

Exploring the Influencing Factors of Human-Robot Trust from the Perspective of Interpersonal Trust

Yunxia Cao

Department of Psychology, Renmin University of China, Beijing
Email: caoyx2015@ruc.edu.cn

Received: Apr. 7th, 2018; accepted: Apr. 21st, 2018; published: Apr. 28th, 2018

Abstract

Trust plays an important role in the process of interpersonal interaction. Is there any difference between the trust mechanism of human and AI system and interpersonal trust? Research shows that the performance of the AI system itself (such as reliability, false alarm rate and failure rate, etc.) and attributes (such as appearance, physical contact) and the different cultural backgrounds of human beings all affect the establishment of human-machine trust. Fortunately, these factors that affect human-machine trust, such as performance and attributes, can be promoted by human design.

Keywords

Human-Machine Trust, Interpersonal Trust

从人际信任的角度探究人机信任的影响因素

曹云霞

中国人民大学心理系, 北京
Email: caoyx2015@ruc.edu.cn

收稿日期: 2018年4月7日; 录用日期: 2018年4月21日; 发布日期: 2018年4月28日

摘要

信任在人际交互过程中起着重要作用, 而人与人工智能系统的信任机制是否与人际信任有区别? 研究表明, 人工智能系统本身的性能(如可靠性、误警率和故障率等)和属性(如外观, 身体接触)以及人所处的

不同文化背景，均会影响人-机信任的建立。幸运的是，这些影响人机信任的因素，如性能和属性可以由人设计而提升。

关键词

人-机信任，人际信任

Copyright © 2018 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

人机交互(Human Robot Interface, HRI)是以特定的语言和交互方式，完成人工智能系统与用户之间信息交换的过程。随着人工智能技术的发展和应用，人与人工智能系统的交互成为如今研究的热点话题，而信任在人机交互良好互动中扮演着重要角色[1]。虽然人机交互领域的专家对发生在用户和机器之间的交互进行了大量研究，但人工智能系统是不同于机器的特殊系统，在有些方面，人工智能系统与人类更类似。由此，人际信任可以为人机信任提供一些见解和启示。本文从参考人际信任研究的角度出发，回顾人机信任相关研究，探索影响人机交互之间信任的因素，希望为提高人机信任给出一些建议和参考。

2. 人际信任与人机信任

信任由于其概念的抽象性和结构复杂性，在社会学、心理学、营销学、经济学、管理学等不同的领域定义有所区别，但是达成共识的观点是：信任是涉及交易或交换关系的基础。人际信任是人际交互的重要条件，在人际互动中，信任帮助人们面对不确定性，适时采取决策和行动。信任他人意味着必须承受易受对方行为伤害的风险，因此，承担易受伤害之风险的意愿亦是人际信任之核心。与此同理，人机信任也涉及风险和决策。关于人际信任，我们已有大量的研究结果。衡量信任程度有不同的方法。如社会综合调查，可以询问人对他人的信任态度。心理学行为实验在预测和量化信任方面尤其有用。

虽然人工智能系统也可以提供建议、做出诊断、以及完成其他的任务，但是人们是否能够像信任他人一样信任人工智能系统？Parasuraman 等人表明，正如人际信任的方式，优化人机信任能够为人机工程学提供有效的解决方案[2]。Pan 等人[3]研究了不同媒体呈现方式在用户顾问关系中对信任行为的影响。比较了视频，头像，和机器人介导的互动。他们要求参与者尝试回答 31 个困难的知识问题，对于每个问题，参与者可以且只能向两位顾问中的一位咨询。顾问的建议是通过视频的方式呈现的，视频中有真人顾问，化身顾问和机器人顾问。结果显示不管是哪种顾问方式，参与者能够识别出专家的建议；但相比于化身顾问，参与者更倾向于向真人及机器人顾问寻求建议，而且，对二者的建议寻求意向没有差异，参与者的主观报告也进一步支持了行为结果。以上实验室研究说明了至少在某些情况下，人们可以像信任他人一样信任人工智能系统。

关于人际信任的实验室研究，或许可以为信任提供一些参考。Stirrat (2010)等认为，面部的非言语暗示对于信任有促进作用，稳定的面部特征，如吸引力、相似性和感知可信度是信任的重要影响因素[4]。实验研究表明，人本身的信息，如吸引力、性别和面部表情可以影响另一个人的信任感知[5]。一些研究表明，当提供视觉信息时，信任游戏中的被信任者会获得更多的信任。视频会议和面对面沟通，比

纯文本和电话通信,能更好地建立对社会困境的信任[6]。研究发现,身体接触对于人际信任也起着至关重要的促进作用,在谈判之前握手,可以得到更多的互利结果[7]。Jack 等研究发现,文化价值观与民族语境限制了人们的行为选择[8]。例如,在不同的文化中,情绪,手势都有不同的表达和接受规则[9]。若跨文化行为理论将从人际交互转移到人机交互,在不同的文化中也有所不同。以上研究表明,外观线索,身体接触,文化背景都是人际信任的影响因素。那么,人际信任是否可以为人机信任提供一些参考呢?

3. 外观影响人机信任?

人工智能系统有各种不同的载体形式,通常比较常见的是各种形状的机器人,也有研究者试图开发出具有特定年龄或性别的人类面孔的机器人[10]。这也就引出我们另一个问题,人类对机器人的信任与否,是否与机器人的外观有关?在人际信任中,被信任者的外貌特征会影响信任判断的准确性。进化理论和神经科学证据表明,人类根据面部外观可以自动推断出其他人的信任度。面孔是第一印象形成的重要信息来源,基于面孔做出信任评价在信任关系的建立中起着重要作用[11]。从进化的角度讲,快速而准确地依据对方面孔做出是否可信任的判断对于个体的趋利避害行为乃至生存发展都至关重要。有研究者指出,基于面孔的信任判断是一个快速且自动化的过程,在面孔呈现的 100 ms 左右即已发生,快于其他人格特征判断。在一个信任游戏中,发现被试可以根据陌生人的面孔照片来判断其是否欺骗,并且这一结果高于随机概率[12]。这说明单纯借助于面孔我们就可以得到信任判断所需要的线索信息。已有研究显示,眉脊的高低、颧骨的浅深和下巴的宽窄等结构特征与面孔信任评价密切相关[11]。那么,作为人工智能系统的载体的机器人的外观是否也会影响人-机信任?

人们确实会认为外观更像人的机器人更加友善,更加可靠,而机械机器人被认为更冷酷。Broadbent 等认为,外观可以深刻地影响机器和人类之间的相互作用,一个具有更人性化的脸部显示器的机器人被认为有更好的个性[13]。他们招募了 30 名年龄在 18 和 38 岁之间的志愿者,与三种不同的外观的机器人进行互动,机器人帮助志愿者测量血压。三种机器人分别有人性化的脸,像人的银色脸或无面,前两个机器人具有人脸特征,眉毛和肉质肤色,非常像欧洲学生。结果表明最受欢迎的助手是具有类人外观的机器人,参与者最不喜欢具有银色人脸的机器人,认为此类机器人最令人毛骨悚然,且具有更低的信任值。这一研究结论符合恐怖谷理论。恐怖谷理论表明机器人越像人则人们越有好感,但当超过临界点时,越像人越反感恐惧,直至谷底。但当机器人的外表和动作和人类的相似度继续上升的时候,人类对他们的情感反应亦会变回正面。外观很可能会影响用户对机器人个性和心灵的想法。Ho 等[14]研究表明,机器人与人某些外部特征的不良混合可能是人感到怪异的原因,与此一致,Bartneck 等[15]研究发现有明显非人外观的机器人比极力像人的机器人更受欢迎。虽然,人们喜欢与具有类人外观的机器人互动,但是,人们却不一定信任这些机器人。在人际信任中,如果喜欢被信任者,则更有可能信任他们。即被信任者的吸引力的大小与他们的可信度之间是一种正相关的关系。而在人-机信任中,机器人外观的吸引力与可信度之间的关系还有待考证。

4. 实体接触影响人机信任?

在人际交互中,身体接触,例如握手,在需要相互合作的信任交互中尤其重要。有人建议,没有实体接触就没有真正的信任。握手是一种非语言的社会仪式,在东西方文化中广泛使用。握手作为一个象征性的社会动机,来表达合作和信任的意愿[16]。人们握手做交易;人们握手来解决他们的差异。许多关于人际关系的研究都发现,与他人适当的社会交往性质的身体接触对于随后的人际交往有积极的影响[17]。一个简单的触摸例如,“拍肩膀”可以传达关注和安慰,而“拍拍背”是普遍公认的信号激励。Hancock 等研究发现在谈判前握手对于得到更多互利的结果有促进作用[18]。Bainbridge 等人[16]探讨了机器人的

实体存在如何影响以机器人为社会伙伴的人的判断。他们发现相对于视频显示,当机器人物理存在时,参与者更有可能实现信任。Rae 等[19]调查了物理实体和控制对交际和协作的影响。他们发现物理实体的呈现增加了合作伙伴之间的信任。Bevan [20]考察了与谈判代表“Nao”人形机器人握手之前和之后分别发起谈判的效果,结果表明握手这一实体接触动作增加了合作几率。Sirkin [21]的研究也表明,增加物理实体接触,对于人机良性互动有促进作用。可见,正如人际信任,身体接触在人-机信任中也可以增强对对方的信任程度。

5. 文化背景影响人机信任?

人-机信任受到文化传统的制约。不同文化间存在巨大差异,无论是语言、动作、表情,在不同文化中的含义也千差万别。也即是说,如果机器人的设计不考虑文化的因素,那么机器人可能并不会被使用者所接受。因此机器人交互过程也需要对不同的文化进行研究。一些研究表明,在不同文化背景中的人,和人工智能系统的互动中也有所不同。Bartneck 等[15]对7个国家的参与者进行了调查,研究发现来自不同国家的参与者对机器人有不同的态度。相对而言,美国的受访者对于机器人态度最积极,墨西哥受访者表现最消极。这项研究进一步证明了不同文化背景影响到人们与机器人的互动相处。Evers 等[22]发现当机器人助手被描述为群体内的成员,相比于美国被试,中国被试描述说感觉更舒服,控制感更强烈。从集体主义和个人主义文化的角度来解释,这一研究结果更有说服力。Rau 等[23]对不同文化背景下的人机交互信任进行过研究。实验邀请中国人和德国人完成选择产品价格区间的任务。实验从三个维度考量了人对机器人的态度:机器人的受欢迎程度,人对机器人的信任程度,以及机器人的可信度。实验的结果发现,中国人和德国人对于机器人有两种截然不同的态度。中国被试会对机器人更加信任,他们一般对机器人有更高的好感度,信任程度,而且他们会根据机器人的观点改变他们的决定。与中国人相反的是,德国人对机器人的好感度并不高,在机器人做出评价后,他们也往往不愿意改变自己的观点。中国人和德国人的这个区别可以用文化理论来解释。德国人往往更加独立自信,因此在面对机器人的建议时,不会轻易地改变自己的决策。而中国人往往更加依赖与他人的关系,相对而言并不是那么独立自主,因而也更加愿意接受它的建议。所以,在不同文化中生活的人们,对人工智能系统的信任程度可能并不相同,影响人-机信任的因素也有差异。

6. 其他影响因素

国内关于人与人工智能系统的研究还在发展当中,研究者们致力于人机信任模型的建立,通过调整机器人性能参数,说明了人,机器人系统,环境三者对人机信任关系的影响,重在讨论模型的优化。在这方面,国外研究者通过元分析法得出了结论。Peter 等[1]定量评估了人、机器人以及环境等三大因素在人机交互中对信任的影响,研究发现与机器人自身相关的因素,尤其是它们的性能(如可靠性、误警率和故障率等)对信任有很大影响,而与机器人属性相连的因素(如亲近性、机器人个性、拟人化等)则仅有相对较小的影响。Tanaka [24]认为,自主性(autonomy)、风险/自愿性(risk/vulnerability)与交互性(interactions)是考察人机信任度的时候要考虑的主要因素。机器的自主性越高,人机信任关系越有可能,同时信任者的风险或者自愿程度也越高。而且交互方式越直接,信任者越倾向于信任被信任者。

人际信任是人类社会的生存法则之一,合理的人际信任帮助我们规避风险和减少威胁,促进合作行为。随着人工智能系统的发展普及,人-机信任已成为人与机器人是否能和谐共处,合作发展的基础。AI时代,人机交互的前提是建立信任机制。信任是“人-机关系”迅速发展的重要部分。如果机器人想要在现实世界中高效运行,机器人必须学会适应周围环境的复杂性。诸多研究表明,人们试图以理解人的方式来定位机器人在人类世界中的角色,那么,人际信任的所遵循的规则可以为信任提供

一定的参考。已有研究证明, 机器人物理实体呈现, 实现适当的人-机接触可以促进人-机信任。外观在人际信任的判断中至关重要, 当然, 机器人的外观能一定程度上影响人类对机器人的喜恶, 但外观与信任的关系, 这方面还需研究考证。目前的研究结果表明, 机器人的性能和属性是人机交互信任发展的最大贡献者。幸运的是, 这些因素(例如, 机器人性能, 机器人属性)可以直接由人类设计而提升, 这样我们能够在某种程度上发展建设对现有人机合作伙伴关系有利的信任系统。然而, 具备合适性能和属性的机器人, 能否在世界范围内推广得到普遍认可, 文化差异也是需要考虑的因素之一。

参考文献

- [1] Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y.C., Ewart, U.S., de Visser, J. and Parasuraman, R. (2011) Human Factors. *The Journal of the Human Factors and Ergonomics Society*, **53**, 517. <https://doi.org/10.1177/0018720811417254>
- [2] Parasuraman, R., Sheridan, T.B. and Wickens, C.D. (2008) Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making*, **2**, 140-160. <https://doi.org/10.1518/155534308X284417>
- [3] Pan, Y., Steptoe, W. and Steed, A. (2014) Comparing Flat and Spherical Displays in a Trust Scenario in Avatar-Mediated Interaction. SIGCHI, ACM, Toronto, 1397-1406. <https://doi.org/10.1145/2556288.2557276>
- [4] Stirrat, M. and Perrett, D.I. (2010) Valid Facial Cues to Cooperation and Trust Male Facial width and Trustworthiness. *Psychological Science*, **21**, 349-354. <https://doi.org/10.1177/0956797610362647>
- [5] Oosterhof, N.N. and Todorov, A. (2009) Shared Perceptual Basis of Emotional Expressions and Trustworthiness Impressions from Faces. *Emotion*, **9**, 128-133. <https://doi.org/10.1037/a0014520>
- [6] Drolet, A.L. and Morris, M.W. (2000). Rapport in Conflict Resolution: Accounting for How Face-to-Face Contact Fosters Mutual Cooperation in Mixed-Motive Conflicts. *Journal of Experimental Social Psychology*, **36**, 26-50. <https://doi.org/10.1006/jesp.1999.1395>
- [7] Schroeder, J., Risen, J. and Gino, F. (2014) Handshaking Promotes Cooperative Deal Making. Harvard Business School NOM Unit Working Paper No. 14-117. <https://doi.org/10.2139/ssrn.2443674>
- [8] Jack, R., Blais, C., Scheepers, C., Schyns, P. and Caldara, R. (2009) Cultural Confusions Show That Facial Expressions Are Not Universal. *Current Biology*, **19**, 1-6. <https://doi.org/10.1016/j.cub.2009.07.051>
- [9] Mahdi, K., Ashwini, S., Stéphane, R., Bardy, B.G. and Aude, B. (2016) Role of Gaze Cues in Interpersonal Motor Coordination: Towards Higher Affiliation in Human-Robot Interaction. *PLoS ONE*, **11**, 1-21, 21 p.
- [10] Ogawa, K., Nishio, S., Koda, K., Balistreri, G., Watanabe, T. and Ishiguro, H. (2011) Exploring the Natural Reaction of Young and Aged Person with Telenoid in a Real World. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, **15**, 592-597. <https://doi.org/10.20965/jaciii.2011.p0592>
- [11] Todorov, A., Said, C.P., Engell, A.D. and Oosterhof, N.N. (2008) Understanding Evaluation of Faces on Social Dimensions. *Trends in Cognitive Sciences*, **12**, 455-460. <https://doi.org/10.1016/j.tics.2008.10.001>
- [12] Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H. and Hagita, N. (2007) Android as a Telecommunication Medium with a Human-Like Presence. *Human-Robot Interaction*, Arlington, 9-11 March 2007, 193-200. <https://doi.org/10.1145/1228716.1228743>
- [13] Broadbent, E., Kumar, V., Li, X., Stafford, R.Q., Macdonald, B.A. and Wegner, D.M. (2013) Robots with Display Screens: A Robot with a More Humanlike Face Display Is Perceived to Have More Mind and a Better Personality. *PLoS ONE*, **8**, e72589. <https://doi.org/10.1371/journal.pone.0072589>
- [14] Ho, C.C., MacDorman, K.F. and Pramono, Z.A.D. (2008) Human Emotion and Theuncanny Valley: A GLM, MDS, and ISOMAP Analysis of Robot Video Ratings. *3rd ACM/IEEE International Conference on Human-Robot Interaction*, ACM, Amsterdam, 169-176.
- [15] Bartneck, C., Nomura, T., Kanda, T., Suzuki, T. and Kato, K. (2005) A Cross Cultural Study on Attitudes towards Robots. *Proceedings of the 11th International Conference on Human-Computer Interaction*, Las Vegas, 22-27 July 2005, 577.
- [16] Bainbridge, W.A., Hart, J.W., Kim, E.S. and Scassellati, B. (2011) The Benefits of Interactions with Physically Present Robots over Video-Displayed Agents. *International Journal of Social Robotics*, **3**, 41-52. <https://doi.org/10.1007/s12369-010-0082-7>
- [17] Gallace, A. and Spence, C. (2010) The Science of Interpersonal Touch: An Overview. *Neuroscience & Biobehavioral Reviews*, **34**, 246-259. <https://doi.org/10.1016/j.neubiorev.2008.10.004>

-
- [18] Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y. and de Visser, E.J. (2011) A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors: The Journal of the Human Factor*, **53**, 517-527. <https://doi.org/10.1177/0018720811417254>
- [19] Rae, I., Takayama, L. and Mutlu, B. (2013) In-Body Experiences: Embodiment, Control, and Trust in Robot-Mediated Communication. *Interaction*, **15**, 36. <https://doi.org/10.1145/2470654.2466253>
- [20] Bevan, C. and Stanton Fraser, D. (2015) Shaking Hands and Cooperation in Tele-Present Human-Robot Negotiation. In: *HRI*, ACM, Portland, 247-254.
- [21] Sirkin, D. and Ju, W. (2012) Consistency in Physical and On-Screen Action Improves Perceptions of Telepresence Robots. In: *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction*, Boston, 5-8 March 2012, 57-64.
- [22] Evers, V., Maldonado, H.C., Brodecki, T.L. and Hinds, P.J. (2008) Relational vs. Group Self-Construal: Untangling the Role of National Culture. *Proceedings of the ACM Conference on Human Robot Interaction*, Amsterdam, 12-15 March 2008, 255-262.
- [23] Rau, P.P., Li, Y. and Li, D. (2009) Effects of Communication Style and Culture on Ability to Accept Recommendations from Robots. *Computers in Human Behavior*, **25**, 587-595. <https://doi.org/10.1016/j.chb.2008.12.025>
- [24] Tanaka, K., Nakanishi, H. and Ishiguro, H. (2014) Comparing Video, Avatar, and Robot Mediated Communication: Pros and Cons of Embodiment. In: *Collaboration Technologies and Social Computing*, Springer, Berlin, 96-110. https://doi.org/10.1007/978-3-662-44651-5_9

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2169-2556, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>
期刊邮箱: ass@hanspub.org