Digital Design for Trust and Trustworthiness

Jeffrey Chan[1]

Digitalization in design is likely to mean that people are increasingly interacting with machines—especially intelligent machines that can learn from and respond to human behaviors. Trusting these intelligent machines will be critical to their successful deployment. However, what does trust in intelligent machines mean? The trust observed in cooperative and sustained human relationships at least relies on (i) encapsulated (shared) interest (i.e., A trusts B because B's interest encapsulates A's), or (ii) a moral commitment for trustworthiness, or (iii) strong psychological disposition to be trustworthy. At least for the immediate future, intelligent machines can only demonstrate reliable and robust functioning for building trust. Although reliable functioning is a necessary attribute, it is insufficient because people tend to trust intelligent machines when they also know, inter alia, why these machines make a particular decision or prediction. To cement trust between people and intelligent machines, there should be formative conditions for trust and trustworthy attributes embodied by intelligent machines—conditions and attributes that can come about through design. This chapter discusses different theories of trust and then explains why they are limited when applied to the context of interacting with intelligent machines. This knowledge gap suggests the need to consider how to design conditions for trust and attributes of trustworthiness in artificial artifacts, processes, and systems. In sum, this chapter aims to highlight the emerging gap between digital design and artificial intelligence, and demonstrate why the design of trust and trustworthiness will be vital to bridging this gap.
— *Dr. Jeffrey Chan, Singapore*

1 Singapore University of Technology and Design, Singapore.

Digitization can be defined as a shift, either in part or in full, from what once required in-situ human interaction with physical artifacts to virtual, automated, or even autonomous machine interaction. Consider a visit to the bank. Customers used to enter a lofty-looking building, were greeted by a bank-teller, attended to their transactions, and when in doubt, queried the teller. Even in the span of this brief visit, customers were able to gauge many things for trusting the bank with their transactions: for instance, knowing something about the reputation of the bank before the visit, observing its physical upkeep, noting the countenance of the bank-teller, and perhaps, evaluating if their complaints had been empathetically addressed. Uncertainty and vulnerability may be the core elements of trust relations;[2] but trust only begins when the apprehension of uncertainty and vulnerability is somewhat assuaged—for instance, when customers' expectations and their experiences of in-person banking align, reinforcing the bank's real or perceived trustworthiness.

But digital banking has upended conventional conduits of building trust. Through a digitized assemblage consisting of information storage, transmission, networking, processing hardware and associated software and interface capabilities—presented via a virtual banking platform—a customer now interacts with a set of prefigured choices on the screen while being assisted by a disembodied bot.[3] This customer can no longer observe the bank's upkeep, verify a query, or gauge the helpfulness of tellers. Complaints to customer service are often directed to an automated system, where replies are all but uncertain. Yet a customer is nevertheless asked to trust the bank. The intangibility of digitization clouds judgment and phishing scams lure even savvy customers to surrender their banking credentials on fake bank websites where vast sums of money are lost and where mistrusting scammers for the bank has escalated into open distrust of digital banking and even reputable banks.[4]

---

2 Carol A. Heimer, "Solving the Problem of Trust," in *Trust in Society*, ed. Karen S. Cook (Russell Sage Foundation, 2001), 43.

3 William J. J. Mitchell, *E-Topia* (Cambridge: The MIT Press, 1999), 71–72.

4 Bryan Tan, "Commentary: Banking Scams and Phishing Attacks – New Measures Should Apply to Other Business Sectors," *ChannelnewsAsia*, January 27, 2022, https://www.channelnewsasia.com/commentary/ocbc-phishing-banking-scams-digitalisation-monetary-authority-singapore-cybersecurity-2459961.

This example highlights how digitization can often diminish, if not directly, undermine trust. Instead of designing for trust, digital designers promise even more sophisticated (cyber-)protection or impregnable security architectures. The distributive networked technology of blockchain has been described as an ironic architecture of "trustless trust"—an example of how digital design sidesteps the many vulnerabilities of trust.[5] Blockchain technology shares information among parties that may not necessarily trust one another, but nevertheless generates an output that can be considered trustworthy.[6] Yet the blockchain, while guaranteeing greater security, cannot enjoy the many pleasures of trust that generate benefits for social life.[7] A bank receives the benefit of a good reputation from the lavished trust of its customers, just as customers who are trusted by a reputable bank can access greater credit than those who are not. Subsequently, a trusted bank is incentivized to do better, which benefits shareholders and customers, and by extension, also the community. Conversely, trusted customers can afford to invest in new businesses, creating jobs that in turn empower individuals and their communities. This "regard-seeking" element of trust may appear trivial but when compounded, provides regenerative civic possibilities.[8] The alternative "trustless trust" architecture of blockchain has perhaps taken away far more than it gives.

Gaps and Key Questions

How digital designs are shaped can moderate trust—with significant technical, social, political, and ethical consequences. Even so, focused discussions on trust remain anemic in design studies.[9] Questions of trust

5 Kevin Werbach, *The Blockchain and the New Architecture of Trust* (Cambridge, MA: The MIT Press, 2018), 246.

6 Werbach, 7, 96–98.

7 Philip Pettit, "The Cunning of Trust," *Philosophy & Public Affairs* 24, no. 3 (1995), 218, https://doi.org/10.1111/j.1088-4963.1995.tb00029.x.

8 Pettit, 222.

9 Notable exceptions are: Pieter E. Vermaas et al., "Designing for Trust: A Case of Value-Sensitive Design," *Knowledge, Technology & Policy* 23, no. 3 (December 1, 2010): 491–505, https://doi.org/10.1007/s12130-010-9130-8; Philip J. Nickel, "Design for the Value of Trust," in *Handbook of Ethics, Values, and Technological Design*, ed. Jeroen van den Hoven, Pieter E. Vermaas, and Ibo van de Poel (New York: Springer, 2015), 551–67,

have regularly surfaced with the rise of Information and Communication Technologies (ICT).[10] Early consolidation in this direction has culminated in the notion of e-trust, which is trust developed in digital contexts and/or involves artificial agents. An example of e-trust is found in the practice of commercial content moderation. Can users trust a social media platform to protect them from traumatizing materials uploaded by other users of this platform? Without e-trust, users will be disinclined to use this platform. The path-breaking work of e-trust has sketched out four key vectors, namely: (i) the distinctive aspects of e-trust; (ii) the relation between trust and e-trust; (iii) how e-trust may emerge; and (iv) the extent to which artificial agents are involved in an e-trust relationship.[11] Debates on how e-trust may be similar or different from interpersonal trust persist.[12] Discussions of e-trust with autonomous artificial agents now span disciplines of AI, robotics, and ethics.[13]

These philosophical analyses and reflections cover important ground. However, they have largely sidestepped the design questions of trust and trustworthiness. Designers are increasingly tasked to reinforce trust relations rendered vulnerable by digitization. But how should trust be designed into digital systems?[14]Are greater regulations and more securitized features the answer? Or are there other approaches, by design, that can also build trust? To be certain, digital systems that lean heavily on security features will

---

https://link.springer.com/book/10.1007/978-94-007-6970-0; Melvin Chen, "Trust and Trust-Engineering in Artificial Intelligence Research: Theory and Praxis," *Philosophy & Technology* 34, no. 4 (December 1, 2021): 1429–47, https://doi.org/10.1007/s13347-021-00465-4.

10 Batya Friedman, Peter H. Khan, and Daniel C. Howe, "Trust Online," *Communications of the ACM* 43, no. 12 (December 1, 2000), 34–40, https://doi.org/10.1145/355112.355120.

11 Mariarosaria Taddeo and Luciano Floridi, "The Case for E-Trust," *Ethics and Information Technology* 13, no. 1 (March 1, 2011), 1–3, https://doi.org/10.1007/s10676-010-9263-1.

12 Jonathan Tallant, "You Can Trust the Ladder, But You Shouldn't," *Theoria* 85, no. 2 (2019): 102–18, https://doi.org/10.1111/theo.12177.

13 Helga Nowotny, *In AI We Trust: Power, Illusion and Control of Predictive Algorithms* (Medford, MA: Polity, 2021); Marcello Pelillo and Teresa Scantamburlo, eds., *Machines We Trust: Perspectives on Dependable AI* (Cambridge, MA: MIT Press, 2021), 2; Claudia Hauer, "Should We Trust Robots? The Ethical Perspective," in *Trust in Human-Robot Interaction*, ed. Chang S. Nam and Joseph B. Lyons (London: Academic Press, 2021), 531–33.

14 See, for example, James Clayton and Jasmin Dyer, "Roblox: The Children's Game with a Sex Problem," *BBC News*, February 15, 2022, https://www.bbc.com/news/technology-60314572.

drastically differ from others characterized by fostering trust relations.[15] Furthermore, building trust often requires counter-intuitive approaches; for instance, the design of constructive d istrust.[16] Here, conflicts are intentionally designed into digital systems to provoke active questioning and critique.[17] Constructive distrust prevents individuals from relying on institutions or systems that would abuse them if they had acted as though they trusted them.[18] Drawing on the earlier example of phishing scams again, an example of constructive distrust is to remind users that all digital messages claiming to be legitimate should be distrusted until further authenticated.[19] This approach of deliberately interleaving d istrust for trust in digital design remains under-explored.

On the other hand, the sporadic literature that foregrounds design for trust has significantly advanced conceptual and practical knowledge of the intersection between trust and design. Nearly all are built on the important framework of value-sensitive design, where trust is defined as a key human value.[20] Nevertheless, differences also abound. In reviewing the trust that exists between organizations, Pieter Vermaas and his collaborators highlight how different types of trust—calculus-based trust, knowledge-based trust, identification-based trust—call for different design considerations.[21] On the other hand, Philip Nickel draws out key factors when designing for the value of trust.[22] Among many, salient factors include first, providing evidence that meets acceptable standards of trust to people, and second, building in social

15 Batya Friedman and Peter H. Kahn, "Human Values, Ethics, and Design," in *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, ed. Julie A. Jacko and Andrew Sears (New York: Lawrence Erlbaum Associates Inc., 2002), 1183.

16 Mireille Hildebrandt, "Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning," *Theoretical Inquiries in Law* 20, no. 1 (March 16, 2019): 83–121, https://doi.org/10.1515/til-2019-0004.

17 Carl Disalvo, *Adversarial Design* (Cambridge, MA: The MIT Press, 2012), 4–7.

18 Karen S. Cook, Russell Hardin, and Margaret Levi, *Cooperation Without Trust?* (New York: Russell Sage Foundation, 2005), 62.

19 Rei Kurohi, "4 Common Types of Scams and How to Recognise Them," *The Straits Times*, January 20, 2022, https://www.straitstimes.com/singapore/courts-crime/4-common-types-of-scams-and-how-to-recognise-them.

20 Batya Friedman and David G. Hendry, *Value Sensitive Design: Shaping Technology with Moral Imagination*, (Cambridge, MA: The MIT Press, 2019).

21 Vermaas et al., "Designing for Trust."

22 Nickel, "Design for the Value of Trust," 564.

and linguistic attributes that invite interpersonal trust. Other works focus on how to build trust in socially inclusive and democratic co-design processes.[23] These studies all underscore that trust is a value and relation amenable to design interventions.

Nevertheless, trust is also a value with ethical import.[24] Rightly, Nickel acknowledges that trust is not always good; instead, it is also a psychological state that represents the trusted as trustworthy.[25] Trust renders the trustor vulnerable, who can be made substantially worse off if they trust a scammer.[26] Designing for trust when the trusted entity is in fact untrustworthy is unethical. How should designers justify the design of trust relations? What are the assurances provided as a warrant to trust? Influencing or persuading individuals to trust through design is a moral matter.[27] The need to secure users' scarce attention to increase profits has been normalized in digital design.[28] All things equal, attention is lavished on platforms that invite trust, and these platforms are more frequently used. However, advertisers can also exploit this trust for objectives that do not always align with users' best interests. Where then is the redline beyond which building trust on digital platforms by design becomes patently unethical? Digitization has rendered the perplexing ethics of "design for trust" palpable.

The subsequent discussions will address these identified gaps through the following three clusters of questions. First, what is trust in digital design? What are other concepts closely associated with trust? Second, why is trust important, and how is it relevant to what designers do? Third, what are the considerations in the design of trust and trustworthiness—and how are they different? And beyond design considerations, what are preliminary ethical considerations of designing trust and trustworthiness? Working answers to these questions provide a primer to the examination of trust in digital design.

---

23 Ezio Manzini, *Design, When Everybody Designs: An Introduction to Design for Social Innovation*, trans. Rachel Coad (Cambridge, MA: The MIT Press, 2015).

24 Friedman and Hendry, *Value Sensitive Design*, 28.

25 Nickel, "Design for the Value of Trust," 552.

26 Russell Hardin, *Trust and Trustworthiness* (New York: Russell Sage Foundation, 2004), 37.

27 James Williams, "Persuasive technology," in *Future Morality*, ed. David Edmonds (Oxford: Oxford University Press, 2021), 137.

28 Williams, 137.

What is trust—in digital design?

Reliability and confidence are concepts that bear a close resemblance to trust. Yet, neither is trust, even though each overlaps trust in important ways. The following discussion serves to highlight key distinctions and, through them, elicit the notion of trust.

First, a reliable thing is competent in performing a particular task, and in everyday parlance, people are said to trust reliable things. Yet, to count on a reliable thing to do something is different from trusting a person to do something.[29] To trust a person to do something is not only to count on this individual's competency for a certain task but also on their motivation when performing this task. This is a categorically different matter from merely expecting an outcome from a reliable artifact. Insofar as a person is concerned, competency alone is necessary but not sufficient for trust: after all, it is possible to rely on a highly competent or intelligent person to perform a specific task without trusting them. To trust a person, in other words, is to accept the vulnerability of being exposed to the power of this trusted individual.[30]

In the context of design, a reliable artifact can only be interpreted as *functionally trustworthy*. But it is not necessarily trustable.[31] This is because users may not completely understand the working of this artifact or the motivations driving its design.[32] For example, users of a reliable online search engine are quick to discover that it is functionally trustworthy for showing a robust range of results, but they become quickly apprehensive when they discover certain results tend to be prioritized before others.[33] Similarly, deep learning systems can produce accurate and reliable results, but their opacity precludes straightforward explanations that can aid trustability.[34] These examples demonstrate that the functionally trustworthy artifact is

29 Richard Holton, "Deciding to Trust, Coming to Believe," *Australasian Journal of Philosophy* 72, no. 1 (March 1, 1994): 63–76, https://doi.org/10.1080/00048409412345881.

30 Annette Baier, "Trust and Antitrust," Ethics 96, no. 2 (January 1986): 231–60, https://doi.org/10.1086/292745.

31 John D. Lee and Katrina A. See, "Trust in Automation: Designing for Appropriate Reliance," Human Factors 46, no. 1 (March 1, 2004): 50–80, https://doi.org/10.1518/hfes.46.1.50_30392.

32 Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2016), 7.

33 Pasquale, 66, 75.

34 Mark Coeckelbergh, *AI Ethics* (Cambridge, MA: The MIT Press, 2020), 115.

merely the first step toward trust. Ideally, the goal is to attain high functional trustworthiness alongside high trustability. But there are often design trade-offs between functional trustworthiness and trustability.[35] For example, one way to improve trustability in artificial intelligence (AI) systems is to ensure as little bias as possible in their computations. Yet reducing bias in certain cases is likely to mean compromising accuracy, which in turn reduces reliability or functional trustworthiness.[36]

The second concept is confidence.[37] Positioned more strongly, confidence is akin to faith. Faith is usually one-sided and is maintained by the trustor independent of the actual trustworthiness of what is being trusted. An individual that professes faith in something or someone does not necessarily require reasons that can explain their trust—if only because trust at the level of faith is unverifiable even when it can be unreservedly accepted as true and real.[38] In faith, one simply trusts.[39] In this way, faith can be distinguished from trust because the latter usually requires some degree of rational assessment on why trust is warranted even when both presume some vulnerability on the trustor's part.[40] According to Hardin, trust should be defined as a three-part relation: A trusts B to do X, and B's (the trustee) interest is encapsulated in A's (the trustor) interest on task X.[41] Furthermore, A is able to explain why B can be trusted on task X. To render this relation in a concrete example, I can trust my neighbor's daughter Sally to babysit my toddler James, because I have observed their interactions in the past year and concluded that Sally really has the best interest of James at heart, which encapsulates my own interest for James. In contrast, faith is mostly a two-part relation, where A simply trusts B.

Is there any room for faith in design? The emergence of critical design theory can be interpreted as a deliberate attempt to question the inbuilt optimism of design action, tantamount to a kind of unquestioned faith in

35 Lee and See, "Trust in Automation," 74–75.

36 Michael Kearns and Aaron Roth, *The Ethical Algorithm: The Science of Socially Aware Algorithm Design* (New York, NY: Oxford University Press, 2019), 19.

37 Vermaas et al., "Designing for Trust," 497.

38 Josef Pieper, *Faith, Hope, Love* (San Francisco, CA: Ignatius Press, 1997), 15.

39 Russell Hardin, "Conceptions and Explanations of Trust," in *Trust in Society*, ed. Karen S. Cook (New York: Russell Sage Foundation, 2001), 13.

40 Hardin, 12.

41 Hardin, 4–6.

design for solving wicked problems.[42] On the contrary, democratic design institutions require not faith but trust-building exercises experienced through participatory and inclusive co-design processes.[43] Trust is cultivated through the germinating conditions of mutual respect, publicity, transparency, and autonomy. Under these conditions in participatory processes, designers and stakeholders can repeatedly encounter one another, find opportunities for cooperation, and discover reasons to trust one another.[44] Design for trust should at least provide cues and evidence that can allow users to ground their trust.[45]

The Importance of Trust in Digital Design

Trust is important in digital design for two primary reasons. First, trust can mitigate concerns about risks, especially when users are adopting a new design.[46] Maintaining some constructive distrust when encountering a novel platform or digital artifact is prudent. For instance, over-trusting the Autopilot in one's automobile, even when this self-driving technology is incomplete, could be fatal.[47] Yet, under-trusting can lead to an overestimation of risk, resulting in premature disuse. Increasing autonomous digital designs require locating that sweet spot between over-trusting and under-trusting. Second, recognizing the integral roles of trust in digital design can enable designers to avoid designing artifacts, institutions, or systems that will reduce trust, or else

42 Anthony Dunne and Fiona Raby, *Speculative Everything: Design, Fiction, and Social Dreaming* (Cambridge, MA: The MIT Press, 2013), 2.

43 Manzini, *Design, When Everybody Designs*, 173–75.

44 Manzini, 175.

45 Nickel, "Design for the Value of Trust," 556.

46 Ahmed Shuhaiber and Ibrahim Mashal, "Understanding Users' Acceptance of Smart Homes," *Technology in Society* 58 (August 1, 2019): 101110, https://doi.org/10.1016/j.techsoc.2019.01.003.

47 Faiz Siddiqui, "Tesla Driver Faces Felony Charges in Fatal Crash Involving Autopilot," *Washington Post*, January 20, 2022, https://www.washingtonpost.com/technology/2022/01/20/tesla-autopilot-charges.

drive trust out completely.[48] The practice of sharing users' data, for instance, has reduced trust and renders the task of rebuilding users' trust arduous.[49]

Trust is also connected to digital design in the following three broad ways.

First, there are artifacts, institutions, and systems that are designed with trust as one of their primary goals. Where uncertainty is present, trust as a design goal is paramount. As discussed, blockchain technology has been designed, for instance, to resist censorship or tampering by any one party using this system.[50] This feature aims to build trust, even though it is an ironic form of "trustless trust." Here, consider the case of Singapore's TraceTogether app. This app was designed for contact tracing during the COVID-19 pandemic. At the time of writing, this app has become a "catch-all" official platform for contact tracing, demonstrating proof of vaccination, and the portal to all pandemic-related protocols. Designers recognized early that to get as many people as possible habituated to this new app, its design had to "gain the trust of the people."[51] Among the different design considerations for building trust are assuring users that this app does not track their movement and location, using neutral white and teal color to indicate unvaccinated and vaccinated status, respectively, instead of the more common symbolic red colors to avoid discrimination (because everyone has their personal reason for or against vaccination), and the use of a swimming otter animation as a distinguishing mark of legitimate vaccination credentials.

Second, there are artifacts that mediate trust between people. Designs that specify how users encounter or interact with one another can bring people closer together, which can popularize the use of a certain artifact, or else render indifference and push users apart over time, leading to technological disuse and abandonment. For example, consider the reputation system for online transactions.[52] A clear and transparent reputation rating system on an

48 Pettit, "The Cunning of Trust," 202.

49 Elizabeth Schulze, "Facebook Says It Got Users' Permission to Share Data. Those Users Might Say Differently," *CNBC*, December 20, 2018, https://www.cnbc.com/2018/12/20/facebook-data-sharing-with-amazon-microsoft-netflix.html.

50 Werbach, *The Blockchain and the New Architecture of Trust*, 47.

51 Clement Yong, "Designer behind Otter on TraceTogether Check-in Page Wanted It to Dance," *The Straits Times*, https://www.straitstimes.com/singapore/designer-behind-otter-on-tracetogether-check-in-page-wanted-it-to-dance.

52 Paul Resnick and Richard Zeckhauser, "Trust among Strangers in Internet Transactions: Empirical Analysis of EBay's Reputation System," in *The Economics of the Internet and E-Commerce*, ed. Michael R. Baye, vol. 11, Advances in Applied Microeconomics

e-commerce platform, where sellers or buyers who have either been praised or criticized can respond, is instrumental for cementing a three-way trust: between users, the user and the platform, and a community of users and this digital platform. The greater the trust engendered by this reputation system, the more trustworthy the platform will be, and more users will want to use it.

The TraceTogether App in Singapore.[53]

(Emerald Group Publishing Limited, 2002), 127–57, https://doi.org/10.1016/S0278-0984(02)11030-3.

53 Jeffrey Chan, Photograph, 2021.

There are also more tangible examples. For instance, users of the BlueSG electric shared car in Singapore must ensure that after their use, the shared car is properly plugged into the charging station so that the next user will find this car fully charged and ready to use.[54] Unless the shared car is returned and properly plugged into the charging station, the user's credit card charges will not terminate. By constraining prior users to be responsible, subsequent users can be guaranteed a fully charged car ready to be driven. Through this design that mediates how strangers encounter each other in the BlueSG system, users can be said to trust one another on the matter of a fully charged car.

Third, an entirely novel category of autonomous artifacts and systems now requires trust. People depend on things to do something, yet increasingly, things are also deciding what to do and when to do it.[55] Trust is required when it is no longer possible for users to completely predict or control autonomous technologies. From autonomous vehicles and urban robots to the "Just Walk Out" automated stores, the effectiveness of these AI-powered artifacts depends on users trusting them. Users of future autonomous vehicles must trust them enough not to intervene at every conceivable opportunity during fast-moving traffic. Likewise, individuals who interact with urban robots and use "Just Walk Out" automated stores such as Amazon Go must sufficiently trust automated processes in these artifacts not to infringe their privacy, collect confidential data, and influence their behaviors without consent.[56] Failure to offer sufficient warrant for trust either in the autonomous vehicle or the "Just Walk Out" stores is likely to precipitate an escalation of distrust, which is penultimate to disuse and abandonment. Stated plainly, autonomous systems that are not trusted will not be used.[57]

---

54 "How It Works," BlueSG, September 20, 2017, https://www.bluesg.com.sg/how-it-works.

55 Rachel Botsman, *Who Can You Trust?: How Technology Brought Us Together and Why It Might Drive Us Apart* (New York: PublicAffairs, 2017), 319.

56 Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs, 2019), 7–8.

57 Kristin E. Schaefer et al., "A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems," *Human Factors* 58, no. 3 (May 1, 2016): 377–400, https://doi.org/10.1177/0018720816634228.

Design and Ethical Considerations of Trust and Trustworthiness

The primary task in design for trust is to transform trust as a human value into socio-technical design requirements.[58] This is a three-part process where the key value is first identified (i.e., trust), followed by explicating the norms or conditions that can support trust. These conditions can be made explicit by asking the following question, "under what norms or conditions is trust enabled and sustained?" Finally, these norms or conditions are further specified into various socio-technical requirements that are amenable to design. In turn, this produces a socio-technical feature, attribute, or artifact. This three-part process not only comprises different stakeholders' participation but is also iterative and integrative. This framework recalls the tripartite iterative methodology of value-sensitive design, which integrates conceptual, empirical, and technical investigations.[59] Nevertheless, this framework assumes that trust is good or desirable, and does not question why trust should be designed in the first place.

Consider Singapore's TraceTogether app again. Applying the three-part design method retroactively to this case, it may be possible to see how trust could be designed. Trust was likely first identified as an important goal of this app. As aforementioned, users that trust this app are more likely to use it; a higher usage leads to greater protection of public health. This was then followed by defining the conditions under which trust can be sustained. One defining condition for trusting this app is integrity. Integrity must be satisfied from at least two directions. On the one hand, officers performing gatekeeping functions outside of restaurants, shops, and offices must be able to ascertain that the individual requesting access has a legitimate vaccination certification. On the other hand, patrons of these venues must be reassured that the people they meet there have legitimate vaccination status. Therefore, no one should be able to access these venues using illegitimate credentials, for instance, in the form of a static screenshot of vaccine certifications borrowed from elsewhere. To counter this illicit use of a static screenshot, an animation of a swimming otter was programmed as the distinguishing feature of legitimate vaccination status.[60] A dynamic animation is far harder to counterfeit than a

58 Evgeni Aizenberg and Jeroen van den Hoven, "Designing for Human Rights in AI," *Big Data & Society* 7, no. 2 (July 1, 2020): 1–14, https://doi.org/10.1177/2053951720949566.

59 Friedman and Hendry, *Value Sensitive Design*, 32.

60 Yong, "Designer behind Otter on TraceTogether Check-in Page Wanted It to Dance."

static screenshot. This moving animation of a swimming otter gliding back and forth across the smartphone screen presents a hard-to-replicate socio-technical artifact that can secure integrity. The preservation of integrity, in turn, builds trust.

If this is a snapshot for the design of trust, what about the design of trustworthiness? Trust has been defined as a psychological disposition primarily consisting of people's expectations, whereas trustworthiness is a quality of a person, system, or artifact.[61] A reliable artifact is at least functionally trustworthy; an autonomous vehicle that is not only reliable but can further explain its decision to take a lengthier route exhibits an attribute of trustworthiness that one may accord to another human being. Yet, the design of trustworthy attributes in artifacts and systems is necessary but not sufficient for trust relations. For example, the city of San Francisco, California, has banned the use of facial recognition technology (FRT) by the police and other public agencies.[62] But this ban on FRT as a trustworthy attribute of a publicly accountable AI on its own does not mean that people now trust law enforcement agencies. Social norms and other political and cultural factors impinge on the perception of trustworthiness. In other words, it may be important to establish preconditions of social trust before directly designing trustworthy features.

Even so, trustworthiness must remain the kernel of trust relations. Acting on trust in the absence of trustworthiness can bring harm.[63] In philosophy and political theory, trustworthiness has been assessed as a moral matter.[64] Socio-technical artifacts are primarily designed with functional trustworthiness in mind (i.e., they perform a certain function reliably). Yet, they are increasingly also expected to meet the standards of moral trustworthiness—for instance, integrity, fairness, honesty, autonomy, and nonmaleficence. Nevertheless, the simultaneous fulfillment of functional and moral trustworthiness can be "tricky."[65] For example, the Apple AirTag trackers appear to have satisfied

61 Nickel, "Design for the Value of Trust," 559.

62 Kate Conger, Richard Fausset, and Serge F. Kovaleski, "San Francisco Bans Facial Recognition Technology," *The New York Times*, May 14, 2019, https:// www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html.

63 Hardin, *Trust and Trustworthiness*, 30.

64 Hardin, 36.

65 Tom Fisher and Lorraine Gamman, "Introduction: Ways of thinking tricky design," in *Tricky design: The ethics of things*, ed. Tom Fisher and Lorraine Gamman (New York: Bloomsbury, 2019), 2.

all performance criteria of functional trustworthiness. However, anti-social use of AirTags can violate privacy, which conflicts with values integral to moral trustworthiness.[66] Should designers now consider every way that a device could be used? Or should designers preempt paternalistic features that automatically report anti-social uses to the law enforcement agencies? If the former quickly becomes intractable, then the latter easily corrodes trustability. At the very least, designers should have considered a counterfactual scenario that the AirTag could be used to infringe privacy, and then equipped this device with an alerting feature that reminds users of unwanted AirTags in their vicinity. Designing this feature into the AirTag before it was introduced might have been a more responsible strategy than introducing a privacy update after public furor following privacy violations.

Toward an Interdisciplinary Research of Trust

As social relations are increasingly mediated by digital design, the design of trust and trustworthiness has become salient and urgent. But an efficacious and ethical design of trust and trustworthiness is hardly straightforward. Nevertheless, this chapter demonstrates that designing for trust is not only possible but there are also different design choices for building trust. One may opt for more securitized architecture, for instance, observed in blockchain technologies. Conversely, one can also consider morally trustworthy attributes that can result in a more trustable artifact. However, one thing is clear: interdisciplinary research intersecting digital design, artificial intelligence, and ethics on trust and trustworthiness has only just begun. The discussion in this chapter offers a primer for further work in this important area.

66 See Brett Molina, "Apple AirTag Trackers to Receive Privacy Update amid Stalking Concerns," *USA TODAY*, February 10, 2022, https://www.usatoday.com/story/tech/2022/02/10/apple-airtags-privacy-update-stalking/6738071001.