

面部特征对 社会机器人感知可信度的影响*

Benedict Barrow¹ and Roger K. Moore²

Abstract—信任和可信度的感知在决策和我们对他人的行为中起着重要作用，这不仅适用于人与人的互动，也适用于人机互动。虽然近年来社会机器人领域取得了显著进展，但在完全理解影响人类对机器人信任的因素方面还有很长的路要走。本文介绍了基于假设“婴儿脸”会引发信任的社会机器人的面部特征给人的第一印象的研究结果。通过操纵 Furhat 机器人的背投影面孔，研究证实眼睛的形状和大小对可信度感知有显著影响。这项工作有助于理解在开发社交机器人时需要做出的设计选择，以优化人机互动的有效性。

I. 介绍

信任是任何社会正常运转的基本构建模块。我们不断被要求信任我们的政府和官员，我们的基础设施和机构，在日常生活中打交道的人们，以及我们日常生活所依赖的技术。反之，我们也被他人信任去完成任务，保守秘密，保护财产，捍卫边缘群体的权利，并保护弱势群体。没有信任，生活将充满风险和困难。

信任因此似乎具有明显的哲学兴趣，然而，这一领域的现代研究直到 1950 年代才真正开始 [1]。作为一种多方面的构念，关于信任的概念没有统一的定义或分析框架。从社会科学的研究来看，信任根植于社会经验，并远远超越了简单的依赖 [2]。关键的一点是将自己置于相对于他人的脆弱位置这一概念。这需要依赖互动伙伴的良好意愿和善意 [3]。因此，由于信息的不完善，信任总会涉及一定程度的风险。

那么，人类关系中的信任概念是如何在人机关系中表现出来的呢？随着机器人能力的增强，我们看到社交机器人进入了我们的公共场所、家庭和工作场所。因此，探索与这些新的社会实体所期望的关系本质将有助于为未来富有成效且有效的人机交互（HRI）奠定基础。

为了解决这些问题，本文介绍了关于面部特征对社会机器人可信度感知影响的研究结果。章节 II 探讨了影

响人机交互中信任的因素，包括第一印象的重要性。章节 III 和 IV 阐述了本研究的目标以及如何实现这些目标。章节 V 描述了实验过程，而章节 VI 则呈现了主要发现。最后，章节 VII 讨论了研究中遇到的一些问题。

II. 人机交互中的信任

A. 影响人机交互信任的因素

有许多因素影响人机交互中的信任。根据桑德斯等人提出的“人类-机器人信任模型” [4]，这些因素可以分为三类：(i) 与个人相关的因素，(ii) 与外部环境相关的因素，以及 (iii) 与机器人本身相关的因素。

人为因素：人类对机器人能力的期望可以极大地影响他们如何看待这个机器人。如果期望与机器人的实际能力不一致，这会导致信任的重大损失，尤其是在出现错误或误解时 [5]。此外，研究表明，接触过更多机器人的用户比那些经验较少的人更信任机器人 [6]，而且用户的性格（如内向/外向）也起到作用 [7], [8]。

外部因素：人机交互中的信任会受到所执行任务性质的影响。例如，在救援行动等高风险情况下，信任是影响人类决策及相应结果的关键因素 [9]。此外，文化因素也可以导致对机器人态度的差异 [10]。

机器人因素：由于信任是随着时间的一系列互动中建立起来的，任何不可预测的行为或系统故障都会削弱对关系的信心 [4]。此外，机器人的物理外观、声音、拟人特质、整体举止及其可能带来的怪异感也产生影响 [11], [12], [13], [14], [15]。

B. 第一印象

第一印象很重要，研究表明人们会根据面部特征评估来对人格特质（包括可信度）做出非常迅速的¹初步判断 [16]。这意味着，无论这些判断是否准确，面部刻板印象都会引导决策和人际互动 [17]。

当然，这种隐性偏见在特定情境下可能会有问题。例如，在法庭判决中已经证明面部刻板印象具有现实的法律后果 [18]。特别是，研究表明面部成熟度是法院裁决结

¹小于 100 毫秒。

*This study was conducted by B. Barrow (supervised by Prof. R. K. Moore) as part of his MComp. degree in AI & Computer Science at the University of Sheffield.

¹Benedict Barrow has now graduated from the University of Sheffield, UK benedict.barrow@gmail.com

²Roger K. Moore is with the School of Computer Science, University of Sheffield, UK r.k.moore@sheffield.ac.uk

果的预测因素之一,而那些“稚气”的被告被判有罪的可能性较低 [19]。这些发现在一个 2020 年的研究中得到了重复,在该研究中,被认为是“不可信外表”的被告被判定为有罪的概率高出 8.03% [20]。

C. 面部特征

如上所述,影响对社交机器人信任的因素是复杂且数量众多的。与机器人相关的特征已被证明具有最大的影响,并且由于面部在理解社会线索和意图中的重要性,考虑如何设计面部特征以创造可信赖的第一印象是很有趣的。

修改机器人的物理特征可能会很困难和昂贵。然而,新技术(如背投技术)允许在屏幕上动画并渲染一张脸,这使得探索这些设计特性变得更加可行 [21]。这为研究如何设计面部特征(例如眼睛、眉毛和脸部的形状)以创造更强且更值得信赖的社会机器人的第一印象开辟了新的途径。

III. 本研究的目的

调整面部特征的大小、形状和位置的效果已经在社会机器人学的背景下进行了探索。例如,在宋等人 [22] 的一项研究中,参与者被展示了一系列社交机器人的潜在设计方案,并随后对其可信度进行评估。然而,这些设计是二维的,并通过计算机屏幕向参与者展示。因此,可能存在对二维表示的社交机器人和三维、共处一地的具体化代理之间的感知差异。

一些工作已经使用了通过调整面部表情 [23] 或社交线索 [24] 的具身社交机器人来完成,但据作者所知,迄今为止尚未进行过关于尺寸调整、形状和面部特征定位对激发具身社交机器人信任的影响的研究。因此,本研究的主要目标是探讨一个问题——“婴儿脸”特征在人机交互中是否能重现其对信任的影响?

婴儿模式(德语:幼儿图式)是一种现象,其中某些婴儿般的特征可以积极影响对一个人性格的看法。根据 Zebrowitz 和 Montepare [25] 的研究,婴儿脸的人通常具有大眼睛、圆脸、细眉毛和小鼻梁,并且这些身体特性的组合与感知的信任度呈正相关。因此,在这些研究结果的基础上,本研究测试了这样一个假设:更大的眼睛和/或更圆的眼睛和/或更细的眉毛会增加对社交机器人初始信任感的印象。

IV. 方法论

A. Furhat 机器人

Furhat 是由 Furhat Robotics 开发的一款社交机器人²。Furhat 机器人拥有一张类似人类的头部,它使用背投技术在半透明的物理面具上实时显示数字面孔。这意味着面部可以高度定制、动画化并即时调整。有多种物理面具可供选择,从而使得投影出的面部大小和形状存在广泛的差异。

Furhat 预装了多个预配置的脸部,这些脸部可以非常容易地互换。还可以创建自定义脸部或操纵这些预设,它们本质上是映射到 3D 模型上的纹理,然后再投影到一个面具上。

虽然这些纹理可以在照片编辑软件中手动调整,也可以通过编程来操纵面部特征。例如,可以调整变量值如嘴更窄或眼睛_更大。因此这种灵活性使 Furhat 成为研究上述假设的理想平台。

B. 信任游戏

在社会机器人学领域,信任的测量方式种类繁多。自我报告量表很常见,通常使用非标准化的定制量表,或行为任务,如游戏 [26]。

这样一个行为措施的例子是从博弈论中借鉴而来的。“信任游戏”[27] 是一个经济博弈,它将信任建模为参与者愿意接受的金融风险的数量。在游戏中,有两个活跃玩家——投资者和伙伴。投资者会获得一些代币,这些代币被赋予了货币价值。然后告诉他们可以给伙伴任何数量的代币(包括不给)。他们给出的代币将乘以某个因子,通常是三倍或四倍,而未送出的代币则归他们自己所有。随后,伙伴决定要退还给投资者多少个这些代币。在整个游戏中,投资者和伙伴之间不允许进行沟通。

两个当事人之间的信任程度影响了转移的代币数量。如果合作伙伴互相信任并且合作良好,双方最终拥有的钱比开始时更多。然而,存在伙伴滥用信任的机会;通过首先转账,投资者暴露了这种风险。研究表明,在信任游戏中,与具有更值得信赖面孔的伙伴玩时,参与者更有可能表现出冒险的金融行为 [28]。

V. 实验

A. 设置

上述的信任游戏作为自定义 Furhat “技能”的一部分得以实现,在该技能中调整了与眼睛和眉毛形状相关的参数,以创建可能的“可信赖”和“不可信赖”的面孔。所选的基础面部网格是默认的 Furhat “Alex”。游戏规则

²<https://www.furhatrobotics.com>

以书面形式呈现，因此其传达方式没有差异。参与者在游戏过程中不得与机器人交谈，例如询问机器人的意图。要求参与者与一系列展示不同面孔的机器人伙伴玩游戏，并记录了冒险的资金数额。

实验采用被试内因子设计。自变量是眼睛大小、眼睛形状和眉毛宽度，每个因素有两个水平：

- 眼睛大小：小或大
- 眼形：狭长或圆润
- 眉毛宽度：细或粗

这导致了一个 $2 \times 2 \times 2$ 的因子设计，总共有八种不同的机器人面部设计——参见图 1。

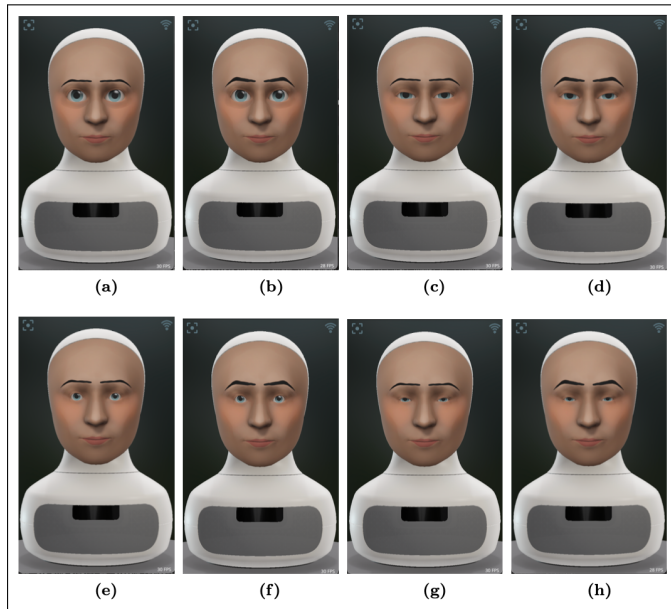


Fig. 1. 本实验中使用的八个机器人面孔（详见表 I）。请注意，图像来自 Furhat SDK，而受试者接触的是实际的机器人。

B. 参与者

共招募了 32 名参与者来自谢菲尔德大学，年龄范围从 19 岁到 50 岁，平均年龄为 25 岁。十名参与者认同为女性，21 名为男性，一名为非二元性别。参与者有机会赢得两张在线零售商的 £25 礼品券中的一张，以提供参与和认真玩信任游戏的动力。

一项使用 G*Power 软件进行的事先功率分析显示，为了使研究具有 95% 的把握识别中等大小的效果（使用 Cohen 的 f 值为 0.25），并且显著性水平为 0.05，所需的最小样本量为 23。因此，最终样本量 $n=32$ 满足这些要求。

C. 措施

可信度使用两种不同的方法进行评估。第一种是一系列与信任相关的自我报告量表，这些量表是从“人与自动化之间信任的检查清单”量表中修改而来的。这是一

个广泛使用的、包含十二个项目的七级李克特量表，通常用于评估人与自动化之间的信任 [29]。这些项目被修改以用单词“机器人”替换单词“系统”，以便适应实验的背景。某些条目未被包含（例如，“系统提供安全”），因为它们对机器人来说不相关。其他一些则因减少冗余和减轻参与者疲劳的原因而被移除。这导致了以下五个陈述的七级李克特量表：

- 机器人具有欺骗性。
- 我对机器人的意图、行为或输出表示怀疑。
- 机器人具有完整性。
- 我可以信任这个机器人。
- 机器人是可靠的。

感知可信度的第二个衡量标准是通过经济信任游戏来进行的。参与者被要求选择要给每个机器人多少代币。他们没有被告知每个机器人如何回应他们的初始投资，因为这会导致不希望出现的顺序效应——前一个机器人的结果可能会影响未来的投资。案例的独立性是用于分析数据的方差分析 (ANOVA) 方法的一个绝对关键假设。

D. 程序

在实验当天，参与者被邀请到谢菲尔德大学的 HRI 实验室，在那里他们坐在一张桌子前面对 Furhat 机器人。向他们展示了参与信息表和同意书，他们需要阅读并填写完才能继续。

每个机器人都是通过网页应用中的滑块进行评分的，并提供了一个文本框用于输入在信任游戏中投资的金额。当参与者按下下一个时，Furhat 投影仪会变暗，滑块会被重置为默认值，并且会在网页应用中显示一条定时消息，通知他们机器人面部会发生变化。三秒钟后，投影仪重新开启并显示出不同的面部。这一过程对所有八个机器人都重复执行，实验至此结束。

VI. 结果

因素分析显示李克特量表的响应数据是一维的。因此，可以通过计算对所有五个陈述（上述介绍）的响应的平均值和方差来接受信任度的综合评分。结果如表 I 所示，并可以看出具有小圆眼睛和细眉毛的机器人 (e)（见图 1）被认为是最值得信赖的。最不值得信赖的机器人是具有小窄眼睛和粗眉毛的机器人 (h)。这些发现大致与假设相符，不同之处在于机器人 (e) 和机器人 (f) 被评为比机器人 (a) 更积极。所有四个大眼睛的机器人倾向于变化，表明眼睛大小可能是极化因素。

三因素方差分析 (ANOVA) 在三个固定因子之间进行：眼大小、眼形状和眉宽，以及综合可信度得分。这些

TABLE I
总体可信度意见回应

条件	眼大小	眼睛形状	眉宽	μ	σ^2
Robot (a)	big	round	thin	4.28	1.16
Robot (b)	big	round	thick	4.02	1.12
Robot (c)	big	narrow	thin	4.01	1.22
Robot (d)	big	narrow	thick	3.87	1.22
Robot (e)	small	round	thin	4.55	1.06
Robot (f)	small	round	thick	4.38	0.91
Robot (g)	small	narrow	thin	3.78	1.23
Robot (h)	small	narrow	thick	3.33	1.01

结果总结于表 II 中。可以看出，眼形状以及眼大小+眼形状的结果具有显著性。

TABLE II
可信度意见的 ANOVA 结果

来源	F	p
Eye Size	0.074	0.786
Eye Shape ***	18.066	0.001
Brow Width	3.724	0.055
Eye Size + Eye Shape **	7.096	0.008
Eye Size + Brow Width	0.192	0.662
Eye Shape + Brow Width	0.088	0.767
Eye Size + Eye Shape + Brow Width	0.556	0.456

同样的三因素方差分析被应用于信任游戏的反应，结果汇总在表 III 中。同样，眼睛形状发现了显著的结果。

TABLE III
信任游戏的 ANOVA 结果

来源	F	p
Eye Size	0.589	0.443
Eye Shape ***	10.756	0.001
Brow Width	1.520	0.219
Eye Size + Eye Shape	0.156	0.693
Eye Size + Brow Width	0.589	0.443
Eye Shape + Brow Width	0.156	0.693
Eye Size + Eye Shape + Brow Width	0.005	0.944

眼形：分析支持了圆形眼睛比狭长眼睛更可信的假设。这一发现通过自我报告措施显示，并在信任游戏中得到了证实。该研究结果与关于眼部区域增加人类吸引力感知的研究一致，这与诚实和可信度的感知相关 [19], [30]。

眉宽：虽然 ANOVA 结果未达到被视为统计学显著性的正式要求，但仍有一种倾向认为薄眉的机器人面孔比浓眉的更值得信赖。

眼睛大小：可信度随着眼睛大小增加的假设被 不 支持。鉴于许多研究已经证实这一点，这是一个出乎意料的结果 [22], [19]。

交互效应：眼睛大小 + 眼睛形状：本研究中唯一显著的交互效应是眼睛大小与形状之间的关系。当眼睛又小又窄时，它们远不如大而窄的眼睛受信任。这可能是因为小而窄的眼睛几乎看不见。然而，对于圆眼来说，这种效果被逆转了；当眼睛又小又圆时，它们比大而圆的眼睛更受信任。一个可能的解释是，小圆眼相对于面部结构更加协调（相比之下，大的圆眼看起来不自然——下面将讨论这一点）。

VII. 讨论与结论

由于 Furhat 面具成型的性质，研究遇到了一些实际限制。特别是，调整面部特征的位置会导致不自然的表演。因此决定不移动特征，而是仅调整特征大小和形状。然而，尽管有这些缓解措施，对参与者的非结构化访谈显示，大眼睛被感知为“非常强烈”或“外观类似外星生物的”，这表明对面部网格的操作可能过于极端，并且眼睛变得不相称的大。同样，在特定面部网格中眉毛的变形有些角度，不够现实或不能代表典型的眉毛，尤其是婴儿的眉毛。

还应注意，不同人群的眼睛形状和大小存在差异，因此在参与者不同的其他国家可能无法重现这一结果。有可能参与者更倾向于信任那些设计上与他们具有相似特征的机器人，这将是未来研究的一个有趣方向。

尽管存在这些问题，本研究已经表明面部特征的设计在初次见面面对信任感的印象有显著影响，特别是眼睛。这种最初的积极印象在某些领域非常重要，例如医疗和治疗环境。社会机器人将长期存在，并且对于我们希望与这些新型代理在未来能够自主导航我们复杂世界的关系进行讨论对于社会来说至关重要。机器人的外观是这一讨论中的一个巨大因素，对人机交互和社会需求的仔细理解对于最终的设计及其接受度至关重要。

REFERENCES

- [1] M. Deutsch, “Trust and suspicion,” *Journal of Conflict Resolution*, vol. 2, no. 4, pp. 265–279, dec 1958.
- [2] L. Cominelli, F. Feri, R. Garofalo, C. Giannetti, M. A. Meléndez-Jiménez, A. Greco, M. Nardelli, E. P. Scilingo, and O. Kirchkamp, “Promises and trust in human – robot interaction,” *Scientific Reports*, vol. 11, no. 1, p. 9687, may 2021.
- [3] A. Baier, *Trust and antitrust*. New York, NY, USA: Routledge, 2014, ch. 31, p. 26.

- [4] T. Sanders, K. E. Oleson, D. R. Billings, J. Y. C. Chen, and P. A. Hancock, "A model of human-robot trust: Theoretical model development," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 55, no. 1, pp. 1432–1436, sep 2011.
- [5] M. Kwon, M. F. Jung, and R. A. Knepper, "Human expectations of social robots," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Christchurch, New Zealand: IEEE, mar 2016, pp. 463–464.
- [6] T. L. Sanders, K. MacArthur, W. Volante, G. Hancock, T. MacGillivray, W. Shugars, and P. A. Hancock, "Trust and prior experience in human-robot interaction," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 61, no. 1, pp. 1809–1813, sep 2017.
- [7] L. P. Robert, "Personality in the human robot interaction literature: A review and brief critique," in *Proceedings of the 24th Americas Conference on Information Systems*. New Orleans, LA, USA: SSRN, 2018.
- [8] G. Huang and R. K. Moore, "Get off on the right foot with whom?: How users' profiles affect their perception and experience with a social robot," in *IEEE International Conference on Robotics and Automation (ICRA)*, London, 2023.
- [9] E. Park, Q. Jenkins, and X. Jiang, "Measuring trust of human operators in new generation rescue robots," *Proceedings of the JFPS International Symposium on Fluid Power*, no. 7-2, pp. 489–492, 2008.
- [10] L. Wang, P.-L. P. Rau, V. Evers, B. K. Robinson, and P. Hinds, "When in Rome: The role of culture and context in adherence to robot recommendations," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Osaka, Japan: ACM Press, 2010, pp. 359–366.
- [11] J. Goetz, S. Kiesler, and A. Powers, "Matching robot appearance and behavior to tasks to improve human-robot cooperation," in *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003*. Millbrae, CA, USA: IEEE, 2003, pp. 55–60.
- [12] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 53, no. 5, pp. 517–527, oct 2011.
- [13] F. Eyssel, D. Kuchenbrandt, S. Bobinger, L. de Ruiter, and F. Hegel, "If you sound like me, you must be more human," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. New York, NY, USA: ACM, mar 2012, pp. 125–126.
- [14] M. Mori, "Bukimi no tani (the uncanny valley)," *Energy*, vol. 7, pp. 33–35, 1970.
- [15] R. K. Moore, "A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena," *Nature Scientific Reports*, vol. 2, no. 864, 2012.
- [16] J. Willis and A. Todorov, "First impressions: Making up your mind after a 100-ms exposure to a face," *Psychological Science*, vol. 17, no. 7, pp. 592–598, jul 2006.
- [17] M. van 't Wout and A. Sanfey, "Friend or foe: The effect of implicit trustworthiness judgments in social decision-making," *Cognition*, vol. 108, no. 3, pp. 796–803, sep 2008.
- [18] D. S. Berry and L. Zebrowitz-McArthur, "What's in a face? Facial maturity and the attribution of legal responsibility," *Personality and Social Psychology Bulletin*, vol. 14, no. 1, pp. 23–33, mar 1988.
- [19] L. A. Zebrowitz and S. M. McDonald, "The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts," *Law and Human Behavior*, vol. 15, no. 6, pp. 603–623, 1991.
- [20] B. Jaeger, A. T. Todorov, A. M. Evans, and I. van Beest, "Can we reduce facial biases? Persistent effects of facial trustworthiness on sentencing decisions," *Journal of Experimental Social Psychology*, vol. 90, p. 104004, sep 2020.
- [21] A. Kalegina, G. Schroeder, A. Allchin, K. Berlin, and M. Cakmak, "Characterizing the design space of rendered robot faces," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. New York, NY, USA: ACM, feb 2018, pp. 96–104.
- [22] Y. Song, A. Luximon, and Y. Luximon, "The effect of facial features on facial anthropomorphic trustworthiness in social robots," *Applied Ergonomics*, vol. 94, p. 103420, jul 2021.
- [23] A. S. Ghazali, J. Ham, E. I. Barakova, and P. Markopoulos, "Effects of robot facial characteristics and gender in persuasive human-robot interaction," *Frontiers in Robotics and AI*, vol. 5, jun 2018.
- [24] D. DeSteno, C. Breazeal, R. H. Frank, D. Pizarro, J. Baumann, L. Dickens, and J. J. Lee, "Detecting the trustworthiness of novel partners in economic exchange," *Psychological science*, vol. 23, no. 12, pp. 1549–1556, 2012.
- [25] L. A. Zebrowitz and J. M. Montepare, "Impressions of baby-

faced individuals across the life span.” *Developmental Psychology*, vol. 28, no. 6, pp. 1143–1152, nov 1992.

- [26] M. Chita-Tegmark, T. Law, N. Rabb, and M. Scheutz, “Can you trust your trust measure?” in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. New York, NY, USA: ACM, mar 2021, pp. 92–100.
- [27] J. Berg, J. Dickhaut, and K. McCabe, “Trust, reciprocity, and social history,” *Games and Economic Behavior*, vol. 10, no. 1, pp. 122–142, jul 1995.
- [28] R. van den Brule, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, and P. Haselager, “Do robot performance and behavioral style affect human trust?” *International Journal of Social Robotics*, vol. 6, no. 4, pp. 519–531, nov 2014.
- [29] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, “Foundations for an empirically determined scale of trust in automated systems,” *International Journal of Cognitive Ergonomics*, vol. 4, no. 1, pp. 53–71, mar 2000.
- [30] L. A. Zebrowitz, J. M. Montepare, and H. K. Lee, “They don’t all look alike: Individual impressions of other racial groups.” *Journal of personality and social psychology*, vol. 65, no. 1, p. 85, 1993.