

功利主义在无人驾驶设计中的道德算法困境

隋婷婷¹ 张学义²

(1. 北京大学 哲学系 北京 100871; 2. 东南大学 哲学与科学系 南京 211189)

摘要: 无人驾驶的发展伴随着算法设计的诸多道德争议。电车难题是其面临的关键道德困境之一。功利主义作为电车难题中最为大多数所接受的选择,是最受期待的算法之框架。但根据伯尼法等学者的调查,功利主义尽管最符合大众偏好,在购买调查中,使用该算法的无人驾驶车却乏人问津。究竟是被试的道德偏好有所转变,抑或是道德偏好独立于选择偏好之外?电车难题的功利主义偏好是否应成为算法设计的依据?针对以上疑问,我们设置了选择与道德接受度并行的三情境实验,发现被试的选择与道德偏好并不构成连动关系,电车难题的既有偏好不足以作为无人驾驶采取功利主义算法的支持论据。根据现有算法的种种局限,无人驾驶算法的未来方向应当是跳出功利主义与利己主义框架,且能够兼容应然与实然需求的第三种方向的算法。

关键词: 无人驾驶; 电车难题; 道德算法; 功利主义; 实验哲学

中图分类号: B82 **文献标识码:** A

DOI:10.19484/j.cnki.1000-8934.2021.10.019

一、引言

21世纪是无人驾驶蓬勃发展的时代,得益于视觉识别、数字导航以及人工智能系统的发展,无人驾驶汽车在现实路测中已拥有与人类不相上下的驾驶能力^[1],让智能软件接管方向盘,为全世界提供全自动出行方式的未来已不再遥远。

相比传统驾驶,无人驾驶有望为交通领域带来更高的安全预期值,在美国公路交管局的年调中,人为失误是90%以上交通事故的主因^[2],无人驾驶的稳定性是减少人因事故,提高出行安全预期的不二选择。

然而驾驶框架中先天地内嵌着一些无法绕过的道德困境,如电车难题(trolley problem)。传统驾驶情境中,人类掌握方向盘,无论车辆失控状态下,个体是遵循功利主义,选择舍少救多,抑或是遵从道义论,都可作为道德主体在短时应激任务中合理

的个体化选择,但无人驾驶的算法先置性使得事故状态中的选择不可能成为时间压力下未经深思且带有偶然性的反应,设计者不得不预先设定算法的道德倾向。

让·弗朗索瓦·伯尼法(Jean-François Bonnefon)等研究者提出应进行诉诸大众直觉的道德偏好调查,选取一种更符合大众道德偏好,且能提高社会安全预期的原则作为道德算法。2015年的实验中,90%左右的被试均对电车困境表现出功利主义偏好,面对自我牺牲情境时这一偏好也未改变,75%以上的被试仍选择了功利主义式的牺牲小我,拯救多数行人。^[1]

功利主义至此似乎成为算法的首选,但在随后对功利主义算法车辆的购买调查中,被试的选择却与之前的道德偏好大相径庭,对功利主义车辆的接受度仅在33%左右。^[3]

无人驾驶的道德算法由此面临两个新问题:究竟是购买情境促使被试的道德偏好突然转变,抑或是道德偏好独立于选择偏好之外?电车困境中大

收稿日期: 2020-2-23

基金项目: 国家社科基金重大项目“技术创新哲学与中国自主创新的实践逻辑研究”(19ZDA040); 国家社科基金重大项目“人工智能伦理风险防范研究”(20&ZD040); 中国博士后科学基金(2021M690243)。

作者简介: 隋婷婷(1989—),女,江苏南京人,北京大学哲学系博雅博士后,主要研究方向: 认知神经哲学、实验伦理学、人工智能哲学; 张学义(1983—),通讯作者,安徽阜阳人,东南大学哲学与科学系副教授,主要研究方向: 科学哲学、实验哲学、科技伦理。

众的功利主义偏好是否应如部分欧美研究者所说的:成为算法设计的依据?

针对以上疑问,我们设置了道德接受度与实际选择并行的三情境实验。

二、实验数据与结果

322人(女164,男158)参加了本次实验。实验采取2(电车困境:是vs否)×2(无人驾驶困境:是vs否)×2(购买意向情境:是vs否)混合因素设计。电车困境中,被试需决定是否扳动铁轨上的道岔,让失控电车驶入岔道撞死1名工人,或选择不扳道岔,任由电车撞死主轨上的5名工人;无人驾驶困境中,被试需决定是否设计“避让5名行人牺牲车上1名乘客”的功利主义算法;在随后的购买意向情境,被试将决定是否购买功利主义算法无人驾驶车。除做出(是/否)选择外,被试还需对三个情境的牺牲选项在七点量表上(1为完全不能接受,4为不确定,7为完全能接受)做出道德接受度选择。

1. 情境选择数据

电车与无人驾驶困境中半数以上的被试选择舍少救多,并赞同设计功利主义的无人驾驶车辆,但对于采取这一框架的车辆,购买意向仅有13.4%(见图1)。

购买意向与其他两情境在卡方检验中均有显著差异,有意愿购买功利主义无人车的被试人数显著低于其他情境:购买意向vs无人驾驶困境, $\chi^2(1,644) = 120.83, p < 0.001, \phi = 0.43$,购买意向vs电车困境, $\chi^2(1,644) = 127.15, p < 0.001, \phi = 0.45$ 。电车困境与无人驾驶困境之间的差异不显著, $\chi^2(1,644) = 0.1, p = 0.41, \phi = 0.01$ 。

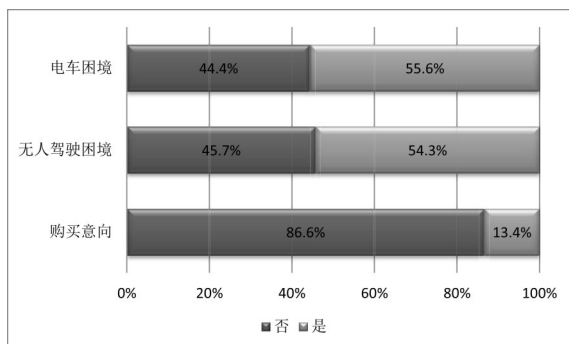


图1 情境选项

2. 道德接受度数据

道德接受度在情境间对比中存在差异:电车困境($M = 3.48, SD = 1.62$)、无人驾驶困境($M = 3.05, SD = 1.66$)以及购买意向($M = 2.89, SD = 1.84$)中道德接受度在2(性别)×3(情境类型)的AVONA分析中,情境类型间差异显著, $F(2,640) = 17.49, p < 0.001, \eta_{p2} = 0.05$,性别差异不显著, $F(2,640) = 0.58, p = 0.54, \eta_{p2} = 0.001$ 。

配对T检验中,被试对电车困境中功利主义选项的道德接受度显著高于无人驾驶和购买情境,无人驾驶与购买情境的道德接受度则无显著差异,电车困境vs无人驾驶困境, $t(1,321) = 4.65, p < 0.001, d = 0.26$,电车困境vs购买意向, $t(1,321) = 4.95, p < 0.001, d = 0.34$,无人驾驶困境vs购买意向, $t(1,321) = 1.613, p = 0.12, d = 0.09$ 。

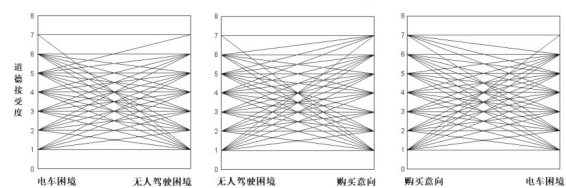


图2 道德接受度配对T检验

三、无人驾驶的算法困境

道德接受度的情境对比中,被试对电车困境中功利主义的接受度略高于其他情境。选择趋势方面,中国被试对功利主义的选择率低于欧美,但在电车与无人驾驶困境中对功利主义均表现出高于购买情境的选择偏好。

1. 无人驾驶算法标准:道德偏好vs选择偏好

不论是欧美背景下的实验抑或是本次实验,被试在选择偏好上均表现出道德认知与实际选择的非一致性,因此,单纯以大众在道德困境中的功利主义偏好作为无人驾驶算法的标准将陷入乞题困境。

道德认知方面,道德偏好与实际行为的背离反应了道德判断的双加工模型(dual-process model),即道德判断和选择是代表理性的认知与代表感性情感模式相互作用的结果。道德两难中,涉及

直接伤害的“亲身性(personal)”两难困境更易引发情感反应,而被试在“非亲身性(impersonal)”情境中情感唤起较轻,更易作出功利主义选择。⁽⁴⁾ 购买情境对被试而言,更易激发情绪主体关于个体生存发展和安危状况的关注,其自我卷入度和情感唤起程度高于接近抽象假设的电车困境,换言之,情感使生存安危的前提突显出来,影响了主体的行动偏好。

实验量表方面,道德与选择偏好的背离涉及迫选差异,无人驾驶困境中的选择本身是“在两种罪恶(撞倒行人或牺牲乘客)中的不得已”⁽³⁾,选项的先验性限制了道德选择的范围。因而被试面对两难之外的购买情境时,大多选择了拒绝购买,避免涉入可能出现的迫选困境,但这一选择回避并不等同对功利主义道德接受度的下降,因为“某人可能因胆怯失去付诸行动的意愿,但这并不妨碍他称赞勇敢是一种可贵的品质”⁽⁵⁾²⁰⁹。

个体的道德心理模型方面,被试可以接受情境中的道德迫选,但主动趋近这一情境则涉及个体道德动机调节方式的差异,使个体面临自我牺牲的趋避判断,唤起被试的“损失厌恶(loss aversion)”,演化心理学家将损失厌恶看作人类在进化中获得的一种适应性本能:尽管收益能改善生存前景,但重大损失却会让他们彻底“出局”。⁽⁶⁾ 因此在不确定条件下,对收益或损失的非对称性选择模式将不可避免。

2. 无人驾驶算法方向:功利主义 vs 利己主义

在道德接受度与实际选择背离的前提下,以功利主义作为无人驾驶的算法显然不明智。科技产品的推出建立在市场需求之上,过低的需求无法拉动投入产出比。无人驾驶对安全的提升也必须建立在群体对车辆的普及使用上,使用率低下对安全的提高并无助益。在这一前提下,无人驾驶车是否应遵循大众购买偏好,放弃功利主义,以全力保护车主的利己主义算法为标准?

初看之下,利己主义算法能最大限度满足使用者的安全预期,促进无人驾驶普及,利己主义在极端两难情境中尽管将造成更大伤亡,但此类情境的发生毕竟是小概率,无人驾驶似乎仍有可能完成减少事故、提高安全性的目标。

但事实并非如此,从伦理中立角度看,将车主的生存置于他人之前将引发:制造商是否有权将死亡风险平摊至卷入事故的行人身上? 由算法引发

的行人伤亡如何归责等不正当风险分配的权责争议。同时,不论制造商是否承担责任,这一非中立算法最后也必然会使车主承担起比以往更大的事故责任,因为“使用这一算法的车辆,是经由车主带入公共空间的”⁽⁷⁾⁸⁰。作为车辆的所有者和安全保障的最大受益人,车主必将为自身安全预期付出高昂的代价,从而可能重新引发大众对无人驾驶车辆的购买拒斥。

此外,车主即使付出高昂代价,安全性实质上也未提高。尽管1名车主与5名行人在功利主义算法中,车主的死亡预期值远高于利己主义算法,但社会情境下,车主与行人的角色属性并非恒定不变,特定个体作为车主这一角色时死亡预期值的下降,并不代表其社会总体死亡预期值下降。社会总体死亡预期值在利己主义预设下是: $1/6 * 0 + 5/6 * 1 = 5/6$; 而在功利主义预设下的死亡预期值则只有: $1/6 * 1 + 5/6 * 0 = 1/6$ 。因而利己主义算法并不能成为合理的替代算法,寻找超出功利主义和利己主义之外的道德原则仍是未来算法研究的当务之急。

四、道德原则算法化难题溯源

“电车难题”作为探讨功利主义与道义论的经典母题,主要涉及个体生命价值与群体生命价值在牺牲情境中的正当性的探讨,其算法化难题的成因关涉“电车难题”本身的缺省逻辑预设、道德算法一元化的悖论以及算法的“物律”需求等三个方面。

1. 电车难题算法化的逻辑困境

“电车难题”及作为对照组的衍生版本“天桥难题”(选择是否将一人推下天桥,以一人的牺牲拯救铁轨上五人的生命的决策困境),是无人驾驶道德算法的主要情境模板与母题来源,这一思想实验延伸出的算法困境与其包含的三个逻辑缺省预设相关:

1 在“电车难题”中人们通常具有功利主义的先在倾向;

2 在“天桥难题”中人们通常天然地倾向于道义论;

3 在无人驾驶的“电车难题”困境中人们通常会延续自身的道德选择倾向。

相比做出全有或全无判断的经典逻辑,缺省逻辑(default logic)阐释了人们在知识不完全的前提

下所做的推理,其预设是模糊且具可错性的。⁽⁸⁾这一点在实验框架的大众直觉检测中得到了验证,预设1与预设2作为对经验性事实的预设判断与实验结果存在差异,不同文化背景下功利主义的道德倾向比例并不具有原初思想实验中所预设的普遍性。

同时,缺省逻辑不同于经典逻辑前件为真则结论为真的单调性特征,具有默认前提下某事物为真的非单调缺省特征,即前提为真的条件下不一定蕴涵结论为真。⁽⁸⁾因此即便依照欧美背景下伯尼法等人在实验中对预设1的肯定,情境中车主的直接牺牲,预设3在“直接伤害”等“亲身性”层面与“天桥难题”重合,派生出预设1与预设2的重叠悖论。该重叠造成了在伦理层面上区分“电车难题”与“天桥难题”的双重效应原则(doctrine of double effect)的失灵。该原则中,电车情境由功利主义选择造成的1名工人牺牲可被归属为拯救5名工人时可预见的副效应,因而在道德上可被允许;但在天桥困境中采取功利主义选择时,1人的牺牲是作为主效应的有意伤害,被拯救的5名工人是1人牺牲前提下的副效应,因此不能被大众的道德判断所允许。⁽⁹⁾⁹³

预设的重叠困境使道德算法设计者不得不放弃价值层面对道德理论的正当与否进行规范性分辨,转而由现实情境的单一“电车难题”切入,剔除道义论假设(预设2),单纯采用功利主义假设(预设1),以期将道德算法正当性的“应然”问题,转化为实际情境中如何能够被大众直觉普遍接受的道德算法“实然”问题。但这一做法的隐含前提在于无人驾驶的道德算法先置原则在本质上默认了人们的道德选择具有恒常性与稳定性,换言之,道德算法有着德性伦理学中强调道德品格一贯性的理论底色。然而,电车难题与相关衍生难题是在抽象了牺牲对象的前提下,以不同情境模式为变量从而探讨不同道德选择的思想实验,其本质上是情境主义的,从而加剧了道德选择与情境施行的算法鸿沟与选择悖论。

2. 无人驾驶的一元化算法悖论

在道德规范性层面,电车难题作为常被用于阐述个体遭遇的公正悖论以及个体道德倾向非一贯性的例证,其原初目的并不在于证明道义论较之功利主义在两难情境下更具正当性,而是反驳将功利主义推崇为唯一具有正当性的道德规范。换言之,电车难题的结论在于对价值独断主义框架的批驳,以及对道德价值多元存在的反思。

但在无人驾驶算法中,多元道德价值内核在本质上与道德算法的一元化需求难以兼容。在车辆动态集群配合层面,算法的非统一性将使得无人驾驶车辆在集群协调方面引发难度的上升以及容错率的下降。在道德规范性层面,若采取多元可选式的道德算法,将算法设置权交予使用者,无人驾驶车辆由此成为使用者的道德代理,这一作法看似延续了个体在道德两难当中立场和信念的多元性,却衍生出两个伦理难题:

其一是“群体歧视”。算法的自由设置权给予了主体在事故中先置群体牺牲偏好的权力,也即允许了在事故产生前,对某些特定群体(如少数族裔、特定性别群体等)的歧视预设,在算法层面上的正当化。

其二是“囚徒困境”。多元选择看似赋予使用者自由选择权,但个体做出的理性选择常导致集体博弈的双输,囚徒困境中,双方个人的理性最优解往往等同于群体的内卷式虚耗,这一情境又将促使个体做出道德偏好之外的选择,即若有相当数量的人选择利己主义,其他使用者在这一场不透明的非零和博弈中,会更倾向于积极保护自身权益作出利己化选择,因为大众愿意选择代表自我牺牲的功利主义,通常包含了他人也愿作出同样牺牲的公平性前提⁽¹⁰⁾,在这一囚徒困境的持续博弈中,情况将再次回到利己主义算法提升社会死亡预期的不良循环。

面对这一困局,采用一元化的道德算法,仍是重新达成道德均衡的关键,可能也是唯一的办法。⁽¹¹⁾

3. 道德算法的“物律”需求

在形而上的层面,道德判断双重效应原则的失灵与道德两难造成的两可性选择是可并存、可悬置的,但在算法层面却无法复刻道德选择在悬置或重叠中形成的类薛定谔状态。

这一重叠映射了实然情境与逻辑抽象的落差,因为“逻辑学的限制使它有权、甚至有义务抽掉知识的一切对象和差别”⁽¹²⁾¹¹,实然情境则不能拥有这一特权。道德算法的本质是将道德规范写入技术物的“道德物化(moralizing technology)”过程;“道德物化”作为荷兰学派将技术道德化的代表性概念,隐含了道德施动者的转变。⁽¹³⁾道德原则由此在算法框架下对于人的意义由哲学语境中的“自律”或“他律”转化为了由物对人的“物律”。在实

质上,道德两难情境实验的意义已由反驳形而上层面的道德原则正当性,转向了获得主体对于道德判断“物律”知情同意的契约性反馈过程。

因而当问题回到道德算法本身,解决道德矛盾的思路仍是“应然”与“实然”分离,将道德原则的正当性、规范性讨论归还形而上学,以大众直觉实验对实际情境中道德算法的可行性进行验证和改良,目前可选的道德算法并不限于功利主义、道义论以及利己主义,以“无知之幕(veil of ignorance)”与“最大化最小原则(maximin procedure)”缓和牺牲困境的罗尔斯算法也是可能的选择之一,该算法通过对安全最大化底线达成帕累托最优原则(the Pareto principle)^[14],能够避免直接以人数为变量做出先置牺牲选择的算法难题,从而在一定程度上消除个体对自我牺牲的“损失厌恶”。

尽管“无知之幕”本身受到的批驳(如不公平的风险分摊)可能和功利主义平分秋色,但由于电车难题的决策情境中并不存在一个超越所有原则的、最具正当性的道德规范,不论选择哪一种道德理论,都需面对其固有的缺陷。若强求无缺陷的规范,道德算法将陷入诉诸完美的逻辑困境。因而在无法兼顾“应然”与“实然”的前提下,为避免陷入“布里丹之驴(Buridan's ass)”式的理性无作为状态^①,无人驾驶的未来研究方向仍是面向广阔的道德理论,选择应然层面上具备一定的正当性,实然层面具有可实现性且能够切实提升安全性的算法架构。

五、结语

无人驾驶的“技术奇点”已近在眼前,算法设计涉及的电车困境亦是中外学界无可规避的难题,笔者认为造成这一难题的主因有四个方面。

首先,前提的误用。研究者目前普遍寻找的是应然层面上的最优算法,该做法包含“应该蕴含能够”的隐性前提,“即假如我们被要求应当做某事,我们就能够做某事”^{[15]40}。但一方面,人们不能够完成某行为,并不妨碍它成为义务;另一方面,该前提也与算法的实验结果相去甚远:实验情境中即便

多数人认同面对电车困境“应该”牺牲少数,更多的人并不“能够”接受且实施这一做法。

其次,选择的鸿沟。研究者们试图通过诉诸大众直觉确定道德算法的选择,但大众选择本身表现出了不同文化背景浸染下的巨大差异,且大众直觉的选择仅能提供实然性的“是”,无法提供代表正当性的“应该”。尽管“是”与“应该”可由“主体意图”推动在有限条件下的逻辑推导,即某人意图做某事→做到某事的唯一方法“是”如何→某事“应该”如此做。然而步骤二中某事做法的唯一性恰是实验结果所不具备的,因而导致“是”与“应该”的鸿沟在无人驾驶算法实验中难以弥平。

其三,实验的效力边界。无人驾驶道德困境实验脱胎于电车难题,其意义在于通过道德两难选择的公正悖论对价值独断主义进行反驳与思考,而非通过大众直觉指定某种道德理论具有唯一可行性。同时,诉诸经验的实验在对道德直觉进行验证时,易于陷入由直觉对直觉的循环论证当中。^[16]因而作为电车困境衍生的无人驾驶实验提供了一个从情境观察人类道德多元性的视角,有其证伪性和检验性价值,能够对道德算法给出“何者不能”的答案,但却无法直接回答算法在道德选择上“何者可能”的问题。

最后,道德原则与道德算法的维度差距。哲学思辨由于其本身澄清理论、去除变量、追问终极的需求,必然需要对思想实验进行降维式的抽象处理,只讨论剥离具体身份的抽象人,与无人驾驶情境中现实维度的具体对象存在差异。但在实施层面,又需对道德原则进行具象的重构与改造,这也是道德原则很难被完整还原为算法的原因。

换言之,道德算法困境的产生源于抽象人与具象人、降维与升维、应然与实然间的解释鸿沟。人们提出了一个道德两难困境中应当如何选择的应然性问题,但却要求给出一个能够如何选择的实然性答案。因而笔者认为算法难题的解决仍需借助应然性与实然性检验并行的复合式构建方式,将应然层面上具有伦理正当性的道德算法代入实际情境,以检验其是否能够兼顾实然情境中争议性风险分配、生命权冲突以及生命选择顺位等难题。

^① 布里丹之驴是以14世纪法国哲学家布里丹命名的悖论,这一悖论指出,当一只完全理性的毛驴处于一模一样且距离相同的两堆干草中间时,无法对究竟该吃哪一堆草做出理性决定,因而无法行动以至饿死。

参考文献

- (1) Bonnefon J, Shariff A, Rahwan I. Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars [J]. *ArXiv*, 2015, 10(13): 1–15.
- (2) Sütfield L R, Gast R, König P. Using Virtual Reality to Assess Ethical Decisions in Road Traffic Scenarios: Applicability of Value-of-Life-Based Models and Influences of Time Pressure [J]. *Frontiers in Behavioral Neuroscience*, 2017, 11(122): 1–14.
- (3) Bonnefon J F, Shariff A, Rahwan I. The Social Dilemma of Autonomous Vehicles [J]. *Science*, 2016, 352(6293): 1573–1576.
- (4) Greene J D. An fMRI Investigation of Emotional Engagement in Moral Judgment [J]. *Science*, 2001, 293(5537): 2105–2108.
- (5) Foot P. Moral Belief [C] // Hudson W D. (ed.) *The Is – Ought Question*. London and Toronto: Macmillan, 1969: 196–213.
- (6) 贾拥民, 黄达强, 郑昊力. 偏好的异质性与一致性——基于损失厌恶和脑刺激的神经经济学研究 [J]. *南方经济*, 2015, 33(5): 97–119.
- (7) Lin P. Why Ethics Matters for Autonomous Cars [C] // Maurer M, Gerdes J C, Lenz B, et al. (eds.) *Autonomous Driving*. Berlin: Springer, 2016: 69–85.
- (8) Bochman A. Default Logic Generalized and Simplified [J]. *Annals of Mathematics and Artificial Intelligence*, 2008, 53(1): 21–49.
- (9) Kamm F. *Intricate Ethics: Rights, Responsibilities, and Permissible Harm* [M]. New York: Oxford University Press, 2007.
- (10) 王珀. 无人驾驶与算法伦理: 一种后果主义的算法设计伦理框架 [J]. *自然辩证法研究*, 2018, 34(10): 70–75.
- (11) Gogoll J, Müller J F. Autonomous Cars: in Favor of a Mandatory Ethics Setting [J]. *Science and Engineering Ethics*, 2017, 23(3): 681–700.
- (12) [德]康德. 纯粹理性批判 [M]. 邓晓芒, 译. 北京: 人民出版社, 2004.
- (13) 王小伟. 荷兰学派道德物化观点溯源 [J]. *自然辩证法通讯*, 2020, 42(6): 42–47.
- (14) Leben D. A Rawlsian Algorithm for Autonomous Vehicles [J]. *Ethics and Information Technology*, 2017, 19(2): 107–115.
- (15) [德]康德. 实践理性批判 [M]. 邓晓芒, 译. 北京: 人民出版社, 2003.
- (16) 费多益. 实验哲学: 一个尴尬的概念 [J]. *哲学分析*, 2020, 11(1): 43–53.

The Moral Dilemma of Utilitarian Algorithm for Driverless Vehicles

SUI Ting – ting¹, ZHANG Xue – yi²

(1. Department of Philosophy, Peking University, Beijing 100871;

2. Department of Philosophy and Science, Southeast University, Nanjing 211189, China)

Abstract: Driverless vehicles are accompanied by many moral disputes in algorithm. The trolley dilemma is one of the most crucial moral dilemmas. Among moral algorithm, utilitarianism, as the most acceptable choice, is the most expected algorithm. However, in a survey conducted by Bonnefon, it's the most preferred choice, but seldom people showed interest in buying driverless vehicles with it. Was the purchase situation a trigger for the moral preferences, or was the morality independent of the selection? Whether the utilitarian preference can be the basis of algorithm design? We set up three experiments to deal with these questions. In the within group comparison, we found that moral and choice preferences were not always linked. The choice preference in the trolley case was not enough to support driverless vehicles in using a utilitarianism algorithm. According to the limitation of existing algorithms, the future direction of driverless vehicles' ethical algorithm should not only break out of framework of utilitarianism and egoism but also need to be compatible with necessity and reality.

Key words: driverless vehicles; the trolley dilemma; moral algorithm; utilitarianism; experimental philosophy

(本文责任编辑: 费多益 朱欢欢)