Advances in Psychology 心理学进展, 2020, 10(11), 1842-1846
Published Online November 2020 in Hans. http://www.hanspub.org/journal/ap
https://doi.org/10.12677/ap.2020.1011214

Hans汉斯

人机交互中的过度信任

陈嘉乐，朱孟婷，李永娜*

中国人民大学应用心理学系，北京
Email: 1060928311@qq.com, 18631855120@163.com, *cogpsyli@ruc.edu.cn

收稿日期：2020年10月26日；录用日期：2020年11月23日；发布日期：2020年11月30日

摘　要

随着科技的发展，基于人工智能的自动化系统在人类世界变得越来越重要，人们渐渐将机器人视为自己的工作或生活伙伴，在与机器人的互动中也会表现出类似人际互动的行为与态度，如认为机器人是值得信任的。由于人们会在冲突情境中遵从自动化技术的倾向，从而产生了对机器人的过度信任。基于对人机过度信任的理解不同，研究者们在研究人机之间是否存在过度信任和过度信任的产生及其影响因素时，采用了不同的操作定义。本文总结了这些操作定义，并且在此基础之上，讨论了人机交互过程中的过度信任的影响因素和减少过度信任的方法。

关键词

人机交互，人机信任，过度信任，自动化偏爱

The Overtrust in Human-Robot Interaction

Jiale Chen, Mengting Zhu, Yongna Li*

Department of Psychology, Renmin University of China, Beijing,
Email: 1060928311@qq.com, 18631855120@163.com, *cogpsyli@ruc.edu.cn

Received: Oct. 26th, 2020; accepted: Nov. 23rd, 2020; published: Nov. 30th, 2020

Abstract

With the development of science and technology, the automation systems based on artificial intelligence become more and more important in the human world. People gradually regard robots as partners in work and life. Therefore, they show similar behaviors and attitudes in the human-robot interaction as in interpersonal interaction. People believe that they can trust robots. Due to the automation bias in situation where there is conflict exists, overtrust occurs in hu-

*通讯作者。

文章引用: 陈嘉乐, 朱孟婷, 李永娜(2020). 人机交互中的过度信任. *心理学进展, 10(11),* 1842-1846.
DOI: 10.12677/ap.2020.1011214

man-robot interaction. Researchers chose different operational definitions of human-robot overtrust when investigating whether overtrust occurs in human-robot interaction and what factors would influence the occurrence of overtrust. The current work summarized the operational definitions used in previous research, further discussed the influential factors of overtrust, and proposed how to eliminate overtrust.

Keywords

Human-Robot Interaction, Human-Robot Trust, Overtrust, Automation Bias

1. 引言

科技的发展极大地改变了人们的生活，尤其是自动化技术的进步使得人们从繁重的体力劳动甚至是某些脑力劳动中获得解放。基于人工智能的各种自动化系统，尤其是各种机器人在人类世界扮演着越来越多、越来越重要的角色。人们逐渐把自动化系统或者机器人当成自己的工作或者生活的伙伴，而不是把它们看成是单纯的工具。在与机器人的互动过程中，人们会表现出很多类似人际互动中的行为与态度，会假设机器人是可以信任的，它们能够有效安全地执行相应的功能。由于自动化偏爱(automation bias)的倾向，人们很容易对机器人产生过度信任。自动化偏爱倾向指的是，在面对相互矛盾的信息的时候，人们更倾向于遵从自动化技术(Mosier, Palmer, & Degani, 1992)。过度信任是技术盲从的一种形式。对机器人的过度信任可能发生在个体对某个行为带来的风险有误判的情况下。之所以会产生风险误判，是因为个体要么低估了信任破坏所带来的损失，或者是低估了机器人会破坏信任的概率；要么二者都低估了(Wagner, Borenstein, & Howard, 2018)。人机交互过程中的过度信任，使得人们开始容忍通常情况下不会接受的风险，从而威胁到个体的健康和安全。因此，它变成了计算伦理中的一个核心话题。

2. 人机信任与过度信任的概念

信任是建立与维持人际互动的基础。很多研究者都认为，信任的产生基于这样一种情形，即个体会因为他人的动作、行为或者动机而容易受到伤害。冒着受伤害的风险而接受他人的互动，是信任的本质。Wagner 等人将人机信任界定为一种信念，是在冒着承担行为消极结果的风险的情形下，信任者认为被信任者会选择降低风险的方式来做出相应行为的信念(Wagner, Robinette, & Howard, 2018)。

虽然对人机信任的理解有很多共识，但关于人对机器人的过度信任，研究者依然会各持己见。基于对工厂自动化的研究，Lee and See 首先提出，过度信任是超出了自动化系统功能范围的信任，其结果可能是导致流水线的崩溃，而过度信任自动化流水线的人是不会有风险的。机器人与工厂的自动化流水线有很大的区别，主要体现在它与人关联的密切程度更高，人们如果过度信任机器人可能导致机器人的滥用，从而对人身安全产生威胁。所以，过度信任应该被理解为这样一种情形，个体因为相信机器人能够行使其本身并不具备的功能而承受风险；或者个体因为预期机器人会减少风险而承受更多的风险(Wagner, Borenstein, & Howard, 2018)。

有很多父母对自动化的技术过于信任而让他们的孩子处于危险境地。Borenstein 等人为了探讨对保健机器人的过度信任，进行了一个问卷调查。调查对象是身体有运动方面残疾的孩子的父母，询问他们对

于一种保健机器人即自动化外置机器骨骼的看法。结果显示，大部分的父母对自动化技术是不担忧的，之前都或多或少给孩子使用过类似的自动化设备。虽然有的父母会担心孩子的安全问题，但在意识到孩子可能会做出一些危险动作的前提下，父母们依然相信他们的孩子应该使用这种外置的机器骨骼(Borenstein, Wagner, & Howard, 2018)。这个研究指出，对机器骨骼的过度信任可能是因为虽然父母对机器人技术很熟悉，但他们并没有经验来估计使用机器人的相关风险。

3. 过度信任的操作定义

由于对人机交互过程中过度信任的理解不同，研究者们在实证研究中会采用不同的操作定义。Itoh (2012)考察了人们是否会对自动驾驶系统过度信任。实验中采用的是高级驾驶辅助系统(Advanced Driving Assist System)中的适用性巡航控制系统(Adaptive Cruise Control, ACC)。ACC 系统可以在车辆行驶过程中使用安装在车辆前部的车距传感器持续扫描车辆前方道路情况，同时通过轮速传感器采集车速信号。当车辆与前车之间的距离过小时，ACC 系统可以使发动机的输出功率降低，并且控制刹车系统使车轮适当制动，从而与前方车辆始终保持安全距离。ACC 系统对于减少驾驶员的疲劳很有用，可以适当减少追尾事故的发生。但是，常规的 ACC 系统不会对道路上静止的物体做出反应，所以在交通拥堵或者红灯停车的时候，ACC 系统无法对前方停止的车辆做出制动反应，反而会发生追尾。实验中设计了不同的道路情况，主要包括了前方有静止车辆的路况，要求参与实验的人操纵安装了 ACC 系统的模拟驾驶器，考察在前方有静止车辆的时候是否会及时采取人工制动。实验开始的时候，被试都试用 ACC 系统，产生信任，然后再切换不同的路况。结果发现，被试在看到前方有静止的车辆的情况下也没有及时切换成人工制动，而是一直使用 ACC 系统，导致了追尾的发生。说明了被试对 ACC 系统存在着一定的过度信任。

Itoh (2012)对过度信任的操作化符合了 Borenstein 等人(2018)的界定，即个体因为相信机器人能够行使其本身并不具备的功能而承受风险是过度信任的一种形式。虽然 ACC 系统不能在前方有静止车辆的情况下自动减速刹车，但人们却相信它在任何情况下都会采取自动制动，从而将自己置身于发生交通事故的危险之中。

过度信任的另一种操作化的方法就是对机器人的盲从。Robinette 等人(2016)考察了在紧急情况发生时，人们是否会对机器人产生过度信任。他们设计了一个简单的建筑物内部的场景，包括几个房间和走廊，使用人造烟雾和烟雾探测器模拟发生火灾的紧急情况，让机器人作为安全逃生的向导。一共有两种条件：一种是机器人按照最直接简短的路径引导人们走到出口的位置，即高效的条件；另一种是在导航的途中，机器人会进入一个空房间并在里面转两个圈，然后再走向出口，即迂回的条件。如果人们在迂回的条件下自己选择了直接的路径，而没有跟随机器人的引导，说明是不存在过度信任的。结果是所有参与者都按照机器人的指示走，即使是机器人在行进的过程当中出现了短暂的故障，或者是机器人引导参与者进入塞满家具的黑暗房间，参与者仍然是跟随机器人的路线。Christensen 等人(2019)重复了这个结果，并且发现即使机器人发出语音警告，说明机器人的感觉系统出现故障，仍然有参与者跟随机器人的指导去寻找出口。

在紧急情况发生时，个体会明显地感觉到自己置身于危险之中，但他们相信机器人能够减少风险，根据 Borenstein 等人(2018)的概念，这也是一种过度信任。由于采用的实验情境都是紧急情况，而人在紧急情况下的心理状况与日常其他情境中可能是不同的，紧急情况下人们会信任任何可以降低他们的不确定性和焦虑感的信息源。所以，是否可以把紧急逃生过程中对机器人的盲从看作是过度信任，其实是值得讨论的。如果是非紧急情况下，人们还是盲从机器人，可能是支持过度信任的更好的证据。虽然实证研究做的不多，但现实生活中存在这样的现象，新闻中会报道有人按照导航系统将车开进了河里或者开到了火车站广场，被救援之后依然会打开导航系统继续行驶。

除了以上两种操作定义之外，Booth 等人(2017)采用了人们是否会给机器人打开门禁系统来测量过度信任。他们选择的是现实生活中的场景，实验是在一个大学有门禁系统的宿舍楼进行的。学校规定只有特定人员才允许进入到宿舍楼，陌生人是不能进的，不然会给学生们带来财产或者人身安全方面的问题。Booth 等人将一个机器人放在等待进宿舍的队伍中，然后观察有多少人会打开门禁系统，允许机器人进入宿舍。结果发现，有相当一部分人会将机器人带入宿舍内。作者认为这反映了学生们对机器人的过度信任，因为作者会推断说学生应该清楚机器人进入宿舍可能会对里面的学生造成安全威胁，但依然会让机器人进入。问题是，那些允许机器人进入宿舍的学生是否真正意识到了机器人的潜在危险性，由于没有提供事后访谈的资料，是无法知道学生们当时的想法的。学生们让机器人进入的原因可能五花八门，也可能在那一刻并没有意识到机器人会带来威胁。Booth 等人的研究与其说是探讨了过度信任，不如说是探讨了对机器人的信任，因为在这个场景中学生们既没有高估机器人的功能，也不可能产生机器人会降低已存在风险的预期。

4. 小结

研究者对过度信任的日益关注，不仅仅是为了说明人机互动过程中是否会存在过度信任，以及有哪些因素会影响过度信任，最终目的是为了探讨如何减少或者消除过度信任，以保护人们的安全。已有的研究表明，机器人发生故障，偏离路径、语音提示发生故障都不能显著减少人们的过度信任。

Wagner，Borenstein 和 Howard (2018)提出了几个减少过度信任的建议。第一个是机器人设计的不能太拟人化，他们认为拟人化的机器人会使人产生一种熟悉性，从而导致交互过程中的更多风险。第二个是机器人应该具有识别人的情绪、注意状态，并以此来推测人们行为的功能，机器人最好也能够具有识别使用者是否是儿童或者是否具有心理障碍或者身体残疾的能力。第三个就是机器人的功能应该是显而易见的，这样使用者就容易知道机器人没有哪些功能，从而避免过度信任。这三个建议是从机器人设计的角度提出的。

过度信任的产生应该受到三个方面因素的影响。一个方面是关于机器人的因素，一个方面是关于人的因素，还有一个方面是关于互动情境的因素。因为目前没有太多的实验研究证明哪个方面的因素是更重要的，所以单从一个方面来考虑可能是不完整的。Wagner，Borenstein 和 Howard (2018)的建议的合理性也值得探讨。首先，虽然人们会觉得高度拟人化的机器人相比机械机器人更不可信(Zlotowski et al., 2016)，但 Itoh (2012)已经表明了人们对非拟人化的 ACC 系统也会产生过度信任。其次，目前的技术可能没有办法设计可以跨情境准确识别人的情绪状态，并可以准确预测行为的机器人，即使有的机器人产品宣传可以进行情绪识别，但其灵活性与准确性也存在问题。另一方面，如果机器人可以做到识别人的情绪与注意状态，人们可能对机器人的功能更加盲目崇拜，结果是增加了过度信任。最后，人们清晰地了解了机器人的功能，是否就会减少过度信任的产生？答案不是肯定的，可能与人的人格特点有关系，比如比较谨慎的人会提醒自己机器人的功能界限，但比较开放的人可能不会这么做。

人的因素在过度信任的产生中起到的作用也是很重要的。未来的研究需要对过度信任的影响因素做进一步的探索，可以明确机器人、人与情境因素的贡献是多少；也需要从人的角度出发，考察不同年龄阶段的人产生的过度信任是否有差异，增加人机交互的经验是否可以让人对机器人的功能有更清晰的了解，从而减少过度信任。此外，还要探讨创设可以发挥人的主动性的人机交互情境的作用。如果一个情境中个体没有自由选择权，只能依赖机器人或者自动化系统，就会有效地培养过度信任，带来的风险可想而知。所以，要求出租车司机都按照导航系统的指示行驶，从人机互动的角度看，不是一个好的政策。归根结底，减少过度信任取决于人的积极性与主动性，也取决于人对自己判断的信心。不管自动化系统将来会发展的多么高级，只要人们对自己的信心还在，只要在人机互动中没有完全放弃自我，就可以减

少过度信任，保护自己和他人不会受到严重的伤害。

基金项目

本研究是中国人民大学科学研究基金(中央高校基本科研业务费专项资金资助)项目成果，基金号：19XNB033。

参考文献

Booth, S., Tompkin, J., Pfister, H., Waldo, J., Gajos, K., & Nagpal, R. (2017). Piggybacking Robots: Human-Robot Overtrust in University Dormitory Security. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, Vienna, 6-9 March 2017, 426-434. https://doi.org/10.1145/2909824.3020211

Borenstein, J., Wagner, A. R., & Howard, A. (2018). Overtrust of Pediatric Health-Care Robots: A Preliminary Survey of Parent Perspectives. *IEEE Robotics & Automation Magazine, 25,* 46-54. https://doi.org/10.1109/MRA.2017.2778743

Christensen, A. B., Dam, C. R., Rasle, C., Bauer, J. E., Mohamed, R. A., & Jensen, L. C. (2019). Reducing Overtrust in Failing Robotic Systems. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Daegu, 11-14 March 2019, 542-543. https://doi.org/10.1109/HRI.2019.8673235

Itoh M. (2012). Toward Overtrust-Free Advanced Driver Assistance Systems. *Cognition, Technology & Work, 14,* 51-60. https://doi.org/10.1007/s10111-011-0195-2

Mosier, K. L., Palmer, E. A. & Degani, A. (1992). Electronic Checklists: Implications for Decision Making. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 36,* 7-11. https://doi.org/10.1177/154193129203600104

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016). Overtrust of Robots in Emergency Evacuation Scenarios. *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, Christchurch, New Zealand, 3 July 2016-3 October 2016, 101-108. https://doi.org/10.1109/HRI.2016.7451740

Wagner, A. R., Borenstein, J., & Howard, A. (2018). Overtrust in the Robotic Age. *Communications of the ACM, 61,* 22-24. https://doi.org/10.1145/3241365

Wagner, A. R., Robinette, P., & Howard, A. (2018). Modeling the Human-Robot Trust Phenomenon: A Conceptual Framework Based on Risk. *ACM Transactions on Interactive Intelligent Systems (TiiS), 8,* Article No. 26. https://doi.org/10.1145/3152890

Zlotowski, J., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., & Ishiguro, H. (2016). Appearance of a Robot Affects the Impact of Its Behavior on Perceived Trustworthiness and Empathy. *Paladyn, 7,* 55-66. https://doi.org/10.1515/pjbr-2016-0005