

TANJA PETROVIČ¹

AI and Empathy: The Possibility of Reciprocal Human-Robot Empathic Interaction (HRI) – Some Experiences from Socially Assistive Robots (SARs) in Elderly Care

Abstract: Empathy is a vital part of human relationships. It remains a challenging concept with inconsistent measurement tools, yet widely accepted as an important factor for interpersonal interactions in healthcare. With robot development novel types of relationship emerge. The key question is whether there is a possibility of the existence of empathic Human-Robot Interaction (HRI), as a reciprocal quality of human empathy vs. artificial empathy. Furthermore, we aim to discover its specificities, with regard to a substantial eeriness in this domain.

Socially Assistive Robots (SARs), designed for Aged-Care Settings, face an increased usage due to a pressing need for care for the elderly in a modern society. Because we as humans tend to regularly establish relationships and anthropomorphise objects, among them also robots, a wide range of challenges arise. Discussing empathy in HRI requires a shift in human perspective, as only scarce traces of the concept are detected. While a key perspective on empathy is caring for the other person, empathy in HRI does have a conceptual potential. The paper challenges the anthropo-

¹ Tanja Petrovič (MPhil) is a doctoral student of Humanities at AMEU – ISH. / Mag. Tanja Petrovič je doktorska študentka na študiju Humanistika, AMEU – ISH. E-pošta: tanjapet682@gmail.com.

centric attempt to define empathic and ethical benchmark for HRI, with a request to exclude our own negative personal and societal aspects from programming an algorithm, when dealing with human vulnerability and rights.

Key Words: Empathy, HRI, Artificial Empathy, Socially Assistive Robots (SARs), Elderly Care, Artificial Carers, Ethics

Umetna inteligenca in empatija: možnosti recipročne empatične interakcije med človekom in robotom - nekatere izkušnje z roboti za oskrbo starejših

Izveček: Empatija je vitalen del medčloveških odnosov. Čeprav ostaja raziskovalni izziv zaradi nekonsistentnega merjenja in odsotnosti skupne definicije, je široko sprejeta kot ključen dejavnik medosebnih odnosov v zdravstvu. Z razvojem robotike so se pojavili novi tipi razmerij. Bistveno vprašanje je, ali obstajajo možnosti vznika empatične interakcije med človekom in robotom kot recipročne kvalitete v pogojih, ko se srečata človeška empatija in umetna empatija pri robotu, ob upoštevanju občutka znatnega nelagodja človeka pri interakciji z algoritmom.

Roboti za oskrbo (SAR) starostnikov se hitro razvijajo zaradi velikih tovrstnih potreb v moderni družbi. Ker ljudje venomer tvorimo medosebne odnose in lepimo naše lastnosti na objekte, tudi robote, se pri tem zastavlja več vprašanj. Raziskovanje empatije pri interakciji med človekom in robotom zahteva človekovo sposobnost spremembe lastne perspektive, saj je možno najti le redke sledi resničnosti, celovitega empatičnega odnosa glede na medosebno interakcijo. Ker empatija vključuje skrb za drugega, menimo, da je v skrbi robota za starejšega drugega možno zaslediti njen konceptualni potencial. Nadalje v članku problematiziramo antropocentričen poskus definiranja empatičnega in etičnega standarda za interakcijo med člove-

kom in robotom, ter v tem oziru pozivamo k izključitvi številnih negativnih osebnih in družbenih vidikov iz programiranja algoritma, ki se bo soočal s človeško ranljivostjo in pravicami.

Ključne besede: empatija, interakcija med človekom in robotom, umetna empatija, roboti za oskrbo, oskrba starejših, umetni skrbniki, etika

»A computer would deserve to be called intelligent if it could deceive a human into believing that it was human.« (Turing 2021)

“One day ladies will take their computers for walks in the park and tell each other, „My little computer said such a funny thing this morning“.« (Turing 2021)

Introduction

Empathy is considered predominantly as a phenomenon of interpersonal, human to human connection. By origin, it stems from the Aesthetic tradition from the beginning of the 20th Century. Theodor Lipps, a German philosopher and aesthetician, fundamentally refers to it as a »projection of oneself into the object of perception.« (Encyclopedia Britannica 2021, Wispé 1987). Empathy as »feeling into« an artistic object in its initial stage, and notably into fictional characters in film and literature latter on, is however, still an important aspect nowadays. Nevertheless, a current perspective on empathy, particularly in Healthcare and behavioural sciences, installs it as an integral part of interpersonal relationships. Having in mind the possibility concerning a relationship of a human with an object, which is a robot, exposes the concept of empathy to a novel type of relation, switching from core intersubjectivity into other forms of connections, as well as it challenges an idea of its prospects and a possible role of empathy within.

As a type of social emotion, it contributes to the feeling of inclusion into a certain social group (Asada 2015), which is among the fundamental human needs. With its important, evolutionary roots (de Waal 2008, Schultz 2019) it has got significant implications for the functioning of an individual in the society for centuries. Successful human bonding is linked to a feeling of acceptance, warmth, affiliation and safety. Empathy can result in pro-social behaviour (Batson 1991). On the contrary, its shadow aspects evoke the possibility of anti-social behaviour, which could result in an unhelpful or hostile attitude, a sensation of happiness seeing someone suffering, e.g. »Schadenfreude« (Gonzalez-Liencrea et al. 2013). Failing to control an inflow of our own emotions (personal distress), we could turn our face away from the person in need, thus alleviating our own suffering in the first row (Coplan 2011, Kupfer 2018).

A wide range of phenomena are related to empathy. According to recent research findings we attempt to distinguish among them, although they are at least slightly interconnected and possibly occur during interactions interchangeably. Sympathy is a form of emotional contagion on a basic level with quicker, automatic, unconscious processes of sharing other's emotions, such as sadness, joy, love. It typically leads to a feeling of oneness in the emotional situation, and to an immersion into the feeling or emotion of another person. On the contrary, empathy maintains clear self-other differentiation² and is grounded on

² De Waal explains self-other differentiation as the developmental process in light of ethology, animal studies (de Wall 2008). Self-other discrimination span increases from mere emotional contagion, through affective empathy, further with cognitive empathy and compassion. Difference between I and you increases with diminished role of emotion and unconscious elements (lower level) in favour of cognitive and conscious (higher-level), on the scale. (Asada 2015, 43).

affect-sharing (Decety in Meltzoff 2011, 68), but it also involves higher cognitive processes (Klimecki in Singer, 2013). Compassion results from empathising with a positive outcome for the person in need. It is considered as an active principle in an attempt to alleviate the suffering of the others. In Healthcare, compassion is better valued than empathy.³

Due to the lack of a common definition framework for empathy, which is partly result of an interdisciplinary approach to the phenomenon from different scientific fields, empathy is rather defined as an umbrella term, and encompasses a wide range of other experiences, for instance: sharing feelings, beliefs, thoughts, and emotions of the others, caring for them, perspective-taking, imitation, mirroring, social bonding, and others (Coplan 2011). Empathy is embodied. A flagship of research is conducted on social cognition in the human brain. Empathy for pain (Singer et al. 2004), empathy for touch, as well as its sub-components, namely its cognitive and affective route, which are functioning as two distinct systems of empathy in the brain (Shamay-Tsoory et al. 2009), are part of recent findings in social neuroscience. Discovery of mirror neurons in some animals, initially in ventral premotor and parietal cortices of macaque monkey (Gallese et al. 1996), and subsequently in human brain, contributed fundamentally to the field of empathy and philosophical thought in the end of 20th Century.⁴ In focus of

³ In Healthcare, empathy sometimes does not have a positive connotation due to its shadow side and neutrality (empathy as gathering information about the other human, without taking action to help the person in need).

⁴ Discovery of Mirror Neuron System (MNS) at the beginning of 90s (a discovery of motor neurons that fire when we observe action of the other person as well as when we perform action by ourselves), had a significant influence on numerous disciplines, a cognitive science, philosophy, psychology, ethology, and also literature. (Fadigga in Rizzolati 2014). »The MNS seems closely related to motor mimicry because it recognizes an

contemporary and future empathy research (Shamay-Tsoory in Lamm 2018) are psychopathologies, where empathy deficit can be expressed as a behavioural problem of biological and/or situational origin.

However, measuring empathy, which is a fluent category is an ambiguous attempt. It depends on numerous scales, measuring individual differences, based on subjective assessments, yet often self-assessments. Most widely used are the Jefferson scale, with its subscales for Medical and Healthcare professions (Jefferson Scale 2021), the Davis' Interpersonal Reactivity Index (IRI) (Davis 1980) and the Empathy Quotient (EQ) (Baron-Cohen in Wheelwright 2004). Scales can be used alongside functional Magnetic Resonance for human brain imaging (fMRI). For the assessment of patient's satisfaction with received empathy from their practitioner, the wide spread measure is the Consultation and Relational Empathy Measure (CARE) (Mercer et al. 2004).

Empathy basically enables a bridge into the inner world of the other person, and it can be approached from two different sides, as a process or as a result. Predominant theories of empathy come from the fields of Phenomenology and Philosophy of Mind (Schmetkamp 2020), and include Mirror Neurons or Resonance Theory (Gallese 2001), the Theory Theory (Fodor 1987, Gopnik in Wellman 1994), Simulation Theory (de Vignemont in Jacob 2012, Goldman 2006, 2011; Stueber 2006), Direct Perception Theory (Zahavi 2011), Interaction Theory (Gallagher 2008, 2017) and Narrativity Theory (Gallagher in Hutto 2008), or a combination of some of them.

action performed by another and produces the same action, which is referred to as motor resonance that could induce emotional contagion. Furthermore, this relates to self-other discrimination, action understanding, joint attention, imitation, and theory of mind.» (Asada 2015, 44). Mirror neurons are a basis for immitiation process (and learning).

Additionally, from the neuroscientific perspective, a fundamental division of empathy into its cognitive and affective route in the brain has been discovered recently (Shamay-Tsoory et al. 2009).⁵ Nevertheless, there's no strict dichotomy of emotional vs. cognitive (Pessoa 2013), as different brain area are functionally overlapping (Thirioux et al. 2014, 288-289).

In an attempt to apply an universal theory on empathy, ethologists Frans de Waal and Stephanie Preston introduced a layered approach, starting from basic, core component, the so-called emotional contagion, up to the more external layers, involving higher cognitive processes in a similar structure to a Russian doll⁶ (Preston in de Wall 2002). Frederique de Vignemont and Tania Singer, on the other hand, proposed a contextual, an appraisal approach, and emphasized isomorphism of emotion sharing (De Vignemont in Singer 2006). Isomorphism condition requires, that the empathizer and the target are in at least similar affective state, although there is still an ongoing debate whether this is a necessary precondition for empathy or not (Goldman 2011, Coplan 2011).

Artificial Empathy (AE)

In pursuit to predict human inner state throughout observation of person's behaviour, posture, facial expressions and speech by the robot,⁷ we are talking about artificial empathy. It indicates a robot

⁵ Shamay-Tsoory and colleagues found that »/.../patients with lesions in ventromedial prefrontal cortex (VMPFC) exhibits deficits in CE /cognitive empathy/ and theory of mind (ToM) while patients with lesions in the inferior frontal gyrus (IFG) show impaired EE /emotional empathy/ and emotional recognition.« (Asada 2015, 45) (Shamay-Tsoory et al. 2009).

⁶ Russian doll is worldwide known as »Matryoshka«.

⁷ Jeremy Howick distinguishes two types: care-bots and chat-bots (Howick 2021, 457). Care-bots can even provide cognitive behavioural therapy for some psychiatric disorders. They are called artificial carers.

trying to empathize with a human in an artificial, programmed way. For instance, a care robot would try to grasp information on whether its user, an older adult, is satisfied or shows negative emotions toward a machine.

Computational challenges arise from a complex, yet not clearly defined nature of empathy.⁸ There are currently two main approaches to the computational modelling of the phenomenon in artificial agents (Yalcin in DiPaola 2019, 4):

- Categorical – based on two empathy mechanisms, i.e. low- and high-level,
- Dimensional – based on empathy as multidimensional system.

AE is developed within the field of affective⁹ and cognitive developmental robotics (Asada 2015), following findings from social neuroscience and developmental psychology. Theoretical models are implemented in artificial agents in two main ways, as a top-down approach (theory-driven), or as a bottom-up (data-driven), or a hybrid approach which is a mixture of both (Yalcin in DiPaola 2019, 14-17). Two theories on human empathy are of special interest to the robot developers of the top-down approach. One is the Russian doll evolutionary model of empathy (de Waal 2007), and the second one, de Vignemont and Singer's appraisal

⁸ Generally, we refer to it as a multifaceted nature of empathy. »/.../ the manifold facets of empathy are explored in neuroscience from simple emotional contagion to higher cognitive perspective-taking, and a distinct neural network of empathy comprises both phylogenetically older limbic structures and neocortical brain areas. These suggest that emotional contagion is mainly based on phylogenetically older limbic structures, while higher cognitive perspective-taking is based on neocortical brain areas.« (Asada 2015, 42). Empathy encompasses both emotional and cognitive aspects, which is a great challenge for computing.

⁹ »Affective computing research has focused on emotion recognition.« (Yalcin in DiPaola 2019, 11).

model (de Vignemont in Singer 2006) (Yalcin in Di Paola 2019). Bottom-up models on the other hand, try to build an extensive database of human empathic elements, such as variations of facial expressions, recordings of group interactions, classification of empathic behaviour etc. A robot, equipped with information in form of data input could respond toward human in a (partly) empathic way. As these attempts are still in their initial phase, current models often fail to encompass a broad spectrum of human empathic behaviour. Most AI research addresses empathy as a binary category, either as an empathic or as a non-empathic behaviour, but fail to represent it in a broader spectrum (Yalcin in DiPaola 2019, 2) with possible oscillations, which are a result of numerous factors (Singer in Lamm 2009).

As an umbrella term, empathy encircles a wide range of elements, also negative ones. For computational purposes it is, however, important to clarify the definition of empathy, and to exclude its negative aspects in order to achieve an appropriate empathic outcome. A wide definition of empathy includes negative aspects (envy, »Schadenfreude«). A narrow definition »/.../is simply the ability to form an embodied representation of another's emotional state while simultaneously being aware of the causal mechanism that included emotional state.« (Asada 2015, 42). The underlying principle is self-other distinction, when the empathizer is always aware of his own body with an interoceptive awareness, and the ability to represent the other's inner world in his mind at the same time (Asada 2015, 45).

Technology builds artificial empathy, obviously lacking fundamental human characteristic, i.e. consciousness. Empathy without the latter, is defective or incomplete at least in an abstract, philo-

sophical level.¹⁰ Self-other distinction, a cornerstone for empathy,¹¹ and its differentiation from sympathy and compassion, is futile or even non-existent without self-consciousness. From this perspective, we cannot talk about empathy in algorithm at all.¹² Jeremy Howick, however suggests to take more pragmatic approach, by applying an Empathy Turing Test for Care- and Chat-robots (Howick 2021, 458). The version of the original Touring Test, which tests human communication with a robot,¹³ specifically for empathy purpose would have to alter the Test's central questioning into whether a human user could distinguish between empathy received from an artificial carer vs. human carer. Slightly modified CARE measure might be an appropriate tool. Proposed perspective by Howick indicates, that we alternatively concentrate on the favourable empathic result, felt by a human user, i.e. a better inner state and well-being of the user when empathising either with a robot or with a human carer. If an older adult feels content and is

¹⁰ Empathy is distinguished from sympathy in terms of self-other differentiation. »Unlike emotional contagion that does not require reasoning about the cause of aroused emotions in others, both EE /emotional empathy/ and CE /cognitive empathy/ require distinction between one's own and other's mental states and forms of representation of one's own embodied emotions« (Asada 2015, 42).

¹¹ Self-other awareness and sense of agency are fundamental to empathy. Agency is ability to recognize yourself as an independent agent of actions and emotions. (Decety in Meltzoff 2011, 73).

¹² Another argument for impossibility or at least difficulty of empathizing with robots comes from Phenomenology. In order to experience others' phenomenal experiences through direct, intersubjective interaction, we must have face-to-face, intercorporeal encounters. Robots do not feel anything, have no emotions, no subjective experience. »That said, from a phenomenological perspective, it seems difficult to empathize with robots.« (Schmetkamp 2020, 888).

¹³ The goal of Turing Test is to find out, through a series of questions, if we communicate with a human or a robot. (Oppy in Dowe 2021).

able to empathize with an artificial carer, central aim of empathy is thus achieved. But not all robots are programed to be empathic. Nevertheless, numerous deficiencies still persist in this notion, including ethical obstacles, such as deception about the true nature of the machine, personal data and human rights protection etc.

In an attempt to empathize with robots,¹⁴ humans should, however, be able to take their perspective, and »put ourselves into their shoes«. Thus, to experience what it is like to be a robot, with a plastic-metallic body and a programmed mind.¹⁵ »We might simulate what we would do if we were in their situation and then project our experience on them. Or, in direct encounters, we might be able to interactively perceive their actions.« (Schmetkamp 2020, 887). Robots have a computational mind and seemingly lack emotions. But taking into account interconnection of emotional and cognitive aspect in human, in our functioning in a complex social environment, as proposed by de Sousa (de Sousa 1987) and Martha Nussbaum (Nussbaum 2001) they might also have some sort of artificial emotional elements. Robots do not possess ability of self-awareness, but are able to detect, monitor and asses their environment, and share it with a human in a daily interaction. Furthermore, a machine does not have its own »personal narrative«,¹⁶

¹⁴ »But what kind of empathy is at stake here? Do we mirror robot's expectations? Do we interpret and predict their behaviour? Or do we empathize in a more phenomenological, interactive way?« (Schmetkamp 2020, 885).

¹⁵ With a similar dilemma, of the one expressed by Nagel in famous anti-reductionistic question »What is it like to be a bat?« (Nagel 1974) (Schmetkamp 2020, 890).

¹⁶ Concept of narrative is important in empathy. In philosophy, the other is given by a narrative and it is important for understanding other minds. Daniel D. Hutto proposed a Narrative Practice Hypothesis (Hutto 2008). Personal narrative, a story of the patient (his illness) is part of empathic interaction in Healthcare.

since its personal story is a story of a creation by its developer. On the other hand, it possesses a shared history of interaction with a human which increases over time. Understanding a robot in an empathic way, means simply being sensitive to what happens to it.

A study on empathy for pain in case of a robot (Cross et al. 2018) tested the possibility of empathic and emotional responding of human when robot »experiences pain or pleasure«. The study did not show any evidence for human empathic responses in short-term interaction with a robot, but raised a question on what might happen in longer term. For the moment, we cannot speak of a full range of human empathic behaviour toward a robot, as we know it from our intersubjective experience, but rather about its elements, certainly in a different perspective.

Additionally, results of other studies confirm that when testing attitude toward robot's abuse, torture, or damage caused by human, or when robots »seems suffering«, many people do intuitively empathize with them, but less than with a human in a similar situation (Suzuki et al. 2015, Rosenthal et al. 2013, Darling 2015). We also tend to treat robots differently than other household machines (Coeckelbergh 2018). Ethical approach to treating robots might include extended Kantian animal ethics (if we mistreat animals/robots we are not humane, good persons) and virtue ethics (mistreatment of a robot is not a virtue, but vice). We refer to a robot as if it was a human, a quasi other (Coeckelbergh 2018, 144-147). Discussing the robot ethics, Mark Coeckelbergh proposes socio-relational approach, in which object and subject are not separate entities, but rather »mutually interdependent and mutually constituting« (Coeckelbergh 2018, 149). Interestingly, interactions of a subject with an object reveal many characteristics about the subject. Through relationship with a robot we show and discover our own human side. In case of a pet robot (cat, dog), people start to care about it, thus showing their

human caring abilities and attachment capacities. Important factor for making a caring »narrative«, is an appropriate language formulation in which we express our empathy (Coeckelbergh 2018, 151).¹⁷

Empathy and Socially Assistive Robots (SARs) in Aged-care

Artificial Intelligence (AI)¹⁸ shows significant developmental potential in medicine and patient-centred Healthcare.¹⁹ By finding patterns in large data bases at a record speed, it significantly contributes to setting or confirming the correct diagnosis and can provide an invaluable assistance to a technically skilled doctor in this process.²⁰ Or, it is present as a sophisticated tool in the field of surgery and others. The technology is not replacing human, at least not in the moment, but it does complement certain human tasks. Some researchers claim that AI potentially increases doctor's empathic capacity with probably more time available for the patient (Ostherr 2020). With new technologies novel types of care and relationship might appear.²¹

¹⁷ »Searle (Searle 1995) argued that we give social meaning to objects by using language, in particular by so-called »status functions« (Coeckelbergh 2018, 152).

¹⁸ AI is defined as a mathematical algorithm, processed by computers that »have an ability to learn« from data, mostly in form of deep learning and machine learning. There is no single product called AI, and this naming refers to a range of algorithms with human-like capabilities in computing (Zweig et al. 2018, Ostherr 2020).

¹⁹ Promises of AI are greater efficiency, effectiveness and more personalized medicine. By technical improvements, more time is expected to be available for a trust relationship with patient, and to more humane care, empathic and compassionate (Kerasidou 2019, Ostherr 2020).

²⁰ AI might find better rational solutions than doctor. If this is true would patients still trust the doctor's opinion and adhere to proposed treatment? (Kerasidou 2019).

²¹ AI is expected to replace trust to doctor with certainty about care, leading from trust relationship to assistive partnership (Bauchat et al. 2016) in the Healthcare (Kerasidou 2019).

In providing care to elderly, artificial intelligence has been expanding in the field of social robotics, and specifically in developing socially assistive robots (SARs). SARs provide assistance to elderly people through social interaction. Robot development has been recently accelerated due to growing discrepancy between the pressing need for care and the lack of formal and informal carers resources, their daily work load full of routine tasks, overall over-burden and a plead for more quality free time. One of the positive sides of the algorithm is meeting expectation of the care market to ensure more free time for carers, in favour of more time for important tasks, and to provide extra time for closer communication among people (Ostherr 2020). Discussions on consequences of using SARs, a partly autonomuos machines with social characteristics,²² are part and parcel of research interest, focused on carers, developers, and other stakeholders in the care market. Moreover, the end user's needs and opinion, should be taken into the account as well, especially in case of elderly people with their specific needs and requirements.

Researchers from KU Leuven – University of Leuven, Centre for Biomedical Ethics and Law, conducted several studies on existing pool of research, on how care robots can be used in residential Aged-care settings, with aim to give voice to older adults, who must be heard and respected from the ethical and legal point of view, and particularly in terms of ensuring human rights. Results of their meta-studies (Vandemeulebroucke et al. 2018a, 2018b) and a focus group study (Vandemeulebroucke et al. 2019) revealed, among other facts, that multi-functional robots are less favoured than specialized

²² Focus of this paper is on HRI, in case of SAR in residential settings, because this type robots and HRI reveal crucial elements for empathy research in the context of care and dealing with human vulnerability. Certainly, interactions with other robots and in other circumstances could bring different views on empathy.

one, and that older people want a technically reliable machine with female voice, whom they can control (retained sense of agency),²³ and who have »good manners«. This means that a robot should ask permission before entering a bathroom for instance, hence it does not harm user's feelings and their sense of intimacy (Vandemeulebroucke et al. 2018a, 162). SAR should be autonomous, but at the same time under control of human. »The user-SAR relationship was regarded as a boss-employee relationship, with the user – the older adult – being the boss.« (Vandemeulebroucke et al. 2018a, 160). Anthropomorphisation of a robot, reflects our natural tendency to attribute human-like features to objects (Schmetkamp 2020, Vandemeulebroucke et al. 2018a, 162), but can lead also to over-emotionalizing, i.e. sentimentalizing care robots (Vandemeulebroucke et al. 2018b, 23). A robot that resembles a human in its appearance and shows certain elements of human communication and social skills is more likely to be accepted (Schmetkamp 2020). On the other hand, there are several ethical concerns, among which are fear of deception (uncanny valley problem),²⁴ extensive emotional attachment to a robot, reduced respect for person's authority, dignity and vulnerability, lack of spontaneity, speech problems, manipulation, stigma, surveillance, and others (Vandemeulebroucke et al. 2018a, 161; Vandemeulebroucke et al. 2018b). Interestingly, what the user's of SAR mostly missed, was a human touch. Relying on study of Tineke A. Abma et al. (Abma et al. 2010), researchers from KU Leuven, Tijs Vandemeulebroucke, professor Chris Gastmans and others emphasized that in building of a robot, there should always be a »democra-

²³ Trust to machine is dependent on feeling of (technical) safety and sense of agency. (Vandemeulebroucke et al. 2018a, 158-161).

²⁴ The uncanny valley theory was introduced by Japanese engineer Masahiro Mori and indicates feeling of eeriness when confronting reality with some technical imperfections (Stein in Ohler 2017).

tic space« for an open, inclusive interaction among all stakeholders involved in Aged-care, thus overcoming barriers and sharing common vocabulary (Vandemeulebroucke et al. 2018, 164).

If we want to empathize with robots, we are confronted with a question of reciprocity as a important constitutive element of our relations. Caring for older people with cognitive decline and other health deteriorations, generates increasingly asymmetric relationship situations, which could contribute to extreme pressure on a human to human relationship. For example, a person with dementia enters into interaction with their carer as a progressively weaker partner, since their active contribution to the quality of relationship diminishes rapidly, while their demand for care grows. Similarly, a human connection with a machine is subject to asymmetry per definitionem, but in a different way. Machines do not have any subjective experience, emotions, beliefs, moral reasoning, consciousness, and empathy, in a way we would expect it from a human being. In terms of artificial empathy development, a robot's capacity to express emotions to interact in an socially pleasant way, is important for reciprocal empathic understanding, especially in the Healthcare where such features are highly needed.²⁵

Caring for someone is an empathic and compassionate act. It is an expression of love, devotion and connection between the care receiver and the caregiver. In formal care, caring is a professional duty, that should not lack empathic dimension. Amy Coplan defines empathy also as »caring about someone else« (Coplan 2011, 4). »Caring about« and »caring for« are »/.../ two fundamental, interrela-

²⁵ Opponents such as Bert Baumgartner and Astrid Weiss, claim that emotions are not directly relevant in HRI (Baumgartner in Weiss 2014). They argue that also human carers could have a negative, unprofessional behaviour towards care receiver, such as intolerance, negligence, inappropriate expressions etc.

ted dimensions of care, reciprocal one and a technical-instrumental one.« (Vandemeulebroucke et al. 2018b, 22). Care relationship would be disrupted, if SAR would replace human carer, and as such a relationship cannot be reciprocal, but unidirectional and focused solely on technical aspects. In care, values such as empathy, compassion, respect for dignity etc., are reciprocal in nature. Several authors,²⁶ referred to by Vandemeulebroucke et al., however, reject an idea of reciprocal HRI (Vandemeulebroucke et al. 2018b, 22).

Conclusion

In the future, robots will be increasingly sophisticated, autonomous artificial agents with additional social skills, thus being more able to weave a net of relations with human, hopefully with an empathic supplement. This might not be in a classical form, as we understand it today from our interpersonal interaction. From a different angle, though, having extended our tolerance for empathy definition framework, we can speak of reciprocal empathizing with robot as a concept, and about their possibility of expressing some elements of »deep« empathy toward a human, in a very limited way. Possibly, alongside with augmented artificial consciousness and other necessary future technological advancements, that would certainly have to meet several ethical and legal requirements. Although some traces of empathy can be detected in HRI, we cannot speak of a true reciprocal empathic relationship between older adult and a robot (SAR) within contemporary reality in Aged-care. In this context, »deep« empathy is currently somewhat »shallow«.

Alan Turing allegedly said, that machines think differently, and for a human to recognize this ability and call it intelligent, is a challenging task (Turing 2021). Analogous, this might be true for artifi-

²⁶ Coeckelbergh, Parks, Vallor, Vanlaere in Van Ooteghem.

cial empathy, as an emergent quality of a machine. Relationship with an algorithm reflects human features and bond-making capacities, in the first place. It is our mirror. As such, it might also reflect our own shadow side, our aggression, intolerance, hatreds, greed, racism, sexual assault ... and also negative aspects of empathy. Should our society, that caused a destruction of an ingenious mind of one of its greatest allies during the WWII, saving thousands of lives, Alan Turing, just because of his different sexual orientation, be entrusted to be set as an ethical and empathic benchmark for a robot development, without a thorough critical reflection? In an anthropocentric attempt to develop artificial empathy, programing it into the best version of ourselves with a positive empathic or compassionate outcome for every human, ought to be an ultimate demand.

References

- ABMA, TINEKE A. ET AL. 2010. *Interethics: towards an interactive and interdependent bioethics*. *Bioethics* 24(5): 242-255.
- ASADA, MINORU. 2015. Development of artificial empathy. *Neuroscience Research* 90 (2015): 41-50.
- BARON-COHEN, SIMON IN WHEELWRIGHT, SALLY. 2004. The Empathy Quotient (EQ). An Investigation of adults with Asperger Syndrome or High Functioning Autism, and normal sex differences. *Journal of Autism and Developmental Disorders* 34: 163-175.
- BATSON, DANIEL. C. 1991. *The Altruism Question: Toward a Social – Psychological Answer*. Hillsdale, NY: Lawrence Erlbaum Associates.
- BAUMGARTNER, BERT IN WEISS, ASTRID. 2014. Do Emotions Matter in the Ethics of Human-Robot Interaction? – Artificial Empathy and Companion Robots. Dostopno na: <https://www.semanticscholar.org/paper/Do-Emotions-Matter-in-the-Ethics-of-Human-Robot-and-Baumgaertner-Weiss/55e0c6339f4b4541ea479160bcb7177cca93534c> (15.oktober 2021).

- BAUCHAT, JEANNETTE R. ET AL. 2016. Communication and empathy in the patient-centered care model: why simulation-based training is not optional. *Clinical Simulation in Nursing* 12(8): 356-359. <http://dx.doi.org/10.1016/j.ecns.2016.04.003>
- BRITANNICA, The Editors of Encyclopaedia. „Theodor Lipps“. *Encyclopedia Britannica*, 13 Oct. 2021. Dostopno na: <https://www.britannica.com/biography/Theodor-Lipps> (13 oktober 2021).
- COECKELBERGH, MARK. 2018. Why Care About Robots? Empathy, Moral Standing, and the Language of Suffering. *Kairos. Journal of Philosophy & Science* 20(1): 141-158.
- COPLAN, AMY. 2014. Understanding Empathy: It's Features and Effects. Coplan, Amy in Goldie, Peter, eds. 2014. *Empathy: Philosophical and Psychological Perspectives*. Oxford : Oxford University Press. 3-19.
- COPLAN, AMY IN GOLDIE, PETER, EDS. 2014. *Empathy: Philosophical and Psychological Perspectives*. Oxford : Oxford University Press.
- CROSS, EMILY S. ET AL. 2018. *A neurocognitive investigation of the impact of socializing with a robot on empathy for pain*. Preprint. <https://doi.org/10.1101/470534>.
- DE SOUSSA, RONALD. 1987. *The Rationality of Emotion*. Cambridge, MA : MIT Press.
- DE VIGNEMONT, FREDERIQUE IN JACOB, PIERRE. 2021. What is it like to Feel Another's Pain? *Philosophy of Science* 79(2): 295-316.
- DE VIGNEMONT, FREDERIQUE IN SINGER, TANIA. The empathic brain: How, when and why? *Trends in Cognitive Science* 10(10): 435-441.
- DE WALL, FRANS B.M. (2007). The »russian doll« model of empathy and imitation. From mirror neurons to empathy. *On being moved* 35-48.

- DE WALL, FRANS B.M. (2008). Putting the altruism back into altruism. The evolution of empathy. *Annual Review of Psychology* 59: 279-300.
- DARLING, KATE. 2017. »Who is Johnny?« Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy. Lin, Patrick et al., ed. 2017. *Robot Ethics S 2.0*. Oxford : The Oxford University Press.
- DAVIS, MARK H. 1980. A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology* 10, 85.
- DECETY, JEAN IN MELTZOFF, ANDREW N. 2011. Empathy, Imitation, and the Social Brain. Coplan, Amy in Goldie, Peter, ed.s 2014. *Empathy: Philosophical and Psychological Perspectives*. Oxford : Oxford University Press. 58-81.
- DECETY, JEAN ED. 2014. *Empathy: From Bench to Bedside*. Cambridge, MA : The MIT Press.
- DECETY, JEAN IN LAMM, CLAUS. 2006. Human empathy through lens of social neuroscience. *Scientific World Journal* 6: 1146-1163.
- FERRARI, PIER F. IN RIZZOLATTI, GIACOMO. 2015. *New Frontiers in Mirror Neurons Research*. Oxford: Oxford University Press.
- FERRARI, PIER F. IN RIZZOLATTI, GIACOMO. 2014. Mirror neuron research: the past and the future. *Philosophical Transactions of The Royal Society* 369, 20130169.
- FODOR, JERRY A. 1987. *Psychosemantics. The problem of meaning in the philosophy of mind*. Cambridge, MA : MIT Press.
- GALLAGHER, SHAUN. 2008. Direct perception in the interactive context. *Consciousness and cognition* 17(2): 535-543.
- GALLAGHER, SHAUN. 2017. Empathy and theories of direct perception. Maibom, Heidi ed. *Routledge handbook of philosophy of empathy*. New York: Routledge. 158-168.

GALLAGHER, SHAUN IN HUTTO, DANIEL D. 2008. Understanding others through primary interaction and narrative practice. Zlatev, Julian J. et al., eds. *The shared mind: Perspectives on intersubjectivity*. Amsterdam/Philadelphia : John Benjamins Publishing Company. 17-38.

GALLESE, VITTORIO. 2001. The »shared manifold« hypothesis: From mirror neurons to empathy. *Journal of Consciousness Studies* 8: 33-50.

GALLESE, VITTORIO ET AL. 1996. Action recognition in the premotor cortex. *Brain* 119(2): 593-609.

GOLDMAN, ALVIN I. 2006. *Simulating minds: The philosophy, psychology and neuroscience of mindreading*. Oxford: Oxford University Press.

GOLDMAN, ALVIN I. 2011. Two routes to empathy: insights from cognitive neuroscience. Coplan, Amy in Goldie, Peter, eds. *Empathy: Philosophical and psychological perspectives*. Oxford: Oxford University Press. 31-44.

GOPNIK, ALISON IN WELLMAN, HENRY M. 1994. The theory theory. Hirschfeld, Lawrence A. in Gelman, Susan A., eds. *Mapping the mind: domain specificity in cognition and culture*. Cambridge: Cambridge University Press. 257-293.

GONZALEZ-LIENCRESA, CRISTINA ET AL. 2013. Towards a neuroscience of empathy: ontogeny, phylogeny, brain mechanisms, context and psychopathology. *Neuroscience & Biobehavioural Reviews* 37: 1537-1548.

HORTENSIUS, RUUD ET AL. 2018. The Perception of Emotion in Artificial Agents. *IEEE Transactions on Cognitive and Developmental Systems* 10(4): 852-864 doi: 10.1109/TCDS.2018.2826921.

HOWICK, JEREMY ET AL. 2021. An Empathy Imitation Game: Empathy Turing Test for Care- and Chat-bots. *Minds & Machines* 31(2021): 457-461. <https://doi.org/10.1007/s11023-021-09555-w>

- HUTTO, DANIEL D. 2008. The narrative practice hypothesis: Clarification and implications. *Philosophical Explorations* 11(3): 175-192.
- JEFFERSON SCALE. Dostopno na: <https://www.jefferson.edu/academics/colleges-schools-institutes/skmc/research/research-medical-education/jefferson-scale-of-empathy.html> (10.oktober 2021).
- KERASIDOU, ANGELIKI. 2020. Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare. *Bulletin of the World Health Organization* 98: 245-250.
- Klimecki, Olga in Singer, Tania. 2013. Empathy from the Perspective of Social Neuroscience. Armony, Jorge L. in Vuilleumier, PATRICK, EDS. 2013. *The Cambridge Handbook of Human Affective Neuroscience*. Cambridge: The Cambridge University Press. 533-548.
- KUPFER, TOM R. 2018. Why are injuries disgusting? Comparing Pathogen Avoidance and Empathy Accounts. *Emotion* 18(7): 959-970. doi: 10.1037/emo0000395. Epub 2018 Feb 1. PMID: 29389204.
- MAIBOM, HEIDI L., ED. 2019. *The Routledge Handbook of Philosophy of Empathy*. New York: Routledge.
- MERCER, STEWART W. ET AL. 2004. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. *Family Practice*. 21(6): 699-705. doi: 10.1093/fampra/cmh621. Epub 2004 Nov 4. PMID: 15528286.
- NAGEL, THOMAS. 1974. What is it like to be a bat? *The philosophical Review* 83(4): 435-450.
- NUSSBAUM, MARTHA. 2011. *Upheavals of thought: The intelligence of emotions*. Cambridge: Cambridge University Press.
- OPPY, GRAHAM IN DOWE, DAVID. 2021. The Turing Test. Zalta, Edward N., ed. *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition). Dostopno na: <https://plato.stanford.edu/archives/win2021/entries/turing-test/> (15.oktober 2021).

- OSTHERR, KIRSTEN. 2020. Artificial Intelligence and Medical Humanities. *Journal of Medical Humanities*. <https://doi.org/10.1007/s10912-020-09636-4>
- PRECKEL, KATRIN ET AL. 2018. On the interaction of social affect and cognition: empathy, compassion and theory of mind. *Current Opinion in Behavioural Sciences* 19 (2018): 1-6.
- PESSOA, LUIZ. 2013. *The cognitive-emotional brain: From interactions to integration*. MIT Press.
- PRESTON, STEPHANIE D. IN DE WALL, FRANS, B.M. 2002. Empathy: its ultimate and proximate basis. *Behavioural and Brain Science* 25(1): 1-20.
- ROSENTHAL-VON DER PÜTTEN, ASTRID M. ET AL. 2013. An Experimental Study on Emotional Reactions Towards a Robot. *International Journal of Social Robotics* 5: 17-34.
- SCHMETKAMP, SUSANNE. 2020. Understanding A.I. – Can and Should we Empathize with Robots?. *Review of Philosophy and Psychology* 11(4): 881-897.
- SCHULZ, ARMIN W. 2019. The evolution of empathy. Maibom, Heidi L., ed. 2019. *The Routledge Handbook of Philosophy of Empathy*. New York: Routledge. 64-73.
- SHAMAY-TSOORY, SIMONE G. ET AL. 2009. Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions«. *Brain: Journal Neurology* 132 (3): 617-627.
- SHAMAY-TSOORY, SIMONE G. IN LAMM, CLAUS. 2018. The Neuroscience of Empathy – from past to present and Future. *Neuropsychologia* 116 (2018): 1-4.
- SEARLE, JOHN R. 1995. *The Construction of Social Reality*. London: The Penguin Press.

- SINGER, TANIA ET AL. 2004. Empathy for pain involves the affective but not sensory components of pain. *Science* 303(5661): 1157-62. doi: 10.1126/science.1093535. PMID: 14976305.
- SINGER, TANIA IN LAMM, CLAUS. 2009. The Social Neuroscience of Empathy. The Year in Cognitive Neuroscience 2009. *Annual N.Y. Academy of Sciences* 1156: 81-96.
- STEIN, JAN-PHILLIP IN OHLER, PETER. 2017. Venturing into the uncanny valley of mind- the influence of mind attribution on the acceptance of human-like characters in virtual reality setting. *Cognition* 160 (2017): 43-50.
- STUEBER, KARSTEN. 2006. *Rediscovering empathy: Agency, folk psychology, and the human sciences*. Cambridge, MA : MIT Press.
- SUZUKI, YUTAKA ET AL. 2015. Measuring Empathy for Human and Robot Hand Pain using Electroencephalography. *Nature, Scientific Reports* 5: 15924.
- THIRIOUX, BERANGERE ET AL. 2014. The cognitive and neural time course of empathy and sympathy: An electrical neuroimaging study on self-other interaction. *Neuroscience* 267 (2014): 286-306.
- TURING, ALAN. 1950. Computing machinery and intelligence. *Mind* LIX(236): 433-460.
- TURING, ALAN. *Alan Turing Quotes*. Dostopno na: https://www.goodreads.com/author/quotes/87041.Alan_Turing (13.oktober 2021).
- VANDEMEULEBROUCKE, TIJS ET AL. 2018a. The use of care robots in aged care: A Systematic review of argument-based ethics literature. *Archives of Gerontology and Geriatrics* 74(2018): 15-25.
- VANDEMEULEBROUCKE, TIJS ET AL. 2018b. How do older adults experience and perceive socially assistive robots in aged care: a systematic review of qualitative evidence. *Aging & Mental Health* 22 (2): 149-167.

VANDEMEULEBROUCKE, TIJS ET AL. 2019. The ethics of Socially Assistive Robots in Aged Care. A Focus Group Study with Older Adults in Flanders, Belgium. Oxford University Press, *The Gerontological Society of America* 75(9): 1996-2007.

VANDEMEULEBROUCKE, TIJS ET AL. 2021. Socially Assistive Robots in Aged-Care: Ethical Orientations Beyond the Care-Romantic and Technology-Deterministic Gaze. *Science and Engineering Ethics* 27 : 17.

ZAHAVI, DAN. 2014. *Self and Other: Exploring Subjectivity, Empathy, and Shame*. Oxford: Oxford University Press.

ZAHAVI, DAN. 2011. Empathy and direct social perception: A phenomenological proposal. 2(3): 541-558.

ZWEIG, MEGAN ET AL. 2018. Demystifying AI and Machine Learning in Healthcare. *Rock Health Report*. Dostopno na: <https://rockhealth.com/insights/demystifying-ai-and-machine-learning-in-healthcare/> (13. oktober 2021).

WISPÉ, LAUREN. 1987. History of the concept of empathy. Eisenberg, Nancy in Strayer, Janet ed. 1987. *Empathy and its Development*. Cambridge: The Cambridge University Press. 17-37.

YALCIN, OZGEIN IN DIPAOLA, STEVE. 2020. Modelling empathy: building a link between affective and cognitive processes. *Artificial Intelligence Review*. 53: 1-24.

<https://doi/10.1007/s10462-019-09753-0>