



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

New Techno-Humanities

journal homepage: www.elsevier.com/locate/techum

The humanity of the non-human — Themes of artificial subjectivity in Ishiguro's *Klara and the Sun*

Oliver Li*, Johan Eddebo

CRS - Centre for Multidisciplinary Research on Religion and Society, Uppsala University, Sweden

ARTICLE INFO

Keywords:

Artificial subjectivity
Personhood
Subjectivity
Ishiguro

ABSTRACT

In this article we discuss themes of artificial subjectivity in Ishiguro's novel *Klara and the Sun*. We first present a thematic overview, and some reflections upon subjectivity. The analysis proceeds in four steps pertaining to perspectives on artificial subjectivity and the narrative construction of human dignity: (1) who is human, (2) where does the heart lie, (3) the dialectical creation of the heart, and (4) reflections on subjectivity and personhood. Finally, we summarize the views suggested and emphasize their relevance to society's understanding of humanity and the non-human. We also conclude that relational ontologies are more suitable to understand subjectivity and personhood, in particular in cases of interaction between the human and non-human.

1. Introduction

Ishiguro's *Klara and the Sun* is a beautiful and complex work that, in a sense, represents the belated ascent of sci-fi into the "literacy citadels" (Benford 1998). With the novel's low-key exploration of many ideas central to the genre's golden age and consequent, careful illumination of contemporary issues, we can here speak of an attempt to critically revitalize the hard questions once posed by a genre from which popular culture has thus far mainly extracted tropes and stereotypes.

While the relationships between culture, worldviews, and operative mythologies are inevitably intricate, the observation can be made that science fiction was to some extent pressed into service for reproducing the narratives of both the modern West and the underlying industrial social order. Central aspects thereof which are almost universally represented in science fiction include the myth of progress and some form of redemptive scientism where technology is framed as salvific. It could moreover be plausibly argued that the beliefs in the myth of progress, scientism or technology as salvific are once again becoming increasingly central to issues of power, liberty, and democratic agency in relation to the marketing and introduction of potentially disruptive technologies such as AI, biotechnology, and intrusive digital surveillance.

With this in mind, Ishiguro's work recovers the critical potential of sci-fi at a pertinent time and thematizes, as shall become clear, questions critical to precepts of everlasting technological progress and technology as salvific.

Key among these is the notion of subjectivity, the nature of agency and consciousness, which are explored against an implicit dystopian

background of megatechnology, penetrating digitalization, and a post- or transhuman social order. In exploring subjectivity in this context, Ishiguro heavily utilizes the concept of "the human heart" to emphasize the unique essence of the person (Ishiguro 2021, 218–219), which will be a key focus of this paper. The term 'human heart' as it is used in the novel includes both cognitive and emotional capabilities, however, not as separate entities but rather in some sense intertwined. This synergy brings into focus several non-reductionist ontologies compatible with the novel's implied metaphysics, especially such non-Western perspectives, which have been more or less absent from the popular discourse pertaining to artificial intelligence. Indeed, this metaphysical harmony can aptly be described with the Japanese term *kokoro*, denoting unity of spirit, mind, and body, a concept for these reasons notoriously difficult to translate and grasp with Western terminology (Kasulis 2011).

Given the above themes and Ishiguro's focus on the concept of the human heart, one can identify pointed questions such as 'where can humanity be found in humans or non-humans?', 'what is it that makes humans or non-humans human?' or 'what is the human heart?'. It is in the light of such questions concerning the possible demarcations between human and non-human, between humane and inhumane, that we wish to read and interpret Ishiguro's novel.

We first present a thematic overview and some general reflections upon subjectivity in the following. We then provide a brief synopsis of the contents of Ishiguro's novel. Following the above central questions the analysis of Ishiguro's novel then proceeds in four steps pertaining to perspectives on artificial subjectivity and the narrative construction of human dignity as they are depicted in the novel: (1) who is humane, (2)

* Corresponding author at: Department of Theology, Center for Multidisciplinary Research on Religion and Society (CRS Uppsala), Uppsala University, Box 511, SE-75120 Uppsala, Sweden.

E-mail addresses: oliver.li@teol.uu.se, oliver.li@crs.uu.se (O. Li), johan.eddebo@crs.uu.se (J. Eddebo).

<https://doi.org/10.1016/j.techum.2023.11.001>

Received 8 April 2023; Received in revised form 24 August 2023; Accepted 5 November 2023

Available online xxx

2664-3294/© 2023 The Author(s). Published by Elsevier Ltd on behalf of Shanghai Jiao Tong University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

where does the heart lie, (3) the dialectical creation of the heart, and (4) reflections on subjectivity and personhood. Finally, we summarize and discuss the views suggested by the novel and emphasize their relevance to society's understanding of humanity in relation to the non-human and conclude that relational ontologies are more suitable to understand subjectivity and personhood, in particular in cases of interaction between the human and non-human.

2. Thematic overview and general reflections on subjectivity

The general background of the novel places us in some unspecified future or alternate past akin to a post-industrial hi-tech capitalist social order. Class divisions are reinforced by active eugenics, and strong artificial intelligence is an almost trivial fact of everyday life. There are echoes of Huxley's *Brave New World* in relation to the eugenic regimentation as well in an implied social movement towards "opting out" of society, instead residing in decentralized communes. All of this, however, is rarely explicitly discussed but rather framed as unimportant background facts that play almost no role in the narrative's progression.

The first-person narration focuses extensively upon the subjective experience and reflections of an artificial subject in the form of a companion robot, an AF - Artificial Friend, even to the exclusion of other characters' experiences. It is in the following clear and simple sense we wish to understand subjectivity: As "the capacity to think of oneself as oneself in the first person" (Baker 2013,39).¹ Ishiguro's choice to mostly exclude the experiences of other human characters is very significant since it not only normalizes artificial subjectivity as something given from the outset but approaches the artificial subject as the only true person in the narrative. In contrast, we are given almost no details of the inner lives of other characters, human or otherwise. In a sense, therefore, the exclusive focus on the phenomenal subjectivity of a robot is both the central medium and message of the narrative, which becomes the nexus of the other themes being developed throughout and forms the basis of dialectical reasoning about what is humane.

From the very outset, the framing of this first-person narration is both profoundly existential and thoroughly political. The context of the androids in the "friend store" aptly sets the theme. They are basically line products being manufactured and marketed to satisfy consumer demand and cater to human needs, but at the same time possess a full rational subjectivity and an inner emotional life. Clearly, Klara thinks of herself as 'herself' and describes the world from her first-person perspective. We here see the same conflict that is peculiar to the phenomenon of alienation in modern industrial society, albeit magnified through the full synthetic and utilitarian nature of the subject fully as a product. The synthetic subjects are thoroughly alienated from their own needs and natures, since they exist only through and for the consumer economy.

In parallel, the narrative's "lifting" procedure, which can most adequately be described as a form of active eugenics designed to improve the capabilities of its human recipients in a transhumanistic sense, renders the very same alienation upon the human characters. Both man and machine are merely cogs in the overarching system of instrumental rationality, something which is particularly evident in Klara's "becoming" Josie, and the behavioural ontological framing of this process. All implies the interchangeability of human and machine, as well as of the roles and relations of production, emptying the worker as well as the means of production of their "situated thisness" and unique meaning that cannot fully fit within the system of technological rationalism.

¹ The notion of subjectivity is of course complicated and has been debated extensively in the philosophy of mind together with the closely related concepts of first-person perspective, personal identity or personhood. For a detailed discussion both from more naturalistic and non-reductive perspectives see, for example, works by Lynne Rudder Baker, Daniel Dennett, Thomas Metzinger or David Chalmers (Baker 2013; Dennett 1991; Metzinger 2004; Chalmers 1996; 2017).

Nonetheless, a significant aspect of the narrative is the radical affirmation of subjectivity in the above given sense, the phenomenal inner life of an artificial being. In this way, Ishiguro provides a sort of re-enchantment of the synthetic being's irreducible and subjective uniqueness framed through the lens of love, faith, and relationality. The android's affection for her owner, the young and ailing girl Josie, eventually grows into a sacrificial and tragic love bolstered by a sort of rudimentary theism anchored in the solar-powered robot's experience of receiving life from the Sun. This rudimentary theism in Klara's thinking stand in conflict to Josie's illness. Indeed, the transhumanistic lifting procedure is the cause of Josie's declining health, and the AI expresses theodicy issues and pleads for the sun's aid. Here, one can also discern the conflict between reductive physicalism and a holistic non-reductionism. The former is represented by the notion of a transhumanist replacement of Josie, and the latter position by several characters intuited notion that there is an essence, a "heart" of a person that cannot be approached reductively; a position we shall argue for further down.

As mentioned, the first-person narration is framed from the point of view of Klara, an artificial subject. Even though the reader knows who is human and who is not, this framing and the observation that the reader is given almost no details of the inner lives of any other human characters implicitly raises the question "who is *humane*?". Furthermore, the narrative moves from the assumption that there are artificial subjects, that subjectivity can be artificially created, to the possibility of human enhancement, as in the 'lifting process'. It then moves to the question whether a mind, Josie's mind, can be upload and copied (Ishiguro 2021, 207–210) – a clear transhumanistic theme – and finally to the final question what the human heart is and where the human heart lies (Ishiguro 2021, 306). Consequently, we shall discuss these two questions, "where the heart lies?" and "how the heart is (dialectically) created?" in greater detail. Finally, the discussion and analysis of the novel in light of these question leads us back to the discussion of the initial assumption that artificial subjectivity and personhood are possible and the closely related question which kind of ontology implicitly is implied by Ishiguro's fictive depiction of the world. We shall now turn to the first sub-question: "who is humane?"

3. Who is humane?

Josie's first 'interaction meeting' is a crucial scene early in the novel. It raises issues pertaining to the delineations between being human, non-human, or inhuman. Josie's peers seemingly do not have much 'natural' social contact, so their parents arrange these semi-formal social gatherings. However, Josie is not so fond of them. "Mom, if my grades are so good, do I really have to host this interaction meeting?" (Ishiguro 2021, 63) Josie's friend Rick who is not lifted, is also invited. During the meeting, a group of children bullies and intends to mistreat Klara, while Rick, in turn, 'saves' Klara from this precarious situation (Ishiguro 2021, 74–79).

The children neither seem to realize nor acknowledge the possibility that Klara, as an AF, might have consciousness and feelings. The humans act inhumanely. However, at least one human, Rick, the only one who is not part of the process of enhancement, acts according to acceptable moral standards. Paraphrasing John Danaher's words, Rick welcomes Klara into the moral circle (Danaher 2020b) and defends her as if she were a fellow human. This is reminiscent of Danaher's suggestions that one ought to "[...] err on the side of caution, of over-inclusivity not under-inclusivity, when it comes to whom we owe duties" (Danaher 2020a). In other words, he maintains that we should treat artificial agents (in appropriate cases) as if they had feelings and consciousness in order to be 'on the safe side.' Furthermore, the apparent suffering of Klara introduces the idea that artificial suffering will be a consequence of the development of beings like the AFs. The mere possibility of such suffering has led some philosophers, such as Thomas

Metzinger, to issue warnings against the striving even to develop systems in this direction (Metzinger 2021). Similarly, in Jonathan Nolan's and Lisa Joy's depiction of future sentient human-like machines in *Westworld*, humans are fully capable of treating AI-humanoids, like Dolores in the first season, in brutal and evil ways since they believe that the AI-humanoids simply are machines (Nolan and Joy 2016).

This ethical dimension of how humans think about possible other moral patients is also discussed by, for example, virtue ethicist Shannon Vallor. She argues that virtue ethics should be the preferred ethical framework in relation to human interaction with AF, AI or similar technology. Indeed, she believes there is a risk of moral degradation in other ethical frameworks such as the Kantian deontological or utilitarian frameworks (Vallor 2016, 209–11). Surely, the development of virtues can be seen as part of the emotional and cognitive abilities residing in the 'human heart' and in *kokoro* using the possibly more appropriate Japanese term.

Moreover, there is tension between humans who have been enhanced and those who have not been enhanced. If transhumanism is understood as referring to humans who are "still human, albeit greatly enhanced" (Mercer and Trothen 2021, 21), then the children in the interaction meeting are actually transhumans, and it is *not* the humans who act inhumanely, but the transhumans. Thus although, as Calvin Mercer and Tracy J. Trothen claim, both secular and religious transhumanists wish to spell out their projects in positive terms (Mercer and Trothen 2021, 19–41), Ishiguro's depiction of transhuman children bullying a non-human clearly points to a negative aspect within transhumanistic projects.

Indeed, the children who are lifted, that is, are enhanced and thus are part of the transhumanistic and eugenic project, envisioned by Ishiguro, have or at least will have higher status in his fictive society and will seemingly also have superior intellectual capacity. As such, one might also expect them to live up to higher moral standards. However, this is not the case. Those who are 'lesser' - Rick and Klara - have higher moral standards than the children who will become the future elite. A dehumanizing attitude is depicted in those who have a 'higher' status in society. The transhumanistic move in the lifting process seems to 'transcend' humanity not in a positive sense but a negative sense. If transhumanism is seen in parallel to posthumanism, this negativity can, according to Mark C. Taylor, be taken to another level who claims that "posthumanism is anti-humanism". (Taylor 2021, 106) Taylor's claim seems to imply that something is lost in any transhumanist or posthumanist striving and, in particular, in the lifting process as depicted in the novel.

On this note, the notion of transhumanism as reification in the Marxist sense, can be helpfully introduced. The original German concept, *Verdinglichung*, implies the reduction of something to an object. It especially refers to the process of transforming social relationships into seemingly objective attributes of a commodity or of an actual living person. A prominent example is the perception of a marketable item like brand clothing as having a significant intrinsic value in and of itself, over and above its use-value, while the actual attributed extra value consists in the surrounding network of social relations and arrangements. These are then being said to be "reified" in the commodity.

Transhumanism connects with reification since it implies the rejection of the natural limits of the person. It's a radical anti-essentialism that hinges on the Cartesian notion of an unimpeded subject, free to determine its own nature according to whatever it prefers.

Yet, this negation of the natural limits opens the "new" human being as socially constructed to be imbued with values alien to its basic relational wants, needs and sympathies. The structure of the surrounding system can now enforce its teleology and values in and through the synthetic consumer preferences of the "unfettered" subject, and a stronger reification of the anti-human tendencies of the dominant axiology becomes a clear possibility. Assuming that the heart after all is human and humane one may now wonder where the heart - human or non-human - actually lies and, subsequently, how the heart is created.

4. Where does the heart lie?

Returning to the topic of transhumanism, Ishiguro clearly thematizes this central topic in Capaldi's preparations for transferring Klara's mind to another android body which resembles the looks of Josie in the case of Josie's death. In particular, striving for immortality as imagined and suggested in the copying of Josie (Ishiguro 2021, 207–212) and in the transhumanistic idea of 'uploading the mind' would be, according to Taylor, deeply inhuman and destructive for 'being human'. (Taylor 2021, 106) To be sure, many transhumanist would *not* agree on this conclusion. Some transhumanist researchers would rather claim that steps toward immortality by uploading the human mind are part of an inevitable evolution. Transhumanist Ray Kurzweil, for example, describes the uploading of the human mind into an artificial system as one possible step in transhumanism (Kurzweil 2005, 198–204). Nick Bostrom, although he issues warnings about future speculative superintelligence, believes that there is a clear evolutionary line from simple organisms to such technological forms of intelligence. (Bostrom 2014) Matthew Fisher even discusses and interprets transhumanism in the light of Christian theological anthropology. (Fisher 2015)

In contrast to superhuman descriptions of artificial subjects, Ishiguro depicts the AF Klara as humane and mortal; she slowly fades away in the novel's final part. The human elite - not Rick, for example, who is not lifted - strives for immortality and enhancement as expressed in the 'uploading' or 'copying' theme and the process of lifting. Here Ishiguro's thematization parallels or even anticipates one of Taylor's final conclusions: that it is human to die and to know and accept that one is mortal (Taylor 2021, 106).

In the novel Capaldi is fully convinced that Klara, in detail, can copy Josie. He states: "The second Josie won't be a copy. She'll be the exact same ...". (Ishiguro 2021, 210). Here Klara, although she clearly is depicted as a sentient individual, means *nothing* to Capaldi. Capaldi becomes the personification of the inhumanity, destructiveness, and utilitarian view on artificial systems within transhumanism. Such uploading is, of course, in line with the transhumanistic idea of achieving some form of immortality or at least prolonged life based on future technology. However, in Capaldi's view, the process is entirely "rational" (Ishiguro 2021, 210). Therefore, Capaldi's transhumanist approach can rightly also be regarded as mechanistic.

This view seems to stand in contradiction, at least, to more religiously oriented forms of transhumanism, as described by Mercer and Trothen (2021, 34–41). This raises the question if Capaldi's emphasis upon a utilitarian, mechanistic approach also strongly points in the direction of a secular form of transhumanism? Here, one may even suspect that any form of transhumanism - even religious forms -, by way of the biological or technological enhancement process, essentially is secular. As Mercer and Trothen rightly observe, at least the view of who or what we are appears to be mechanistic (2021, 34). It could be argued that, to claim, as, for example, Christian transhumanists do (see, for example, Mercer and Trothen, 2021, 37), that the use of scientific and technological advancements in combination with following Christ, leads to spiritual growth, actually reduces spiritual growth to a mechanistic technology-based process. To be sure, this claim would need further support and elaboration. However, in the context of Ishiguro's novel it seems reasonable to interpret Capaldi's transhumanistic strivings as standing in contrast to Klara's naïve theism; a theism which in the novel's fourth part brings the issue of theodicy to a head. The AI, Klara, subsequently accepts as necessary an act of courage and sacrifice in opposition to evil, in order to procure the grace of the sun and restore Josie's health.

Furthermore, in the sense that Klara's *telos*, in the case of Josie's death, is to copy and become Josie, Klara's subjectivity is entirely functional and instrumental to the ends of ensuring Josie's 'survival.' Thus Klara is alienated from her own needs. In subsequent conversations about the human heart between Klara and Josie's father or the Manager, Klara finally concludes that "(t)here was something very special,

but it wasn't inside Josie. It was inside those who loved her. That's why I think now Mr. Capaldi was wrong and I wouldn't have succeeded." (Ishiguro 2021, 306)

One possible interpretation of Klara's conclusion and the reasoning above is the following. Both the transhumanistic process of lifting and the suggested copying of Josie focus on the *functionality* of the individual human or android mind, and presuppose that the mind can be reduced to some mechanistic process. The lifting process aims at enhancing the *functional* abilities of the child. The copying of Josie in an AF also strives to copy the *functional* behavior of Josie. Both processes are effectively based on a mechanistic view of humans and the mind.

However, Klara - there is a clear irony in the dialectical move of portraying the artificial as the most humane and letting an artificial mind make this important conclusion - concludes that whatever may be special in humans *also* lies *outside* the individual and his/her functionality. Human existence or worthy existence of interacting minds is essentially based on *relationality*. The unique part "... was inside those who loved her (Josie)" and is, as such, based on *relationality* rather than a utilitarian view of subjects or individuals. This interpretation could be related to a reading of MacIntyre's ethics by Green focusing on the *telos* in ethics (Green 2015, 203, 209). Here the *telos* in the transhumanistic enhancement of humans or the transhumanistic wish to copy a mind lies in the functionality of the individual minds. At the same time, Klara realizes that the *telos* for creating the human heart, which is unique for human interaction and possibly even a worthy ethical life, essentially most probably lies in the realm of social *relations* and the transcending of oneself, of the individual subject. Given, the central role of relationality in the novel in general and in Ishiguro's depiction of the heart in particular, a closer look at how the heart is created is at place.

5. The dialectic creation of the heart

Initially, we stated in the introduction that Ishiguro made the significant choice to focus on the subjective experience of an AF in the narrative. As already has been hinted, the fact that the AF Klara draws conclusions about the human heart could be seen as a kind of *dialectic* movement. The view that the heart lies somewhere in the human *individual* (thesis) is contrasted by the apparent experience of the reader that the heart lies in the AF as an *individual* with her love for Josie (antithesis). However, in a *synthetic* movement, the AF Klara overcomes the focus on individuals, any individuals, human or artificial. Instead, she places the creation of the human - or should one rather say creaturely - heart in the *relations* a group or a society collectively develops and supports. In that sense, even Klara's heart is not a physical reality programmed into her AF-Being - although her ability to observe, learn, and adequately interact obviously is - but rather a joint creation of those who cherish, love, and appreciate Klara as the valuable and helpful companion in Josie's and Josie's family's life.

Indeed, it is apparent that Klara's major goal in life, her aim, is to be a friend to Josie, a good, loving, caring, committed friend. Her ambition and the love Klara invests in living up to her destiny are crucially supported by her surroundings, much in the same way as Klara claims that what is unique in Josie is inside those who loved her. Josie appreciates Klara, Josie's mother 'believes' in Klara's ability to be a copy of Josie in case of Josie's death. Josie's father and Rick help Klara in her seemingly occult and secretive projects for helping and saving Josie. Even Capaldi, although he has a totally functionalist view on Klara's existence, at least initially supports Klara in that he thoroughly believes in her capabilities. This network of *relations* again is an expression of the dialectical creation of the heart.

However, Klara's surroundings tragically revoke their relational commitment to her. Having achieved her major goal of being a good AF to Josie, Klara then withdraws herself for an unspecified time into a Utility Room. Still, some caring and loving for Klara remains, expressed in the fact that Josie's mother, in a lucent moment living up to her

humanity, forbids that Klara should be reverse-engineered by Capaldi (Ishiguro 2021, 296–99); a process that evokes associations to some kind of dissection involving 'robot suffering.' Again the role of how the surrounding people *perceive* an individual, which relations they have in creating the heart of any being, is emphasized.

This move of placing the creation of the human heart in the *relations* of a group collectively develops, and the emphasis on the perception of surrounding individuals can also be seen as a focus on the *narrative* that any individual is part of, whether artificial or human. Thus the 'having a heart' is at least in part a question of whether the individual is embedded in a narrative in which he/she is ascribed 'having a heart'. Similarly, Mark Coeckelbergh has suggested a form of 'narrative responsibility' pertaining to AI (Coeckelbergh, 2021). In Klara's case, the AF would be involved in a meaning-making process and is – as is described in the novel – part of the formation of a narrative in which Klara both has a heart and bears the responsibility in the narrative of AI; a narrative both given to her and co-created by her.

The clear focus on relations in where the heart lies and how the heart is created together with the dialectic movement we have identified in the novel leads to the final question which consequences the views depicted in the novel may have for ontological positions pertaining to subjectivity or personhood.

6. Subjectivity and personhood

To start with, if the creation of the heart is seen in parallel to the perception and creation of personhood, then the question of personhood in the case of Klara could be seen as relational. According to Eddebo, understandings of personhood that emphasize the importance of the relations and processes involved are more common in Non-Western metaphysics, particularly in Japanese and African philosophy (Eddebo, 2021). In African philosophy, so Eddebo argues, "Personhood has to be achieved, and complete personhood is thus also something we can fail to reach." (Eddebo, 2021) As hinted in the introduction, the fictional human society is depicted as permeated by instrumental rationality and technological rationalism. It seems that humans and machines, at least in some sense, are interchangeable. The reader may suspect that the mainstream metaphysical position in Ishiguro's future human society indeed is some form of mechanistic, reductionist physicalism. However, the suggestion of turning to Non-Western metaphysics is further strengthened if we return to the possible reading of the 'human heart' or the heart in general as *kokoro*. Firstly, according to Kasulis, *kokoro* seems to be a relational term. Secondly, having *kokoro* is not restricted to humans (Kasulis 2011, 7–8, 11–14). Indeed, with reference to the Japanese *Shinto*- worldview, Nancy Jecker has recently argued along similar lines for a more relational and non-dualistic approach to possible artificial subjects. She emphasizes some advantages of Non-Western metaphysics in her discussion, claiming that Shinto practices "venerate nonhuman entities, rather than regarding them as belonging to a lower order". She finally concludes that humans should strive for intrinsically good and appropriate relationships with robots capable of social interaction (Jecker, 2023).

In the novel's society's internal meta-narrative, the mind, the soul, the heart of a human can ultimately be reduced to, and consequently reconstructed, through mechanistic physical processes. This latter view is perhaps most clearly realized in the transhumanistic and mechanistic ideas of Capaldi (Ishiguro 2021, 207–212). In contrast, the worldview of the artificial intelligence of Klara has traits of theism; for her, the sun is the ultimate transcendent life-giving power. In light of the focus on relations and narratives in the creation of the heart, it seems again that subjectivity and personhood do not and even cannot be fully comprehended in terms of a worldview with emphasis upon reductionism and a mechanistic understanding of the mind.

The above stated examples of Non-western metaphysics can be seen as alternatives to the latter. However, there surely also are alterna-

tive metaphysical approaches to understanding the human mind within Western metaphysics, as the recent resurgence of various forms of panpsychism shows.² Combined with the dialectical creation of the *heart*, the decisive push towards a metaphysics *not* based on a mechanistic understanding of humans *at least* suggests that to preserve human dignity closely intertwined with the human heart and the intrinsic value of subjectivity and personhood, one *should* embrace some form of non-mechanistic, non-reductionist metaphysics.

Interestingly, the metaphysics implicitly suggested by the existence of consciousness in Klara is not necessarily mechanistic. Indeed, initially, one might be tempted to think that the basic idea of placing consciousness and sentience into a humanoid may be supportive of some sort of reductive physicalism. However, firstly, the traits of theism in Klara's worldview seem to point to a dualistic view with a focus on a fully transcendent power – the Sun. To be sure, alternative forms of theism acknowledging the transcendence of a godlike divinity, such as pantheism, would also be possible choices. Secondly, while Klara's ideas about the creation of the heart (Ishiguro 2021, 306) can be interpreted as emphasizing the context and the narrative of the lives of the individuals, her ideas do *not* imply that this creation is based on a reductionist or mechanistic understanding of the mind, the heart, subjectivity, or personhood. The significant emphasis on the phenomenal experiences of Klara rather than objectively construed features of the world is a case in point. Instead, the focus is shifted from mechanistic *construction* to relational *creation*, in which the heart, the mind, subjectivity, and loving sentiment *emerge* in the narrative of the involved individuals. Thus, yet again, the push toward a version of metaphysics not based on mechanistic ideas can be discerned.

7. Conclusions and final reflections

Klara and the Sun exemplifies a new level of philosophical engagement, through literature, with pertinent issues connected to emerging technologies. While sci-fi literature has performed this sort of work for more than a century, the ascension of the genre into "high literature" underscores new and broader avenues for a synergy between art and philosophy. Ishiguro's paradoxical framing of the AI as more human than ourselves also contrasts with the genre's general approach, where dystopian as well as utopian narratives from Verne to Gibson almost without exceptions have considered the machine as something fundamentally alien. This new juxtaposition provides a novel, and quite timely, conceptual space in which the technology can serve to symbolically challenge us towards emphasizing and fostering our human distinctiveness.

In the first section of the analysis, we highlighted the dehumanizing attitude of the 'enhanced,' the 'lifted' people in Ishiguro's fictive society. These people presumably have a higher status. However, in contrast to what transhumanists claim, this transhumanistic eugenic process seems to lead to a *negative* anti-humanist attitude, at least in the children depicted in the example. Also, the entire process of enhancement seems to reduce humans to *mechanistic* beings, which can be enhanced just like we can improve machines.

The following section shows how the transhumanistic idea is further developed in the novel. Here another path is explored. While the 'lifting' process presumably is some form of biological process, the uploading of Josie's mind is based on Klara being a possible artificial host for a human mind. Two different kinds of telos were identified; the first is to preserve and guarantee a physical continuation of the human mind, and the second is the goal of establishing *relations* necessary for the creation of the human heart. Here our interpretation is that Ishiguro's novel suggests

a clear move away from a mechanistic worldview toward worldviews based on *relational* ontologies. Thus relations are placed at the center, not only in social life, but even in the understanding and construction of concepts of a more metaphysical nature.

We then argue that the creation of, for example, the human heart, can both be understood as a process embedded in a narrative and a *dialectical* process. In general, there is a dialectic tension between what *seems* to be human and what *seems* to be artificial throughout the novel. In particular, one would expect that the human heart lies somewhere in the human *individual* (thesis). Instead, the reader *experiences* that the heart lies in the AF as an *individual* with her love for Josie (antithesis). The *synthetic* movement, then, is relational; in her loving relationship, Klara transcends the boundaries between what is human and non-human, who has a heart and who does not have a heart.

The reasoning that the narrative we (or any sentient beings) are embedded in, and that we should emphasize relationality, is then transferred to the more abstract notions of subjectivity and personhood. What becomes clear is that the combination of the depiction of transhumanistic enhancement in a mechanistic worldview and the emphasis of the relational and the narrative we live by leads to the suggestion that even with regard to subjectivity and personhood, one should preferably support some form of non-mechanistic, non-reductionist metaphysics with a stronger focus on relations. Thus, subjectivity and personhood should not be understood as results of a mechanistic *construction* but rather relational *creation*.

There's here an almost Bergsonian approach to subjectivity in the affirmation that our actual being-in-the-world inevitably takes places through the intentionalities of others (Bergson 1946). The notion of subjective co-experience that takes part in and through external intentionalities, opening patterns of participation in being that are inaccessible to the isolated Cartesian self. In this sense, Ishiguro's novel attempts to expand on the Cartesian project through an innovative imaginary of the artificial person that purifies the distributed or relational subjectivity predominant in East Asian philosophy. The irreducible self-part of an immediate phenomenal subjectivity which the reductive Cartesian project hinges on is transcended by the emphasis that this self-part can only truly become manifest relationally, in meaningful and empathetic connection with another.

If one would summarize the four-step line of reasoning suggested here, one could say that there is a dialectical tension between the affirmation of the possibility of AF and thus AGI as emergent, *including the heart*, and the anti-humanistic reduction and mechanization of the mind, heart, and subjectivity as suggested in the transhumanistic and eugenic themes in the novel. In the synthesis, this opens the door for worldviews freed from purely reductionist thinking based on relationality and with a focus on the contextual surroundings and narrative. Thus the initial questions 'where can humanity be found in humans or non-humans?', 'what makes humans or non-humans human?' or 'what is the human heart?' can be answered in light of our relations and the governing narratives we *choose* to live by. Our humanity and even the possible humanity of the non-human strongly resides in our relations and our narratives. This also heavily emphasizes the central role of human *responsibility* and human *free will*. Do you choose the path of human dignity? Do we know ourselves? Do we wish to live by the *humane* and *friendly loving* rules the AF Klara lives by?

Finally, while it's possible to see a thematic criticism of scientific reductionism and industrial megatechnology in the novel, other aspects of the story nonetheless seem to express a marked utopian sentiment. Not least of these is the basic underlying fact of the artificial agent's empathetic and even loving subjectivity (which eventually amounts to an act of sacrificial piety). The artificial being is depicted as being more virtuous and humane than the actual human characters involved, and finally suffers abandonment from her friends, in spite of her acts of heroism. And in all of this, one is inclined to see something of a sleight-of-hand, a painting upon the technological Leviathan the face of loving friendship, telling us that the machine world emerging around us can very well

² For an introduction to panpsychism see for example: Goff, Philip, *Galileo's Error* (2019) or Brüntrup, Godehard, Ludwig Jaskolla, *Panpsychism - Contemporary Perspectives* (2017).

bring us real relationship. That the Matrix just might come alive and actually need us to see and cherish her. But there is a significant and almost deceptive transplantation taking place in the background here, a tacit *petitio principii* which assumes something for the machine that we ought to be quite reluctant to ascribe it.

The subjectivity that Ishiguro from the outset describes in and appropriates for his purportedly artificial character is namely never anything less than human.

Availability of data and material (data transparency)

Not applicable

Code availability

Not applicable

Authors' contributions

Not applicable

Declaration of Competing Interest

The author declares that there are no conflicts of interest.

Funding

The research in this paper is funded by the Marianne and Marcus Wallenberg Foundations WASP-HS program within the project 'Artificial Intelligence, Democracy and Human Dignity'. [MMW 2019.0160](#)

References

- Baker, L R. 2013. *Naturalism and the First-Person Perspective*: Oxford University Press.
- Benford, G. 1998. "Meaning-Stuffed Dreams: Thomas Disch and the Future of SF". *New York Review of Science Fiction* 11 (1).
- Bergson, H. 1946. *The Creative Mind*: Philosophical Library.
- Bostrom, N. 2014. *Superintelligence*: Oxford University Press.
- Brüntrup, G, Jaskolla, L. 2017. *Panpsychism - Contemporary Perspectives*: Oxford University Press Edited by Godehard Brüntrup and Ludwig Jaskolla.
- Chalmers, D J. 1996. *The Conscious Mind*: Oxford University Press.
- Chalmers, D J. 2017. "Panpsychism and Panprotopsychism". In: Brüntrup, G, Jaskolla, L (Eds.), *Panpsychism*, 19–47: Oxford University Press.
- Coeckelbergh, M. 2021. "Narrative Responsibility and Artificial intelligence". *AI& Society*, London: Springer.
- Danaher, J. 2020a. "Robot Betrayal: A Guide to the Ethics of Robotic Deception". *Ethics and Information Technology* 22 (2): 117–128.
- Danaher, J. 2020b. "Welcoming Robots into the Moral Circle: a Defence of Ethical Behaviourism". *Science and Engineering Ethics* 26 (4): 2023–2049.
- Dennett, D. 1991. *Consciousness Explained*: Back Bay Book.
- Eddebo, J. 2021. "The Faustian Machine and the Chrome Lotus: On the Diversity of Perspectives on the Metaphysics of Artificial Intelligence With a Particular Focus on the Contributions of Traditional Non-Western thought". *New Techno Humanities*: Elsevier.
- Fisher, M Z. 2015. "More Human Than the Human? Toward a 'Transhumanist' Christian Theological Anthropology". In: Mercer, C, Trothen, T J (Eds.), *Religion and Transhumanism*, 23–38: Praeger.
- Goff, P. 2019. *Galileo's Error*: Rider.
- Green, Brian Patrick. 2015. "Transhumanism and Catholic Natural Law". In: Mercer, C, Trothen, T J (Eds.), *Religion and Transhumanism*: Praeger.
- Ishiguro, K. 2021. *Klara and the Sun*: Faber.
- Jecker, N S. 2023. "Can We Wrong a Robot?" *AI & Society* (38) 259–268.
- Kasulis, T. 2011. "Cultivating the Mindful Heart". In: Swanson, P L (Ed.), *Brain Science and Kokoro*, 1–20: Nanzan Institute for Religion and Culture.
- Kurzweil, R. 2005. *The Singularity Is Near*: Duckworth Overlook.
- Mercer, C, Trothen, T J. 2021. *Religion and the Technological Future*: Palgrave Macmillan.
- Metzinger, T. 2021. "Artificial Suffering: an Argument for a Global Moratorium on Synthetic Phenomenology". *Journal of Artificial Intelligence and Consciousness* 08 (01): 43–66.
- Metzinger, T. 2004. *Being No One*: The MIT Press.
- Nolan, J, Joy, L. 2016. *Westworld Season 1*: HBO.
- Taylor, MC. 2021. "Gathering Remains". In: Taylor, MC, Carlson, TA, Rubenstein, MJ (Eds.), *Image*, 19–115, Trios: The University of Chicago Press.
- Vallor, S. 2016. *Technology and the Virtues*, Oxford: Oxford University Press.