

AI, but Make it Cute: How Cute Design Fosters Trust between Human-Robot Relationships

Gerard Samson

School of Information, Pratt Institute

INFO 601-05: Foundations of Information

Dr. Irene Lopatovska

October 26, 2023

Introduction

As artificial intelligence (AI) advances and is increasingly ingrained throughout our daily lives, the concern for whether or not AI can be trusted as a resource for information and an agent to carry out given tasks becomes more salient amongst the whole population, and especially within the information professions. Trust, however, is a complex mechanism, and can be achieved either by design or through more natural means. The value of cuteness, as an aesthetic, has a biological basis and is a factor in building trust with AI. There's a greater need to demystify these relationships as robots and automation are being compared to human intelligence. Understanding why designs are implemented and the perceptual consequences of these designs can remind users and designers alike that AI is no more than machinery and algorithms. Furthermore, understanding factors that build trust can help users be cognizant of what is being manipulated in AI systems to produce trust regardless of the actual competence of the AI.

Artificial intelligence is any automated machine that uses data to base their decision making. AI is implemented in many forms: chatbots, transportation, and algorithms (Kaplan et al., 2023). AI can also be embodied by social robots, a system designed to communicate and interact with human beings (Song and Luximon, 2020). Social robot systems are often designed with human-like features to facilitate human-robot interaction (Song and Luximon, 2020). According to Caudwell et al., (2019) social robots must be contained within a physical body. Social robots, however, are often designed in various degrees ranging from cute baby-like features to having an appearance most similar to a human being (Wodehouse and Duffy, 2018). There is a current need to research the implications for cute design in AI systems as the subject

hasn't been extensively studied, let alone cuteness and its effect on trust (Chen and Jia, 2023; Lv. et al., 2021)

Trust: It's Complicated

One consensus researchers can arrive at when examining trust: there isn't one. Across disciplines, trust is defined in many different ways and often context-dependent (Granatyr et al., 2018). There are many precursor traits leading to trust that are either dependent upon the situation and setting or dependent upon the traits of the trustor (e.g. human) and the trustee (e.g. robot). Aesthetics alone is a strong variable in influencing trust as people can perceive whether an object or person is trustworthy or not by a 100 milliseconds (Song and Luximon, 2020). The concept of trust is often based on human-to-human social relationships and secondarily applied to human-to-non-human relationships under somewhat altered criteria. In general, however, Jian et al. (2000) argues human-to-human trust and human-to-machine trust to be almost indistinguishable from each other (Chen and Jia, 2023). To apply trust to human-robot relationships, trust can be defined using Mayer et al.'s (1995) concept of trust in which trust is a cognitive stance or attitude of willingness to engage in actions. Actions in the context of human-robot relationships can then be considered any action the robot takes or the human-robot interactions themselves.

Trust is important to understand in human-to-machine relationships as AI systems become more prominent within our society. Trust in human-to-machine relationships alone is important to examine as distrust and or overtrust can lead to life-threatening risks for users. Overtrust is especially a concern when some studies indicate people are still willing to trust AI, even if the system makes flawed suggestions (Kaplan et al., 2023). When it comes to

human-robot relationships, trust is important as trustees, such as social robots, are designed to help the trustor with emotional and physical tasks and can be a significant influence on the trustors' decisions. Users must thus be cognizant of antecedents influencing trust, when actual competence and range of abilities of the robot is taken out of the account.

The “Aww” Factor

Cuteness can be a “small” vehicle in driving the large impact of achieving trust. The perception of cuteness, when it comes to AI design, is a byproduct of the baby schema effect. Animal behaviorist Konrad Lorenz (1971) proposed Kindchenschema, also known as the “baby schema” effect, as a tendency for humans to associate certain physical features such as round face, high forehead, small nose, small mouth, etc. to cuteness as it relates to human and animal infants (Chen and Jia, 2023; Song and Luximon, 2020). Baby schema effect isn't limited to aesthetics. According to a study conducted by Mara and Appel (2015), certain movements, such as head-tilts, are also associated with cuteness. The baby schema effect offers an explanation as to why consumers and users are enamored by cute designs. Cuteness can evoke positive feelings and behavioral responses. The baby-schema effect is an innate response serving to increase survivability of offspring (Caudwell et al., 2019; Lorenz 1971) Cute designs can thus inhibit emotional connection through the instinct to protect and nurture a cute agent (Chen and Jia, 2023).

High prices, low efficiency in task-completion, and planned obsolescence are barriers that could shorten the lifespan of social robots in the market. However, it may be too soon to say social robots are “gimmicky” as artificial intelligence is advancing by increasing in complexity, becoming ubiquitous as it's being embodied across all types of devices, and unintentionally

personified by users or intentionally personified through design. Cuteness as a design strategy isn't a commodity either – robot designers are actually leveraging cuteness for desired effects like trust.

To demystify cuteness and its implications in human-to-robot relationships, Caudwell and Lacey (2020) examine three social robots: Mayfield Robotics' Kuri, Emotech's Olly, and Jibo Inc.'s Jibo. Reminiscent of the baby schema effect, Kuri is small, round, has limited movement, and large eyes in proportion to their head. Etherington (2017) notes, "Kuri needed to inspire trust in its users, and so the design process involved eliminating any motion, sound or type of movement that would potentially unnerve its users." Both Olly and Jibo challenge the notion that cuteness must be exemplified through rigid human-like features. Olly and Jibo embody the most abstract definition of cuteness as it exhibits the least amount of human-like features, compared to social robots like Kuri, but they both leverage baby schema through limited movements, small size, roundness, and features reminiscent of a large eye.

Cuteness is also associated with anthropomorphization as explained by the baby schema effect. Anthropomorphization of AI systems, or the design of AI systems to be human-like (Chen and Jia, 2023), develops trust as previous research supports that trust is fostered when the trustee looks similar to the trustor. When compared to a mechanical face vs. an anthropomorphic face, previous studies support a higher perception of trustworthiness for anthropomorphic faces (Song and Luximon, 2020). However, a human-like appearance can only achieve trust to a certain extent, as the Uncanny Valley effect, conceptualized by robot designer Masahiro Mori (1970), proposes there's a certain degree of human similarity that users might find "strange, unlikable, and even repulsive" (Caudwell and Lacey, 2020; Chen and Jia 2023; Gn, 2018). Cuteness can be viewed as a slight departure from the Uncanny Valley to avoid any features that might inhibit

feelings leading to distrust for the user. It rounds out the harsh features of what a human might see as an untrustworthy design, and leads the user to engage with the robot more.

Anthropomorphization and the baby schema effect are effective as it can become an easy shorthand to communicate trust. From an evolutionary perspective, humans are designed to recognize faces quickly. As previously noted, it only takes 100 ms for humans to perceive traits like trust in humans and objects (Song and Luximon, 2020). Human tendency to recognize faces or bodies in ambiguous shapes and features is thought to help establish who is a friend and a foe in the context of tribal alliance and recognizing predators (Damiano and Dumouchel, 2018).

Due to the baby schema effect, cuteness communicates innocence and vulnerability, and thus establishes trust through these traits. Because of limited features and movements, cuteness implies a lack of power and ability to harm the user. The perception of an imbalance of power (i.e. superiority) the trustor has on the trustee leads to user perception that the trustee can't harm the user, and thus leads to the user developing trust with the machine. Furthermore, research suggests users perceive highly human-like robots to be menacing, especially if the robot seemingly performs better than humans (Yogeeswaran et al., 2016). A suggestion of powerlessness can thus mitigate negative user perceptions brought about by highly competent, human-like robots.

Through anthropomorphization, cuteness establishes a form of intimacy to develop trust. The agent is essentially “lending their hand” and asks the user to trust the agent through the design. The design suggests intimacy as it is in form next to human design. Anthropomorphism is a gesture of trust as it tries to achieve proximity to something more human-like and disarms barriers distancing between human-to-robot relationships. Cute design allows enough disparity to avoid user discomfort brought about by human verisimilitude. Too much human similarity offers

a suggestion of the grotesque as the highly human-like robot may exhibit a form of deformity and maneuver unlike a human would. Users can process a robot's imperfect human-like features to be alien-like, and experience discomfort due to the agent's otherness. Mori (1970) further argues that the response of repulsion is a result of these "not-quite" human features being reminiscent of a dead body (Caudwell and Lacey, 2020).

Both anthropomorphization and the baby schema effect achieve to communicate goodness and safety. Traits communicating in-group similarities facilitate trust (Song and Luximon, 2020). The human-like and baby-like design affords an opportunity for the agent to communicate trust as it is similar to the user. In other words, the cute design applied to an agent communicates: "I am like you, you are like me, and thus I wouldn't harm you. In conclusion: I should be trusted."

The Dark Side of Cute

Cuteness has evident shortcomings as it relates to durability and longevity. In behavioral psychology and neurobiology research, cute stimuli is shown to activate the nucleus accumbens, lyona region in the brain that modulates reward processes. Cute stimuli therefore provides a "quick hit" as the reward processes can be compared to the outcomes felt from exercise, gambling, euphoric drugs, and video games. Short term reward responses are especially concerning as social robots have been failing in the market (Caudwell et al., 2019). The limited lifespan of social robots and short-term neural responses to cuteness brings to question whether or not cute design is enough to uphold a long-term sense of trustworthiness in a relationship between humans and social robots. Based on the research, feelings of trust brought about by cute design may then be short lived.

No matter how endearing cuteness is, it is argued to be a manipulative design principle as it could lead the user to disregard the actual competency of the autonomous system. Despite “cuteness” exhibiting an essence of honesty and transparency (Chen and Jia, 2023), two associated traits associated with trust, it can be argued that the agent is dishonest as cuteness conceals and compensates for the robot’s limitations. The human-robot relationship is thus dishonest itself. The relationship is fabricated largely by design and not by the robot’s efficiency in performing its abilities. The cute aesthetic leads to overtrust as the factors developing to trust have nothing to do with the agent’s competencies and brings the question as to whether or not the agent can actually be trusted. Overtrust in the autonomous system could produce risk especially if the task may harm the user emotionally and physically. While cuteness could be considered a manipulative design principle, it's essential when applied to social robots as nonetheless it still excels in its purpose of facilitating human-robot interactions.

There are further philosophical and ethical concerns critiquing the emotional connection and trust between human-to-machine relationships, especially with social robots leveraging cute design. Since cuteness communicates a lack of power, human relationships with cute robots could imply a toxic social hierarchy of humans towering over robots, but a social hierarchy relies on the presupposition that AI is closer to human essence rather than a mechanical tool meant to achieve tasks. In truth, human-AI relationships are more nuanced, as AI is developing to be thought of as “partners” rather than just simply “tools” (Dumouchel and Damiano, 2018; Lyons et al., 2019). Therefore, social dominance is less of concern if the human-robot relationships are regarded as partnerships.

Future Implications and Conclusion

In the future, social robots may become more prominent in use within the general population and across different purposes and contexts. Agents that previously lacked embodiment, such as text chatbots or voice-user interfaces, may be anthropomorphized with a digital human-like interface or expand embodiment by putting more complex large language models into mobile social robots. In the context of anthropomorphizing social robots, cuteness may be the ideal design choice as simplistic features are easier to design and easier to achieve trust as opposed to more human-like or machine-like attributes. Caudwell et al. (2019) claims that the role of user-experience, product, and industrial designers as of late are limited to seeing aesthetics as an “added value” of robots. As the role of designers expands and human-robot interactions become ubiquitous, the role of designers consequently could expand beyond seeing external robot design as an added value, but rather designers are more integral in the design of interactive attributes necessary to building long-term, trustworthy human-to-robot relationships in the real world. Due to complexity, the ethical implications, and risk of human-to-robot interactions, there’s an increased need for robot design teams to include, not just designers, but an eclectic combination of engineers, philosophers, computer scientists, storytellers, and psychologists (Caudwell et al., 2019). With future implications in mind, it is evermore important to examine trust as AI systems evolve to carry out life-risking tasks (e.g. self-driving cars) and are designed to interact more with humans.

Cuteness is a design strategy that’s often implemented by robot designers to facilitate trust between the human-to-robot interaction. Cuteness is effective to foster trust. It’s important to examine antecedents of trust as human-to-robot interactions can produce emotional and physical risks. Additionally, robots can influence user decisions. There's a biological explanation

of why cuteness is effective in fostering trust as it relates to human-to-human relationships and our positive affect towards infants. Through anthropomorphization and the baby schema effect, cuteness is associated with human-like traits that facilitate trust.

References

- Caudwell, C., & Lacey, C. (2020). What do home robots want? The ambivalent power of cuteness in robotic relationships. *Convergence: The International Journal of Research into New Media Technologies*, 26(4), 956–968.
<https://doi.org/10.1177/1354856519837792>
- Caudwell, C., Lacey, C., & Sandoval, E. B. (2019). The (Ir)relevance of Robot Cuteness: An Exploratory Study of Emotionally Durable Robot Design. *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, 64–72.
<https://doi.org/10.1145/3369457.3369463>
- Chen, C.-H., & Jia, X. (2023). Research on the influence of the baby schema effect on the cuteness and trustworthiness of social robot faces. *International Journal of Advanced Robotic Systems*, 20(3), 172988062311684. <https://doi.org/10.1177/17298806231168486>
- Damiano, L., & Dumouchel, P. (2018). Anthropomorphism in Human–Robot Co-evolution. *Frontiers in Psychology*, 9, 468. <https://doi.org/10.3389/fpsyg.2018.00468>
- Dumouchel, P., and Damiano, L. (2017). *Living with Robots*. Cambridge, MA: Harvard University Press.
- Gn, J. (2018). *On the Curious Case of Machine Cuteness*.

Granatyr, J., Osman, N., Dias, J., Nunes, M. A. S. N., Masthoff, J., Enembreck, F., Lessing, O.

R., Sierra, C., Paiva, A. M., & Scalabrin, E. E. (2018). The Need for Affective Trust

Applied to Trust and Reputation Models. *ACM Computing Surveys*, 50(4), 1–36.

<https://doi.org/10.1145/3078833>

Etherington, D. (2017). Home robot Kuri is like an Amazon Echo designed by Pixar |

TechCrunch. <https://techcrunch.com/2017/01/03/>

home-robot-kuri-is-like-an-amazon-echo-designed-by-pixar/

Jian, J. Y., Bisantz, A. M., and Drury, C. G. (2000). Foundations for an empirically determined

scale of trust in automated systems. *Int. J. Cogn. Ergon.* 4, 53–71. doi:

10.1207/S15327566IJCE0401_04

Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2023). Trust in Artificial

Intelligence: Meta-Analytic Findings. *Human Factors: The Journal of the Human Factors*

and Ergonomics Society, 65(2), 337–359. <https://doi.org/10.1177/00187208211013988>

Lorenz, K. (1971). *Studies in Animal and Human Behavior*. Harvard University Press.

Lv, X., Liu, Y., Luo, J., Liu, Y., & Li, C. (2021). Does a cute artificial intelligence assistant

soften the blow? The impact of cuteness on customer tolerance of assistant service

failure. *Annals of Tourism Research*, 87, 103114.

<https://doi.org/10.1016/j.annals.2020.103114>

Lyons, J. B., Wynne, K. T., Mahoney, S., & Roebke, M. A. (2019). Trust and Human-Machine Teaming: A Qualitative Study. In *Artificial Intelligence for the Internet of Everything* (pp. 101–116). Elsevier. <https://doi.org/10.1016/B978-0-12-817636-8.00006-5>

Masahiro, M., (1970) The Uncanny Valley. *Energy* 7.4: 33–35.

Mayer, R. C., Davis, J. H., and Schoorman, F. D. (1995). An integrative model of organizational trust. *Acad. Manag. Rev.* 20, 709–734. doi: 10.5465/ amr.1995.9508080335

Mara, M., and Appel, M. (2015). Effects of lateral head tilt on user perceptions of humanoid and android robots. *Computers in Human Behavior* 44 (2015), 326–334.
<https://doi.org/10.1016/j.chb.2014.09.025>

Song, Y., & Luximon, Y. (2020). Trust in AI Agent: A Systematic Review of Facial Anthropomorphic Trustworthiness for Social Robot Design. *Sensors*, 20(18), 5087.
<https://doi.org/10.3390/s20185087>

Wodehouse, D. A., & Duffy, A. (2018). *PAREIDOLIA: CHARACTERISING FACIAL ANTHROPOMORPHISM AND ITS IMPLICATIONS FOR PRODUCT DESIGN.*

Yogeeswaran, K., Złotowski, J., Livingstone, M., Bartneck, C., Sumioka, H., and Ishiguro, H. (2016). The interactive effects of robot anthropomorphism and robot ability on perceived

threat and support for robotics research. *J. Hum. Robot Interact.* 5, 29–47. doi:
10.5898/JHRI.5.2.Yogeeswaran