

订阅DeepL Pro以编辑此演示文稿。  
访问[www.DeepL.com/pro](https://www.deepl.com/pro?cta=edit-document)，了解更多信息。

**选择行为或规则功利主义来激励自主车辆的防撞协议？一个分析性评论。**

**梁子怡**

**新南威尔士大学**

如今，自动驾驶汽车（AVs）还不是一项成熟的技术，存在着相当大的争议；没有人知道是否应该把人类的生命安全交给机器。然而，在未来几十年，完全自动驾驶汽车的市场潜力相当突出。完全自动驾驶汽车将在提高公路安全、缓解交通拥堵和减少空气污染等领域带来颠覆性的改进。AVs不是单一的新技术，而是各种技术的整合。它不仅需要硬件开发，更重要的是，工程师应该如何将人类社会的道德标准和规则设计到算法中。也就是说，在遇到意外情况时，AV应该如何做出合理的判断。伦理因素 "对于在司机、行人和公共财产之间分配风险至关重要。对于工程师和公众来说，AVs的决策系统必须在每一个决策中考虑人类的伦理。当面临模糊的道德层面时，我们一般会专注于遵守法律，同时尽量减少具体后果造成的损失。这种策略更受欢迎，因为它是基于法律规定的决策，而开发者几乎没有努力去解释AV的行为。这样一来，定义道德行为的责任就转移到了立法者身上。然而，法律一般不会对 "在紧急情况下该怎么做 "作出具体指示，尤其是在仅有的两种选择都危及生命的情况下。在这样一个特殊的情况下，要由机器来决定哪种选择会产生最好的结果。因此，本文将讨论选择不同的功利主义来激励自动驾驶汽车的防撞协议会如何导致不同的协议，以及为什么两种功利主义的激励协议看起来都有缺陷。首先将介绍行为功利主义和规则功利主义的细节和它们的区别，然后比较用它们来指导避免碰撞协议的后果，最后，将给出为什么两种功利主义都不适合的理由。

边沁（1789）认为，幸福或快乐是同质的，因此可以在人们之间进行比较和总结。因此，功利主义的基本思想是，道德行为或制度应该促进 "最大多数人的最大幸福，这是衡量正确与错误的标准"。它的一个比较突出的特点是，允许并提倡牺牲个人幸福甚至生命来增加集体幸福。例如，手推车问题就是一个典型的道德困境的例子，牺牲一个人救五个人是功利主义所提倡的举措。总而言之，功利主义是以利益最大化为基础的，在这个过程中谁的利益会受到损害，并不是它考虑的因素。用边沁自己的话说，功利主义原则就是根据事情增加或减少利益相关者的快乐的倾向来表达对任何行动的赞同或不赞同，也就是说，要看这个行动会给他们带来快乐还是痛苦（Bentham & Bowring 1843）。另一个例子是，许多科幻电影定义机器人消灭人类的原因是，机器人通过计算发现，最好的拯救地球的最直接方式是消灭人类。这是功利主义的一个生动的例子。

后来又发展了不同的功利主义分支，如行为功利主义。与其他功利主义流派不同，行为功利主义强调的问题是："此时此刻，在这种情况下，怎样做才能促进所有人的幸福价值？"而不是问把这个道德法则推广到每个人身上会对整体的幸福价值产生什么影响。一个简单的例子是，在日常生活中说谎被认为是不道德的。然而，在某些情况下，行为功利主义者会认为说谎是合乎道德的行为。一个例子是一个代理人为了保守国家机密而撒谎。这就是否认道德规则的意义，认为所有的人和他们的情况都是不同的，不可能为行为制定统一的道德规则：在选择行为时，人们必须衡量自己的情况，直接按照功利主义原则行事，即选择一个不仅对自己有利，而且对所有与之相关的人中的最大数量有利的结果。

与此相反，还有一个 "规则功利主义 "的分支。规则功利主义认为，如果每个人都始终遵守一套道德规则，如道路规则，就能产生最大的快乐值。如果每个人都遵守道路规则，那么每个人都可以快乐和安全地驾驶。强调一个行为的道德性是指它对大多数人有利，同时遵守特定的规则或行为准则，导致人们的最大利益或幸福。

行为功利主义和规则功利主义的一个重要区别是，在行为功利主义中，后果在 "行为 "中，而在规则功利主义中，后果在 "规则 "中。行为功利主义通过当前行为的实际后果来评价当前的行为。相反，规则功利主义通过一般行为实践的后果（通过所有的人，包括过去和未来）来评价当前的行为。例如，如果在考试中作弊能让你得到高分，让你的父母高兴，那么行为功利主义者就会作弊，因为它能使所有人的利益最大化。规则功利主义者不会这样做，因为如果每个人都作弊，就没有人有资格上大学，成绩也就没有意义了。这将导致利益的总价值减少。一个规则功利主义者遵循一个规则，并能通过遵循该规则做出收益最大化的决定。在这种情况下，就是任何人都不能作弊的规则，这在短期内可能不会导致最大的收益，但从长期来看是最好的决定。请注意，规则的功利性将集中在长期利益上。如果遵循这个规则在当下带来最大的利益，但在未来没有保证的利益，那就不好了。所以，规则功利主义者在某些情况下，可以不把现在的收益最大化，从而保证长期收益稳定，不会减少。

有一类经典的伦理困境是AV可能遇到的。那就是，如果一辆AV正在路上行驶，但前面突然有三个孩子在过马路，而车辆的左右两边都有沉重的石墩。AV现在必须选择是撞向前面的孩子，还是不顾车主的生命安全，急速转弯撞向石墩。从而牺牲唯一的乘员来避免三个孩子的死亡。这时，行为功利主义动机的协议会考虑你撞上石墩后的生存概率和三个孩子死亡的后果。将会做出权衡，牺牲车主的利益。这是因为行为功利主义选择的结果是在考虑了当前的情况和行为的后果后，对处境中最多的人来说是最大的好处，也是对生命的最大保护。

另一方面，规则功利主义协议还考虑，如果目前的情况在未来多次发生，选择牺牲车主的总体后果是否会带来最大的好处。也就是说，如果在这种情况下牺牲车主的选择成为每个AV都会做的普遍行为，那么对社会和所有人类会有什么后果。在这种情况下，如果选择牺牲车主，那么可能会对AVs行业造成打击，消费者的购买欲望也会降低。这是因为一部分人不会接受这种机器在危机情况下未经用户同意选择牺牲自己的选择。根据世界卫生组织（2018年）的数据，2016年道路交通伤害在全球范围内估计造成135万人死亡。虽然目前还没有证明自动驾驶汽车比人类驾驶员更安全，但自动驾驶汽车的持续发展最终可能会导致车祸率大幅下降。然而，如果反车辆未能普及，从长远来看，似乎也会导致更多的死亡和社会总体福利的下降。任何阻碍AVs发展的选择都会间接伤害到那些未来可能被普通车辆杀死的人。在规则功利主义的眼中，对人类未来的整体利益超过了目前不道德的车辆所带来的任何风险。规则功利主义的协议需要确保反车辆的行为为车主所接受，这对于获得公众对这些新技术的信任也至关重要。

在涉及道德困境的碰撞中，纯粹的行为功利主义协议只是将社会损失降到最低，而纯粹的规则功利主义协议在个别情况下可能是不道德的，两者似乎都不是最佳选择。上面提到的规则功利主义启发的避免碰撞协议认为，为了人类未来的整体利益，可以忽略偶尔的不道德的AV，对此Goodall（2014）提出批评：如果一个群体的幸福可能是以另一个群体为代价的，那么即使是对整个人类的改善，也是社会不能接受的。此外，在Mayer等人（2021）的实验中，他们向参与者描述了反车辆的经典道德困境，他们必须选择是 "以牺牲行人的利益 "还是 "以牺牲乘客的利益"。结果是60%的参与者倾向于拯救最大数量的生命，而也有40%的人倾向于自我保护的选择，即使这意味着牺牲行人（Mayer等人，2021）。他们还提到，视角强烈地决定了参与者的选择，被提示为乘客视角的人始终表示不愿意牺牲乘客，而行人则表示愿意牺牲乘客。这又一次表明，无论是纯粹的功利主义的行为协议还是规则，都可能被大众的道德价值观所接受。

从另一个角度思考，我们可能不需要让出指挥权，AV的道德选择应该由乘坐车辆的人事先选择。也就是说，应该为AVs提供一个道德偏向的设置。这种设置将允许乘客在紧急情况下，在合法的范围内选择自己的道德偏好。例如，选择会更倾向于行为功利主义或规则功利主义或利己主义。因此，在面对道德困境时，AV的决定将取决于用户的自定义选择，责任将被分配给用户，而不是工程师和制造商。根据Contissa等人（2017）的建议，如果用户将AV的道德偏见设置为极端利己主义模式，那么那些预先选择这种偏好的人应该支付更高的保险费。当然，这里的AV的极端利己主义行为也应该是法律框架下允许的AV行为，不会使用户承担刑事或民事责任。一个可以设定道德偏见的防撞协议可以更好地推动AV行业的发展。立法者可以为道德困境设定界限，让用户在这个范围内选择自己的偏好。

总而言之，在道德困境中，自动驾驶汽车必须两害相权取其轻，这就提出了许多伦理和法律问题。在这篇文章中，探讨了避免碰撞协议的两种不同的功利主义动机，即它们会做出什么样的选择和行为。它们是行为功利主义和规则功利主义，前者是在当前情况下对所有人的福祉产生最大利益的行为，后者则强调遵守规则和考虑大局。两者在各自的背景下似乎都有意义，但并不被所有人接受。把道德选择留给用户，似乎是一个可以中和这两种策略的缺点的解决方案。至少有这种协议的AV会在市场上更普遍，对整个AV行业的发展有好处。因此，无论是行为还是规则功利主义动机的AV防撞协议似乎都不是最佳选择。

**参考文献**

Bentham J. & Bowring J. (1843).Jeremy Bentham的作品。W. Tait; *Simpkin Marshall.*

Bentham, J. (1789).道德和立法原则简介》。*Clarendon Press,* Oxford. <http://dx.doi.org/10.1093/oseo/instance.00077240>

Contissa, G., Lagioia, F., & Sartor, G. (2017).伦理旋钮：伦理上可定制的自动驾驶汽车和法律。*人工智能与法律，*25（3），365-378。

Goodall, N. J. (2014).机器伦理和自动驾驶汽车。*In Road vehicle automation* (pp. 93-102).Springer, Cham.

Mayer MM, Bell R, Buchner A (2021) 行人和乘客在涉及自主车辆的道德困境中的自我保护和自我牺牲偏好。*PLoS ONE* 16(12): e0261673. [https://doi.org/10.1371/journal. pone.0261673](https://doi.org/10.1371/journal.%20pone.0261673)

世界卫生组织。(2018).2018年全球道路安全状况报告：摘要。*世界卫生组织*，日内瓦。