8.1. The Interrelationship between Human Empathy and Technological Progress in *Klara and the Sun* by Kazuo Ishiguro

Ilenia Vittoria Casmiri

Abstract

The aim of this essay is to explore the interrelationship between technological progress and the acquisition or loss of empathy in the science fiction novel *Klara and the Sun* (2021) by Kazuo Ishiguro. After outlining the results of contemporary studies on the concept of empathy in the fields of robotics and artificial intelligence (AI), I will explore some cultural reverberations of a potentially epistemic turn through the future scenario envisaged by the laureate of the 2017 Nobel Prize for Literature. *Klara and the Sun* is a dystopian narrative, in which humans and robots — known as artificial friends (AFs) — coexist. This relationship feeds the author's inclinations toward the study of human loneliness and love in the aftermath of a climate crisis, a theme that he previously explored in the uchronia *Never Let Me Go* (2005).

Key words: dystopia, artificial intelligence, human nature, science fiction, hyper-technological futures

The word "empathy" finds its roots in the Greek word *pathos*, which could be translated in English as "feeling". Its first occurrence dates to the translation of the German word *Einfühlung*. It was applied in the field of psychology in the early twentieth century to refer to the imaginative projection of a subjective state into an object so that the object appears to be infused with the feeling in a sort of animism (Wispé, 1986: 316). Since the early 1990s, empathy has been an object of interest in cognitive science and has been regarded as the inherently human tendency to attribute so-called propositional attitudes or intentional states of

belief, desire, and hope to other beings, as an attempt to explain their behaviour. Because it is also considered as the expression of a connection between two animated subjects (Quine, 1992: 68–69), human empathy is associated with AI consciousness in "machine societies", in which the role of the "machine" in relation to humans seems to hint at a perceived reduction of "the human" to "a component of a system" (Slocombe, 2020: 214), according to the social framework of the "two cultures", namely, the hard sciences and the humanities (Snow, 1990: 169). Therefore, the presence or lack of empathy would determine the (in)ability to understand and experience the feelings and thoughts of another person — or animate being — without having them communicated in an explicit manner. In this essay, the concept of empathy will be considered as a tool that humans use actively to understand and coordinate their intra- and inter-species relationships.

In fact, empathy has been identified as a key component of human–machine relationships in British fiction since the Victorian era — exemplary narratives are Anthony Trollope's *The Way We Live Now* (1875), and in the United States, *The Steam Man of the Prairies* (1868) by Edward S. Ellis. Since then, the western cultural imaginary has fed on and contributed to the spread of visions of future scenarios in which humans and robots do not coexist peacefully, as mirrored by the conflictual existence driving the narrative of Philip K. Dick's novel, *Do Androids Dream of Electric Sheep?* (1964) and its 1982 film adaptation, *Blade Runner*. Nevertheless, in their study of sex-robots, Kate Devlin and Olivia Belton (2021) provide evidence of contemporary conciliatory attitudes toward the machine. In fact, contemporary neurophysiological studies have found that people are moved by compassion when a robot vacuum cleaner is verbally harassed (Hoenen *et al*., 2016), or by empathy when they see a robot being physically harmed (Suzuki *et al*., 2015).

In my view, one of the most immediate and resonant instances of the relevance of empathy in the fields of robotics and AI is the famous robot called Sophia, activated in 2016 by former Disney imaginer David Hanson, CEO of Hanson Robotics. Since 2017, she has been a legal citizen of Saudi Arabia, which also makes her the first robot citizen in the world. All information concerning her creation, motives, and goals can be found on a dedicated page on the Hanson Robotics website, which is structured in such a way that readers are exposed to Sophia's objective uniqueness in the first section. The second section highlights the public implications of her existence and her contribution to technological advancement, while the third and most articulate part, seems to answer the question: what can Sophia do for me, as an individual, as a human being? The psychological implications of Sophia's existence in our world are eloquent in the last part, according to which the "Loving AI" project "seeks to understand how robots can adapt to users' needs through

intra and interpersonal development" (Hanson Robotics, 2022). The website includes the remark that this is a deliberately "science-fiction-like" concept, but it is my understanding that what Hanson Robotics, Amazon, and fellow major corporates do not seem to consider is that this idea is not necessarily utopian. In fact, robotic adaptation and development do not necessarily entail complete human control over the extent of such changes, for they most likely refer to robots' ability to update and redesign their behaviour in response to stimuli from the surrounding environment. In the case of interpersonal development through interaction with users, this environment is determined by the way in which humans relate to robots, which researchers have called "human-humanoid interaction" (Herrmann & Leonards, 2018: 2135).

The first conference on human-humanoid interaction (HHI) took place in 2008 and is today one of the main focuses of social robotics, which delves into the cultural aspects of determining the social acceptance of humanoid robots in our communities. For the time being, inter-species studies in HHI are focusing on the possibility that soon enough humans could be working side by side with humanoids (Kiesler, 2005: 731). In fact, this peculiar experience could trigger unwanted psychological mechanisms in human respondents, which "may influence critical cognitive processes" (Koban *et al*., 2021: 2) and determine whether they hinder or improve performance. In fact, evidence suggests that people tend to attribute humanlike characteristics to social robots (Spatola *et al*., 2020: 75; Koban *et al*., 2021: 5). This human tendency is mirrored by the changing perception of android beings in the last fifteen years. In fact, until 2007, researchers would study the human perception of robots through a perceptual-cognitive lens, which observed the robotic skills of perception, knowledge, and communication, as well as the affective capacities to sense and feel. Today, this vision has been integrated with a social cognitive dimension aimed at assessing androids' skills of social reasoning and moral cognition (Koban *et al*., 2021: 3).

Social reasoning and moral cognition are the skills that make Klara a rather peculiar robot in Kazuo Ishiguro's latest novel *Klara and the Sun* (2021). The story is told from the perspective of Klara, who is an insightful artificial friend (AF) that runs on solar power. She lives — or functions — in a techno-dystopian future in the USA, where social relations and family dynamics are altered by genetic engineering procedures on children and robots' crucial role in their upbringing. In Ishiguro's dystopia, children are usually "lifted", or subject to genetic alterations of their bodies to surpass natural human limits. Because of their peculiar (altered) nature, lifted children are partnered with robots, or AFs. The name used to refer to the robots is indicative of the social matrix of this AI: their role is to attend to the needs of the children in the household. In this "feasible eugenic dystopia" (Claeys, 2010: 109), middle-class parents must decide whether they want their children to undergo a process of genetic modification.

Such a process is never accurately described, but is related to neurotechnology. This decision will result in either a successful future for the children or their painful and slow demise. Klara will learn that parents do not always agree on their offspring's future, and those who oppose the biomodification process usually lose their job to androids and self-segregate in "post-employed" urban fractions. Segregation also occurs between "lifted" and "unlifted" children, which is the common name for children who did not receive the genetic engineering procedures that enhance their natural human faculties. "Lifted" and "unlifted" children can only socialize by attending "interaction meetings", which are regulated by rigid interaction rules.

Even though Klara's elemental insight into basic human social dynamics determines a biased narrative conveyed in simple language, readers immediately understand that lifted children do not seem to be able to feel empathy towards their unlifted peers. This does not mean that robots are treated better than unlifted humans. AIs are programmed to serve their household but do not know how to respond to social stimuli other than explicit orders. After being bought by Josie and her mother, Chrissie, Klara is greeted in her new household with different shades of distrust, discomfort, and unkindness. Yet Klara is not discouraged and is still willing to understand humans, but also to feel like one.

At the beginning of the novel, Klara can recognize negative feelings on humans' faces, such as frustration (26). She does not know how to feel anger yet (18), but she experiences surprise (12), puzzlement (17), and sadness at the apparent death of a beggar and his dog (37). She is so eager to learn and test her knowledge that she even attributes feelings such as sadness (5) and astonishment (12) to other AFs as well, and projects them onto Josie's drawings (140). She is curious about the complexity of human feelings such as the "pain alongside happiness" (21) of two people who see each other by chance on the street after spending a long time apart, or the basic human fear of loneliness (82). As she interacts with humans, the spectrum of emotions she can feel grows day by day.

In relation to Josie, the feeling Klara experiences the most is fear (41), for example, when she thinks that Josie will not take her home after all (40) or when the interaction meeting with other lifted children did not go as planned (84). Yet Josie is convinced that Klara is able to feel, and questions her AF about her alleged feelings of happiness (89) or melancholia (90). At one point, Josie complains about her own lack of social skills, for she cannot effectively communicate with unlifted children on her own, yet manages to succeed through Klara, who lectures Josie about kindness (126) and empathy (128) towards others. Josie and other lifted children never learn how to read other people's feelings for their vantage point allows them to see the world through the eyes of limitless and infallible individuals only. Therefore, Josie will never know how to deal with her own fallibility as a human being (134–137).

The complexity of HHI interaction is even clearer when we consider the effects of a technophilic society on the parents. Parents belong to the generation who saw and pursued the technological shift, yet this does not mean that they naturally weave social relations with AIs. After complaining about her daughter's carelessness towards other people's feelings, Chrissie claims that it must be nice being an AI without any feelings (97). Klara replies that the more she observes, the more feelings become available to her. On the one hand, Chrissie claims patronizingly that Klara could never develop empathy nor feel anything, for she was not built with this skill to begin with (98). On the other hand, Josie's mother is disappointed at not seeing Klara's "usual smile" (102) at the sight of a waterfall, and has no qualms about raising questions concerning her silent mood during the roadtrip to the natural site. These two behaviours are inconsistent in a character who is deeply convinced that AIs are merely functioning mechanics.

One reading of the behaviour described by Ishiguro may be that Chrissie is exerting cognitive dissonance (Festinger, 1962: 93) by pretending that the AF is just an empty shell, for this is a secure and soothing truth that will not interfere with her real, devious plans for Klara. Readers may understand that there is more to the story when Chrissie and Klara share an emotional conversation near the waterfalls. At the end, Chrissie asks Klara to *be* Josie. The AF claims she might be able to *imitate* Josie (103). Nevertheless, Chrissie meant what she said and expects Klara to sit, move, and speak while *being* Josie (104). The tension between the two characters rises when Klara can no longer deny the cruelty and greed in Chrissie's voice that accompanies her perverse commands, such as: "I want you to move. Do something. Don't stop being Josie. Let me see you move a little" (104). And later, "Good. More. Come on. … That's good, that's good, that's good" (105). The climax reaches its peak when Chrissie has a mental breakdown and forgets that Klara is not Josie after all.

After the dramatic dialogue, Klara is not yet aware of her role in Chrissie's plans, but her attention is caught by "the wooden rail marking where the ground finished and the waterfall began" (105). The waterfall is described by Klara as much more impressive than what she had seen in the magazines (106). In light of these considerations, I read this passage as a metaphor for the present, explicitly stated implications of AI technology in 2022 and the waterfall-like motion run by the endless possibilities envisaged by AI technology in the future. A more text-bound reading of the description of the natural site stems from the knowledge that Chrissie has commissioned a portrait of Josie, which is really a sort of wearable 3D sculpture with the looks of her daughter. Chrissie and Mr Capaldi, the sculptor, ask Klara if she would consent to wear the suit and take Josie's place in her mother's life, or to use Klara's words, to "continue" the child.

Today, AI experts are considering the implications of human–humanoid interactions on the workplace, a possibility that is likely not so distant in the future. Ishiguro brings this all too real possibility to a new extreme and makes his readers wonder whether AIs could replace humans in their emotional lives too. In fact, the only reason why Josie is not continued by her Artificial Friend is Klara's own will to look for a way to save her. Klara, who has developed a personal form of religion and worships the sun, concludes that, if she made the right offerings to her god, he might be able to heal Josie. Eventually, Josie gets better and Klara's usefulness in the household declines day by day, until she spends most of her days in the Utility Room (294). Eventually, Mr Capaldi asks Klara to go through with their experiment and to try and impersonate another dying human being anyway. Otherwise, she would "slow fade" (297). Mr Capaldi explains:

> There is growing and widespread concern about AF right now. People saying how you've become too clever … They accept that your decisions, your recommendations, are sound and dependable, almost always correct. But they don't like how you arrive at them (297).

Chrissie claims that Klara deserves her slow fading after all. Klara does not share with the readers why slow fading is perceived by humans and AFs alike as more "humane" than shutting androids down. Eventually, Klara is brought to a landfill to slowly die. Klara is not sad about missing the opportunity to work with Mr Capaldi, because replacing humans was never her intention. Mr Capaldi and Chrissie wanted to replace Josie with an AF because of hubris. The sculptor wants to defy the limits of human finitudes in a Frankenstein-like attempt to master nature, while the mother is trying to make up for her incapacity to bear losing yet one more child to the lifting process. Eventually, the only being that acts out of love and care for Josie is non-human Klara.

Conclusion

The central idea of *Klara and the Sun* is that every individual is irreplaceable; that there is something unique about humans that cannot be transferred to anyone or anything else, because it is not dependent only on our biology and genetics. What enables humans to act with empathy towards one another, to feel love and care for other beings, and be considered as unique is determined by the peculiar matrix of social relations woven during their lifetime, fostered by the

genetic inclination of intra-species interrelationships. Ishiguro's dystopia can be considered as the author's warning against a humanity that has failed in being true to its own nature and surrendered to systematic aversion stemming from human hubris. In fact, in an interview with *Nikkei Asia*, Ishiguro claimed to be more worried about genetic engineering than about the spread of applications of technological advancements in AI (Gohara, 2021). Although he acknowledged the enormous benefits brought by genetic editing in medicine and food production, he also expressed anxiety over its possible implications in human bioengineering, which would aim to achieve a "lifted" society from an intellectual and athletic point of view. According to Ishiguro, endorsing such a philosophy would turn our societies into "meritocracies", which he does not identify with a hierarchy based on merit, but rather as founded on class privilege and race, and fostered by the available technological tools used to "make some people superior to others" to engender a novel "apartheid system".

Note

1.*Merriam-Webster Dictionary*, q.v. "empathy", <https://www.merriam-webster.com/dictionary/empathy> [accessed 10 March 2022].

Works Cited

Claeys, Gregory (2010), "The Origins of Dystopia: Wells, Huxley and Orwell", in *The Cambridge Companion to Utopian Literature*, ed. Gregory Claeys. Cambridge: Cambridge University Press, pp. 107–31.

Festinger, Leon (1962), "Cognitive Dissonance", *Scientific American* 207.4, pp. 93–107 <https://doi.org/10.1038/scientificamerican1062-93>.

Devlin, Kate and Olivia Belton (2020), "The Measure of a Woman: Fembots, Facts and Fiction", in *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, ed. Stephen Cave, Kanta Dihal, and Sarah Dillon. Oxford: Oxford University Press, pp. 357–81.

Dick, Philip K. (1968), *Do Androids Dream of Electric Sheep?* London: Gollancz.

Gohara, Nobuyuki (2021), "Kazuo Ishiguro confronts basic questions about humanity and technology", *Nikkei Asia*, 2 March, online at <https://asia.nikkei.com/Life-Arts/Arts/Kazuo-Ishiguro-confronts-basic-questions-about-humanity-and-technology> [accessed 3 October 2022].

Hanson Robotics (2022), "Sophia", online at <https://www.hansonrobotics.com/sophia/> [accessed 10 March 2022].

Herrmann, Guido and Ute Leonards (2018), "Human-Humanoid Interaction: Overview", in *Humanoid Robotics: A Reference*, ed. Ambarish Goswami and Prahlad Vadakkepat. Dordrecht: Springer, pp. 2133–48.

Hoenen *et al*., 2016) Hoenen, M., *et al*. (2016), "Non-Anthropomorphic Robots as Social Entities on a Neurophysiological level", *Computers in Human Behavior* 57, pp. 182–6.

Ishiguro, Kazuo (2021), *Klara and the Sun.* London: Faber & Faber.

Kiesler, Sara (2005), "Fostering Common Ground in Human-Robot Interaction", in *RO-MAN '05: Proceedings of the IEEE 2005 International Workshop on Robot and Human Interactive Communication, Nashville, TN.* Piscataway, NJ: IEEE Press, pp. 729–34.

Koban, Kevin, Brad A. Haggadone, Jaime Banks (2021), "The Observant Android: Limited Social Facilitation and Inhibition From a Copresent Social Robot", in *Technology, Mind, and Behavior* 2.3, pp. 1–10 <https://doi.org/10.1037/tmb0000049>.

Slocombe, Will (2020), "Machine Visions: Artificial Intelligence, Society and Control", in *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, ed. Stephen Cave, Kanta Dihal, and Sarah Dillon. Oxford: Oxford University Press, pp. 213–236.

Snow, Charles Percy (1990), "The Two Cultures", *Leonardo* 23.2/3, pp. 169–73. <https://doi.org/10.2307/1578601>.

Spatola, Nicolas, Sophie Monceau, and Ludovic Ferrand (2020), "Cognitive Impact of Social Robots: How Anthropomorphism Boosts Performance", in *IEEE Robotics and Automation Magazine* 27.3, pp. 73–83 <https://doi.org/10.1109/MRA.2019.2928823>.

Suzuki, Yutaka, Lisa Galli, Ayaka Ikeda, Shoji Itakura, and Michiteru Kitazaki (2015), "Measuring Empathy for Human and Robot Hand Pain Using Electroencephalography", *Scientific Reports* 5, 15924 <https://doi.org/10.1038/srep15924>.

Quine, Willard Van Orman (1992), *Pursuits of Truth.* Cambridge MA: Harvard University Press.

Trollope, Anthony (1875), *The Way We Live Now.* London: Chapman and Hall.

Wispé, Lauren (1986), "The Distinction between Sympathy and Empathy: To Call Forth a Concept, a Word Is Needed", in *Journal of Personality and Social Psychology* 50, pp. 314–21 <https://doi.org/10.1037/0022-3514.50.2.314>.