

DOI: 10.16276/j.cnki.cn51-1670/g.2019.10.009

对相容论操控难题的分析

余日青

(厦门大学人文学院 福建厦门 361005)

摘要: 20世纪现代物理科学的发展,出现了决定论对自由意志和道德责任的挑战。相容论者认为,即使决定论为真,自由意志也可与其兼容。相容论主张,当一个行为真正出自行为者时,行为者对该行为负道德责任。然而,不相容论者提出了操控难题,即被外在操控的对象也可满足“自我性”条件,此时被操控的人或机器人是否有道德责任?这个难题不仅是相容论目前面临的重大挑战,而且对司法判定也有重要影响,因为存在类似的现实操控案例。因此,解决难题的关键就是要深入分析何为操控难题,阐明相关操控条件和因素,才能为相容论的辩护提供有力条件。

关键词: 决定论; 外在操控; 道德责任; 相容论; 自我性

中图分类号: B82-02 **文献标志码:** A **文章编号:** 1672-612X(2019)10-0053-06

操控难题是相容论理论目前有待克服的最重要的挑战。裴若步的操控案例不仅证明了相容论的“自我性”条件不是道德责任的充分条件,而且动摇了相容论理论的基本立场,即决定论和道德责任相容的观点。如果说运气难题(即一个偶然的的结果如何可以被称为是行为者自己的)是自由意志论在回答自我性问题上面临的最大挑战,那么相容论者在回答自我性问题上,也面临着他们自己的“运气难题”——一个被外在因素完全决定的行为如何可以被称为是行为者自己的?

一、操控难题的理论背景——相容论的道德责任观

彼特·施特劳森(Peter Strawson)认为,道德责任体现在社会人际交往或实践中。他认为,当我们说行为者对行为X有道德责任时,当且仅当该行为者因行为X而作为赞扬或责备等回应态度(Reactive attitude)的合适对象^{[1]6-10}。

(一) 道德责任和决定论相容

20世纪牛顿力学和拉普拉斯定理^[2]的出现,使许多科学家相继站在了决定论的立场。他们认为,

物理世界任何一个时刻的事态(至少是在宏观、分子及以上层面的事态)都已经为自然定律以及之前的事态所决定。如果外部世界和我们的思想真的是遵循确定的规律运行,就不存在真实敞开的其他可能,那么自由意志不存在,因此道德责任也就不存在。显然,决定论对自由意志和道德责任构成了挑战。而且决定论的挑战产生了相容性问题,即道德责任所要求的自由是否和决定论相容或在何种程度上相容?

针对这一问题的讨论,出现了相容论和不相容论的分流。相容论者认为,自由意志和决定论可以相容,而不相容论者反对这个观点。相容论者通过对“自由”进行假言式的解读,即当没有约束或障碍阻止我们做想做的事时,我们是自由的,这里强调有能力和机会采取其它行动。在这种意义上,决定论和自由意志相容,因此决定论和道德责任相容。

(二) 道德责任的必要条件——“自我性”条件

相容论者强调“自我性”是道德责任的必要条件。“自我性”是指“行为者是行为的源头或始因”。那么一个行为如何才是行为者自己的?源头派不相容论者认为,一个行为真正出自行为者自己的,当且

收稿日期: 2019-06-19

基金项目: 中央高校基本科研业务费专项资金资助科研项目“证言知识论”(20720181076)。

作者简介: 余日青(1994-),女,江西九江人,硕士研究生,研究方向:道德哲学、政治哲学。

仅当行为者是该行为的“终极原因”,即该行为在因果上不是由外在于行为者的因素所完全决定的^[3]。然而,相容论者反对自我性与因果决定论相矛盾的观点。法兰克福、沃特森、苏珊·沃尔夫从欲望和理性评价的契合来构建自我性;而史卓森、瓦莱士、费舍尔、马克拉维扎则从回应态度和机制角度来刻画自我性^[1]。

1. 法兰克福的欲望分层说。法兰克福将人的欲望分为两个层次:一阶欲望和二阶欲望。一阶欲望是指行为者对具体对象的欲望,二阶欲望是对一阶欲望的反思和评价。法兰克福认为,只有一阶欲望和二阶欲望相符,且二阶欲望没有和其他二阶欲望相冲突,或者与最高阶欲望相符时,我们的意愿才是“行为者自己的意愿”,即行为是他自己的行为^[3]。如一染上毒瘾的父亲法兰克福有戒毒的一阶欲望,法兰克福认同戒毒的欲望,他认为吸毒会导致他没能力照顾家人,所以,此时他的二阶欲望和一阶欲望都要求拒绝吸毒。因此,法兰克福戒毒的行为是自由的。然而,沃特森认为,法兰克福理论仅仅关注欲望或者动机维度,而没有足够重视理性评价的维度^[3]。沃特森区分人类心理的动机系统(Motivational System)和评价系统(Valuation System)。动机系统指的是引起行为者行动的心理因素;评价系统是指对行为产生判断的心理因素。沃特森认为,行为者根据理性原则和目标(Principles and Ends)来调整行为和欲望,当动机系统和评价系统相符合时,便构成了真实的自我^[3]。比如,法兰克福欲求读书,且法兰克福认同读书是件有价值的事情,那么法兰克福就是自由的。法兰克福的欲望分层理论关注的是一个人内部不同层面的欲望的一致性;而沃特森关注的是一个人的欲望和对欲望的价值判断的一致性。

2. 苏珊·沃尔夫的“真实自我”观点。苏珊·沃尔夫持“真实自我”(Real Self View)的观点,她认为,只有当行为出自“真实自我”时,我们才能认为一个人有道德责任。所谓“真实自我”这里指的是反思认同后的真实或深层的自我,一个行为出自真实自我,当且仅当行为者有能力根据评价系统自由的支配自己的行为^{[4]23-35}。苏珊·沃尔夫把她的价值理论称为“理性观点”(the Reason View),即要求人的价值信念建立在客观价值“真和善”(the True and the Good)的基础上^[5]。因为,有些时候,人的欲

望和意志是分离的,我们对某些欲望的掌控不依赖于我们的主观选择,一个人在考察自己应该具有什么欲望时,并不能通过主观感受而给出正确的判断。苏珊·沃尔夫强调道德责任判断的两个重要因素,即精神健全(Sanity)和规范能力(Normative Competence)。规范能力是指,鉴别自身的行为并将其与道德、法律规范相契合的能力^[3]。根据“理性观点,道德责任依赖于一个人具有正确的理由做正确的事的能力”^[5]。精神不健全者、犯罪狂热分子等没有能力出于正确的理由做正确的事情。人只有具备了客观的理性评价能力,才能基于正确的理由去做正确的事。显然,沃尔夫和沃特森都强调基于价值来解释自由和道德,都认同一个人对自己的行为负责,当且仅当这个人出于自身的价值信念行事。然而,二者的区别在于,沃特森强调的是主观的价值,沃尔夫强调的是客观普遍的价值。

3. 费舍尔和马克·拉维扎的认责型回应理论。费舍尔和马克·拉维扎的回应理论包括两个方面:适度理性回应机制和认责条件。理性回应论者主张道德责任所必须的控制力存在于依据不同的外在理由或刺激,做出与之相应的回应的能力中。比如瘾君子想要戒毒,实际做出决定的那个认知机制是行为者在非瘾症起作用时的理性机制。费舍尔和拉维扎认为这种控制力是一种“适度理性回应机制”(Moderate Reason-responsibility),该机制能够对某些外在的道德相关理由做出正确认知,并在此基础上形成行动的意图和动机。那么我们如何保证这个理性回应机制属于行为者自己呢?费舍尔和拉维扎进一步提出了认责条件。除了行为必须是经由行为者的“适度理性回应机制”所做之外,道德责任所需的控制力还必须满足一个历史性条件——行为者必须在某个时刻对引起行为的相关的“适度理性回应机制”认责,从而成为这个机制的所有者。认责是指行为产生于行为者自身的理性回应机制。认责(Take Responsibility)包括三个条件:其一,行为者知晓自己的选择和行为能够从因果上改变外在世界的状况;其二,行为者必须视自己为相应回应态度(比如赞扬、批评、感激和怨愤)的合适对象。这意味着他必须觉悟到他所在道德社群中有赞扬、批评、感激和怨愤这类回应性态度,而他可以因自己成为这类回应态度的对象。同时,他必须对该态度有适当的

主观上的反应。比如,如果他因为做出了值得批评的行为而受到批评,他会因此感到羞愧。其三,上述两种认知必须经由适当的方式,建立在行为者对周围的环境的观察和分析的基础上^[1]。

二、相容论面临的操控难题

裴若步设计了一组操控案例,论证上述相容论者提出的“自我性”条件作为道德责任的判定条件是不充分的。理由是:裴若步认为一个被外在操控的对象,也可以满足相容论的自我性条件,而我们通常的直觉是外在操控的对象是没有道德责任的,因此,道德归责仅有自我性条件,似乎是不充分的。裴若步的操控案例对相容论构成了挑战。下面我们来对比分析下列裴若步设计的一组经典案例。

(一) 裴若步的操控案例介绍

案例1:普拉姆是由神经科学家创造的。神经科学家可以通过无线技术直接操纵普拉姆,使其最大程度上和正常人一样。神经科学家通过直接产生或修改普拉姆的欲望,创造他的每一种大脑状态,实时控制着普拉姆的理性权衡活动,并通过一些按钮控制普拉姆使其保持自利的性格。最终普拉姆杀死乔治。

案例2:普拉姆看起来和正常人一样(除了是由神经科学家创造的)。神经科学家不直接控制普拉姆,而是通过间接的程序控制普拉姆的理性权衡,从而保证普拉姆通常都是自利的。最终普拉姆杀死乔治。

案例3:普拉姆是一个普通人,由于从小受到家庭和社区的严格训练,形成了自利的性格。最终普拉姆杀死乔治。

案例4:普拉姆是一个普通人。成长环境和经历影响着普拉姆的性格和思考方式,使其形成自利的性格。最终普拉姆杀死乔治。(DeekPereboom 2003, p115)

说明:在上述四个案例中,普拉姆没有受到不可抵抗的欲望的影响,没有任何行为上的限制,他的思考和行为都和他的性格相符合。普拉姆想杀死乔治的有效的一阶欲望和他的二阶欲望相符合;普拉姆的理性权衡来自适度理性回应机制;普拉姆能根据不同的外界刺激做出不同的反应和选择。普拉姆不是永远自利的。

首先,上述裴若步的案例的特点可以简要概括如表1所示:

表1 裴若步的操控案例特点

案例	普拉姆案例情境	特点
案例1	由科学家创造,神经科学家实时操控大脑活动,形成特定性格,产生杀乔治的意图和行动	科学家直接、实时控制
案例2	科学家创造,被植入特殊程序,形成特定性格,使其产生杀乔治意图及行动	科学家间接、预先控制
案例3	是普通人,从小一直受到家庭灌输或严格训练,形成特定性格,使其产生杀乔治意图及行动	受到家庭循序渐进的控制
案例4	是普通人,受成长环境和经历影响,形成特定性格,使其产生杀人意图及行动	后天环境的影响

(二) 裴若步操控案例对相容论“自我性”的挑战

首先,裴若步认为,无论行为者的行为是被自然因果决定(Determinism Universe)还是被外在操控决定,二者在“被外在因素所决定”这一事实上没有本质的区别。因此,案例1、案例2、案例3和案例4在“被外在因素决定”上没有本质的区别。其次,裴若步认为,依照直觉,案例1中的普拉姆的行为是被神经科学家这一外在因素操控所决定的,所以普拉姆没有道德责任。根据普及策略(Generalization Strategy)推理出案例4中普拉姆也没有道德责任。最后,裴若步的操控论证(Manipulation Argument)的结构

可以分析如下:

a.无责任直觉:直觉上,被外在操控的行为者没有道德责任,尽管行为满足所有相容论自我性条件;

b.无差别判断:外在操控的行为和自然因果决定的行为,在道德归责上没有本质的区别;

c.因此,自然因果决定的行为也没有道德责任。

裴若步通过上述4个案例得出结论:相容论的“自我性”条件不是道德责任的充分条件。与此同时,裴若步的操控案例还揭示出相容论理论内在不一致问题。一方面,相容论的基本立场是,行为者需要为一个被通常环境因素(Determinism Universe)所

决定的行为负道德责任;另一方面,相容论又承认行为者不需要为一个被外在操控所决定的行为负道德责任。不相容论者可能会问,一个有自我性的被操控的行为者是否有道德责任?裴若步认为,在回答这个问题之前,相容论者需先说明两个问题:第一,外在操控和因果决定在道德归责上的区别;第二,如果相容论者无法解释二者的区别,那么相容论者需要解释,为什么同样被因果决定,却会出现一个有责和一个无责的判断。裴若步认为,在道德归责上,被外在操控的行为与被通常的环境因素所决定的行为在道德归责上没有本质的区别。由于被外在操控的行为没有道德责任,所以被通常环境因素所决定的行为者也没有道德责任^[6]。显然,对于这个结论,相容论是难以接受的,这就构成了相容论有待解决的难题。

三、经典操控案例的对比分析

裴若步的经典操控案例首次对相容论提出了挑战。此外,随着操控话题的讨论愈演愈烈,相继出现了很多类似的更加精致的案例,下文列举四种有代表性的操控案例进行对比分析。

(一) 贝茨案例对比分析——被催眠的贝茨和被洗脑的贝茨

(被催眠的贝茨) 贝茨在一次诊所探访中,她的精神科医生对她实施了催眠,并将一种新的价值观和杀死乔治的欲望植入她的体内。催眠是非常巧妙的,贝茨仍然能够像她正常的自我一样理性权衡。贝茨运用她的理性能力,判断出这种欲望完全符合她的价值观,并全心全意地认同这些价值观。然后贝茨设计了一个计划,杀死了乔治。第二天,贝茨恢复了正常,她对昨天的事什么也不记得了。(Mathe-son 2014)

(被洗脑的贝茨) 贝茨睡着的时候,一组心理学家对贝茨植入了一种不同的价值观,并抹去了她原有的价值观,但是保留她的记忆。第二天,贝茨醒来时想杀死乔治。虽然她总是觉得乔治不讨人喜欢,但她对这种想法感到很惊讶。但经过反思,贝茨认为这种欲望完全符合她的价值观,她全心全意地支持这些价值观。然后贝茨制定了一个计划,毫无保留地遵循这个欲望杀死了乔治。(Mele 2013 p.169)

说明:贝茨是一个非常温和、善良的人,她的性格和价值体系没有任何故意伤害他人的欲望。她也在很大程度上对拥有这种性格和价值体系负有道德责任。

根据上述案例,可以总结出两个案例的共同点:(1) 贝茨原本心性纯良,且具有善(Good)的价值体系;(2) 贝茨被植入新的外来的特定的价值观和欲望;(3) 贝茨具有自我性,可以做出理性权衡判断和行为选择;(4) 恢复正常的贝茨认同并接受新的外来的价值观;(5) 贝茨将杀乔治的想法付诸了实施。两个案例依旧存在一些不同点:(1) 贝茨被洗脑后抹去了她原有的价值观,而催眠并没有刻意抹去贝茨原有的价值观;(2) 贝茨洗脑后的主体身份没有变化,而催眠中的贝茨和催眠后的贝茨主体身份发生了变化(贝茨醒来后不记得自己的行为)。

密里认为,贝茨全心全意地认同杀乔治的欲望,满足法兰克福的自我性条件,贝茨的行为是自由的,然而贝茨的行为是操控的结果,因此,贝茨没有道德责任,相容论的自我性条件是不充分的^[7]。然而,相容论者反对贝茨无责的直觉。马西森认为,案例中贝茨的主体身份存在混淆。洗脑或催眠前的贝茨是一个完全善良的贝茨,而操控违背了贝茨原有的善良的价值体系。被操控后的贝茨具备邪恶价值观,此时主体身份发生了变化,不再是原初的贝茨,而是充满邪恶欲望的怪物,尽管邪恶的贝茨完全接受了新的价值观,满足自我性条件^[8]。假设用Y代表被操控后的贝茨,用X表示原初的贝茨,用Z表示邪恶的观念,那么 $Y=X+Z$ 。显然,Y和X不能直接等价,因为Y是由X加入Z的元素后转化为Y。因此,被操控杀人的实施者是被科学家植入邪恶观念后的主体Y,归责的对象应该是Y。然而,被操控前的贝茨X,本身没有邪恶的杀人观念Z,Z是外来加入的因素。X与Y的主体身份不同,因此贝茨X没有道德责任。

(二) 密里的受精卵案例与裴若步的案例2对比分析

受精卵案例(The Zygote Case):戴安娜在玛丽的体内创造了受精卵Z,受精卵中被加入了特殊原子,该原子确保事件E在30年后发生。由于戴安娜事先就确知世界的状态是被因果决定的,戴安娜推测受精卵Z将会成长为一个精神健全的厄尼,他有自我控制的意识和能力,且能对自己的行为理性权衡。然而,30年后,Ernie在没有任何强迫或控制的情况下实施了行为E。(Alfred.R.Mele 2008 p.279)

回顾前文第三节的裴若步的案例2,将其与受精卵案例进行对比可以总结出共同点:(1) 均属于间接的控制手段;(2) 具备道德归责的自我性条件;(3) 最终实施了外在操控所指定的行为;(4) 杀乔治的欲望和行为E的欲望都不是不可抵抗的。此外,依旧能发

现两个案例间存在着微小的不同之处:(1)从产生方式上分析。普拉姆是由科学家直接创造的,而厄尼是由带有特殊程序的受精卵成长而成的。根据日常直觉,相比于前者厄尼更接近正常人的受精发育的成长过程,更加接近自然因果决定的外在环境。(2)直觉上,我们会认为厄尼的自我性更充分,而普拉姆的自我性在一定程度上被削弱了。

密里设计受精卵案例的目的在于使戴安娜创造的受精卵与正常受精卵的产生很难找出本质的区别,从而得出因果决定论排除道德责任的结论。密里对受精卵案例的论证结构进行了总结,如下所示:

a. 由于受精卵是戴安娜创造的,所以厄尼不是自由的行为者,厄尼对自己的行为没有道德责任;

b. 考虑到受精卵发育后的自由行为和道德责任,被戴安娜创造的厄尼的存在和因果决定论的正常人的受精卵的产生没有本质的区别;

c. 因此,因果决定论排除自由的行为和道德责任^[9]。

不相容论者认为,厄尼是被戴安娜故意创造的用来执行特定的行为或戴安娜的某个意图,如同机器人被工程师设计执行电脑的程序命令,所以厄尼没有自由和道德责任。而且受精卵被戴安娜创造和我们通常由自然因果创造的受精卵,在产生方式上很难说有本质的区别,因此,受精卵论证表明,自然因果决定论削弱行为者的自由和道德责任。

四、操控案例与现实案例的联系

相容论的操控难题,在理论层面上,给相容论构成了重大挑战。然而,在现实层面上,我们同样会遇到类似操控的社会案例,法院在处理某些案件时应该考量操控的因素进行判处,违法行为者是否存在外在操控,影响归责对象的指向与归责程度,进而影响刑罚审判。如果在理论层面,操控难题悬而未决,那么,在现实生活中,关于一些相关情景的道德责任的判断,也会产生争议。

在日常生活中,存在类似洗脑、催眠的情况,如邪教信徒、传销人员等,行为者被循序渐进地灌输特定的价值观念。类似案例中的贝茨被催眠或被洗脑而植入了不同的价值理念,认同新的价值观和特定的欲望,进而执行了特定欲望导向的活动。在密里的贝茨案例中,贝茨被外在操控,她的行为是外在操控决定的结果。同样,现实生活中的被传销的人或接受特定理念灌输的人,是否也可视为被操控,他们的行为是否也是外在操控决定的结果?根据不相容

论者的观点,如果被传销的人是被操控,那么他们的行为是被外在操控决定的结果,所以传销者对自己的行为没有道德责任。然而,根据日常直觉,我们很难接受这样的结论。

下面我们针对一个具体的现实案例——“女神教”案例来分析不相容论者的操控案例和社会现实案例的联系。

“女神教”是一个邪教组织,信徒遍布世界各地,其信徒宣扬“不相信法律,只信女神”,肆意破坏法律法规。某日,“女神教”信徒向路边的一位女士索要电话,遭到女士拒绝,信徒便将女士视为“恶魔”“邪灵”殴打致死。最终,法院判决涉案信徒死刑^[10]。

“女神教”是最具危害力的邪教组织之一,该组织通过成立严密的邪教机构,对信徒灌输邪恶的价值观达到精神控制。“女神”信徒和贝茨案例中的贝茨一样,被灌输了邪恶的价值观念,导致一系列恐怖暴力的行为,而且贝茨和信徒被洗脑后都认同了新的价值理念。显然,贝茨的行为是被外在科学家操控的结果,而信徒的恐怖行为是被邪教头目灌输邪恶价值观操控的结果。根据不相容论的观点,贝茨和信徒虽然都有自我性,但是他们的行为是受到外在操控所决定的,因此,没有道德责任,自我性不是归责的充分条件。毫无疑问,这个结论是很难被日常直觉认同的。然而,相容论目前还没有给出自我性如何容纳操控情境的最有力的解释。

如果相容论者不能合理地解释操控的(有自我性的)行为者有道德责任这个问题,那么不相容论者就会坚持认为,被操控的(有自我性的)行为者没有道德责任。那么,针对“女神教”案例,会有人倾向于认为,信徒杀害女士是邪教头目的责任。如果接受这样的观点,就要接受被恐怖分子胁迫杀人的行为者、被洗脑的传销者、被催眠的杀人者均没有道德责任。毫无疑问,这个观点是很难被日常直觉所容纳的。

五、总结

操控难题是相容论目前面临的最棘手的挑战。不相容论者以一系列操控案例来论证相容论“自我性”条件不是道德归责的充分条件,进而论证决定论和道德责任不相容。通过深入分析裴若步和密里的操控案例,揭示出相关的操控元素,进而联系到近年来重大的社会操控案例,揭示出操控难题的解决不仅在理论层面意义重大,而且对于现实社会的道德、法律的归责也有重要影响。因此,操控难题是一个

值得关注的、迫切的、不得不解决的难题。

参考文献:

- [1] TIMPE K. Free Will: Second Edition Sourcehood and Its Alternatives [M]. New York: Continuum International Publishing, 2012.
- [2] MELE A R. Moral Responsibility and History Revisited [J]. Ethical Theory & Moral Practice, 2009, 12(5): 463-475.
- [3] KANE R. A Contemporary Introduction to Free Will [M]. Oxford: Oxford University Press, 2005.
- [4] WOLF S. Freedom Within Reason [M]. Oxford: Oxford University Press, 1994.
- [5] 姚大志. 我们为什么对自己的行为负有道德责任: 相容论的解释及其问题 [J]. 江苏社会科学, 2016(6): 14-22.
- [6] PETTIT G. Living Without Free Will [J]. Faith & Philosophy Journal of the Society of Christian Philosophers, 2013, 19(3): 368-71.
- [7] MELE A R. Manipulation, Moral Responsibility, and Bullet Biting [J]. Journal of Ethics, 2013, 17(3): 167-84.
- [8] MATHESON B. Escaping Heaven [J]. International Journal for Philosophy of Religion, 2014, 75(3): 197-206.
- [9] MELE A R. Manipulation, Compatibilism and Moral Responsibility [J]. Journal of Ethics, 2008, 12(3/4): 263-86.
- [10] 王雨. 邪教的特点和产生原因: 山东“全能神教”的调查与分析 [J]. 西北民族研究, 2017(2): 135-41.

Analysis of the Manipulation Problem in Compatibilism

YU Riqing

(College of Humanities, Xiamen University, Xiamen, Fujian 361005)

Abstract: With the development of modern physical science in the 20th century, Determinism challenges free will and moral responsibility. Compatibilists argue that even if Determinism is true, free will can be compatible with it. The compatibilist holds that when a behavior is truly from the actor, the actor is morally responsible for the action. However, incompatibilists raise the manipulation problem, that is, objects manipulated externally can also satisfy the self conditions. In this way, is the person or robot being manipulated still morally responsible? This conundrum is not only a major challenge to compatibility theory, but also has important implications for judicial decisions, as there are similar real-life manipulation cases. Therefore, the key way to solve the problem is to deeply analyze what the manipulation problem is and clarify the relevant manipulation conditions and factors, so as to provide powerful conditions for the defense of the compatibility theory.

Keywords: Determinism, Compatibilism, moral responsibility, manipulation, self conditions

(责任编辑:冯芳)