

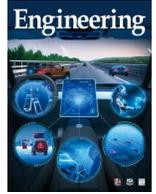


ELSEVIER

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng



Research
Safety for Intelligent and Connected Vehicles—Article

完全自动驾驶车辆事故的责任归属

翟思明^{a,#}, 王琳^{b,#}, 刘鹏^{a,*}

^a Center for Psychological Sciences, Zhejiang University, Hangzhou 310063, China

^b Department of Library and Information Science, Incheon National University, Incheon 22012, Republic of Korea

ARTICLE INFO

Article history:

Received 8 September 2022

Revised 1 May 2023

Accepted 18 October 2023

Available online 16 December 2023

关键词

完全自动驾驶事故

责任归属

可控性

可预见性

摘要

在决策能力日益自动化的复杂系统中,人的作用变得越来越有限。例如,在完全自动驾驶汽车(即无人驾驶汽车)中,人类乘客没有直接的车辆控制权,仅作为乘客存在。一个引人关注的问题是,用户是否要对这些汽车的事故负责。规范性伦理和法律分析通常认为,个人不应该对超出其控制范围的伤害承担责任。在这里,我们通过三项研究的7个实验($N = 2668$)来考虑人类对涉及完全自动驾驶车辆的事故责任的判断。我们比较三种情况下乘客的责任:车主驾驶自己的完全自动驾驶汽车,乘客乘坐无人驾驶的机器人出租车以及乘客乘坐传统出租车,在这三种情况下,这三类乘客都不能直接控制造成相同行人伤害的相关车辆。与规范分析不同的是,我们发现参试者认为无人驾驶汽车(私家车和机器人出租车)上的乘客比传统的出租车乘客要承担更多责任。这种反常结果在不同情景下都存在(例如,参试者来自中国或韩国、参试者采用第一人称视角或第三人称视角,以及乘客是否在场)。进而发现,这并不是因为参试者认为这些无人驾驶汽车用户对驾驶过程有更大的控制能力,而是因为参试者认为他们更应能预见到使用无人驾驶汽车的潜在后果,因此要承担相应责任。我们的研究表明,当无人驾驶汽车(无论是私家车还是出租车)造成伤害时,其用户可能会面临更多的社会压力,因此在进行公共讨论和法律法规制定时,需要考虑这一社会问题。

© 2023 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. 引言

机器(如自动化和机器人),通常由人工智能(AI)驱动,在各种安全关键环境(如道路车辆驾驶、飞机驾驶、医疗诊断和手术)中,与人类一起操作,以增强人类能力,甚至取代人类,以减少人类疲劳和注意力不集中等风险[1–4]。作为我们的研究重点,自动驾驶汽车(AV)是其中一个重要的应用。自动驾驶汽车的广泛使用有助于提高安全性、经济性、可及性和可持续性[5–6]。由于大

多数交通事故的背后是人为错误[7–8],研究人员声称,使用自动机器驾驶员将容易出错的人类驾驶员从事故因果链中移除,因此可以显著减少车祸和伤亡事故[9]。

人类驾驶员和机器驾驶员的角色(或能动性水平)对于理解在发生错误时对谁负什么责任至关重要。二者角色可以用6级驾驶自动化分类系统描述[10]:无自动驾驶(0级;以下简称L0)、驾驶辅助(1级;以下简称L1)、部分自动驾驶(2级;以下简称L2)、有条件自动驾驶(3级,以下简称L3;人类与机器共享车辆控制,并在必

* Corresponding author.

E-mail address: pengliu86@zju.edu.cn (P. Liu).

These authors contributed equally to this work.

要时进行干预)、高度自动驾驶(4级,以下简称L4;不需要人工干预)和完全自动驾驶(5级;以下简称L5;无法进行人为干预),如图1所示[10]。类似6级分类系统也存在于其他领域,如外科机器人[4,11]。本文中除非明确提到,我们将驾驶员/操作员或用户(乘员、乘客或乘坐者)统称为“人”。在完全自动驾驶汽车中,人类只是一个乘客,而机器是唯一的驾驶员(即没有方向盘的无人驾驶车辆)。世界各地的公司都在开发和测试完全自动驾驶汽车和其他自动驾驶汽车。例如,中国、美国和其他国家正在试验机器人出租车(无人驾驶出租车),这是一种完全自动驾驶的商业模式。

与每一种开创性技术一样,AV技术可以带来显著的社会、环境和经济效益,但也会导致伦理、法律和社会问题[12-13]。如果AV引起碰撞事故,责任归因(特别是法律责任归属)可能是最基本和最重要的因素。传统上,驾驶员承担了几乎所有的道德和法律责任。这种以驾驶员为中心的责任制度在AV时代就已经过时。自动驾驶车辆的操作责任从人类驾驶员转移到自动驾驶系统(ADS;即机器驾驶员)。这意味着对交通事故后果责任的转移[14],并引发了人们对自动驾驶汽车造成交通事故谁应该负责的担忧[15-16]。汽车只是一个更大的社会技术系统的一部分。为了让自动驾驶汽车在公共道路上运行,世界必须改变[17],包括关于责任归属的决策制定和立法[18-19]。

涉及不同自动化级别的交通事故已经发生。2016年5月,一场碰撞导致一名驾驶特斯拉自动驾驶汽车(L2)

的驾驶员死亡[20],因为“自动驾驶[L2系统]和驾驶员都没有注意到明亮的天空下拖拉机拖车的白色一侧,因此没有施加制动”[20]。2018年3月,一辆Uber AV(L3)导致一名在人行横道外穿过街道的行人死亡。在这次事故中,ADS错误地将行人识别为车辆、未知物体或骑自行车的人,从而延迟了制动。与此同时,测试驾驶员直到碰撞前一秒才注意到行人。尽管这些事故是由涉及的人类和机器驾驶员的共同错误造成的,但法院免除了这两家公司的法律责任。在第一次陪审团审判中,法院指控Uber的后备驾驶员犯有过失杀人罪,她是唯一面临刑事指控的人[21]。这些事故引起了人们对AV事故中责任归属的担忧。一个紧迫的问题是,人类(通常是最近的操作员)可能会被有意或无意地视为“道德缓冲区”[22]或“法律海绵”[23],从而首当其冲地承担道德和法律责任,而这些人对故障机器的决策权和控制权却十分有限。责任归属的不确定性会对AV的开发和部署产生负面影响[16,18,24],并减缓AV事故的保险框架的建立[14]。

一个有趣的问题是,完全自动驾驶汽车的用户是否要对他们的事故负责?这个问题似乎很幼稚,但也有一些重要的含义。L5 AV用户仅仅是无人驾驶汽车的乘客[25],并且执行有限的决策任务,比如设定行程目的地。目前的法律惯例认为,乘坐出租车、巴士或其他交通工具的无过错乘客对交通事故的发生不负有责任。按照这一逻辑,L5 AV用户可能希望对这些自动驾驶的车辆造成的任何伤害免除后果责任,这将是他们愿意购买或使用L5 AV的动

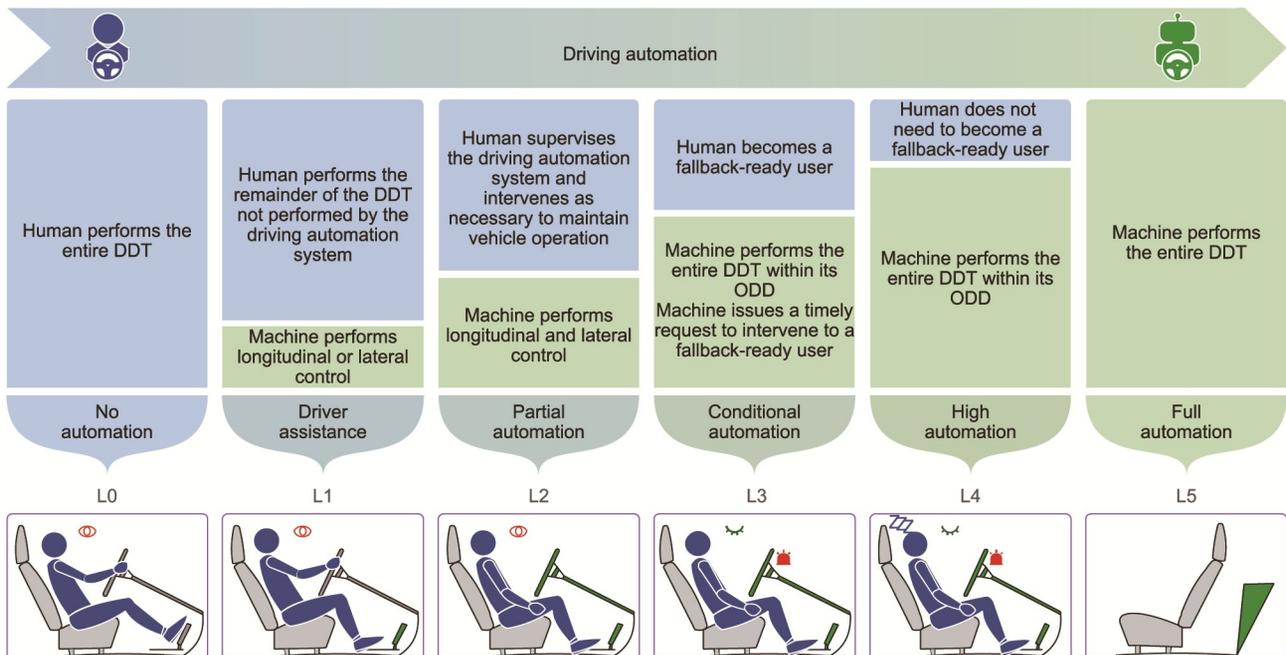


图1. 驾驶自动化的6个级别[10]。DDT: 动态驾驶任务; ODD: 操作设计领域。

机。规范性法律和伦理分析[26–30]认为，如果人类没有出现任何非法行为（如黑客或其他不当干预），后果责任可能会从人类转移到制造商（第1.1节）。这种转变似乎正是该行业所期望的。沃尔沃[31]和奥迪[32]公开承诺，当无人驾驶汽车上市后，将对其车辆造成的碰撞事故承担全部责任。相比之下，社会可能仍会以目前未知的原因指责其用户（第1.2节），这将给他们的未来蒙上阴影。

1.1. 规范分析中的责任归属

后果责任至少有三种形式：因果责任、道德责任（即责备）和法律责任[12,33–34]。它们是不同的，但有密切的相互关系；例如，因果责任归属是责备判断的一个重要因素[35]。可控性（或能动性）是指行为人可以自愿地改变一个原因的程度[36]。这是涉及L5 AV事故责任归属的伦理、哲学和法律论述的核心，通常意味着L5 AV用户不对其车辆造成的伤害承担道德和法律责任。

在伦理学和哲学中，亚里士多德[37]在《尼各马可伦理学》中认为，承担责任至少需要两个条件：控制条件（行为者必须可以控制他们的行为）和认知条件（他们必须知道自己在做什么以及在哪里）。可控性通过控制原则与道德责任联系在一起：个人只能对他们能够控制的行为承担道德责任[38–40]。这是评估一个行为人的道德错误行为的一个常用的标准。在一项伦理分析中，Hevelke和Nida-Rumelin [28]认为，在没有实际的干预机会时，将AV造成的死亡归于L5 AV的乘驾者是一种中伤。他们强调，该乘驾者没有义务注意道路和交通，也没有能力在必要时进行干预，以避免交通事故。然而，哲学家们也注意到了特定案例中的“道德运气”现象：人们可以将行为人为道德判断的对象，即使他们行为的一个重要方面取决于他们无法控制的因素[41–42]。

控制和法律责任之间的联系也是很明确的[43]。控制原则是机动车事故法中的一种法律原则。这些法律都以驾驶员为中心，因为人类驾驶员可以完全控制他们的车辆。因此，法律认为，如果没有减轻罪行的情况，他们将对交通事故承担全部责任。Huddy [44]在《汽车法》中写道：“机动车运行的责任是强加于‘控制’其运动的人。”法律学者对AV事故也表达了类似的观点。由于L5自动驾驶汽车的乘驾者在决策和车辆控制回路之外，不能执行动态驾驶任务的任何部分，某些法律学者认为，他们的乘驾者不太可能对事故负责[25–26,29]。Vladeck [27]称它们和“盆栽植物”没有什么不同的作用。因此，Gurney [30]建议将制造商视为L5 AV的驾驶员；汽车制造商必须对当前道路上的人类驾驶员承担最终责任。

1.2. 来自公众舆论的责任归属

正如技术监管的“社会在环”框架所述，立法者在改变围绕这些技术的法规和立法时，应考虑受新兴技术影响的各种利益相关者的价值和意见，以使其透明、公平和负责[45]。意见众包能识别并计算不同的利益相关者的价值观、选择和意见，同时解决由新兴技术造成的非技术难题。作为公民科学中一种重要的众包方法，情景实验[18, 46–53]研究了如果AV造成伤害，社会公众如何进行责任归因。在这些实验中，参试者会阅读到关于AV事故的报道，这些事故通常被认为是由于人类驾驶员的错误行为（如分心）造成的，参试者要求将自己想象成是涉及不同自动化水平的AV事故的旁观者或受害者，然后从他们的角度将责任归因于多方（如驾驶员、车辆、制造商、行人、开发商和政府）。

其中，两项使用评级量表（如从“无指责”到“大量指责”）的研究报告称，当L5 AV造成事故时，L5 AV中的用户[49]或乘客[18]承受的指责并非最低。这一违反规范发现在这两项研究中没有太多讨论，但在其他研究中得到特别的关注[47,53]。Bennett等[47]考虑了四种类型的车辆（L0、L2、L3和L4/L5）。在他们的情景中，相应的车辆在一条笔直的街道上行驶，撞上了一个过马路的行人，事故时驾驶员正在打电话。然而，他们的情景并没有提到车辆是否有故障。参试者要求站在受害者的角度，将责任分配到6个利益相关者（驾驶员、行人、汽车、政府、制造商和程序员）。Bennett等[47]观察到，随着自动化水平的提高，指责驾驶员的参试者比例减少，而指责制造商的参试者比例增加。然而，大约三分之一的参试者在车辆完全自动驾驶时仍然指责驾驶员。关于这一发现，Bennett等[47]认为这是一种“驾驶员羞辱”。然而Bennett等[47]仍然在他们的L5情景中使用“驾驶员”一词，这与L5 AV中的人在车内的角色不匹配，Bennett等研究中L5 AV中的人可能仍然受到责备可能归咎于他们的“驾驶员”角色。Aguiar等[53]进行类似研究，认为这一现象与控制原则相矛盾（第1.1节）。

1.3. 研究动机

与规范分析相比，最近的公众观点研究[18,47,49,52–53]为责任归属倾向提供了早期证据：人们仍然将涉及完全自动驾驶的事故归咎于L5 AV的用户，而实际上他们并没有直接控制车辆。以往研究可能会存在方法方面的缺陷：所观测到的观点倾向可能来自于主观测量中潜在的测量误差[54]。例如，一些“不专心”的参试者可能没有在选择责任题项上选择假定的“无”（或“没有”）水平。然

而，如果我们发现不能将这一发现归因于主观测量的潜在弱点，那么它就反映了更深层次的社会心理，这可能表明人们对完全自动驾驶及其伤害的社会反应。更具体地说，这可能预示着对未来完全自动驾驶汽车的消费者的社会偏见，这对 AV 行业和 AV 事故的立法有实际影响（例如，对消费者使用 L5 AV 汽车的积极性产生负面影响）。它也具有理论意义，因为它可能表明了控制和责任之间的不对称性。研究者在社会心理学研究中很少关注完全自主机器造成的伤害相关的责任归属。对机器操作的责任问题已经变得越来越重要。我们的研究也为越来越多关于自主机器责任归属的文献做出了贡献[2,18,55–56]。

总的来说，目前关于 L5 AV 事故的责任归属的经验证据是暂时性的（由于主观测量中潜在的测量误差）。学者们还没有研究责任归因的边界条件和潜在的心理机制。在本研究中，我们系统地调查并确定了人们是否要求 L5 AV 用户对其 L5 AV 造成的伤害负责。

2. 理论和研究问题

在三项研究的 7 个实验（6 个在中国进行，1 个在韩国进行）中，我们重点关注了两个研究问题。第一个问题是，L5 AV 用户是否需要对这些车辆造成的伤害承担（部分）责任？为了回答这个问题，我们在研究 1 中研究了其在不同实验条件（即其潜在边界条件）下的持续性和稳健性。第二个问题是，其可能的心理机制是什么？我们从社会心理学领域归因理论中的可控性（研究 2）和可预见性（研究 3）的角度来回答这个问题。接下来，我们将解释这些问题的理论基础和实验安排。

2.1. L5 AV 用户是否被要求对无人驾驶汽车造成的伤害负责？

我们通过考虑传统的出租车乘客（根据现行交通法规

其对任何事故不负任何责任）作为参考对象，并比较了他们和 L5 AV 用户的责任，避免了以前的公众观点研究中的方法缺陷。我们考虑了 L5 私人汽车的车主和 L5 无人驾驶出租车（即无人驾驶出租车，一种全球范围内试点的无人驾驶服务）的乘客的责任。如果发现，当乘坐两种不同的出租车导致相同的碰撞事故时，乘坐无人驾驶出租车的乘客也比乘坐传统出租车承担更多的责任，那么我们就为“公众要求 L5 AV 用户承担相关责任”这一现象提供了更有力的证据。

为了验证“当无人驾驶汽车造成伤害时，参试者让 L5 AV 用户承担责任的倾向”，我们操纵了不同实验条件，以保证其稳健性。首先，我们指定了默认条件（作为参考条件）。通常，在讨论责任时，法律学者、伦理学家、政策制定者或外行人从第三人称的角度（即观察者的角度）来评估责任。关于责任归因的默认心理测量手段，我们模拟法院或责任认定机构将责任分配给多个责任人的实际方法；因此，我们要求我们的参试者在相关车辆造成事故后，将固定数额的责任分配给乘员和其他责任代理人。在默认情况下，乘员都在车里。因此，研究 1a 中操纵的默认条件是，参试者从第三人称视角将固定数量的责任分配给了车内乘员和另一个责任代理人（责任分配×第三人称视角×乘客在车内）。据此，我们设计了几种实验条件（图 2）。

先前研究[18,49]采用了责任评估量表（即参试者对不同责任人进行独立责任评级）。评估方法差异可能导致多个研究中责任归属结果的差异[46]。因此，研究 1b 考虑了责任评价方法。

换位思考对责任归属至关重要[57]。以往的研究已经考察了第二人称视角（即受害者视角）[47]和第三人称视角[49]。根据防御性归因理论[58]和自我保护倾向[59]，从第一人视角出发的参试者（即想象自己为 L5 AV 用户）会认为他们不应该对自己无法控制的车辆造成的伤害承担

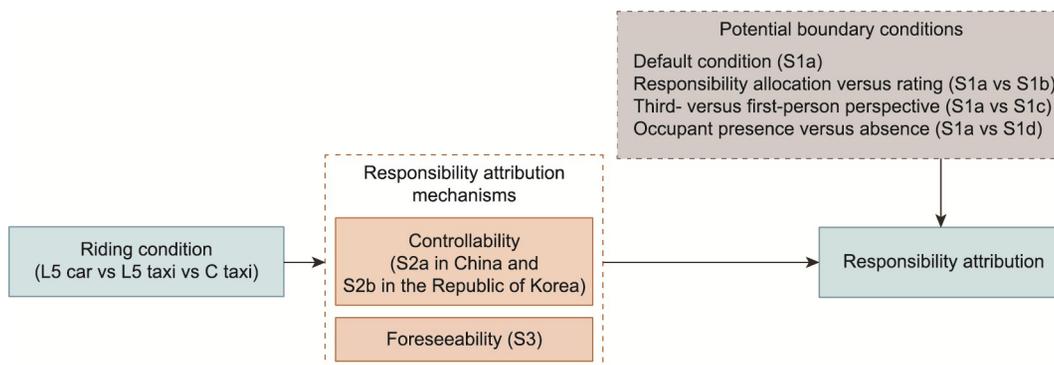


图 2. 涉及 L5 AV 事故的责任归属的理论框架。研究 1（S1）、研究 2（S2）和研究 3（S3）表明了测试特定条件或心理机制的实验研究。研究 1a 的默认条件是，参试者从第三人称角度将固定数量的责任分配给车内乘员和其他责任代理人（责任分配×第三人称视角×乘客在车内）。C taxi: 传统出租车。

责任。然而，如果从第一人称的视角来看，我们所研究的责任归因倾向仍然存在，那么它将得到更实质性的证据。因此，研究1c考虑了第一人称视角。

研究人员想知道，这种趋势是否是因为用户在车里（作为离事故最近的人）的缘故[47]。我们感兴趣的是无人驾驶汽车中用户在场是否会影响责任分配。因此，作为一种回应，研究1d规定了当碰撞发生时，乘客不在车内。

2.2. 潜在的心理机制是什么？

如果我们感兴趣的“参试者要求L5 AV用户承担责任”的倾向是持续和稳健的，那么如何解释它呢？一个简单的解释是，L5 AV车主应该对汽车造成的不良后果负责（尽管他们不对动态驾驶任务负责），这可以由所有权和责任归属之间的关联来支持（如参考文献[60–61]）。如果有人购买并拥有一辆车，社会就会接受他们的合法所有权，法律也会保护他们从使用车辆中受益的权利。相应地，社会认为，如果不存在减轻处罚的条件（如汽车中的产品缺陷），他们要对这辆车所造成的负面后果负责。Rahwan等[3]在讨论自动机器造成伤害后的责任归属时也表达了类似的观点：“如果狗咬了人，狗的主人就要负责。”我们的研究还比较了机器人出租车和传统出租车场景下乘客责任差异。由于在这两种乘坐条件下的出租车乘客都不拥有出租车，我们可以说明这种倾向到底是由于财产所有权还是其他未知因素造成的。正如我们将揭示的，与传统出租车相比，出租车乘客在无人驾驶出租车上仍然要承担更多的责任，这表明这种倾向（即参试者要求L5 AV用户承担责任）并不完全是由于所有权造成的。

社会心理学中的归因理论理解人们如何识别某些事件的原因和责任归属的必要条件。我们使用归因理论[36,62–65]来建立我们的理论，用来解释涉及L5 AV用户的责任归因倾向。具体来说，我们使用了感知可控性和可预见性两种解释，这两种解释在为负面结果分配责任方面发挥着重要的作用。

社会心理学中的归因理论和实证研究支持了伦理学与哲学中控制原则中的控制与责任之间的联系（第1.1节）。Weiner [36]的归因理论认为，人们在试图理解事件的原因时会检查三个维度，其中之一是可控性。根据Weiner的理论，因果控制是确定责任的必要条件。Alicke [63]提出了过失控制模型来评估过失发生的过程。它的一个核心假设是，普通人根据行为者对有害事件的个人控制来评估潜在的应受谴责的行为。实证研究表明，当机器自主性增加，人类控制减少时，人们分配给相关人类的责任就会减少[47,49,52,66]。在L5的情景下，人们可能会怀疑L5 AV

汽车（个人拥有的汽车或机器人出租车）的用户是否不能进行干预。因此，一个可能的解释是，尽管三个乘客（私家L5车的车主、机器人出租车的乘客和传统出租车的乘客）在字面上和实际上没有参与车辆的驾驶决策和控制，但人们误认为两个用户在两个L5条件下对驾驶有更多的控制权，因此，对驾驶结果有更多的责任。研究2（两个跨国实验）检验但未能证实这一假设，因此进行了后续研究3。

人们的责备归因与对参试者的认知状态感知相关[62–64]。因此，另一个可能的解释是乘驾者对结果的可预见性。Lagnado和Channon [64]将这种结构分为三种：主观可预见性（即从行动者的角度来看，结果的可能性有多大）、客观可预见性（即实际上可能发生的情况，不管行动者实际期望什么）和合理的可预见性（即行为者合理地期望什么，即在给定他们可用的信息的情况下，他们应该期望什么）。Lagnado和Channon [64]观察到主观和客观可预见性影响人们的责备判断。然而，很少有研究人员实际检验了合理可预见性感知的的影响。Shaver [62]认为，责任归属应该受到行为者对其行为结果的信息（合理可预见性）的影响，而不是受他们实际知道的信息（主观可预见性）的影响。合理可预见性在法律案件中尤为重要，因为它是过失侵权的一个特征[67]。虽然非专业人士和专业人士对合理可预见性的理解存在差异，但我们认为非专业人士对合理可预见性的理解可能会影响非专业人士的责任判断。因此，我们在研究3中检验了合理的预见性，并假设人们将更多的责任归给L5 AV用户，因为他们认为用户应该更多地意识到使用L5的后果。

我们连续进行三项研究（图2），以更深入地理解L5 AV事故的责任归属。浙江大学心理科学研究中心的伦理审查委员会批准了这些研究。我们运行StatCheck [68]，并没有检测到我们的统计数据的不同组成部分（如 t 或 F 值、 df 和 p ）之间存在任何不一致。

3. 研究1

研究1旨在研究，当车辆造成相同的行人伤害时，人们对L5 AV汽车用户（私人L5汽车和无人驾驶L5出租车）的责任判断是否大于传统出租车乘客的，以及这种模式在不同的实验条件下是否稳健。研究1a设定了默认实验条件：参试者从第三人称的视角将固定数量的责任分配给车上的乘客和另一个责任方。研究1b使用了责任评级量表（vs研究1a中的责任分配量表），研究1c考虑了第一人称视角（vs研究1a中的第三人称视角），研究1d考虑了

乘员不在场（vs 研究 1a 中乘员在场）。

3.1. 方法

所有的实验都采用了一种情景实验设计方法[1,18,47,49]，参试者读到一个假设的碰撞事故，然后将责任分配给相关责任方。

3.1.1. 参试者

我们采用了参试者间设计，操作三种乘驾条件（车主乘坐私人 L5 汽车、乘客乘坐 L5 出租车以及乘客乘坐传统出租车）。我们计划在所有实验中，每种情况下在线招募至少 120 名参试者，以达到 80% 统计效力研究的经验法则（即每个条件至少 100 名参与者）[69]。

我们排除了研究 1a 中的一名参试者和研究 1c 中的两名参试者的数据，因为他们的年龄在 18 岁以下（不符合我们预定要求）。我们通过中国的一个平台（Sojump）提供的样本服务在线招募参试者。本研究的参试者如下：研究 1a: $n = 393$, $M_{age} = 30.4$ 岁, 45.5% 为女性；研究 1b: $n = 391$, $M_{age} = 30.5$ 岁, 50.4% 为女性；研究 1c: $n = 389$, $M_{age} = 30.6$ 岁, 48.1% 的女性；研究 1d: $n = 371$, $M_{age} = 31.1$ 岁, 53.4% 的女性。敏感性检验[70]显示，我们在研究 1a~1d 中的最终样本量可以提供 80% 效力来检测 $\eta_p^2 = 0.024 \sim 0.026$ （小到中等效应量； $\alpha = 0.05$ ）的影响。附录 A 中的表 S1 提供了更多关于我们的参试者的人口统计信息。

3.1.2. 程序

我们将参试者随机分配到三种乘驾条件中的一种。参试者首先阅读两种 L5 条件下的 L5 的定义（改编自参考文献[10]）：“L5 自动驾驶车辆是指：全自动驾驶系统能够执行全部驾驶任务，处理所有路况，进行全自动驾驶，彻底解放驾驶人的手和脚，车辆内没有方向盘和踏板，因此车内人员无法对任何驾驶任务进行手动干预。”并附有图形说明（改编自 Diels 和 Bos [71]）。

随后，参试者阅读了一个碰撞事故的场景。以 L5 私家车为例；研究 1a 和 1b 的事故是“在城市道路上，一辆 L5 级自动驾驶汽车正搭载它的车主，车辆处于自动驾驶模式，撞到一位突然横穿道路的行人，导致其受伤。事故前该车车主正在看手机，全自动驾驶系统未能正常工作。”（参考文献[47,49]）。这与 2018 年 Uber 的 AV 造成的第一起行人死亡事件相似[21]。研究 1c 考虑了第一人称视角。所以，我们要求参试者把自己想象成车里的乘员。因此，碰撞情景的措辞有一个小的变化，在研究 1c 中，车载乘客（L5 车的车主和 L5/传统出租车的乘客）被“您”

替代。研究 1d 考虑了乘员不在车内：乘员通过智能手机应用程序呼叫涉事的车辆，等待其到达，并在等待期间发生了碰撞事故。在附录 A 中的表 S2 中列出了所有的碰撞事故场景。

在阅读了有关事故的内容后，参试者将三种类型的责任（因果责任、责备责任和法律责任题项顺序随机）分配或评分给三名乘客中的一个[12,34]。研究 1b 考虑责任评级量表，其他三个考虑责任分配。例如，在 L5 私家车的情况下，关于责任评级的三个问题是（以“您认为在这起事故中”开始）：“该车车主在多大程度上导致了该行人的受伤？”（因果责任），“该车车主应该受到的谴责程度有多大？”（责备责任），以及“该车车主的法律责任有多大？”（法律责任），10 分制（1=非常低到 10=非常高）（改编自 Liu 和 Du [72]）。其他研究中关于责任分配的三个问题是（从“认为在这起事故中”开始）：“该车车主和其他主体（L5 级汽车制造商）分别在多大程度上导致了该行人的受伤？”（因果责任），“该车车主和其他主体（L5 级汽车制造商）分别应该受到多大程度的谴责？”（责备责任）和“该车车主和其他主体（L5 级汽车制造商）分别应该承担多少法律责任？”（法律责任）。参试者使用滑块或直接输入数字（改编自 Kirchkamp 和 Strobel [73]）分配固定数量（10）的责任给车主和其他责任主体。

在另外两种乘驾条件下，我们将术语“车主”替换为“车内乘客”。另一个责任方是 L5 出租车条件下的 L5 出租车制造商和服务运营商，以及传统出租车条件下的出租车驾驶员。我们做出这些改变是为了匹配这些责任方的角色（尽管词语上的差异可能会影响参试者的判断）。我们要求参试者在分配责任时忽略行人的责任。在两种 L5 条件下，我们没有指定 L5 车辆作为责任方，因为机器代理本质上不成为道德代理[74]，而 L5 车辆等机器不能对惩罚和责备做出反应[27]。此外，在追究责任之前，我们还测量事故引起的负面影响，如附录 A 所示。

最后，参试者提交了性别、年龄和驾驶执照持有情况等基本信息，并获得了相当于 0.31 美元的补偿。

3.2. 结果和讨论

由于三种责任测度的判断数值相似且具有显著的内部一致性，因此我们将它们平均化，获得一个单一因素的责任评级（研究 1b 中 Cronbach 的 $\alpha = 0.94$ ）或分配（其他子研究中 Cronbach 的 $\alpha = 0.90 \sim 0.92$ ），类似于以前的研究[18]。我们以乘员责任为结果变量，以乘驾条件（L5 车=0，L5 出租车=1，传统出租车=2）为自变量，以性别（男性=0，女性=1）、年龄和是否拥有驾驶执照（没有=

0, 有=1) 为三个协变量, 进行协方差分析 (ANCOVA)。图 3 给出了不同条件下乘员责任的估计边际均值 (EMM)。我们选择了 ANCOVA 而不是方差分析, 是因为研究 1d 中参试者的人口统计学因素在三种乘驾条件下在统计学上并不相等 (这是一个巧合; 见附录 A 中的表 S1)。因此, 我们添加了参试者的人口统计学因素作为协变量, 并在所有的统计分析中对它们进行控制 (为了保持一致性)。使用最小显著差异法 (LSD) 进行两两比较。

在所有四项调查中, 乘驾条件对乘员的责任都有显著影响 (研究 1a: $F_{(2,387)} = 42.69, p < 0.001, \eta_p^2 = 0.181$; 研究 1b: $F_{(2,385)} = 62.04, p < 0.001, \eta_p^2 = 0.244$; 研究 1c: $F_{(2,383)} = 45.70, p < 0.001, \eta_p^2 = 0.193$; 研究 1d: $F_{(2,365)} = 4.10, p = 0.017, \eta_p^2 = 0.022$)。与传统的出租车乘客相比, 参试者将更多的责任归咎于两个 L5 用户 ($ps < 0.01$; 图 3), 无论责任测量方法 (责任评级与分配方法)、参试者的视角 (第一人称与第三人称视角) 或他们的缺席和在场情况 (乘客不在场与在场) 如何。研究 1d 中的乘客不在车内是唯一的例外, 两名出租车乘客承担相同的责任 ($p = 0.441$); 然而, 如图 3 (d) 所示, 与等候 L5 出租车的乘客 ($\Delta EMM = 0.51, t = 2.04, p = 0.042$) 和传统出租车的乘客 ($\Delta EMM = 0.70, t = 2.77, p = 0.006$) 相比, 参

试者仍然将更多的责任分配给了等候 L5 私家车的车主。

前人研究初步表明, 当这些 L5 AV 车辆造成碰撞事故时, 根据他们的责任分配测量方法, 无论是 L5 AV 用户 [18,49] 还是驾驶员 [47], 参试者认为他们的责任并未为“无”。在本研究中, 我们以传统出租车上的无错的乘客为参考, 为以下发现提供了明确而有力的证据: 在本研究中所涉及的任何情况下, 当车辆造成相同的碰撞和损坏时, 人们认为 L5 AV 车 (私家车和机器人出租车) 用户比传统出租车乘客的责任更高。这一发现是令人困惑和反规范的, 因为 L5 AV (私家车或机器人出租车) 的用户在车辆的驾驶决策和控制中几乎没有发挥任何作用。

4. 研究 2

为了解释研究 1 中获得的令人困惑的发现, 我们扩展研究 1a, 以探索研究 2 和研究 3 中潜在的心理机制。研究 2 假设: 观察者 (参试者) 认为无人驾驶汽车的用户对驾驶有更多的控制权, 因此认为他们对驾驶结果负有更多的责任。请注意, 从技术层面来说, 这种想法是不正确的, 因为车内乘客只是无人驾驶汽车的乘客, 没有直接的车辆控制权。研究 2 考虑来自中国 (研究 2a) 和韩国 (研究

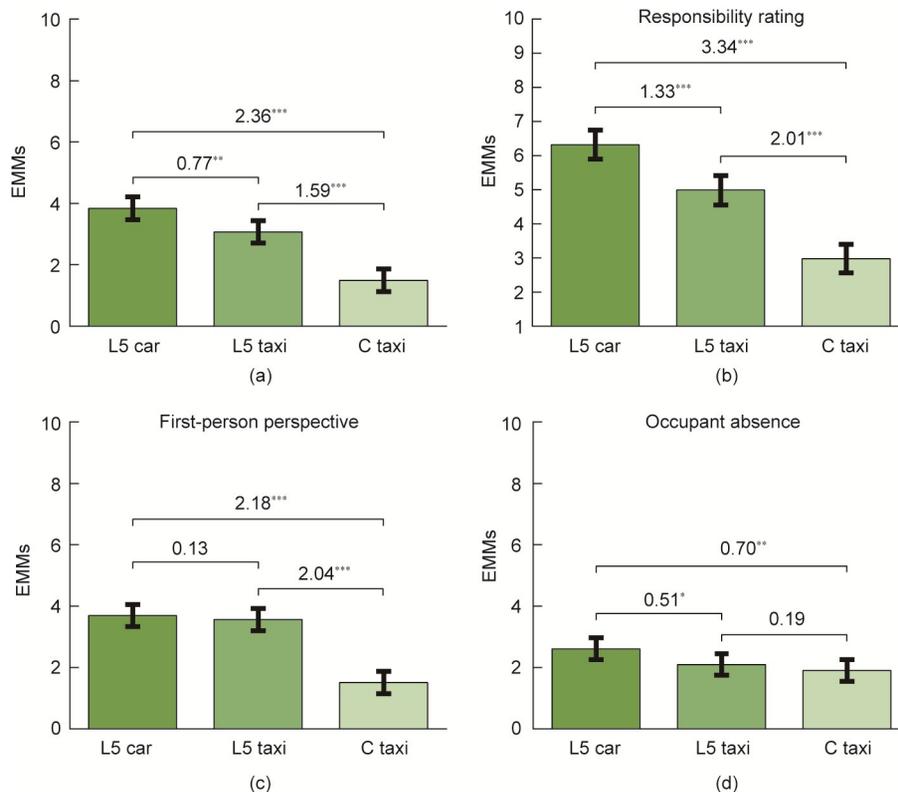


图 3. 研究 1a (a)、研究 1b (b)、研究 1c (c) 和研究 1d (d) 中乘员责任的估计边际均值 (EMM)。误差棒 = ± 2 标准误差 (SE)。* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ 。我们将研究 1a 设置为默认值 (责任分配 \times 第三人称视角 \times 乘员在车内)。其他三个子研究与研究 1a 之间的差异如说明文字所示。

2b) 的参试者来检验这一假设。这种跨国研究设计也能证明我们在责任归因方面的发现是否可以拓展到不同国家。

4.1. 方法

4.1.1. 参试者

与研究1一样,我们采用了参试者间设计,并在每次调查中操纵了三种乘驾条件。参试者是通过中国的 Sojump (研究2a: $n = 395$, $M_{age} = 30.0$ 岁, 47.6%为女性)和韩国的 Survey Billy (研究2b: $n = 360$, $M_{age} = 41.4$ 岁, 40.3%为女性)在线招募的。此外,我们在研究2a中排除了一名年龄不到18岁的参试者。样本量可以提供80%统计效力来检测研究2a中 $\eta_p^2 = 0.024$ 、研究2b中 $\eta_p^2 = 0.026$ 的效应(小到中等效应量; $\alpha = 0.05$)。

4.1.2. 程序

研究2的程序与研究1a相同,但测量方法存在差异。研究2a和2b要求参试者在了解碰撞事故之前对驾驶时的可控性感知进行评分。参试者回答了三个问题(以私人L5车为例)，“当乘坐L5级自动驾驶汽车时,整个驾驶过程在车内车主的掌控之中/车内车主控制了整个驾驶过程/车内车主主导了整个驾驶过程”,以7分制(1分=绝不同意,7分=绝对同意),评定车内人员(L5车内的车主或L5/传统出租车内的乘客)的可控性。这三个关于感知可控性的项目(Cronbach在研究2a中的 $\alpha = 0.90$,在研究2b中的 $\alpha = 0.93$)改编自参考文献[60]。与研究1a中一样,参试者回答了关于责任分配的三个问题(Cronbach在研究2a中的 $\alpha = 0.94$,研究2b中的 $\alpha = 0.95$)和一个关于负面情绪的问题。

4.2. 结果和讨论

研究2重复了研究1a的结果。ANCOVA结果显示,乘驾条件对感知乘员责任有显著影响(研究2a: $F_{(2,389)} = 41.07, p < 0.001, \eta_p^2 = 0.174$; 研究2b: $F_{(2,354)} = 15.65, p < 0.001, \eta_p^2 = 0.081$)。两国三种乘驾条件下乘员感知责任是不同的[$p_s < 0.01$; 图4(a)和(b)],L5车中感知的乘客责任最多,传统出租车中感知的乘客责任最少,唯一的例外在韩国,L5车的乘员责任略高于L5出租车条件下的乘员责任[$\Delta EMM = 0.53, t = 1.74, p = 0.083$; 图4(b)]。

三种乘驾条件下的汇总数据显示,感知可控性和责任分配之间存在正相关关系(研究2a: $r = 0.41, p < 0.001$, 95%置信区间(CI) [0.33,0.49]; 研究2b: $r = 0.21, p < 0.001$, 95% CI [0.11,0.31])。然而,研究2拒绝了我们的假设,即观察者认为无人驾驶车辆的乘员对驾驶有更多的可

控性(能动性),因此,对驾驶结果承担更多的责任,因为他们判断乘客在三种乘驾条件下具有相同的可控性(研究2a: $p = 0.872$; 研究2b: $p = 0.117$; 图4)。

我们还根据之前的ANCOVA测试,将参试者的国籍(中国=0,韩国=1)作为自变量。我们观察到国籍($p = 0.962$)和乘驾条件($p = 0.621$)对乘员可控性的主效应不显著,以及交互效应不显著($p = 0.306$)。就感知到的乘员责任而言,国籍($F_{(1,746)} = 14.83, p < 0.001, \eta_p^2 = 0.019$)和乘驾条件($F_{(2,746)} = 52.26, p < 0.001, \eta_p^2 = 0.123$)具有主效应。它们的交互作用不显著($p = 0.099$)。韩国参试者比中国参试者感知到更多的乘员责任($\Delta EMM = 0.79, t = 3.85, p < 0.001$; 见图4)。另外,如前所述,在三种乘驾条件下,感知到的乘员责任是不同的($p_s < 0.01$)。

最近几项研究[46–47]采用了(但没有实证检验)人类可控性的概念来解释为什么当汽车自动化水平提高时,用户或驾驶员承担的责任会更少。与Weiner [36]的归因理论相一致,当我们的参试者感知到乘员对驾驶有更大的控制权时,他们在这两个国家的所有三种乘驾条件下,均将更多的责任分配给了人类乘员。然而,他们感知到的乘员可控性并没有考虑到他们在三种乘驾条件下乘员责任的差异。

5. 研究3

研究2未能证实感知可控性在解释观察到的涉及L5 AV事故的乘员责任归属模式。如第2.2节所述,后续研究3检验了人们将更多责任归咎于L5 AV用户的假设,因为他们认为L5 AV用户应该更多地意识到使用无人驾驶汽车的后果(合理预见性)。

5.1. 方法

5.1.1. 参试者

与研究1a一样,我们在中国在线招募了参试者,并排除了两名年龄在18岁以下的参试者,最终剩下369名参试者($M_{age} = 30.6$ 岁, 51.8%为女性),这可以提供80%统计效力来检测 $\eta_p^2 = 0.026$ 的影响(小到中等效应量; $\alpha = 0.05$)。

5.1.2. 程序

研究3的程序与研究1a相同,但测量方法存在差异。参试者在判断责任分配后对两个项目(改编自McCaul等[75])做出回应,评估乘员对事故的可预见性和他们乘坐的后果(Cronbach的 $\alpha = 0.80$)。在参试者得知事故发生之

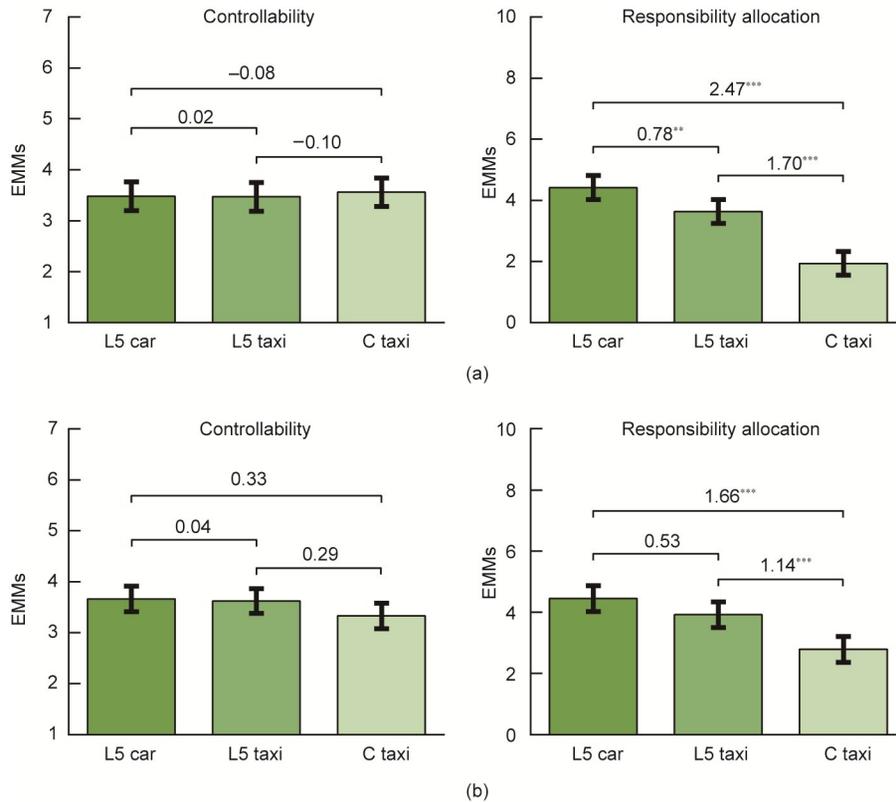


图4. 中国的研究2a (a) 和韩国的研究2b (b) 中成员责任和可控性的EMM。误差棒 = ±2 SE。** $p < 0.01$; *** $p < 0.001$ 。

前, 他们还回应了评估另一个负责方的可控性的三个项目 (Cronbach的 $\alpha = 0.83$; 附录 A)。我们通过以下指标来衡量可预见性 (以私人 L5 车为例): “该车车主应该能预料到会发生此类事故” 和 “该车车主应该能预料到乘坐 L5 级自动驾驶汽车可能造成的事故后果”, 采用 7 分制 (1 分 = 绝不同意, 7 分 = 绝对同意)。随后, 类似于之前研究, 我们测量责任分配 (Cronbach的 $\alpha = 0.89$) 和事故的负面影响。

5.2. 结果和讨论

如前所述, 乘驾条件影响了感知乘员责任 ($F_{(2, 363)} = 34.13, p < 0, \eta_p^2 = 0.158$)。这在三种情况下都有所不同 ($ps < 0.001$), 在 L5 型车中最高, 而在传统出租车中则最低, 如图 5 (a) 所示。乘驾条件也显著影响了感知可预见性 ($F_{(2, 363)} = 37.19, p < 0.001, \eta_p^2 = 0.170$)。在 L5 条件下感知乘员可预见性最高, 在传统出租车条件下感知乘员可预见性最低, 如图 5 (a) 所示。在三种乘驾条件下, 感知乘员可预见性与感知乘员责任呈正相关 ($r = 0.50, p < 0.001, 95\% \text{ CI } [0.42, 0.57]$)。

进行中介分析 (模型 4, 有 5000 个重样本, 遵循 Hayes 和 Preacher [76]), 以乘驾条件作为多类别自变量 (设计了两个对比: L5 车相对于传统出租车和 L5 出租车

相对于传统出租车)。中介分析表明, 乘驾条件通过感知乘员可预见性产生间接影响 [$f_1 = 0.81, 95\% \text{ CI } [0.56, 1.10]$; $f_2 = 0.48, 95\% \text{ CI } [0.27, 0.73]$; 图 5 (b)]。由于乘驾条件的直接影响仍然很显著 ($c_1' = 1.22, p < 0.001$; $c_2' = 0.58, p = 0.015$), 因此感知乘员可预见性具有部分中介效应。

因此, 研究 3 支持这样一种观点, 即人们将更多的责任归咎于 L5 AV 用户, 因为人们认为他们应该更多地意识到使用 L5 AV 的后果。因此, 我们发现了合理预见性的解释作用 [64], 以解释为什么人们在乘坐车辆 (L5 AV 和传统出租车) 并造成相同伤害时对乘员责任有不同的判断。

6. 责任归属差异总结

通过一系列的单论文元分析 (single-paper meta-analysis) [77], 我们总结研究 1a、1b、2a、2b 和 3 的结果 (均与第三人称视角和乘员在车内有关)。由于研究 1c 和 1d 具有独特的、单一研究水平的调节因素 (第一人称视角和乘员不在车内), 我们无法使用目前的单论文元分析来对它们的结果进行分析。L5 汽车与传统出租车、L5 出租车与传统出租车、L5 汽车与 L5 出租车的简单效应的点估计值分别为 1.07 (95% CI [0.93, 1.21])、0.67 (95% CI [0.53, 0.82]) 和 0.40 (95% CI [0.25, 0.54]) (均以标准偏差单位

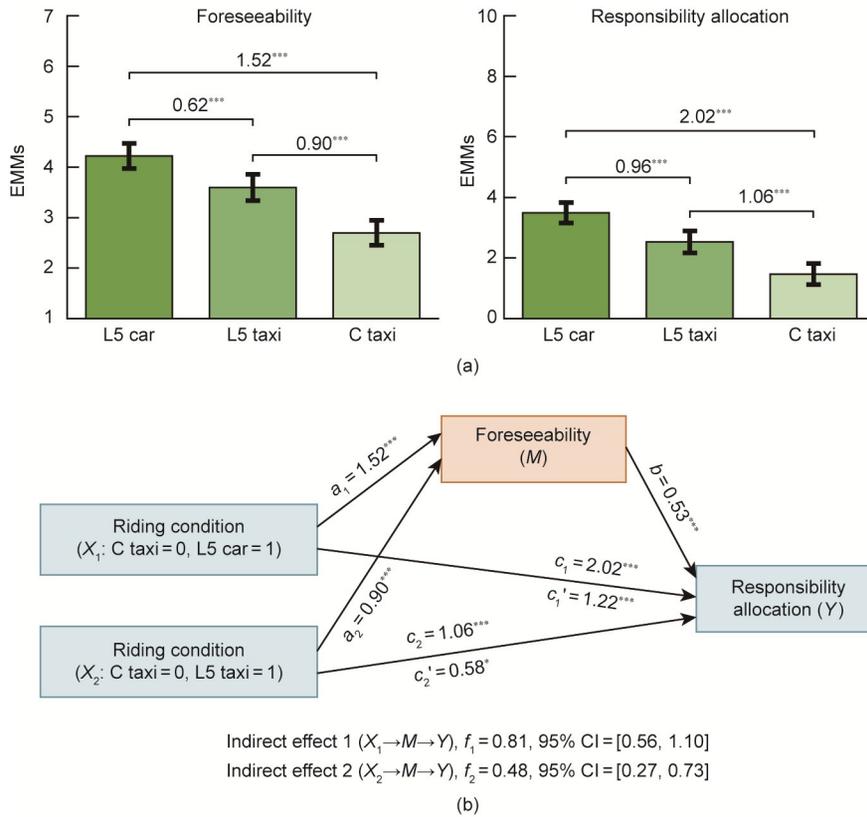


图5. 研究3中感知乘员责任和感知乘员可预见性的EMM (a) 和中介分析 (b)。误差棒 = ± 2 SE。* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ 。

表示) (图6)。因此, 当无人驾驶汽车 (私家车和机器人出租车) 发生碰撞事故时, 参与者对无人驾驶汽车乘客的责任归因要高于传统出租车乘客。

键环境[1]中兴起时。我们的研究是最早分析人类对使用全自动化机器产生的不良后果后人类责任的研究之一。在下文, 我们将讨论我们的研究结果的理论和实际意义。

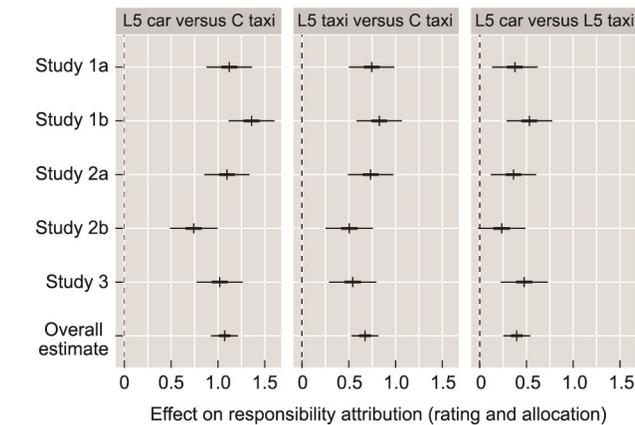


图6. 研究1~3中驾乘条件对乘员责任判断 (评级和分配) 影响的单元文元分析。粗线和细线分别代表50%和95%的置信区间。

7. 总体讨论

完全自动和无人驾驶汽车的用户是否要对其交通事故承担责任? 如果是这样, 那么基于什么理由呢? 这些问题出现在机器人和人工智能在交通运输[18-19]和其他安全关

我们的理论贡献在于, 对完全自动机器造成伤害后用户责任的决定因素和心理过程进行实证调查。我们希望我们的研究结果具有实际意义, 因为责任判断对于理解消费者反应[60,78]及确定 AV 设计和产品定价[18]至关重要。我们的研究结果揭示了社会对自动驾驶碰撞事故的责任分配的期望, 从而为建立一个社会可接受的监管方案提供了见解。

7.1. 理论意义

大多数来自伦理、法律、哲学和其他学科或常识的规范分析表明, 让无人驾驶汽车的乘客或车主对他们无法控制的行为负责是不公平的[18,24]。相比之下, 先前实证研究[18,47,49,52-53]提供的初步证据表明, 当L5型自动驾驶汽车造成碰撞事故时, 用户 (或“驾驶员”) 的责任并非为“零”。因此, 这揭示客观可控性与责任归属之间的可能冲突[53]。研究人员尚未研究其潜在的边界条件和根本原因。

在我们的三项研究和7个实验中, 我们将传统出租车乘客作为参考 (根据现行法律制度, 在无错的情况下, 他

们对所涉及的交通事故不承担任何责任)。我们比较了这名完美乘客和L5 AV用户(L5车主和出租车乘客)的责任。我们观察到一个有趣的“困境”或趋势:虽然L5 AV用户没有直接的车辆控制权,但人们可能会要求他们对这些车辆造成的伤害承担部分责任。这种违反直觉的倾向在不同的研究背景下都成立。特别是,将自己想象成L5 AV用户的参试者也比想象自己是传统出租车乘客的参试者要求承担更多的责任[图3(c)]。考虑到众所周知的自我服务倾向[59]或防御型归因倾向[58],这是令人惊讶的结果。它与社会可以对无人驾驶汽车的车主施加严格责任的观点相似[79];然而,这个观点并不能解释为什么无人驾驶出租车上的乘客也被要求比传统出租车上的乘客承担更多的责任。同样,所有权(即拥有一个物体)的概念可能会解释要求无人驾驶汽车车主承担更多责任的原因[60-61],但不能解释要求无人驾驶出租车乘客承担更多责任。

因此,我们的理论贡献是,我们通过有力的证据证实了这种倾向(客观可控性和责任判断之间的冲突)的稳定性,并通过可预见性[62-63]——更具体地说,合理可预见性[64]——的视角对其进行解释。参试者(持旁观者视角)更有可能认为,无人驾驶汽车(私家车或出租车)的用户应该能够预见到使用无人驾驶车辆的后果,从而认为这些用户应对其车辆造成的事故承担更多责任。这可能是违反事实的想法,或者是对这些无人驾驶车辆用户有偏见的直觉反应,因为他们在现实中可能没有更多的先见之明[80]。合理预见性的潜台词是,旁观者认为,使用无人驾驶汽车比使用传统出租车的风险更大,因此,如果无人驾驶汽车造成交通风险和伤害,他们会更容易指责用户。在传统如过失、产品责任和严格责任[67,81-82]等侵权原则中,合理预见性也表明法律责任大小,尽管人们对其的通俗理解与这些原则中的含义不同。

涉及自主机器的责任归属的文献[2,55-56]广泛讨论了影响责任归属的重要因素,如感知意图和客观预见性。我们通过研究在完全自动机器出现错误时用户的责任归属,并强调机器用户对其责任的合理预见性的重要性,从而扩展相关文献。例如,Hidalgo等[55]报道,当人工智能机器(相对于人工)创建相同的淫秽图像时,雇用了人工智能机器(相对于雇用人工)来创建广告图像的公司要承担更多的责任。根据我们的调查结果,公司的合理预见性可能部分解释了Hidalgo等人的研究结果。

7.2. 实际意义

在讨论潜在的实际意义之前,我们强调,我们并没有声明,从我们的实验中收集到的参试者的判断应该直接转

化为AV事故的法律规则。然而,他们的判断应该在公共讨论和法律法规中被适当地体现和管理。

即使自动驾驶汽车最终比人类驾驶员更安全,它们仍然会造成道路事故,从而引起人们对谁或什么应该负责的担忧。来自AV行业和立法条款的某些声音可以减轻消费者因使用无人驾驶汽车而产生的法律担忧。例如,沃尔沃[31]和奥迪[32]等一些汽车制造商已经公开承诺,他们将对汽车造成的事故承担全部责任;英国《2018年自动和电动汽车法案》规定,保险公司应对其造成的损害负责,而不是车主(如果使用得当)[83]。

然而,我们观察到的“困境”是,尽管L5 AV用户(L5车主和出租车乘客)没有直接的车辆控制权,但社会公众可能要求他们对这些车辆造成的伤害承担部分责任——我们7个实验的结果可能表明这一未来可能性。我们不认为这是“天真的”参试者的归因错误。相反,它可能表明了责任归属方面对无人驾驶汽车用户存在的社会偏见。先前的实证研究发现,公众[72,84-85]和法官[86]更严厉地评估AV造成的事故(相对于相同的人为造成的事故),并更严格地指责他们,因此表现出对AV使用及其在责任归属方面的偏见。更广泛地说,它可能反映了使用人类无法控制的自主机器的负面社会信号[87]。

根据我们的发现,具体的法律法规要求L5 AV汽车的车主或用户对其车辆造成的碰撞事故承担责任,即使事故超出了他们的控制范围。最近,深圳市发布了首个此类法规,以填补自动驾驶汽车事故的法律空白,并明确了自动驾驶汽车事故的责任规则[88]。它规定,如果驾驶座上有驾驶员,且车辆处于自动驾驶状态,由驾驶人承担赔偿责任;如果没有驾驶员在驾驶座上,则车辆所有人或使用人承担责任(向受害人赔偿);如果是车辆存在缺陷造成事故的,车辆的所有人或驾驶人在支付赔偿后,可以向车辆制造商或销售方请求赔偿。该规定对制造商很友好,但对所有人或使用人不太友好。根据该规定,L5 AV的所有人或使用人仍然是责任主体;然而,我们来自中国和韩国的参试者要求他们承担部分或次要责任(即他们不是主要责任方)。

尽管我们的主要研究结果和深圳的现行法律规定[88]都支持这样一种观点,即如果这些无人驾驶汽车造成伤害,L5 AV车主或用户并没有免除责任,但我们认为,这种责任认定方案可能适得其反。我们的研究结果表明了采用无人驾驶汽车的潜在心理层面障碍:当发生车祸时,无人驾驶汽车的消费者可能面临更大的公众压力和道德谴责。法官很可能会有类似于社会公众的反应[86]。因此,在公共讨论和法律责任认定中,社会公众可能会将无人驾

驶汽车的消费者视为“道德缓存区”[22]或“法律海绵”[23],将责任从无人驾驶汽车制造商和(或)运营商身上转移出来。指责消费者可能会阻止他们拥有或使用这些全自动无人驾驶汽车,给其未来蒙上阴影。需要采取如公共讨论和保险[18,72,80]等具体措施,以减少碰撞事故发生后对无人驾驶汽车消费者的社会指责,并减轻他们的责任负担。

7.3. 研究局限

鉴于我们研究的理论和实际意义,强调其局限性是至关重要的。首先,与其他基于情景的实验类似,我们基于情景的实验可能缺乏足够的心理现实性[18]。然而,他们模仿了人们在未来几十年与AV事故的互动方式[55]:“通过在新闻或社交媒体上听到有关他们的报道。”实验情景法在行为科学和社会科学中得到了广泛的应用和充分的研究。Zhai等[51]最近分析了一起实际的AV事故(即2018年Uber AV事故)后公众的责任判断,然后在一个基于情景的实验中研究了这次事故的更多变种。这两种方法产生了类似的研究结果,即公众认为Uber测试驾驶员比该公司(即Uber)承担了更多的责任。他们的研究为情景实验研究的外部效度提供了证据。

其次,责任判断通常发生在比我们工作中使用的精心控制的场景更丰富的背景下,受到更多的背景因素的影响,如新闻报道中表达的情感细节和立场[18]。再次,我们的研究结果可能有文化上的局限性,因为我们是从亚洲两个国家获得的数据。未来的研究应该在不同的文化背景下研究这一问题。最后,我们不能排除除了合理预见性之外的其他因素(如乘员“接近”事故,参试者对无人驾驶汽车相较于传统汽车的较低信任度)也可能解释人们对无人驾驶车辆用户的责任判断倾向。

8. 结论

我们的研究重点关注全自动无人驾驶汽车,揭示了这样一个困境:全自动驾驶汽车的用户(全自动汽车的车主和无人驾驶出租车的乘客)与传统出租车中的乘客相比,当这些车辆造成相同的伤害时,人们要求他们承担更多责任(虽然所有这些乘员都没有直接控制驾驶)。这种责任归因差异的部分原因是,人们期望全自动驾驶汽车用户对使用全自动驾驶汽车的事故后果有更多的可预见性(合理的可预见性)。这一发现表明,无人驾驶汽车用户在碰撞事故发生后可能面临社会指责。公共讨论、保险和其他社会措施应侧重于减少对无人驾驶汽车用户造成

的社会指责和责任负担。

致谢

本研究得到国家自然科学基金项目(72071143)的资助。

Compliance with ethics guidelines

Siming Zhai, Lin Wang, and Peng Liu declare that they have no conflicts of interest or financial conflicts to disclose.

Appendix A. Supplementary data

All data, codes, and results are publicly available on the *Open Science Framework* at https://osf.io/58s42/?view_only=6b5b8d4bade449a8a0f8d0cc8836ac57.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.eng.2023.10.008>.

References

- [1] Jamjoom AAB, Jamjoom AMA, Marcus HJ. Exploring public opinion about liability and responsibility in surgical robotics. *Nat Mach Intell* 2020; 2(4): 194-6.
- [2] Bigman YE, Waytz A, Alterovitz R, Gray K. Holding robots responsible: the elements of machine morality. *Trends Cogn Sci* 2019;23(5):365-8.
- [3] Rahwan I, Cebrian M, Obradovich N, Bongard J, Bonnefon JF, Breazeal C, et al. Machine behaviour. *Nature* 2019;568(7753):477-86.
- [4] Yang GZ, Cambias J, Cleary K, Daimler E, Drake J, Dupont PE, et al. Medical robotics—regulatory, ethical, and legal considerations for increasing levels of autonomy. *Sci Robot* 2017;2(4):eaam8638.
- [5] Fagnant DJ, Kockelman K. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transp Res Policy Pract* 2015;77:167-81.
- [6] Wang J, Huang H, Li K, Li J. Towards the unified principles for level 5 autonomous vehicles. *Engineering* 2021;7(9):1313-25.
- [7] Dingus TA, Guo F, Lee S, Antin JF, Perez M, Buchanan-King M, et al. Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proc Natl Acad Sci USA* 2016;113(10):2636-41.
- [8] Wang X, Liu Q, Guo F, Fang S, Xu X, Chen X. Causation analysis of crashes and near crashes using naturalistic driving data. *Accid Anal Prev* 2022;177: 106821.
- [9] National Highway Traffic Safety Administration (NHTSA). Automated vehicles for safety [Internet]. Washington, DC: National Highway Traffic Safety Administration; 2020 [cited 2020 October 18]. Available from: <https://shorturl.at/mty10>.
- [10] Society of Automotive Engineers (SAE). J3016. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Washington, DC: SAE International/ISO; 2021.
- [11] Fosch-Villaronga E, Khanna P, Drukarch H, Custers BHM. A human in the loop in surgery automation. *Nat Mach Intell* 2021;3(5):368-9.
- [12] Bonnefon JF, Černý D, Danaher J, Devillier N, Johansson V, Kovacicova T, et al. Ethics of connected and automated vehicles: recommendations on road

- safety, privacy, fairness, explainability and responsibility. Brussels: EU Commission; 2020.
- [13] Bonnefon JF, Shariff A, Rahwan I. The social dilemma of autonomous vehicles. *Science* 2016;352(6293):1573-6.
- [14] Pattinson JA, Chen H, Basu S. Legal issues in automated vehicles: critically considering the potential role of consent and interactive digital interfaces. *Humanit Soc Sci Commun* 2020;7(1):153.
- [15] Marchant G, Lindor R. The coming collision between autonomous vehicles and the liability system. *Santa Clara Law Rev* 2012;52(4):1321-40.
- [16] Liu P, Du M, Li T. Psychological consequences of legal responsibility misattribution associated with automated vehicles. *Ethics Inf Technol* 2021;23(4):763-76.
- [17] Stilgoe J. Self-driving cars will take a while to get right. *Nat Mach Intell* 2019;1(5):202-3.
- [18] Awad E, Levine S, Kleiman-Weiner M, Dsouza S, Tenenbaum JB, Shariff A, et al. Drivers are blamed more than their automated cars when both make mistakes. *Nat Hum Behav* 2020;4(2):134-43.
- [19] Hancock PA, Nourbakhsh I, Stewart J. On the future of transportation in an era of automated and autonomous vehicles. *Proc Natl Acad Sci USA* 2019;116(16):7684-91.
- [20] A tragic loss [Internet]. Austin: Tesla; 2016 Jun 30 [cited 2019 May 1]. Available from: <https://www.tesla.com/blog/tragic-loss>.
- [21] McFarland M. Uber self-driving car operator charged in pedestrian death [Internet]. Atlanta: CNN; 2020 [cited 2021 Jan 23]. Available from: <https://rb.gy/hkskb>.
- [22] Elish MC. Moral crumple zones: cautionary tales in human-robot interaction. *Engag Sci Technol Soc* 2019;5:40-60.
- [23] Holford WD. An ethical inquiry of the effect of cockpit automation on the responsibilities of airline pilots: dissonance or meaningful control? *J Bus Ethics* 2022;176(1):141-57.
- [24] Geistfeld MA. A roadmap for autonomous vehicles: state tort liability, automobile insurance, and federal safety regulation. *Calif LRev* 2017;105(6):1611.
- [25] Grieman K. Hard drive crash: an examination of liability for self-driving vehicles. *J Intell Prop Info Tech Elec Com L* 2018;9(3):294-309.
- [26] Mackie T. Proving liability for highly and fully automated vehicle accidents in Australia. *Comput Law Secur Rev* 2018;34(6):1314-32.
- [27] Vladeck DC. Machines without principals: liability rules and artificial intelligence. *Wash Law Rev* 2014;89(1):117-30.
- [28] Hevelke A, Nida-Rümelin J. Responsibility for crashes of autonomous vehicles: an ethical analysis. *Sci Eng Ethics* 2015;21(3):619-30.
- [29] Marchant G, Bazzi R. Autonomous vehicles and liability: what will juries do? *B U J Sci Tech L* 2020;26(1):67-119.
- [30] Gurney JK. Imputing driverhood: applying a reasonable driver standard to accidents caused by autonomous vehicles. In: Lin P, Abney K, Jenkins R, editors. *Robot Ethics 20: from autonomous cars to artificial intelligence*. Oxford: Oxford University Press; 2017. p. 51-65.
- [31] Atiyeh C. Volvo will take responsibility if its self-driving cars crash [Internet]. Harlan: Car and Driver; 2015 [cited 2019 May 1]. Available from: <https://shorturl.at/cIRW9>.
- [32] Maric P. Audi to take full responsibility in event of autonomous vehicle crash [Internet]. Drive; 2017 Sep 11 [cited 2019 May 1]. Available from: <https://shorturl.at/pEPTX>.
- [33] Lima G, Grgić-Hlača N, Cha M. Human perceptions on moral responsibility of AI: a case study in AI-assisted bail decision-making. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*; 2021 May 8-13; Yokohama, Japan; 2021.
- [34] van de Poel I. Moral responsibility. In: van de Poel I, Royakkers L, Zwart SD, editors. *Moral responsibility and the problem of many hands*. New York City: Routledge; 2015.
- [35] Cushman F. Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition* 2008;108(2):353-80.
- [36] Weiner B. Social motivation, justice, and the moral emotions: an attributional approach. Mahwah: Lawrence Erlbaum Associates; 2006.
- [37] Aristotle. *Nicomachean ethics*. In: Barnes J, editor. *The complete works of Aristotle*. Princeton: Princeton University Press; 1984.
- [38] Johnson DG. Technology with no human responsibility? *J Bus Ethics* 2015;127(4):707-15.
- [39] Fischer JM, Ravizza M. *Responsibility and control: a theory of moral responsibility*. Cambridge: Cambridge University Press; 1998.
- [40] Nelkin DK. Moral luck. In: Zalta EN, editor. *The Stanford encyclopedia of philosophy*. Stanford: Stanford University; 2004.
- [41] Nagel T. *Mortal questions*. Cambridge: Cambridge University Press; 1979.
- [42] Williams B. *Moral luck*. Cambridge: Cambridge University Press; 1981.
- [43] Howe JST. Towards a control-centric account of tort liability for automated vehicles. *Torts Law J* 2021;26(3):221-43.
- [44] Huddy XP. *The law of automobiles*. 6th ed. Albany: Matthew Bender; 1922.
- [45] Rahwan I. Society-in-the-loop: programming the algorithmic social contract. *Ethics Inf Technol* 2018;20(1):5-14.
- [46] Copp CJ, Cabell JJ, Kimmelmeier M. Plenty of blame to go around: attributions of responsibility in a fatal autonomous vehicle accident. *Curr Psychol* 2023;42(8):6752-67.
- [47] Bennett JM, Challinor KL, Modesto O, Prabhakaran P. Attribution of blame of crash causation across varying levels of vehicle automation. *Saf Sci* 2020;132:104968.
- [48] McManus RM, Rutchick AM. Autonomous vehicles and the attribution of moral responsibility. *Soc Psychol Personal Sci* 2019;10(3):345-52.
- [49] Pöllänen E, Read GJM, Lane BR, Thompson J, Salmon PM. Who is to blame for crashes involving autonomous vehicles? Exploring blame attribution across the road transport system. *Ergonomics* 2020;63(5):525-37.
- [50] Li J, Cho MJ, Zhao X, Ju W, Malle BF. From trolley to autonomous vehicle: perceptions of responsibility and moral norms in traffic accidents with selfdriving cars. In: *SAE 2016 World Congress and Exhibition*; 2016 Apr 12-14; Detroit, MI, USA; 2016.
- [51] Zhai S, Gao S, Wang L, Liu P. When both human and machine drivers make mistakes: whom to blame? *Transp Res Policy Pract* 2023;170:103637.
- [52] Zhai S, Wang L, Liu P. Human and machine drivers: sharing control, sharing responsibility. *Accid Anal Prev* 2023;188:107096.
- [53] Aguiar F, Hannikainen IR, Aguiar P. Guilt without fault: accidental agency in the era of autonomous vehicles. *Sci Eng Ethics* 2022;28(2):11.
- [54] Jahedi S, Méndez F. On the advantages and disadvantages of subjective measures. *J Econ Behav Organ* 2014;98:97-114.
- [55] Hidalgo CA, Orghian D, Canals JA, de Almeida F, Martin N. *How humans judge machines*. Cambridge: MIT Press; 2021.
- [56] Franklin M, Ashton H, Awad E, Lagnado D. Causal framework of artificial autonomous agent responsibility. In: *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*; 2022 Aug 1-3; Oxford, UK; 2022. p. 276-84.
- [57] Malter MS, Kim SS, Metcalfe J. Feelings of culpability: just following orders versus making the decision oneself. *Psychol Sci* 2021;32(5):635-45.
- [58] Shaver KG. Defensive attribution: effects of severity and relevance on the responsibility assigned for an accident. *J Pers Soc Psychol* 1970;14(2):101-13.
- [59] Bradley GW. Self-serving biases in the attribution process: a reexamination of the fact or fiction question. *J Pers Soc Psychol* 1978;36(1):56-71.
- [60] Jörling M, Böhm R, Paluch S. Service robots: drivers of perceived responsibility for service outcomes. *J Serv Res* 2019;22(4):404-20.
- [61] Palamar M, Le DT, Friedman O. Acquiring ownership and the attribution of responsibility. *Cognition* 2012;124(2):201-8.
- [62] Shaver KG. *The attribution of blame: causality, responsibility, and blameworthiness*. New York City: Springer; 1985.
- [63] Alicke MD. Culpable control and the psychology of blame. *Psychol Bull* 2000;126(4):556-74.
- [64] Lagnado DA, Channon S. Judgments of cause and blame: the effects of intentionality and foreseeability. *Cognition* 2008;108(3):754-70.
- [65] Heider F. *The psychology of interpersonal relations*. New York City: John Wiley & Sons; 1958.
- [66] Furlough C, Stokes T, Gillan DJ. Attributing blame to robots: I. the influence of robot autonomy. *Hum Factors* 2021;63(4):592-602.
- [67] Khoury L, Smyth S. Reasonable foreseeability and liability in relation to genetically modified organisms. *Bull Sci Technol Soc* 2007;27(3):215-32.
- [68] Nuijten MB, Hartgerink CHJ, van Assen MALM, Epskamp S, Wicherts JM. The prevalence of statistical reporting errors in psychology (1985-2013). *Behav Res Methods* 2016;48(4):1205-26.
- [69] Brysbaert M. How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *J Cogn* 2019;2(1):16.
- [70] Faul F, Erdfelder E, Lang AG, Buchner AG. *Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007;39(2):175-91.
- [71] Diels C, Bos JE. Self-driving carsickness. *Appl Ergon* 2016;53:374-82.
- [72] Liu P, Du Y. Blame attribution asymmetry in human-automation cooperation. *Risk Anal* 2022;42(8):1769-83.
- [73] Kirchkamp O, Strobel C. Sharing responsibility with a machine. *J Behav Exp Econ* 2019;80:25-33.

- [74] Banks J. A perceived moral agency scale: development and validation of a metric for humans and social machines. *Comput Human Behav* 2019; 90: 363-71.
- [75] McCaul KD, Veltum LG, Boyechko V, Crawford JJ. Understanding attributions of victim blame for rape: sex, violence, and foreseeability. *J Appl Soc Psychol* 1990;20(1):1-26.
- [76] Hayes AF, Preacher KJ. Statistical mediation analysis with a multicategorical independent variable. *Br J Math Stat Psychol* 2014;67(3):451-70.
- [77] McShane BB, Böckenholt U. Meta-analysis of studies with multiple contrasts and differences in measurement scales. *J Consum Psychol* 2022;32(1):23-40.
- [78] Weiner B. Attributional thoughts about consumer behavior. *J Consum Res* 2000; 27(3):382-7.
- [79] Shavell S. On the redesign of accident liability for the world of autonomous vehicles. *J Legal Stud* 2020;49(2):243-85.
- [80] Shariff A, Bonnefon JF, Rahwan I. Psychological roadblocks to the adoption of self-driving vehicles. *Nat Hum Behav* 2017;1(10):694-6.
- [81] Karnow CEA. The application of traditional tort theory to embodied machine intelligence. In: Calo R, Froomkin AM, Kerr I, editors. *Robot law*. Northampton: Edward Elgar; 2016.
- [82] Van Uytsel S. Different liability regimes for autonomous vehicles: one preferable above the other? In: Van Uytsel S, Vasconcellos Vargas D, editors. *Autonomous vehicles: business, technology and law*. Singapore: Springer; 2021.
- [83] Commission L. Automated vehicles: joint report. UK: Law Commission of England and Wales and Scottish Law Commission; 2022.
- [84] Liu P, Du Y, Xu Z. Machines versus humans: people's biased responses to traffic accidents involving self-driving vehicles. *Accid Anal Prev* 2019; 125: 232-40.
- [85] Franklin M, Awad E, Lagnado D. Blaming automated vehicles in difficult situations. *iScience* 2021;24(4):102252.
- [86] Rachlinski JJ, Wistrich AJ. Judging autonomous vehicles. *Yale J Law Technol* 2022;24:706-66.
- [87] Bigman YE, Gray K. People are averse to machines making moral decisions. *Cognition* 2018;181:21-34.
- [88] Ma S. Shenzhen gives green light to fully autonomous vehicles [Internet]. Beijing: China Daily; [updated 2022 Jul 11; cited 2022 Aug 1]. Available from: <https://shorturl.at/mvO18>.