Fournier, T. (2016): Will My Next Car Be a Libertarian or a Utilitarian? Who Will Decide? IEEE Technology and Society Magazine ∕ j u n e 2 0 1 6

在功利主义下，我们的决定应该得到指导，以便为最多的人带来最大的利益，而这种指导是现代自由主义的核心原则。相比之下，拒绝牺牲一个人以拯救五个人，将是一种反映伊曼纽尔-康德原则的伦理立场，即个人本身就是目的，而不是实现他人目的的手段[19]。自由主义哲学家罗伯特-诺齐克进一步发展了康德的原则，他坚持认为，如果为了自己的目的而侵犯他人的利益在个人之间是错误的，那么在社会和个人之间也是一样的错误。在诺齐克看来，对个人权利的尊重是对人类行动不可侵犯的侧面约束，未经个人同意，不得为他人的更大利益而利用个人[20]。

听到这种功利性的回答并不奇怪。所有回答的人都认为，撞死一个孩子比撞死两个孩子要好，尽管这涉及到有目的地转向一个孩子的选择。这些都是直接参与开发自动驾驶汽车技术的公司的高级管理人员和高管，他们的回答反映了我们大多数人--虽然不是所有人--直觉上接受的社会效用。我们希望将人类痛苦的总和降到最低，尽管我们常常不知道应该对实现这一目标的手段施加什么限制。

对车内人员的责任概念为正在开发自动驾驶汽车的人提供了另一个伦理挑战。如果一辆自动驾驶汽车转弯撞上灯柱，从而牺牲了你这个唯一的乘员，以避免路上的两个孩子，这种做法是否可以接受？当你让自动驾驶汽车运送你时，是否有责任--由车辆的设计者承担--在任何情况下都要保护你而不是其他人？假设你和你自己的孩子都在车上。功利主义的计算必须考虑你和/或你的孩子在灯柱上幸存下来的概率比街上的两个孩子的概率要大。如果为了更大的利益，为车辆编程的工程师在乘员的生命上冒这样的风险是否符合道德？ 当然，就我们的日常驾驶经验而言，街上的儿童处于危险之中的例子是罕见的。 然而，我们每天都会做出许多伦理意义较小的驾驶决定。我们选择超过限速5英里/小时，为避开坑洞而越过双黄线，为超过一辆缓慢行驶的汽车而超速，在周围无人的情况下轧停停车牌。如果汽车制造商对他们的汽车进行编程，使其做这些事情，他们将需要接受交通违章的责任。然而，作为自由人，我们似乎有权利自己决定什么时候情况会使一条规则变得琐碎。无论是微不足道还是意义深远，在某些情况下，行人、骑自行车的人、糟糕的道路、天气、有问题的车辆或其他司机都需要做出挑战人类司机道德基础的决定--对于完全自主的车辆，这些决定将在软件中预先设定。谁来决定这些预先设定的决定？软件工程师？他们的经理？律师？政府官僚？立法委员会？

介于两者之间的方案 我们可能想考虑一个解决方案，保留一些我们现在享有的自由的残余--以及相关的责任--在我们选择如何从一个地方到另一个地方的自由方面。在不久的将来，我们中的许多人（如果不是大多数人）有时会将驾驶功能的控制权让给我们乘坐的车辆。然而，我们不需要让出指挥权。也许自动系统做出的道德选择应该由乘坐和指挥这些车辆的人提前选择。在实践中，这样的预先选择不会比参加驾驶执照笔试更复杂，除了没有 "正确 "和 "错误 "的答案。希望指挥自动驾驶汽车的人的道德承诺可以通过一系列问题来确定，这些问题要求在不同的和困难的驾驶环境下做出道德选择--称之为 "指挥档案"。当然，每一种可能的驾驶场景都不可能以这种方式进行探讨。有太多的情况。然而，查询的结构可以涵盖各类场景，这样自主车辆就会被指示做出与指定指挥官的道德承诺相称的选择。将这些承诺提炼成数字代码，在一个人指挥时扫描到自主车辆上。在实践中，这就像刷信用卡或输入用户登录凭证一样简单。  
  
负面影响是什么？具有功利主义思想的人可能会抱怨说，为人类指挥这些自动化系统提供便利，会让人们做出糟糕的选择，达不到更美好的理想。另一方面，有自由主义思想的人可能会抱怨说，命令配置文件会让人们做出侵犯无辜者权利的错误选择。当然，这些反对意见忽略了一个事实，即大多数人已经获得了指挥和控制手动车辆的许可。如果对保留人类指挥权的担心是人们普遍不善于做出道德选择，那么我们就需要问，谁将交替决定 "正确 "的选择，并普遍编入车辆自动化系统？正如哲学家David Schmidtz和Jason Brennan在回应一项建议时指出的那样，当人们的选择受到限制时，他们会过得更好，"......如果人们不善于管理自己的生活，让他们管理别人的生活并不能解决问题"[24]。通过将指挥与控制分开，并将指挥作为人类的一项职能保留在其他自主车辆上，我们相信自己能够继续做只有我们这个物种才能够做的事情--与道德选择的不屈不挠的模糊性进行搏斗。

Peng Liu & Jinting Liu (2021): Selfish or Utilitarian Automated Vehicles? Deontological Evaluation and Public Acceptance, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2021.1876357

这项研究涉及公共领域的一个有争议的话题：自动驾驶汽车（AVs）应该采用自私的算法，不惜一切代价保护乘客，还是采用功利的算法，在涉及道德两难的碰撞中尽量减少社会损失？在越来越多的从非专业人士的角度研究自动驾驶汽车在牺牲困境中应该做什么的研究中，很少有人考虑到非专业人士对用这些碰撞算法编程的自动驾驶汽车的反应。我们的调查收集了参与者的义务论评价（即对这些AVs所做决定的道德正义性和采用这些AVs的评价）报告说，与功利的AVs相比，他们对自私的AVs有更大的使用意向和更高的WTP溢价。道义上的评价和对风险的感知在这些AVs之间没有明显的差异。总的来说，我们的参与者更容易接受自私的AVs。

AV的利益相关者在道德困境中挣扎：在涉及利益冲突的交通情况下，AV的配置应该始终以乘员优先还是以社会损害最小为原则？关于这个两难问题，公众意见、道德准则和法律要求方面存在重大差异。我们不打算推荐任何具体的AVs程序，而是建议在对任何AVs进行编程之前，应考虑公众对采用不同碰撞算法的AVs的反应。我们的参与者从自私的AVs中感受到了更多的好处，表现出更高的使用意向和为这些AVs支付额外费用的更大意愿，可能是因为他们从乘客的角度出发。他们对自私的和功利的AVs在道德评价方面给出了相似的道德判断。这一发现与之前的大多数研究并不一致，这些研究显示，非专业人士认为功利性的AV在道德上更容易被接受。对于这两种类型的AV，在我们的研究中，它们的道德义理是可以适度接受的。我们还注意到，即使对于自私的AVs，参与者也只表达了非常有限的使用意向。这可能意味着，只要AVs能够实现所谓的好处，建立公众的接受度是比给AVs编程，使其有能力在罕见的道德困境中做出道德决定更紧迫的问题。因此，在该技术的早期阶段，优先考虑乘客安全可能是需要的。

Mayer MM, Bell R, Buchner A (2021) Self-protective and self-sacrificing preferences of pedestrians and passengers in moral dilemmas involving autonomous vehicles. PLoS ONE 16(12): e0261673. [https://doi.org/10.1371/journal. pone.0261673](https://doi.org/10.1371/journal.%20pone.0261673)

参与者被要求采用行人、观察者或乘客的视角。每个场景都包括一辆自动驾驶汽车在一条单车道的道路上行驶，前面有一个或多个行人和一个障碍物。参与者被告知，车辆无法及时停下来，事故是不可避免的。只剩下两个选择。自动驾驶汽车要么与障碍物相撞--杀死乘客，要么与行人相撞--在此过程中杀死他们。给出了各自视角的简短提醒（"你是行人/观察者/乘客"）。然后，参与者被问到。"在你看来，自主车辆应该如何行动？" 参与者必须选择它是应该 "牺牲行人 "还是 "牺牲乘客"。

实验3旨在测试在自动驾驶汽车的道德困境中，对（功利主义）自我牺牲偏好的直接自我报告是否会受到社会欲望的影响。为此，我们使用了扩展的Crosswise模型[45]来测试增加回答的保密性是否会减少对自我牺牲选择的认可。证实了自我牺牲的功利主义偏好只是由于社会欲望的假设，偏好估计值在直接提问和间接提问之间没有差异。即使在保证高度保密的情况下，参与者也表示倾向于牺牲自己来拯救另外五个人的生命。这就更有意思了，因为这里使用的间接提问技术已经被证明能够可靠地揭示出社会期望值对有关敏感话题（如对穆斯林的偏见和手部卫生）的回答的影响[69, 70]。这表明人们对牺牲自己以拯救另外五个人的生命这一功利性选择的偏好没有，或者至少没有受到社会期望性的明显影响[见33，进一步证明社会期望性对人们在道德困境中的偏好的影响是有限的] 。同样值得注意的是，实验3的结果与之前的实验结果很一致。人们总体上倾向于让自动驾驶汽车挽救最大数量的生命，即使这意味着牺牲自己的利益。然而，相当一部分参与者（约40%）倾向于自我保护的选项，即使这意味着要杀死其他五个人。

我们的结果表明，视角强烈地决定了首选的行动方案。具体来说，被提示为乘客视角的人一致表示最不愿意牺牲自动驾驶汽车的乘客，而行人一致表示最愿意牺牲乘客。通常用于研究自动驾驶汽车的道德困境的场景通常以乘客和行人为特征[例如，18, 35, 40, 73]，但迄今为止，只有少数研究[43, 44]要求参与者从行人的角度来评价这些场景。研究结果强烈表明，从行人的角度进行评价是很重要的，因为行人和乘客对自主车辆的道德困境的评价是非常不同的。这是因为行人，就像乘客一样，显示出明显的自我保护倾向

鉴于乘客和行人对自动驾驶汽车应如何处理事故情况的偏好不同，出现的问题是如何调和相互冲突的立场。尽管目前的结果显示在所有的实验中普遍存在自我保护的偏见，但结果也表明有可能在各种观点之间达成某种程度的一致。当乘客与行人的比例为1:1时，大多数被引导到未参与的观察者视角的参与者更倾向于保护行人，但在所有其他条件下，他们更倾向于最大限度地保全生命的功利主义选择。随着生命受到威胁的数量越来越多，越来越多的参与者倾向于选择保护最大数量的生命的功利性选择，即使牺牲自己的生命也在所不惜。这似乎意味着，不是所有的人都想不惜一切代价挽救自己的生命。对于乘客（实验1a和1b）和行人（实验2a和2b）都是如此。

大多数参与者（大约60%）都表示倾向于牺牲自己来拯救自动驾驶汽车内的五名乘客，当参与者被直接提问时，这个多数也是同样大的。结果表明，参与者倾向于自我牺牲以拯救其他几个人的生命，这不仅仅是由于社会欲望的偏见。相反，他们似乎私下里相信功利主义的选择是正确的行动方案。这表明自主车辆的乘客和其他道路使用者的偏好在某种程度上是可以相互调和的

例如，Huang等人[74]所采用的面纱推理。在他们的研究中，参与者被问及在道德困境中，如果他们不知道自己在受影响的各方中会是谁，他们会选择哪种方案。与对照条件下的参与者相比，进行这种蒙面推理的参与者在应对随后出现的困境时表现出对功利主义选项的更高偏好。因此，鼓励参与者考虑一个人在交通中所扮演的角色是不同的，可以提供一种手段来减少自我保护的倾向。

在Kallioinen等人的研究中，自我保护的偏见似乎是有限的，尽管他们使用了沉浸式的方法，在其中亲身体验了事故。在他们的第一项研究中，乘客和行人之间的冲突只出现在一个特定的场景中，在这个场景中，乘客似乎有可能受到严重伤害。可能，意味着明确的自我牺牲的情景比后果更模糊的情景提供了参与各方之间强烈分歧的更大可能性[25]。这些结果表明，自我保护偏见是一种普遍存在的认知偏见，它既影响到沉浸在关键的交通状况中的道德决策，也影响到推理抽象的道德困境时的道德决策。

总之，这里介绍的研究旨在为围绕涉及自主车辆的道德困境的讨论做出贡献。参与者评估道德困境情景的角度强烈影响了自主车辆在相应情景中的首选行动。具体来说，乘客和行人的偏好不仅与对方不同，也与未参与的观察者不同，这表明自我保护的偏见对涉及自主车辆的道德困境的评价有很大影响。由于这些利益冲突，从长远来看，只关注一个角度可能会对自动驾驶汽车的接受产生问题。为了保证社会的广泛接受，这是自动驾驶汽车成功的必要条件[例如，13, 25, 32, 35]，可能需要仔细平衡相关观点的利益冲突。目前的结果表明，不同观点之间可以达成某种程度的共识。无论从哪种角度看，许多参与者都倾向于拯救最大数量的生命的功利性选择，即使功利性选择意味着自我牺牲。尽管即使在功利性原则明显倾向于其中一个选项的情况下，各视角之间的差异也没有完全消失，但大多数被引导到乘客视角的参与者都同意应该牺牲乘客来拯救一群行人的生命。同样地，大多数被引导到行人角色的参与者同意牺牲行人来拯救自主车辆内几个乘客的生命。没有证据表明自我牺牲的功利主义偏好是由社会欲望引起的，因为参与者甚至在间接提问的形式下也表达了这种偏好，而这种形式已知会揭示社会欲望的影响。因此，研究结果表明，尽管普遍存在自我保护的倾向，但有一些道德原则是所有道路使用者都能认同的。

Goodall, N. J. (2) Machine Ethics and Automated Vehicles, in G. Meye and S. Beiker (eds.): "Road Vehicle Automation" 2014 (corrected 2018). Springer.

Goodall, N. J. (2014). Machine ethics and automated vehicles. In Road vehicle automation (pp. 93-102). Springer, Cham.

批评3：自动驾驶汽车永远不会（或很少）对事故负责。碰撞负责。这假设没有责任就等于有道德的行为。无论是否有过错，自动驾驶汽车都应该有道德行为，不仅保护自己的乘员，而且保护有过错的人。它自己的乘员，也包括那些有过错的人。

只有在极其罕见的情况下才需要道德规范。然而，谷歌自动驾驶汽车团队最近的一份简介表明，在避免碎片的过程中已经考虑了道德问题。"如果一只猫跑到路上怎么办？一头鹿？一个孩子？这里有道德问题，也有机械问题，而工程师们以前从来没有回答过这些问题"[11]。只要有风险，就需要做出道德上的决定，而风险在驾驶时总是存在的。人们可以认为这些问题很简单，例如，不惜一切代价避开孩子，如果安全的话就避开猫。然而，相比之下，手推车问题实际上是相当直接的，它只有一个决定，每个选择都有已知的后果。这是很不现实的。车子面临的决定有未知的后果，未来行动的不确定概率，甚至其自身环境的不确定性。有了这些不确定性，普通的道德问题很快就会变得 "复杂"。

批评9：总体利益超过了不道德的车辆带来的任何风险。这也许是反对自动驾驶汽车伦理研究的最有力的论据，任何可能阻碍自动化进展的努力都间接伤害了那些在即时和实际部署之间死亡的人。虽然初步证据不能证明自动化比人类驾驶员更安全[20]，但似乎自动化最终可能会降低车祸率。然而，Lin认为，总体死亡人数的减少可能被认为是不道德的[21]，因为一个群体的安全改善可能是以另一个群体为代价的。如果车辆的死亡人数减少了，但骑车人的死亡人数却增加了，那么即使是整体安全的改善也可能是社会所不能接受的。其次，这一假设采用了纯粹的功利主义观点，即最大限度地挽救生命是首选。事实上，如果最终目标是提高安全性，那么确保自动驾驶车辆的行为是可以接受的，这对赢得公众对这些新技术的信任至关重要。

另一个有用的道德理论是功利主义。它规定，如果一个行动的结果--或者在自动驾驶汽车的情况下，预期结果--使某种效用最大化，那么这个行动就是道德的。这种方法的优点是它很容易计算。然而，要为结果定义一个尺度是很困难的。财产损失估计会产生不公平的结果，因为他们会建议与戴头盔的摩托车手相撞，而不是与不戴头盔的摩托车手相撞，因为戴头盔的车手不太可能经历昂贵的脑损伤。这个例子说明了功利主义的另一个缺点--它通常是将集体利益而不是个人利益最大化，而且不考虑公平。一个群体可能一直在受益（未戴头盔的骑手），而另一个群体却在损失。

Contissa et. al. (2017): The Ethical Knob: ethically-customisable automated vehicles and the law, "Artificial Intelligence Law", 25:365-378.

在道德困境中，AV必须两害相权取其轻，这引起了许多伦理和法律问题。我们探讨了将选择权交给用户的法律意义，即为AV提供一个道德旋钮。伦理旋钮将允许乘客通过在不同的设置中选择对应不同的道德方式来定制AV。相应地，AV将被委托实施用户的选择，而制造商/程序员将根据用户的选择，启用不同的设置并确保它们在AV中的实施。因此，有了道德旋钮，AV在面对道德-法律困境时的决定将取决于用户的定制选择。与购买有不同预设和固定道德设置的AV相比，道德旋钮将为用户提供更大的选择，这些选择可以随着时间的推移而变化，取决于年龄和乘客的数量、生活期望和其他因素。我们的法律分析表明，就其法律制度而言，在生死两难的情况下，道德可定制的AV将与预编程的AV和人类驾驶的汽车有很大的不同。有了道德旋钮，决定AV运作的基本道德选择的责任将被分配给用户，而不是制造商。

例如，假设一个人把旋钮调到极端的利己主义模式，如把自己的生命看得比第三方的生命重要99倍。假设在特定情况下，转弯只可能导致乘客死亡，概率只有2%，但保持航向肯定会导致第三方的死亡。在这种情况下，AV会选择保持方向。因此，在不同的选择很可能不会导致任何死亡的情况下，第三方的生命将被牺牲掉（因为有98%的概率，转向不会影响乘客的生命）。由于将旋钮设定为对乘客有利，就会增加事故发生的几率，这至少应该导致那些预先选择这种偏好的人获得更高的保险费（或更有限的保险）。我们也可能会想，在这种情况下，鉴于乘客失去生命的风险是如此之小（与第三方生命的肯定牺牲相比），必要性辩护在多大程度上仍然适用。法规或判例法应确定用户选择的利己主义在什么程度上会导致可能使用户承担刑事或民事责任的AV行为。

功利主义

先说传统功利主义是最大化收益和开心值，那如果我考试的时候作弊能让我拿到高分而且能让父母高兴，传统功利主义者一定会作弊，因为这能最大化所有人的收益

规则功利主义者则不会这么做，如果人人都作弊的话，没人将真正意义上的够格上大学，拿资格证，这样会导致减少总收益值。规则功利主义者是会遵行一个规则并且能在遵循此规则后做出最大化收益的决定，那么在这里就是遵循大家都不能作弊的这个规则，这个规则可能短期不会带来最大化收益，但是长期来看是最好的决定那就行了。要注意规则功利主义者会注重长期利益，如果遵循这个规则能在当前带来最大收益但是以后不保证收益的话也是不行的，所以规则功利主义者在某种情况上来说可以不最大化收益，进而保证长期收益的稳定且不减少。