

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

姚大志

内容提要 人们通常认为,只有人是自由的,他们对自己的行为才负有道德责任。但是,从科学的观点看,人不是自由的,因为人的行为服从于因果解释,而这种解释的性质是决定论的。这样便产生一个重大的道德理论问题:自由(以及道德责任)与决定论是不是相容的?一些道德哲学家主张两者是不相容的,而很多道德哲学家则主张“相容论”。相容论者面对的一个根本难题是:如果决定论是真的,那么人如何能够对自己的行为负有道德责任?大体上我们可以归纳出三种版本的回答,即反应理论、等级理论和价值理论。我们的分析和批评表明,这些版本的相容论都存在一些难以克服的困难。

关键词 道德 责任 自由 决定论

姚大志,吉林大学哲学基础理论研究中心暨哲学社会学院教授 130012

如何从道德的观点解释人的行为,我们通常持有两种直觉。一方面,基于现有的知识体系(其中特别是生物学、心理学、社会学和行为科学),我们能够给予人的行为以科学解释,而这种科学解释归根结底是一种因果解释。而且我们也相信,这种因果解释在性质上是决定论的。另一方面,基于实际的社会生活和道德生活,我们认为人是自由的,并且应该对自己所做的事情负有道德上的责任。几乎所有的道德哲学家都承认,道德责任依赖于自由。但是,这两种直觉之间存在紧张:如果决定论是真的,那么一个人的行为在因果上就是被决定的;如果一个人的行为是被决定的,那么我们如何能够要求他为自己的行为负责?因此,一些道德哲学家主张,决定论与自由是不相容的,从而与道德责任也是不相容的。然而很多道德哲学家认为,决定论以及它的支持基础(知识体系)是很难反驳的,所以更合理的主张是,自由以及道德责任与决定论是相容的。这些主流道德哲学家的观点被称为“相容论”(compatibilism)。这样,相容论者便面对一个根本的难题:如果决定论是真的,那么人如何能够对自己的行为负有道德责任?不同的相容论者对这个问题给予了不同的回答。我们可以把这些回答归纳为

本文为国家社科基金项目“制度正义的理念研究”成果(项目批准号为15BZX022)以及国家2011计划司法文明协同创新中心研究成果。

三种主要的观点,即反应理论、等级理论和价值理论。我们下面的分析和批评表明,这些版本的相容论都存在一些难以克服的困难。

一、反应理论

斯特劳森 1962 年发表了他的经典文章《自由与忿恨》。这篇文章不仅提出了关于道德责任的反应理论,而且它也是当代自由与道德责任争论的思想根源之一。斯特劳森试图为人的道德责任规定条件,即在什么条件下一个人对自己的行为是负有道德责任的。他主张,基于这些条件,我们能够有理由认为某人对自己的行为是负有道德责任的。这些条件的解释依赖于斯特劳森所说的“反应性态度”(reactive attitudes)。

所谓“反应性态度”,是指人们对其他人针对自己的态度和行为所做出的自然反应,而其他人针对自己的态度和行为可能是善意的,也可能是恶意的,还可能是中性的^[1]。比如说,我们对善意的自然反应通常是感激或称赞,对恶意的自然反应通常是忿恨或谴责。按照斯特劳森的观点,所谓道德责任就是我们相互接受这样的反应性态度,而道德上的赞扬或责备实质上就是这种态度的表达。这种理论是以群体实践的观点来理解和解释道德责任:在我们的社会生活和道德生活中,这样的反应性态度表达了我们的关切以及如何相互对待的要求。

问题在于,如果决定论是真的,那么我们还能够继续坚持这种反应性态度吗?决定论是真的,这意味着一个人的行为是由某些先在的原因决定的,这些先在的原因或者是外在的环境,或者是内在的基因,而它们都不是行为者本人能够控制的。如果导致某种行为的原因是不受行为者本人控制的,那么我们很难要求行为者对该行为负责。另外,按照斯特劳森的观点,一个人在道德上是负有责任的,是因为我们认为他在道德上是负有责任的。与“他是负有责任的”相比,“认为他是负有责任的”具有优先的地位,而这种优先性内在于反应性态度之中。但是,我们在“认为”的事情上可能出错。比如说,我因某个人做了某件错事而感到义愤,但我不知道实际上这件错事是另外一个人做的。如果基于我认为他做了这件错事而断定他对此是负有责任的并谴责他,这是没有道理的。

鉴于这样一些问题,以瓦莱斯为代表的一些道德哲学家在 20 世纪 90 年代阐发出更精致的反应理论。瓦莱斯赞同斯特劳森的这种观点,即一个人在道德上是负有责任的,这意味着我们认为他在道德上是负有责任的。但是,他对这种观点增加了一个规范性条件:一个人对某种行为在道德上是负有责任的,当且仅当我们认为他对这种行为在道德上是负有责任的,这是公平的^[2]。这个规范性条件包含两个方面。一个方面是认识的,在确认道德责任的时候,我们对相关事实的了解是正确的,没有发生错误。另一方面是评价的,关于一个人对所做事情所负有的责任,我们给予的评价是公平的和适当的。

瓦莱斯与斯特劳森另外一个不同的地方涉及到对反应性态度之性质的看法。对于斯特劳森,反应性态度本质上是情感的和心理的,它们内在于人类的本性之中。对于瓦莱斯,反应性态度基于实践的理由,而这些理由表达了我们所承诺的道德义务。这里的关键是道德义务,而道德责任与道德义务相关。那么什么是道德义务?瓦莱斯认为,我们有义务去做的事情是那些我们所把握的理由推动我们去做的事情,而这些理由表达了我们所信奉的道德原则^[3]。一个人对某种行为是否在道德上负有责

[1]Peter F. Strawson, "Freedom and Resentment", in *Free Will*, edited by Gary Watson, Oxford, UK: Oxford University Press, 1982, p.67.

[2][3]R. Jay Wallace, *Responsibility and the Moral Sentiments*, Cambridge, MA: Harvard University Press, 1994, p.91, p.131.

任,这取决于他是否违反了道德义务,而道德义务则是由表达了道德原则的理由规定的。这样,只有我们基于相关的理由认为某个人对某种行为在道德上是负有责任的,我们的评价才是公平的。

瓦莱斯反应理论的核心是一种公平的原则——“没有错误不应责备”(no blameworthiness without fault)^[1]。这个原则有两种含义。首先,当一个人没有违反道德义务(“没有错误”)的时候,如果我们责备他,那么这是不公平的。例如,我们误以为某个人做了某件事情而责备他,这就违反了公平的原则。其次,当一个人违反了道德义务但是他没有能力来理解道德义务并且也没有能力按照道德义务行事的时候,如果我们责备他,那么这是不公平的。比如说,因受到强迫、心理疾病、神智不健全或被操控,这样的人被免除了道德责任。

瓦莱斯的观点以及广义上的反应理论面临一些理论上的困难,而且也受到了很多批评。特别是这种版本的相容论会面临两种重要的反驳,一种反驳来自于相容论内部,另外一种来自于不相容论。

从相容论的观点看,虽然反应理论对道德责任给予了解释,但是在这种解释中自由不仅没有发挥任何作用,甚至也没有任何地位。反应理论的意图是为道德责任确认条件,而这些条件或者是文化的和心理的(如斯特劳森),或者是道德理由和道德义务的(如瓦莱斯)。只要行为者的行为符合这些条件,那么他们对自己的行为在道德上就是负有责任的。但是,人们通常认为(其中也包括相容论者),道德责任以自由为前提。没有自由,就没有道德责任。如果这样,那么相容论者需要论证的不仅是道德责任与决定论是相容的,而且也包括自由与决定论是相容的。

从不相容论的观点看,如果决定论是真的,那么我们认为一个人对其行为在道德上负有责任,这就是不公平的。决定论是真的,这意味着一个人的行为是被决定的。按照决定论的观点,在给定的外部环境条件和内部生理条件下,他只能做他所做的事情,他没有其他选择的可能性。如果一个人没有其他选择的可能性,那么我们就很难认为他对自己所做的事情是负有责任的。这样,相容论者就需要提供这样的论证:在决定论是真的情况下,我们认为一个人对其行为在道德上是负有责任的,并且这种判断也是公平的,即使他没有其他选择的可能性。

这意味着,相容论要想站得住脚,就必须回应这两个批评:一个是如何解释自由的问题,另外一个涉及其他选择的可能性。在这种意义上,法兰克福提出的等级理论正是对上述两种困难的回答。

二、等级理论

法兰克福1971年发表了一篇经典文章“意志自由与人的概念”。这篇文章提出,人与其他动物的区别既不在于人有理性,也不在于人是政治动物,而在于人有意志,并且对人的意志结构给予了分析和界定。这篇文章也对相容论的观点给予了坚定的辩护,并为此提出了等级理论。如果我们说这篇文章是迄今为止对自由与道德责任的当代辩论产生了最大影响的文献,恐怕也不为过。

就我们关心的问题来说,法兰克福等级理论的意义在于它回应了古典相容论(反应理论)所面对的难题。这些难题主要有两个:一个是古典相容论缺乏对自由的解释,特别是没有关于意志自由的解释;另外一个则是古典相容论只解释了为什么人要负有道德责任,而没有论证道德责任与决定论是相容的。在古典相容论的框架下,前者是“意志自由”的问题,后者是“其他选择的可能性”的问题。

我们首先讨论“意志自由”。在西方传统哲学中,意志是一种形而上学的实体。与此不同,法兰克福提出,意志是某种特殊的欲望。我们有很多欲望,但是只有一些能够实现出来;我们也可能有相互冲突的不同欲望,而只有其中一些是真正驱动我们行动的力量。特别是我们经常会遇到这样的现象,

[1]R. Jay Wallace, *Responsibility and the Moral Sentiments*, Cambridge, MA: Harvard University Press, 1994, p.135.

某个人决心做某件事情,结果却做了另外一件事情。因此,虽然法兰克福主张意志就是某种欲望,但是他强调意志与欲望的范围不是一样的,欲望的范围比意志更大。在这种意义上,他把意志看作是一种“有效的欲望”,一种始终驱动人们行动的力量^[1]。这种直接驱动人们行动的欲望或意志也被法兰克福称为一阶的欲望或意志。

法兰克福的相容论被称为“等级理论”,这是因为它主张,人不仅具有一阶欲望,而且还有二阶欲望。所谓二阶欲望是指,一个人不仅拥有某种欲望,而且他在反思中也想要拥有这种欲望^[2]。比如说,我想喝茶,便去烧一壶开水。在这里,“喝茶”是我的一阶欲望,同时作为驱动我去烧水的力量,它也是我的一阶意志。假设当我想喝茶的时候,也有其他饮料供我选择,如咖啡、果汁和水等等,经过考虑之后,我还是觉得自己想喝茶。这个时候的“喝茶”欲望就是我的二阶欲望,它表现为“我想我要喝茶”。在反思中,在考虑是喝茶还是喝其他饮料的时候,如果我不仅想喝茶,而且决定要喝茶,这时“我想我要喝茶”不仅是我的二阶欲望,而且也是我的二阶意志。“我想做某事”,这是我的一阶欲望;“我想做某事”成为驱动我去做某事的原因,这是我的一阶意志。“我想我要做某事”,这是我的二阶欲望;“我想我要做某事”驱动我把“我想做某事”变成行为时,这种二阶欲望同时也是我的二阶意志。

一阶欲望表达的是人的本能,二阶欲望是对一阶欲望的确认。因为人有各种各样的一阶欲望,而人在具体的时间和地点只能做一件满足欲望的事情,所以需要一阶意志来驱动满足某一种欲望的行动。因为人们所拥有的一阶欲望可能是相互矛盾或相互冲突的,所以需要二阶欲望来确认某一种欲望是自己的真正欲望,也需要二阶意志来驱动自己确认的某种欲望。法兰克福认为,拥有二阶意志,这是人的本质所在^[3]。当然,一个人也可能有各种不同的二阶欲望,而这些二阶欲望之间可能是相互冲突的,在这种情况下,他也可能具有三阶或更高阶的欲望。正是在这种意义上,法兰克福的理论被称为“等级理论”。

对于相容论者来说,问题的关键不在于人拥有意志,而在于人的意志是否是自由的,因为他们主张的是自由(以及道德责任)与决定论的相容性,而他们反对的是不相容论。那么自由意味着什么?自霍布斯以来,这样一种自由观念便广为流传:一个人是自由的,这是指他能够做他想做的事情。法兰克福反对这种传统的自由观念。在他看来,自由地做想做的事情,既不是自由的充分条件,也不是其必要条件。它不是充分条件,因为一个动物也能够自由地做它想做的事情(如奔跑),但是我们不认为动物具有意志自由;它也不是必要条件,因为即使一个人的行动自由被剥夺了(锁在一个房间内而不能自由行动),他仍然拥有意志自由。法兰克福认为,人的自由存在于他的一阶意志与二阶意志的一致之中^[4]。我们可以把两者的一致理解为二阶意志对一阶欲望和意志的认同。也就是说,一个人在认同自己的一阶欲望和意志时,他行使的是意志的自由。

法兰克福的等级理论有些复杂,起码涉及到四个因素(一阶的欲望和意志,以及二阶的欲望和意志),以及相互关系。让我们举一个毒瘾的例子来加以解释。一些人具有强烈的毒瘾,他的这种欲望如此强烈以致根本就没有考虑是否应该戒掉的问题。另外一些人也有毒瘾,在反思中认为它是有害的,但是他们的意志还不足以使他们戒掉它。还有一些人也受到毒品的诱惑,但是知道它是有害的,并且能够坚定地抵制它。第一种人只有一阶的欲望(或许还有意志),但是没有二阶的意志。第二种人是所谓“不情愿的有毒瘾者”,他们有二阶意志,但是他们的二阶意志与一阶的欲望和意志是不一致的,在这种意义上,他们没有意志自由。第三种人是正常的人,他们不仅具有二阶意志,而且他们的意

[1][2][3][4]Harry Frankfurt, “Freedom of the Will and the Concept of a Person”, in *Free Will*, edited by Gary Watson, Oxford: Oxford University Press, 1982, p.84, p. 86, p.86, pp. 90-91.

志是自由的。

我们再来看“其他选择的可能性问题”(alternate possibilities)。在当代关于自由与道德责任的讨论中，“其他选择的可能性”通常被看作是一个原则：只有当一个人能够做其他事情的时候，他才能对自己的行为负有道德责任。比如说，一个人经常酗酒，但是他本来也可以不酗酒的，因此，他要为自己的酗酒负责。如果一个人只能做他所做的事情，而没有其他选择的可能性，那么他就不应对自己的行为负责。对于很多道德哲学家，“其他选择的可能性”不仅对道德责任来说是一个原则，而且对自由也是一个原则。一个人是自由的，这意味着他有“其他选择的可能性”。如果他没有“其他选择的可能性”，那么他就不能是自由的。

问题在于，“其他选择的可能性”与决定论是不相容的。按照决定论，在外部社会环境和内在生理构造已定的条件下，一个人只能做他所做的事情，而不可能做其他的事情。也就是说，决定论意味着只会有一种行为(他所做的事情)，不会有其他的行为(其他选择的可能性)。如果“其他选择的可能性”与决定论是不相容的，并且决定论是真的，那么人就没有“其他选择的可能性”。按照通常的理解，如果人没有“其他选择的可能性”，那么他就不应对自己的行为负责。因为相容论者一般都接受“决定论是真的”，所以“其他选择的可能性”对相容论构成了一个挑战。

法兰克福认为，“其他选择的可能性”原则是虚假的。为了证明这个原则是虚假的，他在1969年发表了一篇著名文章“其他选择的可能性与道德责任”。法兰克福在该文中提出了一个假设的例子来表明，一个人即使没有其他选择的可能性，他也应该对自己的行为负责。后来，其他道德哲学家也提出了各种与其相似但更为精致的例子来证明自己的观点。这类例子如此之多，但又与法兰克福提出的例子相似，以致所有这类例子现在都被称为“法兰克福式案例”。

法兰克福提出的例子是这样的^[1]。一个叫琼斯的人打算做某件事情(去法国旅游)，但是也存在做其他事情的可能性(去夏威夷旅游)。这时有一个叫布莱克的人在暗中监视琼斯，而这个人具有操纵琼斯的能力(比如说他已经在琼斯的大脑中安装了一个芯片)。他的意愿是要琼斯去法国，并且为此密切监视琼斯。如果琼斯决定去法国，布莱克就会按兵不动，不做任何事情。如果琼斯决定去夏威夷，那么他就会出手干预，通过琼斯大脑中的芯片来操纵他改变主意。这个暗中操纵者的存在决定了琼斯只能去法国，排除了他去夏威夷的可能性。

现在我们假设，琼斯最终决定去法国，而布莱克则按兵不动，没有进行任何干预。在法兰克福看来，琼斯是自己决定去法国的，他应该为此承担道德责任。但是，由于操纵者布莱克的存在，琼斯没有“其他选择的可能性”，即没有去夏威夷的可能性。在去哪里的问题上，琼斯没有选择，只能按照布莱克的意愿行事，无论其行为是由自己决定的，还是在别人的操纵下决定的。但是，如果琼斯最终决定去法国，那么他就应该为这个行为负责，即使他没有做其他事情的可能性。换言之，一个人是否对自己的行为负有道德责任，这与“其他选择的可能性”是无关的。

对于法兰克福，自由和道德责任与决定论都是相容的^[2]。自由与决定论是相容的，因为即使人的行为在因果上是被决定的，人仍然拥有自由。所谓自由无非是人的一阶欲望与二阶意志的一致性，即他认可自己的欲望并按照自己的欲望行事，而这与决定论并不冲突。道德责任与决定论也是相容的，因为人即使没有“其他选择的可能性”，他仍然要为自己的行为负责。只要一个人所做的事情是出于

[1]Harry Frankfurt, "Alternate Possibilities and Moral Responsibility", in *Moral Responsibility and Alternative Possibilities*, edited by David Widerker and Michael McKenna, Burlington, Vermont: Ashgate Publishing Company, 2003, pp. 21-22.

[2]Harry Frankfurt, "Freedom of the Will and the Concept of a Person", in *Free Will*, edited by Gary Watson, Oxford: Oxford University Press, 1982, p.95.

自己的意志,那么他就应该为此负有道德责任,无论他是否有可能做其他的事情。

法兰克福提出的等级理论以及法兰克福式案例引起了大量的讨论和争论,产生了广泛和重要的影响。但是,他的理论也面临一些严重的困难。

首先,法兰克福在其等级理论中把人的欲望分为一阶的和二阶的,他的目的是想表明,二阶欲望是人的真正欲望,而一阶欲望则有可能不是。问题在于,所谓二阶欲望无非是人在反思中确认的欲望,而反思中确认的欲望并不能保证它必然是自我的真正欲望^[1]。在这个方面,一阶欲望与二阶欲望之间没有法兰克福所想象的那种重大差别。比如说,一个吸毒者有吸毒的一阶欲望,他也可能有吸毒的二阶欲望,即在反思中他仍然认可自己吸毒。也就是说,没有什么东西能够确保二阶欲望属于真正的自我。

其次,法兰克福的等级理论主张,一个人只要认同自己的欲望并控制自己的行为,他就应该为自己的行为负责。但是,这种主张会产生某些反直觉的结果。比如说,某个人认为自己不应该吃含有大量食糖和脂肪的蛋糕,并且也在反思中认同自己这种不吃蛋糕的欲望。然而出于意志薄弱,他每次在面对美味的蛋糕时都禁不住大快朵颐。按照法兰克福的理论,既然他不认同吃蛋糕的欲望并且不能控制自己的行为,那么他就无需为此负责。但是我们的直觉告诉我们,这个人应该为吃蛋糕负责。

最后,法兰克福认为,他提出的案例能够证明,一个人即使没有其他选择的可能性,他仍然要对自己的行为负责。对此,道德哲学家们之间存在大量的争议。一些道德哲学赞同法兰克福的观点,但是更多的道德哲学家则反对他的观点。即使是赞同法兰克福主张的人,也认为他的例子过于粗糙。关于法兰克福案例产生了大量的讨论和争论,以致于它几乎变成了当代道德哲学争论的一个单独主题。

三、价值理论

等级理论是一种内在的理论,它关注的东西是一个人内部不同层面之间的欲望一致性,比如说一阶欲望与二阶欲望的一致性。这种理论主张,如果一个人的二阶欲望与一阶欲望是一致的,那么这种欲望就是他自己的真正欲望。从道德的观点看,这种等级理论存在一个重要缺欠,即它没有考虑价值问题。人们追求实现的欲望不仅应该是自己真正的欲望,而且也应该是值得追求的欲望。这样就会出现一个问题:一个真正属于自我的欲望有可能不是一个值得追求的欲望,譬如一个无可救药的瘾君子的欲望。正是出于这种考虑,一些相容论者提出了价值理论,其影响较大的代表有沃森和沃尔夫。

沃森的理论思路基于柏拉图的灵魂三分说。柏拉图主张人的灵魂由三个要素构成,即欲望、理性和精神。与这里讨论相关的主要是欲望和理性。欲望是人的本能,它表达了人直接想要的东西。但是,人直接想要的东西并不一定是好的或有价值的。什么东西是好的或有价值的,这要由理性来判断。重要的地方在于,这种价值判断为人的行为提供了理由:如果这些人们想要的东西是好的或有价值的,那么人们就应该追求它们。

基于这种欲望与理性的区分,沃森提出了他的价值理论。我们欲望的东西(如健康)可以是我们的重要价值,同样,我们赋予重要价值的东西(如职业成就)也可以是我们欲望的东西。在这种情况下,一个人的行为就是自由的。相反,如果一个人所欲望的东西(如吸烟)不是他赋予重要价值的东西,或者一个人赋予重要价值的东西(如读书)不是他所欲望的东西,那么他的行动就是不自由的^[2]。按照沃森的观点,人作为自由的主体由两套系统构成,一套是动机系统,另外一套是评价系统^[3]。动机

[1][2][3] Gary Watson, "Free Agency", in *Free Will*, edited by Gary Watson, Oxford: Oxford University Press, 1982, p.108, p.100, pp.105-106.

系统的功能是引发行为,它作为本能直接驱动我们去做某件事情。与此不同,评价系统的功能是产生判断:在给定的环境下,所有的情况都被考虑后,我应该做的是这件事情。这种评价系统是自由之源。当我们把“自由”归之于某个人时,我们假定,他是一个做出了这种价值判断的人。

沃森的这套理论有三个要点。第一,人的评价系统应该支配动机系统,也就是说,理性应该支配欲望。这一点明显带有柏拉图主义的色彩。第二,人是不自由的,这产生于他的动机系统与评价系统的不一致,产生于他所欲望的东西与他赋予价值的东西之间的不一致。比如说,一个不自愿的瘾君子是不自由的,因为他所欲望的东西(吸毒)不是他赋予重要价值的东西。第三,一个自由的人有能力把他的价值转化为行动,也就是说,他的行动产生于他的评价系统。因此,对于一个自由的人,他所欲望的东西与他赋予重要价值东西是一致的,他的动机系统与评价系统是一致的。

即使对于相容论者,甚至是从价值的观点看,沃森的价值理论也存在一些明显的问题。首先,按照沃森的理论,与欲望和动机系统相比,理性和评价系统代表了真正的自我。但正如佛洛伊德的心理分析所表明的那样,起码对某些人,他们的欲望和动机系统所表现的是更真正的自我。其次,沃森主张,如果一个人的欲望与价值是一致的,那么他就是自由的,从而应该对其行为负责。问题在于,一个人在评价自己的欲望这件事情上不能被看作权威:这个人的判断有可能出错,他可能被不明的因素所影响(如酒精或毒品)。也就是说,一个人有可能在什么是自己的真正欲望问题上出现判断错误。最后,对于沃森,如果一个人的行为源自其真正的自我,那么他对自己的行为就是负有责任的。但问题在于,这个人对于自己的真正自我,对于自己目前具有的评价系统,他可能是没有责任的。也就是说,在一个人的理性或评价系统的形成过程中,外部环境可能发挥了决定性的作用。

鉴于沃森的理论存在着上述问题,沃尔夫提出了自己的价值理论。沃尔夫把自己的价值理论称为“理性观点”(the Reason View),而它包含有两个特别的观点:首先,她主张,自由意味着一个人能够出于正当的理由去做正确的事情;其次,关于道德责任,她的理论对正确行为与错误行为的解释是不对称的。

从沃尔夫的观点看,沃森基于价值来解释自由和道德责任是正确的,但是问题在于,在沃森的理论中价值是主观的。沃尔夫主张,只有当一个人有能力在其价值的基础上行事的时候,他对自己的行为才是负有责任的;而只有当一个人在客观的“真和善”的基础上形成其价值的时候,他的价值才是客观的^[1]。一个人出于自己的价值信念行事,他才对自己的行为负责,在这点上,沃尔夫与沃森是一致的。两者的差别在于,沃尔夫要求人们的价值信念建立在客观价值——大写的“真和善”(“the True and the Good”)——的基础上。

人们通常认为,自由是道德责任的前提条件:一个人只有是自由的,他才能对其行为负有道德责任。但是,自由意味着什么?在沃尔夫看来,自由是人们认识善、理解善、追求善的能力,是一种出于正当理由去做正确事情的能力^[2]。那么什么样的理由是正当的以及什么样的事情是正确的?符合客观的“真和善”的理由和事情是正当的或正确的。在涉及道德责任的问题上,自由的意思肯定不是任意行事。因此,一些道德哲学家主张自由的行为应该建立在价值的基础上,而沃尔夫更进一步,要求行为建立在客观价值的基础上。一个人只有具备评价客观的“真和善”的规范能力,他才能出于正当的理由去做正确的事情。对于沃尔夫,自由和道德责任都基于客观的价值。

按照沃尔夫的“理性观点”,道德责任依赖于一个人出于正当理由做正确事情的能力。假如决定论是真的,这意味着一个人在心理上被决定去做某件事情。如果一个人在心理上被决定出于正当理

[1][2]Susan Wolf, *Freedom within Reason*, New York: Oxford University Press, 1990, p.75, p.77.

由去做正确事情,那么这意味着他有这种必要的能力,并且与决定论是相容的。但是,如果一个人在心理上被决定出于某种理由去做错误的事情,那么这意味着他没有这种必要的能力。一个人不得不做错误的事情,或者没有能力去做正确的事情,这意味着他缺少按照“真和善”行事的能力。基于以上观点,沃尔夫提出了一个自己也认为“古怪的”主张:如果一个人在心理上被决定出于正当的理由做了正确的事情,那么这是值得赞扬的,而且行为者对此也是负有道德责任的;如果一个人在心理上被决定做了错误的事情,那么他对此是没有道德责任的,并且也不应受到责备^[1]。

关于道德责任,这种解释在正确事情与错误事情上是不对称的。它们之所以是不对称的,在于一种假定,即决定论是真的。如果决定论是真的,那么人们都是必定做某种事情的,无论这些事情是正确的还是错误的。在这种情况下,人们没有其他选择的可能性。如果人们没有其他选择的可能性,那么人们即使做了错误的事情,他们也不应该受到责备,并且也对此不应负有个人的责任。与此不同,即使人们没有其他选择的可能性,如果他们做了正确的事情,那么他们也应该受到赞扬,并且对此也应负有个人的责任,因为他们是出于正当理由去做正确事情的。

沃尔夫价值理论的这两个特别观点都存在明显的问题。首先,沃尔夫主张,自由和道德责任依赖于一个人能够出于正当的理由去做正确的事情。这意味着,如果一个人不能出于正当的理由去做正确的事情,那么他对自己的行为就是没有责任的。这种观点明显违反了我们的道德直觉。比如说,某个人在首饰店趁人不备偷了一件珍贵的首饰。他知道偷首饰是不对的,但是利益的考虑压倒了道德的考虑,使他铤而走险。在这个场合,他不能出于正当的理由做正确的事情,但是,我们都会认为他对此是负有责任的。如果这样,那么沃尔夫的理论为什么会这样明显的错误?相容论的价值理论和等级理论通常也被称为“契合理论”(mesh theories)。等级理论要求一阶欲望与高阶欲望之间的契合,价值理论要求行为与价值信念的契合。对于价值理论,这意味着“正当的理由”与“正确的事情”之间的契合,只有两者的一致,人对其行为才是负有责任的。问题在于,沃尔夫没有看到,还存在另外一种契合,即“错误信念”与“错误行为”之间的一致,在这种情况下,人对于其行为也是负有责任的。

其次,沃尔夫提出,在没有其他选择的可能性的条件下,道德解释在正确事情与错误事情上是不对称的。道德解释是不对称的,这是因为做了正确的事情是基于正当的理由,所以它是值得赞扬的,而在做了错误事情的场合,由于不存在其他选择的可能性,所以它是不应得到责备的。这种不对称性面临强烈的质疑。一方面,如果决定论是真的,那么人们对自己所做的任何事情都是没有责任的,无论它是正确的事情还是错误的事情,就像决定论者所主张的那样。另一方面,如果一个人做了正确的事情是值得赞扬的,那么他做了错误的事情也是值得责备的,即使在没有其他选择的可能性的条件下。比如说,我们可以设想一个法兰克福式案例:某个人在首饰店里思考是否趁人不备偷走一件珍贵的首饰,这时有一个暗中的操纵者在监视他;如果这个人决定偷走首饰,那么他就按兵不动;如果这个人决定不偷,那么他就会出手干预(启动早已安装在该人大脑中的芯片);最终这个人决定偷走首饰,而操纵者没有做任何事情。在这里例子中,虽然这个盗窃者没有其他选择的可能性,但是我们仍然会认为他对偷盗是负有责任的。

四、简短的结论

综上所述,我们对上述三种版本的相容论可以给出一个总的分析和批评。

首先,相容论主张人们应该对自己的行为负有道德责任,而这种主张能够成立,就需要以另外一

[1] Susan Wolf, *Freedom within Reason*, New York: Oxford University Press, 1990, p.79.

个主张为前提,即人是自由的。人是自由的,这包含两方面的含义:一方面,一个人做什么,这取决于他自己的选择;另一方面,一个人行为的最终根源在他自己本身。但是,在所有三种版本的相容论中,自由的观念都是太弱了:在等级理论中,自由意味着一阶意志与二阶意志的一致;在价值理论中,自由意味着行为与理由的一致;而在反应理论中,自由干脆被忽视了。几乎所有的道德哲学家都承认(无论是相容论者还是不相容论者),道德责任依赖于人的自由。由于所有三种版本相容论的自由观念都太弱了,所以无法支持这种主张,即人们应该对自己的行为负有道德责任。

其次,在缺乏强有力的自由观念条件下,一个人做什么,这或者是决定论的,或者是非决定论的。所谓决定论的,这意味着行为是由行为者所不能控制的外部环境或生理构造决定的;所谓非决定论的,这意味着行为是或然的,即一个人做什么,这纯属一件运气的事情。如果是前者,即人的行为是由他不能控制的外部环境或生理构造决定的,那么相容论要求他为自己的行为负责就是不正义的;如果是后者,即一个人做什么是一件运气的事情,那么相容论要求他为自己的行为负责就是不公平的。无论是哪一种情况,相容论都不能要求人们对自己的行为负道德责任。

最后,相容论承认决定论是真的,并且在这个前提下主张自由以及道德责任与决定论是相容的。但是,相容论无法证明两者是相容的。按照决定论,一个人的选择或行为是由先前的事件决定的,这个事件的因果锁链可以一直追溯到他的出生之前,而一个人出生之前的事情不是他能够控制的。这意味着,相容论者所说的选择或行为:或者服从这种因果关系的后退追溯,从而与决定论是相容的,但是它们在这种意义上根本就不是自由的,以致行为者不能为其行为负有道德责任;或者不服从这种后退的追溯,其因果锁链能够在某个地方停下来,即其选择或行为的最终“起源”在行为者本身,但是这样的行为与决定论又是不相容的。在这个问题上,相容论者很难摆脱这种两难的处境。

[责任编辑:曾逸文]

Why Should We Bear Moral Responsibility for Our Own Behavior? — A Compatibilist Explanation and the Problems with It

Yao Dazhi

Abstract: It is usually believed that only with freedom do humans bear moral responsibility for their own behavior. However, from a scientific perspective, humans are not free because human behavior submits to a cause-and-effect explanation, which is by nature determinist, thus generating a significant question on moral theory: Is freedom (as well as moral responsibility) compatible with determinism? Some moral philosophers maintain that they are incompatible whereas many argue for compatibilism. Thus, compatibilists are confronted with a question fundamentally hard to answer: If determinism is true, how can humans have moral responsibility for their own behavior? Generally speaking, we can sum up three versions of answers, namely response theory, hierarchy theory, and value theory. The present critical analysis suggests that there exist insurmountable difficulties in compatibilism in these versions.

Keywords: moral; responsibility; freedom; determinism