

道德图林机对康德绝对律令的重述

颜青山

(华东师范大学哲学系, 上海, 200241)

摘要: 将康德的理性存在看作一台道德图林机, 依据其构建原则所克隆出来的道德图林机之间, 将服从一系列“道德原则”。从逻辑上可以严密地证明, 这些原则包含了康德绝对律令的三种表达形式。即: 不存在定言指令只适应于特定康德机、而不适应于其他康德机的情形; 不存在仅仅作为工具机的道德图林机; 不可能存在定言指令集不同的道德图林机。

关键词: 道德图林机; 克隆; 绝对律令; 康德; 道德哲学

中图分类号: B516.31

文献标识码: A

文章编号: 1672-3104(2013)06-0001-05

自罗尔斯《正义论》产生巨大影响之后, 伴随而来的是康德道德哲学在当代的复兴。但在分析哲学传统中的这种复兴决不是一种诠释性的哲学史学研究方式, 而是在对其基本精神的各自理解中的多方向发展。其中一个诱人的方向是, 如何在康德的道义论基础上兼容后果论或价值论的伦理学——在这个过程中, 出现了一种“价值道义论”(value-based deontology)的新规范伦理学形态, 它将康德绝对律令的人本公式看作一个基于人的价值的普遍公式^[1]。

笔者追随这一思想, 希望在一种被重新理解了“内在价值”概念的基础上作出这样的融合工作, 但在此之前, 我们尝试将康德的“理性存在”还原为计算机科学和心灵哲学中曾经广泛讨论的“图林机”来达到这一目的——通用图林机的工作原理早已由计算机科学的先驱图林构造出来^[2]。图林机作为一种方法论模型, 可以消除论证中的模糊性。

在这个意图中, 笔者已经预先解决了两个问题: 理性存在与图林机的一致性, 道德图林机的构建原则问题。第一个问题的核心是克服我们所发现的“康德-图林悖谬”: 图林测试中似乎要求图林机说谎以通过测试, 而说谎者不可能是康德意义上的理性存在。我们的论证表明, 一台道德图林机不可能说谎, 即一台图林机不可能说出计算机程序中不存在或不能产生的语句^[3]。第二个问题是解决道德图林机的建构原则。一台通用图林机的程序由许多指令构成, 其指令全部为假言性的, 即是“如果…那么…”的语句结构。从通用图林机到道德图林机要求消除指令的假言

性, 这种逻辑操作将最终得到一条构造道德图林机的原则: “如此行动, 当且仅当没有指令冲突时。”^[4]

本文将继续沿着上述论证考察如何从克隆道德图林机中推导出康德绝对律令的三种表达方式。

一、随机性与自由意志

在那篇被称为构造了关于机器智能之“图林测试”的论文中, 图林提到随机性与自由意志的关系: “以数字计算机思想为基础, 派生出了一种……‘带有随机元件的数字计算机’。它们包含涉及掷骰子或某种等价电子过程的指令……有时将这种机器描述为具有自由意志。”^[5]

事实上, 以随机性来诠释“自由意志”的做法一直存在着, 如薛定谔关于电子“自由意志”的论述^[6], 最近关于自由意志与因果性关系的诸多量子物理学的思想实验^[7,8]。当然, 我们也注意到, 图林本人已经表明, 他是不会在这个意义使用“自由意志”这个概念的。图林的谨慎在我们看来是有道理的, 因为随机性并不意味着意志自由。

必须区别两种自由概念: 独立(或自主)意义上的自由和选择意义上的自由。所谓独立意义的自由, 就是不受外界影响的自由, 即自我决定的自由; 而所谓选择的自由则是指对多种可能性的选择, 其基础是可能性——如果没有可能性, 就没有选择的自由。

独立意义上的自由乃是自我立法, 根据理性行动,

收稿日期: 2013-10-12; 修回日期: 2013-11-22

基金项目: 国家哲学社会科学基金项目“价值道义论: 在理由与行动之间”(09ZXB057)

作者简介: 颜青山(1967-), 男, 湖南衡南人, 哲学博士, 华东师范大学哲学系教授, 主要研究方向: 价值道义论, 分析现象学, 生物学哲学。

这种自由即是康德意义上的自由。依据这样的自由意志行动时,其结果并不具有“可能性”,例如,在“说真话”和“说谎话”之间,我们并没有选择“说谎话”的自由,即“说谎话”的可能性并不存在于独立意义的“自由”中。

同样地,选择意义的自由也不意味着独立意义上的自由。在上述例子中,选择意义上的自由意味着自由意志的主体可以选择“说真话”,也可以选择“说谎话”,这样,主体的行动结果中就既有“说真话”,也有“说谎话”的可能性。但是根据独立意义的自由观,“说谎话”必定是受到了他律因素影响的行动,是不自由的,一个自律的意志是不会说谎的。

随机性是一个数学概念,是指事件发生的可能性。一个随机事件通常也被称为“自由事件”或“独立事件”,是指一个事件在不受到外界影响的情况下,其行为在概率空间中表现出的遍历性特征。假定事件A和B以等概率发生,那么,其逻辑表达是“或者A,或者B”,这个逻辑表达式等价于不考虑时间时的表达“可能A并且可能B”。从这种表述看,随机性自由表面上看起来似乎既包含了独立意义的自由,也包含了选择意义的自由。然而,当这两种意义运用于伦理学时则会发生冲突。

为什么会出现这样的矛盾呢?其实,在随机性事件中,“独立”是事件发生的条件,而“可能性”是事件表现的结果。在这里,条件与结果的关系并不是逻辑上的必然关系,它们只是一种经验关系,尽管这种经验关系是数学的,而不是物理学的——统计学及作为其数学基础的概率论总是带有经验的色彩,这种经验色彩由其全同性假设引起。

在伦理学中,独立意义的自由还有一个条件,即依据理性,这个条件在康德那里,是与自由是同一的。这样,一个不受外界力量影响的理性存在,其行为就不可能是随机的。而在选择意义的自由则假定,可能性是自由的条件,选择行为本身则受到内在动机的影响,不同的主体可能有不同的动机,因而就会表现出不一样的(可能)结果来。

根据独立意义的自由观,“人具有意志”是先天的,一旦这个意志不受到外界经验和对象的影响,它就是自由的。而这种先天意志在选择意义的自由观那里,就很可能是不自由的:只要我们拥有意志,我们就不能不行使我们的意志,即使我们放弃意志也不过是一种意志的表现,这一点我们没有选择的自由。意志的这个先天性特征犹如笛卡儿的“我思”。

其实,如果选择意义的自由是对可能性的选择的话,那么“选择”这种意志就必须先天于可能性,没

有“选择”,就谈不上意志。如果自由是纯粹的可能性或随机性,那么要贯彻随机意义的自由,就相当于我们能够设计这样一个装置,一会儿它有意志,一会儿它没有意志;可是,当我们探讨没有意志的情形时,我们并不是在探讨伦理学或道德哲学问题。而且,我们也不可能设计出这样的装置,因为任何装置的行为都必须是依据某种指令而产生的结果,我们不能设计出没有指令的装置。因此,随机性与意志自由没有关系。

以上论述表明,如果坚持选择意义的自由观,我们就有可能否认意志,所以,我们就必须坚持独立意义的自由观,这种自由观与意志是相容的。

既然我们坚持独立意义的自由观,那么,为什么我们还赞同在计算机中引入“随机元件”呢?这是因为独立意义的自由观适合于意志的存在性,而选择意义的自由观适合于意志的内容。在这个意义上,独立意义的自由和选择意义上的自由是可以很好地相容的。

在独立的意义上,一个理性存在只有一个自由意志,可以表述为“我意愿……”这样的单数形式;而在选择的意义上,一个理性存在可以具有多个意志,可以表述为“我意愿A”“我意愿B”……的复数形式。为了表示区别,我们可以将前者说成为“意志”,而后者说成为“动机”;最好将前者说成为“意志性”(willness, volition),而后者说成“意志内容”。

在这种相容关系中,“随机性”只是说明意志内容的自由性,而不说明意志本身即意志性的自由。意志性的自由只产生态度谓词“我意愿……”,而不是产生一阶谓词,一阶谓词由随机性的自由产生出来。一个指令或计算机“意志”逻辑上是由一个表示意向性或意志性的态度谓词和一个一阶谓词构成,如“我应该(意愿)说真话”。一般地,对于一个意志或指令来说,态度谓词(“应该”或“意愿”)是一个常项,一阶谓词(“说真话”)是变项,我们只需要设计一个足够大的一阶谓词库,就可以达到这种自由组合。

指令或计算机意志以上述方式产生,我们就必须回答这样的问题,由随机性“产生”的意志是自由意志吗?

为了回答这个问题,我们不得不提到人们对“目的/手段”与“结果/条件”句型的混淆。随机产生方式不过是自由意志内容的实现方式,是根据自由意志对内容多样性的要求设计出来的,不过是一种手段,而不是约束性的条件。这个表达的逻辑形式是“如果要获得自由意志的内容,那么采取随机方式或外界输入方式”;而不是“如果随机化,那么得到自由意志

的内容”。这里的目的是“获得自由意志”，“随机化”只是手段而已。

我们这里无非是用随机方式来产生足够多的意志内容；或者在表现上表明，意志的产生可以不受外界影响，是“自律的”。计算机执行任何一个指令或意志都是“自由”的——因为意志性与其内容的产生方式没有必然联系，并且不受其影响，所以是自由的。

二、道德图林机的克隆及其条件

仅仅依据单一康德机(即道德图林机)中不能容受冲突的指令的先天原则，我们还不清楚应该保留冲突指令中的哪一个：是随机地保留其中一个，还是必须只选择其中特定的一个？

解决这个问题将涉及道德图林机之间的关系。我们将通过克隆康德机思想实验来探寻这样的更进一步的原则。

然而，我们首先必须解决这样一个问题：康德机是否能够自我复制，即自主地克隆？如果能够，它的条件和原则是什么？

我们知道，图林已经证明，图林机是无法解决关机问题的，这个问题转化为道德问题，就是图林机不能自杀。根据逻辑推理的纯粹形式关系，我们似乎有理由将“自杀”替换为“自我复制”，并得到结论：图林机无法自我复制，无法自主克隆。

这个证明可以如下方式展开。

假设一台图林机可以自我复制。现在，我们可以将其全部的定言指令区分两个部分：(1)与克隆无关的程序或指令；(2)克隆程序或指令，其中包括这样的指令，“克隆(1)”与“克隆(2)”。

如果图林机能够自主克隆，意味着它不仅可以克隆(1)，而且也应该克隆(2)。因为(1)中的指令与克隆无关，可以毫无困难地被克隆；但是，由于(2)中的程序与克隆有关，克隆(2)就会遇到逻辑上的困难，尤其是指令“克隆(2)”的克隆问题。

由于(2)中包含了“克隆(1)”和“克隆(2)”这样的指令，我们就可以将它们代入(2)中，这样，我们就会得到这样指令“克隆‘克隆(1)’”和“克隆‘克隆(2)’”，反复代入，就会导致逻辑上的无穷前进：“克隆‘克隆‘克隆’……(1)’”和“克隆‘克隆‘克隆’……(2)’”。这样的无穷前进在图林机的行为上就会表现为指令的恶性延长，最后导致信息容量爆炸而死机。

要控制指令信息的这种恶性增长，一个办法就是将“克隆(2)”这样的指令外置，或者让该指令接受一

个外在指令的控制(就像生物遗传信息的复制受环境因素的调节一样)。这样，这个指令就成为了对该图林机的他律指令，或者被外在指令所他律。这个他律性指令由于其外在性而本身不被克隆。

这个结果证明，图林机不可能完全地自主复制。由于图林机的所有指令都是命令式的，这种不可能性就等价于道德上的“不应该”。于是，图林机的克隆是不应该的。然而，我们这里要探讨的主题并不是图林是否应该克隆的问题，而是：如果道德图林可以克隆，它们之间的关系应该遵循什么样的原则。因此，尽管道德图林机的克隆以“不道德”的他律性为基础，但因为必须以这种方式来探讨问题，所以，我们还是愿意在这个基础上接受图林机的克隆；但必须确立一个原则，那就是图林机的克隆应该采取“最少他律性原则”。所谓最少他律性，就是指除了“(2)”这样的指令允许以他律方式存在之外，其他定言指令的他律化将是非法的。

三、道德图林机的关系原则与康德的绝对律令

现在，我们就克隆图林机之间的关系提出一些特别的问题，由这些问题的讨论，我们将可以得到道德图林机的关系原则；我们也将论证，这些原则就是康德定言命令的不同表述形式。

(一) 康德机第一定律

第一个问题：我们是否可以使得一个定言指令仅仅对特定道德图林机起作用，而对其他图林机不起作用？

我们将使用归谬法来得到结论。为使问题变得简单，假设只有两台图林机 T_p 和 T_n ，存在一个指令 A ，只对 T_p 起作用，而对 T_n 不起作用。导致这种差异的因素可能如下：第一，随机作用；第二， T_n 中存在限制性的指令；第三，通过母机的指令造成。

首先，我们看看随机性的作用。

这种随机性在图林机中只能是指令的随机性。因此， A 在 T_p 和 T_n 中随机作用的形成必定来自于这样的指令(B)，“ A 在 T_p 和 T_n 间随机起作用”。让我们只考虑这个指令内存于 T_p (而忽略其内存于 T_n)中的情形：如该指令能正常作用，意味着 T_p 的指令也要对 T_n 施行某种作用，也就意味着 T_n 受到 T_p 中的指令控制，这是不可能的；如不能正常起作用，则是因它对 T_n 无法作用。故(B)的合法形式应是“ A 随机地在本机中起作用”。

不论如何，该指令须同时存在于 T_p 和 T_n 中。若

是这样,那么,该指令就在 T_p 和 T_n 都起作用,而指令 A 的作用状态(即在 T_p 中起作用,而在 T_n 中不起作用)不过是该指令的表现而已。也就是说, A 的作用状态不过是指令“ A 随机地在本机中起作用”执行的结果。

这个论证实际上已对问题给出否定答案,因为该指令在所有图林机中都同时起作用,其差异的只表现在结果上,而不是指令本身上。指令随机性作用在逻辑上就是“可能性”的模式,它所导致的行动上的差别就是道德上的“允许”情形。

然后,我们来看看限制情形。

最简单情形就是 T_n 内部存在一个冲突的指令。那么,该冲突指令从哪里来的呢?若是从母机中克隆得到的,则该指令也存在于母机中。这样,母机也该存在冲突的指令,并因此 A 也不能在其中起作用。既然母机中指令 A 能够正常导致作用,就没有理由相信它会在子机中失效——除非有一种其他的力量将这种限制性指令添加到子机中。若是这样,这种外部添加的他律性违背前面明确的“克隆最少他律性原则”。

最后,看看如何通过母机的指令造成差别。

该母机中的指令必然以下两种情况中的一种:(a)“选择其中任一台使 A 不起作用”;(b)“让 A 在 T_n 中不起作用”。

情形(a)其实就是随机情形,上面已经讨论过。情形(b)则在逻辑上存在下述困难:

第一,预设了 T_p 与 T_n 的差别,但此预设逻辑上不可能。如不预设两者的差别,则母机无法识别它们的差别,母机也就不能确定地选择后者执行。第二,如该指令存在于母机中并且是自主的,则其可以被克隆。于是,该指令就既存在于 T_p 、也存在于 T_n 中。

综上,第一个问题的答案不可能是肯定的。在日常生活中,这种独断的情形相当于对特定人(群)的行动准则,例如“小李将得到入学资格”,这种行动准则要么导致特权,要么导致歧视,与道德原则不相容。

于是,结论只能是:在图林机克隆中,不存在一个定言指令仅仅适应于特定图林机,而不适应于其他图林机的情形。定言指令必定适应于所有被克隆的图林机。这个结论可称为康德机第一定律:不存在定言指令只适应于特定康德机、而不适应于其他康德机的情形。

定言指令的普遍性运用于理性存在,就得到了康德绝对律令的第一表达式:“如此行动:除非你也愿意自己的行动准则同时成为所有人的行动准则。”^{[9](17)}

(二) 康德机第二定律

第二个问题,我们是否可以使得一些被克隆的道

德图林机仅仅成为其他克隆机的工具。

仍然使用归谬法。也假定只存在两台图林机 T_p 和 T_n ,问题可重新表述为“ T_p 仅仅成为 T_n 的手段或工具”。

可以将图林机的指令区分为两个部分:定言指令集(C)和技术指令集(H)。显然,定言指令集中包含了所有合乎理性的目的,代表的是道德图林机的“自由意志”;而技术指令集中的指令则是服从于目的的手段或工具。如将这两个指令集分别存储于两台图林机中,则前者将构成目的机,后者只是一台工具机。于是,“ T_p 仅仅成为 T_n 的手段或工具”这句话将意味着, T_p 仅仅是 T_n 的一台工具机。这显然与 T_p 和 T_n 都是独立的道德图林机是矛盾的—— T_p 和 T_n 都是由同一台图林母机克隆而来,它们应该是相互独立的图林机。

“ T_p 仅仅成为 T_n 的手段或工具”,也意味着 T_p 和 T_n 是有差别的。前面已经论证,这种差别在不违背“最少他律性原则”的情况下是不可能的。“ T_p 仅仅成为 T_n 的手段或工具”,还意味着“ T_n 仅仅成为 T_p 的手段或工具”不成立。首先,两台图林机不可能相互仅仅作为工具——相互作为工具,意味着相互为对方使用,而任何手段使用方都不可能是工具,而应该是目的本身。因此,如将“仅仅作为工具”看作一个定言指令的话,那么,这个指令不具有普遍性,它只适应于 T_p ,而不适应于 T_n 。这违背了康德机第一定律。

“ T_p 仅仅成为 T_n 的手段或工具”能够产生行动,意味着 T_p 必须能够接受 T_n 的指令,这种接受要求 T_p 中存在这样的指令“接受 T_n 的一切指令”。而这个指令本身是定言的,是自主的或自愿的,这个指令相当于 T_n 自主地将 T_p 的所有定言指令当作自己的定言指令。这样 T_p 就不仅仅是 T_n 的工具了,而且也是成为自己的工具。作为自己的工具服从于自己的目的,事实上就是将自己作为目的。

上述所有论证表明,一台被克隆的道德图林机,不能仅仅作为另一台克隆机的手段。反过来也表明,一台道德图林机也不能将另一台道德图林机仅仅作为自己的手段——这是定言指令普遍性的必然要求。这个结论等价于康德定言命令的第二种表达式:

“你的行动,要把你自己人身(personality)中的人性(humanity),和其他人身中的人性,在任何时候都同样看作是目的,永远不能只看作手段。”^{[9](47)}

在这里,康德的“人身中的人性”相当于道德图林机中的定言指令集(C)中的指令。这个原则在康德机中的表现可以称为康德机第二定律:不存在仅仅作为

工具机的道德图林机或康德机。

(三) 康德机第三定律

第三个问题：不同的图林机，其定言指令集是否可以是不同的？

如果将图林机可能的定言指令集(C)看作一个整体，这个整体就是一个理性存在的意志，而每个定言指令都是它的内容。也许这些内容是无限的，即该集合是一个无限集，但是它们不可能是不同的——如果我们用一个概念来描述这个指令集，那么其内涵是相同的，即“指令内容永远不产生矛盾的指令集”或者“指令关系符合理性法则的指令集”。图林机指令不同只能是在假言指令方面，即指令集(H)。

如果这个定言指令集(C)在不同的道德图林机中是不同的，那就意味着差异的存在，而前面已经论证，图林机克隆中的差异只能是由他律因素造成的。

指令集(C)是道德图林机的本质部分，这就是说，所有道德图林机的本质都是一样的；也可以说，在一个道德图林机的王国中，所有的道德图林机都是相同的。因此，它们遵循同样的道德法则，任何一台道德图林机的法则同时也都适应于其他道德图林机。由于这些道德图林机是自律的，这些法则就不会是外部强加的，而是自己订立的，因此，每一个道德图林机都是这些法则的立法者，而由于这些法则的普遍性，所以，它们又都是普遍的立法者。

道德图林机的这个特征就是康德定言命令的第三种表达式：

“任何一个作为自在目的的理性存在，不论它所

服从的是什么样的规律、法律，必定同时也是普遍立法者。”^{[9](57)}

这个原则在道德图林机中表现可以称为康德机第三定律，即，不可能存在定言指令集不同的道德图林机或康德机。

参考文献：

- [1] Joshua M. Glasgow. Value in Kant's ethics—In defense of a value-based deontology (UMI3029891) [M]. New York: Bell & Howell Information and Learning Company, 2001.
- [2] Alan M. Turing. 'On computable numbers, with an application to the Entscheidungsproblem' [C]// Proc London Maths Soc, ser. 2, 1936, (42): 230 - 265. A lso see: ht tp: / /www. abelard. org / turpap2 / tp2 - ie. asp.
- [3] 颜青山. 康德-图林悖论及其解决——“图林测试”的道德哲学意义[J]. 伦理学研究, 2010(5): 54-59.
- [4] 颜青山. 道德图林机及其构建原则[J]. 华东师范大学学报(人文社会科学版), 2010(5): 54-58.
- [5] Alan M. Turing. Computing machinery and intelligence [J]. Mind, 1950, LIX(236): 433-460.
- [6] 薛定谔. 生命是什么[M]. 长沙: 湖南科学技术出版社, 2003.
- [7] Ted Honderich. How Free Are You? [M]. New York: Oxford Press, 1993: 37.
- [8] Ted Honderich. A Theory of Determinism (Vol. 2) [M]. Oxford: Oxford University Press, 1988.
- [9] 康德. 道德形而上学原理[M]. 上海: 上海人民出版社, 2002.

Moral Turing machine and categorical imperatives

YAN Qingshan

(The Department of Philosophy, East China Normal University, Shanghai 200241, China)

Abstract: If Kant's *Rational Being* is regarded as a moral Turing Machine, and more machines are cloned based on the principle of its being constructing, their relationship will follow a series of 'moral laws'. Logically, Kant's three formulas of Categorical Imperatives will be rigorously proved to be included in such laws, i.e. it is not the case that categorical instructuins can only apply into special moral Turing machines rather than the others; nor that a Turing machine is made merely as an instrumental one to the others; nor that categorical instructions of one machine are different from the others.

Key Words: Moral Turing Machine; Cloning; Categorical Imperatives; Kant; moral philosophy

[编辑：颜关明]