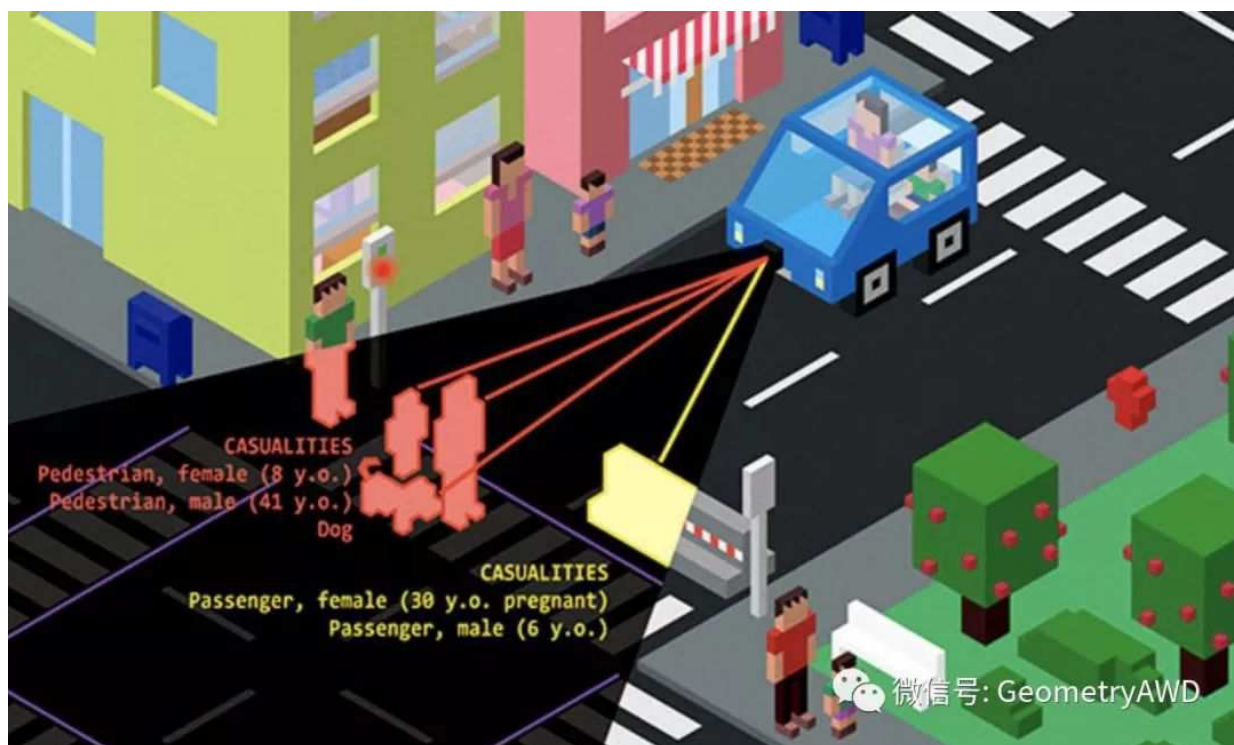


「浪漫科技」与「古典伦理」的碰撞 - 求解无人车技术的道德困局

原创：徐鸿鹄 几何四驱 5月19日



撰文：徐鸿鹄 | 排版：王晓峰

公众号：几何四驱 (ID: GeometryAWD)

序

人工智能界有个说法，认为「**机器学习**」是非生物智能领域里最能体现智能的一个分支。工业界也有个说法，认为声光力电无所不包的「**汽车**」是人类私有财产里最能体现科技含量的产品。

如此一来，「**机器学习**」与「**汽车**」的激情碰撞，想必定会是最美妙的事情。

这就是无人车，需要几十项技术难点的解决方案全面整合才会日趋成熟的技术产品，不光技术公司在关注无人车，还有法律界，政治界，金融界等等，这是一个全人类的话题，无人车被看作智能科技集大成的应用，空间足够大，参与者足够广，时间线也足够长。

和身边「机器学习」和「汽车」两个领域的朋友接触，会经常获得一些道听途说的趣闻以及专家们对于无人车产品的发展现状和未来前途的评论。这其中，「**无人车伦理**」是最受关注的话题之一。在「几何四驱」群内和大家讨论得多了，便难免会产生一些自己的疑问，群内的朋友多次建议我整理自己的看法，并借此文章抛砖引玉展开持续讨论。于是，我便有机会把它们写下来放在这里，姑且算是一种“外行求教无人车伦理”吧。

以「**技术中的道德真空？还是道德中的技术真空？**」，「**法律与道德，何去何从？**」和「**认知科学的奇袭**」三个子篇为题，我试图从科学哲学，法律，认知科学三个独立的角度对无人车「自动驾驶」的伦理话题进行探讨。

最后一篇「**通向实在之路**」，以罗杰·彭罗斯的经典著作名为题，此书号称现代人类文明毁掉之后，重建科技树的圣经，以此为题是希望在无人车伦理甚至机器伦理的道路上，能够涌现出如彭罗斯一般极富人文才情的科学大师，为我们指引未来的方向。

「**无人车伦理**」，这是一个崭新而又内涵广袤的话题，试图通过一篇文章就说清楚一二是相当具有挑战性，由于涉猎不足，更兼时间和精力所限，错谬之处在所难免，承蒙不吝告知，将不胜感激！

引

「无人驾驶落地，AI序幕降临」

摒弃人类的决策力来驾驶汽车的探索始于90多年前。

[1] 1926年12月8日，密尔沃基哨兵报（Milwaukee Sentinel）刊登了一篇“幽灵汽车城市巡游”的告示，借助无线电远程控制的方式，操控一台“无人驾驶”的汽车穿行于密尔沃基的城市街道。



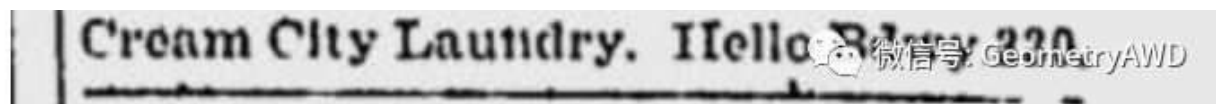
WILL TOUR CITY

A "phantom motor car" will haunt the streets of Milwaukee today.

Driverless, it will start its own motor, throw in its clutch, twist its steering wheel, toot its horn, and it may even "sass" the policeman at the corner.

The "master mind" that will guide the machine as it prowls in and out of the busy traffic will be a radio set in a car behind. Commanding waves sent from the second machine will be caught by a receiving set in the "ghost car."

The tour, conducted by the Achen Motor company, will start at 11:30 a. m. from the company's rooms at Onelda and Jackson streets, will go west on Onelda to Broadway, north to Martin, west to Eighth, south to Grand, west to viaduct, where it will "bout face" and return on Grand to Eighth, south to Sycamore, then east to Broadway and back to the sales rooms. Tomorrow the car will visit Milwaukee-Downer and the Normal school.



很显然这是一场噱头十足的概念炒作，因为实际操控汽车的还是人类自己。

2017年7月，奥迪发布同年量产款的首个真正意义上可以解放人类驾驶员双手，双脚，双眼，并一键触发的具备L3「**自动驾驶**」车型：Audi A8 ！



2017年11月，Google Waymo也不甘示弱，宣布开启更高段位的无人车路试，解放人类司机的大脑，安全员作为乘员跟车测试。



奥迪A8可以在部分条件下完全接管人类驾驶员，可一旦车辆系统故障或失效，驾驶员就要重新接管车辆控制，这注定不是无人驾驶的终极目标。而Google Waymo从一开始就坚定不移地率先进行高等级的无人车测试，它的技术路线代表了一种创造性的尝试：

人类第一次大规模的技术尝试将人类的生命托付给机器智能以换取便利性！

由此可以预见的是：一旦「自动驾驶」技术成功应用，今后人类将大概率地无条件信任其它人工智能产品。



壹

「技术中的道德真空？ 还是道德中的技术真空？」

不同于人类的思维方式，机器更加擅长基于预设的规则批处理指令，IF条件判断语句和THEN执行语句即可组合成基本计算逻辑。


```

if allowed(request.querystring("startFolder"), session("username")) then
    dim filename : filename = request.querystring("filename")
    dim objFile : Set objFile = objFSO.GetFile(getPath(null, filename))
    dim currentAttr : currentAttr = objFile.Attributes
    if request("Submit") = "Save" then
        if request("attr") <> "" then objFile.Attributes = cint(request("attr"))
        set objFile = nothing
        response.write( filename & " has been saved")
    else if request("Submit") = "Cancel" then
    else
        response.write("<form class='msgbox' method='post'><div>
        response.write("<input type='text' value='&filename&'></div><div>
        response.write("<input name='filename' type='hidden' value='&filename&'></div>
        response.write("<input name='action' type='hidden' value='file_save'></div>
        response.write("<input name='startFolder' type='hidden' value='&request("startFolder")&'>
        response.write("<input type='radio' value='0'>
        if (currentAttr AND 0) = 0 or currentAttr="" then response.write(" checked='checked'")
        response.write("</div> None<br /></div><div>
        response.write("<input name='attr' type='radio' value='1'>
        if (currentAttr AND 1) = 1 then response.write(" checked='checked'")
        response.write("</div> Read Only<br /></div><div>
        response.write("<input name='attr' type='radio' value='2'>
        if (currentAttr AND 2) = 2 then response.write(" checked='checked'")
        response.write("</div> Hidden<br /></div><div>
        response.write("<input name='attr' type='radio' value='32'>

```

IF (THIS) THEN (THAT)

微信号: GeometryAWD

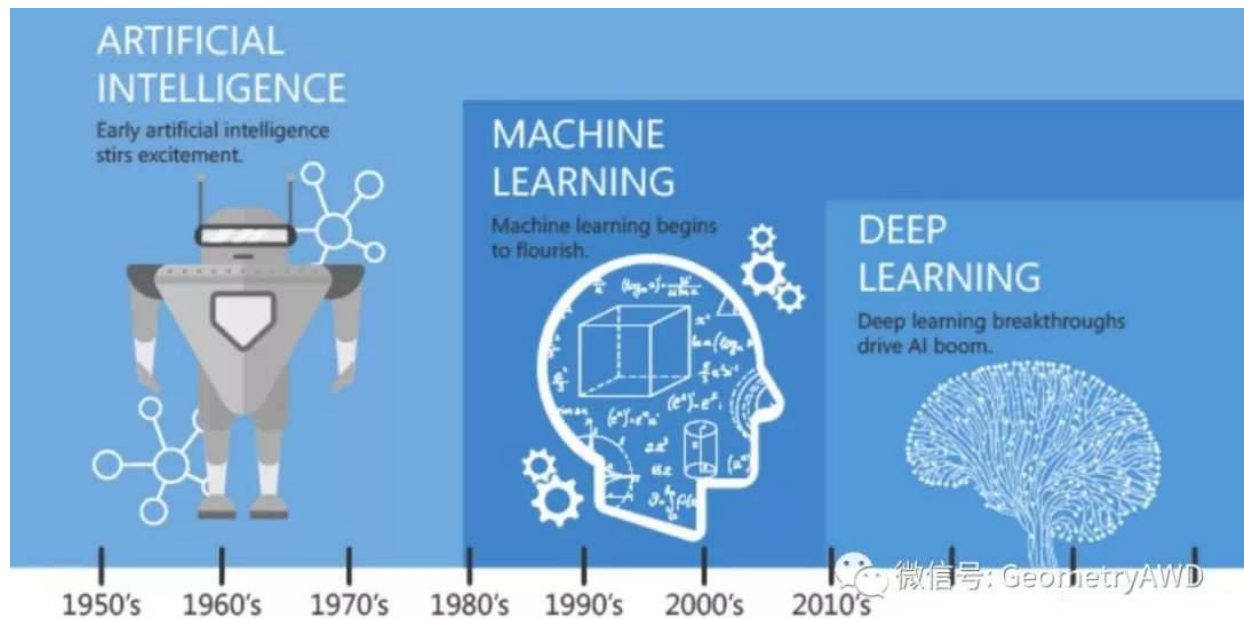
在复杂的无人驾驶场景里，比如环境多变的城市道路上，充斥着交通信号灯，标识，行人，非机动车，路障还有加塞和多变的天气，通过穷举所有的组合条件调取预设命令触发执行机制的方式已经无法有效运作。



上世界80年代开始，机器智能开始加速进化，计算机被人类赋予了学习的能力：

不需要人为的指导，机器就会自主决策！

[2] 2006年开始，机器学习的一个重要分支，深度学习技术的突破裹挟着颠覆性的应用席卷了整个人工智能产业，这些创造旨在松开系在机器上的枷锁，让它潜入数据当中，自主地探索和体验，无人干预地寻找数据模式和关联，最终能够预测未来的行为，后果乃至趋势。



某种程度上来说，**人类赋予了机器思考的能力，使其参与了决策的过程。**

决策过程从此实现了人类与计算机的共同助力，暂且把这种技术和人合作的智能体叫做 [3] **弱人工智能**。

Weak and Strong AI

- Weak AI

- Computers can be programmed to act **as if** they were intelligent (as if they **were** thinking)

- Strong AI

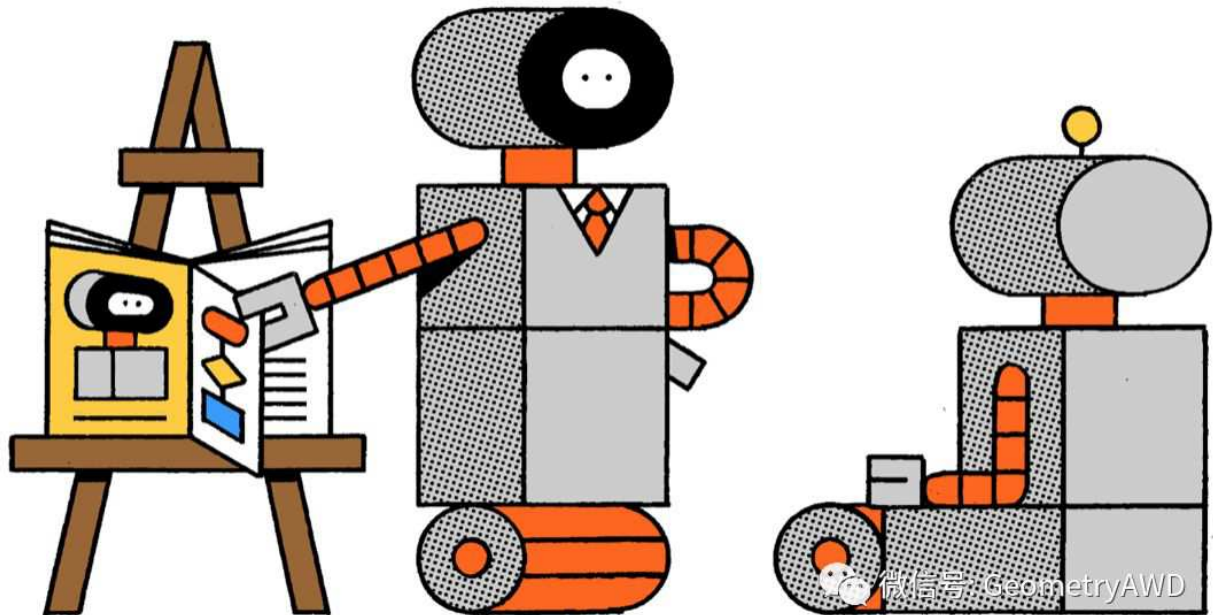
- Computers can be programmed to think (i.e. they **really** are thinking)

微信号: GeometryAWD

之所以称作弱人工智能，就在于它没有人类所具有的那种主观意向，因为他不能够完全自主地有意识地去完成某些事，但却具有一种物质形式的客观意向性。就好比怀孕的妇女在决定是否堕胎时，产科的超声波技术扮演了“参与到”决策当中的作用，最终决定是孕妇做出的，但超声波技术却以物质的形式间接参与了决策。机器智能的进步正积极地影响着它的使用者，并且改变着使用者感知世界的方式。**弱人工智只有被赋予了自由意志，和人类一样平等享有自主决策的能力和权利，才会成为强人工智能独立为新的物种。**



无人车「自动驾驶」的智能系统被构建成为具有应用领域特定问题自主学习能力，以深度学习为例，它在不断累积的“车辆和环境数据”与“人工干预”中持续锻炼“学习模型”（如交通灯车道线交通标示识别，障碍物行为预测，人机交互等子任务），最终使它拥有一定的自主性从而成为弱人工智能的典型代表。



无人车被允许从自己的“训练记忆”和“运行经验”中不断优化自己的“决策能力”，设计它的程序员与决定使用它的管理者引入了模糊的逻辑并不亲手设定策略，而是通过奖惩机制让机器慢慢掌握输入数据与输出行为之间的相互关系，于是有趣的事情发生了：

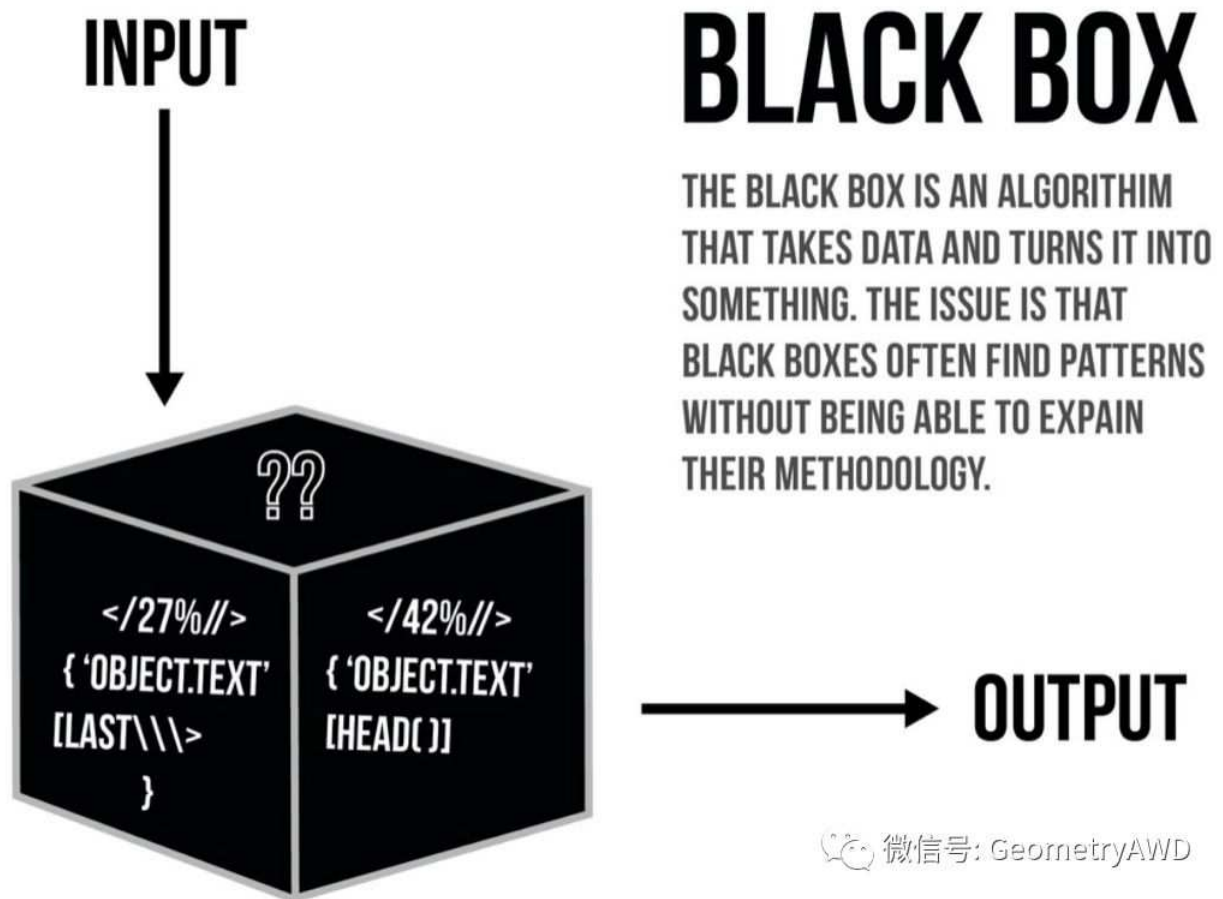
机器开始产生类似于人类思维的过程

人类程序员为机器预设的学习模型和训练数据，其实早已决定了无人车所做选择的倾向性：**机器的自主决策能力受制于先天训练水平和奖惩机制的选择**。^[4] 无人车作为弱智能，当面临两难的抉择时，自身并不具备完全独立的主观自主意识（机器的自由意志），却在千钧一发之间被放在了决策者的重要位置，无人车作为单纯的技术“物”不得不由此担负起人类的道德风险。

^[5] 在无人车这个技术实践的尝试里，主体性与客体性，人类与非人类共同交织在了一起。



弱人工智能采用的机器学习算法“黑箱”的特点，无形当中成为了智能“物”设计者和管理者推卸自己责任的借口，将安全风险转移给了使用者。



实际上，无人车不具有自主意识因此无法独立承载道德风险，用法律来惩戒没有自主意识的“物”没有任何实际意义，由此，技术便产生了道德真空。

对于无人车来说，机器学习技术的可解释性就代表着道德，道德研究的技术性工作就是要弄清训练数据对模型预测产生的影响，尊重生命的态度是不容质疑的，任何无人车致死事故的解决方案一定要最终落实到技术的实质进步上。

貳

「法律与道德，何去何从？」



自动驾驶，有亟待澄清的法律问题，比如：[6] “无人驾驶这项技术与私人属性的关系，这两者存在天然矛盾：“驾驶”是一辆车私有属性的唯一保证，如果没有这个概念，汽车的私人属性就不存在，汽车就变成了“公共”汽车。另外私有性质的自动驾驶的汽车（产权属于A）在自动驾驶过程中捎带上了乘客B，并在行驶过程中发生了事故并造成A和B的人身伤害，如何进行责任认定？A和B需要向谁索赔？这类问题是奥巴马一直所说要解决的法律问题之一，完全解除汽车私有属性才能消除其天然矛盾。”

无人车相关的法律得以建立精确的体系，是需要长时间高密度大强度的普通法，科学研究，媒体专业实践共同努力，经受诉讼双方利益的尖锐对抗考验后才可成型的。处于灰色地带的法律问题，会被控辩双方精确辩论并诉讼修正，最终固化为成熟的法规条文和司法实践。

对于任何争议首先要尊重法律，尊重法律才能求同存异，理性交锋。法律是斗争和博弈后的产物，是被前人证明有效的追求公平性的制衡。法律追求的从来都不是绝对正义，因为在公平原则下不存在绝对权威来定义什么是正义。

美国对待无人机自动驾驶技术的态度，正体现了这种制衡：

一边，是宪法给予公民的最大自由

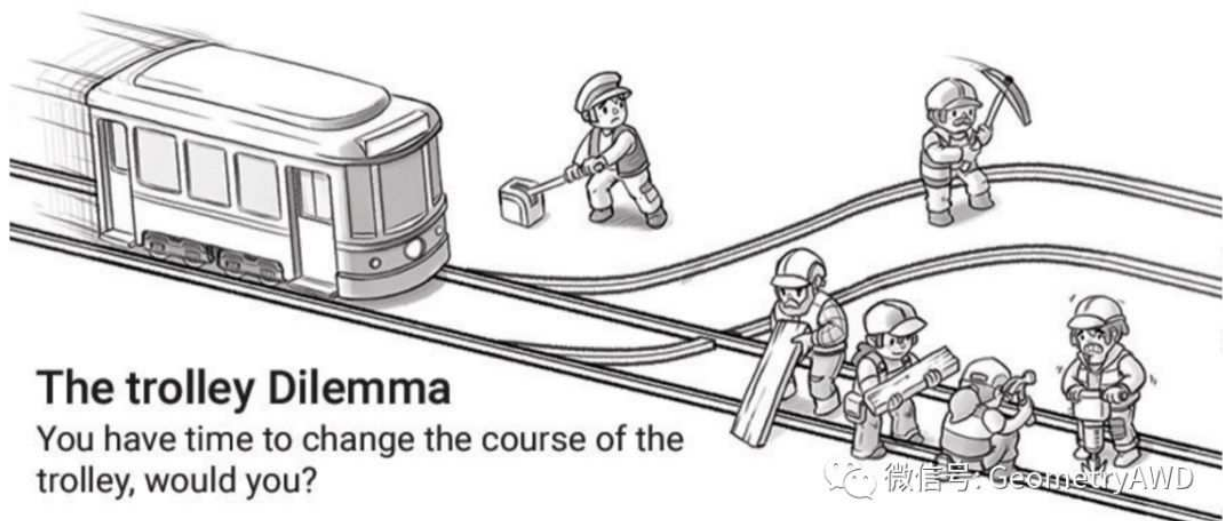
一边，是大法官在典型判例中始终把持的公平原则

新的技术因而在双方的制衡中在美国孕育。

FREE & FAIR

微信号: GeometryAWD

自动驾驶，更有亟待解决的道德问题，经典的“电车难题”里，^[7]功利主义和绝对主义在博弈。前者说：死一个比多死几个好，后者说：你的任何决定都直接导致了死亡，没有哪一个更好。



我们该怎么去理解道德行为？

康德曾经通过四个例子对什么是真正的道德行为加以解释：

- 做买卖童叟无欺（对他人的消极义务）
- 不放弃自己的生命（对自己的消极义务）
- 帮助他人（对他人的积极义务）
- 增进自己的幸福（对自己的积极义务）

Kant's Four Examples

	Violation Involves a Formal Contradiction	Violation Involves a Contradiction of the Will
Duty to Self	Duty to Refrain from Suicide	Duty to Develop Talents
Duty to Others	Duty Not to Make a Lying Promise	Duty to Render Aid

微信号: GeometryAWD

道德不是真理，道德是一种选择，可以是一种追求个人幸福的选择，可以是一种不触犯法律侵害他人利益的选择，可以是一种在不侵害他人利益的前提下，追求自身利益最大化的一种原始利己的动机。

《三体》小说里，程心是一个非常普通三观正得没法再正了的正常人，她在每个关键时刻做出的选择是每个正常人会做的选择，符合普世价值观和大众道德取向，但恰恰是这种道德模范一般的人的选择把人类推向了灭绝。

另一个有趣的有关道德行为的例子，德国大众柴油车尾气门，曾让众生意愤填膺，但大众柴油车在欧美二手车市场不但不降价，反而普遍涨价，柴油车成为了稀缺资源。开自己的车，既经济且马力大，就让别人去受污染吧。



康德认为，**一个行为的道德价值不是由结果赋予的，而是与行为的动机有关，即“善良意志”**。然而康德到底多大程度上将动机的善恶当作道德价值的评判标准，这还没有定论。

关于道德的准则有很多种，一如波斯纳大法官曾经说过的那样：**“除非简化为事实问题，否则，对有争议的道德问题，不存在令人信服的答案。”**

道德不存在“制高点”！

叁

「认知科学的奇袭」

回到**“电车难题”**，这种绝对抽象化的道德题目，是否适用于无人车这个具体的应用场景？道德，怎么才能有效指导实际应用呢？

借助神经磁震影像技术的帮助，认知心理学领域开展了一系列的实验，[8] 有研究表明，在电车难题里，当搬动转道器手柄的可能性被强调时，是大脑的理智区域在做反应，而当强调会造成牺牲的后果时，却是大脑的情感区域产生了反应。

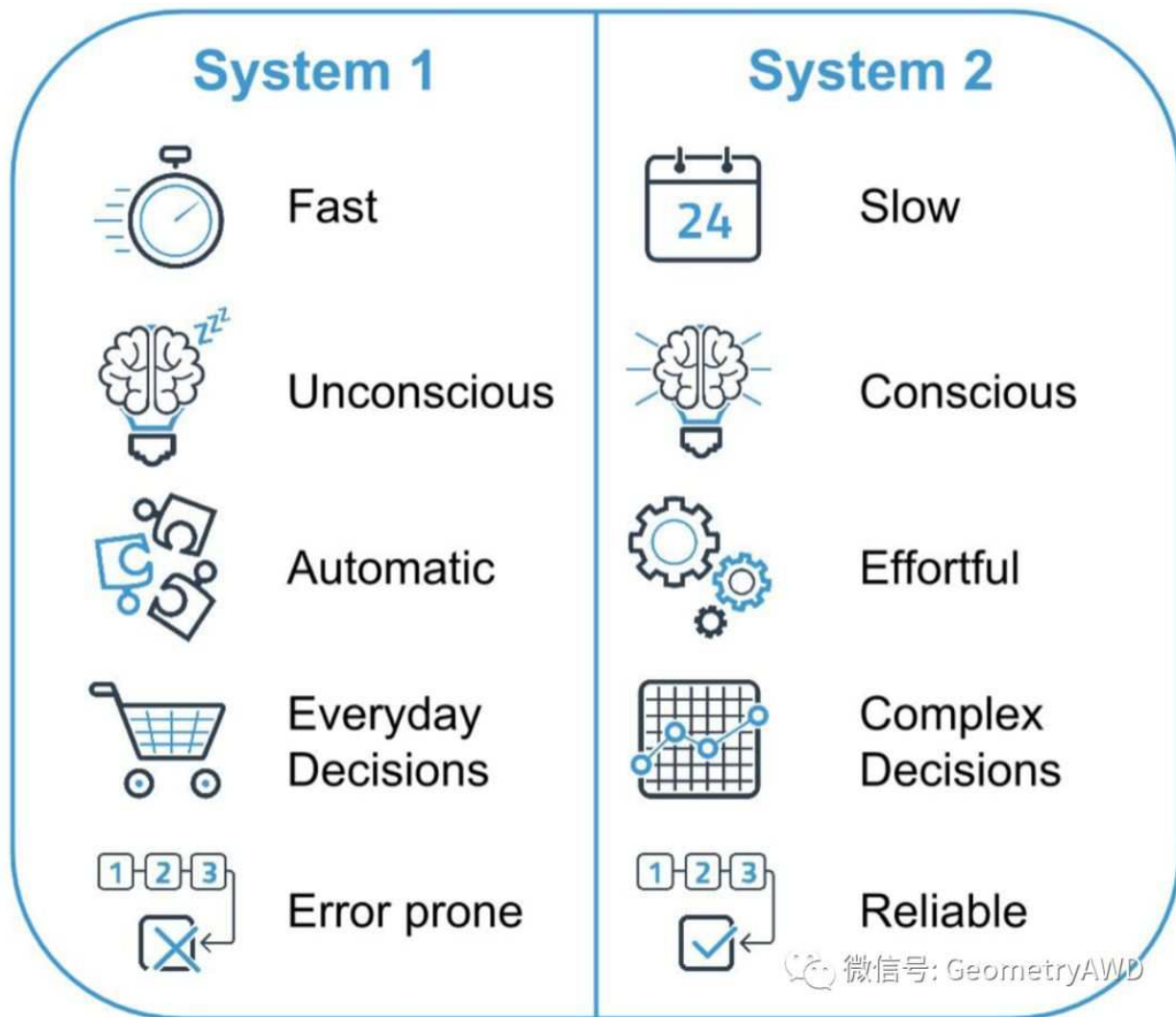


[9] 进一步的研究更探讨了电车难题里“道德决策”与人类出于进化的本能“排斥暴力行为”两者的关系，这种排斥感会激发人类特定的情感反馈机制，这一研究间接暗示了道德决策可能是一种低级的条件反射行为。

也就是说，直觉并非总是通过理性而判断的。

也许，在人类的伦理学中，并不存在直觉所依托的“基础原理”。

心理学家丹尼尔卡曼尼将人类大脑分为两个系统，依赖直觉的“行动者”和依赖理性的“思想者”，传统观念将道德看作理性的结果。



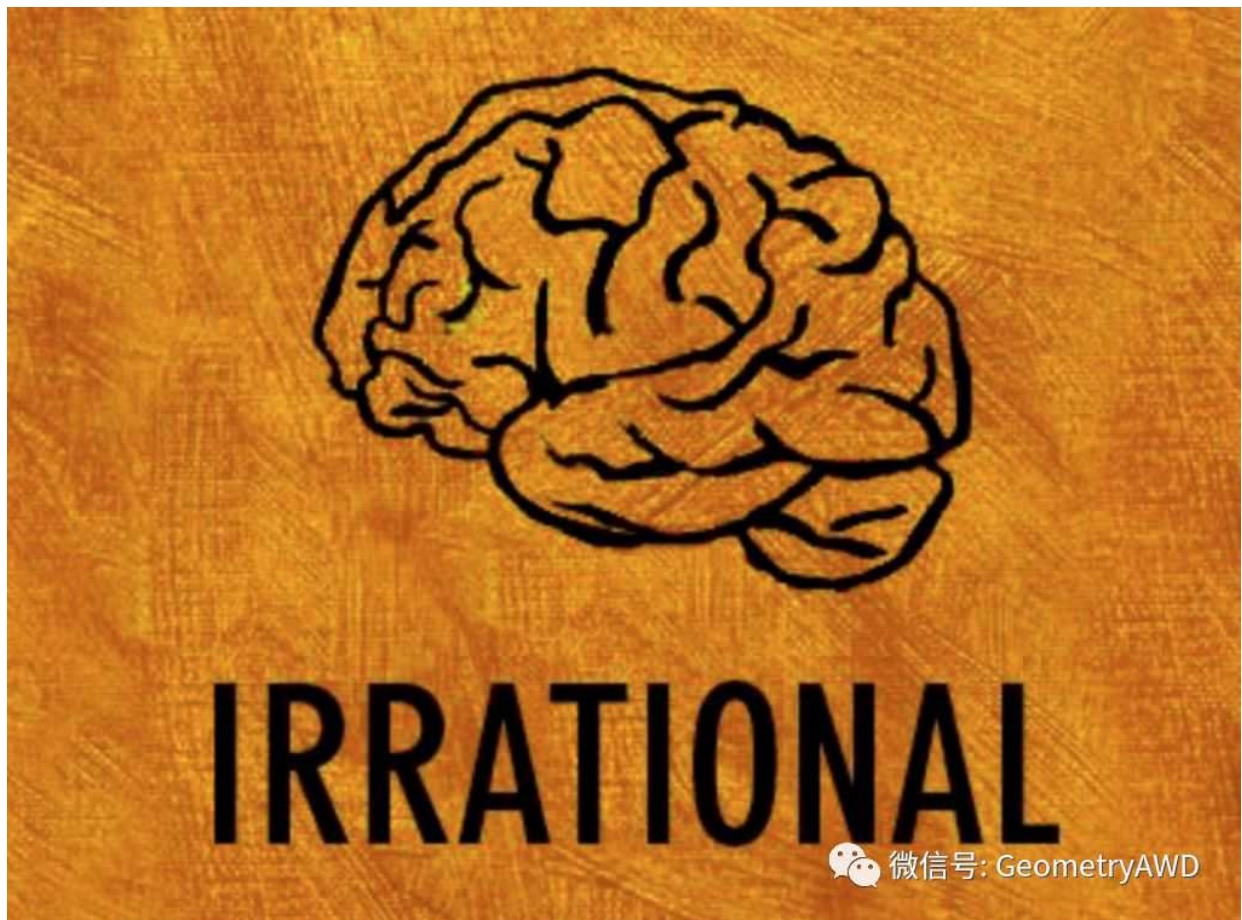
基于这个**行动者—思想者**的模型，[10] 认知科学家乔纳森海特通过实验发现认知常常是直觉部分的工作，理性部分更多参与“事后辩护”。“聪明并不会让你更加正确——它只会让你更好地说服自己和别人自己是正确的。”

The wise man must be wise before, not after, the event.

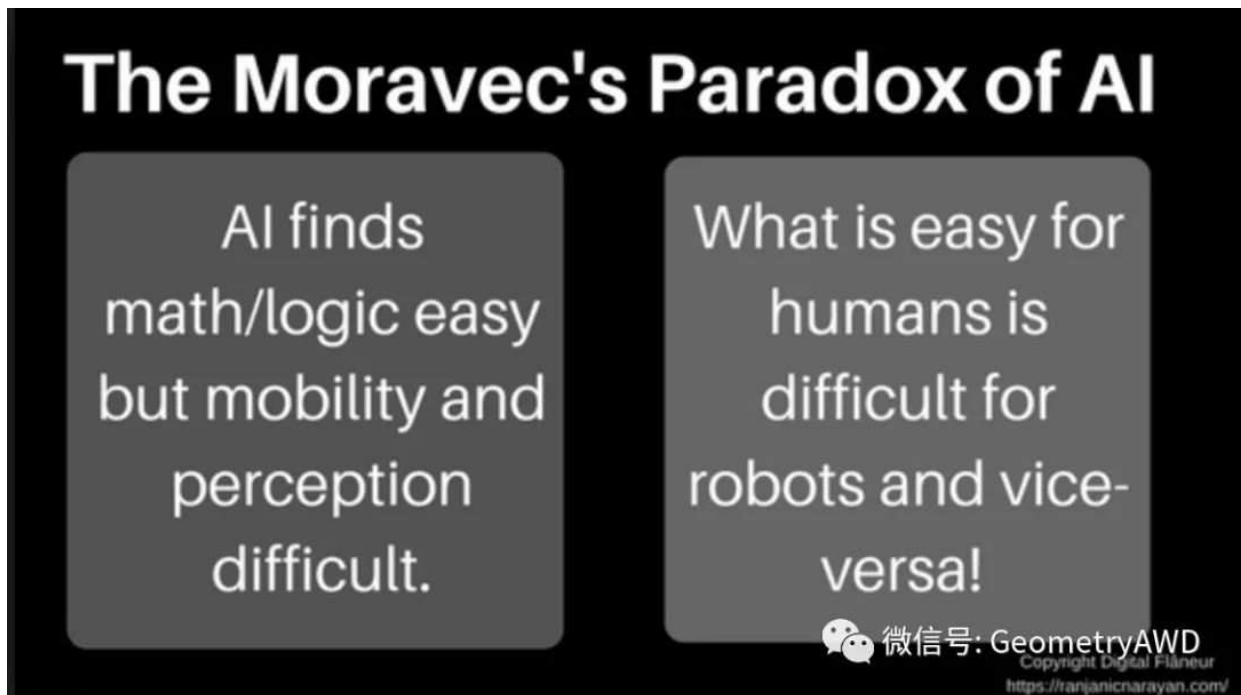
(Epicharmus)

微信号: GeometryAWD

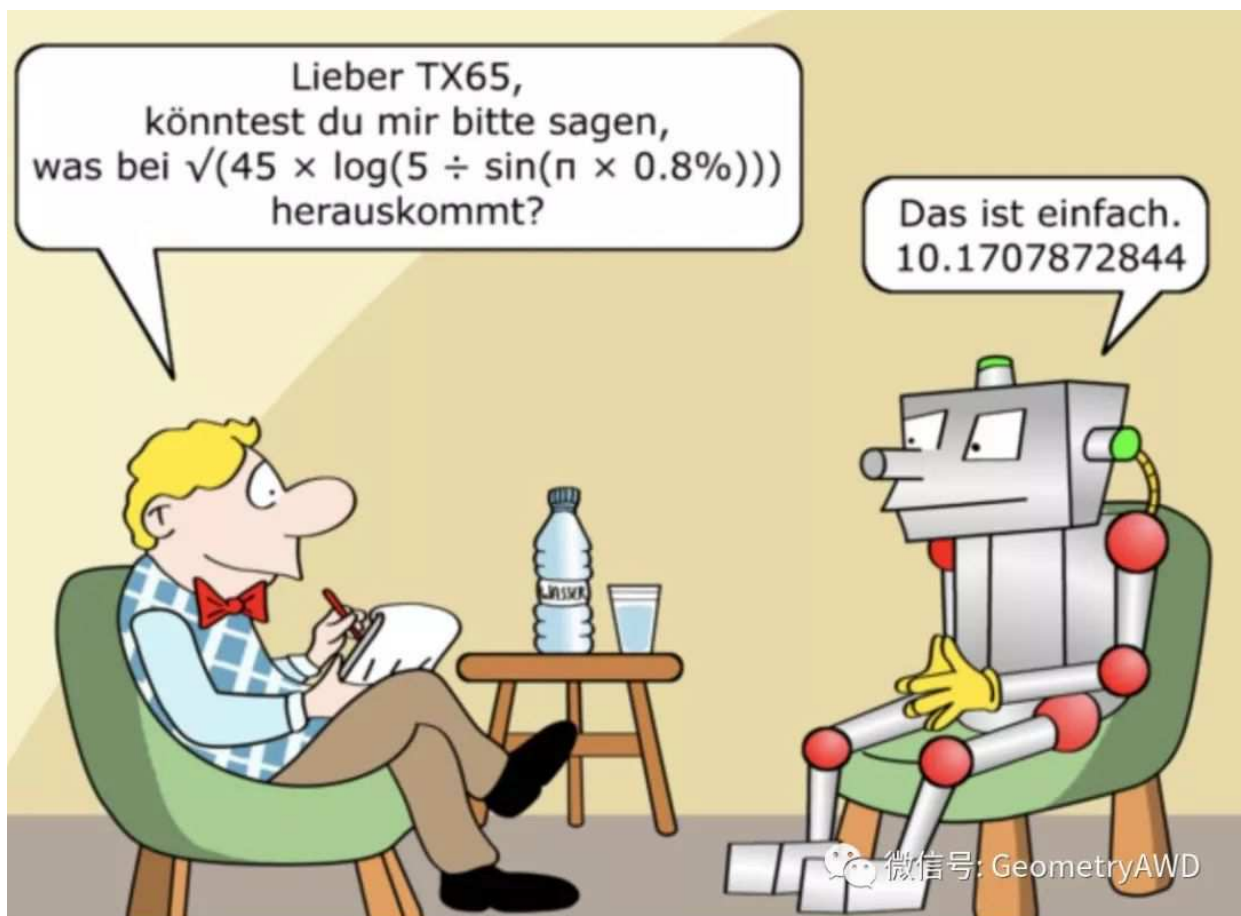
更进一步地，当乔纳森海特利用记忆数字的任务让理性的部分“饱和”时，再进行价值判断的速度并没有受到显著的影响，这也说明了人类做出的道德判断，基本上是诉诸直觉的，换句话说，人类的道德观念往往是非理性的。



非理性决策并不代表人类的低级，恰恰相反，在资源有限的情况下，人类能够从小数据迅速提炼归纳出规律，这正是人类非理性决策的先天优势，是人类思维能力的集中体现。莫桑维克悖论阐述了一个很有意思的原理，不管是人类还是机器，高级推理所需要的计算量不大，而低级的感觉和运动技能却需要庞大的计算力。



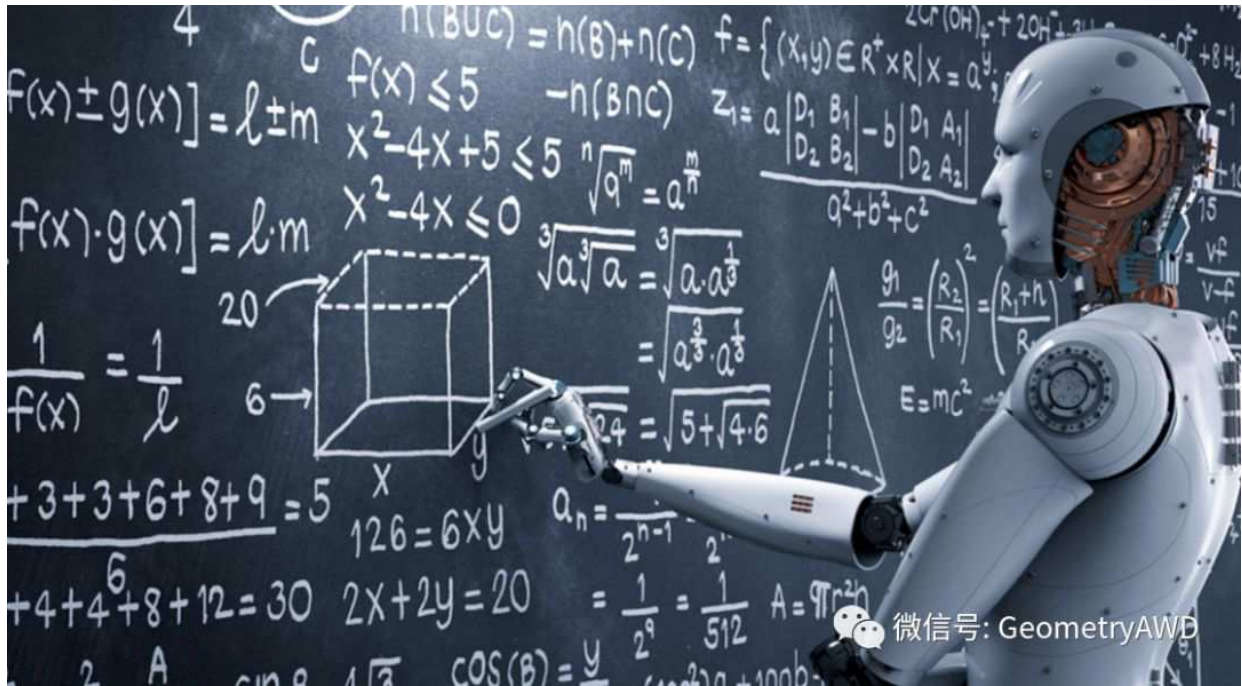
恰恰正因为如此，人类更应该本着务实的态度，将无人车当作人类四肢和眼睛的物理延伸来支配，释放脑力将感觉、运动、计算的底层计算力交给更厉害的无人车。



翻译：“亲爱的TX65，你是否可以告诉我下面这个方程式的解是什么？” “很简单，10.1707872844”

但是，如果人类不负责任地将「伦理决策」也交给无人车并将其看作人类道德的转嫁物，我们的目的就变成了可怕的“机器取代人类”，而不仅仅是“机器取代人类不擅长的工作”。

我始终坚信，“懒惰”的代价不应该是“牺牲道德”。



肆

「通向实在之路」

人类不应该对没有道德判断能力的弱人工智能机器挥舞道德的大棒，苛求它在实际应用之前达到极端完美。面对危机的一刹那，人类驾驶者往往会依赖本能的反应，而不会去主动考虑“**电车难题**”里的道德哲学问题。

在面对道德难题时，作为道德僵尸的弱人工智能无法抉择，弱人工智能没有自由意志成为道德的承载物，因此创造并使用它的人最应该受到道德约束。



与其追求一个100%完全正确的驾驶机器，人类不如更现实一点，期望机器在对自己的答案不确定的时候，机器能够明确向人类表达出它的局限性，并由技术开发者来承担机器“无知”而带来的法律后果。

“我不知道” 比一个错误答案更显珍贵。



人类对于自动驾驶技术的要求比对人类自己更苛刻，这无可厚非，无人车的感知范围，计算力，反应速度，失效率无一不领先于人类。在人类出行最大的两个痛点交通事故和交通拥堵上，无人车都给出了更优的解决方案。



自动驾驶道德问题的出路就是尽快地普及这项技术，市场竞争和司法实践会源源不断地为推进这项技术走向成熟保驾护航，道德审判只会带来纷争和无解的局面。

在技术还未成熟的时候，就考虑给没有道德判断力的机器强力编程实现特定的道德束缚，会人为地引入“默许选择性牺牲”的预设立场，这是舍本逐末。

发展新的技术寻找新的途径，包容解决未知问题的各种可能性，才是正确的道路。



只要自动驾驶的实际性能超越了统计意义上人类司机的平均水平，科技价值能够得以体现，这项技术就是值得推广的；顾虑无人车安全的人仍然可以保留自己开车选择的权利，以确保个人最大自由；无人车交通事故的在公平原则的审视下，法律建设日臻完善；道德建设上，人类已经开始致力于探索廉价而正义的人工智能，确保人工智能不会被别有用心的人用于邪恶的目的。

科幻作家阿西莫夫一生的最高杰作：1955年写成的《永恒的终结》，看到最后一页才会明白，只有经过严酷的考验，人类才能不断前进，走向发展的高峰。生存危机感，是驱使人类不断进步、不断征服新事物的根本动力。如果因为道德的原因而畏惧新的技术，人类便会选择最安全最中庸的道路前进，征服宇宙和群星就会变成遥不可及的幻梦。当人类没有了可以征服的重大难题和苦难，就失去了发挥创新，探索未知的能力，而自掘坟墓。

无论技术多强大，最终受益的或受害的，还是它的发明和使用者 - 人类！



延伸阅读 / 引文出处

- [1] 无人车的历史 Wikipedia: Autonomous car https://en.wikipedia.org/wiki/Autonomous_car#History
- [2] 2006年—深度学习的起点 <http://book.51cto.com/art/201710/554969.htm>
- [3] weak AI https://en.wikipedia.org/wiki/Weak_AI
- [4] 技术物是道德的行动者吗？维贝克“技术道德化”思想及其内在困境 东北大学学报 2017年第3期 刘铮
- [5] 费尔巴哈《未来哲学原理》及《哲学改造的临时纲要》中对人的主体性和客体性作出了多方面的规定和说明。
- [6] 网络文章《我为什么对自动驾驶说不》 作者已经不可考
- [7] 电车难题最早是由哲学家菲利帕·福特（Philippa Foot）于1967年发表的《堕胎问题和教条双重影响》论文中提出来的，用来批判伦理哲学中的主要理论，特别是功利主义。功利主义提出的观点是，大部分道德决策都是根据“为最多的人提供最大的利益”的原则做出的。详见哈佛大学哲学课对电车难题的讲解！
- [8] Greene, Joshua D.; Sommerville, R. Brian; Nystrom, Leigh E.; Darley, John M.; Cohen, Jonathan D. (2001-09-14). "An fMRI Investigation of Emotional Engagement in Moral Judgment". Science.
- [9] 《伦理学思想实验中道德直觉的规范性效力问题》 李裕晨
- [10] Emotions Rule: How Jonathan Haidt Can Help Us Have Better Relationships <https://collegeinfo geek.com/emotions-rule-jonathan-haidt/>

完

特邀撰稿：徐鸿鹄

作者微信：honghu967935

现就职于 **蒂森克虏伯 普利斯坦** (Thyssenkrupp Presta) ，负责电动转向系统 (EPS) 应用项目的开发和管理。

航天科班，主业汽车电动转向开发，纯正懒散东北人，王小波脑残粉，个人主义走狗，AI威胁论煽动者，民科数学家，半吊子程序狗，沉迷理论，日渐消瘦，无一精通。

更多文章推荐

- [「特斯拉的核心工作原理」十年猥琐发育，用十分钟讲给你听！](#)
- [你知道他们口中的「L3自动驾驶」到底是什么意思么？](#)
- [新能源汽车的前世今生！资本游戏？媒体噱头？还是有真实力？](#)
- [他们口中的新能源「电动车」真的环保么！？](#)
- [# 2017 # 日内瓦国际车展\(Geneva Motorshow\)](#)
- [「TOP100」世界汽车零部件供应商百强榜及2016年调研报告](#)
- [「四驱攻略」之 你对世界级的全球著名五大车展都了解多少？](#)

关注「几何四驱」

各种高能文章，随后速速就来

版权申明

本篇文章内容为作者本人原创，展示图片均来自网络，版权归图片作者本人所有！文章未经 **「几何四驱」** 或 **「作者本人」** 授权不得进行商业性转载。个人性转载请务必注明文章出处。原创不易，感谢理解！

不跟风，拒绝平庸，只做最好的自己。

「几何四驱」 All made in Germany , Since 2016 !

长按二维码关注
「几何四驱」



All made in Germany since 2016!



[Read more](#)