命逻辑”是人类作为一种生命形态在人—环境、人—人互动过程中所采用的逻辑规则。这种逻辑规则首先体现了生命活动的基本追求,这就是生命本身的存活和种族的存活。因而这样的逻辑规则具有有限适用性、任务指向性以及领域特殊性等特征,即这样的逻辑规则可能仅仅针对某一种任务有效,或者仅仅在特定的环境中有效。于是,为了应对不同环境中的生存任务,每一个人不得不拥有许多套不同的“生命逻辑”,比如,专门用来指导觅食活动的逻辑、专门用来指导求偶活动的逻辑,等等(巴斯,2015)。自然选择学说为这样一些专门化逻辑的形成提供了理论上的解释。这样安排的好处是,由于每一种逻辑都是有限制的,所以它的规则很简单,个体只需要经过很短的学习经历,甚至不需要学习经历就可以使用它;因为它很简单,所以在使用时人的应答速度会很快。而相应的缺点是,这种逻辑体系是有限的,所以人必须要具备很多种不同的逻辑以适应不同任务。尽管如此,人依然会遇到超出已有逻辑种类的新任务。每当遇到这样的新任务,人可能采取两种策略,一是用某种原有的逻辑来将就对付,另一种是将某种原有逻辑经过一定修改来应对新任务。无论是哪种策略,都可能会有很大的误差,所以,一旦遭遇新任务,人犯错误的概率就会很大。

“科学逻辑”则是人类进入文明纪元以后被理智发明出来,用以规范所有人的思维与行动的规则。这种逻辑的特征是:希望用一个统一的规范来限定人类全部活动;对各种任务都具有一般性的适用性;拥有稳定的真值评判标准和推理规则,以及尽可能地实现对所有任务的形式化表征。我们可以看

出,“科学逻辑”具有比“生命逻辑”广泛得多的适用性,我们几乎可以用这样一套逻辑规则来指导所有的活动;并且它能够有效地消除人与人之间的交流障碍。交流的双方分别使用不同的逻辑是人际交流最大的障碍。但是这一套规则非常复杂,需要人进行长时间专门化的学习才能掌握。实际上,文明纪元的人类从幼年时期开始就必须接受这套逻辑规则的训练(奚家文等,2014),但是仍然有很多人直到成年也不能正确地使用它。“科学逻辑”的最重要优势是它适合于实现机器的智能化工作。因为“科学逻辑”是一套形式化的规则,而机器智能能够处理的正是经过了形式化处理的任务。我们可以看到,随着计算机的运算能力的提升,其执行“科学逻辑”的效能也越来越强,终于出现能够在围棋中战胜人类的计算机程序。但是这是否代表了计算机智能达到甚至超越人类的智能呢?

如前所述,在“科学逻辑”之外,人类还使用着另一套“生命逻辑”。“生命逻辑”和“科学逻辑”不同,所以目前计算机能够很好地执行“科学逻辑”并不能代表它们的智能全面地达到了人类的水平。而我们只能说,在执行“科学逻辑”时,计算机智能具有比人类更快的速度和精确性。我们还要认识到,人类发明“科学逻辑”一开始并不是为了适用于计算机的工作,人类开始研究“科学逻辑”时,计算机的出现还未被当时的人们所预计。事实上,是后来的人们依据“科学逻辑”的形式化计算特征而设计了计算机。这再一次体现了前文所论述的人—机关系,即人类首先对自身的心智发展具有某种预期,当自己的心智能力不能完全实现事先的预期时,人类制造了某种机器来增强心智的能力,这就是计算机。人类发明了“科学逻辑”,因为人类对自身的心智活动具有规范性和形式化的预期,而当人类自身的心智在执行“科学逻辑”时遭遇瓶颈时,人类制造了计算机来帮助自己更加快速和精确地执行这一套“科学逻辑”。所以,计算机在执行“科学逻辑”时所表现出来的智能化特征正是人类在发明“科学逻辑”时所预期的,如果说计算机有“心智”的话,这种心智并不是计算机自己拥有的,而是人类设计出来的。与此同时,人类之所以发明“科学逻辑”,正是希望自身的“心智”朝向更加一般化、形式化的方向发展,这也正是计算机“心智”的特征。

在心智活动的规则,即逻辑的意义上,人类对逻辑的依赖以及逻辑自身的演变特征等都揭示了人类心智与计算机“心智”正在朝着同样的方向发展。

## 四、小结

只有我们承认“心智”是公共交流的建构,才可能走出“我思故我在”所设置的“唯一心智”悖论。但是这种出走是对“自我”的叛逆,即只有当人不再将自己当作全世界的唯一中心,不再以自己作为衡量世界上其他的心智存在的标准时,才能够认识到心智作为公共交流的建构而具有的一般性存在意义。这种叛逆是对“自我中心”的出离,是皮亚杰式“去自我中心化”的发展(皮亚杰,1997)。

当人类走出了“自我中心”之后,又遭遇了机器智能的挑战。这一次,是人—机交流——即人与物之间的互动——建构了存在于人与物之间的公共心智。这将成为人类心智的第二次“去自我中心化”。第一次“去自我中心化”是人类出离了个体而建构人类的公共心智,而第二次“去自我中心化”则是人类出离人类自身并与机器一起建构“人—机心智”。

这将是一次彻底的叛逆与出离。

## 参考文献:

[美]巴斯,2015:《进化心理学》,张勇等译,商务印书馆。

陈晓平,2015:《功能主义与物理主义的趋同与分异——评当代心灵哲学的走向》,《现代哲学》第3期。

[法]笛卡尔,2000:《谈谈方法》,王大庆译,商务印书馆。

高申春, 2005:《心理学:危机的根源与革命的实质——论冯特对后冯特心理学的关系》,《吉林大学社会科学学报》第45卷(第5期)。

蒋柯, 2016:《身心统一的功能哲学进路》,《南京师大学报》(社会科学版)第1期。

蒋柯, 2017:《计算机模拟大脑的功能性计算策略》,《南京师大学报》(社会科学版)第1期。

蒋柯, 2018:《阿尔法元动了谁的奶酪》,《中国社会科学报》2月27日第6版。

[美] 詹姆斯·黑, 2013:《心理学史, 心理学思想的主要流派》(第6版), 蒋柯等译, 上海人民出版社。

刘晓力, 2014:《当代哲学如何面对认知科学的意识难题》,《中国社会科学》第6期。

[美] 塞尔, 2006:《心、脑与科学》, 杨音莱译, 上海译文出版社。

[瑞士] 皮亚杰, 1997:《发生认识论原理》, 王宪钊译, 商务印书馆。

王佳、朱敏, 2010:《对强人工智能及其理论预设的考察——基于中文屋论证的批判视角》,《心智与计算》第1期。

奚家文、蒋柯, 2014:《进化而来的心理是功能还是结构》,《华东师范大学学报》(教育科学版)第1期。

熊哲宏, 2002:《认知科学导论》, 华中师范大学出版社。

杨晓爱、魏屹东, 2013:《中文屋论证究竟证明了什么? ——由“中文屋论证”引发的计算机思维问题探析》,《山西大学学报》(哲学社会科学版)第36卷第2期。

周晓亮, 2005:《自我意识、心身关系、人与机器》,《自然辩证法通讯》第4期。

Minsky, M., 1985, *The Society of Mind*, New York: Simon & Schuster, Inc.

Nagel, T., 1965, “Physicalism”, *The Philosophical Review*, vol.74, pp.339—356.

Nagel, T., 1974, “What is it like to be a bat?”, *The Philosophical Review*, vol.83, pp.435—450.

Turing, A. M., 1950, “Computing Machinery and Intelligence”, *Mind*, vol. 59, pp.433—460.

Wallach, W. & C. Allen, 2009, *Moral Machines: Teaching Robots Right from Wrong*, Oxford: Oxford University Press.

(责任编辑: 蒋永华)

## Turing Test, Reversal Turing Test and the Meaning of Mind

JIANG Ke

**Abstract:** Turing test was designed to determine whether the computers have possessed the same intelligence as human beings. From the day that Turing test was established, it has also become the guide for the projects of artificial intelligence (AI). It has been thought that every project of AI should pass the Turing test before it is accepted as successful. This article analyzes the historical meaning of Turing test and argues that human mind is actually a historical construction, and it is the social communication of human beings that shapes the computational logic of human mind. Since the computer emerged, the interactions between the mind of human beings and AI have also produced a so-called “human-computer mind.” The construction of “human-computer mind” is sure to become a post-modern approach to the integration between human mind and AI.

**Key words:** artificial intelligence; Turing Test; Reversal Turing Test; mind; construction