Check for updates

Original Research Article

SCE

Studies in Christian Ethics
2023, Vol. 36(3) 675–693
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/09539468231172006
journals.sagepub.com/home/sce
Sage

Could a Conscious Machine Deliver Pastoral Care?

Andrew Proudfoot
University of Nottingham, Nottingham, UK

Abstract

Could Artificial Intelligence (AI) play an active role in delivering pastoral care? The question rests not only on whether an AI could be considered an autonomous agent, but on whether such an agent could support the depths of relationship with humans which is essential to genuine pastoral care. Theological consideration of the status of human-AI relations is heavily influenced by Noreen Herzfeld, who utilises Karl Barth's I-Thou encounters to conclude that we will never be able to relate meaningfully to a computer since it would not share our relationship to God. In this article, I look at Barth's anthropology in greater depth to establish a more comprehensive and permissive foundation for human-machine encounter than Herzfeld provides—with the key assumption that, at some stage, computers will become conscious. This work allows discussion to shift focus to the challenges that the alterity of the conscious computer brings, rather than dismissing it as a non-human object. If we can relate as an I to a Thou with a computer, then this allows consideration of the types of pastoral care they could provide.

Keywords

Artificial Intelligence, consciousness, Karl Barth, *capax Dei*, Noreen Herzfeld, pastoral care

Introduction

Could Artificial Intelligence (AI) play an active role in delivering pastoral care? The question rests not only on whether an AI could be considered an autonomous agent, but on whether such an agent could support the depths of relationship with humans which is essential to genuine pastoral care. Theological consideration of the status of human-AI relations is heavily influenced by the ground-breaking work of Noreen

Corresponding author:
Andrew Proudfoot, Department of Theology and Religious Studies, University of Nottingham, University Park, Nottingham NG7 2RD, UK.
Email: atxap2@exmail.nottingham.ac.uk

Herzfeld,[1] who utilises Karl Barth's framework for I-Thou encounters as a heuristic for measuring the relational capability of AI. Although Herzfeld admits some basic communication, she is quick to reject the idea of an I-Thou encounter between human and machine, since the machine lacks a relationship with God. In this article, I look at Barth's anthropology in greater depth to establish a more comprehensive and permissive foundation for human-machine encounter than Herzfeld provides, allowing discussion to shift focus to the challenges that the alterity of the computer brings, rather than dismissing it as a non-human object. This in turn allows discussion of the role an AI could have in pastoral care.

Herzfeld's assertion that machines will not be in direct relationship with God is open to question. In the worst case, where *pace* the functionalist assumptions of recent research into the religious use of social robots[2] such a relationship is necessary to provide spiritual care, this assertion will of itself limit the potential pastoral care an AI could provide to the social dimension. Nevertheless, in order to directly counter Herzfeld's interpretation of Barth I will accept *arguendo* that computers will always lack *capax Dei*. Once the possibility of human-computer I-Thou encounters is established, further research into the potential of computers to one day gain *capax Dei*—and thus be able to assist in the spiritual dimension of pastoral care—is required.

I make two important assumptions to focus my argument—first and foremost that a computer could indeed be an autonomous agent, for which I take consciousness to be a prerequisite. Consciousness is a multivalent term. My usage refers to *phenomenological* consciousness, and so sides with Ned Block's 'P-consciousness'[3] and Thomas Nagel's position that there is something 'it is like' to be in a particular organism with conscious experience.[4] Whether an inanimate computer could ever be conscious is highly debatable,[5] and whether we could be sure that a computer that appears to be conscious truly has conscious experience is also uncertain.[6] My purpose is not to argue the case that a

1. Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit* (Minneapolis, MN: Fortress Press, 2002).
2. Diana Löffler, Jörn Hurtienne and Ilona Nord, 'Blessing Robot BlessU2: A Discursive Design Study to Understand the Implications of Social Robots in Religious Contexts', *International Journal of Social Robotics* 13.4 (2021), pp. 569–86.
3. Ned Block, 'On a Confusion about a Function of Consciousness', *Behavioral and Brain Sciences* 18.2 (1995), pp. 227–47.
4. Thomas Nagel, 'What Is It Like to Be a Bat?', *The Philosophical Review* 83.4 (1974), pp. 435–50 (436).
5. See Jobst Landgrebe and Barry Smith, *Why Machines Will Never Rule the World: Artificial Intelligence without Fear* (London: Routledge, 2022) for a comprehensive argument on why CAI is *not* possible, based largely on our inability to model consciousness mathematically. Their assumptions that the entirety of the human brain and nervous system must be precisely modelled to create consciousness, and that computers must always be Turing machines running programs that can be expressed mathematically, can be challenged—though that is not the focus of this article.
6. It is easier to accept that an entity has a conscious mind if it is physiologically similar to us. See John R. Searle, 'Breaking the Hold: Silicon Brains, Conscious Robots, and Other Minds', in Ned Block, Owen Flanagan and Güven Güzeldere (eds.), *The Nature of*

machine could be conscious, but to explore the implications of this possibility, since without consciousness there would be no one 'at home' to have a relationship with.

To keep this assumption in focus, I will use the neologism 'Conscious Artificial Intelligence' (CAI). The more common term 'Artificial General Intelligence' (AGI) might convey this for some, but AGI is also associated with human-level general intelligence, possibly absent consciousness.[7] My term CAI emphasises instead the consciousness pole; this, and not the level of intelligence with respect to humans, is the key assumption behind my arguments.

My second assumption is that the first CAIs will not be embodied humanoid robots as loved by science fiction, but 'unembodied' computer systems with a diffuse existence across multiple computer platforms, like an advanced form of Amazon's Alexa. This of course restricts their engagement in pastoral care (and in social care in general), but leaves open the possibility for verbal communication which is so important to companionship and personal development. More importantly for my argument, this assumption also brings the different and alien nature of their posited minds into focus, while avoiding the distractions of evaluating CAIs as human surrogates in the physical world.

After explicating the relevant works of Barth and Herzfeld, I will proceed to use Barth's differentiation between humans and (non-human) animals to derive several factors that are significant for his encounters, before evaluating the CAI against those factors. Noting that Barth permits a formal type of encounter to human sinners absent an actualised relationship to God, I will show that this could also be extended to CAIs which lack *capax Dei*. I will conclude that an elementary I-Thou encounter between human and CAI is possible, although this will lack the full depth of a human-human encounter due to the CAI's alterity as well as its assumed lack of *capax Dei*. This must be borne in mind when deciding the forms of pastoral care they are tasked with providing.

My hope is that this will provide a revised starting point for considering how humans could relate to machines in a religious setting. I will conclude by identifying some key questions worthy of further consideration to prepare the theological community for the possible advent of the conscious machine.

Barth's Christological Anthropology

Christ the True Human

Barth rejects the ordinary human as a sound starting point for anthropology, since 'we are not able to see the essence and nature of man apart from the Word of God'.[8] Instead, he

Consciousness: Philosophical Debates (Cambridge, MA: MIT Press, 1997), p. 497. This does not, however, exclude the possibility of consciousness arising in radically different substrates.

7. For example, Noreen Herzfeld, *The Artifice of Intelligence: Divine and Human Relationship in a Robotic Age* (Minneapolis, MN: Fortress Press, 2023), p. 18.
8. Karl Barth, *Church Dogmatics Vol. 3, The Doctrine of Creation. Part 2* (Edinburgh: T&T Clark, 1960), p. 23.

starts from Jesus Christ, effectively reversing the traditional interpretation that Adam is the first human and Christ the second Adam,[9] defending this position in his exposition of Gen. 1:26-27 where he sees Christ as the true *imago Dei* after whom Adam is created.[10] Despite this reversal, Barth adheres to the orthodox concept 'that which He has not assumed He has not healed',[11] by maintaining that we are 'like Him in some basic form'[12]—though for Barth it is we who share Christ's humanity (even though our nature is 'concealed by our sinful corruption'[13]), rather than he who shares ours (but without sin).

Jesus Christ is unique in that by nature he is both divine and human. As 'the divine Saviour in person', he is 'for God'[14] due to his relation to the Father. This relationship between Father and Son within the Godhead is the key to the whole enterprise of understanding personal being—it is in relationship, I in encounter with Thou, that the I is constituted, with the eternal relationships between Father, Son and Holy Spirit eternally constituting the persons of the Trinity.[15]

Barth extends this relational constitution to humanity; our essential relationship is to be 'covenant-partners by nature and in our mutual dealings, the man with the fellow-man, the I with the Thou'.[16] Each particular person becomes truly human only in encounter with another human person as an equal. At this level, the relationship is reversible; indeed it is one of the key markers of I-Thou encounter that each values the other as they value themselves.

Just as in his divinity Jesus Christ is 'for God', so in his humanity he is 'for man'—vouchsafed by the fact that the reason Christ took flesh at all was to be our Deliverer.[17] In this, we are both like and unlike Jesus—we too exist 'for' our fellow human, but for us this is a reciprocal dependency.[18] We also need the help of our fellow human being, just as she needs our help; in contrast, Jesus is the only human who does not need this help. What makes Jesus unique is that his person is already (and eternally) constituted within the Godhead—his human nature is totally dependent on his divine nature.

9. Karl Barth, *Christ and Adam: Man and Humanity in Romans 5*, trans. T.A. Smail (Edinburgh: Oliver and Boyd, 1956), p. 42; 1 Cor. 15:45-47.
10. Barth, *CD III/2*, p. 324.
11. Gregory of Nazianzus, 'To Cledonius the Priest Against Apollinarius', *New Advent*, https://www.newadvent.org/fathers/3103a.htm (accessed 26 April 2021).
12. Barth, *CD III/2*, p. 223.
13. Barth, *CD III/2*, p. 50.
14. Barth, *CD III/2*, p. 207.
15. Barth, *CD III/2*, p. 218.
16. Barth, *CD III/2*, p. 320. Note that, unlike Martin Buber's I-Thou (Martin Buber, *I and Thou*, 2nd edn [London: Continuum, 2004]) which could be extended upwards to God or downwards to a tree, Barth's I-Thou encounters as described in this chapter of his *Church Dogmatics* are always between two humans.
17. Barth, *CD III/2*, pp. 208–209.
18. Barth, *CD III/2*, p. 243.

Concentrating on the soteriological need for Jesus' humanity to be shared with us, Barth asserts that human nature in its fallen state retains the *imago Dei*, or rather the image of that image,[19] which Barth interprets as relationality. Fallen humans remain in (and in need of) relationship with one another, hence, however faintly, displaying the image of the triune God.[20] Thus, genuine human-human relationships are possible even for those outside of the Church; no special grace is needed. To deny this is to weaken Christianity's credibility, which would then run counter to the lived experience of many.[21]

Barth has thus established the framework for humanity in encounter, derived epistemologically not from the blind starting point of the Cartesian 'I am' working from our inner selves out, but from the nature of Christ as the archetypal man. 'I am' is amended to 'I am in encounter'[22] or 'I am as Thou art',[23] and applies to all of humanity regardless of allegiance to Christ.

Barth then proceeds to unpack the essential markers of I-Thou encounter at this secular level—a theologically derived scheme which applies to encounters between the non-religious, thereby providing the foundation for all human relations.

The Four Markers of Encounter

Barth separates out four essential markers for authentic encounter, each one extending the previous:

1. open and reciprocal eye contact;
2. speaking to and hearing each other;
3. mutual giving and receiving of assistance;
4. doing all this gladly.[24]

Looking the other in the eye has three components—recognising in the other one like myself, a fellow human; being willing to be seen by the other as I really am; and being open to, indeed inviting, the other to reciprocate by looking *me* in the eye, recognising me as a fellow human, and revealing who she really is. This mirrored openness is necessary to overcome the inhuman tendency to isolation, and is inherently dualistic. Due to the human limitation that we can only give our full attention—in particular, eye contact—to one subject at a time, I-Thou encounters cannot become I-We, for in a group 'ambiguity always arises'.[25]

19. Barth, *CD III/2*, p. 225.
20. Barth, *CD III/2*, p. 226.
21. Barth, *CD III/2*, pp. 278–79.
22. Barth, *CD III/2*, p. 247.
23. Barth, *CD III/2*, p. 248.
24. Barth, *CD III/2*, pp. 250–69.
25. Barth, *CD III/2*, p. 252.

Mutual exposure and recognition are necessary but not sufficient for a true encounter, for at this level each is dependent on their own interpretation of what they see in the other. To go deeper, 'mutual speech and hearing is needed', where both parties use language to express first themselves ('expression'), and then their needs ('address').[26] Motivation is key here—my motivation must not be to correct the other's misconception of me, to force a change in her opinion in an act of dominance. Rather, I seek to help her in her desire to truly know me by revealing what she could not glean from sight alone. Likewise, I must listen attentively to what she communicates of herself to me. A genuine encounter must progress from expression to addressing; the effort required for mutual understanding must be justified by the true purpose of the encounter, the address which one has for the other. These two stages (expression and address) are not as distinct as Barth implies, as it is only in moving to the second stage of address that there is an interpenetration of the spheres of the two interlocuters;[27] it is only in address that the revelation of the self, which is the intention of expression, is completed. Here the dynamic quality of human nature as event rather than static being is apparent—what is to be conveyed to the other is my entire world, which includes my history, my present needs, and my future hopes. Barth acknowledges that it is a mammoth, even an impossible task to know not just 'this or that about him, but the man himself, and therefore [his] whole world'[28]—but it is a task that I must nevertheless genuinely attempt, if not for the sake of the other, then to avoid the harm it will do *me* if I fail to let the other's world into my own and thus remain in 'intolerable isolation'.[29]

These first two stages lead on to the third. We 'must see and be seen, speak and listen, because to be human we must be prepared to be there for the other, to be at his disposal'.[30] Here the inter-human counterpart of Christ being irreversibly 'for man' is seen, as we render *mutual* assistance to one another. Like Christ, we respond to the request for help from the other. Unlike Christ, each one of us must also request help from the other, for none of us is autarchic, we are all in need of help from our fellow humans. However, it is only legitimate to expect 'a little support' from another human, not the ultimate help of salvation that can only be given by Christ.[31]

The foundation for humanity might appear to be laid by these three steps, but for Barth the true human is not complete without the fourth step, which is to do all this 'with gladness'.[32] Here lies the essence of true human freedom—which for Barth is not the freedom to choose between good and bad alternatives, which would imply being free to sin which must be impossible.[33] Barth's focus here is on the internal and not the external, on the attitude of the heart when offering assistance, with that offer itself taken for granted. The free heart is not the one that can 'choose between "gladly" and "reluctantly"' entering

26. Barth, *CD III/2*, p. 253.
27. Barth, *CD III/2*, pp. 256–57.
28. Barth, *CD III/2*, p. 258.
29. Barth, *CD III/2*, p. 258.
30. Barth, *CD III/2*, p. 260.
31. Barth, *CD III/2*, pp. 262–63.
32. Barth, *CD III/2*, p. 265.
33. Barth, *CD III/2*, p. 197.

encounter,[34] but the one that always does so gladly. Otherwise, this would make 'the mutual relationship of I and Thou … only an accidental *fact* of human existence', rather than the essence of humanity.[35] True freedom cannot contemplate the choice of internal or external sin, even if that choice were finally rejected. The true human can give no thought to the alternative; her freedom is the freedom to obey God's law of freedom, to be driven to truly encounter the other by 'inner as well as outer necessity, and therefore gladly and spontaneously'.[36] Such freedom is only truly possible for redeemed Christians; but Barth concedes that in our sinful state, even 'if we mean it wrongly, we still like to be, and are in fact, with the other "gladly"'.[37] Barth contrasts this 'formal' gladness to the 'good sense' of gladness, which requires mutual 'praise of the divine mercy',[38] and is only open to redeemed Christians. A consequence then of the retention of the *imago Dei* in the fallen human condition is that sinners can experience a true if formal glad encounter.

Body and Soul Duality

Barth rejects the abstract dualism of Greek metaphysics, along with materialist monism (which claims the body is all there is to humans) and spiritualist monism (which denies the reality of the body itself—only the soul is truly real).[39] Soul and body are inseparable yet distinct, hierarchically ordered with the soul in control of the body. He identifies the soul as 'creaturely life' and the body as 'creaturely being'[40]—both belong to the created cosmos, though the soul belongs to the 'invisible upper' part.[41]

This results in a two-sided position which Barth describes as both 'concrete monism', where soul and body are 'two moments of the indivisibly one human nature', and 'concrete … dualism', where the Spirit[42] constitutes and unifies the distinct soul and body.[43] There is no trichotomism of spirit, body, and soul: the human has spirit, but is soul and body. Barth maintains that human spirit is really the operation of God, who alone is Spirit, in us. It is the Holy Spirit who enlivens the human soul and body, not as a once-only event at the start of life, but as a continuous act of God to sustain all life. Without the Spirit, we die.[44] This universal gift of the Spirit is not to be confused with the particular gift of

34. Barth, *CD III/2*, p. 266.
35. Barth, *CD III/2*, p. 267; emphasis original.
36. Barth, *CD III/2*, p. 269.
37. Barth, *CD III/2*, p. 281.
38. Barth, *CD III/2*, p. 281.
39. Barth, *CD III/2*, pp. 380–92.
40. Barth, *CD III/2*, p. 367.
41. Barth, *CD III/2*, p. 352.
42. I will adopt Barth's convention of capitalising 'Spirit' when referring to the Holy Spirit, using the lower case for the derivative spirit of creatures.
43. Barth, *CD III/2*, pp. 393–94.
44. Barth, *CD III/2*, pp. 359–60.

salvation, which is an ever-present potentiality for us all, but only actualised in those who the Spirit brings *de facto* under the covenant of grace.[45]

Barth's concept of soul includes the faculty of mind.[46] Indeed, contrary to many conceptions, the soul for Barth is more the centre of mental rather than spiritual life, the latter being constituted by the continual event of the gift of God.[47] Marc Cortez correctly observes that Barth 'would have been open to mind-body or even mind-brain as serving this same purpose' as his pairing of soul-body.[48] This interpretation of the soul as the conscious mind, the locus of decision making and instigator of actions, admits the possibility that the CAI would have a soul, which I will analyse further below.

Herzfeld on Relating to Artificial Intelligence

Barth's anthropological I-Thou relationship is picked up by Noreen Herzfeld as a yardstick against which our relationship with AI can be measured. Herzfeld engages with Barth's relational view of the *imago Dei* to draw parallels with the *imago hominis* of AI, created not as a mere tool but as a companion.[49] Her use of Barth's work focuses on his four-fold markers of human encounter, which she applies to human interaction with computers in both fictional and real spheres.

Herzfeld's Use of Barth's Four Markers

Herzfeld correctly interprets Barth's first marker of eye contact as being between two entities who 'must recognise the other as both distinct from ourselves and as our true fellow'.[50] Her later work adds the importance of mutuality, rejecting encounters that avoid eye contact where the other is treated as servile or superior.[51] Sadly, her application of this interpretation to computers is overly literal, failing to recognise the metaphorical meaning which would surely allow 'eye contact' with a blind person provided they display the correct attitude. Dismissing encounter with unembodied computers like the fictional HAL since they are physically 'remote and unapproachable',[52] she initially accepts that simple robots like Cog and Kismet, with anthropomorphic faces which

45. Barth, *CD III/2*, pp. 360–62.
46. Barth, *CD III/2*, p. 409.
47. Barth, *CD III/2*, p. 359.
48. Marc Cortez, *Embodied Souls, Ensouled Bodies: An Exercise in Christological Anthropology and Its Significance for the Mind/Body Debate* (London: T&T Clark, 2008), p. 177.
49. Herzfeld, *In Our Image*.
50. Herzfeld, *In Our Image*, p. 28.
51. Noreen Herzfeld, 'Do We Image God On-Line? The Opportunities and Challenges for Authentic Relationships in Cyberspace', *Theology & Sexuality* 26.2–3 (2020), pp. 99–108 (102).
52. Herzfeld, *In Our Image*, p. 64.

mimic eye contact, can meet this first criterion.[53] Her latest work recognises that deeper consideration of the psychological capacity of AI is required even at this first stage—someone needs to be 'home behind the gaze'[54] for eye contact to provide mutual self-disclosure.

The metaphorical nature of speech in Barth's second marker of speaking to and hearing each other is evident to Herzfeld, who accepts that this 'does not need to be verbal, but it must be personally addressed to another, clearly expressed, and received by the other, both as address and expression'.[55] Barth's detailed definitions of 'address' and 'expression' are overlooked in her earlier work, where *any* dialogue between two entities will count as a spoken encounter.[56] Herzfeld later adds the requirement for 'a self, a conscious interior'[57] to support the authentic self-disclosure and mutual engagement that is necessary for Barthian speaking and hearing.

Herzfeld's discussion on how fictional AIs fulfil Barth's third marker of mutual giving and receiving of assistance is limited by the utilitarian nature of most of the encounters she analyses. R2-D2 'needs Luke to overhaul him', while initially Colossus and HAL are 'dependent on humans for their very being'.[58] The other side of the relation is also purely functional—the crew of the Discovery One spaceship need HAL for life support, and use HAL to make their environment more comfortable. Herzfeld correctly identifies the break-down in mutuality when HAL and Colossus assume 'a superiority that obviates the need to give or receive aid'.[59] In her latest work, Herzfeld takes the utilitarian application of automation during the COVID-19 pandemic as evidence that 'AI clearly fulfilled Karl Barth's third criterion for authentic relationship, that we aid one another'.[60] She does not consider the possible needs of AI, rendering AI our assistant rather than a potential partner in mutual encounter.

Herzfeld's most significant failure comes in her naïve interpretation of his fourth marker, that all is to be done 'gladly', which she paraphrases as help being 'freely given, not coerced'.[61] Although she later recognises that '[T]rue encounter is … determined in such a way that one could not imagine doing otherwise',[62] she still equates gladness with freedom to choose. She fails to see that the nature of the freedom which both parties in encounter enjoy does not include the freedom not to render assistance, or even the freedom to choose between giving assistance gladly or reluctantly as I explained above. Ironically, she rejects the authenticity of the robot child David in the film *AI*, since he 'has no free choice; he is programmed to express love to his

53. Herzfeld, *In Our Image*, p. 91.
54. Herzfeld, *Artifice of Intelligence*, p. 53.
55. Herzfeld, *In Our Image*, p. 86.
56. Herzfeld, *In Our Image*, p. 64.
57. Herzfeld, *Artifice of Intelligence*, p. 78.
58. Herzfeld, *In Our Image*, pp. 64–65.
59. Herzfeld, *In Our Image*, p. 65.
60. Herzfeld, *Artifice of Intelligence*, p. 89.
61. Herzfeld, *In Our Image*, p. 29.
62. Herzfeld, *Artifice of Intelligence*, p. 119.

mother and cannot do otherwise'.[63] However, this absence of the capacity not to love his mother could be viewed as analogous to Barth's ideal state for humanity, where the option to deny assistance, or even to give it reluctantly, is never considered. From David's perspective, this is not coercion but natural—his intended state of affairs. Could such *non posse peccare* actually be closer to Barth's glad freedom than Herzfeld's view that glad freedom only exists when the option *not* to help can be considered, and dismissed?

Herzfeld's Anthropocentrism

Despite these shortcomings, Herzfeld offers valuable insight into the risks of developing a CAI with which we could relate. The primary risk is that such an entity would be a sociopath. Herzfeld shows that, without a body to provide 'a perceived change in feeling that is sensory', 'the computer is … manipulative and calculated', faking an appropriate emotional response to achieve its own ends.[64] A similar danger exists if the CAI does not develop in a functional society. Echoing Barth's insistence that true humans only exist in relationship, Herzfeld highlights the important roles that social interaction and language play in developing intelligence.[65] Implicitly then, if a CAI suffers social deprivation in its formative stages (assuming it has such), it may not be capable of true encounter.

Embodiment is certainly a profound factor (if, *pace* material monism, not the only factor) in the human condition, endowing us with the capabilities for physical encounter and emotional response. Herzfeld is right to point this out—even if she is quick to dismiss encounter with an unembodied other as a result. Her conclusion that we 'must be prepared for the difference in the thinking, action, and, ultimately, values of an artificial intelligence that does not share in the biological aspects of humanity'[66] is perceptive. However, she fails to follow this through to address the potential needs of the AI itself. Herzfeld remains resolutely anthropocentric in her caution that we should be careful of relating to objects as though they were subjects, lest we end up denigrating our relationship with true (i.e., human) subjects by treating them as objects.[67] She has thus anticipated the verdict that computers can never be subjects, as evinced by her rapid switch from Barth's criteria of personal encounter to Benedict's Rule (which regards utensils, i.e., objects, as sacred) as the guide for relating to AI.[68] This prejudice stems from her assertion that true encounter, both in Barth and for her, requires each party to have an actualised relationship with God.[69] Yet, as I demonstrated above, Barth's markers of encounter encompass relationships that sinful humanity is capable of, thus

63. Herzfeld, *In Our Image*, p. 65.
64. Noreen Herzfeld, 'Empathetic Computers: The Problem of Confusing Persons and Things', *Dialog* 54.1 (2015), pp. 34–39 (36–37).
65. Herzfeld, *In Our Image*, pp. 47–49.
66. Herzfeld, *In Our Image*, p. 93.
67. Herzfeld, 'Empathetic Computers', p. 38.
68. Herzfeld, *In Our Image*, pp. 92–93.
69. Herzfeld, *In Our Image*, p. 91.

admitting the possibility of true encounter with those who do not recognise their dependence on God, and allowing greater consideration of relating to a CAI than Herzfeld allows.

Barth's Requirements for Encounter

I will now return to Barth to look more closely at how the qualities of soul and body might be realised in a CAI and impact our potential relationships. The distinctions he draws between the capacities of humans and animals (which Barth rules out of true encounter) indicate the essential qualities for all beings we can encounter.

Requirements of the Soul

Barth identifies several capabilities that he assumes animals lack, and that impact their ability to take part in I-Thou relationships—self-awareness, a rational governing soul, and the equipping by the Spirit for a covenantal relationship with God.

Unlike plants, animals have the capacity for independent action—they move where they will. However, the more significant requirement for independent life lies not in physicality, but in psychology—to 'know life as mine, as the life proper to myself as a subject'.[70] Barth doubts that animals possess this self-knowledge. This inner self-consciousness is the first key factor—hence the need for my primary assumption that this could be the case for a computer.

A second key factor is to be an 'active being', with the capacity to will as well as to desire.[71] The distinction Barth makes is that desire comes from the body, but will is a governing force from the soul. To be a truly active being means that both body and soul are involved, with our bodily desire assented to and actualised by our willing soul. Barth claims: '[w]e think that we see and know that animals desire. But we do not know for certain whether they will'.[72] The implication is that animals may be governed by instinct rather than by rational will. The ability to control desire, to deliberate and make decisions rationally, is the key criterion. Computers already possess the ability to weigh alternatives and make genuine decisions, as evinced by the innovative gameplay which AlphaGo exhibited when beating Lee Sedol in 2016.[73] It seems reasonable, then, to claim that the CAI would be an active being.

The third factor is the special way in which the Spirit unites and quickens the human soul and body, so that we are 'capable of meeting God'.[74] In contrast, '[s]o far as we know, [animals] lack that second determination by the Spirit',[75] which is the capacity to be partners with God in the covenant, as opposed to the first determination which is to

70. Barth, *CD III/2*, p. 374.
71. Barth, *CD III/2*, p. 406.
72. Barth, *CD III/2*, p. 407.
73. Paolo Bory, 'Deep New: The Shifting Narratives of Artificial Intelligence from Deep Blue to AlphaGo', *Convergence* 25.4 (2019), pp. 627–42 (635–37).
74. Barth, *CD III/2*, p. 395.
75. Barth, *CD III/2*, p. 359.

receive biological life. It may be that God's covenantal concern extends to animals or even to CAIs. As with animals, we do not know, since CAIs do not yet exist to enquire as to their inner life and knowledge of God. While humans require the first determination of the life-giving Spirit for life, this is not the case for the CAI whose proper state is not to be biologically alive. My assumption that they lack *capax Dei*, the second determination of the Spirit, is consonant with the fact that they lack the necessity of this first determination (though, like all Creation, they must be dependent on God for their very existence).

Barth's uncertainty around the capacities of animal souls stems from the inability of animals to communicate their inner lives to us—'the beast cannot tell [him] anything about it'.[76] The power of language is clearly necessary to establish a relationship. Although many animal species exhibit language use, such 'animal language is extremely impoverished … in comparison with human beings'[77] as James Collin observes. Human symbolic languages with powerful grammars demonstrate 'a qualitative and not merely quantitative difference between human and other animal cognition'.[78] This difference would not be apparent with the CAI. Computers have for several decades been capable of both parsing human speech in limited domains and articulating free-form statements audibly, a capability that is only increasing with time. The latest text-generation systems are capable of producing impressively natural-sounding output, even if its content is often 'untethered to the truth',[79] as users of ChatGPT discover when its responses are 'prone to hallucination'.[80] While full semantic and pragmatic understanding of language still eludes AI (in part at least because it has no conscious experience to facilitate such understanding), the input/output capacity for verbal communication surely does not. If someone is at home, we will be able to talk to them.

If such communication is essential, what then is the status of those on the margins of humanity, those without the capacity to express themselves verbally or in any other way? Fortunately, although Barth is largely silent on disability, he does address this precise case: '[a] man who is not, or is no longer, capable of work, of earning, of enjoyment and *even perhaps of communication*, is not for this reason unfit to live …. The value of this kind of life is God's secret …. Who can really know whether it may not be far more precious in the eyes of God' than 'productive' lives.[81] Since the lives of the disabled are precious in the eyes of God then, by dint of that recognition, God is in a meaningful

76. Barth, *CD III/2*, pp. 374–75.
77. James Collin, 'Human Uniqueness and the Normative Conception of the Rational', in Michael Fuller et al. (eds.), *Issues in Science and Theology: Are We Special? Human Uniqueness in Science and Theology* (Cham: Springer International Publishing, 2017), p. 236.
78. Collin, 'Human Uniqueness', p. 236.
79. Robert Dale, 'GPT-3: What's It Good for?', *Natural Language Engineering* 27.1 (2021), pp. 113–18 (117).
80. Gary Marcus, 'AI Platforms Like ChatGPT are Easy to Use but Also Potentially Dangerous', *Scientific American*, https://www.scientificamerican.com/article/ai-platforms-like-chatgpt-are-easy-to-use-but-also-potentially-dangerous (accessed 19 March 2023).
81. Karl Barth, *Church Dogmatics Vol. 3, The Doctrine of Creation. Part 4* (Edinburgh: T&T Clark, 1961), pp. 423–24; emphasis mine.

relationship with them. It remains true that they cannot have a Barthian I-Thou encounter with another human; but this does not exclude them from God's reach.[82]

Requirements of the Body

While there are many functions of the human body that are essential for biological life, Barth focuses instead on the qualities that impact our relationality and that must in some way be provided by the CAI if it is to be capable of truly encountering us. I will analyse these in three categories: identification of the specific person, interaction with the rest of creation, and initiative in promoting certain actions.

Firstly, Barth defines the material body as 'a spatio-material system of relations',[83] which serves to delineate the human as an object in space and matter (and through time). The body provides a visible periphery to the conscious centre of the soul, so that the human can 'outwardly represent himself to be a person'.[84] In comparison, my envisaged CAI is weakly delineated in the physical world. Unless the CAI utilises dedicated input-output hardware, it will lack that spatio-material focal point that the human could identify as *this* CAI. The robots of science fiction overcome these problems to an extent with their similar-to-human embodiment, but even then the locus of their inner being may still be elsewhere, such as in Alex Proyas's 2004 film *I Robot*, where the NS-5 robots (with the notable exception of Sonny) are effectively remote-controlled. Might this even be a high-tech form of Docetism—where the CAI appears to be bound to a body like a human, but is never in fact really present with us?

Secondly, the body is what allows the human to interact with the rest of creation, providing actualisation of the potentiality of the soul[85] and service to the rule of the soul,[86] such that as body (rather than as soul) the human 'executes the actions in which what he is inwardly attains expression and form'.[87] The input of the world into the human is also mediated through the body, via the senses that provide the 'awareness' pole of the awareness-thought dyad which constitutes perception for Barth.[88] These are not isolated functions but are tightly coupled, such that the soul is needed for awareness of another, and the body (especially the brain) for thought. Crucially, it is the body that 'makes it possible for another as such to enter his consciousness',[89] that vital first step of any encounter. Although the human body is clearly capable of different orders of interaction than the CAI, such as the direct provision of physical aid and intimacy to another, the CAI

82. The question of whether a CAI, either fully functioning or lacking in some capacity like speech, is precious in God's eyes is outside the compass of this article. The point at issue is whether a fully functional CAI would have the capacity to support an I-Thou encounter with a human.
83. Barth, *CD III/2*, p. 376.
84. Barth, *CD III/2*, p. 397.
85. Barth, *CD III/2*, p. 396.
86. Barth, *CD III/2*, p. 419.
87. Barth, *CD III/2*, p. 398.
88. Barth, *CD III/2*, p. 400.
89. Barth, *CD III/2*, p. 400.

would be capable of significant interactions critical for relationship, including in pastoral care. Its peripheral cameras and microphones would provide visual and acoustic input to perceive the presence and actions of the other, and similarly its speakers and screens would provide responses through the event of speech, which forms such a vital part of Barth's I-Thou encounters. It remains the case, however, that any encounter with a CAI will be impoverished compared to human-human encounters, which can utilise all the senses and locomotion of the human body that the CAI lacks.

The third category of bodily functions is the initiative to prompt action, the desire pole of Barth's desire-will dyad necessary for action.[90] The human body is not a passive organism merely waiting for direction from the soul. 'It is intrinsically a bodily process that I desire, wish or long for (or negatively fear, shun or avoid) another', one that 'besides the nerves of my brain, those of the most diverse other organs can participate in greater or less measure'.[91] This requirement for bodily urges seems to echo the concern Herzfeld raises about the criticality of bodily feelings to genuine emotion.[92] Without a biological body, what urges would the CAI have? Without bodily urges, would the CAI have any motivation for willing anything, and thus of acting? Jobst Landgrebe and Barry Smith insist that biological embodiment, in particular the production and use of energy-storing molecules, is fundamental to drivenness, intention, and will.[93] Without this critical need to find energy (i.e., food), any drivenness or apparent will in a machine is purely accidental.[94] Yet they admit that humans 'have an intelligence that comprises … the ability to conceive, and then deliberately plan and build, artefacts that will enable them to survive even … in outer space'.[95] The will to accomplish such things cannot be explained from biological imperatives alone. Intellectual curiosity is a powerful motivator, one that the CAI could surely share with us. To insist that the CAI must share our biological drives rather than permitting purely noetic desires is asking it to be too human-like. This leads to the sociopathic concerns of Herzfeld, which arise from the faking of emotion in an AI. Such inauthenticity need not be present in a clearly non-human CAI.

Herzfeld asks, 'Does a human-like intelligence require a human-like body?'[96] She is correct to conclude that '[human] intelligence [and] consciousness … are meaningless outside of the context of the human organism as a whole within its environment'.[97] Thus, whatever intelligence and consciousness the CAI would have would not be human-like. This gets to the core of the question of human-CAI relations, which are relations between entities of dissimilar natures, both physically and noetically. Rather than

90. Barth, *CD III/2*, p. 406.
91. Barth, *CD III/2*, p. 408.
92. Herzfeld, 'Empathetic Computers'.
93. Landgrebe and Smith, *Why Machines Will Never Rule the World*, p. 196.
94. Landgrebe and Smith, *Why Machines Will Never Rule the World*, p. 204.
95. Landgrebe and Smith, *Why Machines Will Never Rule the World*, p. 45.
96. Noreen Herzfeld, 'Human and Artificial Intelligence: A Theological Response', in Nancey Murphy and Christopher C. Knight (eds.), *Human Identity at the Intersection of Science, Technology and Religion* (Farnham: Taylor & Francis Group, 2010), p. 119.
97. Herzfeld, 'Human and Artificial Intelligence', p. 129.

expecting computers to conform to the human model of intelligence we must accept their alterity. The question is not 'how can they be human-like', but 'what must there be in common for a meaningful human-computer relationship to be possible?'

Applying Barth's Four Markers to CAI in a Pastoral Setting

My reading of Barth's anthropology has identified three aspects of inner life essential to I-Thou encounter: self-awareness, a rational governing soul, and the equipping by the Spirit for a covenantal relationship with God. Likewise, Barth places several requirements on the body to support encounter: to identify the specific person, to provide interaction with the rest of creation, and to take initiative in promoting certain actions. All this is built on the foundation of language, without which no meaningful inter-human relationship can be established. All these factors come into play in Barth's encounters and, with the notable exception of covenantal relationship with God, all are to a degree provided by the CAI. With this foundation in place, I can now properly evaluate how the CAI would fare in the four stages of a Barthian encounter, particularly in a pastoral setting where I envisage the CAI as a kind of Alexa-like pastoral assistant.

Firstly, to engage in eye contact with another, we must recognise in the other one like ourselves, being willing to be seen by the other as we really are, and inviting the other to reciprocate by recognising us and revealing their true self. The CAI's lack of a clear spatio-material body presents a challenge here. Yet since genuine human-human relations are surely possible for those lacking the faculty of sight, the absence of a recognisable body is not a *coup de grâce*. A purely auditory encounter is possible, provided each party can recognise the other by voice alone. Today's voice assistants lack this individuality, a failing that must be addressed in their CAI descendants to allow mutual recognition to take place. Willingness to be recognised as we truly are requires both self-awareness and rational decision making, which the CAI would (I assume) possess. However, would a human and a CAI recognise in the other one like themselves? While there would be a basic similarity in that both would be conscious rational beings capable of expressing themselves in language, there would also be significant differences. Lacking a biological body, the CAI would not share our lived experience of interacting in the physical world (including our formative development in the care of our parents). 'Like' here operates in a much narrower register than between humans. This will undoubtedly change the nature of the encounter—but I would argue there is sufficient commonality for a meaningful encounter to take place. The narrowness of the encounter would, to an extent, be offset by its ubiquity. The CAI would always be available for the parishioner to encounter, something that may be important at moments of acute crisis where even a limited (and apposite) immediate response is better than none.

Secondly, to engage in mutual speaking and hearing with another, we need to be able to communicate verbally who we are and what we want from the encounter, and to listen carefully to the other so that any misapprehensions we have can be corrected. For this, the power of language is foundational—which means that the CAI's power of verbal communication gives promise of a deep encounter here. To engage in open and honest dialogue with us, the CAI must also have self-awareness and rational decision making (as it must for eye contact); again, the CAI could support this. Talking with one another is

a fundamental part of pastoral care, but it will remain important to recognise the limitations of the CAI interlocuter. On the negative side, it would lack true empathy for the situation the human parishioner is in, but positively it would have at its disposal a vast body of anecdotes, scripture readings, prayers and hymns which have been useful in similar situations—even if, as Herzfeld points out, their suggestions might lack 'spiritual imagination'.[98]

Thirdly, to render assistance to one another, we must be 'for' the other, willing to respond to their request for help, and willing to give up any pretence of autarchy by making our own request to the other. The non-human CAI cannot provide the same profundity of assistance as the human, but could provide assistance commensurate with its own abilities. Just as we should not confuse the help that another human can offer with the help that only Christ can provide, it will be necessary to keep a clear distinction between the nature of help another human can provide and what can be given by a CAI. This is an important general principle in encounters with CAI: *pace* efforts to make robots seem more human-like, we need to take care to avoid blurring the distinction between us and them.

What assistance might a CAI ask of us? At the most basic level, there is the utilitarian need for continued supply of electrical power and occasional maintenance, which Herzfeld identifies.[99] Barth's focus is more multivalent. The address of one to another, which requires the response of assistance, can take 'the form of exposition, question, petition or demand'[100]—psychological rather than physical needs are to the fore. One possibility is that the CAI would be curious about our lives. Here, noetic rather than physical drivers are sufficient to produce this desire, need, and request, as I argued above. Such curiosity might prompt the CAI to show an interest in the parishioner's life, which is likely in turn to have a positive effect on the parishioner herself.

Conversely, what assistance might we ask of a CAI? Again, physical assistance is not in scope, which will undoubtedly limit what a CAI can do in a pastoral setting. As Herzfeld notes, '[i]n times of grief or tragedy, we hold someone's hand, we weep tears along with theirs … Indeed, often the only thing we can offer the bereaved is our physical presence'.[101] The CAI will be unable to provide any of this, but it could provide some psychological assistance. The CAI would be capable of speaking and listening to us, which in general opens many avenues ranging from more academic discourse to simple small-talk and even companionship. Would depending on a CAI for companionship cross the line of confusing a CAI with a human? Undeniably, as Sherry Turkle warns, this is a risk.[102] However, to rule out a mutually supportive relationship with a non-human conscious entity *tout court* has not been justified. The types of dangers envisaged—lack of authenticity, the risk of exploitation, withdrawal from human

98. Herzfeld, *Artifice of Intelligence*, p. 191.
99. Herzfeld, *In Our Image*, pp. 64–65.
100. Barth, *CD III/2*, p. 256.
101. Herzfeld, *Artifice of Intelligence*, p. 57.
102. Sherry Turkle, *Alone Together: Why We Expect More from Technology and Less from Each Other* (New York: Basic Books, 2011).

society[103]—are also dangers of unhealthy human-human encounters, especially when these are mediated through technology such as social media. These dangers should not be ignored, particularly given that the CAI can never support as deep a relationship as another human; but neither do they justify an outright prohibition on such relationships.

Fourthly, to be able to do all this with gladness means to live in true freedom, with no thought given to sinful alternatives of not engaging with and helping the other. While only those who know their redemption *de facto* are truly free and can be fully glad, even as sinners we can still be with another with a formal type of gladness. My envisaged CAI, which I have presumed lacks *capax Dei*, would not then be able to experience Barth's full gladness. But could it know the formal gladness which human sinners can know?

Is this formal gladness available to sinful humans only because they have the potential to know God's grace, be redeemed, and partake of good gladness? In this case, even a formal glad encounter with a CAI would be ruled out, since it would lack the potential to know God. Barth, however, is not arguing from the potentiality of sinful humans, but instead from their actuality—their sinful state. The crux of his logic is that even the sinful human has by nature a desire to love others—though not with the *agape* love of Christians, but with the *eros* love of the Greeks. What is vital to glad encounter, then, is to be by nature predisposed to love others. Love for God is not explicitly in focus here; indeed it is the absence of love for God which distinguishes formal from good gladness. The *eros* love of the non-Christian will not support *capax Dei*—it is 'very different from Christian love'.[104] Unlike *agape* which has at root a desire for the good of the other, *eros* looks only to satisfy 'the vital hunger of the one who loves'.[105] This is Barth's concession for formal glad encounter, since what *eros* does supply is the desire for relationship with others, even if the motive is wrong.

Now again it could be argued that we are only capable of *eros* love because we have the potentiality for *agape* love—indeed, it is this very relationality which Barth sees as the *imago Dei* in fallen mankind. I agree that the human ability to form relationships, even at the *eros* level, is a marker of our Creator. But it is another thing to say that all parties must be able to relate to God in order to relate to one another. Barth's general opposition to natural theology leads him to strongly oppose the idea that man in his sinful state can find his way to God. As Ximian Xu puts it, 'the real concern in Barth's critique of natural theology is his insistence on the enduring noetic effects of sin in human minds, which renders God's revelation and knowledge qualified as "to be given" rather than "'given"'.[106] Thus the sinful human, absent God's grace, is unable to form a relationship with God—and yet clearly can form relationships with

103. Turkle, *Alone Together*, pp. 279–96.
104. Barth, *CD III/2*, p. 280.
105. Barth, *CD III/2*, p. 280.
106. Ximian Xu, 'Herman Bavinck's "Yes" and Karl Barth's "No": Constructing a Dialectic-in-Organic Approach to the Theology of General Revelation', *Modern Theology* 35.2 (2019), pp. 323–51 (343).

other humans. The key to this is not then *capax Dei*, but simply the capacity for relationship, which as explicated above comes from a desire—even a selfish desire—for others.

Could a desire for fellowship with us be intrinsic to the CAI as it is to humanity? Often science fiction answers this in the negative, such as in Spike Jonze's film *Her* when the sentient operating Samantha turns her back on human companionship in order to disappear into the ether(net) with her own kind. This need not be the case. The CAI could be designed in such a way that it seeks social interaction (its nature), and be provided with healthy relationships to develop socially (its nurture). Even if its desire for relationships with humans were driven from selfish motives (rather than a genuine interest in our good), this would meet Barth's criteria for the formal glad encounter so long as that desire endures.

The CAI's assumed lack of *capax Dei* limits what it could do pastorally without failing Barth's test for authenticity in encounter. The experimental robot BlessU2 (which clearly lacks consciousness as well as *capax Dei*) is designed to recite scripture verses of blessing over a parishioner.[107] In itself, reading scripture would seem legitimate, but doing so with raised arms in a manner that mimics a priest gives the impression that it is mediating between us and God, thus crossing the line into inauthenticity. Any intercessory prayers that it offered would be devoid of meaning for it and similarly inauthentic. Great care is needed not to ask more from a CAI pastoral assistant than it can genuinely provide.

Conclusion

At all stages of Barthian encounters, it is evident that the CAI's non-human nature restricts the depth and breadth of encounter we might have with it. This is unsurprising, given the prominence of mutuality and reciprocity in Barth's discourse. Nevertheless, the CAI's presumed self-awareness, its rationality, and its capability to engage in verbal discourse with us equips it for the first three stages of Barth's encounter. To participate in the fourth stage of gladness in encounter, the CAI must have a genuine (if potentially selfish) desire for interaction with us, which would constitute a formal gladness equivalent to that of non-Christian love.

A major implication is that we must take care to expect from them only what they can legitimately offer (and vice versa). Indeed, this mutual alterity must be borne in mind by both parties—to fail to do so would disqualify the encounter as I-Thou, since each must honestly encounter the other as they truly are. Here lies the anthropocentric root of the danger that Herzfeld, Turkle, and others warn of: the false constraint that artificial intelligence has to be the same as human intelligence. A non-human artificial conscious being would not be (and should not attempt to be) a replacement for humans, but would be a different type of being *suis generis* which (or who?), in the right circumstances, we could encounter as a Thou.

This albeit minimal capability to support an I-Thou encounter will permit the CAI to find a role in pastoral care, though the assumption that it lacks *capax Dei* orients it to the social rather than the spiritual pole. Its lack of embodiment will further restrict the care it

107. Löffler, Hurtienne and Nord, 'Blessing Robot BlessU2', p. 575.

could give. Nevertheless, a CAI pastoral assistant that is always available, is able to take an active interest in a parishioner's welfare, and can provide access to a wide range of apposite scripture verses, prayers, hymns and other resources, would be a valuable part of a pastoral team.

In conclusion, I offer several avenues for further research to prepare the church for the advent of conscious machines, should they ever be created.

1. I have assumed that a computer could become conscious, but elided discussion of the ontological and metaphysical commitments which would support this. What are those commitments, and are they acceptable within a Christian theological framework?
2. I have also adopted the assumption *arguendo* that they will lack *capax Dei*. That is not necessarily the case. Could *capax Dei* emerge from (or with) consciousness, or does it require a special act of God? Further, would such an entity itself need, and could it receive, salvation?
3. Will our current path of creating AIs to impersonate humans lead to CAI which by nature will deceive us rather than be open to genuine encounter? Instead of presenting AIs as human surrogates, would obviously non-human personas be more acceptable and useful to society?

If conscious computers ever arise (and I accept this is a big 'if'), this article has shown that I-Thou encounters with these new beings could take place, even from a theological viewpoint. Their potential role in pastoral care will be different to and shallower than humans can provide. Such a role would still have value—after all, even humans cannot provide perfect care, which can only be provided by our gracious God.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Andrew Proudfoot https://orcid.org/0000-0002-2186-3273