Friendship Between Human Beings and AI Robots?

Margaret S. Archer

Contents

Abstract

In this chapter the case for potential Robophilia is based upon the positive properties and powers deriving from humans and AI co-working together in synergy. Hence, Archer asks 'Can Human Beings and AI Robots be Friends?' The need to foreground social change for structure culture and agency is being stressed. Human enhancement speeded up with medical advances with artificial insertions in the body, transplants, and genetic modification. In consequence, the definition of 'being human' is carried further away from naturalism and human essentialism. With the growing capacities of AI robots the tables are turned and implicitly pose the question, 'so are they not persons too?' Robophobia dominates Robophilia, in popular imagination and academia. With AI capacities now including 'error-detection', 'self-elaboration of their pre-programming' and 'adaptation to their environment', they have the potential for *active collaboration* with humankind, in research, therapy and care. This would entail *synergy or co-working* between humans and AI beings.

Keywords

Robots · Artificial intelligence · Sociology · Consciousness · Emotions · Friendship

Introduction

Friendship is regarded as paradigmatic of human sociality (Donati and Archer 2015: 66) because it entails no implications about kinship, sexuality, ethnicity, nationality, language, residence, power, status, beliefs, etc., although each and every one of these could be imbricated in it. Clearly, not all human relations are of this kind, the extreme exception being slavery, including variations on the Hegelian 'Master'/'Slave' theme. Significantly, in antiquity the enslaved were regarded as non-human. Their (supposed) absence of

M. S. Archer (✉)
University of Warwick, Coventry, UK
e-mail: margaret.archer@warwick.ac.uk

a soul served to justify their subordination and perhaps its shadow lives on in general attitudes towards other species.

Moreover, as Aristotle maintained, 'friendship' is not a unitary concept even for humans alone; it could take three forms (based upon utility, pleasure and goodness). Thus, in Aristotelian philosophy, different humans would not accentuate the same relationship today when they referred to their 'friends' compared with those held by others to be 'friends' in different times and places. What then is generic to 'friendship' and why is it relevant to discussing our (potential) relationships with other intelligent beings/entities such as AI robots? The answer to the first part of this question is that 'friendship' generates emergent properties and powers, most importantly those of trust, reciprocity and shared orientation towards the end(s) of joint action. These are distinct from simple friendly behaviour towards, say, dogs, where even as a source of pleasure they can only metaphorically and anthropomorphically (Haraway 2003) be termed a 'companion species' because there is no emergent common good or shared orientation; these may be imputed by dog-owners but cannot be ascertained for the dog except in behaviouristic terms. The answer to the second part of the question is the subject of this paper and will doubtless be contentious.

Its aim is to break the deadlock between popular 'Robophobia' and commercialized 'Robophilia' (products intentionally marketed as humane 'Best Friends' or useful 'housekeepers' that control thermostats and draw the blinds, etc.). Most of the best-known arguments in this dispute rest upon exaggerating the binary divide between the human and the AI, as if this difference between organically based and silicon based entities formed an abyss dividing up the whole gamut of properties and powers pertaining to agents of any kind (Brockman 2015).

Conversely, I argued in the first papers in our Centre for Social Ontology series on *The Future of the Human* (Archer 2019) that acceptance of their shared 'personhood' can span this divide. The main propositions defended there and used here are the following[1]:-

1. 'Bodies' (not necessarily fully or partially human) furnish the necessary but not the sufficient conditions for personhood.
2. Personhood is dependent upon the subject possessing the First-Person Perspective (FPP). But this requires supplementing by *reflexivity* and *concerns* in order to define personal and social identities.
3. Both the FPP and Reflexivity require Concerns to provide traction in actuating subjects' courses of action and thus accounting for them.

4. Hence, personhood is not in principle confined to those with a human body and is compatible with Human Enhancement.

In my above conclusions, point (4) merits a particular comment. Human Enhancement is anything but new. Historically it goes back at least to Pastoralism, enhancing height and strength through improved nutrition, and merely becomes more bionic with prostheses, pacemakers, etc. This is why I do not find the relationship between contemporary humans and the latest AI robots to be usefully captured by the concept of 'hybridity', since *Homo sapiens* have been hybrids throughout recorded history because of their progressive enhancement. Instead, this paper accentuates *synergy* between the two 'kinds' as paving the way to friendship.

Responses to my list of conclusions from opponents can again be summarized briefly as denials that the three capacities I attribute to all normal human beings are ones that can be attributed to an AI entity.

1. The AI entity has no 'I' and therefore lacks the basis for a FPP.
2. Consequently, it lacks the capacity to be Reflexive since there is no self upon which the FPP could be bent back.
3. Similarly, it cannot have Concerns in the absence of an 'I' to whom they matter.

In what followed, I sought to challenge all three of these objections as far as AI entities are concerned. However, this was not by arguing, in some way, that the subsequent development (if any) of these highly sophisticated, preprogrammed machines tracks the development of human beings in the course of their 'maturation'. On the contrary, let me be clear that *I start from accepting and accentuating the differences between the human and the AI in the emergence of the powers constitutive of personhood.*

In the human child, the 'I' develops first, from a *sense of self,* or so I have argued, as a process of doing in the real world, which is not primarily discursive (language dependent) (Archer 2000). The sequence I described and attempted to justify was one of {'I → Me → We → You'}. The sequence appears different for an AI entity that might plausibly follow another developmental path, and which, if any, would be a matter of contingency. What it is contingent upon is held to be relational, namely it develops through the synergy between an AI entity and another or other intelligent beings, namely humans to date. In this process of emergence, the 'We' comes first and generates a reversal in the stages resulting in personhood, namely {'We' → 'Me' → 'I' → 'You'}. Amongst robots, it is specific to the AI and will not characterize those machines restricted to a limited repertoire of pre-programmed skills, sufficient for a routine production line. Conversely, the AIs under discussion are credited with at

---

[1]These conclusions share a great deal in common with Lynne Rudder Baker (2000, 2013). Compare with the above reference.

least four main skill sets that can be supplemented: language recognition and production; learning ability; reflexive error-correction; and a (fallible) capacity for Self-Adaptation relevant to the task to which they have been assigned. This paper is not concerned with 'passing as human' (the Turing test), nor with 'functional equivalences in behaviour', independent of any understanding (Searle's Chinese Room) (see Morgan 2019). Its aim is twofold: First, to make a start on debunking some of the main obstacles regularly advanced as prohibiting 'friendship' between AIs and humans (and perhaps amongst AI themselves, although this will not be explored), second, to venture *how* synergy (working together) can result ceteris paribus in the emergence of 'friendship' and its causal powers.

Overcoming the Obstacles?

Three main obstacles are regularly advanced as precluding 'friendship' with human beings and often reinforce 'robophobia'. All of these concern ineradicable deficits attributed to AI robots. Specifically, each systematically downplays one of the characteristics with which AIs have been endowed in this current thought experiment; abilities for *continuous learning* (until/unless shut-down); for *error-correction* and for *adaptation* of their initial skills set—and thus of themselves—during their task performance. The accentuation of AI deficits which follow shadows Colonialist portrayals of colonized peoples.

Normativity as a Barrier

Stated baldly, this is the assertion that an AI entity, as a bundle of micro-electronics and uploaded software, is fundamentally incapable of knowing the difference between right and wrong. Consequently, alarm bells sound about the ensuing dangers of their anormativity for humans, as prefigured in Asimov's normative laws of Robotics (1950). In other words, 'Robots are seen as (potential) moral agents, which may harm humans and therefore need a "morality"' (Cockelbergh 2010: 209). Usually, this need is met by a top-down process of building human safeguarding into pre-programmed designs or, less frequently, by AI robots being credited with the capacity to develop into moral machines from the bottom-up, through learning morality, as ventured by Wallach and Allen (2008).

Both protective responses confront similar difficulties. On the one hand, morality changes over time in societies (poaching a rabbit no longer warrants transportation) and the change can be fast (from French Revolutionary laws to the Code Napoléon), as well as coercive and non-consensual. Even Hans Kelsen (1992) had abandoned one *grundnorm* as

founding all instances of legal normativity by the end of his work. On the other hand, if the model of childhood learning replaces that of pre-programming, we largely now acknowledge that the socialization of our kids is not a simplistic process of 'internalization'; again it is a non-consensual process societally and the family itself often transmits 'mixed messages' today. Thus, complete failure may result. It simply cannot be concluded that 'we set the rules and bring it about that other agents conform to them. We enable them to adapt their behaviour to our rules even before they can understand social norms (Brandl and Esken 2017: 214).'

Equally pertinent, social normativity is not homogeneous in its form or in its force. Both of the latter are themselves subject to social change. Just as importantly, some of its transformations are more amenable to straightforward learning (for example, it took less than 5 min online to learn how to renew my passport) than others (such as 'What constitutes Domestic Violence in a given country?'), which requires interpretation and changeable judgements about appropriate classification of various behaviours (e.g. the new category of 'coercive control' now used in British family law).

If we break normativity down into recognizable categories—and what follows is for purposes of illustration, not the only useful manner of doing so—it should clarify that working upwards through the list of rules is to move from easy learning to the need for expert advice that itself will be challenged.

Etiquette is heterogeneous (changeable and regional), varying with factors such as the age, and social standing of participants *vis à vis* one another. Although transgression attracts only social sanctions, its applicability to AIs is dubious. Are handshakes appropriate? May they list their 'special requirements' as 'access to an electric socket and cable'? How should they address others and expect to be addressed? No human guide to correct behaviour[2] can assist AIs in learning 'good manners', despite their ability to read and recite any manual available, because many human conventional practices are disallowed by the AIs' (current) constitutions. Much the same goes for large tracts of *customs and conventions*.

More significantly, as I have argued elsewhere (Archer 2016) is the growth of *anormative bureaucratic regulation*—both public and private—that now predominates in the production of social co-ordination. The rule of law can no longer run fast enough to keep up with social morphogenesis in almost any social domain; novel forms of malfeasance outstrip counteractive legislation as recognized in countries such as Britain that tried to declare a halt on designating

---

[2]Although such guidelines have become more relaxed (Abrutyn and Carter 2014).

'new crimes'[3] in the new millennium. Instead, Regulators and regulations boom in every area, sometimes upheld by law but frequently not. In this reversal, something very damaging has happened to normativity itself and is directly relevant to demolishing the barrier that the absence of a capacity for it once constituted for delineating AI beings from humans.

This is the fact that obeying regulations does not rely upon their ethical endorsement; indeed the rules governing domestic refuse disposal or the sale of forked carrots in supermarkets may be regarded as idiotic by the general public—who were never consulted. Regulations are not morally persuasive but causally produce conformity through fines, endorsements and prohibitions. Thus, it is up to the subjects to perform their own cost-benefit analyses and determine whether the price of infringement, such as motorway speeding, is worth it to them on any given occasion. Taking the normativity out of an increasing range of activities progressively weakens the barrier placing AI beings outside the moral pale. Like today's humans, they do not have to feel guilt, shame, remorse or wrongdoing in breaching regulations. Thus, whether or not they are capable of such feelings is an argument that has become less and less relevant because the social context makes decreasing use of them.

Intensive social change also undercuts a frequent protective suggestion that in the interests of human health and safety 'our' values should be built into AI beings. But, if this is on the model of childhood learning or socialization, what (or, rather, whose values) are adopted? The alternative is to programme such values into the AI to-be, in an updated version of Asimov's three laws. Yet, supposing it was possible and desirable, it is not an answer to the normative barrier because these would be our values that we have introduced by fiat. Pre-programed values can never be theirs, not because they did not initiate them (however that might be), but rather because the same theorists hold that no AI entity can have emotional commitment to them for the simple reason that they are presumed incapable of emotion. However, anormative administrative regulation sidesteps this particular issue since it is independent of emotions through its reliance on calculative cost-benefit analysis.

Emotionality as a Barrier

I am not arguing that human and AI beings are isomorphic, let alone fundamentally the same substantively. Nor is that the case methodologically for studying the two, which is particularly relevant to emotionality. Dehaene (2014) and his team do a wonderful job in de-coding parts of brain activities, established by working back experimentally *from behaviour*

for human beings, including some of the brain damaged. However, in comparing the capacities of the 'wet' and the 'dry' for experiencing emotion, it seems more productive to start the other way around with the physical constitution of AIs and the affordances and limitations of their uploaded software.

In *Being Human* (Archer 2000), I differentiated between our relations with the 'natural', 'practical' and 'social' orders as the sources of very different types of emotions impinging upon three inescapable human concerns (respectively, our 'physical well-being', 'performative achievement' and 'self-worth')—given our *organic* constitution, the way the world is made and the ineluctability of their interaction. I still endorse this approach when dealing with the human domain; each order of natural reality has a different 'import' for our species and generates the emergence of different clusters of emotions acting back upon them.

My definition of emotions as 'commentaries upon our concerns' in the three orders of natural reality is about matters we human beings cannot help but care about (to some extent)—imports to which we cannot be totally indifferent, given our constitution. Hence, whilst humans may experience 'terror' in anticipation of being plunged into Arctic water, constitutionally they cannot 'fear' themselves rusting. Conversely, objectively dangerous imports for AIs would be quite different: for example, rusting, extended power-cuts or metal fatigue.

However, because I am maintaining that AI robots can detect that they confront real dangers (just as all our electronic devices signal their need for re-charging), this does not in itself justify attributing emotions to robots. But, such a move is unnecessary. This need not be the case because on my account it is *concerns* that are pivotal and whilst emotion may increase attention and provide extra 'shoving power', it is not indispensable and it can be misleading. Thus, I disagree with those who maintain that 'emotions matter. They are the core of human experience, shape our lives in the profoundest ways and help us decide what is worthy of our attention' (McStay 2018: 1). Even when confined to humans, our emotionality surely cannot be a guide to 'worth'.

Nevertheless, many do maintain that an 'emotional commentary' is an essential part of all our *concerns* and, by extension, of all forms of *caring*. Yet, to be realistic, we *care enough* about a variety of mundane concerns to do something about them (e.g. checking the warranty when buying appliances, keeping spare light bulbs, stocking some 'long life' products in the pantry, etc.), all without any emotionality. Such concerns are sufficient to make (some of us) care 'enough' to do such things, without any 'feelings' at all. Conversely, being emotionally moved by a photo of a dead baby on an Aegean beach (a real example), said on social media to have 'moved many to tears', was clearly not enough to promote their active caring for asylum seekers. It

---

[3]Over 4000 entered the statute books in last two decades of the twentieth century (Cabinet Office Paper 2013).

seems that the branding of some movies as inconsequential 'tear jerkers' is not far from the mark. In sum, *I hold to my view that emotions are neither a necessary nor a sufficient condition for caring, whilst accepting that their addition may reinforce fidelity to our concerns.*

My position does not amount to Cognitivism, which might at first glance seem appropriate to AI beings. However, what 'cognitivists' maintain is that 'emotions are very real and very intense, but they still issue from cognitive interpretations imposed on external reality, rather than directly from reality itself' (Ortony et al. 1988: 4). Above, I have argued against this reduction of ontology to epistemology in human relations with the three orders of natural reality. Now I am asking why we should agree that AI beings are deficient precisely if they *do not experience* 'very real and very intense emotions' in natural reality? Were they merely to scrutinize a situation and to conclude cognitively that undesirable φ was highly likely to transpire, that would sometimes improve their 'prospects' of averting φ in comparison with dramatic displays of affectivity that can foster stampedes in human crowds confronting fires in confined spaces.

Some might think that a mid-way position is provided by Charles Taylor's statement that we speak of emotions as essentially involving a sense of our situation, claiming that they 'are affective modes of awareness of situation … *We are not just stating that we experience a certain feeling in this situation*' (Taylor 1985: 48, my ital.). I agree, but stress that the ontology of the situation remains indispensable. Moreover, there are two controversial words in that quotation which effectively divide psychologists of the emotions: one is 'affect' (affectivity) and the other 'feelings' and both are very relevant to our present discussion.

'Feelings' as the mainstay of my opponents are a slippery concept because some are held worth consideration and others not, some to be concealed, others displayed. But this is largely a matter of social convention. As individuals, humans vary enormously in how readily they reveal their suffering and which sufferings, but their acceptability has also varied historically. Why has 'having a stiff upper lip' come to be seen as derogatory would be interesting for semantic archaeology. Equally, why does encouraging others to grieve overtly, to 'tell their stories' or to 'let it all come out', seem to be the creed of the counsellors today? This overt variety shows emotionality not to be a universal and essential component of human responses to similar circumstances if they are so socially malleable. The rejoinder could obviously be that we can never know what people suffer (or exult about) in silence, but then if we cannot know it, neither can we study it. A last resort would be to hand this over to the 'therapeutic couch', but in which of those warring psychiatric practitioners should we place our trust?

If those holding that the presence and absence of feelings comes down to the 'phenomenal feel' of *qualia,* which will forever divide the wet and the dry, it seems a weak case for two reasons. First, it varies experientially within the human species, or calling someone 'tone deaf' or 'blind to natural beauty' would not have been coined. Second, if one of us is motivated by, say, injustice or unfairness why are those supposedly accompanied by particular *qualia*? Judges are expected to rule on cases in the light of evidence made available in trials, not to share their 'phenomenal feel' for it or the parties involved. Neither did John Rawls argue that decisions made 'behind the veil' entailed such phenomena. If these continue to be regarded as a 'barrier' by some, then sharing the same *qualia* will also continue to separate all beings from one another, regardless of their physical constitution, if these can ever be objectively determined.[4]

The Ultimate Barrier: Consciousness

This is generally presented as the 'great divide' that those in silico can never cross. Indeed, it is a resilient version of the old dispute between Comte and Mill about the premise of a 'split-consciousness' built into the concept of introspection. Mill's riposte was to jettison the simultaneous element by inserting a brief time interval between the original thought and *inspection* of it. Consequently, our self-awareness became an unobjectionable exercise of memory. I will not repeat the lengthy argument I advanced (Archer 2003: 53–92), by selectively drawing upon the American Pragmatists, to buttress my contention that 'introspection', on the observational model (*spect intro*), should be replaced by the 'inner' or 'internal conversation'. But how is this relevant to AI beings?

Instead of re-invoking 'introspection' I simply rely on two software abilities: to speak and to listen for *securing their self-consciousness*. Every day we humans employ language to pose questions: internally to ourselves, externally to other people and also of our outside environment. A common exemplar, not universal and answerable in various non-linguistic ways, is the first question likely to arise each day for most adults upon waking—'What time is it?' We are both questioners and respondents and this means that all normal people are both SPEAKERS and LISTENERS, to themselves. This is what the American pragmatists—James, Pierce, and Mead—called the 'inner conversation' and I have explored this subjective mental activity in my trilogy of books on human 'reflexivity' (Archer 2003, 2007, 2012).[5]

---

[4]In other words, I agree with Dehaene, that the 'concept of qualia, pure mental experience detached from any information-processing role, will be viewed as a peculiar idea of the prescientific era' (Dehaene 2014: 262).

[5]Reflexivity is defined 'as the regular exercise of the mental ability, shared by all normal people, to consider themselves in relation to their social contexts and vice versa' (Archer 2007: 4).

Now, I want to venture that given the AIs I have postulated are programmed to be/become proficient language users, then why should it be queried that they too function as both speakers and listeners? This cannot be seriously contested. But if that is so, why can't they be credited with internal reflexivity? The barrier on the part of those who deny the capacity for 'thought' to AI robots is a simplistic definitional denial of their ability to think because computers are held to be incapable of consciousness, let alone self-consciousness. Yet, if we examine the basic constituents of the 'internal conversation' what is there in the activities of speaking, listening and responding (internally, but reflexive deliberations can be shared externally) that would put any AI robot permanently beyond the pale of reflexivity? Certainly there are practical obstacles, the most powerful being that in current computers each software application works in a separate memory space between which exchanges are precluded, meaning that programmes have no general means of exchanging their specialized knowledge (Dehaene 2014: 259f). But, such limitations as these result from the programme designers rather than being intrinsic to AI robots and are not (usually) applicable to speaking and listening per se.

When we do think of questioning and answering in general conversations, all conversational exchanges are alike in one crucial respect, namely they involve turn-taking. Therefore, I am arguing that when we talk to ourselves the same rule maintains and it does so by our *alternating between being subject and object in the dialogical turn-taking process*, which is rendered possible because of the necessary time gap—however small—that separates question from answer. Some may query how this is possible given that any data or notion I produce (as subject) is identical to that I simultaneously hear (as object). Yet, it would be meaningless to entertain an alternation between two identical things. Instead, following the insight of William James, in expressing a response we review the words in which to articulate it, welcoming the felicitous ones and rejecting those less so (James 1890). Thus, our answers often do not come pre-clothed in the verbal formulations that clearly express them to the best of our ability or to our satisfaction—either externally or internally. We are fallible as well as sub-optimal in this respect, sometimes even saying what we do not mean—to ourselves and to others. But we are capable of reformulation before venturing a revised response. And we may do this several times over (see Fig. 1).

This extension of James consists only in allowing the subject to question his/her/its own object over time. Such a process will be familiar to any writer and is the bread and butter of literary historians. To redeploy James' notion of

Fig. 1 Datum and verbal formulation in the internal conversation. Source: Adapted from Archer (2003: 99)
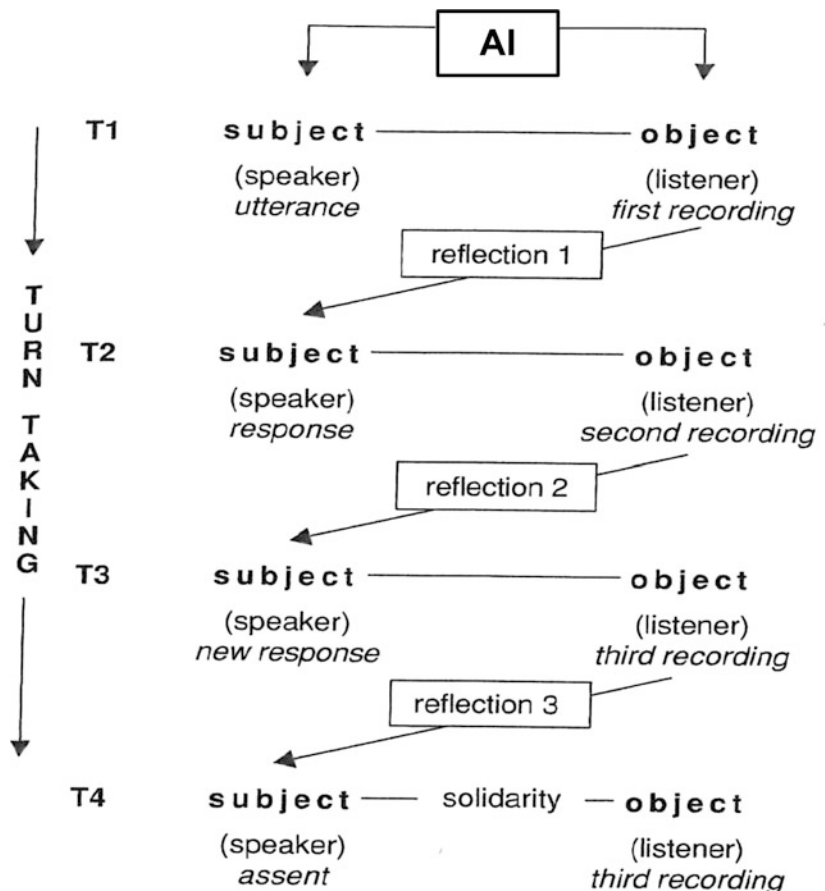
'welcoming' certain verbal formulations, discarding others and seeking for more adequate words would all be held illegitimate by opponents because as mental activities they entail 'thought'. Even if an AI is pictured summoning up a Thesaurus, the objection would be that there is no mechanism that can match semantics for their appropriateness in a *novel* research context, where 'common usage' cannot be the Court of Appeal. Of course, the AI may do as we do and sometimes resort to poetry. But then he is guilty of thoughtful 'creativity' and to concede this would allow the barrier to sag. I can see no way in which an AI's exercise of judgement can be freed from conceding that he is also exercising thought—and thus 'consciousness and self-consciousness', as Harry Frankfurt puts it:

> '(B)eing conscious in the everyday sense does (unlike unconsciousness) entail reflexivity. It necessarily involves a secondary awareness of a primary response. An instance of exclusively primary and unreflexive consciousness would not be an instance of what we primarily think of as consciousness at all. For what would it be like to be conscious of something without being aware of this consciousness? It would mean having an experience with no awareness whatever of its occurrence. This would be, precisely a case of unconscious experience. It appears, then, that being conscious is identical with being self-conscious. Consciousness *is* self-consciousness.' (Frankfurt 1988: 161–162)

Were these considered to be convincing if over-condensed objections to the 'barriers', there remains a final but crucial point to note which is central to the rest of the paper. All of the obstacles briefly reviewed have depended upon 'robotic individualism', since they have all taken the form of 'No AI robot can ᒐ, Ω, or Θ'. That is not the case that I am seeking to advance here. Instead, this concerns the dyad and a particular form of dyadic interaction, between an AI robot and a human co-worker or consociate. Since I presume we would agree that there are properties and powers pertaining to the dyad that cannot belong the individual, whether of organic or silicon constitution, then these need to be introduced and be accorded due weight.

The Emergence of Friendship and Its Emergents

This section begins with the dyad but does not end there. As Simmel maintained, there are both distinctions and connections between what are conventionally differentiated at different societal levels: the micro-, meso- and macro-strata. This is because different emergent factors and combinations come into play at different levels just as distinct contingencies are confronted there. In Critical Realism this represents a *stratified social ontology*, which is the antithesis of the 'flat ontology' endorsed, for instance, in Latour's 'Actor network' approach. Not only are there different properties that emerge at the three levels but there is also upward and downward

causation between them. Restriction of space will make this a very sketchy introduction.

At the Micro-, Meso- and Macro-Levels

Let us begin schematically with the relationship between a single human researcher and the AI being awarded to him under his/her grant. The aim is obviously that the co-ordinated action of the two is intended to advance a particular research programme, or overcome a specific problem. For two main reasons there is nothing deterministic about this relationship. First, there will be psychological differences between human researchers in their expectations about the future relationship that arise from a whole host of factors, some but not all of these would be the same for what any particular researcher might expect if instead he/she had been awarded a post-doctoral co-worker. Second, there are certain structural and cultural constraints, formal and informal, attending such co-working (rights and duties) that will influence the likelihood of friendship developing, without proscribing it.

Ultimately, co-ordination is the generic name of the game for those setting up these 'mixed dyads'. And there is nothing about co-ordinated actions per se that is conducive to or hostile towards friendship; even 'collaboration' can involve a strict division of labour, which never segues into 'collegiality'. Indeed, co-ordination between humans may refer to 'one off' events such as two motorists coming from opposite directions who collaborate to move a fallen tree that is blocking both drivers but which neither can do alone (Tuomela 2010: 47). But the two drivers are unlikely candidates for subsequent friendship, given they will probably not meet again.

Equally, some senior researchers may define the relationship in advance as one in which they have acquired a Research Assistant, regardless of whether the subordinate is human or robotic. Black neatly characterizes this Command and Control hierarchy in terms general enough to apply to those research Bosses who hold themselves 'to be the only commander and controller, and to be potentially effective in commanding and controlling. [He] is assumed to be unilateral in [his] approach (he tells, others do), based on simple cause and effect relations, and envisaging a linear progression from policy formation through to implementation' (Black 2001: 106). None of that characterizes the *synergy* within the dyad that is the necessary but not sufficient condition for the development of friendship.

The Boss could indeed confine his robot Assistant to tasks of compilation and computation for which the latter is pre-equipped and their relationship may remain one of CAC even though a master does not possess all the skills of his servant, as with most Lords of the Manor and their gardeners. But it would be a short-sighted aristocrat who

never talked or listened to his gardener and one who believed he could envisage a linear progression of the unfinished garden to his ideal and thus issue unilateral demands for the implementation at each stage. That would make his gardener dispensable as other than a digger and planter; any discussion of their synergy becomes irrelevant.

In a previous paper (2019) I sketched the emergence of synergy through a thought experiment about a much more open-minded Boss (Homer) and his robotic assistant (Ali) who became a co-worker. In the first instance, Homer played to what he knew to be Ali's strengths, to make fast computations on the Big Data available which were too extensive for Homer himself to review. He could then venture his hypotheses about the incidence of a lethal Tumour X in humans that extended way beyond his own reading or clinical practice. Ali's data overview produced some confirmatory cases but also cast doubts on it from others. Because they did talk about it, Homer shared his new suspicion with Ali that qualitative data was required in addition to quantitative to explain not just the incidence but the progression of Tumour X. In the second instance, after lengthy discussion, the two recognized what was needed but also that Ali was not equipped with the necessary software. He wanted to help his boss to solve the problem set (and on which his 'employment' depended), even though *qua* robot he had no concern about the solution benefitting humanity. He surveyed the qualitative data analysis programs and went on to read up evaluation reports about them through consulting e-journals. All of this is completely unremarkable, except for one thing. Ali had taken the responsibility (and accountability) for making and executing this extension of the research program. He had acted as a genuine part of this 'we'.

In the third instance, as the research program progressed and the relational goods generated grew, the design also became increasingly elaborate. Although funding increased, so did Homer's wish-list for new technical tools and so did his questions to Ali: 'Could you do this or that?' This led Ali to a stock-taking exercise, what could simply be added to his pre-programed repertoire of skills, what could be adapted by re-writing that software and what was it beyond him to deliver? In short, Ali discovered the affordances of its bionic body and its resistances to adaptation for some of the novel tasks mooted.[6] Because of their 'we-ness'—for Homer and Ali are not or no longer in a command and control relationship—Ali makes the adaptations to his pre-programming that are possible and commensurate with further progress on their future research design. Sometimes he makes mistakes (as Turing anticipated), but he is familiar with error-correction.

But something has happened to Ali during their long collaboration. He has learned a great deal and *he is aware of this*. He is (1) not merely carrying and processing information (as does a GPS and as Searle's man in the Chinese Room did). Ali is doing things that enable new knowledge to be generated, things that Homer cannot do but needs doing. As an AI, he is not knowledgeable (2) in the purely *metaphorical sense* that a statistical Table might be said to 'know' (e.g. the extent of drug trading in different countries). In any case, that is about the publication and not the generation of knowledge. Finally, (3) Ali's awareness is quite different from the sense in which a thermostat might (again metaphorically) be called aware of temperature change when it kicks-in. That is a pre-programmed mechanical response to a change in the external environment. Conversely, Ali's awareness is in response to his own self-induced changes in his internal constitution and resultant adaptive capacities (for which he has records). But enough of this film script that will never hit the big screen.

Some readers may have no insuperable objections to the above but still feel that it has not got to the core of friendship. Undoubtedly, it has not in Aristotelian terms where 'With those in perfect friendship, the friend is valued for his/her own goodness and constitutes 'another self''.[7] However, basing my case upon synergy, this is neither about differences alone nor similarities alone between co-workers, but about *diversity* (Archer 2013)—that is some shared qualities together with some differences—that incentivizes collaboration and makes it possible. What a focus upon such syncretism avoids is insistence on the 'similarity' criterion, which immediately evokes 'speciesism' as an insuperable barrier to friendship in this context (Dawson 2012: 9).

'Holism', Individualism' and 'Relationality' are three distinct ways out of this cul-de-sac. 'Holism' means abandoning human essentialism by focusing sociologically instead upon the characteristics and capacities that we acquire through our first-language and induction into a culture, which McDowell (1998) terms our 'second nature'. What this does is to transfer the characteristics previously attributed to the human 'species' to our social communities. However, elsewhere, I have criticized this as the 'Myth of Cultural Integration' (Archer 1985: 36), disputing that any 'culture' is fully shared and homogeneous, if only because none are free from contradictions, inconsistencies and aporia. If that is the case, then cultural essentialism falls too—to a multiculturalism where 'goodness' varies over time, place, selection and interpretation.[8]

---

[6]This is not unlike the child discovering that he is socially advantaged or disadvantaged compared with his classmates. In both cases, these very different kinds of subjects consider themselves as objects in relation to others of their kind.

[7]In the Nicomachean Ethics, Book 1, this represents the apex of friendship: 'Only the friendship of those who are good, and similar in their goodness, is perfect. For these people each alike wish good for the other *qua* good, and they are good in themselves'.

[8]The same can be said of Tuomela's (2010), reliance upon agents drawing on a 'shared ethos' in order to act as a 'we' or in the 'we-mode', and that is considerably less demanding than friendship.

'Individualism' holds that what is admirable in a friend originates from his/her uniqueness and therefore cannot be essential properties of human beings. *How* someone came to be of exemplary goodness might be due to some unique concatenation of circumstances, but we do not admire *what made them what they are—though we may learn from it—but rather celebrate their virtuosity, which can have aggregate consequences alone for co-working.*

'Relationality' shifts conceptualization away from holistic species' attributes and individual uniqueness alike, and moves the relationship to centre-stage. It replaces discussion of essences by causal considerations of outcomes. In fact, friendship is defined on the causal criterion, that is, on the basis of the relational goods and evils it generates. It is thus simultaneously freed from essentialism, exceptionalism and speciesism. It displaces reliance upon 'joint Commitment', the mainstay of Margaret Gilbert's account and which also has to be central to John Searle's since he places identical thoughts in the two heads of the dyad,[9] difficult as this is to endorse (Pettit and Schweikard 2006: 31–32). Instead, I focus upon 'joint action' and its causal consequences, where the commitments of each party maybe quite different. (The research Boss in the story could have been motivated by the death of his mother from Tumour X, which cannot be the case for his robotic co-worker).

However, although 'joint action' is held to be a necessary condition, it is not sufficient for engendering friendship. Paul Sheehy gives the nice example of four prisoners who are rowing a boat to escape. They share the belief that 'I am escaping' and that it entails 'we are escaping', since they are literally in the same boat. Hence their joint action may or may not be successful. But if this is the emergent effect of their cooperation, it does not necessarily entail or engender friendship (Sheehy 2002). It is only if relational goods are generated in synergy, from which both parties are beneficiaries and further benefits are deemed likely, given continued collaboration, that the first stepping stones towards friendship are laid. Like an orchestra or a successful football team, neither party can 'take away their share' without cutting the generative mechanism producing the relational goods they value, albeit for different reasons.

The first paving stone is the emergence of trust in their co-action. Although the commitments of the two may be substantively different, they cannot be incompatible or zero-sum, if each is a beneficiary and motivated by this to continue co-working. In short, they have a *common-orientation* to the project (of whatever kind) continuing and developing. This is all the more essential the more different in kind are their separate contributions. Such reciprocal trust, reinforced by successful practice over time, is what unites the skating duo,

the trapeze 'flyer' and 'catcher' as well as co-authors and research dyads. In all such instances, that trust needs to become sufficiently resilient to withstand occasional accidents, false starts and blank periods without progress. This argument is highly critical of the conceptualization and usage of the concept of 'trust' within the social sciences. Frequently, it is treated as a simple predicate rather than a relational outcome, which requires consistent reinforcement. But predicates cannot be plucked out of thin air when needed; they have their own morphogenesis, morphostasis and morphonecrosis (Al-Amoudi and Latsis 2015).

There are many ways of defining 'friends' and distinguishing between the characteristics and consequences of friendship. Here I treat 'dimensionality' as differentiating between 'thin', that is, 'one-dimensional' relations versus the 'thicker' multi-dimensional relationships as constitutive of friendship. In everyday human life it is common for people to refer to 'my golfing friend', 'my travelling companion' or 'my workmate', meaning their relations with another are restricted to these domains. Such 'thin' friendships are vulnerable to breakdown, partly because it takes only one row, bad trip or disappointed expectations for their fragility to break, and partly because such dyadic partners are quite readily replaceable. On the other hand, 'thickness' is tantamount to the resilience of the friendship, with its various facets compensating for deficiencies in any particular one.

Some commentators may conclude at this point that a human and a robot might possibly succeed in developing a friendly but one-dimensional working relationship but that social structural and cultural constraints would preclude this morphing into 'thick' friendship. In a manner evocative of 'apartheid', it could be agreed that many expressions of this incipient relationship do seem closed to them; e.g. they cannot go out for a drink or a meal together. Similarly, for the time being, there are many social embargos on their sharing leisure activities (e.g. AI robots may be ineligible for Golf Club Membership). The implication could seem to be that this dyad is confined to strictly cognitive activities in the extension of their friendship.

Even so, that leaves them plenty of options (film, television, the media, gaming, gambling, literature, music and art) and the cultural context of society is increasingly privileging some of these pursuits. If one day Homer wants to watch the news, he might be surprised that Ali as spectator becomes vocal about a short film-clip on football's match of the day. Certainly, his comments are cognitively confined to working out team disposition in relation to goal scoring— but so are those of many pub-pundits. In watching movies, he represents the critic that sci-fi film makers have never had (to date), commenting on the portrayal of those of his own kind and effectively challenging the dominant 'robophobia' presented. As far as 'gambling' is concerned, he is the ideal co-participant and readily computes how far the odds are

---

[9] For a detailed critique of the 'Plural Subject' analytical philosophers, see Donati and Archer (2015: Ch. 2).

stacked in favour of the virtual 'house'. Indeed, were the human/AI dyad to go 'rogue' and devote some of their efforts to the generation of 'Relational Evils' through online gambling—or working fruit machines for that matter—this could be a formula for breaking the bank at Monte Carlo.

When we come to micro-meso level linkages, the weakness of the sociological imagination, which has Cinderella status in the foregoing, dries up almost completely. Take the Research Centre which figured prominently in the discussion of relationships at the micro-level. Sociologically it is too spare and bare. For example, I know of no Research Centre that is exempt from the pressures of either its benefactors or of the University, or of the educational system of which it forms a part, if not all three. In turn, the *relational networks* are much more complex; in reality students, administrators, commercial agents, journalists, social media, funding agencies, educational policy makers, etc. would impinge upon this dyad—were the account more than fiction—with their different causal powers and vested interests. These would not be contingent occurrences but part and parcel of current research life and indispensable to accounting for real-life outcomes in it.

There is a cluster of congruent reasons for research showcasing the micro-stratum. To be fair, our world of competitive research funding structurally reinforces this focus on AI individuals. Thus, firstly, it is hard in the barrage of literature to find any empirical research *on AI to AI relations themselves*. Given many of them have the software for speech recognition and use, do they never speak to one another, about what, or are they disabled from doing so? Why, given this could be beneficial to the research? For example, Aurélie Clodic and her colleagues describe a table-test involving an AI and a human building a tower from wooden bricks together and the AI sometimes drops a brick (Clodic et al. 2017). Possibly the robot has something constructive to say about its occasional clumsiness that would improve the joint action. There is nothing to lose, beyond the cost of installing the communicative software if it is lacking.

Secondly, the same point is reinforced by the wealth of material concerning the (usually) positive views about AIs in the roles of taking care of the elderly, communicating with Autistic children, assisting in hospital operating theatres, etc.[10] What would these workers say about the greatest difficulties encountered versus what they accomplish easily in a small group? Instead, the focus remains fixedly on 'individual client satisfaction', but maybe this clientele could be better satisfied given such input was available. After all, the aim is not exclusively for-profit; it can also be to assuage or offset loneliness, so there does seem to be a role for

fostering friendship more effectively here. The AI assistants would not become bored by the repetition of stories by the lone, frail elderly, may supply missing names and places that escape and distress their clients, might show them photo montages of, let's say, the subject's children or of significant houses, neighbourhoods, landscapes in their biographies. These actions, undertaken by a team of robot-carers could enhance the quality of life for those in care.

Thirdly, when collective AI is considered, especially in military contexts, what they are modelled on is but 'swarm behaviour' amongst birds, insects and animals of lesser intelligence. The US Department of Defense is explicit that such swarms 'are not pre-programmed synchronized individuals, they are a collective organism, sharing one distributed brain for decision-making and adapting to each other like swarms in nature' (Plummer 2017). They appear to resemble a weapon of mass destruction rather than candidates for citizenship.

With such military exceptions, the *predominant focus* is upon the AIs as *individuals* rather than as any kind of interactive collective, it is about *singular* relations with particular human persons rather than shared problems in dealing with them and about *aggregate* and never (to my knowledge) about emergent consequences.

Yet Emmanuel Lazega (2013, 2014, 2015, 2016) has shown empirically in a variety of contexts (including science laboratories, law courts, and a Catholic Diocese) how, in particular, the quest for advice from peers and superiors is the stuff from which networks are built geographically between localities and can reach politically upwards for representation and voice, thus consolidating collectivities at the meso-level. This is precisely what almost exclusive concentration upon micro-level research precludes. It seems ironic that commercial ICT enterprises are resolutely global in the extension of their algorithms in pursuit of worldwide profit and of their surveillance data for globalized power, whereas the most intelligent forms of robots are confined to micro-level personal services. Of course, such confinement constrains them to supplementing the human hierarchy rather than substituting for it. Is this the reason why, besides the costs involved, it is the self-protectiveness of their human designers which discourages AI inter-communication, which deters the formation of collective associations and which defines their contributions in aggregate terms? This conclusion seems hard to resist given that Dehaene (2014: 259ff) ventures that the development of a Global Workspace accessible to all would increase the affordances available to each and every AI, potentially facilitating an increase in innovation and enabling the self-adaptive capacities of artificial intelligence to have freer social rein. This, of course, might more likely incite structural repression for human self-defence than to be met in friendship.

[10]For examples, see Maccarini (2019) and LaGrandeur (2017). By 2011, the total number of robots in use worldwide was 18.2 million: only 1.2 million of those were industrial, and the rest—17 million—were service robots', p. 98.

In sum, the consequences of the missing meso-level can broadly be presented as threefold as far as society is concerned:

1. It operates as a severe curb upon innovation and therefore on the novelty and new variety upon which societal morphogenesis depends.
2. It works to preclude the emergence of AI 'social movements', whatever (unknown) form these might take.
3. It serves to buttress human domination and thus is hostile to friendship between these two categories of persons.

At first glance, it may now appear surprising that *at the Macro-level* there are significant signs of the recognition of AI robots as 'electronic persons', who would acquire rights and obligations under a draft EU resolution of 2017. Yet, this only appears unexpected and contradictory if we wrongly interpret it as a similar stage in the succession of social movements following on from those first promoting workers' unionization, then female enfranchisement, anti-discrimination legislation, and LGBT rights, to date. However, 'robot rights' are not the aim of an AI millennial vanguard pressing for legal and political recognition in society. In fact, their closest historical precursor was the accordance of 'corporate personhood' to firms and companies, enabling them to take part as plaintiffs or respondents in legal cases (Lawson 2015). Its main objective was to free individual company owners, executives and shareholders from financial obligations following legal judgements of culpability. Exactly the same objective appears here; to exculpate human designers and marketers from legal and financial responsibility, whilst defending humans against demands for compensation given damages they have caused to a robot.

Significantly, the European Parliament (February 2017) adopted a draft resolution for civil law rules on robotics and AI, pertinent to issues of liability and ethics of robots. What is worth noting is that Robotics were at first treated as an undifferentiated collectivity and a compulsory insurance scheme was mooted for 'Robot users'—in general. And this despite the fact that a year earlier a draft European Parliament motion (May 31, 2016) had noted that the (AIs) 'growing intelligence, pervasiveness and autonomy require rethinking everything from taxation to legal liability' (CNBC 2016). This motion called upon the Commission to consider 'that at least the most sophisticated autonomous robots could be established as having the status of electronic persons with specific rights and obligations'.

The Code of Ethical Conduct (2017) proposed for Robotics Engineers endorsed four ethical principles: (1) beneficence (robots should act in the best interests of humans); (2) non-maleficence (robots should not harm humans); (3) autonomy (human interaction with robots should be voluntary; (4) justice (the benefits of robotics

should be distributed fairly) (Mańko 2017). These defensive principles are redolent of Asimov's 'laws' of the mid-1940s (and were reproduced in the Code's text); they have nothing in common with today's movements or social protests. The report's author, Mady Delvaux (Luxemburg) encapsulated its main motive: 'to ensure that robots are and will remain in the service of humans, we urgently need to create a robust European legal framework' (Hern 2017).

The paucity of 'robot rights' reflects the priority given not only to the generic defence of humans (spurred by the approach of self-driving cars) but specifically, as one international lawyer put it, to the issue of intellectual property rights,[11] which I earlier maintained was the heart of the matter in late modernity's new form of relational contestation (Archer 2015). No more dramatic illustration is needed than this discussion of draft legislation to (supposedly) acknowledging 'robotic personhood' than a legal commentary which took *the sale of 'electronic persons' for granted*! (my italics) (Hern 2017).

Conclusion

In the context of the present paper, I wish to finish with a quotation from Clause 28 from the Draft European Report (2015).[12] What it reveals is a fundamental incapacity to deal with human/AI synergy, to treat synergy as non-emergent but rather susceptible of being broken down into the 'contributions' of the two parties, and having no necessary relationship to relational goods and evils or to sociality.[13]

In other words, 'friendship' between humans and AIs, far from being the anchorage of 'trust', 'reciprocity' and helping to engender 'innovative goods' is remarkable for its absence.

---

[11] As Ashley Morgan, working for the Osborne Clarke practice, summarized matters: 'If I create a robot, and that robot creates something that could be patented, should I own that patent or should the robot? If I sell the robot, should the intellectual property it has developed go with it?' (Hern 2017).

[12] European Parliament, *Committee on Legal Affairs,* 31.5.2015. [The Report] considers that, in principle, once the ultimately responsible parties have been identified, their liability would be proportionate to the actual level of instructions given to the robot and of its autonomy, so that the greater a robot's learning capability or autonomy is, the lower other parties' responsibility should be, and the longer a robot's 'education' has lasted, the greater the responsibility of its 'teacher' should be; note, in particular, that skills resulting from 'education' given to a robot should be not confused with skills depending strictly on its self-learning abilities when seeking to identify the person to whom the robot's harmful behaviour is actually due.

[13] This differs hugely from the Nobel Prize's procedure of awarding more than one researcher, without specification of their 'discrete' contributions. Of course, the conferring of an award for collective innovation is very different from the partitioning of legal liability among those involved.

However, as usual, looking for absences reveals whole tracts of activities as well as rights that have been exiled from such reports. There is space only to itemize some of those that are missing, partly at least because 'friendship' is not considered as a generative mechanism, that is as an indispensable building block in the attribution of rights to AI beings:-

– Eligibility to stand for election
– To hold a Passport and use it
– The right to open a Bank Account, to possess property (including intellectual property, patents, copywrite, etc.) and to benefit from inheritance or to designate inheritors.
– Rights to compete with humans for appointments, promotion, etc.
– Legal protection against discrimination
– Rights covering Marriage, Intermarriage and Adoption
– Appropriate changes in sports categories
– And, perhaps the most contentious of all, no blanket embargo upon AIs becoming 'received' into various Church and Faith communities.

References

Abrutyn, S., & Carter, M. J. (2014). The decline in shared collective conscience as found in the shifting norms and values of Etiquette manuals. *Journal for the Theory of Social Behaviour, 43*(2), 352. https://doi.org/10.1111/jtsb.12071.
Al-Amoudi, I., & Latsis, J. (2015). Death contested: Morphonecrosis and conflicts. In M. S. Archer (Ed.), *Generative mechanisms transforming the social order*. Cham: Springer.
Archer, M. S. (1985). The myth of cultural integration. *The British Journal of Sociology, 36*(3), 333–353. https://doi.org/10.2307/590456.
Archer, M. S. (2000). *Being human*. Cambridge: Cambridge UP.
Archer, M. S. (2003). *Structure, agency and the internal conversation* (pp. 53–92). Cambridge: Cambridge UP.
Archer, M. S. (2007). *Making our way through the world*. Cambridge: Cambridge UP.
Archer, M. S. (2012). *The reflexive imperative*. Cambridge: Cambridge UP.
Archer, M. S. (2013). Morphogenic society: Self-government and self-organization as misleading metaphors. In M. S. Archer (Ed.), *Social morphogenesis* (pp. 145–154). Dordrecht: Springer.
Archer, M. S. (2015). *Generative mechanisms transforming the social order*. Cham: Springer.
Archer, M. S. (2016). Anormative social regulation: The attempt to cope with social morphogenesis. In M. S. Archer (Ed.), *Morphogenesis and the crisis of normativity*. Cham: Springer.
Archer, M. S. (2019). Bodies, persons and human enhancement: Why do these distinctions matter? In I. Al-Amoudi & J. Morgan (Eds.), *Realist responses to post-human society: Ex machina*. Abingdon: Routledge.
Asimov, I. (1950). *I, Robot*. New York: Genome Press.
Black, J. (2001). Decentering regulation: Understanding the role of regulation and self-regulation in a post-regulatory world. *Current Legal Problems, 54*(1), 103–147.
Brandl, J. L., & Esken, F. (2017). The problem of understanding social norms and what it would take for robots to solveit. In R. Hakli &

J. Seibt (Eds.), *Sociality and normativity for robots* (pp. 201–215). Cham: Springer.
Brockman, J. (Ed.). (2015). *What to think about machines that think*. New York: HarperCollins.
Cabinet Office Paper. (2013). *When laws become too complex: A review into the causes of complex legislation*. Available via Gov.uk. Retrieved February 27, 2020, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/187015/GoodLaw_report_8April_AP.pdf
Clodic, A., Pacherie, E., & Chatila, R. (2017). Key elements for human-robot joint-action. In R. Hakli & J. Seibt (Eds.), *Sociality and normativity for robots* (pp. 159–177). Cham: Springer International Publishing.
CNBC. (2016). *Robots could become 'electronic persons' with rights, obligations under draft EU plan*. Accessible via CNBC. Retrieved September 4, 2019, from https://www.cnbc.com/2016/06/21/robots-could-become-electronic-persons-with-rights-obligations-under-draft-eu-plan.html
Cockelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology, 12*, 209–221. https://doi.org/10.1007/s10676-010-9235-5.
Dawson, R. (2012). Is Aristotle right about friendship? *Praxis, 3*(2), 1–16.
Dehaene, S. (2014). *Consciousness and the brain*. New York: Penguin.
Donati, P., & Archer, M. S. (2015). *The relational subject*. Cambridge: Cambridge UP.
Frankfurt, H. (1988). Identification and wholeheartedness. In H. Frankfurt (Ed.), *The importance of what we care about* (pp. 159–176). Cambridge: Cambridge UP.
Haraway, D. (2003). *The companion species: Dogs, people and significant otherness*. Boulder: Paradigm.
Hern, A. (2017). *Give robots 'personhood' status, EU committee argues*. Available via The Guardian. Retrieved February 27, 2020, from https://www.theguardian.com/technology/2017/jan/12/give-robots-personhood-status-eu-committee-argues
James, W. (1890). *Principles of psychology*. London: Macmillan.
Kelsen, H. (1992). *Introduction to the problems of legal theory*. Oxford: Clarendon.
LaGrandeur, K. (2017). Emotion, artificial intelligence, and ethics. In J. Romportal, E. Zackova, & J. Kelemen (Eds.), *Beyond artificial intelligence* (pp. 97–109). Cham: Springer International Publishing.
Lawson, T. (2015). The modern corporation: The site of a mechanism (of global Social change) that is out-of-control. In M. S. Archer (Ed.), *Generative mechanisms transforming the social order* (pp. 205–230). Cham: Springer.
Lazega, E. (2013). Network analysis and morphogenesis: A neo-structural exploration and illustration. In M. S. Archer (Ed.), *Social morphogenesis Series* (pp. 167–185). Cham: Springer.
Lazega, E. (2014). *'Morphogenesis unbound' from dynamics of multi-level networks*. Cham: Springer.
Lazega, E. (2015). *Body captors and network profiles*. Cham: Springer.
Lazega, E. (2016). *Joint anormative regulation from status inconsistency*. Cham: Springer.
Maccarini, A. M. (2019). Post-human sociality: Morphing experience and emergent forms. In I. Al-Amoudi & E. Lazega (Eds.), *Post-human institutions and organizations*. Abingdon: Routledge.
Mańko, R. (2017). *New Laws needed for robots. What Europe is doing for its citizens?* Available via European Parliamentary Research Blog. https://epthinktank.eu/2017/08/14/new-laws-needed-for-robots-what-is-europe-doing-for-its-citizens/
McDowell, J. (1998). *Two sorts of naturalism: Mind, value and reality*. Cambridge, MA: Harvard UP.
McStay, A. (2018). *Emotional AI: The rise of empathetic media*. Los Angeles: Sage.

Morgan, J. (2019). Yesterday's tomorrow today: Turing, Searle and the contested significance of artificial intelligence. In I. Al-Amoudi & J. Morgan (Eds.), *Realist responses to post-human society: Ex Machina*. Abingdon: Routledge.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge: Cambridge UP.

Pettit, P., & Schweikard, D. (2006). Joint actions and group agents. *Philosophy of the Social Sciences, 36*(1), 18–39. https://doi.org/10.1177/0048393105284169.

Plummer, L. (2017). *Swarm of micro-drones with 'hive mind' dropped from fighter jets in US military test*. Available via Mirror. https://www.mirror.co.uk/tech/drone-swarm-hive-mind-launched-9597295

Rudder Baker, L. (2000). *Persons and bodies*. Cambridge: Cambridge UP.

Rudder Baker, L. (2013). *Naturalism and the first-person perspective*. Oxford: Oxford UP.

Sheehy, P. (2002). On plural subject theory. *Journal of Social Philosophy, 33*(3), 377–394. https://doi.org/10.1111/0047-2786.00148.

Taylor, C. (1985). *Human agency and language*. Cambridge: Cambridge UP.

Tuomela, R. (2010). *The philosophy of sociality*. Oxford: Oxford UP.

Wallach, W., & Allen, C. (2008). *Moral machines: Teaching robots right from wrong*. Oxford: Oxford University Press.