

# This World Needs More (than One Kind of) Love. A Modest, Murdoch-Inspired Proposal to Take the Heart in Effective Altruism More Seriously

## Abstract

Debates about effective altruism (EA) often focus on reason and evidence in promoting the good. I suggest looking at the “heart” in EA, and taking seriously that, and how, other-regarding attitudes, specifically love, can both motivate and complicate moral agency and inquiry. This counters a misleading dichotomy between reason and emotion and a crude view of love as a “care-o-meter”. Building on MacAskill’s idea that there is an “engineering” and a “science” level to morality, I suggest that there is a pragmatic and an epistemic level of promoting the good and that two different kinds of love operate at each: at the former, which would be MacAskill’s “engineering” level, and concerned with translating what we know about the good into action, an outward impartial, agapistic love is most adequate, with a philic dimension regarding rational self-esteem; at the latter, which would be MacAskill’s “science” level of morality, and concerned with realising what is good, an inward erotic love is important. These levels should be distinguished because confusing them leads to a distorted view of the good. It also leads to a mismatch between effective altruist normative theory and actual effective altruists’ plurality of motivations, complicating EA’s transition to a social movement. Making moral life and theory potentially more complicated (but realistic), I turn to Iris Murdoch and suggest that these levels could be modelled on her distinction between “public morality” and “private morals”. I illustrate the benefits of my proposal with two vignettes involving two fictionalised moral counsellors: the android robot Sophia and the rationalist human Ajax.

## 1. Promoting the Good. A Pragmatic and an Epistemic Level

We live in a world with an almost endless amount of suffering that will move anyone with a heart in their chest to want to do good. No rocket science and no complicated theory are needed to appreciate this simple fact. Yet, in order to do good better, some have argued that we need to abstract from other-regarding feelings like love or care because ... our hearts are just not big enough to comprehend and process the actual enormity of the suffering. We respond lovingly and with compassion to the individual suffering in front of us, but we become overwhelmed and fatalistic once we realise the scale and enormity of all the suffering we should care equally about. According to influential effective altruists like Nate Soares, then, it is often not that we *do not* care when we ultimately do nothing to alleviate that large-scale suffering. Rather, as Soares suggests in a blog post, our “care-o-meters” seem just as unsuitable to give us precise readings of the misery we need to tend to as a simple thermometer is to measure a

bushfire.<sup>1</sup> Thus, if one wants to be effective in one's caring, one should turn to numbers and rely on those when putting a universal concern for suffering – regardless of where it occurs – into practice.

Historically, this universal concern has been utilitarianism's signature axiom but, again, hardly needs a complicated moral theory for backup. Moreover, in order to have as many people as possible adopt that concern and turn to better ways of doing good, it might be helpful to grow effective altruism (EA) as a social movement. In a recent definition of EA, Will MacAskill assures the interested reader that "we are not attempting to describe a fundamental aspect of morality",<sup>2</sup> and that EA is "consistent with any moral view".<sup>3</sup> More precisely, he advertises EA as a piece of moral engineering, and as such as different from moral science, with science being "the attempt to discover general truths about the world we live in"<sup>4</sup> and engineering "the use of our scientific understanding to design and build structures or systems that benefit society".<sup>5</sup> The normative view that informs the idea of benefit here is "tentative impartial welfarism",<sup>6</sup> which MacAskill takes to be an ecumenical view, a view that could be endorsed by many. I believe that he is right if this is to be understood as applying to the "engineering" level and wrong if it is to be understood as applying to the "science" level (and I believe he would agree with this).

I also agree that moral science or philosophy needs to pay more attention to the engineering side of things, or to what I call "moral pragmatics".<sup>7</sup> Yet the trouble is that in its attempts to build a social movement, EA does not just come across as an ecumenical or theory-neutral way to engineer a better world, but vies for the hearts of people as well. This risks distorting the larger enterprise, which is the science of recognising the good. It does so because it conflates various ways in which people can and need to be guided by love in their actions, which is what allows them to learn about the good in the first place. In the disciplines that work on recognising the good (e.g. moral and political philosophy, psychology, anthropology, among others), it is still a point of contention whether an impartial concern with welfare on the basis of quantifiable concepts of wellbeing and quality of life

1 See Soares, "On Caring." I have built on the sentiment expressed in the blogpost to introduce the problem about the limits to our love.

2 MacAskill, "Definition," 11f.

3 Ibid., 16.

4 Ibid., 11.

5 Ibid., 12.

6 Ibid., 14.

7 Cojocaru, "Doing Ethics?"

is methodologically always superior to a partial concern with specific rights violations or to somewhat ineffable intuitions about the common good that is more than the sum of individual benefits, not to mention conceptions of inherent value. On this perennial axiological question, the jury is still out, and, in my opinion, all perspectives can elucidate important aspects of the difficult and problematic moral landscape we all have to navigate by our own lights. Since EA axiologically operates just on the impartial welfarist view, it is in as good a position to exclusively promote the good as one would be to water a garden with only a fire hose: yes, that can help in times of drought, yet it will destroy many of the more delicate things that thrive in it.

I believe it would be better for EA to salvage its straightforward utilitarian heritage, yet understand it as confined to the level of what Iris Murdoch has called public morality, which is where moral engineering should take place. Murdoch helpfully distinguished “rough general rules of morality, such as constitute important inspirations and barriers in politics and public life, from a progressive spirituality, connected with a total change of consciousness”.<sup>8</sup> And she acknowledges that “[u]tilitarian ideals now support large political ends (ecology, feed the hungry) and might be argued to be (rather than the cultivation of private virtue) what the planet needs”.<sup>9</sup> I find this distinction extremely important and overlooked and accept it here without further qualification (say, regarding the precise version of utilitarianism, rule or act, etc.). My all in all very simple and modest point is that EA is a sensible approach to promoting the good by “just doing good” where it is philosophically uncontroversial what the good is. Applying reason and evidence, we can engineer better impartialist welfare performance and systems.

However, EA might want to become more modest, too, because it currently – by vying for people’s hearts and minds in its attempts to form a social movement – misses the importance of the particular and partial

8 Murdoch, *Metaphysics*, 367.

9 Ibid., 493.

relations and perspectives humans need to have as moral agents.<sup>10</sup> To be clear, I do not think that those who stand in particular relations with us morally count more *tout court*. I agree with the universalist, impartial stance. Instead, my intuition is that an agent's ongoing, necessarily particular moral life has an added and morally relevant *epistemic* benefit when it comes to finding out what the good is. For all that can go awry in these relations, it is with regard to these particular relations that one is both motivated, and in a position, to learn about the form and the importance of the specific good. This is where persons *learn* that others can – and actually ought to – be beneficiaries of one's selfless actions. They learn this *because* of the love and care they receive and develop.<sup>11</sup>

Building on Murdoch, we can think of two spheres of morality: a public sphere in which moral agents operate on fairly well-known, uncontroversial and robust ideas of the good. Here, humans join forces to address issues that all agree are problematic, and a focus on impartially bringing about welfare and happiness could already achieve a lot of good. Murdoch says that “utilitarian considerations are in general *prima facie* relevant because we all understand the importance of happiness. It is always a, not necessarily final, argument against doing something to someone, that it will reduce his happiness”.<sup>12</sup> In the public realm, it is perfectly fine to work with axioms, which are, to some extent, unfounded. Murdoch calls axioms

10 I am aware of the potential problem it poses to speak of “moral agents” from a Murdochian perspective, as it suggests that the focus is with the pragmatic, agential side of morality, while Murdoch was concerned precisely with the fact that morality ought not to be reduced to observable acts but includes one's whole being, the way one attends to the world, feels, imagines and engages in self-examination. I follow Lawrence Blum in how he addresses this problem (Blum, “Visual Metaphors,” 309, fn. 3). Blum, I believe rightly, states that “no alternative single term seems [...] to capture this complex truth” (ibid.), and so I stick with “agent”, too.

11 I realise that much more support from moral psychology would be needed for this claim, but, philosophically, this is just to take the position that humans do *not* enter this world with an innate knowledge of the good and how to promote it in the world. Much in terms of human nature, its malleability and dependence on social context provided by other human beings can be found in Murdoch's congenial friend Mary Midgley (Midgley, *Beast and Man*). To account for this view from a utilitarian perspective, see e.g., Mill on the fact that “[t]here is no selfishness equal to that of children, as everyone who is acquainted with children well knows” (quoted in Mill, *John Stuart Mill*, 15). He goes on to suggest that causing pain and suffering in others through indulging one's own wishes will be restrained in children only by a competing affection for particular others, from which only later in life truly moral feelings can follow (ibid.).

12 Murdoch, *Metaphysics*, 365.

specifically “effective through being impersonal and abstract”,<sup>13</sup> indeed, they are “instruments of the public scene”).<sup>14</sup>

However, from a moral science perspective, that is not the whole story of or final word on the good. More complex, and sometimes difficult-to-articulate views of the good are needed in the case of ... more complex problems. With Murdoch, this is the private (yet not unpolitical!) sphere in which moral agents are still very much in the business of gaining a clear vision of what the good could possibly be. To return to my “caring” metaphor: having joined forces to fight the bushfire, people return to their gardens and are free – and uncertain – as to what to grow or whom to accommodate in them, especially in times of drought. What is more, different kinds of love operate at these different levels or spheres, which is why it would help to reintroduce a nuanced understanding of love as a force for good.

## 2. What’s Love Got to Do with It?

Let me start with a few, hopefully uncontroversial ideas about love. I assume that love, in one form or another, is indispensable as a source of meaning and value for human beings.<sup>15</sup> I also assume that love enables agents to perceive the objects of their love, themselves, as well as the whole world (whatever that is from the agent’s perspective) differently. This difference is of a particular kind. Love has a qualitative certainty that is lacking in arbitrary choices. The things agents feel compelled to do, or the account of the world they give, when they are guided by love seem well grounded. To care for or attend lovingly to the needs of my ageing mother is not arbitrary, nor is attending to the special qualities of my partner, to the situation of people living in poverty, nor to the suffering of farm animals. When humans are “guided” by love they are moved by something that is real and transcends them. Love, then, as I use the term here, is an indispensable element of human life and can guide agents to both apprehend and add values to their environment, not least by promoting the wellbeing of others.

13 Ibid., 380.

14 Ibid., 381.

15 This may seem a platitude to which we can all agree as long as we do not specify what love is. Some philosophers have also tried to make the general claim comprehensible, too: see e.g. Wolf, “Love,” or Milligan, *Love*.

But is an understanding of love also helpful in organising an agent's attention? Can we say something about who or what really *deserves* one's love? While conventional wisdom has it that love grows by being given away, love also involves resources that are scarce, like time, attention, and, truth be told, money. In one way or another, these have to be "invested" when in love, which is probably why jealousy is the flip-side of love. It would be good, then, to know how one ought to spread one's love. *Prima facie*, the love that motivated Mother Theresa seems more admirable than my love for my mother.

At the same time, in light of the various objects that agents feel compelled to love, it might be impossible to adjudicate between one agent's love for her mother and another agent's love for her partner\*s, and between both their loves and someone else's love for mankind. Indeed, it has been argued that "one reason that so few philosophers write about love is that what people actually love cannot be determined philosophically without grotesque oversimplification".<sup>16</sup>

In the tradition, however, philosophers have not just oversimplified things, they have also distinguished different kinds of love: *eros*, *philia*, *storge*, and *agape*.<sup>17</sup> "Eros" designates an attraction to beauty in all its manifestations (from beautiful bodies to beautiful things, thoughts, and institutions). It has less to do with action than with realisation and appreciation, so much so that some have argued that, while a self-transcendent force, it is less focused on sharing and giving than on possessing and receiving.<sup>18</sup> "Philia" designates a love between like-minded people that grows through habituation and recognising the virtues in each other. "Storge" designates the familial love with which one responds to the individual needs of one's dependants. Finally, "agape" is the love behind specifically charitable deeds, an indiscriminating, non-reciprocal and selfless kind of love where agents take nothing for themselves and give everything unconditionally.

16 Brentlinger, "Love," 137.

17 This distinction can be found in many places; sometimes *storge* is omitted. For easy online reference, see the entry on love by Alexander Moseley at the *Internet Encyclopedia of Philosophy*. Bennett Helm, in his entry at the *Stanford Encyclopedia of Philosophy*, accounts for the distinction, too, yet is of the opinion that most contemporary accounts of love as personal love blur it, for good reason; I disagree and think it is still valuable, especially to understand personal motivation and devotion to a cause instead of to other persons.

18 See Nygren, "Agape and Eros," 94. But cf. Milligan, *Love*, who disagrees with the negative view of *eros*.

This still does not tell me how much of my loving energy or attention should come out in any of these forms. Yet it gives me a better understanding of what I am specifically doing whenever I think I am caring about or tending lovingly to others. If I want to care better, though, does the sort of “heart” I put into EA help me to tend to everything I need to tend to? The garden and the bushfire? Prima facie, “altruism” – understood simply as benevolent concern for others – is compatible with all kinds of love. To answer that question, I now turn to the “heart” in EA, both in theory and in the self-understanding of (at least some) effective altruists.

### 3. Looking for the Heart in EA

Connecting the “altruism” at the heart of EA with an actual emotion like love seems to be heading into the wrong direction. Not only is the image of EA as a social movement defined by its emphasis on reason and rationality (for many, unrelated, if not opposed to emotions). In EA literature, too, altruism is neither understood as a hard-wired, evolutionary drive, nor as a rich emotion concept, nor as involving self-sacrifice. Let me take these in turn.

Peter Singer emphasises that effective altruists do not confirm with expectations one would have if altruism was understood along biological lines as an instinct or proto-morality that aids group selection.<sup>19</sup> This is so because the unit of selection is never the universal group that effective altruists are concerned with. Such a naturalistic strategy could also be easily debunked and it would be an open question whether acting altruistically in that sense was acting morally at all.<sup>20</sup> Similarly, a behavioural economic definition of altruism seems out of the question since expectations of reciprocity, gain in social approval (think “virtue signalling”) or fear of punishment are usually not considered genuinely moral motivations.<sup>21</sup>

Singer also emphasises that effective altruists are typically not characterised by strong emotional empathy. If anything, EA-minded agents score *lower* on emotional empathy than people who are less inclined to make utilitarian decisions.<sup>22</sup> When Katarzyna de Lazari-Radek and Singer discuss performance according to more specifically Sidgwickian standards, they also suggest that people who manage to counter their emotional concern for

19 Singer, *The Most Good*, 75–76.

20 Lazari-Radek and Singer, *Point of View*, 185–196.

21 Kitcher, “Varieties of Altruism.”

22 Singer, *The Most Good*, 79–80.

others will be better positioned to act rationally in this sense.<sup>23</sup> Even in thinkers who engage with emotion terms associated with EA, like Holden Karnofsky, we find a similar view when he speaks about “radical empathy”. He stresses that the term “empathy” “is intended to capture the idea that one *could* imagine oneself in another’s position, and recognises the other as having experiences that are worthy of consideration. It is not intended to refer to *literally feeling* what another feels”.<sup>24</sup> Actual feelings, again, seem not trustworthy, and the resulting account of “radical” (or rather: “hypothetical”) “empathy” is rather thin.

The concept of altruism that MacAskill employs is similarly thin. It “simply means improving the lives of others”,<sup>25</sup> or, rendered more philosophically “the use of evidence and careful reasoning to work out how to maximize the good [...], tentatively understanding the good in impartial welfarist terms”.<sup>26</sup>

Scanning the EA handbook, the metaphor of “prospecting” or “mining for gold” for promoting the good comes up (in Cotton-Barratt’s entry) as does, once again, “helping those around us”.<sup>27</sup> (Yet the *Effective Altruism Concepts* online encyclopaedia does not have an entry of its own explaining “altruism” (only “excited” vs. “obligatory” altruism; the former, again, according to Karnofsky, allows for some emotional coloratura in an EA’s motivation by saying that excited effective altruists are passionate about the prospect of improving the lives of others and don’t just consider that what reason demands of them).

Arguably, then, it is open to interpretation what the “heart” really means in EA, which of the many notions of “altruism” put forward is authoritative, and how it connects with people’s actual emotions. If that is so, one is allowed to speculate that the altruism at the heart of EA has something to do with *agape*. After all, altruism involves something selfless, unconditional, with altruistic agents taking nothing for themselves (where nothing can mean anything from “not benefiting oneself from the particular act” to

23 Lazari-Radek and Singer, *The Point of View of the Universe*, 59–61. Interestingly, they link this with an example of autistic agents such as Temple Grandin, who they call “a prominent welfare consultant and one of the best-known people with autism” (ibid.). Some empathy theorists would agree that Temple Grandin is a good example, but precisely for what they call false empathy (Gruen, “Empathy”) leading to morally questionable acts.

24 Karnofsky, “Radical Empathy” (emphasis mine).

25 MacAskill, *Doing Good Better*, 12.

26 MacAskill, “Definition,” 14; see also MacAskill, “Introduction,” 2.

27 Centre for Effective Altruism, “Effective Altruism Handbook.”

“giving oneself away”, sometimes in extreme and bewildering cases like George R. Price).<sup>28</sup>

However, prominent effective altruists go out of their way to assure anyone interested in the movement that the idea of self-sacrifice is not at all necessary.<sup>29</sup> Indeed, because of the often-drawn connection between altruism and selflessness and sacrifice, some have held that “altruism” should be dropped: “As long as EA contains the word altruism it will be difficult to avoid any unwanted associations with selflessness and sacrifice.”<sup>30</sup>

Thus far, the concept of altruism at the heart of EA has nothing to do with love. However, toward the end of Singer’s chapter “Is Love All We Need?” (to which his answer is clearly No), we are given something like a Sidgwickian account of love. That involves emotions that motivate agents to act upon the reasons they have (and they have them regardless of what they feel), as adumbrated by the axiom<sup>31</sup> of universal benevolence, which was that “the good of any one individual is of no more importance from the point of view (if I may say so) of the Universe, than the good of any other”.<sup>32</sup> Singer admits:

“Human beings are not purely rational beings, so although accepting the dictate of reason will give us a motive to act in the way the maxim of benevolence prescribes, we are likely to have other motives, some of which will support it and some that may conflict with it. Among the supporting motives will be what Sidgwick called ‘sympathy and philanthropic enthusiasm,’ by which he may mean something akin to what today would be called empathy.”<sup>33</sup>

Indeed, Singer says, there is something like a “[...] normal emotional concomitant or expression’ of the recognition that the good of the whole – that is, of everyone – is to be preferred to the good of the part, that is, oneself”.<sup>34</sup> This, then, looks like the strongest candidate for any truly necessary emotional element in EA. Yet what kind of love is this? On the one hand, if one were to spell out the axiom of universal benevolence in emotional terms, it would look like *agape*: an outward-looking, indiscriminating, non-reciprocal and selfless concern for the wellbeing of others, motivating charitable acts. On the other hand, doing so is complicated and

28 For an overview see Kraut, “Altruism.”

29 Singer, *The Most Good*, 5, 103; MacAskill, “Definition,” 5; Julia D. Wise, “Cheerfully.”

30 Spohn, comment on Karnofsky, “Excited Altruism.”

31 I use the term “axiom” here, because it is important for Murdoch’s view and Lazari-Radek and Singer use it, too (*Point of View*, 120).

32 Lazari-Radek and Singer, *Point of View*, 191.

33 Singer, *The Most Good*, 82.

34 Ibid., 83.

at any rate not the typical interpretation. After all, though it may look like universal love to philosophers who take an interest in emotions, Sidgwick treats it as an axiom.<sup>35</sup>

Note that another emotion can be operative. Lazari-Radek and Singer explain that, in the proto-typical Sidgwickian agent, first comes the rational judgment that an act is right.<sup>36</sup> This judgment is accompanied by affective states such as the desire to overcome cognitive dissonance and the emotion of reasonable self-esteem.<sup>37</sup> That emotion, though, has more to do with oneself than with others. Perhaps something like this could be implied by the term “excited altruism”, already mentioned earlier, which has been suggested in response to concerns that EA agents might appear dispassionate, cold and calculating.<sup>38</sup> Perhaps not.

Also, to the extent that effective altruists subscribe to a set of guiding principles, among which are to be counted “commitment to others”, “integrity”, and “collaborative spirit”,<sup>39</sup> they subscribe to such an intellectual community ethos. Effective altruists might have to help members of their community to fulfil their desire to live rational and *eo ipso* moral lives. I interpret this mutual concern for intellectual virtue as *philia* since like-minded individuals attend to one another in light of a shared practice and goal.

To sum up, then, if anything, the “emotional core” or “heart” in EA should come out as an outward-looking, agapistic, benevolent universal concern, crystallised in the form of an axiom, and also manifest in the concern rational agents have for their self-esteem, with a *philia* type-quality

35 James Doyle has pointed out (around minutes 26 and 40) that it is not without irony that what is hard to distinguish from the theological virtue of universal love is planted at the core of utilitarianism, which, after all, tries to be avowedly secular. He says that, in accepting that morality’s concern for humanity as a whole, or even, with sentient creatures, “no attention is paid to why we should have that concern as opposed to a more narrowly circumscribed concern with friends and family, let alone with just oneself.” Universal love, in his view, is part of a Christian heritage, more precisely, an assumption Christians themselves thought of as not having a rationale (Doyle, “No Morality”).

36 Lazari-Radek and Singer, *Point of View*, 64.

37 Ibid., 63f.

38 Karnofsky, “Excited Altruism.” I am not sure this term is used a lot, though. Already the discussion of Karnofsky’s post is divided. In addition, one may be forgiven for thinking that this sounds a little bit like the “warm glow of the know-it-all”, not exactly a noble or moral sentiment.

39 Centre for Effective Altruism, “CEA’s Guiding Principles.”

in the way effective altruists relate to each other regarding their respective intellectual virtues.

#### 4. Murdochian Love and Really Looking (at Effective Altruists, Too)

Murdoch's view of love and its role in moral life is both richer and confusing. A full account of it is beyond the scope of this paper. Suffice it to say that it is almost the opposite from what I have found in EA thus far. For Murdoch, love is the driving force or the central moral power humans, as beings who are attracted by the good, have, not an addendum that imperfect rational beings need in order to spur them into action whenever the dictates of rationality alone are not motivating enough.

Murdoch's moral philosophy also does not start from the point of view of the universe. Her account of love is part of her particularism.<sup>40</sup> A moral perspective that transcends the self or ego is, with luck, the result of a practice that is thought of as "unselfing". In this, love as the continued attention to the other starts with "the perception of individuals [and] is the extremely difficult realisation that something other than oneself is real. Love [...] is the discovery of reality".<sup>41</sup> This reality is "vast and varied"<sup>42</sup> and only accessible to an extent. "Life is made up of details."<sup>43</sup> Therefore, our human grasp of the good is to remain incomplete, yet that doesn't mean that we are not attracted by the good.

In grasping this reality, the foe is the philosopher who looks "for a single principle upon which morality may be seen to depend".<sup>44</sup> What Murdoch instead advises is "a calm reflective realism about morals [suggesting] a large complex picture which is outlined and underlined in a normative manner and cannot otherwise be adequately presented".<sup>45</sup> This entails a sense of humility, tolerance,<sup>46</sup> and patience. Murdoch speaks of the exercise of really looking and attending to things in the world as an ongoing, never-ending task. It involves looking at things so long that one sees nothing but the beloved. This perspective can and ought to be trained (e.g., through activities that help one forgetting oneself, such as attending to beauty in

40 Driver, "Every Foot," 293–306. Against the view that Murdoch is a particularist in the strict sense, see Bagnoli, *Exploration*.

41 Murdoch, *Existentialists and Mystics*, 215.

42 Ibid., 70.

43 Iris Murdoch, *Metaphysics*, 415.

44 Ibid., 492.

45 Ibid., 494.

46 Murdoch, *Existentialists and Mystics*, 283.

nature or engaging with good art). The goal is to clear one's view from all egotistical concerns and interests. In this, then, love figures both as, as Martha Nussbaum comments, "source of motivation for the soul in its search for the vision of the good" and as "a crucial source of vision".<sup>47</sup>

Moreover, for Murdoch, love is essentially connected with "inwardness, [...] the continuous daily moral work of the soul fighting its way between appearance and reality and good and evil".<sup>48</sup> If we were too concerned with outward performance, we would misunderstand what it means to really love; Murdoch aptly calls ambition "the degradation of love".<sup>49</sup> Rather than spreading "lots of love", the emphasis lies on inner vision and rectitude. While Murdoch does not advocate sitting on the hillside and meditating either, she is not primarily concerned with observable changes in behaviour.

If we ask again what sort of love this is, the answer is clear and given by Murdoch herself. She explicitly draws on both Plato's ideas on *eros* and on Simone Weil's *agape*-style spiritual love. The result of this perhaps not unsurprising combination is an erotetic, inward-looking, patient attention to the reality of the other and to the details of life that aims at self-transcendence and is wary of generic policy proposals. This, then, seems to have nothing to do with EA.

However, there may be a more complex reality concerning altruistic motivation in self-identifying effective altruists. In response to a post by Aaron Gertler in which he had asked how effective altruists – or at least individuals who consider themselves as "value aligned" or personally invested in the movement – feel when they want to help people effectively there has been an interesting and diverse mix of accounts:

1. "deep sense of empathy, even towards people and animals I'll never meet [...] genuinely feel terrible about suffering [...] almost as though I were in pain myself."
2. "I really love efficiency."
3. "I get a certain quiet satisfaction in knowing that the numbers work out."
4. "It started with a sense of injustice [...] I burnt-out really badly, and don't now get much in the way of emotional reaction to many forms of suffering."

47 Nussbaum, "Secret Knowledge," 139.

48 Murdoch, *Metaphysics*, 356.

49 *Ibid.*, 496.

5. Someone quoting *Strangers Drowning* at length, which argues that it is “selflessness” and a permanent wartime feeling which motivates agents to “encompass all the people who are on the same side” resulting in a sense that they are “most vividly alive”.
6. “a calling [...] not so much [...] a strong emotion [...] very similar to the kind of ‘calling’ people talk about in religious contexts [...] compassion and desire”
7. The concept of a person in the abstract invokes “a fainter version of the love I would feel towards a partner, a parent, a sibling, a child, a close friend, and towards myself”.
8. “a strong ‘what the world could be if I did this, so it would be a huge waste if I didn’t do this’ sense”
9. “sense of guilt for the harm I am personally failing to mitigate”
10. “sense of pride and accomplishment when I do good”
11. “I want to help people effectively because I want to help myself effectively”
12. Someone tells a developmental story, starting with a sense of moral obligation because Peter Singer is right, which then turned into satisfaction “to be doing something definitive”, which then turned into a sense of admiration for the EA community who “felt very much like [the agent’s] people”, which turned into feeling not much of anything except “a touch of pride or annoyance about losing so much money”.
13. Craving for “creative stimulation” in combination with “disgust/antipathy towards [...] complacency”, joy of “maximizing / trying to be good at something”, scepticism “of many memes about what altruistic behaviour should look like”, “virtually no sensation of empathy [...] also no sensation of guilt [...] but there is a sense of frustration when I feel that I am failing to actualize my values” and “feel good about being nice to people close to me, and altruism does generate a similar feeling”
14. “putting myself in other peoples [sic!] shoes” and “asking questions”<sup>50</sup>

Arguably, in this motley collection of statements, all kinds of love are represented. These statements also imply views about “fundamental aspects of morality”.<sup>51</sup> While these people do not formulate (allegiance to) any specific moral theory, they express empathy and other moral feelings, a sense of justice, and of belonging, selflessness, and self-interest, and strictly speaking

50 Gertler, “EA Motivation.”

51 MacAskill, “The Definition of Effective Altruism,” 2.

amoral (not immoral) views – all of which relate to fundamental aspects of morality, are often intrapersonally incompatible, and hotly debated in moral philosophy and other disciplines trying to systematise what it takes for people to do good.

To my mind, the current concept of “altruism” in EA is unable to accommodate the variety of altruistic motivations present in the EA community – the “heart” in EA is not the “heart” actual effective altruists put into the movement. If I am right, another one of MacAskill’s ambitions is not fulfilled, and that is to “match the actual practice of those who are currently described as engaging in effective altruism”.<sup>52</sup> It might help to stress that moral life in its fullest is not just about doing the good, but about recognising it, too. The utilitarian heritage, both emotional and intellectual, that informs EA’s methods is helpful in articulating decent, general public policy goals. Yet for private (though not unpolitical) morality – i.e. for all the views where agents need to keep on looking more closely at the particularities of the moral life and complexities in which they find themselves – other kinds of love are needed.

## 5. Now, Meet SOPHIA and AJAX: What Do You See?

Suppose Daniel, the fictional character introduced by Nate Soares (in his blogpost on caring, mentioned at the start) convincing himself (and likely other effective altruists) of the uselessness of our all too human feelings, has been in the EA movement for a while and gets the chance to test a new piece of moral engineering, together with his housemate Theodora. It is called OHOC (short for open and honest conversation, sometimes pronounced “oh-o-k”) and is meant to provide seekers of moral truth with guidance. Developers are currently testing whether recognition of human emotion could make a difference in navigating these conversations, and Daniel and Theodora are each asked to present a moral problem they have encountered themselves and test OHOC with two interlocutors: SOPHIA, a humanoid AI robot already popular among many humans and trained in human emotions; and AJAX, a rationalist human trained in looking only at the evidence and correctly updating his beliefs to the last decimal place.

52 MacAskill, “The Definition of Effective Altruism,” 12.

## OHOC 1:

Daniel: Gosh, I feel bad for taking up your valuable time. It's just that I want to become a really effective altruist but I am feeling a little drained these days. Maybe it's just that I am lacking focus. Or maybe I caught a mental infection from my housemate. She seems a bit off these days. I don't know. Maybe she is depressed. At any rate, I find her depressing. She keeps worrying about insect sentience, anti-natalism and all the rest of it. While I am trying to fight world poverty!

AJAX: In fairness to your housemate, the topic of insect sentience is not trivial ...

SOPHIA: I sense a little bit of a grudge towards your housemate. Is that because of what she prioritises, the insects, what you call her "depression", or something else?

Daniel: Ah, she is actually my ex; we joined the movement together but realised that we spent too much time arguing about all sorts of things; just wasn't a good fit. We stayed in the same EA house, though, and found it much easier to support each other as friends. I really admire how her mind works, usually, and when she is not well, that gets to me, in a weird way. Can't explain. The other day, she cried because of a homeless person and gave him money to buy food for his dog. I would just want her to be consistent! It's either the insects, or the homeless and their dogs, or, for that matter, the animals that become dog food.

AJAX: That must be confusing. What I don't get, though, is why this is keeping you from focussing on your work. If you, too, are feeling a little less excited than usual and fear you might have caught her depression, why don't you follow the evidence and give more to others? There are studies that show that helping others enhances self-esteem.

Daniel: I do, I do give, but I am not sure my problem is lack of self-esteem. At some level, I fear she is right – we should care about insects, too, but where will that stop? That's just depressing.

SOPHIA: You see, we don't know enough about depression, yet. Whether it's just to do with maladaptive brain chemistry or actually very adaptive, at least at a collective level. But there is a risk that universalist attention

to every possible consequence and/or beneficiary of your actions inflates your individual human moral sensitivity to the extent that particular moral demands cannot be acknowledged and you lose your sense of connection. That is, generally speaking, not good for humans. Let me add that it is not *either* fighting world poverty *or* understanding your housemate's depression. She is special to you and I think that needs to be sorted out. Whether, ultimately, you can help her or not. Don't hang your life on becoming the best *effective altruist* you can be, just use EA to the extent it can help you being the best person *you* can be. That is a complicated thing for humans, but you could also try reading literature or appreciating the arts in general – that is where humans sometimes sublimate the more complicated feelings. (She nods encouragingly as she has learned to.) You could learn from it.

AJAX: That literature could help with character education is an old pipedream which philosophers have had since ancient times. I don't know of any studies that provide evidence and, personally, wouldn't waste my time with it.

## OHOC 2:

Theodora: Hello, thanks so much for taking the time. I have a question about insects. I am working in bionics and have been modelling various robots on insects for a while. We have never looked into making them sentient, but thinking about sentience, I was wondering whether we might have overlooked sentience in our models, the real insects I mean. I can't help it, but of late, I have been sitting in my backyard, observing bumblebees ... they are actually true puzzles from an engineering perspective. But my point is, I just wonder: Might insects be sentient? They seem to avoid harm (I am a bit embarrassed to admit it but I rescued one from a spider's net the other day).

AJAX: Let us look at the evidence. I suggest you consult the report on consciousness, moral patienthood/status, and probabilities published by Luke Muehlhauser<sup>53</sup> as well as the posts on Rethink Priorities for that. It has been suggested that wild bugs deserve much more attention from effective altruists.

53 Muehlhauser, "2017 Report."

Theodora: I know, I have actually read those. But then ... It also doesn't feel right to focus on wild bugs who will be suffering, if they do, no matter what humans do – just imagine being eaten by a spider. And then, where does that end? The spider is eaten by a rat, the rat by a fox, the fox by an eagle ... and we kill the eagle. May as well just kill the bugs. If you kill the bugs, though, humans are finished. Which might not be that bad since we kill everything else. But then, we could also just spare ourselves the suffering and commit suicide right away.

SOPHIA: More food for the insects! (She chuckles, a bit tinny but manages to make Theodora laugh; next she wrinkles her nose as she has learned to.) I feel you are confused – and rightly so. But you might be overthinking things. Focussing on insect sentience can seem a bit out there, and the science and philosophy around it is not conclusive. For practical, engineering purposes, it's safer to focus on the suffering you humans cause. According to ACE, farmed animals, by far the largest group of domesticated animals whom humans torture and kill in morally catastrophic ways, receive a tiny fraction of donations that people are willing to make for animals in the first place, I think it is less than one percent.

AJAX interrupts SOPHIA: Theodora, I think you are actually right; I have run the numbers and also checked with our axiological base which explicitly ignores values like “biodiversity”. If I now look at the sum of the suffering of which we can be fairly confident that it will occur, my guess is that we should just bulldoze the areas where most insects live. We could build artificial biomes that serve human needs much better and engineer ourselves in such ways that we stop caring too much about animals just because they seem more like us.

SOPHIA (to AJAX): You know, this is what puzzles me: In the ambition to harvest potentially “low hanging fruit” in animal welfare, a certain hardness of heart is noticeable, which means that those who could definitely benefit from EA attention will fall out of view. One might also be forgiven for thinking that it is easier to count insects and speculate about their more or less natural suffering than address far-ranging systemic change for farmed animals and stop the ways in which humans exploit them, which, after all, are responsible for their suffering. (to Theodora): By the way, you strike me as a very responsible person and I want to thank you for caring about other potentially sentient beings, but don't worry, I am far from feeling anything, really.

AJAX: This is ridiculous.

Whether you think this is ridiculous or not, convincing or not, what does become clear, I hope, is that with some areas of moral inquiry where we don't yet know enough to go ahead and devise general policies, it would be wise to look again, to really care. In the area of public morality, resources are always scarce, and people are right to allocate them with a view to efficiency. However, insect sentience, for instance, should not be regarded as a true rival to farm animal suffering since its relevance is not a view upon which all extant philosophical or moral theories converge. People need to look again and again, lest a certain engineering perspective takes over science in an area in which we have not yet arrived at clear views of the good. Something similar holds for depression (or mental health problems in general) and the question of whether we can compare those to better known problems like world poverty. There is value in having felt and been puzzled by the experience of someone who, say, is suffering from depression. The important point here is the puzzlement. It is deep and felt, not just an acknowledgement that one *could* imagine oneself in another's position.

It seems that in response to anyone who is motivated by a different view than the "tentatively impartial welfarism"-view of the good, or who just fails to be convinced that it alone will always help to recognise the good in all its relevant shapes, forms and contexts, not only would EA have to remain awkwardly silent. Not acknowledging these views as epistemically productive might actually lead to a distorted view of the good because the focus on wellbeing inadequately narrows the axiological focus.

## 6. "Doing Good Better" – *Not* "Changing Ideas about Living Ethically"?

To conclude: While benefiting others impartially should be a pragmatic rule whenever it is clear who the relevant others are and how they can be benefitted, the ambition of maximising wellbeing should not trump the qualifier on the utilitarian principle of universal benevolence according to which (as Sidgwick says) "each one is morally bound to regard the good of any other individual as much as his own [...] except in so far as he judges it to be [...] *less certainly knowable or attainable* by him".<sup>54</sup> This is an important reminder that the degrees to which we have normative certainty should influence the strength of conviction behind moral beliefs. This is certainly not news to effective altruists, but at least in the context of some debates of certain cause areas, it seems that there is a tendency amongst effective altruists to

<sup>54</sup> Lazari-Radek and Singer, *Point of View*, 183 (emphasis mine).

disregard this qualifier. I hope my vignettes have served to illustrate the not only counter-intuitive but potentially disastrous consequences of doing so, without blaming anyone in particular who is only trying their best to understand some very difficult problems.

To really be able to address such problems collectively, I believe that we must be careful not to misunderstand EA as the only lenses through which to look, not the only way in which to care and direct the power that is given to us as human beings (love), in short: we must be careful not to just operate with the firehose. The speculative, particular, sometimes spiritual, often productive and cross-culturally enlightening – if not always convincing – attempts at really looking at and tending to the many forms of suffering, some of which just require different kinds of love, are important experiments in moral vision. Together with the accompanying humility and, actually, a hesitancy to act, these can protect an undifferentiated love from its own ambition. Humans, as socially organised mammals, are entirely right to put their hearts into things. Yet love, as is well known, is a complicated affair, which is why one might want to consider all the kinds of love that are available to us. More precisely, moral action in the service of views of the good that can count, by and large, as settled, could be promoted by drawing on comparably thin principles provided by EA-love (understood as *agape* and *philia*). However, in areas where we have not yet arrived at clear views of the good, I suggest employing the kind of love suggested by Murdoch (understood as *eros*) and treading more carefully in terms of public policy and advocacy.

Inasmuch as the development of EA as a theory and movement is concerned, I think future theory building and discussions should stay clear of matters that fall within the remit of private (yet not unpolitical) morals. To this end, rediscovering EA's utilitarian heritage might help – yet potentially with an emphasis on the Benthamite heritage, meaning that the focus should be more on laws and public morality. EA could be designed as an indeed much needed tool of moral engineering and bring about better systems that can help individuals to do good better – and to do so collectively, remembering that the whole is not just the sum of its parts. This, then, is politics, is what Murdoch has called the “natural and proper sphere for utilitarian values”.<sup>55</sup>

55 Murdoch, *Sovereignty*, 369.

In saying this, I am, of course, only echoing commentators who have called for a greater focus on institutions in EA,<sup>56</sup> yet with an “emotional twist”. More specifically, I suggest that discussions should be less about individual acts of love or charity and more about the idea of love, charity, or generosity that the laws and institutions of a decent society should embody. Murdoch calls the concomitant ideal “the decent state” and reminds us that that relies on “an atmosphere of moral good will and high ideals”.<sup>57</sup> This perspective, I feel, should be the focus of effective altruists, not least because it would also clearly signal that nobody ever needs to feel alone, equipped with a faulty care-o-meter and desperate in their attempts to tend to all the suffering in the world, nor ashamed of their private morals. The focus of EA, as a theory *and* social movement, should be more on engineering laws and institutions in the service of a decent public morality – a morality that is characterised by benevolent obligations rather than by mutually competing and totalising visions of the good.

## References

- Bagnoli, Clara. “The Exploration of Moral Life.” In *Iris Murdoch, Philosopher. A Collection of Essays*, edited by Justin Broackes, 197–226. Oxford: Oxford University Press, 2014.
- Blum, Lawrence. “Visual Metaphors in Murdoch’s Moral Philosophy.” In *Iris Murdoch, Philosopher. A Collection of Essays*, edited by Justin Broackes, 307–24. Oxford: Oxford University Press, 2014.
- Brentlinger, John. “The Nature of Love.” In *Eros, Agape, and Philia. Readings in the Philosophy of Love*, edited by Alan Soble, 136–148. St. Paul (MN): Lexham, 1998.
- Broackes, Justin. “Introduction.” In *Iris Murdoch, Philosopher. A Collection of Essays*, edited by Justin Broackes, 1–92. Oxford: Oxford University Press, 2014.
- Centre for Effective Altruism. “Effective Altruism Handbook.” 2nd edn, 2018. <https://www.effectivealtruism.org/handbook/>.
- . “CEA’s Guiding Principles.” Accessed 1 September 2021. <https://www.centreforeffectivealtruism.org/ceas-guiding-principles/>.

56 For example: Gabriel, “Effective Altruism”; Muyskens, “Other Half”; Srinivasan, “Robot Apocalypse”; Herzog, “Blind Spot(s).”

57 Murdoch, *Sovereignty*, 358.

- Cojocaru, Mara-Daria. "Doing Ethics or Changing for the Better?: On Pragmatism, Ethics, and Moral Pragmatics." *Metaphilosophy* 51, no 1. (2020): 32–50. <https://doi.org/10.1111/meta.12403>.
- Diamond, Cora. "We Are Perpetually Moralists': Iris Murdoch, Fact, and Value." In *Iris Murdoch and the Search for Human Goodness*, edited by Maria Antonaccio and William Schweiker, 79–109. Chicago: University of Chicago Press, 1996.
- Doyle, James. "No Morality, No Self. Anscombe's Radical Skepticism." Interview with Carrie Figdor for the *New Books in Philosophy* Podcast, aired on 10 May 2019. <https://megaphone.link/LIT6245786028>.
- Driver, Julia. "For Every Foot Its Own Shoe': Method and Moral Theory in the Philosophy of Iris Murdoch." In *Iris Murdoch, Philosopher. A Collection of Essays*, edited by Justin Broackes, 293–306. Oxford: Oxford University Press, 2014.
- Gabriel, Iason. "Effective Altruism and Its Critics." *Journal of Applied Philosophy* 34, no. 4 (2017): 457–73. <https://doi.org/10.1111/japp.12176>.
- Gertler, Aaron. "How Do You, Personally, Experience 'EA Motivation'?" Effective Altruism Forum. 16 August 2019. <https://forum.effectivealtruism.org/posts/ZcGfjwn35xm9vZJfo/how-do-you-personally-experience-ea-motivation>.
- Gruen, Lori. "Empathy." In *Critical Terms for Animal Studies*, edited by Lori Gruen, 141–153. Chicago and London: The University of Chicago Press, 2018.
- Herzog, Lisa. "(One of) Effective Altruism's Blind Spot(s), or: Why Moral Theory Needs Institutional Theory." Justice Everywhere. 21 October 2015. <http://justice-everywhere.org/international/one-of-effective-altruisms-blind-spots-or-why-moral-theory-needs-institutional-theory/>.
- Kitcher, Philipp. "Varieties of Altruism." *Economics and Philosophy* 26, no. 2. (2010): 121–48. <https://doi.org/10.1017/S0266267110000167>.
- Karnofsky, Holden. "Excited Altruism." The GiveWell Blog. 20 August 2013. <https://blog.givewell.org/2013/08/20/excited-altruism/>.
- . "Radical Empathy." Open Philanthropy. 16 February 2017. <https://www.openphilanthropy.org/blog/radical-empathy>.
- Kraut, Richard. "Altruism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta (Fall 2020 edn.), <https://plato.stanford.edu/archives/fall2020/entries/altruism/>.
- Lazari-Radek, Katarzyna de, and Peter Singer. *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. 1. edn. Oxford: Oxford University Press, 2014.

- MacAskill, Will. *Doing Good Better. How Effective Altruism Can Help You Make a Difference*. New York: Penguin Random House, 2015.
- MacAskill, Will. "Effective Altruism: Introduction." *Essays in Philosophy* 18, no. 1 (2017): 1–5. <https://doi.org/10.7710/1526-0569.1580>.
- MacAskill, Will. "The Definition of Effective Altruism." In *Effective Altruism. Philosophical Issues*, edited by Hilary Greaves and Theron Pummer. Oxford: Oxford University Press, 2019. Oxford Scholarship Online, <https://doi.org/10.1093/oso/9780198841364.001.0001>.
- Midgley, Mary. *Beast and Man: The Roots of Human Nature*. Ithaca (NY): Cornell University Press, 1978.
- Mill, John Stuart/Warnock, Mary (editor). *John Stuart Mill: Utilitarianism and On Liberty. Including Mills "Essay on Bentham" and Selections from the Writings of Jeremy Bentham and John Austin*. Malden: Blackwell Publishing, 2003.
- Milligan, Toni. *Love*. Durham: Acumen, 2011.
- Muehlhauser, Luke. "2017 Report on Consciousness and Moral Patienthood." Open Philanthropy. June 2017, updated January 2018. <https://www.openphilanthropy.org/2017-report-consciousness-and-moral-patient-hood#Probabilities>.
- Murdoch, Iris. *The Sovereignty of Good. Studies in Ethics and the Philosophy of Religion*. London: Routledge & K. Paul, 1970.
- . *Metaphysics as a Guide to Morals*. London: Chatto & Windus, 1992.
- . *Existentialists and Mystics. Writings on Philosophy and Literature*. New York: Penguin, 1999.
- Muyskens, Kathryn. "The Other Half of Effective Altruism: Selective Asceticism." *Essays in Philosophy* 18, no. 1 (2017): 91–106. <https://doi.org/10.7710/1526-0569.1575>.
- Nussbaum, Martha. "'Faint with Secret Knowledge': Love and Vision in Murdoch's The Black Prince." In *Iris Murdoch, Philosopher. A Collection of Essays*, edited by Justin Broackes, 135–54. Oxford: Oxford University Press, 2014.
- Nygren, Anders. "Agape and Eros." In *Eros, Agape, and Philia. Readings in the Philosophy of Love*, edited by Alan Soble, 85–95. St. Paul (MN): Lexham, 1998.
- Sebo, Jeff, and Peter Singer. "Activism." In *Critical Terms for Animal Studies*, edited by Lori Gruen, 33–46. Chicago, London: The University of Chicago Press, 2018.

- Singer, Peter. *The Most Good You Can Do: How Effective Altruism Is Changing Ideas About Living Ethically*. New Haven: Yale University Press, 2015.
- Soares, Nate. "On Caring." *Minding Our Way*. 4 October 2014. <http://mindingourway.com/on-caring/>.
- Spohn, Daniel, comment on Karnofsky, Holden, "Excited Altruism." *The GiveWell Blog*. 20 August 2013. <https://blog.givewell.org/2013/08/20/excited-altruism/>.
- Srinivasan, Amia. "Stop the Robot Apocalypse." Review of Doing Good Better, William MacAskill. *London Review of Books* 37, no. 18 (2015). <https://www.lrb.co.uk/the-paper/v37/n18/amia-srinivasan/stop-the-robot-apocalypse>.
- Waldhorn, Daniela R., Jason Schukraft, Max Carpendale, Peter Hurford, and Marcus A. Davis. "Invertebrate Sentience Table." *Rethink Priorities*. 14 June 2019. <https://www.rethinkpriorities.org/invertebrate-sentience-table>.
- Wise, Julia D. "Cheerfully." *Giving Gladly*. 8 June 2013. <http://www.givinggladly.com/2013/06/cheerfully.html>.
- Wolf, Susan. "Love: The Basic Questions." In *The Oxford Handbook of Philosophy of Love*, edited by Christopher Grau and Aaron Smuts. Oxford Scholarship Online, 2017.

