Care is in the Air
Artificial Medical Care Agents and Environments of Care

Darian Meacham and **Matthew Studley** [1]

Abstract. In this paper we examine the question: 'Can artificial medical care agents care?' We believe that the standard response to such a question is 'No'. This response is premised in the argument that care requires internal emotional states that such agents lack. Counter to this, we explore arguments which belie this conclusion. We argue that artificial medical care agents may create an environment of care through certain types of expressive movement irrespective of the existence of internal emotional states or intention. We address three possible objections to this argument, and argue that none of these objections is lethal to our hypothesis. Finally, we examine evidence that such human-robot interactions do not involve us in regarding artificial entities as independent agents, but we note that this may change as robots become more ubiquitous and we evolve social and cognitive structures to accept them in our daily lives.

1 Introduction

In this paper we examine possible challenges to the application of the terms 'care' or 'caring' to artificial Medical Care Agents. Simply put, our question is can artificial Medical Care Agents care? In the sections that follow we will reframe this question slightly such it is phrased in the following manner: can artificial medical care agents participate in the constitution of an environment of care? For reasons that we will explain in the next section, we think that inquiring into the conditions for the constitution of an environment of care is more appropriate than asking if an artificial medical care agent, and here we have in mind autonomous robotic systems, could actually care in the sense that we think of ourselves as caring for another living entity or comporting ourselves with an inner attitude of care toward a situation. In fact, we think that the relevant question surrounding the idea of care in general pertains to the constitution of environments of care, not inner attitudes, states, or intentions.

To begin with, we are operating with the presupposition that the standard and intuitively convincing response to this question is no: artifical medical care agents can neither themselves care in the relevant sense nor actively contribute to the constitution of an environment of care. As we take this to be a given or at least a widely accepted point in the literature we will not discuss it in detail. It is helpful however to point to arguments in the literature emblematic of this position and to briefly outline what we think to be the backbone of the position. Sparrow and Sparrow [9] argue:

> In most cases, when people feel happy, it will be because they (mistakenly) believe that the robot has properties which it does not These beliefs may be conscious beliefs, as in cases where

people insist that robots really are kind and do care about them, or are pleased to see them, etc. They might also involve unconscious, or preconscious, responses and reactions to the 'behaviour' of the robot (Brezeal, 2002, Ch. 2 as in Sparrow and Sparrow). It is these delusions that cause people to feel loved or cared for by robots and thus to experience the benefits of being cared for.

It is significant here that Sparrow and Sparrow do acknowledge that the patient may indeed have the experience of being cared for. What they object to is that this experience will be founded in a deception and an attribution of certain properties to the artificial agent that it does not really have - namely internal mental and emotional states. We will discuss the question of deception in greater detail below [2]. Here it suffices to sketch why we think that the standard account of such experiences of caring relationships considers them to be founded in a deception. We take this contention to be grounded in several premises: 1. artificial medical care agents are incapable of care because 'caring relations' and 'environments of care' imply not just functional behaviour, but emotional relations between agents. 2. Emotional relations require internal or conscious emotional states that can be ascribed (if not always self-ascribed) to an agent. 3. Artificial agents do not have the required conscious/emotional states and are thus incapable of emotion. Hence artificial medical care agents cannot care and their subsequent participation in the constitution of environments of care is suspect if not outright precluded.

Counter to this argument, we contend that caring relationships and most importantly environments of care are not dependent upon internal emotional, cognitive, or intentional states. Rather, we argue that an environment of care is a particular type of expressive and behavioural context or milieu characterised by certain types of reciprocal relations and attentiveness to the vulnerability and well-being of agents. Most significantly, we contend that the constitution of this type of environment is not dependent on conscious states or intentions of constituting agents, but rather on certain types of movement which are expressive of 'care' within that particular environmental and behavioural context. What particular movements are constitutive of an environment of care are most likely dependent on a universal or species-specific semiotics as well as local and situation specific conditions. It is not the intention of this paper to try to characterize these conditions.

Thus an environment of care is neither purely functional nor is it dependent on the internal states of the constituting or participating agents. The meaningful relationships necessary for the constitution of an environment of care are wholly contained or made manifest in

---

[1] University of the West of England, Bristol, email: Darian.Meacham@uwe.ac.uk

[2] We are grateful to an anonymous reviewer for stressing the importance of the deception argument.

external movement. Meaning is an emergent property of movements within an environment. Hence, we think it is possible that an artificial medical care agent could theoretically function as a constituting agent of an environment care provided it is capable of the required expressive and reciprocal movements. We will refer to this position as the 'Environmental Hypothesis'. We find that the Environmental Hypothesis is compatible with three of the four 'ethical elements of care' presented by Tronto [11]: 'attentiveness', 'competence' and 'responsiveness'. The element of 'responsibility' presents a greater challenge to the possibility of artificial Medical Care Agents as constituting agents of an environment of care. In the following section we will expand upon the theoretical underpinnings of the Environmental Hypothesis (section 2). We will then examine three possible objections to the Environmental Hypothesis: care is future oriented 3, care involves responsibility 4, and the Environmental Hypothesis condones deception in the manner objected to by Sparrow and Sparrow 5. In a final section we will look at some empirical studies that both support and undermine the Environmental Hypothesis.

2   Care is in the air

The Environmental Hypothesis stipulates that what matters in a caring relation is not the internal states of the agents participating in the relation, but rather a meaningful context. This context is generated by expressive movements: meaning is in the movement and in the salient differences that a movement makes to a meaning infused behavioural context. One way to think about this would be to use the phrase of Gregory Bateson that 'the elementary unit of information - is a difference which makes a difference' [1, pp.457-459] . What this means is that some differences, and this is to be taken in the sense of physical, material, or spatial differences in an environment, will enact a shift in the meaningful relations manifest in that environment. Others will not. Those that do convey information or meaning [3]. Our hypothesis is thereby an externalist one with regard to meaning and also to care.

   The artificial medical care agents example serves as an excellent thought experiment in general for the hypothesis that internal conscious states or emotions are not required for constitution of environments of care. Thus the negative aspect of the Environmental Hypothesis can be phrased in terms of a refutation of the claim that internal states are necessary for caring relations. Our refutation of this point can be sketched in the following manner. 1. internal emotional states are not accessible in human-human relations; i.e. I do not have access to the internal mental and emotional states of others (if such states do indeed exist internally) 2. Environments and relations of care are possible between humans who do not have access to each other's inner states. We will go further and state that environments and relations of care are possible between humans and some non-human animals, which even if they also have internal states do not self-ascribe them, i.e. do not have a self-reflexive relation to their internal states; 3. As we accept the existence of environments and relations of care between humans and non-human animals where there is no access to internal states, there is no reason to rule out the possibility of environments of care constituted by reciprocal relations between humans and artificial medical care agents. We can add a caveat to this third condition: provided the artificial agent can adequately mimic the movements necessary for the constitution of an environment of care.

---

[3] A difference can be made between the terms 'information' and 'meaning', but Bateson does not use the term 'information' in a limited sense of neg-entropy.

This caveat raises the question of deception which we will address below, but more fundamentally we think that the caveat is not potentially misleading in its use of the term mimic. It is misleading because the addition of the caveat and especially the introduction of the term 'mimic' could lead to an overly Cartesian or an 'analogical inferentialist' interpretation of our position. In the first instance external sounds (language) or gestures would be understood as the external signs of internal states (thought). Theories like this maintain that a thinking subject has 'privileged access' to her or his own mental states and this justifies the self-ascription of beliefs, sensations, desires, etc. This justification is completely lacking in our experience of others. Our experience of the mental states of others is mediated by external signs - the movement and sounds made by the body. On the basis of these external signs we make an inferential judgement pertaining to the inner states of the other subject. The judgement is grounded in an analogy between the correspondence of our own internal and external states and those of the other. What is particularly important for our purposes here is that the meaning-forming processes that are important to the caring relation are internal and private to the subject. The argument by analogical inference has been strongly critiqued, perhaps best by Max Scheler, but the idea that internal mental states are what matters still seems to carry considerable authority, as the widespread concern with deception illustrates.

In such a case an artificial care agent might provide all the necessary signs of having the requisite mental states. These states might even be sufficient for establishing the environment or relations of care, but they would remain at the level of mimicry. They would not really reflect the thoughts or feelings of the artificial medical care agent, because it would not have any. This does raise a possible issue of deception. The cared-for human subject could falsely attribute internal states to an artificial medical care agent that was literally just 'going through the motions'. As this situation occurs in human to human contexts where we do not rule out the possibility of caring environments, it is not obvious how serious an issue this is when it concerns an artificial medical care agent.

More to the point of this section, our position denies that this is how meaning-formation works in intersubjective contexts. The crux of our position is that the relevant meaning structures are formed in the space between moving subjects. The upshot of this is potentially quite significant. It is not just that our experience of meaningful relations with others does not proceed on the basis of an inferential judgement pertaining to the internal states of others, it is rather that internal states may not play a primary role in the formation of the relevant relations. This point is illustrated well by the French philosopher Maurice Merleau-Ponty:

> Imagine that I am in the presence of someone who, for one reason or another, is extremely annoyed with me. My interlocutor gets angry and I notice that he is expressing his anger by speaking aggressively, by gesticulating and shouting. But where is this anger? People will say that it is in the mind of my interlocutor. What this means is not entirely clear. For I could not imagine the malice and cruelty which I discern in my opponent's looks separated from his gestures, speech and body. None of this takes place in some otherworldly realm, in some shrine located beyond the body of the angry man. It really is here, in this room and in this part of the room, that the anger breaks forth. It is in the space between him and me that it unfolds. [8, p. 83]

This brings our position close to what is today termed 'enactivism'. Enactivism emphasizes how situated and embodied cognition organise and structure the mind. Evan Thompson [10] provides

a helpful sketch of this position: 'cognitive processes emerge from the nonlinear and circular causality of continuous sensorimotor interactions involving the brain, body and environment. The central metaphor for this approach is the mind as an embodied dynamic system in the world, rather than as a neural network in the head' (p. 11). To us this seems to underline the characteristics of the position we take here: meaning structures are formed in the world through the organism's engagement with its environment. This engagement most often occurs 'below' the level of active reflective consciousness. Our emphasis on whole organism engagement with a meaningful context that is its behavioral environment brings our argument also into close proximity with the holism of 'Gestalt biologists' like Kurt Goldstein [6]. Again what matters here is that the behavioral context is a meaningful one, modulated by salient differences or changes within the environment which should again be primarily understood and modulations of meaning. When we spoke above about expressive movement it meant precisely this: movements which enact a salient change in the environment understood in terms of meaningful context.

This leaves a question of what kind of expression is constitutive of an environment of care. Here we think that a minimal threshold condition for an environment of care entails: movements manifesting an attentiveness and appropriate response to perceived expressions or signs of vulnerability present and and movements eliciting a reasonable expectation for continued attentiveness and response. We see no reason in theory that an artificial medical care agent could not fulfill this threshold condition. Whether the relevant movement belong to a kind of universal or at least species specific biosemiotics, a more specific cultural context, or most likely both remains an open question and outside the scope of the current presentation.

In sections 3, 4, and 5 we will address several possible objections to our environmental hypothesis.

3 First Objection: Care and time, care is future oriented

A possible objection to the Environmental Hypothesis and its contention about artificial medical care agents is that the attitude of care is often taken to entail a specific attunement or orientation toward the future, and more specifically the future well-being of the agents whose expressive activity constitutes the environment of care. This futural understanding of care is likely in part derived from Heidegger's analyses of care as the basic temporal structure of human temporality. Human beings inherit a meaningful environment and comport themselves toward it in a manner that manifests an orientation toward the future, if not necessarily explicit planning for the future or even thinking about it. In the applied context of medical care agents, this attitude separates care from maintenance (understood as a kind of engineering work), which is focused on repair of specific functional relations between the organism and its environment. The objection would follow: As an artificial medical care agent presumably does not share the structure of temporal attunement that could be ascribed to humans, it is not capable of manifesting the behaviour requisite for the constitution of an environment of care.

This objection bears a similarity to the initial position against artificial medical care agents that we sketched out in the beginning. It assumes the necessity of an inner state, in this instance for an inner temporal orientation of conscious processes, for the manifestation of a certain type of external behaviour. Our response to this objection thus takes the same form as the general argument sketched in the previous section. We contend that the expressive behaviour manifest in the environment of care bears the necessary hallmarks of futural

concern. These may be best described in terms of expectation of future attention to vulnerability and reciprocation. It is necessary that an artificial medical care agent exhibit a certain canon of reasonable predictability in its behaviour - as we would expect from a biological organism of sufficient complexity - to be an active co-constitutor of an environment of care. Thus the question of the future orientation of care does not in our account present any special obstacle to the Environmental Hypothesis.

4 Second Objection: The Problem of Responsibility. A robust concept of responsibility requires both temporal consciousness and susceptibility to sanction.

The question of responsibility presents a more serious objection. This objection depends on the manifest behavioural expression of obligation or responsibility being a required constituent aspect of an environment of care. One reason this becomes a problem (there may be others) is that it seems feasible that a necessary condition of responsibility is a receptibility to sanction in the case of an breach of responsibility or obligation. Receptivity to sanction however seems to require an existential concern (conscious or not) for the entity's own continued existence. This kind of concern may well be limited to living systems.

This problem could be countered in the fashion of the Environmental Hypothesis in general. Concern for continued existence as a kind of internal attitude is not a necessary condition for the manifestation of responsibility. It is rather a question of expressive behaviour which need not bear any correlation with inner states, and does not in fact require inner states at all. This response does not address the further objection that while susceptibility to sanction may be a necessary condition of responsibility, it is not clear what type of behaviour manifests this susceptibility. Nor does it seem likely that whatever it was that would constitute the manifestation of responsibility in the above scenario would fit the criteria of what we normally mean when we use the term responsibility. It is debateable whether consistent behaviour that could be classed as responsible without the corresponding internal state of sensing obligation or appropriate reasons for acting would fit the bill of what most people mean when using the term responsible.

Accepting that responsibility does indeed pose an obstacle to the environmental hypothesis another easier response presents itself: simply refute the requirement of responsibility as an ethical element of environments of care. The ethical work done by the concepts of responsibility and/or obligation is already done by the establishment of a canon of reasonable predictability in the expressive behaviour of the agents reciprocally involved in the constitution of an environment of care. In other words, reasonable expectation of the continuation of the caring relationship is sufficient.

5 Third Objection: Deception

We have to some extent addressed the issue of deception in section two, but a few more comments are worth making here. What we think to be the more radical end of our argument is that there is no deception in the potential environment of care constituted by an artificial medical care agent and a human. The emotions and relations of care are real, even if the artificial medical care agent is not aware of the presence of the meaning-structures that it is playing an active role in constituting. It is probably not necessary that people

being cared for adopt our externalist thesis in order to accept the relations of care as being 'real' at least in the phenomenological sense of being real experiences, although of course it would help. The Environmental Hypothesis is probably closer in proximity to everyday experience than its opponents. Humans develop attachments and experience emotions such as empathy in all sorts of situations where it is clear that the other agent does not have the appropriate states with which to empathise. Human emotion shows itself to be both robust and promiscuous.

It is also important to note that human carers may be deceptive in their emotional displays. In his review of 'Emotion Work', Zapf [14] notes that nurses engage in both 'emotion work' in which they control their emotional response, and 'sympathy work' in which they attempt to change a patient's emotions in a desired direction in order to facilitate the primary task in which they are engaged, as, for example, a nurse using a soothing voice to calm a frightened child. It is recognised [4] that the degree to which workers are forced to suppress their real emotions in order to perform their job is a contributor to work stress and burnout. Since many medical professionals find themselves on the periphery of extremely traumatic events, they are recognised to suffer high rates of 'compassion fatigue' [13], and may respond to seeing others' pain by suppressing their own feelings, or by experiencing helplessness and anger. It is clear then, that environments of care constituted by human-human relations may of necessity be characterised by emotional pretence.

6   Concluding thoughts on the environmental model of care

So then, artificial agents can 'care' if we accept that the behavioural interface which betokens inner states is all that we can know of others, whether human or machine. Whether the way in which we engage with others is based on our internal simulation, our possession of a theory of mind, or whether it has some other mechanism as reviewed by [5], the authors are unaware of any physical mechanism that would give us direct experience of our partners' inner states. Moreover, we have maintained that the behavioural interface is generative of the meaning-structures which provide the content of inner states: behaviour and specifically movement are primary in generating the meaning required for an environment of care. The upshot of this is that we see no reason to preclude that artificial medical care agents could play an active role in the generation and maintenance of an environment of care.

Is there a dichotomy between this philosophical position and our experience? As the behaviours of machines become more accurate in their mimicry of the semantics of human interactions, will we accept their veracity in the same way we accept that our human partners are emotionally invested in our well-being, or will we always 'know' the falsity of the former as much as we 'know' the truth of the latter? There is some evidence from neural studies that while our emotional responses to artificial emotion might be similar albeit muted [3], interactions with artificial partners do not cause the stimulation of those areas of the brain which it has been suggested are responsible for the inference of mental states of fellow humans, e.g. the medial prefrontal cortex and right temporoparietal junction [2].

This implies that it might be impossible for us to believe that artificial medical care agents care for us, or take responsibility for our well-being; alternately such belief might rely on a better semblance of human behaviour, or on changing attitudes to such agents in society. Studies, e.g. [12] have shown that children who interact with 'relational artefacts' ascribe these devices to an intermediate class of being between the living and non-living, and others [7] have shown strong evidence that robot pets 'also embodied attributes of living animals, such as having mental states'. Perhaps we may believe that artificial medical care agents really care when today's children become tomorrow's elders?

REFERENCES

[1] Gregory Bateson, *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*, University of Chicago Press, 1972.

[2] Thierry Chaminade, Delphine Rosset, David Da Fonseca, Bruno Nazarian, Ewald Lutcher, Gordon Cheng, and Christine Deruelle, 'How do we think machines think? an fmri study of alleged competition with an artificial intelligence', *Frontiers in human neuroscience*, **6**, (2012).

[3] Thierry Chaminade, Massimiliano Zecca, Sarah-Jayne Blakemore, Atsuo Takanishi, Chris D Frith, Silvestro Micera, Paolo Dario, Giacomo Rizzolatti, Vittorio Gallese, and Maria Alessandra Umiltà, 'Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures', *PloS one*, **5**(7), e11577, (2010).

[4] Cynthia L. Cordes and Thomas W. Dougherty, 'A review and integration of research on job burnout', *Academy of Management Review*, **18**, 621–656, (1997).

[5] Shaun Gallagher and Daniel Hutto, 'Understanding others through primary interaction and narrative practice', *The shared mind: Perspectives on intersubjectivity*, 17–38, (2008).

[6] Kurt Goldstein, *The organism: A holistic approach to biology derived from pathological data in man*, Cambridge: MIT Press, 1995.

[7] Gail F Melson, Peter H Kahn Jr, Alan Beck, and Batya Friedman, 'Robotic pets in human lives: Implications for the human–animal bond and for human relationships with personified technologies', *Journal of Social Issues*, **65**(3), 545–567, (2009).

[8] Maurice Merleau-Ponty, *The world of perception*, Routledge, 2004.

[9] Robert Sparrow and Linda Sparrow, 'In the hands of machines? the future of aged care', *Minds and Machines*, **16**(2), 141–161, (2006).

[10] Evan Thompson, *Mind in life: Biology, phenomenology, and the sciences of mind*, Harvard University Press, 2007.

[11] Joan Tronto, 'Feminist theory: a philosophical anthology', chapter An Ethic of Care, 251–263, Malden: Blackwell Publishing, (2005).

[12] Sherry Turkle, Cynthia Breazeal, Olivia Dasté, and Brian Scassellati, 'Encounters with kismet and cog: Children respond to relational artifacts', *Digital Media: Transformations in Human Communication*, 1–20, (2006).

[13] Elizabeth A Yoder, 'Compassion fatigue in nurses', *Applied Nursing Research*, **23**(4), 191–197, (2010).

[14] Dieter Zapf, 'Emotion work and psychological well-being: A review of the literature and some conceptual considerations', *Human Resource Management Review*, **12**(2), 237–268, (2002).