

道德关怀范围的持续扩展

——从非人类动物到无生命的机器人

杨通进*

〔摘要〕 人类道德进步的历史就是道德关怀的范围不断扩展的历史。环境伦理学的重要理论贡献是,区分了道德行为体与道德承受体,把道德承受体的范围扩展到了人类物种之外的自在物。当代机器人伦理学接过环境伦理学扩展道德关怀的接力棒,不仅把道德承受体的范围从自然存在物扩展到了作为人工制品的机器人,还把道德行为体的范围扩展到了具有道德功能的社会机器人。随着道德关怀范围的扩展,出现了三种不同类型的道德行为体和五类不同的道德承受体。为了应对伦理共同体成员的增加所带来的伦理挑战,需要建构具有拓展主义、非人类中心主义、后科学主义与全球主义性质的新型伦理文化。

〔关键词〕 道德行为体 道德承受体 道德关怀 机器人伦理学

〔中图分类号〕B82-058 〔文献标识码〕A 〔文章编号〕1007-1539(2020)01-0106-009

DOI:10.13904/j.cnki.1007-1539.2020.01.016

道德关怀是指道德行为体对伦理共同体中的其他成员所给予的某种符合道德要求的待遇(moral treatment)。纵观各民族以及人类社会发展的历史,我们可以发现一条比较清晰的线索,即道德进步的历史就是道德关怀的范围不断扩展的过程。从原始社会到20世纪六七十年代,人类的道德关怀完成了从家庭到家族、从氏族到种族、从种族到民族同胞再到所有人类成员的扩展过程;从19世纪中叶开始,人们的道德关怀范围开始突破人类的界限,逐步拓展到人类之外的动物、植物以及作为一个整体的大自然。近年来,随着机器人技术的迅猛发展,一些高度智能化的“机器人”已经或即将高度参与并介入人类的私人生活与社会公共生活。于是,机器人是

否应当获得道德关怀的问题开始引起人们的关注,并成为近年新兴的机器人伦理学(robotics ethics)的一个重要主题^①。为厘清道德关怀的扩展逻辑及其伦理基础,本文拟首先探讨道德关怀的具体内涵,说明区分道德行为体与道德承受体的意义,然后在此基础上探讨扩展道德承受体与道德行为体之范围的伦理理据及其可能的限度,最后探讨道德关怀范围的扩展给人类的道德生活所带来的伦理挑战以及可能的应对策略。

一、道德行为体、道德承受体与道德关怀

道德关怀(moral consideration 或 moral concern)这一概念是随着美国圣母大学教授古德帕斯特发表的《论道德关怀》(1978)一文而逐

* 作者简介:杨通进,广西大学马克思主义学院教授,中国特色社会主义道德文化省部共建协同创新中心首席专家,博士生导师(广西南宁 540004)。

① 本文所讨论的机器人是以强人工智能为基础,在外表、行为方式、交往方式、思维与推理方式等各方面都与成熟的人类个体高度相似的社会机器人。获得沙特公民资格的索菲亚和电影《机械姬》中的爱娃是本文所谈论的机器人的原型。能够享有完整的道德地位的机器人将会是未来那些比爱娃更为先进的机器人。

渐成为当代应用伦理学(尤其是环境伦理学)的一个重要概念工具。在《论道德关怀》一文中,古氏试图探讨的问题是:对于所有的行为体A来说,X应当获得A的道德关怀的充分必要条件是什么?在古德帕斯特看来,这里的行为体指的是“理性的道德行为体”(rational moral agent),而道德关怀指的是“实践意义上的最基本的尊重”^[1]。当然,就道德实践而言,尊重还是一个比较抽象的原则;只有确认了尊重在具体的道德实践中的行为要求,我们才能真正把握尊重原则的规范意涵。就本文的论题而言,我们需首先澄清“道德行为体”与“道德承受体”这两个关键概念的内涵,才能真正确认尊重这一道德原则的具体内容。

在当代应用伦理学领域,“道德行为体”(moral agent)与“道德承受体”(moral patient)是两个使用比较广泛且具有紧密联系的概念^①。道德行为体是指“所有具有以下这些能力的存在(being):它能够做出道德的或非道德的行为,能够具有责任和义务,能够对自己的行为承担责任”^②。这些能力可称之为道德能动性(moral agency),其中最重要的包括:判断是非的能力;从事道德慎思的能力,即思考和权衡赞成与反对各种可供选择的行为方案之道德理由的能力;根据这些道德理由做出决定的能力;发挥必要的决心与意志力来实施这些决定的能力;因未能实施这些决定而对他人承担责任的能力^{[2](8)}。正常而成熟的人类个体是典型的道德行为体。但是,并非所有的人类个体都是道德行为体。有些人(如儿童、青少年)是不完全的道德行为体,只对其行为承担部分道德责任。有些人(如婴儿、智障人士、

胎儿)则不是道德行为体,不对其行为承担道德责任。因此,“不能把人类种属简单地等同于道德行为体种属。原因是双重的。首先,并非所有的人都是道德行为体。其次,或许还存在非人类的道德行为体”^{[2](8)}。

道德承受体指的是具有道德承受性(moral patiency)的主体,它们有资格获得道德关怀。道德承受体“或许缺乏道德行为体的能力,但却拥有道德行为体对其负有责任的实体身份”,因为后者的行为能够使前者的福利或生存条件得到改善或恶化^{[2](9)}。拥有正常道德能动性的人类个体既是道德行为体,也是道德承受体,无疑应当获得其他道德行为体提供的“道德服务”。但是,那些只具有部分道德能动性的人类个体(如儿童、青少年以及其他不正常的人类个体)甚至完全不具备道德能动性的人类个体(如植物人、胎儿、婴儿)也是道德承受体。不仅如此,“非人类动物也是道德承受体”,道德行为体在进行道德慎思时有义务把它们利益也考虑进来^[3]。因此,道德承受体的种属比道德行为体的种属要更为宽广。

毫无疑问,道德行为体与道德承受体都是道德主体(moral subject),拥有道德地位,都是道德共同体的成员,有资格获得道德关怀。道德行为体对道德承受体所给予的最基本的道德关怀就是“尊重”。泰勒指出,尊重是一种终极的道德态度。“只要一个人的行为符合某一套规则所制定的要求,他的行为就可以被认为是体现了尊重自然的态度。”^{[2](55)}因此,判断一个道德承受体是否获得了恰当尊重的标准就是,道德行为体对

① 笔者早年曾把 moral agent 与 moral patient 分别译为“道德代理人”与“道德顾客”。在环境伦理学的语境中,人类是唯一具有道德思维能力的存在者,他们代表地球上的所有存在物提出某种能够获得理性辩护的道德体系,并向那些有资格获得道德关怀的“道德顾客”提供道德服务。因此,在环境伦理学的语境中,把 moral agent 与 moral patient 分别译为“道德代理人”与“道德顾客”不会引起太大的歧义(有学人曾对笔者的理解提出异议,参见姚晓娜:《moral agent 是道德代理人吗》,《道德与文明》2010年第1期)。但是,在机器人伦理学的语境中,机器人也可能是 moral agent。因此,本文遵循多数学人的惯例,把 moral agent 译为“道德行为体”,并相应地把 moral patient 译为“道德承受体”。

② 参见[美]保罗·沃伦·泰勒:《尊重自然:一种环境伦理学理论》,雷毅、李小童、高山译,首都师范大学出版社2010年版,第8页。译文略有改动(后文的改动不再注明)。泰勒使用的两个术语是 moral agent 与 moral subject,《尊重自然:一种环境伦理学理论》(Respect for Nature: A Theory of Environmental Ethics)的中译本把这两个术语分别翻译成“道德代理人”与“道德主体”。后者的含义与本文提到的 moral patient 相同。

他/她/它的行为是否遵循了一套恰当的、得到合理证明的道德规则。在泰勒看来,人类(作为道德行为体)要想对自然(作为道德承受体)表现出尊重的态度,那么,他/她的行为就必须遵守四条基本的道德规则:不伤害(nonmaleficence)、不干涉(noninterference)、忠诚(fidelity)、补偿正义(restitutive justice)^{[2](110-122)}。如果一个道德行为体在与道德承受体交往时遵循了这些基本规则,那么,他/她/它就是给予了后者基本的道德关怀。当然,遵守这些规则还只是道德行为体的“显见义务”或“初始义务”(prima facie duty)。在不同的道德境遇中,针对不同的道德承受体,道德行为体所需履行的“实际义务”会表现出一定的差异,对于那些既是道德承受体又是道德行为体的义务对象,“尊重”原则所包含的内容会更为丰富。

二、道德承受体范围的扩展

当代环境伦理学最重要的理论贡献之一,就是明确区分了道德行为体与道德承受体这两个概念。传统伦理学的一个未言明的假设是,所有的人都是道德行为体,只有道德行为体才有资格获得道德关怀并成为道德承受体。但是,人们的道德实践却表明,很多人类个体都不具备道德能动性,但是,人们对这些人的行为也要遵守某些基本的道德原则。同时,随着动物保护意识和环境保护意识的提高,人们越来越发现,传统的主流伦理学很难解释虐待动物与随意破坏自然环境等行为的错误。此外,生命伦理学关于“胎儿”道德地位问题(胎儿尚未出生,但是他/她们仍享有道德承受体的道德地位)的争论更是使得传统主流伦理学的概念工具捉襟见肘。于是,环境伦理学对道德行为体与道德承受体所做的概念区分逐渐被当代应用伦理学所接受,并为我们证明作为道德行为体的人对不具备道德能动性的人(作为道德承受体)所负有的道德义务提供了重要的概念工具。这两个概念的区分表明,拥有道德能动性、成为道德行为体并不是获得道德关怀的必要条件,尽管拥有这些能力是获得道德关怀的充分条件。这意味着,那些具备道德能动性的

存在物可以自动地成为人们道德关怀的对象,但是,我们却不能仅仅因为一个存在物(不管是人类还是非人类)不具备道德能动性就自动地把他/她/它排除在道德关怀的范围之外。

当代环境伦理学的另一个重要理论贡献是扩展了道德承受体的范围。环境伦理学的创始人之一史怀泽重新界定了善恶的含义。在他看来,“善是保持和促进生命,恶是阻碍和毁灭生命”^{[4](19)}，“伦理就是敬畏我自身和我之外的生命意志”^{[4](26)}。人类只有摆脱对自己的偏爱,抛弃对其他生命的疏远,与自己周围的生命休戚与共,才是道德的,只有这样,人才能成为真正的人。这样,史怀泽就把道德关怀的对象扩展到了所有的生命。环境伦理学的另一创始人利奥波德则是通过扩展伦理共同体成员的方式来完成扩展伦理关怀范围这一任务的。在《沙乡年鉴》的最后一章“土地伦理”中,利奥波德首先讲述了《荷马史诗》中的英雄奥德赛的故事。奥德赛在一次战争凯旋后回到家乡时用绳子吊死了12个女奴。利奥波德解释说,奥德赛杀死女奴的行为并未遭受任何谴责,因为女奴只是财产,不是伦理共同体的成员。但是,随着人类文明的进步,人们逐渐把所有的人都纳入了伦理共同体的范围。在利奥波德的时代,人类还未开始考虑大地共同体的问题,而大地伦理学的任务就是继续扩展伦理共同体的范围,把人类和地球上的其他自然存在物都视为同一个伦理共同体的成员,“把人类的角色从大地共同体的征服者改造成大地共同体的普通成员与公民”^{[5](80)}。因此,根据利奥波德的大地伦理学,大地共同体中的所有成员以及大地共同体本身都享有道德承受体的道德地位。在《大自然的权利》一书中,纳什则把人类道德进步的历程明确地阐释成道德关怀与权利主体的范围不断扩展的过程^{[5](3-6)}。

当代环境伦理学的一个基本共识是,虽然人类是唯一的道德行为体,但是,道德承受体的范围却不限于人类。从不同角度扩展道德承受体的范围,构成了当代环境伦理学的重要主题。从不同的伦理学前提与价值预设出发,当代环境伦理学把道德承受体的范围扩展到了动物(至少是

高等动物)、所有的生命存在物以及作为整体的大自然,并形成了三个重要的理论流派:以辛格、雷根为代表的动物中心主义(zoocentrism)认为所有的动物都享有道德承受体的道德地位^①;以泰勒、阿提费尔德为代表的生物中心主义(biocentrism)把所有的生命都视为应当获得道德关怀的道德承受体;以奈斯、罗尔斯顿为代表的生态中心主义(ecocentrism)则把道德承受体的范围扩展到了作为整体的自然生态系统以及这个系统的构成要素^{[6](93-132)}。

如果说当代环境伦理学还只是试图把道德承受体的范围扩展到自然存在物,那么,当代机器人伦理学则接过环境伦理学扩展道德关怀范围的接力棒,试图把道德承受体的范围扩展至人工物——人类设计并制造的机器人。当代机器人伦理学主要从四个角度论证了机器人的道德承受体地位。

(一)间接义务理据

在现代文明社会,绝大多数人认为,虐待动物的行为是错误的。但是,人们用来证明这种错误行为的伦理理据则分为两种:即直接义务论与间接义务论。根据直接义务论,虐待动物的行为之错误之处在于,这种行为或者损害了动物的利益(动物解放论),或者侵犯了动物的权利(动物权利论)。间接义务论则认为,虐待动物的行为之所以是错误的,乃是由于这种伤害会导致对他人的伤害,而伤害他人是错误的;“不伤害动物”的义务是从“不伤害他人”这一直接义务中推导出来的一种间接义务。阿奎那(1225—1274)较早地表达了间接义务论的观点:“基督教的任何教义看起来都禁止我们去残忍地对待那些不能开口说话的动物,比如,禁止杀死幼鸟;这或者是因为人们会把这种思维转移到残忍地对待其他的人——既然会对动物残忍,也就会对他人残忍;或者是因为对动物的伤害会导致对他人的现世伤害——或是实施行为,或是其他暴行。”^②因

此,在阿奎那看来,对动物的残忍行为之所以是错误的,乃是由于这种残忍行为会间接地导致对他人的伤害。康德也明确提出了类似的观点:“对动物残忍的人在处理他的人际关系时也会对他残忍。我们可以通过一个人对待动物的方式来判断他的心肠是好是坏。”^{[7](25)}

上述理据完全适用于人与机器人的关系:人们拥有不虐待机器人的义务,因为,虐待机器人的人人会养成虐待其他人的心理趋向与行为习惯。人们对网络暴力游戏的一个普遍担忧是,这种游戏会让游戏者养成“暴力思维”的习惯,并对现实中的暴力行为变得麻木不仁。网络游戏中的暴力还只是一种“虚拟的暴力”,可是,如果人们在现实生活中以虐待机器人为乐,那么,这种虐待就不是一种“虚拟的虐待”,而是“实实在在的虐待”,是对虐待行为的实施、练习与强化。如果网络中的“虚拟暴力”都值得我们担忧,那么,人机交往中的这种“真实的虐待”就更值得我们担忧了。因此,为了使人们避免养成虐待他人的思维方式与行为习惯,我们就必须禁止对机器人的虐待行为。“即使在一个最为宽容的社会里,对虐待机器人的行为施加某些限制也是合理的。”^[8]

(二)自我建构理据

康德在论证人对动物负有的间接义务时,除了诉诸对他人的伤害理据,还诉诸人性与美德理据。在康德看来,履行对动物的义务,是我们的人性的展现。如果一条狗长期服务于它的主人,那么,“当这条狗老了不能再为它的主人服务时,它的主人应该照顾它,直到它死去”。相反,“如果一个人因为他的狗没有能力再为他服务而把它给杀了……[那么],他的行为就是不人道的,是对他自身的人性的损害,而他有权向他人展现出这种人性”^{[7](25)}。在这里,不承担对动物的义务的行为所涉及的,就不是对他人的间接伤害,而是对人性本身的损害。所以,康德实际上

^① 英美学术界较多使用的是动物解放/权利论(animal liberation/rights theory)这一术语,欧洲大陆学术界则经常使用痛苦中心主义(pathocentrism)一词;近年来,人们逐渐涌现出使用 zoocentrism 的趋势。

^② 参见〔意〕阿奎那:《理性生物与无理性生物的区别》,〔澳〕彼得·辛格、〔美〕汤姆·雷根主编:《动物权利与人类义务》,曾建平、代峰译,北京大学出版社2010年版,第10页。译文略有改动(后文的改动不再注明)。

是把关心动物当作人性的内在要素来加以理解和建构的。换言之,一个正常的理性的人必须要把对动物的关怀当作一种直接的义务来认可与承诺。这关乎的不仅仅是动物的“应得”,更关乎人性的完整。康德还认为,关心动物是人性的内在情感,一个人如果“不想扼杀他的人性情感(human feelings),他就必须要以仁慈(kindness)的方式对待动物”^{[7](25)}。同理,如果人们虐待那些为他们提供服务、能够与他们交流并能遵循人类行为规范的机器人,那么,他们所伤害的就不仅仅是机器人,他们还扭曲了自己的人格,减损了人性本身的光辉,丧失了应有的美德。以一种形象展现在他人与机器人面前的人,建构与认可的是一种有缺陷的、不完美的人格。

(三)行为主义理据

在讨论人对动物的义务时,康德曾指出,一条狗对它的主人提供的服务,类似于仆人所提供的服务;因此,“当这条狗老了不能再为它的主人服务时,它的主人应该照顾它,直到它死去”^{[7](25)}。康德这里提到的实际上是人与动物之间的相互性或互惠性(reciprocity)。在实际的生活中,人们都会对那些长期为其提供服务的家畜或伴侣动物抱有某种感情,人们会把它们当作值得从道德上加以关怀的道德承受体来对待,人们还会为那些“义犬”“勇敢的战马”等树碑立传。未来以强人工智能为基础的机器人给我们提供的服务无疑将比这些动物还要周到细致,而且,它们还能遵循人类交往的基本礼仪与道德规范。根据康德的相互性理念,机器人享有的道德地位不应比动物更低。

在一些学者看来,机器人即使遵循了人类的道德原则,它们也不是道德行为体,因为它们并不理解那些道德原则的内容,而且,它们并不是在依据道德原则行事,而是按照事先编程的软件系统在运行^{[9][10]}。这里涉及几个目前正在展开的理论争论。首先,现代伦理学在评价一个行为或政策(以及法律或制度安排)的道德价值时,大多采取后果主义立场,即更多地关注行为的后果,而不去追问行为主体的内在动机或主观意愿。其次,自然人的主观动机比较容易识别,但

是,法人(如公司、各类政府与非政府组织等)的动机却难以确定,法人没有自然人那样的中枢神经系统,法人本身没有苦乐感受的能力。在对法人的行为或决策进行道德评价时,我们主要根据该行为或决策是否符合或遵循了大多数人所认可的原则或价值,是促进还是伤害了相关主体的合法利益,而不是去追问这个法人的“动机”是什么。再次,根据行为主义理论,我们无法直接把握人们的主观动机与内在想法,只能通过其外在行为来推论其主观动机,我们无法在人们的外在行为与其主观动机之间做出截然的区分。因此,在操作的意义上,我们可以把人或机器人的外在行为等同于其内在动机。在评价人或机器人行为的道德价值时,我们只需关注其行为的后果,无须追问其主观动机。最后,从决定论的角度看,人的道德决策与机器人的道德决策之间的区别并没有人们想象的那样大。人的道德抉择也是由一系列相关的前置条件所决定的。当代的机器人伦理学正是从这种决定论的角度来理解、设计与制造“道德机器人”的。因此,从行为主义的角度看,只要机器人做出了与人大致相当的行为,我们就应当把机器人视为与人大致相当的主体来对待。“如果某个行为体在行走、交谈与行为方式方面都足够与我相似,那么,我即使不能合理地认为它拥有心灵(mind),我也有义务把它当作它好像是道德行为体那样来对待。”^[11]

(四)人机共同体理据

当代机器人伦理学的一个争议热点是,我们是否应当研制性爱机器人(sex robots)。美国学者勒维在2008年曾著书预言:2050年左右,人们不仅能与机器人做爱,他们还想与机器人结婚,与机器人保持浪漫的伴侣关系^[12]。反对研发性爱机器人的英国学者理查森认为,研制性爱机器人有物化女性与儿童的嫌疑;对性爱机器人的使用和消费会助长人类的性虐待行为,长此以往会使人们对其性伴侣失去同情心;使用和消费性爱机器人会导致性交易的泛滥,进而对人与人之间的关系产生负面影响^[13]。此外,目前所研发的性爱机器人缺乏“知情同意”的程序设置。在这种情况下,人们对性爱机器人的使用与消费不仅

带有“强奸”的色彩,而且还会扭曲许多人的心灵,使他们误以为在与潜在的人类伴侣发生性行为时,后者的知情同意可有可无。^{[14](186-212)}

为了避免出现这些负面的结果,我们就必须要用尊重与知情同意的原则来约束人们与其机器人伴侣之间的交往行为。在与作为性爱伴侣的机器人发生性行为时,人们必须要征得对方的同意。^[15]否则,我们认可的就会是一种包含了“强奸文化”的交往规则,我们与性爱机器人之间建立的就是一种剥削的、不平等的关系。而一种包含了剥削的、等级性的共同体是一种不健康的伦理共同体。因此,作为理性的成熟的道德行为体,我们必须要认可和接受性爱机器人的道德承受体地位,把人与性爱机器人都看成同一个人机伦理共同体的成员来看待。否则,我们的共同体就会蜕变成某种接受“强奸文化”的畸形的伦理共同体。在这种畸形的共同体中,作为性别歧视主义的态度、制度与行为模式的结果,人与机器人之间的未经知情同意的性活动将变得常态化,“强奸型人格”就会成为人机共同体中的一颗毒瘤侵蚀伦理共同体的健康肌体。因此,如果我们不得不与性爱机器人“比邻而居”,那么,我们就必须要认可它们的道德承受体地位。如果我们无法否认性爱机器人的道德承受体地位,那么,我们同样无法否认护理机器人、看护机器人、助理机器人、演艺机器人等社会机器人的道德承受体地位。我们与这些社会机器人组成的共同体同样属于伦理共同体,这样的伦理共同体不能奉行否认其他成员之道德承受体地位的伦理歧视主义的文化。

三、道德行为体范围的扩展

如果说当代环境伦理学的工作重心是扩展道德承受体的范围,那么,当代机器人伦理学的追求目标则是扩展道德行为体的范围。事实上,在机器人伦理学诞生之前,即使最激进的环境伦理学也只是认为,某些高等动物可能拥有某种道德情感或初级的道德能动性。泰勒在提到“非人类的道德行为体”时使用的界定词是“或许还存在”,他心目中的非人类道德行为体指的是其他

的高等动物——一种自然的行为体(自然存在物),而非人工的行为体(人工制造物)^{[2](8-9)}。历史上还没有一位思想家严肃认真地思考或探讨过人工道德行为体的可能性问题。当代机器人伦理学的革命性变革表现为,它首先探讨的就是人工道德行为体(artificial moral agents; AMAs)而非人工道德承受体(artificial moral patients)的可能性问题;它对传统伦理学的最大挑战之处在于,它试图扩展道德行为体(而非道德承受体)的范围。

因此,当代机器人伦理学的目标与使命就是探讨人工道德行为体的可能性与必要性。国际著名的机器人伦理学家、《道德机器:如何让机器人明辨是非》(2009)一书的作者之一艾伦明确指出,“研发人工道德行为体(AMAs)的终极目标应当是制造道德上值得称赞的行为体”^[16]。美国康涅狄格大学哲学教授、《机器伦理学》(2011)一书的主编安德森亦认为,机器伦理学的核心目标是制造自主的道德机器^{[17](21-27)}。事实上,人工智能与机器人伦理学的主流声音是,“AMAs是必须的也是必然的”^{[18](33)}。研究与开发人工道德行为体不仅是可欲的,而且势不可挡。

鉴于研发具有道德能力的智能机器人是机器人技术的既定目标,因而很多学者认为,机器人不仅是道德承受体,还能够成为道德行为体。在美国索诺马州立大学的哲学教授萨林斯看来,自主性、意向性、履行责任是成为道德行为体的三个充分条件;将来的机器人能够拥有这三种能力。“可以肯定,只要我们追求这种技术,那么,未来高度复杂的、具有交往能力的机器人将是拥有相应权利与责任的道德行为体。”^[19]土耳其学者达文波特亦认为,“与机器一样,人类及其道德都是可计算的,因而,前者可以被建造成一个道德行为体”^[20]。

事实上,不仅大多数机器人伦理学家认为,机器人可能成为道德行为体,而且,他们还道德行为体的类型做了区分。例如,穆尔曾区分了四种伦理行为体。(1)其行为具有伦理后果的伦理行为体(ethical impact agents),即代替人类从事重复、枯燥或危险工作的行为体。(2)隐性的

伦理行为体(implicit ethical agents),即安全性与可靠性程度较高、对它的使用不会给人类带来负面影响的行为体。(3)显性的伦理行为体(explicit ethical agents),即伦理原则与规范已经编入其软件系统、且能依据“道义逻辑”进行推理的行为体;获得沙特公民资格的索菲亚已初步具备显性伦理行为体的雏形,而电影《机械姬》中的“爱娃”则是合格的显性伦理行为体,已经属于享有道德行为体地位的机器人。(4)完全的伦理行为体(full ethical agents),即能够做出清晰的道德判断、具有较大的道德自主性的行为体。^[21]一般认为,只有具有意识、意向性与自由意志的存在者才能成为完全的伦理行为体。这样看来,第一类、第二类行为体都不能算是道德行为体,因为它们的行为不是出于特定伦理原则或道德规范的引导。第三类行为体符合道德行为体的基本定义,因而能够发展成为人工道德行为体。第四类伦理行为体的标准比较高,只有那些正常的理性的人类个体(而非所有的人类个体)才会成为这类道德行为体。“鉴于我们目前对道德推理、人工智能、认知机制等的理解,我们至多能够制造出显性的伦理行为体,这种伦理行为体能够做出某些道德判断(这些判断不是作为软件预先编入其系统中的),并有能力对它们为何会做出那些道德判断做出解释。”^[22]

一些学者认为,机器人不仅能够成为道德行为体,将来还会是拥有权利的道德行为体。荷兰学者科齐伯格从人机共同体的角度证明了机器人拥有权利的可能性^[23]。美国学者冈克尔根据“能够”与“应当”两个变量,指出了谈论机器人权利的四中不同话语表达方式。

机器人不能够拥有权利,也不应当拥有权利	机器人不能够拥有,但应当拥有权利
机器人能够拥有权利,但不应当拥有权利	机器人能够拥有权利,也应当拥有权利

第一种观点认为,机器人不能够拥有权利。根据“能够蕴含应当”的定律,机器人不应当拥有权利。因为,在机器人无法理解和享有权利的情况下,说它们不应当拥有权利似乎是顺理成章

的。这种观点完全把机器人当作工具来看待。第二种观点认为,机器人不能够拥有权利,但是,我们应当把它们当作拥有权利的主体来对待。第三种观点认为,即使机器人拥有理解和分享权利的能力,它们也不应当享有权利。第四种观点认为,机器人能够理解并看重自己的权利,因而机器人应当拥有权利。^[24]冈克尔认为,第一种观点虽然可以理解,但是缺乏前瞻性;第二种观点值得同情,但不具有可操作性;第三种观点是非理性的。他自己赞成第四种观点,并依据列维纳斯的他者哲学,证明了作为他者的机器人拥有权利的可能性。

四、构建后人类主义的伦理文化

由上可见,在当代环境伦理学与机器人伦理学的推动下,不仅道德承受体的范围得到了扩展,道德行为体的俱乐部似乎也增添了新的成员:具有道德思维能力的机器人。从道德行为体的角度看,我们的道德生活空间似乎出现了三类不同的道德行为体:(1)完全的道德行为体(正常而理性的自然人);(2)不完全(显性)的自然道德行为体(青少年);(3)不完全(显性)的人工道德行为体(社会机器人)。我们可以把第一类道德行为体称为完美的道德行为体(perfect moral agents),把后两类道德行为体称为不完美的道德行为体(imperfect moral agents)。完美的道德行为体享有完整的道德权利;不完美的道德行为体只享有部分道德权利。在道德行为体俱乐部中,完美的道德行为体享有家长的崇高地位与特殊责任,需要承担起对后两类道德行为体的监管责任与教化义务。从道德承受体的角度看,我们至少可以看到五种不同类型的道德承受体:(1)既是道德行为体又是道德承受体的自然人;(2)既是道德行为体又是道德承受体的机器人;(3)作为单纯道德承受体的人类成员;(4)作为单纯道德承受体的自然存在物(动物、植物与生态系统);(5)作为单纯道德承受体的机器人。这五类不同的道德承受体施加给道德行为体的义务强度是有差别的。

很显然,随着道德行为体与道德承受体的范

围的扩展,由多种道德行为体与道德承受体组成的伦理共同体变得更为复杂了,人们的道德生活也因此需要面对更多的挑战^[25]。但是,人类研发更为先进和完美的机器人的步伐是无法阻挡的。人为地阻止道德关怀范围的扩展也是不明智的。因此,人类(作为完美道德行为体)唯一明智的选择就是理性地面对科技进步给人类的道德生活所带来的全新挑战。应对这种挑战的方式之一,就是构建后人类主义(posthumanism)的伦理文化。就本文的论题而言,我们认为,这种新型的伦理文化至少应当包含四个重要元素。

第一,伦理拓展主义(ethical expansionism)。面对日益多元的道德主体,我们需要拓展人类的伦理思维空间。伦理思维空间包含两个维度,一个是作为学术研究的伦理学维度,一个是作为生活实践的道德关怀维度。在道德实践与学术研究方面,近现代文明都只把人与人的关系纳入了伦理思维的空间。但是,从历史角度看,我们会发现,在人类的神话思维中,人类、各种神祇甚至动物都是道德行为体;神与神之间、神与人类之间、神与动物之间、人类与动物之间都存在着某种伦理关系;神、人类与动物之间的交往也遵循着某些伦理原则。因此,在道德主体日益多样化的今天,我们需要重新激发人类的道德想象力,从学术与实践两个维度拓展伦理思维的空间,重新勘定道德生活的空间范围。

第二,非人类中心主义(nonanthropocentrism)。随着伦理思维空间的扩展,我们发现,人类不仅不再是唯一的道德承受体,也不是唯一的道德行为体。毫无疑问,人类仍将是这个扩展了的伦理大家庭的家长 and 监护人。这是他的独特价值所在。但是,人类肩负的道德责任也因此越来越沉重了。他不仅肩负着建构和维护人类内部的伦理秩序的重任,还肩负着建构并维护人类与自然之间、人类与机器人之间、机器人与机器人之间的伦理秩序的全新使命。在世俗的意义上,人类或许会活得“很累”,但在伦理和精神的意义上,他也将因此而变得更加伟大和崇高,成为一个真正生活在“天地境界”中的“天民”。

第三,后科学主义(post-scientism)。科技

的飞速发展把人类“抛入”了一个高度技术化的世界中。在这个高度技术化的世界中,人类的生活既是方便、舒适与高效的,同时又是充满风险的。人类的生活离不开科学理性,但是,我们也不能陷入科学主义的误区。科学理性能够帮助人类更好地实现自己的梦想,但是,它却不能解答“人类应当追求什么样的梦想”的问题。要解决人们之间的价值分歧甚至价值冲突,我们也只能诉诸价值理性,而不能指望科学理性。在对那些将深刻地影响到人类生活的重大科技决策进行选择时,我们需要在技术理性与公共理性之间实现某种平衡。总之,科学技术虽然是社会进步的最重要的助推器,但是,我们必须站在后科学主义的高度理性地把握文明进步的方向盘。

第四,全球主义(globalism)。技术的进步日益把地球缩小成“地球村”。全球气候变暖、全球金融安全、全球经济的相互依赖、全球反恐合作等全球问题使得人类日益成为一个命运共同体。为了有效地应对和解决这些全球问题,人类需要超越狭隘的民族主义立场,认可并践行世界主义的全球共享价值,建构一种超越民族国家体制的世界主义的全球制度。只有在这样一种世界主义的全球制度的保障下,各民族才能真正实现全球性的环保合作。要实现人与自然的和解,就必须首先实现人与人之间(尤其是民族国家之间)的和解。机器人技术的研发同样需要达成某种国际共识。一方面,我们应在联合国的协调下,防止各民族国家(尤其是对国际事务有主导权的国家)把机器人技术的研发演变成军备竞赛的新战场。这意味着,应当禁止研发作为杀人恶魔的杀人机器人(killing robots),应当把“不杀害任何自然人”作为一条“底线伦理”植入机器人的软件程序中。另一方面,对于作为显性道德行为体(不完美的道德行为)的机器人的设计,国际社会也必须达成某些共识;要把人类社会的价值共识当作这类机器人必须遵循的“底线伦理”。只有这样,未来的机器人才能真正成为人类道德事业的合作者,而非人类的终结者。

在全球化进程日益深化的今天,各国的经济交往与文化交流虽然日益频繁,但是,许多人在

价值观上却仍然深陷在狭隘民族主义与伦理相对主义的窠臼中。狭隘民族主义与伦理相对主义不仅是国家间悲剧的主要根源,更是构建人类命运共同体与全球主义文化的重大障碍。要实现从人类中心主义伦理文化向后人类主义伦理文化的转型,人类就必须首先要超越狭隘的民族主义,在价值观上实现从伦理相对主义向伦理普遍主义的转变,实现人类内部的和谐统一,建构价值论意义上的人类命运共同体,并把全球主义文化的培育作为构建后人类主义文化的重要抓手。

参考文献

- [1] Kenneth Goodpaster, On Being Morally Considerable, *Journal of Philosophy*, 1978, 75(6).
- [2] 〔美〕保罗·沃伦·泰勒. 尊重自然:一种环境伦理学理论[M]. 雷毅, 李小童, 高山, 译. 北京:首都师范大学出版社, 2010.
- [3] John Basl, Machines as Moral Patients We Shouldn't Care about (yet): the Interests and Welfare of Current Machine, *Philosophy and Technology*, 2014, (27).
- [4] 〔法〕阿尔贝特·史怀泽. 敬畏生命[M]. 陈泽环, 译. 上海:上海社会科学院出版社, 1992.
- [5] 〔美〕罗德里克·弗雷泽·纳什. 大自然的权利:环境伦理学史[M]. 杨通进, 译. 青岛:青岛出版社, 2005.
- [6] 杨通进. 环境伦理:全球话语 中国视野[M]. 重庆:重庆出版社, 2007.
- [7] 〔德〕康德. 对于动物的责任[M]//〔澳〕彼得·辛格, 〔美〕汤姆·雷根. 动物权利与人类义务. 曾建平, 代峰, 译. 北京:北京大学出版社, 2010.
- [8] Blay Whitby, Sometimes It's Hard to be a Robot: A Call for Action on the Ethics of Abusing Artificial Agents, *Interacting with Computers*, 2008, 20(3).
- [9] 甘绍平. 机器人怎么可能拥有权利[J]. 伦理学研究, 2017, (3).
- [10] 雷瑞鹏, 冯君妍. 机器人是道德行动者吗[J]. 道德与文明, 2019, (4).
- [11] Kenneth Himma, Artificial Agency, Consciousness, and the Criteria for Moral Agency: What Properties Must an Artificial Agent Have to be a Moral Agent?, *Ethics and Information*

Technology, 2009, (1).

- [12] David Levy, *Love and Sex with Robots*, Harper Perennial, 2008.
- [13] Kathleen Richardson, The Assymmetrical Relationship: Parallels between Prostitution and the Development of Sex Robots, *Acm Sigcas Computers and Society*, 2015, 45(3).
- [14] Sinziana Gutiu, The Robotization of Consent, in Ryan Calo R et al eds., *Robot law*, Edward Elgar Publishing, 2016.
- [15] Lily Frank, Sven Nyholm, Robot Sex and Consent: Is Consent to Sex between a Robot and a Human Conceivable, Possible, and Desirable, *Artificial Intelligence and Law*, 2017, 25(3).
- [16] Colin Allen, Gary Varner, Jason Zinser, Prolegomena to Any Future Artificial Moral Agent, *Journal of Experimental and Theoretical Artificial Intelligence*, 2000, 12(3).
- [17] Susan Leigh Anderson, Machine Metaethics, in Michael Anderson and Susan Leigh Anderson eds., *Machine Ethics*, Cambridge University Press, 2011.
- [18] 〔美〕温德尔·瓦拉赫, 科林·艾伦. 道德机器:如何让机器人明辨是非[M]. 王小红, 等, 译. 北京:北京大学出版社, 2017.
- [19] John P. Sullins, When Is a Robot a Moral Agent?, *International Review of Information Ethics*, 2006, 6.
- [20] David Davenport, Moral Mechanisms, *Philosophy & Technology*, 2014, 27(1).
- [21] James Moor, Four Kinds of Ethical Robots, *Philosophy Now*, 2009, 72.
- [22] Vincent Wiegel, Building Blocks for Artificial Moral Agents, 2006, <https://www.researchgate.net/publication/228615030>.
- [23] Mark Coeckelbergh, Robot Rights? Towards a Social-relational Justification of Moral Consideration, *Ethics and Information Technology*, 2010, (3).
- [24] David Gunkel, The Other Question: Can and Should Robots Have Rights?, *Ethics and Information Technology*, 2018, 20(2).
- [25] Erica L. Neely, Machines and the Moral Community, *Philosophy & Technology*, 2014, 27(1).

责任编辑:李建磊