



It's Friendship, Jim, but Not as We Know It: A Degrees-of-Friendship View of Human–Robot Friendships

Helen Ryland¹

Received: 30 September 2020 / Accepted: 24 March 2021 / Published online: 3 April 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract

This article argues in defence of human–robot friendship. I begin by outlining the standard Aristotelian view of friendship, according to which there are certain necessary conditions which *x* must meet in order to ‘be a friend’. I explain how the current literature typically uses this Aristotelian view to *object* to human–robot friendships on theoretical and ethical grounds. Theoretically, a robot cannot be our friend because it cannot meet the requisite necessary conditions for friendship. Ethically, human–robot friendships are wrong because they are deceptive (the robot does not *actually* meet the conditions for being a friend), and could also make it more likely that we will favour ‘perfect’ robots, and disrespect, exploit, or exclude other human beings. To argue against the above position, I begin by outlining and assessing current attempts to reject the theoretical argument—that we cannot befriend robots. I argue that the current attempts are problematic, and do little to support the claim that we can be friends with robots *now* (rather than in some future time). I then use the standard Aristotelian view as a touchstone to develop a new *degrees-of-friendship* view. On my view, it is theoretically possible for humans to have some degree of friendship with social robots *now*. I explain how my view avoids ethical concerns about human–robot friendships being deceptive, and/or leading to the disrespect, exploitation, or exclusion of other human beings.

Keywords Social robots · Human–robot friendship · Robophilosophy · Friendship

1 Introduction

At the time of writing, the world is in the grips of a global pandemic. One outcome of this is social isolation. Many people (particularly those from vulnerable groups) are shielding or in lockdown, and consequently have limited real-world access to human friends and family. Jecker (2020, ‘abstract’) argues that one solution to

✉ Helen Ryland
hryland.philosophy@gmail.com

¹ The Open University, Milton Keynes, UK

this issue would be “...to design robots to function as companions and friends for socially isolated and lonely older people during pandemic emergencies and in aging societies more generally.” On Jecker’s view, the concept of ‘robot friendship’ is increasingly important; it could be a vital lifeline for many. As Jecker (2020, ‘counterfeit companions’) states, human–robot friendships are positive in this way because “...they can protect our health and enrich our lives.”

Jecker’s above argument is timely and inventive and, if actualised, could have potentially revolutionary practical applications. From a philosophical perspective, however, Jecker’s view is vulnerable to both theoretical and ethical objections (Sect. 3). Theoretically, it is argued that we cannot be friends with robots because robots cannot meet the necessary conditions for ‘being a friend’. Ethically, human–robot friendships are supposedly wrong because they are deceptive, and also make it more likely that we will disrespect, exploit, or exclude other human beings.

This article aims to defend the idea of human–robot friendships from these objections. I argue that many of the supposedly necessary conditions for friendship (Sect. 3) are actually optional components of friendship. I then outline a degrees-of-friendship view, according to which we can have stronger or weaker friendships with *x*, depending on how many of the components our friendship with *x* meets. On this view, it is theoretically possible to have some degree of friendship with a robot (Sect. 4). In Sect. 5, I argue that, understood in this way, human–robot friendships are not ethically problematic. They are not deceptive, and they do not necessarily make it more likely that we will disrespect, exploit, or exclude other human beings.

2 Which Robots Could We Be Friends with?

...dating a robot with a human form may not be quite as absurd or weird as a romantic attachment to an iPhone or smart fridge, but it’s arguably on a continuum with them (Tasioulas, 2019, p. 68).

In the above, John Tasioulas highlights an important aspect of human–robot relationships: the *type* of robot that we are talking about *matters*. Jecker’s (2020) proposal—that we create robot friends for the socially isolated—has considerably more impact and appeal if, like her, we understand robot friends in terms of fairly advanced social robots, rather than smart fridges or robotic vacuums. For the remainder of this article, I will thus be focusing on friendships with social robots. Examples include SoftBank Robotics’ ‘Pepper’, and Blue Frog Robotics’ ‘Buddy’; both of whom can purportedly recognise and respond to human emotions and perform social interactions.¹

As Tasioulas (2019, p. 51) explains, UNESCO have recently defined robots “... as artificial beings with four characteristics: mobility...interactivity...communication...and autonomy, in the sense of an ability to ‘think’ for themselves, and make their own decisions to act upon the environment...” Within this general definition

¹ For further details, see Blue Frog Robotics (n.d.), and SoftBank Robotics (n.d.).

of robots, there are different types of robots (ranging from robotic toys and vacuum cleaners, to military robots and robotic surgeons). Social robots are a special class of robot that are designed to have interpersonal relations with human beings. Examples include sex robots, robot teachers, and robot priests.²

Jecker (2020, 'the proposal') explains that, in order to fulfil this social role, social robots have a number of features that make interpersonal connections more likely. First, most social robots have a recognisable human or animal form, as this is supposed to encourage positive responses from human users.³ Second, social robots are designed to perform and respond to basic social interactions (e.g. they can hold conversations, smile, make eye contact, etc.).⁴ Finally, Jecker suggests that, in future developments, social robots will be able to safely touch others, and sense and respond to human touch.

The above emphasises the relative complexity of social robots. A social robot is more complex than a robotic vacuum. Whilst the robotic vacuum is mobile and interactive, and potentially has limited communication and autonomy, it does not look like a human or animal, and cannot perform or respond to basic social interactions.

Further, different social robots can have different levels of complexity. A social robot marketed as a toy may only perform pre-programmed basic social interactions (and these may be fairly rigid and incomplete). In contrast, a social robot used as a teacher or shop assistant may have a larger repertoire and be able to master more complicated, spontaneous social interactions. These different levels of complexity will become important in Sect. 4, where I will suggest that we can have degrees of human–robot friendship. For now, though, let us address the opposition by considering why many think that human–robot friendships are theoretically and ethically problematic.

² For discussions of sex robots, see Danaher and McArthur (2017), Frank and Nyholm (2017), and Nyholm and Frank (2017). For discussions of robot teachers, see Sharkey (2015, 2016). For discussions of robot priests, see Young (2019).

³ This is not to deny the impact of the so-called 'Uncanny Valley' effect, whereby robots that are *too* humanlike are viewed as 'creepy' or 'unnerving' (and so are viewed negatively). The balance between human/animal-like and robot-like thus seems to be important. For discussions of the uncanny valley, see Lay (2015), and Mathur and Reichling (2016).

⁴ Jecker (2020: 'introduction') explains this more fully as follows: "They display increasingly sophisticated emotional intelligence; interact in ways that seem lifelike, such as recognising voices, faces and emotions; interpret speech and gestures; respond appropriately to complex verbal and nonverbal cues; make eye contact; speak conversationally; and adapt to people's needs by learning from feedback, rewards, and criticisms." I would suggest that the extent to which robots are able to do these things *successfully* and *naturally* will depend on how complex and advanced the robot is. In many cases, it is likely that most *current* robots will only be able to perform rudimentary versions of the above, though their abilities are likely to increase as technology advances.

3 The Opposition: We Cannot Be Friends with a Robot

Most philosophical discussions of friendships frame their discussions in terms of a broadly Aristotelian view of friendship. As both Danaher (2019, ‘Robots Can be Our Aristotelian Friends’) and Vallor (2012, p. 188) explain, there are different (sometimes divergent) ways in which the Aristotelian concept of friendship can be interpreted, but the general consensus is that some combination of the following are *necessary* conditions for friendship⁵:

Reciprocity: Both parties must acknowledge the existence of the friendship. There must also be mutual good will between the friends (i.e. it is not a friendship if one party wants to hurt or exploit the other party) (Danaher, 2019; Fröding & Peterson, 2012; Tistelgren, 2018; Vallor, 2012).

Empathy: It must be possible for friends to empathise with one another. An upshot of this is that friends must be relatively similar to one another (in terms of vulnerabilities, needs, etc.). This is because one must be able to know, understand, and relate to someone in order to empathise with them (Vallor, 2012; possibly Coeckelbergh, 2010).⁶

Self-knowledge: The friends must help each other to grow, learn about themselves, and become their ‘best selves’. This is typically supposed to occur through mirroring—we see ourselves through our friend’s eyes and learn about our strengths, weaknesses, etc. (Fröding & Peterson, 2012; Tistelgren, 2018; Vallor, 2012).

Shared Life: Friends must engage in shared activities, across a number of different domains. Friends ought to enjoy doing these activities, at least in part, because the activity allows them to spend time with one another (Danaher, 2019; Fröding & Peterson, 2012; Munn, 2012; Tistelgren, 2018; Vallor, 2012).

Associative Duties: We are typically supposed to have special duties towards friends. We ought to have reasons to treat them well (and better than we would treat strangers), we ought to believe them, see the best in them, etc. (Arrell, 2014; Kawall, 2013; Munn, 2012; Tistelgren, 2018).

Affection and Well-Wishing: Friends ought to care for one another. It seems sensible to suppose that this can range from basic good will to actively wanting the best for a friend and striving to help them achieve this (discussed in Munn, 2012).

Love and Admiration, Based on Virtue: Fröding and Peterson (2012, p. 204) explain this condition as “...the admiration and love the friends feel for each other is based on the virtues they recognise in the other”. In other words, we love our friends because they are kind, charitable, and brave; not because they are cruel, uncharitable, or cowardly (discussed in Fröding & Peterson, 2012).⁷

⁵ Some of the below conditions appear to overlap, e.g. ‘reciprocity’ and ‘affection and well-wishing’ seem to be very similar. In the literature, however, they are sometimes presented as distinct conditions, and for this reason I consider them separately here.

⁶ I say ‘possibly’ because Coeckelbergh (2010) is focusing on companionship, rather than friendship, though there does seem to be some connection between these two concepts.

⁷ ‘Love and admiration, based on virtue’ is also a necessary condition of friendship on Cicero’s account of friendship. For a clear, accessible discussion of Cicero’s views, and how they relate to the possibility of human–robot friendship, see Nyholm (2020: Sect. 5.6).

Honesty: In order to know someone as a friend, we must (largely) know the truth about them. Friends should not deceive one another, or try to gloss over the bad aspects of themselves (discussed in Danaher, 2019).

Equality: There should be basic social equality between friends. There should not be an imbalanced power relation (discussed in Danaher, 2019).

Many (but not all) philosophical accounts of human–robot friendships use the above to argue against the existence and/or morality of human–robot friendships.⁸ Typically, two broad arguments are presented:

- (1) *The theoretical argument.* We cannot be friends with robots because robots do not meet the above necessary conditions for ‘being a friend’. For example, robots cannot reciprocate love, they are not empathetic, they cannot help us to gain self-knowledge, etc. (see Tistelgren, 2018. This argument is also outlined, and ultimately rejected, in Danaher, 2019).
- (2) *The ethical argument.* Human–robot friendships are *wrong* because:
 - (i) They are deceptive (and this deception is inherently wrong). The relationships are deceptive because the robot is either deceiving the human into believing that it can meet the above necessary conditions (i.e. that it can be empathetic, etc.), or the human is deceiving themselves into believing this.
 - (ii) They make it more likely that we will disrespect, exploit, or exclude other human beings. This is because we might end up prioritising ‘perfect’ robots over ‘imperfect’ humans (the ethical arguments are outlined in Coeckelbergh, 2010; Danaher, 2019; Jecker, 2020; Turkle, 2011. They are also discussed in Tasioulas, 2019; Tistelgren, 2018).

I concede that the above arguments do seem to stand if we adopt the above broadly Aristotelian view of friendship, where there are strict, necessary conditions that a robot must meet in order to be our friend. In most (if not all) cases, robots will not *currently* meet all of the necessary conditions for friendship. In the next two sections, however, I will argue that we need not take this Aristotelian view, and that there are more flexible, contemporary ways of understanding friendship which make human–robot friendships both theoretically and ethically acceptable.

⁸ Notable exceptions to this are Danaher (2019) and Marti (2010). Both will be discussed further in Sect. 4. Gunkel (2018, p.174) also accepts that we are ethically obligated to “hold open the possibility that [robots] might become Other”, and so could potentially be our friends.

4 Rejecting the Theoretical Argument

The above theoretical argument depends on the idea that robots cannot be our friends because they do not meet the relevant necessary conditions for ‘being a friend’. This section will outline three ways in which we could attempt to reject this theoretical argument. The first two suggestions would allow us to retain most (if not all) of the standard Aristotelian view (above). I will outline these suggestions, and explain why they are problematic. The third suggestion is my degrees-of-friendship view. Whilst it does require us to adopt a non-standard view of friendship, I will argue that my view avoids the problems with existing attempts to reject the theoretical argument, and so should be favoured.

4.1 Suggestion 1: Remove Some of the Necessary Conditions for Friendship

The theoretical argument only works if the conditions outlined in Sect. 3 actually are *necessary* conditions for friendship. One way in which we could reject the theoretical argument would thus be to demonstrate that some of the conditions are *not necessary* for friendship. Provided that a robot could meet the remaining necessary conditions, we could theoretically be friends with a robot.

The underlying impetus for this view is nicely presented by Cocking and Kennett (2000). They argue that the ‘self-knowledge’ and ‘love and admiration, based on virtue’ conditions are overly moralistic, outdated, and do not reflect our actual experiences of friendship.⁹ They state: “our friends are not morally or constitutively moral exemplars who thus inspire us to moral growth and improvement” (Cocking & Kennett, 2000, p. 296). In other words, most of us do not see our friends as ideally virtuous, nor do we aim to use our friendships to become better or more moral (though this may be an incidental consequence of friendship). On this view, it is thus possible to have friendships that *do not* involve self-knowledge, or love and admiration, based on virtue.¹⁰

⁹ Cocking and Kennett’s (2000) arguments are concerned with friendship in general, rather than human–robot friendships. Their arguments can, however, naturally be extended to human–robot friendships.

¹⁰ An opponent could have two main objections to this view. First, it may be argued that whilst the ideal, moralistic view of friendship is hard to meet, it might be possible for people to come *close* to meeting it. So, perhaps our friends are those who are not perfectly moral but who nevertheless encourage us to be better people. Nyholm (2020: Sect. 5.6) presents a view of this sort in his discussion of Cicero. I am sympathetic to this view, but believe that imperfect virtue/morality is still too demanding. In suggestion 3 (below), I defend a degrees-of-friendship view, according to which it is possible to have degrees of genuine friendship without the friendship having *any* moralising element.

A second objection could be that the contemporary conceptions of friendship being discussed here are less desirable than the older, Aristotelian ideals. In other words, contemporary ‘friendship’ could be what Nyholm (2020: Sect. 5.7) describes as “some fairly watered-down and less interesting sense of friendship.” Again, I am sympathetic to this view, and believe that it is probably the orthodox view. However, it is not a view to which I subscribe. In suggestion 3 (below) I will argue that it is possible to have degrees-of-friendship. On my view, even a low degree of friendship is a genuine friendship that can be both desirable and theoretically and ethically interesting. This is true both of human–human friendships, and human–robot friendships.

A similar argument could be made about the 'equality' condition. To modern sensibilities, this condition can seem outdated, and also has problematic undertones of discrimination (i.e. that you cannot befriend someone who is *not* your social equal). For many of us, lived experience suggests that friendship *can* flourish, despite inequalities between friends. For example, friends can differ in terms of gender, sexuality, race, social class, education level, and a number of other characteristics that could cause them to have different (and unequal) placings on some constructed social hierarchy. A good visual depiction of this is Hughes' 1985 film *The Breakfast Club*, which famously examines the growing relationships between "...a brain, an athlete, a basket case, a princess, and a criminal" forced to spend time together during detention. The film depicts how the five characters become friends, through mutual good will and shared activities, *in spite* of their differences and the social inequalities between them.

There are two main difficulties with this first attempt to reject the theoretical argument. First, the above only potentially removes the 'self-knowledge', 'love and admiration, based on virtue', and 'equality' conditions. Questions can still remain as to whether robots can meet the remaining necessary conditions for friendship. For example, can we have associative duties towards a robot (and they to us)? Can robots be empathetic? And so on.

The second difficulty concerns the shift between saying that a condition is outdated or against modern sensibilities, and saying that it is not necessary for friendship. This is a very large argumentative leap to take, and one that could easily be defeated. For example, an opponent could argue that the self-knowledge condition is only outdated because of its moralistic undertones. It may be possible to revise the condition so that it has a more modern, secular flavour (e.g. by suggesting that friendships reflect who we are, rather than who we ought to strive to become). On this view, the outdated conditions are still necessary for friendship; they simply require modification.¹¹

¹¹ An opponent could also have a more general worry: if we are changing the conditions for some *x* counting as a friend (either by modifying or removing some conditions), then why should we still insist that something that meets these conditions is a friend? Are they not a 'friend' in some non-standard sense, so perhaps they are 'friend-like', or a 'quasi-friend' instead?

There are two general responses to this worry. First, the above objection seems to implicitly accept that the original Aristotelian conditions for friendship are the *correct* conditions, and that any deviation from the standard Aristotelian view can only lead to pseudo-friendship or some approximation of friendship. However, whether or not the Aristotelian conditions are fit-for-purpose is precisely the question being discussed in suggestions 1 and 2. In order to give these suggestions a fair hearing, we should at least entertain the thought that the standard Aristotelian view might *not* present the correct conditions for contemporary friendship. Instead, some modified neo-Aristotelian view might be the *correct* view.

Second, it is worth emphasising that, even if we do remove or modify some of the conditions for friendship, these new conditions would not *only* apply to robots. They would also apply to human friends. For example, as Footnote 9 emphasises, Cocking and Kennett's argument for modification concern friendships in general. On a modified view of friendship, a friendly robot need not be a friend in some other sense than a human can be our friend. Both the robot and the human meet (at least some of) the same (updated or modified) conditions for friendship. In Sect. 4, I will develop this idea by considering how robots and humans can meet the same conditions for friendship, but could meet them in different ways and to different degrees. On my view, degrees-of-friendship refers to degrees of *genuine* friendship, rather than some form of quasi-friendship.

4.2 Suggestion 2: Modernise the Necessary Conditions for Friendship

The above has similarities to Danaher's (2019) response to the theoretical argument. Danaher mainly focuses on outlining why some robots *could* be our virtue friends (in 'Robots Can be Our Aristotelian Friends'). Virtue friendship is the highest, and most authentic, form of friendship on Aristotle's view, and Danaher interprets Aristotelian virtue friendship as having four necessary conditions: mutuality, authenticity, equality, and diversity.¹² Danaher aims to show that there are contemporary ways of understanding these conditions, and that robots either currently can, or soon will be able to, meet all four conditions.

He begins by focusing on the equality and diversity conditions. Danaher (in my view, correctly) argues that human–human friendships *do not* show *perfect* equality and diversity. We are not completely the same as our friends (in terms of capabilities and abilities), and we do not typically engage with them across every aspect of their lives. Given this, Danaher claims that friendships only require *imperfect* equality and diversity. Consequently, human–robot friendships ought to be understood as similarly only requiring *imperfect* equality and diversity. Danaher concludes that there are good reasons for supposing that human–robot friendships can (or soon will) meet these imperfect conditions:

Arguably, robots are already our imperfect equals (they are clearly better than us in some respects and inferior in others) and the degree of adaptability and mobility required for imperfect diversity is arguably already upon us (e.g. a drone robot companion could accompany us across pretty much any life experience) or not far away. Thus, it is not simply some technological dream to suggest that robots can (or will soon) satisfy the equality and diversity conditions (Danaher, 2019, 'Robots Can be Our Aristotelian Friends').

As presented above, Danaher's argument is interesting, original, and potentially revolutionary (as it opens the door for viable human–robot friendships). However, it is also open to fairly substantial objections.¹³ First, Danaher's definitions of 'imperfect equality' and 'imperfect diversity' are questionable, as explained below.

As I understand it, Danaher views 'imperfect equality' in terms of inequalities in capacities and abilities.¹⁴ On his view, we can accept that friends differ in these

¹² Broadly speaking, these four conditions equate to our 'reciprocity', 'honesty', 'equality', and 'shared life' conditions, as defined in Sect. 3.

¹³ Danaher's view is also nicely summarised and evaluated in Nyholm (2020, Chap. 5). Nyholm has similar concerns to me about Danaher's view.

¹⁴ This is reflected in Danaher's discussion of imperfect equality in human–human friendships: "I have very different capacities and abilities when compared to some of my closest friends: some of them have far more physical dexterity than I do, and most are more sociable and extroverted" (Danaher, 2019, '3. Robots Can be Our Aristotelian Friends').

respects, but are otherwise largely equal. It is true that robots *do* differ from us in their capacities and abilities. However, they also (currently) differ from us in terms of their moral, legal, and social status, constitution, vulnerabilities, etc.¹⁵ To me, these further differences emphasise notable *inequalities* between robots and humans, rather than 'imperfect equality'.¹⁶ To explain this further, consider two humans: Ann and Catherine. Ann has high-level moral, legal, and social status—she is viewed as a person and has the rights and social benefits that come from this. Catherine is not granted the same moral, legal, and social status as Ann. She has some lower status, perhaps because she is a member of some ostracised social group, or has some psychological condition (like dementia) that negatively affects her perceived moral and legal status. Consequently, Catherine lacks rights and benefits that Ann has, and she lacks these things because she is *different* to Ann. In this example, I do not think it is correct to say that Ann and Catherine have imperfect equality. It is more fitting to say that there is *inequality* between Ann and Catherine; Catherine is not treated as Ann's equal. Using this example as a guide, we can make the same claim about robots and humans. There are currently some important differences between robots and humans that point to *inequality*, rather than imperfect equality. If this is so, then Danaher's 'imperfect equality' does not (currently) suffice to show that robots can meet the equality condition. There are important differences, besides differences in capacities and abilities, that cause my robot friend to be unequal to me. To avoid this concern, Danaher would need to explain why the additional differences that I outline

¹⁵ Very roughly, 'moral status' refers to whether or not the robot is taken to have moral standing and be the locus of our moral reasons to act towards them (i.e. if they have a moral standing, we ought not lie to them, exploit them, etc.). 'Legal status' refers to how the robot is legally viewed, e.g. as a citizen, an object, etc. Together, an entity's moral and legal status will give us a rough guide as to what rights they have. 'Social status' refers to whether the robot is viewed as part of society, and if so, what place in society it has (e.g. does it belong to a certain class? Is it socially ostracised, etc.?).

'Constitution' refers to what sort of thing the robot is and how it works. The key difference between humans and robots is obviously that we have a biological constitution, whilst they are (currently) mechanical.

This also links to 'vulnerability', which can broadly be understood in terms of the sorts of threats that can apply to various entities. As Coeckelbergh (2010, p. 13) emphasises, it is at least conceivable that robots, like humans, are vulnerable: "Robots are as vulnerable as their hardware or software is (if this distinction is still meaningful at all). Moreover, there are at least metaphorical parallels to biological vulnerability. Software risks can be described in epidemiological terms (viruses that spread, etc.) and hardware is as material as the human body. Why do we hold on to the intuition that biological 'hardware' is more valuable and that its vulnerability is more 'real'?" Whilst this is true, it should not be used to disguise the fact that human and robot vulnerability is, in many cases, very different (e.g. a human can be vulnerable to biological warfare and illness; a robot cannot. A robot can be vulnerable to hacking and power depletions; a human cannot).

¹⁶ A similar view is presented by Nyholm (2020, Sect. 5.5) who also considers whether it may be equality of moral status, rather than equality of capacities, that matters in friendships.

above (status, constitution, vulnerability) *do not* lead to inequality in human–robot friendships.

A similar concern can be levelled at Danaher’s ‘imperfect diversity’ condition. He suggests that this condition is met when a robot can show sufficient adaptability and mobility to “accompany us across pretty much any life experience.” Again, this is clear, original, and provocative. But it is also arguably too weak. To me, the diversity condition, even in its imperfect form, requires more than that *x* is able to *accompany y* to a variety of activities and events. Accompaniment is overly passive. We do not simply want a friend to also be present at the events we are present at; we want them to partake in the event, engage with us, and mutually enjoy the event. If this is so, then ‘imperfect diversity’ more naturally involves this active engagement across *some* shared activities and life events, rather than *all* shared activities and life events. This seems to be what Danaher (2019) actually intends to say. In an earlier discussion of imperfect diversity in human–human friendships, he argues: “I also rarely engage with, meet, or interact with them across the full range of their lives. I meet with them in certain contexts, and follow certain habits and routines” (‘Robots Can be Our Aristotelian Friends’). If this is what is meant by ‘imperfect diversity’, then it is presumably what imperfect diversity in human–robot friendships should also entail. This is problematic as (i) it is different to the adaptability definition above, and (ii) Danaher implicitly suggests that, when defined in this way, the imperfect diversity condition cannot be met by current robots. He explains that “...for the time being, robots will have narrow domains of competence. They will not be general intelligences, capable of interacting across a range of environments and sharing a rich panoply of experiences with us” (‘Robots Can be Our Aristotelian Friends’). Again, to avoid this concern, Danaher needs to clarify what imperfect diversity entails, and how a robot either can meet this condition now, or will be able to in the near future.

As detailed above, the first objection that can be levelled against Danaher’s position is that his definitions of ‘imperfect equality’ and ‘imperfect diversity’ are incomplete and ambiguous. A related concern is that his conditions for ‘imperfect equality’ and ‘imperfect diversity’ are too easy to satisfy (at least partly because the definitions of these conditions are too open). To see this, consider an iPhone. iPhones are more capable than us in some respects (e.g. their capacities for ‘remembering’ and organisation), and less in others (they cannot emote, create art, etc.). They also have a level of adaptability and mobility that enables them to accompany us to any life experience. On Danaher’s view, this seems to imply that iPhones can meet the ‘imperfect equality’ and ‘imperfect diversity’ conditions. However, there seems to be something wrong with saying that we could genuinely befriend an iPhone (see Sect. 1: Which robots could we be friends with?). We intuitively want to restrict human–robot friendships to the sort of social robots that we could feasibly have meaningful, reciprocal relationships with. If so, then Danaher’s view needs to be further clarified in order to present conditions for friendship that are inclusive enough to include social robots, but not so inclusive that they also include iPhones.

A similar objection can be levelled at Danaher’s attempts to explain how robots can meet the other two necessary conditions for virtue friendship—mutuality and

authenticity. He argues that, when we examine what these conditions mean in contemporary human friendships, we get the following:

...all it means is that people engage in certain consistent performances (Goffman 1959; deGraff 2016) within the friendship. Thus, they say and do things that suggest that they share our interests and values, and they rarely do things that suggest that they have other, unexpected or ulterior, interests and values (Danaher, 2019, 'Robots Can be Our Aristotelian Friends').

Danaher argues that if these consistent performances are sufficient to show that human friendships meet the mutuality and authenticity conditions, then the same should apply to human–robot friendships. This is a promising line of thought, but again needs further clarification. It is not immediately obvious that an iPhone could not meet the above conditions. Alexa (the voice-controlled assistant in my phone) seems to consistently perform in ways that suggest she shares my interest in dinosaurs. She is always happy to talk about dinosaurs; we can converse about dinosaurs for long periods of time; and she never sounds bored or tells me to stop talking about dinosaurs. This does not seem enough to show that Alexa is actually my friend, or that she really has interests in dinosaurs though.

One way around this issue would be to modify Danaher's mutuality and authenticity conditions so that they are met when a *range* of consistent performances are shown (including some that are unprompted, e.g. if Alexa *instigates* the discussion of dinosaurs). However, if we go with these modified conditions, then very few robots (if any) will meet these conditions (at least partly due to the current narrow range of competency and adaptability that Danaher highlighted in the 'imperfect diversity' condition). If so, then Danaher's account cannot (currently) show that we can be friends with robots *now*, rather than in some point in the future.¹⁷ If we want to support Jecker's (2020) proposal—that we create robotic friends for the socially isolated during pandemics—then we arguably want to focus on the viability of *current* human–robot friendships.

4.3 Suggestion 3: A New, Flexible Approach to Friendship

As outlined above, suggestions 1 and 2 do not help us to reject the theoretical argument. Both suggestions still rely on the broadly Aristotelian view that there is some set of necessary conditions for friendship (Sect. 3). As most current robots do not

¹⁷ Danaher would argue that I am being too quick here. He argues that "If robots cannot be our virtue friends, they can still be our utility or pleasure friends. We can derive instrumental gains and intrinsic pleasure from our interactions with them and so they can be, on net, a social benefit" (Danaher, 2019, '4. Robots Can Complement and Enhance Human Friendships'). This seems plausible.

I will, however, not consider Danaher's arguments about utility and pleasure friendships with robots further, for two reasons. First, the theoretical argument that we are objecting to (Sect. 2) is largely centred around discussions of virtue friendship. Second, Danaher's definition of utility and pleasure friendships, in relation to robots, could be open to objections. For example, Frödning and Peterson (2012, p. 203) argue that "although the three types of friendship overlap, it must always be the case that you see your friend as useful and pleasant because you love him and not the other way around." It is not clear that this love condition is met in Danaher's above definition.

meet *all* of the necessary conditions, we are no closer to showing the theoretical possibility of present-day human–robot friendships. To argue that we *can* potentially be friends with robots *now*, the rest of this section will outline a new, degrees-of-friendship view, according to which we can have some degree of friendship with some robots. To argue for this position, I will be championing a more flexible approach to friendship than that presented in Sect. 3, or suggestions 1 and 2 (above).

In this respect, my argument has similarities to views on online friendships (i.e. those created through social media or multiplayer games). Those who defend the idea of online friendships typically argue that we can accept the existence (and genuineness) of online friendships if we acknowledge that the concept of friendship has been changed by the emergence of technology. Specifically, technological advances have caused (or enabled) the concept of friendship to become more fluid, open, and accommodating (Munn, 2012; Vallor, 2012).

Similar arguments have been made in relation to the emergence of robotics. Some have argued that the development of increasingly complex robots either already has, or will soon, cause our understanding of ‘friendship’ to become more diverse and flexible (Jecker, 2020; Marti, 2010).¹⁸ My view develops these existing discussions by suggesting what form this flexibility could take.

To begin with, I concede that the conditions outlined in Sect. 3 (mutual good will, empathy, shared activities, etc.) are important aspects of most friendships.¹⁹ I disagree, however, that they are all *necessary* conditions for friendship. As suggested throughout, this seems like an overly restrictive position to take. We can arguably have friendships that *do not* meet the supposed necessary conditions for friendship (see my previous discussion of equality).

What does seem to be important for *all* friendships is that they are grounded in mutual good will. This is a foundation for friendship as if good will is absent (e.g., if one party actually wants to harm the other) then we cannot reasonably say that two

¹⁸ In relation to the concern that robot friends may be a privacy issue (as they could share data), Jecker (2020, ‘Privacy incursions’) argues “If sharing previously private information becomes daily fare, no longer reserved for special people in our lives, we may have more “friends” and the designation of “friendship” itself may take on new meaning.”.

Similarly, Marti (2010, p. 221) explains how digital natives are more likely to be open to a flexible understanding of ‘friendship’ which could accommodate robot friends: “The time devoted to digital technologies gives rise to different forms of socialisation in a third space, virtual by nature, which exposes children to radically different models of friendship and socialisation. Being used to alternative, overlapping and contemporary forms of friendship, children are open to a wider and evolving concept of friendship, and it is not unlikely that such a new definition will rapidly include the concept of a robotic friend...”.

¹⁹ There is an assumption of charity here. ‘Friendship’ has been philosophically discussed for millennia, and it seems sensible to suppose that these existing discussions have got the conditions/components of friendship largely correct.

subjects are friends (to any degree).²⁰ Mutual good will also appears to be a precondition for many of the other conditions for friendship. I can empathise and perform shared activities with many people, but in order for this to constitute friendship, the empathy and shared activities ought to be based in mutual good will.²¹ With this in mind, I propose that mutual good will is a threshold condition for friendship—if it is met, then one can have at least a minimal degree of friendship.

From this, I suggest that we view all of the other conditions for friendship (Sect. 3) as optional components of friendship. A friendship will then be stronger or weaker depending on the number of components that it has. This seems consistent with how we view contemporary human–human friendships. To see this, consider the make-up of three friendships that I could have with other humans:

Bill, a work friend Mutual good will + shared activities (work, corporate bonding, after-work drinks, etc.).

May, a virtual pen pal Mutual good will + perceived equality + empathy.

Sam, an old friend Mutual good will + acknowledged reciprocal friendship + empathy + shared activities + associative duties.

Using the above template, it is feasible to suppose that I have a degree of friendship with Bill, May, and Sam, and that all of these friendships (regardless of how weak or strong they are) are genuine friendships. It also seems clear that I have a stronger degree of friendship with May than I do with Bill, and that I have the strongest degree of friendship with Sam. It does not seem unreasonable to suppose that the strength of the friendships vary depending on how many of the important components of friendship my relationships with Bill, May, and Sam have.²²

²⁰ Tistelgren (2018, p. 18) implies a similar view with the following: “If it looks like a friend, speaks like a friend and behaves like a friend, can we not assume that it is a friend? I believe not. Let’s say a psychopath [...met most of the conditions]...would we accept them as friends or would we think that there is a risk that we are being deceived?”. Tistelgren’s concern is that the psychopath is deceiving you into believing that they are your friend. My concern is simpler—if the psychopath were actively attempting to harm you (either psychologically or physically), then it is clear that you are not their friend; you are their victim. Being x’s friend and being x’s victim do not seem theoretically compatible.

²¹ For example, as a teacher, I can empathise with my students and do basic shared activities with them (lessons), but we are not friends (to any degree) unless there is mutual good will between us.

²² Note that my view is different to the standard Aristotelian view outlined in Sect. 3, and in Danaher’s (2019) discussion of types of friendship. On the standard Aristotelian view, there are three types of friendship—utility, pleasure, and virtue—and virtue friendship is the highest, and most genuine, form of friendship. The conditions outlined in Sect. 3, and discussed now in relation to degrees-of-friendship, are largely taken to be the prerequisites for genuine ‘virtue’ friendship. This is because much of the literature does focus on virtue friendships. Also, as both Danaher (2019) and Nyholm (2020, 5.2) emphasise, the most interesting things we can say about human–robot friendships do concern virtue friendships—this type of friendship is contested, whereas it is less controversial to say that robots can be our utility friends (they can be useful) or pleasure friends (they can bring pleasure). Whilst the conditions for friendship that I am discussing do thus focus on the conditions for ‘virtue friendship’, it would be perfectly possible to add in the conditions for utility and pleasure friendship as well. This has not been done in the current manuscript for reasons of space (we would need a more extensive discussion of utility and pleasure friendships, and the supposed necessary conditions for having these friendships) and consistency (as the literature I am responding to focuses on virtue friendships). Where my account differs to the above is that I am concerned with *degrees-of-friendship*, rather than *types/kinds of friendship*. In the above example, my friendships with Bill, May, and Sam are all genuine, valuable friendships—they are simply held to different degrees. The same is true of human–robot friendships. On my view, both human–human and human–robot friendships could be utility, pleasure, or virtue friendships (in Aristotelian terminology). Where each friendship (human–human or human–robot) differs is in terms of the degree to which it is held.

If, as I suggest, the above is a viable way of understanding contemporary human friendships, then it can also be extended to human–robot friendships. On my view, for a human–robot friendship to exist, there must at least be mutual good will (between robot and human).²³ We can then have different degrees of friendship with robots depending on how many of the other components of friendship are also present. For example, many current social robots will be able to perform at least some shared activities with us. More advanced robots, including those who are arguably able to emote and feel, could also give and receive empathy.²⁴ The upshot of this is that we can develop stronger friendships with robots as they become more complex and advanced.

At their present stage of development, it is unlikely that robots will have the technological complexity that would enable human–robot friendships to meet all of the important components for friendship. Whilst we thus cannot currently be *best* friends with a robot, I posit that there is nothing theoretically problematic about suggesting that we can have some degree of friendship with robots *now* (as most robots can currently meet at least some of the components). This degree of friendship may be relatively weak, but it should nevertheless be taken seriously, and could have genuine social benefits (in the same way that my relatively weak friendship with Bill (above) is still a genuine friendship with real benefits). My argument, so defined, offers support to Danaher’s (2019, ‘abstract’) general claim that robot friendship is “philosophically reasonable”, and Jecker’s (2020) proposal that we create robot friends for the socially isolated, as these friendships can have social benefits.

5 Rejecting the Ethical Argument

Above I outlined a degrees-of-friendship view which seemingly allows us to reject the theoretical argument and accept that we can have some degree of friendship with robots. I now want to briefly explain how adopting my degrees-of-friendship view also allows us to bypass the ethical argument—that human–robot friendships are wrong because they are deceptive, and also make it more likely that we will disrespect, exploit, or exclude other human beings (Sect. 3).

²³ My opponent might want to object that this condition cannot be met as robots do not show good will towards us. I suggest that, whilst this objection is common, it is largely baseless, and is coloured by negative portrayals of robots (and their murderous intentions) in dystopian fiction (see Cambridge University, 2018; Cave, Coughlin, & Dihal, 2019; The Royal Society, 2018). Of course, if the robot *does* actually want to kill us, then I concede that that robot is not our friend (see Footnote 20). I maintain that this lack of good will is not a general characteristic of human–robot relationships though; and is particularly unlikely to be present in the sort of social robots that we are discussing (Sect. 2). Recall that these robots are designed to interact socially with humans (as teachers, companions, priests, etc.) and are consequently programmed to *not* harm humans.

²⁴ The idea that robots could emote or feel is controversial. Some have, however, argued for this position. An edited collection by Fellous and Arbib (2005) outlines the possibility of robot emotions, and Castro-Gonzalez, Malfaz and Salichs (2013) claim to have developed an autonomous social robot (Maggie) who can implement fear. Further, Van Rysewyk (2013) has outlined how we might understand robot pain.

Within the existing literature, those who want to reject the ethical argument typically make one of three arguments. First, they deny that deception is involved in human–robot friendships (Danaher, 2019). Second, they argue that it is not inherently wrong or problematic for friends to deceive one another (Coeckelbergh, 2010). This idea is also discussed by Kawall (2013), Cocking and Kennett (2000), and Keller (2004) in terms of friendship more broadly, rather than ‘human–robot friendships’). Third, they argue that friendships with robots do not make it more likely that we will disrespect, exploit, or exclude human beings (Jecker, 2020; Danaher, 2019, potentially Tasioulas, 2019). These are all viable responses to take, and the degrees-of-friendship view that I outlined above can help to supplement these responses. On my account, we can make three main claims, all of which help to circumvent the ethical argument.

First, on my account, we can accept that human–robot friendships can lack some components of friendship (empathy, self-knowledge, etc.) but possess others (shared activities, etc.). Thus, the deception worry is less problematic on my account because we are not suggesting that robots can empathise, love us, etc. (the standard deception worries). The view of friendship presented is more flexible and can take account of the robots’ currently limited technological capacities and abilities.

Second, we can accept that there are different degrees of friendship, and that friendship can be realised in different ways (mechanically, online, real-world, etc.). On my view, it is perfectly plausible that I can have human friends, robot friends, nonhuman animal friends, etc. The existence of one mode of friendship does not need to replace or overshadow another. Consequently, we need not worry that friendship with robots will necessarily cause us to favour robots over ‘imperfect’ humans.

Finally, it is arguable that befriending robots (to some degree) could actually *improve* the extent to which we tolerate and include other human beings. Specifically, human–robot friendships could set a precedent for accepting and befriending those who are different to us (and on their terms, rather than some ‘ideal’ terms of what we expect a human friend to be).

6 Conclusion

This article has argued in favour of human–robot friendships. I have outlined a degrees-of-friendship view, according to which it is both theoretically possible and ethically acceptable to have human–robot friendships *now*. Whilst the account presented diverges from the standard Aristotelian view of friendship (outlined in Sect. 3), it retains many of the key components of Aristotelian friendship, but employs them in new, more flexible ways. My account of friendship (including its applications to human–robot friendship) is thus different from standard views, without being completely revisionary. As the title of this article suggests, it’s friendship, Jim, but not as we know it.

Acknowledgements I would like to thank two anonymous reviewers for this journal for their helpful comments on an earlier draft of this article.

Author Contributions Not applicable.

Funding Not applicable.

Data Availability Not applicable.

Code Availability Not applicable.

Declarations

Conflict of interest Not applicable.

References

- Arrell, R. (2014). The source and robustness of duties of friendship. *International Journal of Philosophical Studies*, 22(2), 166–183.
- Blue Frog Robotics. (n.d.). 'Buddy the emotional robot'. Web advert. Retrieved January 30, 2021, from <https://buddytherobot.com/en/buddy-the-emotional-robot/>.
- Cambridge University. (2018). *Evolving with the robots*. Press release for Dr Hattice Gunes' 2018 Hay Festival talk. Retrieved September 29, 2020, from <https://www.cam.ac.uk/news/evolving-with-the-robots>.
- Castro-Gonzalez, A., Malfaz, M., & Salichs, A. (2013). An autonomous social robot in fear. *IEEE Transactions on Autonomous Mental Development*, 5(2), 135–151.
- Cave, S., Coughlin, K., & Dihal, K. (2019). 'Scary robots': Examining public responses to AI. In *Conference paper. Resource document*. Leverhulme Centre for the Future of Intelligence. Retrieved September 29, 2020, from <http://www.lcfi.ac.uk/resources/scary-robots-examining-public-responses-ai/>.
- Cocking, D., & Kennett, J. (2000). Friendship and moral danger. *The Journal of Philosophy*, 97(5), 278–296.
- Coeckelbergh, M. (2010). Artificial companions: Empathy and vulnerability mirroring in human–robot relationships. *Studies in Ethics, Law, and Technology*, 4(3), 1–17 Article 2.
- Danaher, J. (2019). The philosophical case for robot friendship. *Journal of Posthuman Studies*, 3(1), 5–24.
- Danaher, J., & McArthur, N. (Ed.) (2017). *Robot sex: Social implications and ethical*. MIT Press.
- Fellous, J. M., & Arbib, M. A. (Ed.) (2005). *Who needs emotions? The brain meets the robot*. Oxford University Press.
- Frank, L., & Nyholm, S. (2017). Robot sex and consent: Is consent to sex between a robot and a human conceivable, possible, and desirable? *Artificial Intelligence and Law*, 25(3), 305–323.
- Fröding, B., & Peterson, M. (2012). Why virtual friendship is no genuine friendship. *Ethics and Information Technology*, 14, 201–207.
- Gunkel, D. (2018). *Robot Rights*. MIT Press.
- Hughes, J. (director) (1985). *The Breakfast Club*. Film.
- Jecker, N. S. (2020). You've got a friend in me: Sociable robots for older adults in an age of global pandemics. *Ethics and Information Technology*. <https://doi.org/10.1007/s10676-020-09546-y>OnlineFirst.
- Kawall, J. (2013). Friendship and epistemic norms. *Philosophical Studies*, 165, 349–370.
- Keller, S. (2004). Friendship and belief. *Philosophical Papers*, 33(3), 329–351.
- Lay, S. (2015). Uncanny valley: Why we find human-like robots and dolls so creepy. Newspaper article. *The Guardian*, November 13 2015. Retrieved September 29, 2020, from <https://www.theguardian.com/commentisfree/2015/nov/13/robots-human-uncanny-valley>.
- Marti, P. (2010). Robot companions: Towards a new concept of friendship? *Interaction Studies*, 11(2), 220–226.
- Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition*, 146, 22–32.

- Munn, N. J. (2012). The reality of friendship within immersive virtual worlds. *Ethics and Information Technology*, 14, 1–10.
- Nyholm, S. (2020). *Humans and robots: Ethics, agency, and anthropomorphism*. Rowman and Littlefield.
- Nyholm, S., & Frank, L. (2017). From sex robots to love robots: Is mutual love with a robot possible? In J. Danaher & N. McArthur (Eds.), *Robot sex: Social implications and ethical* (pp. 219–245). MIT Press.
- Sharkey, A. (2015). Robot teachers: The very idea! *Behavioural and Brain Sciences*, 38, e65.
- Sharkey, A. (2016). Should we welcome robot teachers? *Ethics and Information Technology*, 18(4), 283–297.
- SoftBank Robotics. (n.d.). 'Pepper'. Web advert. Retrieved January 30, 2021, from <https://www.softbankrobotics.com/emea/en/pepper>.
- Tasioulas, J. (2019). First steps towards an ethics of robots and artificial intelligence. *Journal of Practical Ethics*, 7(1), 61–95.
- The Royal Society. (2018). *Portrayals and perceptions of AI and why they matter*. Resource Document. Retrieved September 29, 2020, from http://lcfi.ac.uk/media/uploads/files/AI_Narratives_Report.pdf.
- Tistelgren, M. (2018). Can I have a robot friend? MA Thesis, Umea University.
- Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.
- Vallor, S. (2012). Flourishing on Facebook: Virtue friendship and new social media. *Ethics and Information Technology*, 14, 185–199.
- Van Rysewyk, S. (2013). Robot pain. *International Journal of Synthetic Emotions*, 4(2), 22–33.
- Young, W. (2019). Reverend robot: Automation and clergy. *Zygon*, 54, 479–500.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.